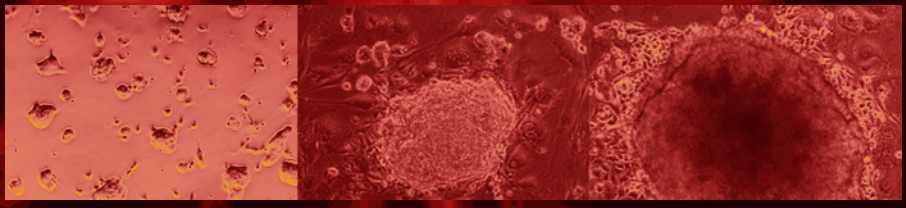
METHODS IN MOLECULAR BIOLOGY™ 360

Target Discovery and Validation Reviews and Protocols

Volume 1
Emerging Strategies for Targets and Biomarker Discovery

Edited by
Mouldy Sioud



HUMANA PRESS

Target Discovery and Validation Reviews and Protocols

Volume 1, Emerging Strategies for Targets and Biomarker Discovery

METHODS IN MOLECULAR BIOLOGY™

John M. Walker, SERIES EDITOR

386. **Peptide Characterization and Application Protocols**, edited by *Gregg B. Fields, 2007*
385. **Microchip-Based Assay Systems:** *Methods and Applications,* edited by *Pierre N. Floriano, 2007*
384. **Capillary Electrophoresis:** *Methods and Protocols,* edited by *Philippe Schmitt-Kopplin, 2007*
383. **Cancer Genomics and Proteomics:** *Methods and Protocols,* edited by *Paul B. Fisher, 2007*
382. **Microarrays, Second Edition:** *Volume 2, Applications and Data Analysis,* edited by *Jang B. Rampal, 2007*
381. **Microarrays, Second Edition:** *Volume 1, Synthesis Methods,* edited by *Jang B. Rampal, 2007*
380. **Immunological Tolerance:** *Methods and Protocols,* edited by *Paul J. Fairchild, 2007*
379. **Glycovirology Protocols**, edited by *Richard J. Sugrue, 2007*
378. **Monoclonal Antibodies:** *Methods and Protocols,* edited by *Maher Albitar, 2007*
377. **Microarray Data Analysis:** *Methods and Applications,* edited by *Michael J. Korenberg, 2007*
376. **Linkage Disequilibrium and Association Mapping:** *Analysis and Application,* edited by *Andrew R. Collins, 2007*
375. **In Vitro Transcription and Translation Protocols:** *Second Edition,* edited by *Guido Grandi, 2007*
374. **Biological Applications of Quantum Dots**, edited by *Marcel Bruchez and Charles Z. Hotz, 2007*
373. **Pyrosequencing® Protocols**, edited by *Sharon Marsh, 2007*
372. **Mitochondrial Genomics and Proteomics Protocols**, edited by *Dario Leister and Johannes Herrmann, 2007*
371. **Biological Aging:** *Methods and Protocols,* edited by *Trygve O. Tollefsbol, 2007*
370. **Adhesion Protein Protocols**, *Second Edition,* edited by *Amanda S. Coutts, 2007*
369. **Electron Microscopy***: Methods and Protocols, Second Edition*, edited by *John Kuo, 2007*
368. **Cryopreservation and Freeze-Drying Protocols**, *Second Edition*, edited by *John G. Day and Glyn Stacey, 2007*
367. **Mass Spectrometry Data Analysis in Proteomics**, edited by *Rune Matthiesen, 2007*
366. **Cardiac Gene Expression:** *Methods and Protocols,* edited by *Jun Zhang and Gregg Rokosh, 2007*
365. **Protein Phosphatase Protocols:** edited by *Greg Moorhead, 2007*
364. **Macromolecular Crystallography Protocols:** *Volume 2, Structure Determination,* edited by *Sylvie Doublié, 2007*
363. **Macromolecular Crystallography Protocols:** *Volume 1, Preparation and Crystallization of Macromolecules,* edited by *Sylvie Doublié, 2007*
362. **Circadian Rhythms:** *Methods and Protocols,* edited by *Ezio Rosato, 2007*
361. **Target Discovery and Validation Reviews and Protocols:** *Emerging Molecular Targets and Treatment Options, Volume 2,* edited by *Mouldy Sioud, 2007*
360. **Target Discovery and Validation Reviews and Protocols:** *Emerging Strategies for Targets and Biomarker Discovery, Volume 1,* edited by *Mouldy Sioud, 2007*
359. **Quantitative Proteomics by Mass Spectrometry**, edited by *Salvatore Sechi, 2007*
358. **Metabolomics:** *Methods and Protocols,* edited by *Wolfram Weckwerth, 2007*
357. **Cardiovascular Proteomics:** *Methods and Protocols,* edited by *Fernando Vivanco, 2006*
356. **High-Content Screening:** *A Powerful Approach to Systems Cell Biology and Drug Discovery,* edited by *D. Lansing Taylor, Jeffrey Haskins, and Ken Guiliano, and 2007*
355. **Plant Proteomics:** *Methods and Protocols,* edited by *Hervé Thiellement, Michel Zivy, Catherine Damerval, and Valerie Mechin, 2006*
354. **Plant–Pathogen Interactions:** *Methods and Protocols,* edited by *Pamela C. Ronald, 2006*
353. **Protocols for Nucleic Acid Analysis by Nonradioactive Probes,** *Second Edition,* edited by *Elena Hilario and John Mackay, 2006*
352 **Protein Engineering Protocols**, edited by *Kristian Müller and Katja Arndt, 2006*
351 ***C. elegans:*** *Methods and Applications*, edited by *Kevin Strange, 2006*
350 **Protein Folding Protocols,** edited by *Yawen Bai and Ruth Nussinov 2007*
349 **YAC Protocols,** *Second Edition*, edited by *Alasdair MacKenzie, 2006*
348 **Nuclear Transfer Protocols:** *Cell Reprogramming and Transgenesis*, edited by *Paul J. Verma and Alan Trounson, 2006*
347 **Glycobiology Protocols**, edited by *Inka Brockhausen, 2006*
346 ***Dictyostelium discoideum* Protocols**, edited by *Ludwig Eichinger and Francisco Rivero, 2006*
345 **Diagnostic Bacteriology Protocols**, *Second Edition*, edited by *Louise O'Connor, 2006*
344 **Agrobacterium Protocols**, *Second Edition: Volume 2*, edited by *Kan Wang, 2006*
343 **Agrobacterium Protocols**, *Second Edition: Volume 1*, edited by *Kan Wang, 2006*
342 **MicroRNA Protocols**, edited by *Shao-Yao Ying, 2006*
341 **Cell–Cell Interactions:** *Methods and Protocols*, edited by *Sean P. Colgan, 2006*
340 **Protein Design:** *Methods and Applications*, edited by *Raphael Guerois and Manuela López de la Paz, 2006*
339 **Microchip Capillary Electrophoresis:** *Methods and Protocols*, edited by *Charles S. Henry, 2006*

METHODS IN MOLECULAR BIOLOGY™

Target Discovery and Validation Reviews and Protocols

Volume 1

Emerging Strategies for Targets and Biomarker Discovery

Edited by

Mouldy Sioud

Department of Immunology,
Institute for Cancer Research,
The Norwegian Radium Hospital,
University of Oslo, Oslo, Norway

HUMANA PRESS TOTOWA, NEW JERSEY

999 Riverview Drive, Suite 208
Totowa, New Jersey 07512

www.humanapress.com

This publication is printed on acid-free paper. ♾
ANSI Z39.48-1984 (American Standards Institute) Permanence of Paper for Printed Library Materials.

Cover design by Patricia F. Cleary
Cover illustrations:*Foreground:* HM-1 ES cells: plated, ready for PCR, and highly differentiated (left to right; Chapter 11, Fig. 5; *see* complete caption on p. 27). *Background:* Overexpression of green fluorescent protein-tagged centrosome/spindle pole-associated protein in HEK293T cells (Chapter 1, Fig. 1; *see* complete caption on p. 3).

For additional copies, pricing for bulk purchases, and/or information about other Humana titles, contact Humana at the above address or at any of the following numbers: Tel.: 973-256-1699; Fax: 973-256-8341; E-mail: orders@humanapr.com; or visit our Website: www.humanapress.com

Printed in the United States of America. 10 9 8 7 6 5 4 3 2 1

eISBN 1-59745-165-7

ISSN 1064-3745

Library of Congress Cataloging-in-Publication Data
Target discovery and validation : reviews and protocols / edited by
Mouldy Sioud.
p. ; cm. -- (Methods in molecular biology, ISSN 1064-3745 ; 360-361)
Includes bibliographical references and index.
ISBN-13: 978-1-58829-656-6 (v. 1 : alk. paper)
ISBN-10: 1-58829-656-3 (v. 1 : alk. paper)
ISBN-13: 978-1-58829-890-4 (v. 2 : alk. paper)
ISBN-10: 1-58829-890-6 (v. 2 : alk. paper)
1. Tumor markers. 2. Biochemical markers--Therapeutic use. 3.
DNA microarrays. 4. Drug targeting. 5. Drugs--Design. 6.
Molecular pharmacology. I. Sioud, Mouldy. II. Series: Methods in
molecular biology (Clifton, N.J.) ; v. 360-361.
[DNLM: 1. Antineoplastic Agents. 2. Drug Design. 3. Tumor
Markers, Biological. W1 ME9616J v.360-361 2007 / QV 269 T185 2007]
RC270.3.T84T37 2007
616.99'4075--dc22

2006020032

Preface

Target discovery is a field that has existed for several years but is so vibrant today because of the recent progress in our understanding of the molecular mechanisms of many human diseases and the technical advances in target identification and validation. More sophisticated gene profiling technologies, such as DNA microarrays and serial analysis of gene expression, permit rapid identification of lead targets. Moreover, analysis of gene networks in living organisms allows the identification of target genes that operate in defined physiological pathways. With the sequencing of several genomes completed and the rapidly growing gene expression databases, there is now greater impetus than ever before for *in silico* discovery of therapeutic targets. Also, recent advances in genetic technologies have increased our ability to generate mouse models for human diseases. The implications of these genetically modified animals in drug development are several, including identification of new drug targets, predicting efficacy, and uncovering possible side effects. Together, these recent technical advances should allow researchers to make the most informed choice early and advance the chosen targets toward clinical studies.

Regarding cancers, any difference between a cancer and a normal cell could potentially be exploited as a therapeutic target. The hope is that drugs targeting specific constituents or pathways in cancer cells will provide more effective therapy, either alone or in combination with other currently used anticancer drugs. In addition to drug targets, identifying new target antigens remains as much of a challenge as improving tumor vaccines already in the clinic. New techniques such as SEREX, phage display, proteomics, and reverse immunology have not only yielded a significant number of new tumor antigens, but they have influenced our understanding of the interaction between the immune system and tumor cells. New adjuvants to improve vaccine potency are also being discovered.

Validating potential drug targets is one of the most critical steps in drug discovery. At present, the currently available target identification technologies appear to yield too many targets. Therefore, programs need to focus on defining the targets that are specific and "druggable" with small inhibitors. While true target validation comes only when a selective inhibitor for the chosen target is tested in patients and exhibits efficacy in the appropriate human disease, the appropriate cell culture and animal studies prior to the trial are not only required but should provide crucial information regarding the validity of the chosen tar-

get. Approaches to target validation might involve the use of small interfering RNAs, antisense oligonucleotides, antibodies, or engineered transgenic or knockout mice. Notably, several human diseases such as cancers and autoimmune disorders can be modeled in the mouse, making it an ideal tool to accelerate the validation process of new agents and the assessment of risk and toxicity.

Because of the rapid technical progress in the target discovery and validation field, we decided to divide *Target Discovery and Validation Reviews and Protocols* into two separates books, each describing specific topics. Volume 1, *Emerging Strategies for Targets and Biomarker Discovery*, provides the recent technological advances for the identification of molecular drug targets, biomarkers, and tumor antigens. In addition, it describes the role of knockout mice in functional genomics and target validation, as well as the clinical impact of gene expression profiling on oncology diagnosis, prognosis, and treatment. Volume 2, *Emerging Molecular Targets and Treatment Options*, concentrates on target validation strategies and on efforts to bring agents against specific targets closer to clinical application. We are presenting very up-to-date overviews and protocols by experts in the field. The covered topics will be of interest to researchers, pharmaceutical companies, and clinicians interested in developing or using new therapies. In addition, *Target Discovery and Validation Reviews and Protocols* will assist academic clinicians and students of biology, medicine, or pharmacy to appreciate the progress of the past few years in our understanding of cancer pathogenesis, cell signaling, gene profiling approaches, tumor immunology, tumor immunoediting, and immune tolerance.

I would like to thank the authors for their contributions, Anne Dybwad for critical reading of the manuscripts, and all those involved in the production of *Target Discovery and Validation Reviews and Protocols.*

Mouldy Sioud

Contents

Contributors

Robert M. Beaty • *Department of Pathology, The Sol Goldman Pancreatic Cancer Research Center, Johns Hopkins University School of Medicine, Baltimore, MD*

Lina Cekaite • *Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, University of Oslo, Oslo, Norway*

Holger Erfle • *Cell Biology and Cell Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany*

Mads Gronborg • *McKusick–Nathans Institute of Genetic Medicine; Departments of Biochemistry and Molecular Biology, University of Southern Denmark, Odense, Denmark*

Guttorm Haraldsen • *Institute of Pathology, Rikshospitalet University Hospital, Oslo, Norway*

Louis-Marie Houdebine • *Biologie du Developpement et Reproduction, Institut National de la Recherche Agronomique, Paris, France*

Eivind Hovig • *Department of Tumor Biology, Institute for Cancer Research, The Norwegian Radium Hospital, University of Oslo, Oslo, Norway*

Seiya Imoto • *Human Genome Centre, Institute of Medical Science, University of Tokyo, Tokyo, Japan*

Per Ole Iversen • *Department of Nutrition, Institute of Basic Medical Sciences, University of Oslo, Oslo, Norway*

Dirk Jäger • *Medizinische Onkologie, Nationales Centrum für Tumorerkrankungen, Universitätsklinikum Heidelberg, Heidelberg, Germany*

Axel M. Küchler • *Institute of Pathology, Rikshospitalet University Hospital, Oslo, Norway*

François Le Naour • *INSERM V602, Hôpital Paul Brousse, Villejuif, France*

Thomas M. Magin • *Institut für Physiologische Chemie, Abteilung für Zellbiochemie, Bonner Forum Biomedizin and LIMES, Universitätsklinikum Bonn, Bonn, Germany*

Anirban Maitra • *Departments of Pathology and Oncology, The Sol Goldman Pancreatic Cancer Research Center and McKusick–Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD*

SAHOHIME MATSUMOTO • *Department of Chemistry and Biotechnology, School of Engineering, Tokyo, Japan*
MAKOTO MIYAGISHI • *Gene Function Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan*
SATORU MIYANO • *Human Genome Centre, Institute of Medical Science, University of Tokyo, Tokyo, Japan*
RAMASWAMY NARAYANAN • *Center for Molecular Biology and Biotechnology and Department of Biology, Florida Atlantic University, Boca Raton, FL*
RAINER PEPPERKOK • *Cell Biology and Cell Biophysics Unit, European Molecular Biology Laboratory, Heidelberg, Germany*
JONATHAN R. POLLACK • *Departments of Pathology and Oncology, The Sol Goldman Pancreatic Cancer Research Center and McKusick–Nathans Institute of Genetic Medicine, Johns Hopkins University School of Medicine, Baltimore, MD*
ØYSTEIN RØSOK • *Department of Immunology, Institute for Cancer Research, The Norwegian Radium Hospital, University of Oslo, Oslo, Norway*
MASAYUKI SANO • *Gene Function Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba, Japan*
CHRISTOPHER J. SAVOIE • *Gene Networks International, Tokyo, Japan*
MOULDY SIOUD • *Department of Immunology, Institute for Cancer Research, The Norwegian Radium Hospital, University of Oslo, Oslo, Norway*
DAG K. SKOVSETH • *Institute of Pathology, Rikshospitalet University Hospital, Oslo, Norway*
GORAN SÖHL • *Institut für Physiologische Chemie, Abteilung für Zellbiochemie, Bonner Forum Biomedizin and LIMES, Universitätsklinikum Bonn, Bonn, Germany*
DAG R. SØRENSEN • *Centre of Comparative Medicine, Rikshospitalet University Hospital, Oslo, Norway*
THERESE SØRLIE • *Department of Genetics, Institute for Cancer Research, The Norwegian Radium Hospital, University of Oslo, Oslo, Norway*
KAZUNARI TAIRA • *Department of Chemistry and Biotechnology, School of Engineering, Tokyo, Japan*
YOSHINORI TAMADA • *Institute of Statistical Mathematics, Tokyo, Japan*
PREETHI VIJAYARAJ • *Institut für Physiologische Chemie, Abteilung für Zellbiochemie, Bonner Forum Biomedizin and LIMES, Universitätsklinikum Bonn, Bonn, Germany*

Contents of Volume 2:

Emerging Molecular Targets and Treatment Options

1

Main Approaches to Target Discovery and Validation

Mouldy Sioud

Summary

The identification and validation of disease-causing target genes is an essential first step in drug discovery and development. Genomics and proteomics technologies have already begun to uncover novel functional pathways and therapeutic targets in several human diseases such as cancers and autoimmunity. Also, bioinformatics approaches have highlighted several key targets and functional networks. In contrast to gene-profiling approaches, phenotype-oriented target identification allows direct link between the genetic alterations and a disease phenotype. Therefore, identified genes are more likely to be a cause rather than a consequence of the disease. Once a gene target or a mechanistic pathway is identified, the next step is to demonstrate that it does play a critical role in disease initiation, perpetuation, or both. A range of strategies exists for modulating gene expression in vitro and in vivo. These strategies include the use of antibodies, negative dominant controls, antisense oligonucleotides, ribozymes, and small-interfering RNAs. In contrast to in vitro assays, mouse reverse genetics such as knockout phenotypes has become a powerful approach for deciphering gene function and target validation in the context of mammalian physiology. In addition to disease-causing genes, the identification of antigens that stimulate both arms of the immune system is the major goal for effective vaccine development. The hope is that target discovery and validation processes will concurrently identify and validate therapeutic targets for drug intervention in human diseases.

Key Words: Bioinformatics; genomics; microarray; protemics; target discovery; target validation; transgenic animals; tumor antigens.

1. Introduction

1.1. Target Discovery

1.1.1. Genomics and Related Technologies

During the last decade, we have witnessed an enormous expansion in our knowledge in understanding the molecular and cellular mechanisms underlying

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols*
Volume I, Emerging Strategies for Targets and Biomarker Discovery
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

disease pathogenesis and perpetuation. Although the human project has made available as many as 5000 potential new targets for drug intervention (http://function.gnf.org/druggable) ***(1)***, yet the main challenge facing the pharmaceutical industry is still to identify and validate novel disease-causing genes and to uncover additional roles for genes of known functions. Molecular profiling, which includes genomics and proteomics, has provided powerful tools with which to analyze gene expression in diseased and normal cells and tissues ***(2–8)***. An obvious example is mRNA expression profiling by using DNA microarray for large-scale analysis of cellular transcripts by comparing expression levels of mRNAs. In all cases, gene expression has been analyzed using clustering algorithms that group genes and samples on the basis of expression profiles. Because the major challenge of genomics studies is the detection of significant changes in gene expression levels, the use of adequate controls and statistical tests is crucial for these technologies. Based on the experimental design, appropriate algorithms should be used.

Microarray technology has been applied to the study of breast cancer to isolate tumor-specific genes as markers for diagnosis and prognosis, monitor changes in gene expression during chemotherapy to set predictors for outcome, classify ductal tumors based on epithelial cell subtypes, and identify candidate molecular targets ***(2,3)***. Also, the technology enabled the discovery of novel functional pathways, prognostic and therapeutic targets in various cancer types. In addition to their value for analyzing patient samples, gene profiling techniques also can provide important insight into disease when applied to experimental models that can be genetically or pharmacologically manipulated. Chapters 4 and 5, Volume 1, review these technologies and their applications to the study of breast and pancreatic cancers, providing a framework for appreciating the promise, while discussing the challenges ahead in translating genomics into real diagnostic, prognostic, or therapeutic products.

Although microarray technology has been successfully used to identify candidate targets, it is limited to detection of whatever is spotted on a slide, making it a "closed" approach for gene discovery. In contrast, other systems-based technologies such as differential display, representational difference analysis (RDA), and serial analysis of gene expression (SAGE) are capable of detecting both known and novel genes for any sample ***(5,9)*** (*see* Chapter 6, Volume 1). By screening for genes associated with high- but not low-grade human non-Hodgkin's B-cell lymphoma by RDA, we have identified a novel centrosome/microtubule-associated coiled-coil protein (CSPP) involved in cell cycle progression and spindle organization ***(10)***. Overexpression of CSPP is observed in human diffuse large B-cell lymphomas and correlated to shorter survival ***(10)***. As illustrated in **Fig. 1**, overexpression of green fluorescent protein-tagged CSPP in human embryonic kidney (HEK)293 protein inhibited bipolar spindle formation and induced aberrant mitotic spindle structures such as the depicted monopolar spindle.

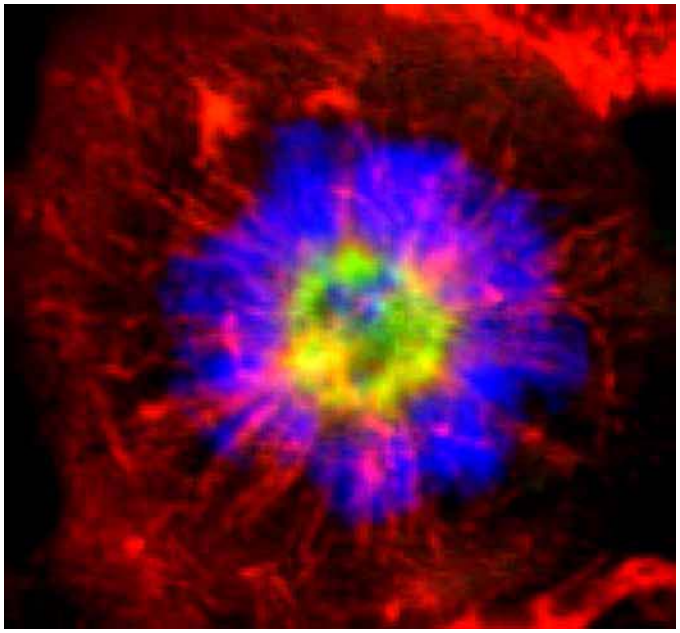


Fig. 1. Overexpression of green fluorescent protein-tagged centrosome/spindle pole associated protein in HEK293T cells (DNA, blue; CSPPegfp, green; and α-tubulin, red).

Expressed sequence tag (EST)-derived expression analysis and SAGE are based on the principle that frequency of sequence tag samples from a pool of cDNA is directly proportional to the expression levels of the corresponding gene. Chapter 4, Volume 1, showed that SAGE could potentially be used to identify prognostic and therapeutic targets in human pancreatic cancer.

The rapidly growing gene expression databases and improvement in bioinformatics tools set the stage for a new *in silico* strategy to discover therapeutic targets ***(2,6)***. Numerous gene expression studies in several cancer types can be downloaded from public databases. Notably, ESTs derived from diverse normal and tumor cDNA libraries offer an attractive starting point for cancer gene discovery. However, it is worth noting that although the *in silico* detection of gene variants holds great promise, it is subjected to the same limitations of all bioinformatics approaches in that the results need experimental validation to avoid false leads derived from noisy data. Chapter 2, Volume 1, highlights the strengths of *in silico* studies to identify diagnostic, prognostic, and therapeutic targets in several cancer types. One of the identified genes, *SIM2*, was validated in vitro and in vivo, and it could predict prostate cancer development. Interestingly, its knockdown induced apoptosis in cancer cells but not in normal cells.

A more fruitful approach for discovering drug targets is to analyze pathways known as genetic networks rather than individual genes. Once a pathway is identified as being relevant for the disease, it is then possible to assess the individual signaling proteins of the pathway for their potential druggability ***(8)***. New bioinformatics tools are being developed that will allow the projection of potential pathway alterations on the basis of the expression data. The analysis of gene networks after gene disruption or drug treatment should play an important role in identifying and validating drug target genes. In addition, this approach allows for the identification of pathways that are directly or indirectly affected by a drug in a cellular system. Chapter 3, Volume 1, provides an excellent illustration of gene network analysis in living organisms and describes the computer tools that can be used to analyze the data. Also, the identification of families of gene targets that are affected by drug treatment would be of great scientific interest as well as considerable pharmaceutical importance. When a signaling protein family already contains known drug targets, algorithms can define additional druggable family members. The advantage for searching for sequence similarity would allow the identification of novel drug targets and the design of selective drugs against only the target under investigation ***(8)***. It is worth nothing that cancer cells do not invent new pathways; they use preexisting pathways in different ways or they combine components of these pathways in a new manner. DNA microarray studies should provide insight into the connectivity of these connections.

Gene profiling and *in silico* analysis of expression databases to discover new target genes also are evaluated for the discovery of tumor antigens ***(6)***. Ideally, these new genomics approaches to antigen discovery in cancer should focus on finding target genes overexpressed in cancer cells with a higher degree of tumor specificity that are capable of inducing cellular immune responses in the host immune system. However, for most antigens defined by genomics approach, it is not yet known whether there is an endogenous immune response in patients. Thus, any identified gene target requires validation to demonstrate immunogenicity.

The main goal of several studies is to identify variants in genes that might affect disease outcome, drug response, or both. For example, human leukocyte antigen (HLA) molecules are characterized by high degree of polymorphism, and numerous allelic variants of both class I and class II molecules have been described. Many immune diseases are associated with certain HLA genes ***(11)***. However, the observed associations are far from absolute because some individuals carrying a given HLA molecule will never develop the associated disease. Therefore, for human diseases, it is important to understand how nongenetic factors interact to influence the phenotype. In addition to genetic changes, a set of human cancers is developed by de novo methylation in genes. Mainly, tumor suppressor genes where promoter methylation diminishes or

abolishes expression give advantage to tumor cells. The challenge ahead is to efficiently translate these epigenetic changes into real therapies. Chapter 2, Volume 2, highlights the epigenetic alterations in cancers and suggests potential therapeutic strategies.

When a gene variant is associated with drug response, there is often an obvious direction to pursue for making clinical use of the polymorphism ***(12)***. However, in complex genetic diseases, risk factors often do not lead to obvious treatment implications. For example, the discovery of *Brca1* and *Brca2* genes involved in hereditary breast cancer has provided critical insight into the mechanisms of carcinogenesis; however, neither gene is a viable therapeutic target. It is worth noting that mutations in a single gene can cause severe disease as exemplified by rare genetic diseases such as cystic fibrosis. With respect to treatment of these genetic disorders, where the gene responsible for the disease may have exerted its effect during the developmental process, gene therapy will play a crucial role.

1.1.2. Forward and Reverse Genetics

The major advantage of using genetic approaches instead of expression profiling genomics is that there is a clear link between a genetic alteration and a disease phenotype. Among the commonly used model organisms, the mouse is the preferred choice because, genetically, it is relatively close to human. Today, most human diseases can be mimicked in mouse ***(13)***. To date forward genetics has been most extensively performed using the chemical mutagen *N*-ethylnitrosouera (ENU) to produce random mutations in the genome ***(14,15)***. ENU induces single base-pair DNA alteration distributed through the genome. In addition to chemical mutagenesis, random integration of retroviruses or transposons could be the most desirable screen because the vector sequence provides a molecular fingerprint for the mutated gene, thereby facilitating its identification.

Although the observed alterations are limited to cellular effects, phenotype screening in mammalian cells can represent a valuable first method for target discovery. Loss-of-function genetic screens by using RNA interference (RNAi) now represents one of the most used technologies ***(16)***. RNAi is a powerful natural cellular process in which double-stranded RNAs (dsRNAs) target homologous mRNA transcripts for degradation ***(17)***. In this process, the dsRNA is recognized by an RNase III nuclease, which processes the dsRNA into small-interfering RNAs (siRNAs) of 19 nucleotides, with two nucleotide 3′ overhangs ***(18)***. Subsequently, these intermediates are incorporated into an RNA-interfering silencing complex, which contains the proteins needed to unwind the double-stranded siRNA and cleave the target mRNA at the site where the antisense RNA is bound. In addition to chemical synthesis, siRNAs can be expressed with a hairpin loop from plasmids and viral vectors, which offers the opportunity for

both transient and permanent cellular transfection ***(16)***. Given the complete sequence of the human genome, it is now possible to design siRNA libraries to target virtually any human gene and to screen for their contribution in human diseases. Large-scale gene knockdown studies using RNAi have been reported recently in mammalian systems. For example, RNAi-based inhibition of 800 genes led to the identification of novel players in the p53 and nuclear factor nuclear factor-κB pathways ***(19,20)***. The results indicate that a phenotype-based approach using RNAi in mammalian cells is feasible and will provide an efficient high-throughput approach for analysis of gene function and target discovery. In Chapters 7 and 8, Volume 1, Taira and colleagues used ribozyme and siRNA libraries to screen human cells for genes with a role in cancer cell migration and apoptosis. These studies verified the involvement of several known genes in these pathways as well as novel genes with no prior involvement.

Reserve genetics based upon functional mutation studies in animals has been successfully used to characterize genes involved in vital cellular pathways such as apoptosis and cell signaling. Moreover, knockout mice have proven to be invaluable tools for the functional analysis of biological process relevant to human physiology and diseases ***(13)***. Compared with other techniques, gene knockout models offer several advantages in drug target discovery. Indeed, the technology enables the discovery of target function as well as the mechanism-based side effects that might result from the inhibition of the target under investigation in mammals. There are several known human genes whose druggability was appreciated only after knockout studies ***(21)***. For example, although the human gene was identified several years ago, interest in acetyl-CoA carboxylase 2 (ACAC2) as target gene was recognized only recently because ACAC2 knockout exhibited a significant increase in fatty acid oxidation and lower glucose level than wild-type mice ***(22)***. This example illustrates the suitability of gene knockout technology to reveal relevant information about gene function and pharmaceutical utility of the target. ACAC2 protein is more likely the target of the antidiabetic drug gluocophage (metformin). Although the ultimate model organism is human, transgenic knockout mice not only facilitate drug target discovery and validation but also have contributed in our understanding of human diseases and normal physiology. Chapters 10 and 11, Volume 1, describe the utility of transgenic animals in target discovery, validation, and production of therapeutic proteins.

1.1.3. Proteomics

Cellular signaling events are driven by protein–protein interactions, posttranslational protein modifications, and enzymatic activities that cannot be predicted accurately or described by mRNA levels. Also, potentially important therapeutic targets might be differentially expressed at the protein but not mRNA level 2. Thus,

information obtained through transcript analysis needs complementation with information on the cellular protein level. Proteomics, the study of the proteins and protein pathways, is a promising technique for the rational identification of novel therapeutic targets ***(23)***. Two-dimensional (2D) polyacrylamide gel electrophoresis, multidimensional liquid chromatography, mass spectrometry, and protein microarray are among the proteomics techniques available for target discovery. Usually, protein extracts from cells or tissues are analyzed by 2D gel electrophoresis to identify differentially expressed or modified proteins. This technique is followed by mass spectrometry to determine the identity of individual protein spots of interest from the silver-stained gels. Also, several subcellular proteins fractions can be separately analyzed by proteomics, thereby reducing the complexity of the proteome. Regarding target and biomarker discovery, recent studies stress the power of combining several approaches to identity informative drug targets. For example, the combination of DNA microarray, SAGE, and proteomics led to an unbiased elucidation of new cancer biomarkers (*see* Chapter 4, Volume 1). Additionally, proteomics approaches can identify posttranslational modifications that can be important in the discovery of tumor antigens. Analysis of protein extract from cancer cells by immunoproteomics has resulted in the identification of novel tumor antigens (*see* Chapter 16, Volume 1).

Several other techniques such as the yeast two-hybrid system, protein chips, and phage display have joined proteomics strategies now available for profiling immune response in patients and identification of tumor antigens. During the past decade, several tumor antigens were discovered using SEREX ***(22,24,25)***, a method that screens the antibody response of cancer patients against a library containing autologous or allogeneic tumor cDNA libraries (*see* Chapter 29). In contrast to SEREX, the selectivity and specificity are expected to be higher when iterative methods are used. By biopanning peptides or cDNA phage-display libraries, we have identified peptides and antigenic proteins to which patients developed humoral immune responses ***(26,27)***. The technique, known as phage display, should substantially accelerate the pace of tumor antigen discovery and could provide a molecular signature for immune responses in different types of cancer. Chapter 14, Volume 1, provides an excellent overview of the immune system and technical advances in tumour antigen discovery. It also highlights the problem of immune tolerance and tumor editing, particularly to cancer vaccine development.

Whereas the principal goal of gene or serum profiling is to identify novel therapeutic targets, or serological markers, predictive screening strategies primary attempt to characterize treatment responses, coupled with the secondary aim of identifying novel therapeutic molecules. By contrast to the analysis of single serum biomarkers, protein arrays can be used to detect group of proteins that, in the aggregate, contain significantly more predictive information than do individual

serum biomarkers. This synergism is also true for the analysis of gene biomarkers detected by DNA microarrys. Given the importance of serological biomarkers, we have developed a protein microarray containing several immunoselected phage-displayed peptides and shown its application in breast cancer diagnosis ***(28)***. Chapter 17, Volume 1, describes this novel approach and highlights the importance of protein arrays in target discovery and validation ***(16)***.

1.2. Target Validation

Once a gene target or a mechanistic pathway is identified, the next step is to demonstrate that it does play a critical role in disease initiation, perpetuation, or both. A range of in vitro assays exists for modulating gene expression in vitro and in vivo. These include the use of antibodies, dominant negative controls, antisense oligonucleotides, locked nucleic acids, peptide nucleic acids, ribozymes, and siRNAs. There are several notable examples in which these technologies have been targeted against genes with already well-characterized roles in disease states and these examples have served as a proof of concept for their use in target validation ***(16)***.

Today, the RNAi technology seems to offer a number of significant advantages over other strategies, including increased specificity ***(16)***. Although prediction algorithms have begun to offer some guidance for siRNA sequence selection, identification of highly effective and gene-specific siRNA and their delivery to human cells is often problematic, and various factors may influence siRNA action. Chapter 9, Volume 2, describes the latest guidelines for designing effective siRNAs for functional genomics and target validation studies. Also, the described rules should potentially enhance our ability to design more effective hairpin siRNA libraries. By using a cell array, target discovery and validation using siRNAs can be performed in a high-throughput format (*see* Chapter 9, Volume 1). The great advantage of protein knockdown through mRNA modulation is that only sequence information about the target gene is required. A critical issue for the use of synthetic siRNAs in target validation is the choice of controls and siRNA sequences. One should therefore verify that the sequence-dependent off-target effects are low and that the designed siRNAs are not immunostimulatory ***(29)***.

Genetically modified animals that either overexpress (transgenic animals) or no longer express the target (knockout animals) are extremely useful in the target validation and drug discovery processes. These models represent the best available tool for examining the complex interaction that underlines pathological responses. When the target is a single gene, target validation will frequently involve producing a transgenic animal that carries either the wild-type gene or its mutated form to demonstrate that this animal is either overexpressing the wild-type gene or that the mutated gene has a phenotype that mimics a certain aspect of the human disease. Subsequently, drug screening can be performed in this animal to show that modulation of transgene expression can have therapeutic effects.

As with transgenic animals, reverse genetics using mouse knockouts played a crucial role in target validation and drug discovery. The lack of the target gene will simulate the effect of complete target inhibition by the desired, future drug lead. Interestingly, the 100 best-selling drugs of 2001 are directed at only 43 host proteins that have been validated in mouse knockouts ***(21)***. Chapter 11, Volume 1, describes the utility of this technology in functional studies to understand how mutations lead to diseases. It also provides excellent protocols and guidelines for the generation of transgenic and knockout mice. Cre-lox system has now been used in several successful studies to generate null mutations in specific genes. Moreover, novel techniques are described that can create "conditional knockout," which is designed to model the more relevant therapeutic intervention. Also, the major techniques to construct transgenic animals are summarized in Chapter 10, Volume 1.

Animal models of autoimmune diseases have been developed that mimic some aspects of the pathophysiology of human disease. These models have increased our understanding of possible mechanisms of pathogenesis at the molecular and cellular level and have been important in target validation and testing of new immunotherapies. The most used model of rheumatoid arthritis is the collagen-induced arthritis model in DBA/1 mice, which has both a cellular and humoral component ***(30)*** (*see* Chapter 13, Volume 2). It seems that local tumor necrosis factor is the driving force behind the chronic inflammation. Given the chronic nature of autoimmune diseases and the complexity and redundancy of the involved cytokines and cells, the therapeutic effects of new molecules should be assessed thoroughly. Overviews of new treatment options for rheumatic diseases are described in Chapter 14, Volume 2. However, despite the discovery of new drugs, corticosteroids, which were life-saving 50 yr ago, remain a stable treatment in several autoimmune diseases, including lupus ***(31)***.

For some diseases, it might not be possible to mimic the complete human pathology in the mouse; for others, it might be more desirable to create a model of specific pathological (e.g., metastasis) or physiological (e.g., angiogenesis) processes. Chapter 12, Volume 1, describes the development of in vivo assay that enables the study of angiogenesis and testing therapies. Also, a murine model for studying hematopoiesis and immunity in heart failure is described in Chapter 13, Volume 1. Using the innovative methods described above, several therapeutic targets have been identified and validated in vitro and in vivo. Other chapters in Volume 1 cover many important aspects of target validation and clinical trials with some of the developed agents.

1.3. Drug Discovery

Once a target is validated, the next step is the drug discovery process in which chemical agents with activity against a specified target or function are identified

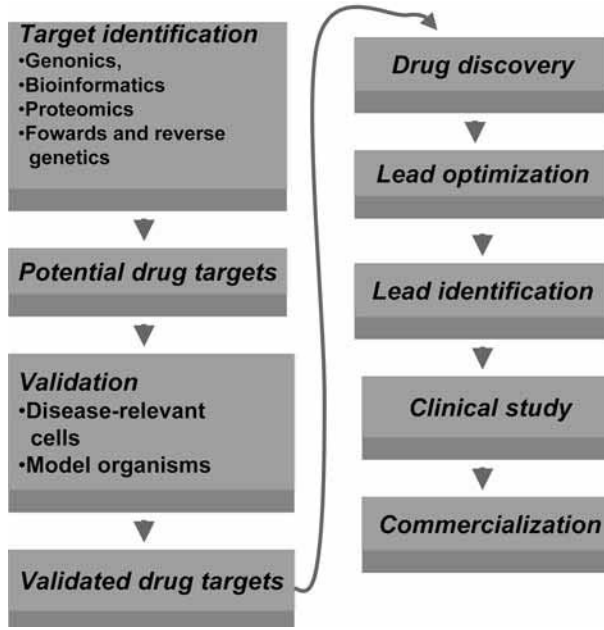


Fig. 2. Schematic of target and drug discovery steps.

through high-throughput screening **(Fig. 2)**. This process requires the use of robust assay of target activity and a good collection of component libraries for testing. Positive "hits" from the primary screening usually are evaluated on the basis of efficacy, specificity, toxicity, and pharmacology in animal models, to identify lead compounds for further optimization. Primary screening for efficacy and specificity can be performed in cell systems. However, active agents in vitro may be inactive in vivo because of in vivo metabolism, elimination, and poor penetration into targets tissues or just because the target was not involved in the pathway under investigation in vivo. During preclinical animal studies (e.g., pharmacology and toxicity), it is important to use biomarkers that can be translated for human use in phase 1 studies. Additionally, it is urgent to use appropriate animal models and experimental design. Indeed, many novel drugs that worked in animal models have failed in humans. Drugs failed because of lack of efficacy.

Given the huge number of potential druggable targets offered by the human genome sequence and new technologies for target validation such as RNAi, new smart drugs for most complex diseases may be developed in the near future. To reach this goal, pharmaceutical companies should examine thoroughly the validity of the target at early stage and define rational drug discovery programs. It does not matter how many targets we select: what really matters is whether the selected targets are involved in the disease under study.

High-throughput screens of small-molecule inhibitors have generated many target-specific drugs that could potentially be used for the treatment of patients with cancer (*see* Chapters 1 and 2, Volume 2). Also, molecules such as antisense oligonucleotides, siRNAs, and antibody inhibitors used to validate gene targets are themselves being developed as drugs (*see* Chapters 7, 10, 12, Volume 2). However, new developed drugs should be given to the right patients because the target might be active or present in only certain patients. There is also a growing list of genetic polymorphisms in drug metabolizing enzymes and transport proteins that have been shown to influence drug response. Drug concentrations in plasma can vary more than 600-fold between two individuals of the same weight on the same drug dosage (*see* Chapter 17, Volume 2). Therefore, molecular diagnostics need to be developed and integrated with drug development and clinical trial design.

References

1. Chanda, S. K. and Caldwell, J. S. (2003) fulfilling the promise: drug discovery in the post-genomic era. *Drug Discov. Today* **8,** 168–174.
2. Segal, E., Fiedman, N., Kaminski, N., Regev, A., and Koller, D. (2005) From signatures to models: understanding cancer using microarrays. *Nat. Genet*. **37,** S38–S45.
3. Perou, C. M., Sørlie, T., Eisen, M. B., et al. (2000) Molecular portraits of human breast tumours. *Nature* **406,** 747–752.
4. Boon, K., Osorio, E. C., Greenhut, S. F., et al. (2002) An anatomy of normal and malignant gene expression. *Proc. Natl. Acad. Sci. USA* **99,** 11,287–11,292.
5. Lash, A. E., Tolstoshev, C. M., Wagner, L., et al. (2000) SAGEmap: a public gene expression resource. *Genome Res.* **10,** 1051–1060.
6. Loging, W. T., Lal, A., Siu, I. M., et al. (2000) Identifying potential tumor markers and antigens by database mining and rapid expression screening. *Genome Res*. **10,** 1393–1402.
7. de Hoog, C. L. and Mann, M. (2004) Proteomics. *Annu. Rev. Genomics Hum. Genet*. **5,** 267–293.
8. Brazhnik, P., de la Fuente, A., and Mendes, P. (2002) Gene networks: how to put the function in genomics. *Trends Biotechnol*. **20,** 467–472.
9. Odeberg, J., Wood, T., Blucher, A., Rafter, J., Norstedt, G., and Lundeberg, J. (2000) A cDNA RDA protocol using solid-phase technology suited for analysis in small tissue samples. *Biomol. Eng*. **17,** 1–9.
10. Patzke, S., Hauge, H., Sioud, M., et al. (2005) Identification of a novel centrosome/microtubule-associated coiled-coil protein involved in cell-cycle progression and spindle organization. *Oncogene* **24,** 1159–1173.
11. McDevitt, H. O. (1998) The role of MHC class II molecules in susceptibility and resistance to autoimmunity. *Curr. Opin. Immunol*. **10,** 677–681.
12. Nebert, D. W., Jorge-Nebert, L., and Vesell, E. S. (2003) Pharmacogenomics and "individualized drug therapy": high expectations and disappointing achievements. *Am. J. Pharmacogenomics* **3,** 361–370.

13. Magin, T. M. (1998) Lessons from keratin transgenic and knockout mice. *Subcell. Biochem*. **31,** 141–172.
14. Hrabe de Angelis, M. H., Flaswinkel, H., Fuchs, H., et al. (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat. Genet*. **25,** 444–447.
15. Balling, R. (2001) ENU mutagenesis: analyzing gene function in mice. *Annu. Rev. Genomics Hum. Genet.* **2,** 463–492.
16. Sioud, M. (2004) Therapeutic siRNAs. *Trends Pharmacol. Sci.* **25,** 22–28.
17. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391,** 806–811.
18. Elbashir, S. M., Harborth, M. J., Ledecknel, W., Yalcin, A., Weber, K., and Tuchl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411,** 494–498.
19. Berns, K., Hijmans, E. M., Mullenders, J., et al. (2004) A large-scale RNAi screen in human cells identifies new compounds of the p53 pathway. *Nature* **428,** 431–437.
20. Zheng, L., Liu, J., Batalov, S., et al. (2004) An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl. Acad. Sci. USA* **101,** 135–140.
21. Zambrowicz, B. P. and Sands, A. T. (2003) Knockouts model the 100 best-selling drug—will they model the next 100? *Nature* **2,** 38–51.
22. Abu-Eheiga, L., Malzuk, M. M., Abo-Hasema, K. A. H., and Wakil, S. J. (2001) Continuous fatty acid oxidation and reduced fat storage in mice lacking acetyl-CoA carboxylase 2. *Science* **291,** 2613–2616.
23. Resing, K. A. and Ahn, N. G. (2005) Proteomics strategies for protein identification. *FEBS Lett.* **579,** 885–889.
24. Sahin, U., Tureci, O., Schmitt, H., et al. (1995) Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc. Natl. Acad. Sci. USA* **92,** 11,810–11,813.
25. Jäger, D., Stockert, E., Gure, A. O., et al. (2001) Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. *Cancer Res*. **61,** 6197–6204.
26. Hansen, M. H., Østenstad, B., and Sioud, M. (2001) Antigen-specific IgG antibodies in stage IV long-term survival breast cancer patients. *Mol. Med*. **7,** 230–239.
27. Sioud, M. and Hansen, M. H. (2001) Profiling the immune response in patients with breast cancer by phage-displayed cDNA libraries. *Eur. J. Immunol*. **31,** 716–725.
28. Cekaite, L., Haug, O., Myklebost, O., et al. (2004) Analysis of the humoral immune response to immunoselected phage-displayed peptides by a microarray-based method. *Proteomics* **4,** 2572–2582.
29. Sioud, M. (2005) Induction of inflammatory cytokines and interferon responses by double-stranded and single-stranded siRNAs is sequence-dependent and requires endosomal localization. *J. Mol. Biol*. **348,** 1079–1090.
30. Williams, R. O. (1998) Rodent models of arthritis: relevance for human disease. *Clin. Exp. Immunol.* **114,** 330–332.
31. Smolen, J. S. and Steiner, G. (2003) Therapeutic strategies for rheumatoid arthritis. *Nat. Rev. Drug Discov*. **2,** 473–488.

2

Bioinformatics Approaches to Cancer Gene Discovery

Ramaswamy Narayanan

Summary

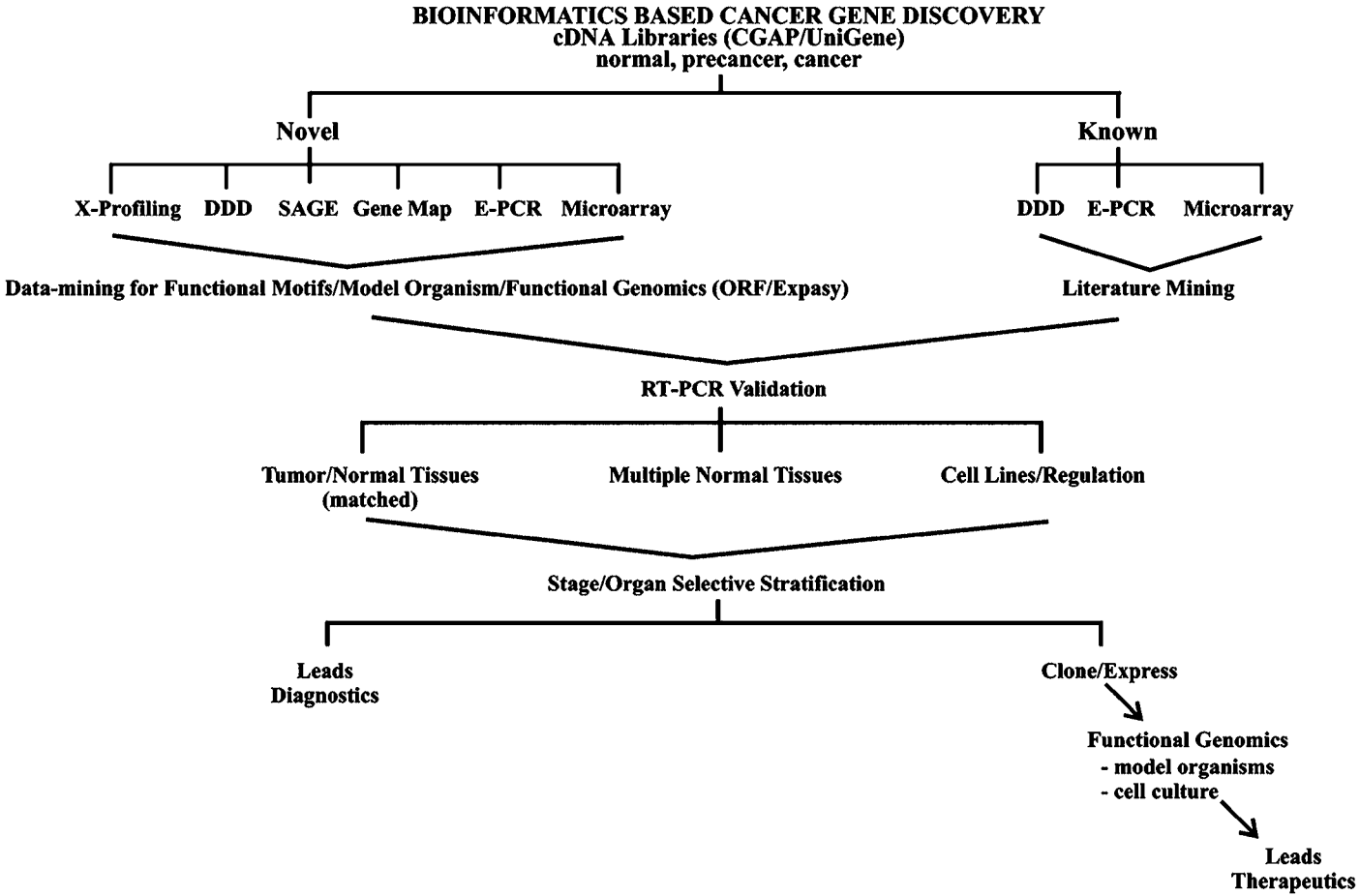
The Cancer Gene Anatomy Project (CGAP) database of the National Cancer Institute has thousands of known and novel expressed sequence tags (ESTs). These ESTs, derived from diverse normal and tumor cDNA libraries, offer an attractive starting point for cancer gene discovery. Data-mining the CGAP database led to the identification of ESTs that were predicted to be specific to select solid tumors. Two genes from these efforts were taken to proof of concept for diagnostic and therapeutics indications of cancer. Microarray technology was used in conjunction with bioinformatics to understand the mechanism of one of the targets discovered. These efforts provide an example of gene discovery by using bioinformatics approaches. The strengths and weaknesses of this approach are discussed in this review.

Key Words: Antisense; digital differential display; Down syndrome; genomics; solid tumors.

1. Introduction

The recent completion of the human genome sequencing efforts offers a new dimension in gene discovery approaches ***(1–4)***. From these vast numbers of new genes, new diagnostic and therapeutic targets for diseases such as cancer are predicted to emerge ***(5)***. Only a subset of genes is expressed in a given cell, and the level of expression governs function. High-throughput gene expression technology is becoming a possibility for analyzing expression of a large number of sequences in diseased and normal tissues with the use of microarrays and gene chips ***(6–9)***. A parallel way to initiate a search for genes relevant to cancer diagnostics and therapy is to data-mine the sequence database ***(10–14)***. A large number of expressed sequences from diverse organ-, species-, and disease-derived cDNA libraries are being deposited in the form of expressed sequence tags (ESTs) in different databases.

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols*
Volume I, Emerging Strategies for Targets and Biomarker Discovery
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ


BIOINFORMATICS BASED CANCER GENE DISCOVERY
cDNA Libraries (CGAP/UniGene)
normal, precancer, cancer
Novel
Known
X-Profiling
DDD
SAGE
Gene Map
E-PCR
Microarray
DDD
E-PCR
Microarray
Data-mining for Functional Motifs/Model Organism/Functional Genomics (ORF/Expasy)
Literature Mining
RT-PCR Validation
Tumor/Normal Tissues (matched)
Multiple Normal Tissues
Cell Lines/Regulation
Stage/Organ Selective Stratification
Leads
Diagnostics
Clone/Express
Functional Genomics
- model organisms
- cell culture
Leads
Therapeutics

The Cancer Gene Anatomy Project (CGAP) database (http://cgap.nci.nih.gov/) of the National Cancer Institute (NCI) is an attractive starting point for cancer-specific gene discovery ***(12)***. The Human Tumor Gene Index was initiated by the NCI in 1997 with a primary goal of identifying genes expressed during development of human tumors in five major cancer sites: (1) breast, (2) colon, (3) lung, (4) ovary, and (5) prostate. This database consists of expression information (mRNA) of thousands of known and novel genes in diverse normal and tumor tissues. By monitoring the electronic expression profile of many of these sequences, it is possible to compile a list of genes that are selectively expressed in the cancers. Data-mining tools are becoming available to extract expression information about the ESTs derived from various CGAP libraries ***(10,13–15)***.

Currently, there are 1.5 million ESTs in the CGAP database, of which 73,000 are novel sequences. These sequences also are subclassified into those derived from libraries of normal, precancerous, or cancer tissues. The CGAP database uses UniGene-based gene clustering. UniGene (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=unigene) is an experimental system for automatically partitioning GenBank sequences into a nonredundant set of gene-oriented clusters. Each UniGene cluster contains sequences that represent a unique gene. An essential weakness of the UniGene is that currently it does not allow for the identification of consensus or contig sequences of a gene. In addition, the UniGene database is continuously updated and hence the ESTs are often removed and reassigned to another UniGene. However, the contig or the consensus (Tentative Human Consensus, TC/THC) information can be obtained from other databases such as TIGR (http://www.tigr.org/). Multiple data-mining tools from the CGAP database can be used to facilitate the gene discovery.

2. Gene Discovery Strategy

An overall strategy for cancer gene discovery by using bioinformatics approaches is shown in **Fig. 1**. Survey sequencing of mRNA gene products can provide an indirect means of generating gene expression fingerprints for cancer cells and their normal counterparts. The cancer specificity of an EST can be predicted using multiple data-mining tools from the CGAP database. These tools include X-profiling, digital differential display (DDD), serial analysis of gene expression (SAGE), electronic-PCR, GeneMap, and microarray databases (*see* Chapters 1, 5, 6, Volume 1). For details of these tools, see the CGAP website.

Fig. 1. *(Opposite page)* Gene discovery strategy. A proposed approach to cancer gene discovery from the CGAP database is shown. Both novel and known ESTs are identified using multiple data-mining tools from this database. Further validation in the wet laboratory provides a rational for diagnostic and therapeutic target discovery.

Recently, a tool called digital analysis of gene expression has replaced DDD in the CGAP database; however, the DDD tool can still be accessed from the UniGene page. Briefly, these tools allow prediction of ESTs specific to each cancer type by comparing the occurrence of an EST between two pools of cDNA libraries in a statistically significant manner. The approaches to doing follow-up studies on novel or known ESTs may differ. The novel ESTs can be subject to additional data-mining to identify functional motifs by using the protein databases from the ExPASy (http://au.expasy.org/) server. For a novel EST, a hint of its function also is obtained from homologue-related information from a model organism database such as Homologene (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db= homologene), whereas the known ESTs can be subjected to literature-mining to develop hints of relevance to specific cancer types. The chosen ESTs (hits) can be validated by reverse transcription (RT)-PCR by using cDNAs from normal and tumor tissues as well as appropriate cell lines for specificity and regulation. The expression specificity of the validated ESTs (leads) provides a rationale for developing a diagnostic application. The drug therapy use can be inferred using numerous techniques, such as antisense, small-interfering RNA, and so on.

3. Example of Cancer Gene Discovery

To identify solid tumor-specific genes, the DDD tool of the CGAP database was used. Both novel and known ESTs that are selectively up- or downregulated in six major solid tumor types (breast, colon, lung, ovary, pancreas, and prostate) were identified. DDD takes advantage of the UniGene database by comparing the number of times ESTs from different libraries were assigned to a particular UniGene cluster. Six different solid tumor-derived EST libraries (breast, colon, lung, ovary, pancreas, and prostate) with corresponding normal tissue-derived libraries were chosen for DDD (N = 110). To identify tumor- and organ-specific ESTs, all the other organ- and tumor-derived EST libraries (N = 327) were chosen for comparison with each of the six tumor types. The nature of the libraries (normal, pretumor, or tumor) was authenticated by comparison of the CGAP data with the UniGene database. Occasionally, the description of the EST libraries in CGAP and UniGene database do not show a match. Those few libraries showing discrepancies of definition between the two databases were excluded. The DDD was performed for each organ type individually. DDD was performed using ESTs from tumors (pool A) and corresponding normal organ (pool B) by using the online tool. The output provided a numerical value in each pool denoting the fraction of sequences within the pool that mapped to the UniGene cluster, providing a dot intensity. An example of DDD output for colon tumor-specific gene discovery by DDD is shown in **Fig. 2**.

A colon tumors	B all others	Gene index	Gene Description	Fold
0.00241 A>B	0.00008 B<A	Hs.25640	claudin 3 (CLDN3)	30 (up regulated)
0.00006 A>B	0.00283 B<A	Hs.75792	hemoglobin, alpha 1 (HBA1)	47 (down regulated)
0.00000 A>B	0.00058 B<A	Hs.118843	creatine kinase, muscle (CKM)	–
0.00020 A>B	0.00000 B<A	Hs.1545	caudal type homeo box transcription factor 1 (CDX1)	+
0.00031 A>B	0.00000 B<A	Hs.75792	ESTs	+

Fig. 2. Example of DDD output. The numerical value in each box is the fraction of ESTs within the pool that mapped to the UniGene cluster (Hs.) shown. The dot is merely a visual aid that reflects the numerical values. If any pool participates in a statistically significant pairwise comparison with another pool, the relationship is indicated. "A>B" indicates a greater amount of ESTs found in colon cancer libraries versus other libraries for a particular gene. If the number of occurrences of an EST in a UniGene is zero, then the EST is predicted to be present only in the colon cancer libraries (shown as +/– fold). If the number is finite, then the ratio is shown as the -fold difference.

Fold differences were calculated by using the ratio of pool A:pool B. Statistically significant hits (Fisher exact test) showing >10-fold differences were compiled, and a preliminary database was created. Novel ESTs were compiled into a separate database. The UniGene database was accessed to establish an electronic expression profile (E-Northern) for each of the hits to facilitate tumor- and organ-selective gene discovery. The cytogenetic map position of the hits also was inferred from the UniGene page. A final database of ESTs that were upregulated, downregulated, and showed absolute differences (+/–) in the six tumor types was created. These hits were functionally classified into major classes of proteins by using gene ontology. Genes belonging to ribosomal proteins, enzymes, receptors, binding proteins, secretory proteins, and cell adhesion molecules were identified to be differentially expressed in these tumor types. A comprehensive database of hits was created, providing additional electronic expression data as well as novel ESTs that were thus identified ***(16)***. This database can be accessed on the World Wide Web at http://www.fau.edu/cmbb/publications/cancergenes.htm.

Colorectal cancer is a commonly diagnosed cancer in both men and women. In 2006, an estimated 106,680 new cases will be diagnosed, and 55,170 deaths from colorectal cancer will occur in the United States alone. About 75% of patients with colorectal cancer have sporadic disease, with no apparent evidence of having inherited the disorder. The remaining 25% of patients have a family history of colorectal cancer that suggests a genetic contribution, common exposures among family members, or a combination (http://www.nci.nih.gov/cancerinfo/pdq/genetics/colorectal). Although tumor suppressor genes such as deleted in colorectal cancer (DCC), adenomatous polyposis coli (APC), mutated in colorectal cancer (MCC), or oncogenes such as k-*ras* offer a promise for diagnosis of colon cancer, additional more specific markers are urgently needed to benefit colon cancer patients. With this in view, we chose two genes from this database predicted to be specific for colon tumors to test the validity of gene discovery by bioinformatics approaches.

4. Discovery of Colon Cancer-Specific Secreted Marker

To date a secreted marker for colon cancer diagnosis has not been identified. Hence, an attempt was made to predict and validate a colon cancer-specific EST that might harbor a signal peptide motif. Sixty-four UniGenes were identified to be upregulated in colon tumors by the DDD approach. Twenty-four of these UniGenes were found to be present only in colon tumor-derived cDNA libraries (CGAP, SAGE, and UniGene). One UniGene, Hs. 307047, which harbored a signal peptide motif, was chosen for further analysis. This UniGene currently has seven different ESTs as a part of the Unigene cluster. The SAGE tag (ACAGTAATGA) identified for this UniGene was also in a colon and gastric tumor-derived library. The TIGR Human Gene Index had a tentative Human Consensus (THC342146) comprising all of the seven different ESTs shown above with the longest EST being AA524300. The EST AA524300 was screened against the National Center for Biotechnology Information database for alu, vector, and bacterial contamination and was found to have no matches. There were significant matches against mouse and rat ESTs, all of them from colon-derived libraries. An *in situ* BLAST library-specific search at TigemNet (http://www.tigem.it) revealed that this EST was not present in any normal colon library. In addition, no match was detected against the Bodymap (http://bodymap.ims.u-tokyo.ac.jp) normal library collections. Comparison against the GeneMap (http://www.ncbi.nlm.nih.gov/genemap) suggested that the EST is located on chromosome 3q13.1. The translated sequence for this EST, compared against the Prosite signature database of ExPASy (http://www.expasy.ch/prosite) showed that this EST encodes a putative signal peptide, prokaryotic lipid binding sites, a prenyl group binding site and membrane glycosylphosphatidylinositol anchor sites. The prediction of a signal peptide sequence and the colon cancer

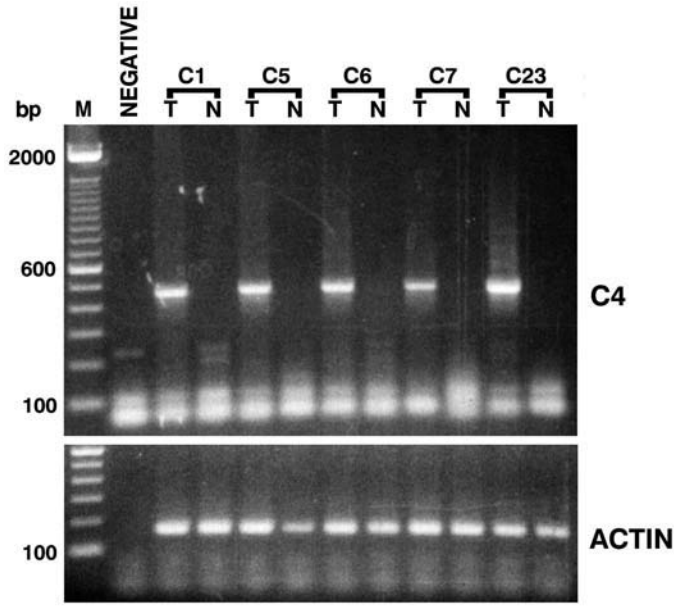


Fig. 3. Discovery of colon cancer-specific secreted marker. The cDNAs from matched sets of tumor (T) and normal (N) colon from five different patients were analyzed by RT-PCR for and for CCRG and actin. M, 100-bp ladder; negative, template minus control.

specificity allowed us to test a rationale of secreted marker discovery for colon cancer diagnosis. Preliminary RT-PCR analysis of a matched set of tumor and normal colon-derived cDNAs by using an exon-specific PCR primer pair detected a product in the tumor but not in the normal colon cDNA **(Fig. 3)**. A comprehensive RT-PCR based expression profiling revealed that this gene is expressed only in normal small intestine among many normal organs. Developmental expression in the fetal brain, kidney, and lung was seen. In addition, cDNAs from matched sets of tumor and normal tissues of breast, lung, ovary, pancreas, and prostate were negative for this EST expression, demonstrating that EST AA524300 is selectively upregulated in colon tumors but not in other major solid tumors. Furthermore, the EST expression was detected in cDNAs from cell lines of colon carcinomas, but not in the cell lines of breast, lung, ovary, pancreas, or prostate carcinomas. The EST expression was detected in three of three adenomas and three of three carcinomas of colon, but not in the polyp, which suggested that the putative gene encoded by this EST may be activated during early stages of colon cancer. Elevated colon carcinoma-related gene (CCRG) protein expression also was detected in the paraffin sections of colon tumors in comparison with the corresponding normal tissues ***(17)***.

The entire cDNA sequence of the *CCRG* is deposited in GenBank under accession AF323921. The putative open reading frame codes for 111 amino acids. The predicted molecular mass of the open reading frame of CCRG is 9.4 kDa, and its theoretical p*I* is 7.6. The C terminus has a unique cysteine-rich motif, 1CX11; 2CX8; 3CX1; 4CX3; 5CX10; 6CX1; 7CX1; 8CX9; 9C10C. A PFAM search (http://pfam.wustl.edu/) indicated that such a motif is present in keratin ultrahigh sulfur matrix proteins, metallothionine, and low-density lipoprotein-receptor-related proteins.

While our work was in progress, several independent laboratories also identified this new family of genes. Holcomb et al. ***(18)*** identified a novel protein in a mouse model of allergic pulmonary inflammation that they called FIZZ1 (found in inflammatory zone). By performing a genomic screening, this group identified two other mouse homologs and two human homologs. mFIZZ2, found in intestinal crypt epithelium, had a human homolog identical to the original EST discovered in our work (AA524300) and was named hFIZZ1. Another group identified a protein in adipocytes that potentially linked obesity and insulin resistance to diabetes in a mouse model and named it resistin ***(19)***. Two other related proteins were identified in mice and humans, termed resistin-like molecule (RELM) alpha and beta ***(20)***. RELM alpha is prevalent in the stromal–vascular fraction of adipose tissue and lung, whereas PEΛM β is found in colonic epithelium. Steppan et al. ***(20)*** demonstrated that in mouse intestine RELM beta is predominant in proliferative epithelia at the base of the crypts and becomes diminished in nonproliferative differentiated epithelia that have migrated up from the crypt towards the luminal surface ***(20)***. Also, in a *min* mouse model, which is a model of familial adenomatous polyposis due to the harboring of a mutated APC gene ***(21)***, increased RELM beta mRNA expression was observed in tumors ***(20)***. Another study showed that RETNLB mRNA and protein expression was restricted to undifferentiated proliferating epithelium. These authors also detected RETNLB protein in the stools of human and mice ***(22)***. Another group ***(23)*** recently described identification of a set of novel genes by using representative difference analysis of myeloid cells from CCAAT/enhancer-binding protein δ knockout mice. A human homolog to one of the novel genes from this study was termed HXCP2, for a gene selectively upregulated in small intestine and colon. Amino acid homology studies indicate that hFIZZ1, RELM beta and HXCP2 are all 100% identical to the gene we discovered, CCRG. Based on all these results, The Human Genome Organization Gene Nomenclature Committee recently assigned the symbol RETNLB (resistin-like β) to the gene encoding CCRG/hFIZZ1/RELM betã/HXCP2. Other members of this newly described family of genes include resistin-like α, which shows specificity to adipocytes ***(24)*** and bronchial epithelial cells ***(25)*** and Resistin, which is specific to adipocytes ***(18)***.

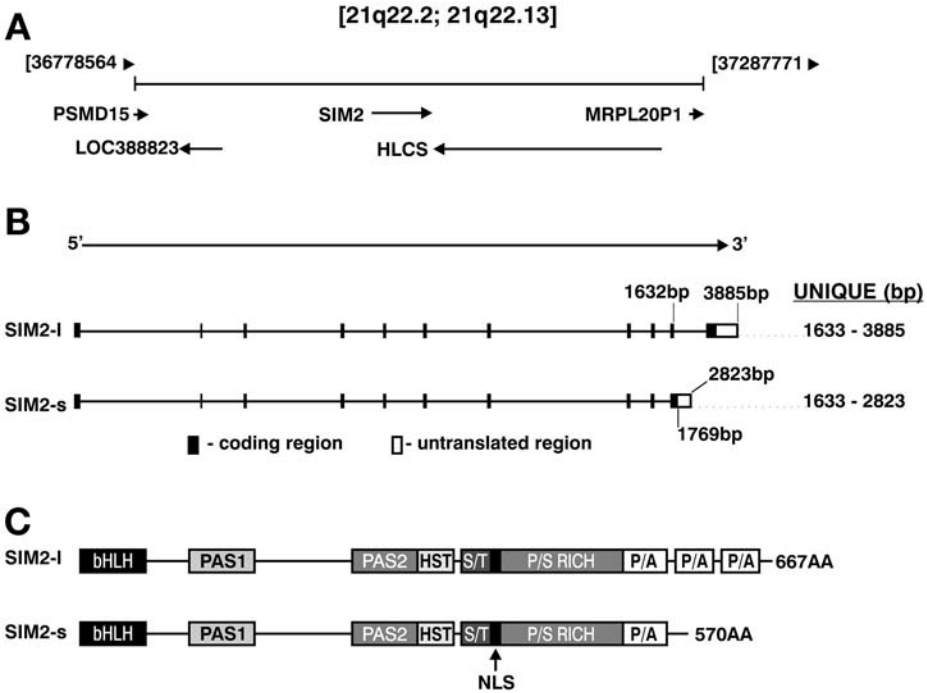


Fig. 4. Structure of human SIM2. **(A)** Location of the *SIM2* gene on a partial map (36,778,564–37,287,771 bp) of chromosome 21. HLCS, holocarboxylase synthetase; PSMD15, proteasome (prosome, macropain) 26S subunit, non-ATPase, 15; LOC388823, novel gene supported by EST alignment; MRPL20P1, mitochondrial ribosomal protein L20 pseudogene 1. **(B)** mRNAs for both isoforms of SIM2 (SIM2-l and SIM2-s) are shown. Boxes represent exons and lines represent introns. The unique 3′ sequences for each isoform are described. bp, base pair. **(C)** Schematic diagram of the conserved domains found in both SIM2 proteins. bHLH, basic helix-loop-helix; PAS1 and 2, PER/ARNT/SIM domains; HST, HIF1-α/SIM/TRH domain; S/T, Ser/Thr-rich region; NLS, nuclear localization signal; P/S RICH, Pro/Ser-rich region; P/A, Pro/Ala-rich region

5. Identification of Single-Minded 2 Gene as a Drug Therapy Target for Solid Tumors

Another EST predicted by the DDD tool to be upregulated in colon tumors belonged to UniGene cluster Hs. 146186, which was homologous to the Down syndrome-associated *SIM2* gene. Members of the human *SIM* gene family include *SIM1* and *SIM2*, which map to 6q16.3-q21 and 21q22.2, respectively ***(26,27)***. The *SIM2* locus spans over a 365-kilobase region on chromosome 21, as shown in **Fig. 4A** ***(28)***. In contrast to other species, the human *SIM2* gene exists in two distinct forms, the long and short forms (*SIM2-l* and *SIM2-s*) as shown in **Fig. 4B** ***(26)***. The original cDNA clone identified was *SIM2-s*, whereas genomic

sequencing subsequently identified *SIM2-l*. The *SIM2-s* mRNA includes a unique 3′ end, encoded by part of intron 10. This unique region, starting at the last nucleotide of exon 10 and including 1191 base pairs (bp) of intron 10, encompasses both coding and 3′ untranslated sequences. The 3′ end of *SIM2-l* mRNA is encoded by exon 11, which is separated from exon 10 by the complete intron 10 sequence (2528 bp). It is unclear whether these two isoforms are a consequence of an alternative use of the 3′ untranslated region contained within intron 10 of *SIM2* or are due to mispriming by using an A-rich sequence within this intron ***(26)***.

The SIM proteins belong to a family of transcription factors characterized by the basic helix-loop-helix (bHLH) and PAS (PER/ARNT/SIM) protein motifs ***(29)***. Regulation of transcription involves the dimerization of two bHLH-PAS proteins within the nucleus, DNA binding by the basic region of the bHLH, and interaction with the transcriptional machinery to modulate gene expression (reviewed in ref. ***29***). The N-terminal bHLH is the key motif for primary dimerization of the two proteins, whereas the PAS domain acts as the interface for selective protein partnering ***(29,30)***. The heterodimerizing partner choice of PAS proteins is highly specific and determines target gene regulation ***(31–33)***. The human SIM proteins are highly homologous to *Drosophila* and murine SIMs in their N-terminal domains, which contain a bHLH motif, two PAS domains, and the HST (HIF1-α/SIM/TRH) domain **(Fig. 4C)**. In the mouse, mSIM1 and mSIM2 can heterodimerize with ARNT, ARNT-2 or BMAL1 ***(31,32,34–36)***, whereas in humans, hSIM1 and hSIM2 can dimerize with either ARNT or ARNT-2 ***(37)***. A novel 23-amino acid nuclear localization signal was recently identified between residues 367 and 389 ***(38)***. This signal is in a region found in both forms of the human SIM2 protein.

The C-terminal part of the human proteins is considerably divergent from other family members; however, there is still high conservation between hSIM1 and hSIM2 compared with mSIM1 and mSIM2, respectively ***(26)***. Both forms of human SIM2 contain a Ser/Thr-rich region (S348–T366), a Pro/Ser-rich region (P385–S503), a Pro/Ala-rich region (P504–P544) and a positively charged region between R367 and R382. In addition, the long-form protein contains two additional Pro/Ala-rich regions (P533–P596 and P611–P644) and a positively charged region (K559–R575). Domains rich in Ser/Thr and proline residues are present in both transcriptional repressors ***(39)*** and transcriptional activators ***(40,41)***, whereas Pro/Ser and Pro/Ala domains are characteristic of repressor motifs (reviewed in ref. ***42***).

At the time of discovery, the *SIM2* gene has not been linked to any cancer types. Reasoning that if the DDD prediction of specificity of *SIM2-s* could be validated in a cancer model, we would be able to link the Down syndrome gene *SIM2* with cancer and hence derive a novel cancer utility for an already known gene, further studies were undertaken.

The mRNA analysis of *SIM2-s* expression of a panel of tumor and normal human tissue-derived cDNAs by using RT-PCR showed tumor-specific expression of *SIM2-s* ***(43–45)***. In contrast to the bioinformatics prediction of colon tumor specificity, the SIM2-s-specific RT-PCR product also was seen in the pancreas and prostate tumor-derived cDNAs but not in the corresponding matched normal tissues. These results underscore the importance of wet laboratory validation of the bioinformatics prediction. Using a peptide-specific polyclonal antibody to the hSIM2-s unique region, a comprehensive immunohistochemical analysis of paraffin section-embedded tumor and normal tissues was undertaken. The majority of tumor sections analyzed stained positive. In colon and pancreatic specimens, early stage adenomas also showed hSIM2-s immunoreactivity. In prostate-related samples, hSIM2-s specific immunoreactivity was detected in almost all tumors of various Gleason scores and in prostatic intraepithelial neoplasia, but it was not detected in most stromal hyperplasia. Interestingly, in the benign prostatic hyperplasia (BPH) samples, some of the sections (20/36) showed positive staining. A subset of these retrospective BPH samples (6/6) that were matched with tumor specimens obtained from the same patient was positive for *SIM2-s* expression. It is tempting to suggest that h*SIM2-s* was activated in these BPH patients before clinical manifestation of prostate cancer. Currently, prostate-specific antigen (PSA) is the only indicator in use for prostate cancer ***(46)***, and additional markers are urgently needed. If validated with a larger cohort, the SIM2-s gene has the potential to become a predictor of risk of prostate cancer development. Assay systems similar to the systems used in PSA can be developed. In an independent prostate cancer profiling study using microarrays, *SIM2* was found to be upregulated in prostate cancer tissues, validating this gene's prostate cancer specificity ***(21)***. An example of colon tumor specificity of SIM2 consistent with the bioinformatics prediction is shown in **Fig. 5**. RT-PCR analysis of 14 different tumors and corresponding normal tissues from colon cancer patients showed the expression of *SIM2-s* gene in the tumors.

A potential drug therapy use of the SIM2-s gene is inferred using antisense knockout studies ***(44,45)***. In both colon (RKO) and pancreatic (CAPAN-1) cancer models, inhibition of SIM2-s expression by antisense resulted in apoptosis in vitro and in nude mice tumorigenicity models in vivo. The induction of apoptosis by the antisense was seen in tumor cells but not in normal renal epithelial cells, despite inhibition of *SIM2-s* expression ***(47)***. Whereas the antisense-treated RKO colon carcinoma cells did not undergo cell cycle arrest, several markers of differentiation were deregulated, including alkaline phosphatase activity, a marker of terminal differentiation. Protection of apoptosis and block of differentiation showed a correlation in the RKO model. In contrast, in normal renal epithelial cells the *SIM2-s* antisense treatment did not cause induction of differentiation. These results suggested that the targets of *SIM2-s* in tumor and

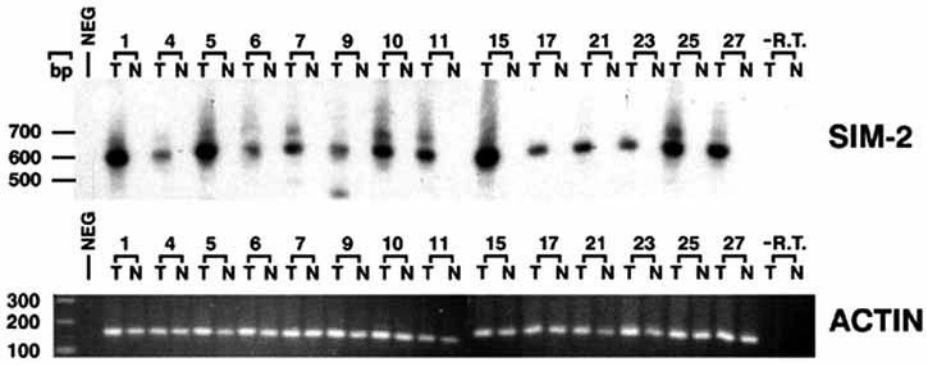


Fig. 5. Colon tumor specificity of the *SIM2-s* gene. Matched tumor and normal tissue-derived cDNAs from 14 independent colon carcinoma patients were analyzed for SIM2-s and actin gene expression by RT-PCR. The *SIM2-s*-specific PCR products were hybridized with an internal oligomer probe. M, 100-bp ladder; negative, template minus PCR control; RT, reverse transcriptase minus control.

normal cells may be different. This finding was consistent with an upregulation of a key stress response gene, Growth Arrest DNA-damage 45α in the tumor but not in the normal cells upon antisense treatment ***(47)***. The discovery of *SIM2-s* and its validation for drug therapy use offers one of the first examples of bioinformatics approaches to cancer gene discovery ***(48,49)***.

6. Identification of SIM2-s Targets by Microarray Technology

The transcription factor function of the SIM family of proteins ***(26)*** suggests a regulatory role. In addition, from the role of SIM in the midbrain an inference can be made of its potential involvement in differentiation ***(50)***. The precise molecular targets of the SIM proteins, however, are not known. The inhibition of SIM2-s expression in the RKO cells induces pronounced apoptosis within 14–24 h ***(44)***. Hence, we used global gene expression analysis to dissect the molecular targets that are affected in the antisense-treated cells. RKO colon carcinoma cells were treated in vitro with either the control or the antisense drug (100 n*M*) and at 10, 14, 18, and 24 h, RNA from the treated cells was analyzed using the Affymetrix U133A (largely known genes) Human Genome array. To develop better reliability, each RNA was analyzed using duplicate chips (chip replication) from two independent experiments (biological replicate). For experimental design, Minimum Guidelines for Experimental design, MGED guidelines (http://www.mged.org/) were followed and the entire data set point can be viewed from the ArrayExpress database (http://www.ebi.ac.uk/arrayexpress/) with the accession E-MEXP-101.

The Human Genome U133 (HG-U133) set, consisting of two GeneChip arrays (A and B), contains almost 45,000 probe sets representing more than

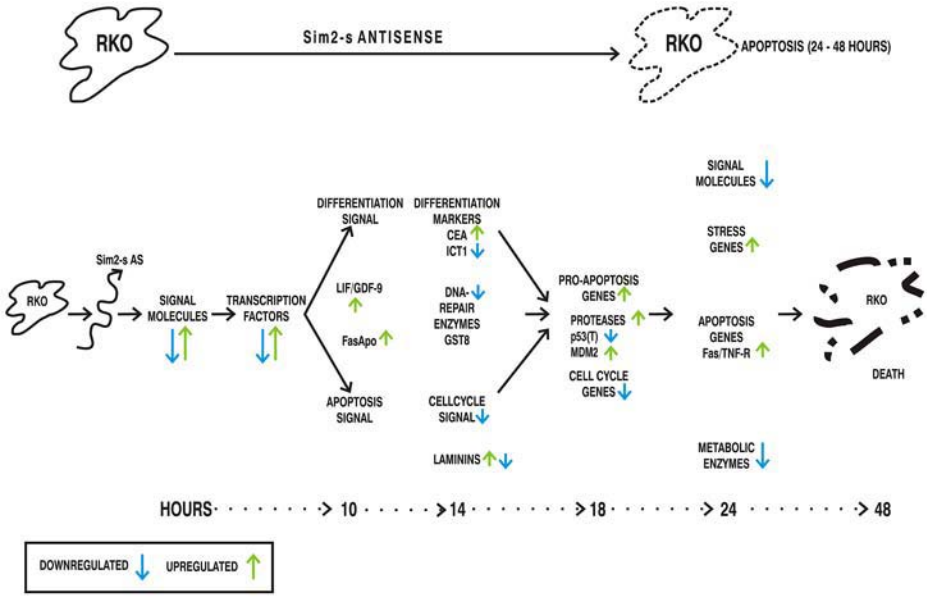


Fig. 6. Mechanism of the *SIM2-s* antisense from the GeneChip analysis. RKO colon carcinoma cells were treated with SIM2-s antisense and at the indicated time, total RNA was analyzed for a global gene expression profile with an Affymetrix U133 A and B microarray. The GeneChip output from U133A (known genes) was used to build the preliminary model as shown.

39,000 transcripts derived from approx 33,000 well-substantiated human genes. The chip output was subjected to statistical filtering (100% concordance of the hits from chip and biological replication), *p* values (>0.05) and -fold changes (>2-fold). The filtered hits were subjected to gene ontology by using GeneSpring (http://www.silicongenetics.com), and a list of genes belonging to distinct families was generated. From the pattern of the unique gene expression profile of the antisense-treated cells, a preliminary working model was generated **(Fig. 6)**. In general, a cascade of gene expression changes was seen that included early perturbation of signal transduction molecules and transcription factors. This cascade was followed by induction of differentiation signals such as leukemia inhibitory factor and growth differentiation factor. In addition, a key apoptotic signal (FasApo) was activated. Induction of differentiation markers succeeded the differentiation signals. At a later time-point, the *SIM2-s* antisense caused activation of proapoptotic genes and downregulation of proteases and tumor p53. This was accompanied by downregulation of cell cycle genes. At 24 h of treatment with the *SIM2-s* antisense, signs of stress were apparent, indicated by upregulation of stress response genes; this upregulation was accompanied by a downregulation of metabolic enzymes.

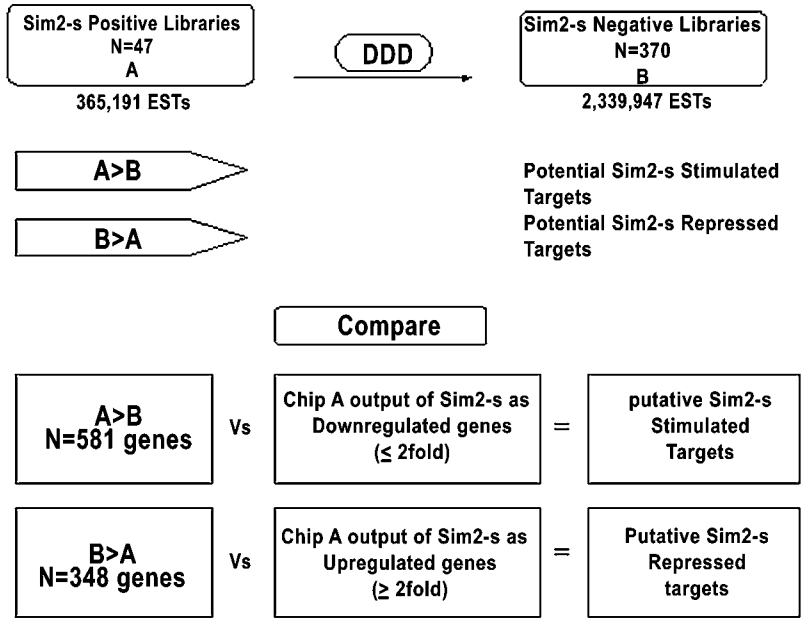


Fig. 7. DDD filter to identify SIM2-s stimulated and repressed genes. SIM2-s-positive libraries (colon, pancreas, and prostate tumor-derived) were included in pool A and SIM2-s-negative libraries (breast, ovary, and lung tumor-derived) were included in pool B. The DDD tool was used to identify SIM2-s coexpressed targets.

Efforts are currently underway to identify direct targets regulated by SIM2-s by means of bioinformatics approaches by using DDD **(Fig. 7)**. The basis for this filter is that genes expressed in tumors where SIM2-s is activated (coexpressed genes) must encompass the putative SIM2-s targets. The SIM2-s-positive (pool A) cDNA UniGene libraries can be compared with SIM2-s negative (pool B) libraries from the CGAP database by using the DDD tool from CGAP to identify SIM2-s coexpressing genes. Genes that are elevated in SIM2-s-positive libraries (A>B) are predicted to encompass SIM2-s-stimulated genes. Correspondingly, genes that are downregulated in the SIM2-s-positive libraries (B>A) are predicted to encompass potential SIM2-s-repressed genes. Among the A>B list of 581 genes, key apoptotic, survival, signal transduction molecules as well as *SIM2-s* were identified, suggesting a potential use of this filter. The antisense fingerprint of genes from the time-course experiments can be compared using this DDD filter. Comparison of A>B output with the fingerprint of antisense downregulated genes from the chip output is predicted to identify SIM2-s-stimulated targets. Conversely, comparison of B>A output with the fingerprint of antisense upregulated genes is predicted to identify SIM2-s

repressed targets. Such a strategy has been recently used to identify interleukin-8 coexpressed genes ***(51)***.

7. Potential Drawbacks

Although the aforementioned two examples provide a proof of concept for cancer gene discovery by using bioinformatics approaches, there are several areas requiring caution.

1. Most of the data sets for ESTs in the CGAP database encompass bulk tissue-derived cDNAs. These tissues are often contaminated with surrounding normal areas as well as necrotic regions that may contribute to both false positive and false negative results. In the future, use of laser capture-microdissected tissues should help alleviate this problem.
2. The diverse data-mining tools of the CGAP database use the UniGene clustering as a basis for partitioning the ESTs. However, due to the dynamic nature of the UniGene clustering, the member ESTs often are reassigned or withdrawn from the cluster. This reassignment or withdrawal may change the predicted expression specificity of an EST.
3. The sources of cDNAs from EST library versus SAGE library are often different; hence, different tools may identify different ESTs. Thus, the choice of a well validated library is essential for more effective gene discovery.
4. Multiple databases such as UniGene, SAGE, and Microarray are available for predicting electronic expression; however, the results from these databases may not show a correlation in the sources of tissues.
5. In the two examples discussed, the *CCRG/RETNLB* gene validation was consistent with the DDD prediction of colon cancer specificity. In contrast, the *SIM2* gene was specific not only to colon but also to prostate and pancreatic cancers. Thus, depending on the user's definition of stringency of target selection, the *SIM2-s* gene might or might not have scored positive.
6. The wet laboratory validation of the chosen EST for expression specificity often involves the use of patient-derived tissues. It is often difficult to define the degree of normalness in the surrounding normal tissues. Further patient-to-patient variation in gene expression contributes to loss or gain of gene expression. Hence, a statistically significant number of samples needs to be analyzed for a chosen EST before establishing the specificity of expression.
7. One of the drawbacks of GeneChip-based experiments is that the chip output can be very large (thousands of genes). Development of multiple filters at the computational level (bioinformatics) and at the biological level (system) is crucial to reducing the number of potential gene targets.

8. Conclusions

The completion of the human genome sequencing efforts promises to offer new ways to discover genes with novel diagnostic and therapeutic potential for diverse

diseases. Data-mining the cancer genome allows us to rapidly discover cancer genes as we have shown with two specific examples. Microarray technology and bioinformatics approaches can be used in conjunction to facilitate target(s) discovery and to clarify the mechanism. Currently, considerable false positives and false negatives are encountered due to the nature of the cDNA libraries. However, better integration of the databases and improvements in the quality of the EST libraries in the future will greatly improve the gene discovery process.

Acknowledgments

I thank members of my laboratory for valuable contributions and Jeanine Narayanan for editorial assistance.

References

1. Andrade, M. A. and Sander, C. (1997) Bioinformatics: from genome data to biological knowledge. *Curr. Opin. Biotechnol.* **8,** 675–683.
2. Cavalli-Sforza, L. L. (2005) The Human Genome Diversity Project: past, present and future. *Nat. Rev. Genet.* **6,** 333–340.
3. Collins, F. S., Patrinos, A., Jordan, E., Chakravarti, A., Gesteland, R., and Walters, L. (1998) New goals for the U.S. Human Genome Project: 1998-2003. *Science* **282,** 682–689.
4. Robbins, R. J. (1996) Bioinformatics: essential infrastructure for global biology. *J. Comput. Biol.* **3,** 465–478.
5. Fannon, M. R. (1996) Gene expression in normal and disease states–identification of therapeutic targets. *Trends Biotechnol.* **14,** 294–298.
6. Elek, J., Park, K. H., and Narayanan, R. (2000) Microarray-based expression profiling in prostate tumors. *In Vivo* **14,** 173–182.
7. Heller, R. A., Schena, M., Chai, A., et al. (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc. Natl. Acad. Sci. USA* **94,** 2150–2155.
8. Khan, J., Bittner, M. L., Saal, L. H., et al. (1999) cDNA microarrays detect activation of a myogenic transcription program by the PAX3-FKHR fusion oncogene. *Proc. Natl. Acad. Sci. USA* **96,** 13,264–13,269.
9. Lockhart, D. J., Dong, H., Byrne, M. C., et al. (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat. Biotechnol.* **14,** 1675–1680.
10. Lal, A., Lash, A. E., Altschul, S. F., et al. (1999) A public database for gene expression in human cancers. *Cancer Res.* **59,** 5403–5407.
11. Nacht, M., Ferguson, A. T., Zhang, W., et al. (1999) Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. *Cancer Res.* **59,** 5464–5470.
12. Strausberg, R. L., Dahl, C. A., and Klausner, R. D. (1997) New opportunities for uncovering the molecular basis of cancer. *Nat. Genet.* **15,** 415–416.
13. Wheeler, D. L., Chappey, C., Lash, A. E., et al. (2000) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **28,** 10–14.

14. Zhang, L., Zhou, W., Velculescu, V. E., et al. (1997) Gene expression profiles in normal and cancer cells. *Science* **276,** 1268–1272.
15. Schmitt, A. O., Specht, T., Beckmann, G., et al. (1999) Exhaustive mining of EST libraries for genes differentially expressed in normal and tumour tissues. *Nucleic Acids Res.* **27,** 4251–4260.
16. Scheurle, D., DeYoung, M. P., Binninger, D. M., Page, H., Jahanzeb, M., and Narayanan, R. (2000) Cancer gene discovery using digital differential display. *Cancer Res.* **60,** 4037–4043.
17. DeYoung, M. P., Damania, H., Scheurle, D., Zylberberg, C., and Narayanan, R. (2002) Bioinformatics-based discovery of a novel factor with apparent specificity to colon cancer. *In Vivo* **16,** 239–248.
18. Holcomb, I. N., Kabakoff, R. C., Chan, B., et al. (2000) FIZZ1, a novel cysteine-rich secreted protein associated with pulmonary inflammation, defines a new gene family. *EMBO J.* **19,** 4046–4055.
19. Steppan, C. M., Bailey, S. T., Bhat, S., et al. (2001) The hormone resistin links obesity to diabetes. *Nature* **409,** 307–312.
20. Steppan, C. M., Brown, E. J., Wright, C. M., et al. (2001) A family of tissue-specific resistin-like molecules. *Proc. Natl. Acad. Sci. USA* **98,** 502–506.
21. Su, A. I., Cooke, M. P., Ching, K. A., et al. (2002) Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. USA* **99,** 4465–4470.
22. He, W., Wang, M. L., Jiang, H. Q., et al. (2003) Bacterial colonization leads to the colonic secretion of RELMbeta/FIZZ2, a novel goblet cell-specific protein. *Gastroenterology* **125,** 1388–1397.
23. Kubota, T., Kawano, S., Chih, D. Y., et al. (2000) Representational difference analysis using myeloid cells from C/EBP epsilon deletional mice. *Blood* **96,** 3953–3957.
24. Blagoev, B., Kratchmarova, I., Nielsen, M. M., et al. (2002) Inhibition of adipocyte differentiation by resistin-like molecule alpha. Biochemical characterization of its oligomeric nature. *J Biol Chem.* **277,** 42,011–42,016.
25. Teng, X., Li, D., Champion, H. C., and Johns, R. A. (2003) FIZZ1/RELMalpha, a novel hypoxia-induced mitogenic factor in lung with vasoconstrictive and angiogenic properties. *Circ. Res.* **92,** 1065–1067.
26. Chrast, R., Scott, H. S., Chen, H., et al. (1997) Cloning of two human homologs of the *Drosophila* single-minded gene SIM1 on chromosome 6q and SIM2 on 21q within the Down syndrome chromosomal region. *Genome Res.* **7,** 615–624.
27. Dahmane, N., Charron, G., Lopes, C., et al. (1995) Down syndrome-critical region contains a gene homologous to *Drosophila* sim expressed during rat and human central nervous system development. *Proc. Natl. Acad. Sci. USA* **92,** 9191–9195.
28. Frazer, K. A., Tao, H., Osoegawa, K., et al. (2004) Noncoding sequences conserved in a limited number of mammals in the SIM2 interval are frequently functional. *Genome Res.* **14,** 367–372.
29. Crews, S. T. and Fan, C. M. (1999) Remembrance of things PAS: regulation of development by bHLH-PAS proteins. *Curr. Opin. Genet. Dev.* **9,** 580–587.

30. Taylor, B. L. and Zhulin, I. B. (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiol. Mol. Biol. Rev.* **63,** 479–506.
31. Ema, M., Morita, M., Ikawa, S., et al. (1996) Two new members of the murine Sim gene family are transcriptional repressors and show different expression patterns during mouse embryogenesis. *Mol. Cell Biol.* **16,** 5865–5875.
32. Michaud, J. L., DeRossi, C., May, N. R., Holdener, B. C., and Fan, C. M. (2000) ARNT2 acts as the dimerization partner of SIM1 for the development of the hypothalamus. *Mech. Dev.* **90,** 253–261.
33. Pongratz, I., Antonsson, C., Whitelaw, M. L., and Poellinger, L. (1998) Role of the PAS domain in regulation of dimerization and DNA binding specificity of the dioxin receptor. *Mol. Cell. Biol* **18,** 4079–4088.
34. Moffett, P. and Pelletier, J. (2000) Different transcriptional properties of mSim-1 and mSim-2. *FEBS Lett.* **466,** 80–86.
35. Probst, M. R., Fan, C. M., Tessier-Lavigne, M., and Hankinson, O. (1997) Two murine homologs of the *Drosophila* single-minded protein that interact with the mouse aryl hydrocarbon receptor nuclear translocator protein. *J. Biol. Chem.* **272,** 4451–4457.
36. Swanson, H. I., Chan, W. K., and Bradfield, C. A. (1995) DNA binding specificities and pairing rules of the Ah receptor, ARNT, and SIM proteins. *J. Biol. Chem.* **270,** 26,292–26,302.
37. Ooe, N., Saito, K., Mikami, N., Nakatuka, I., and Kaneko, H. (2004) Identification of a novel basic helix-loop-helix-PAS factor, NXF, reveals a Sim2 competitive, positive regulatory role in dendritic-cytoskeleton modulator drebrin gene expression. *Mol. Cell. Biol.* **24,** 608–616.
38. Yamaki, A., Kudoh, J., Shimizu, N., and Shimizu, Y. (2004) A novel nuclear localization signal in the human single-minded proteins SIM1 and SIM2. *Biochem. Biophys. Res. Commun.* **313,** 482–488.
39. Madden, S. L., Cook, D. M., Morris, J. F., Gashler, A., Sukhatme, V. P., and Rauscher, F. J., III (1991) Transcriptional repression mediated by the WT1 Wilms tumor gene product. *Science* **253,** 1550–1553.
40. Franks, R. G. and Crews, S. T. (1994) Transcriptional activation domains of the single-minded bHLH protein are required for CNS midline cell development. *Mech. Dev.* **45,** 269–277.
41. Mermod, N., O'Neill, E. A., Kelly, T. J., and Tjian, R. (1989) The proline-rich transcriptional activator of CTF/NF-I is distinct from the replication and DNA binding domain. *Cell* **58,** 741–753.
42. Hanna-Rose, W. and Hansen, U. (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet.* **12,** 229–234.
43. DeYoung, M. P., Scheurle, D., Damania, H., Zylberberg, Z., and Narayanan, R. (2002) Down's syndrome-associated Single Minded gene as a novel tumor marker. *Anticancer Res.* **22,** 3149–3158.
44. DeYoung, M. P., Tress, M., and Narayanan, R. (2003) Identification of Down's syndrome critical locus gene SIM2-s as a drug therapy target for solid tumors. *Proc. Natl. Acad. Sci. USA* **100,** 4760–4765.

45. DeYoung, M. P., Tress, M., and Narayanan, R. (2003) Down's syndrome-associated Single Minded 2 gene as a pancreatic cancer drug therapy target. *Cancer Lett.* **200,** 25–31.
46. Goolsby, M. J. (2001) Use of PSA measurement in practice. *J. Am. Acad. Nurse Pract.* **13,** 246–248.
47. Aleman, M. J., DeYoung, M. P., Tress, M., Keating, P., Perry, G. W., and Narayanan, R. (2005) Inhibition of single minded 2 gene expression mediates tumor-selective apoptosis and differentiation in human colon cancer cells. *Proc. Natl. Acad. Sci. USA* **102,** 12,765–12,770.
48. Ratan, R. R. (2003) Mining genome databases for therapeutic gold: SIM2 is a novel target for treatment of solid tumors. *Trends Pharmacol. Sci.* **24,** 508–510.
49. Touchette, N. (2003) Mouse-to-mouse revelation: genome yields cancer drug target. Genome News Network. http://www.genomenewsnetwork.org/articles/04_03.mouse.shtml. April 18, 2003.
50. Nambu, J. R., Lewis, J. O., Wharton, K. A., Jr., and Crews, S. T. (1991) The *Drosophila* single-minded gene encodes a helix-loop-helix protein that acts as a master regulator of CNS midline development. *Cell* **67,** 1157–1167.
51. Benbow, L., Wang, L., Laverty, M., et al. (2002) A reference database for tumor-related genes co-expressed with interleukin-8 using genome-scale in silico analysis. *BMC Genomics* **3,** 29.

3

Analysis of Gene Networks for Drug Target Discovery and Validation

Seiya Imoto, Yoshinori Tamada, Christopher J. Savoie, and Satoru Miyano

Summary

Understanding responses of the cellular system for a dosing molecule is one of the most important problems in pharmacogenomics. In this chapter, we describe computational methods for identifying and validating drug target genes based on the gene networks estimated from microarray gene expression data. We use two types of microarray gene expression data: gene disruptant microarray data and time-course drug response microarray data. For this purpose, the information of gene networks plays an essential role and is unattainable from clustering methods, which are the standard for gene expression analysis. The gene network is estimated from disruptant microarray data by the Bayesian network model, and then the proposed method automatically identifies sets of genes or gene regulatory pathways affected by the drug. We use an actual example from analysis of *Saccharomyces cerevisiae* gene expression profile data to express a concrete strategy for the application of gene network information toward drug target discovery.

Key Words: Bayes statistics; Bayesian network; Boolean network; drug target; gene network; microarray data.

1. Introduction

In recent years, microarray technology has produced a large volume of genome-wide gene expression data under various experimental conditions, such as gene disruptions, gene overexpressions, shocks, cancer cells, and so on (Chapters 4 and 5, Volume 1). Based on this new data production, there have been considerable attempts to estimate gene networks from such gene expression data, and several computational methods have been proposed together with mathematical models for gene networks, such as Boolean networks ***(1–3)***, differential equation models

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarker Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

(4,5), and Bayesian networks *(6–10)*. Although the paradigm based on microarray technology with the clustering technique has made tremendous impacts on biomedical research and practice, the strategy enhanced with computational gene network analysis has not yet been well examined for practical applications. In this chapter, we focus on a real application of the gene network analysis aimed at understanding responses of the cellular system against a drug *(11–13)*.

For identifying and validating drug target genes, the following three elements are essential: (1) drug-affected genes, (2) druggable genes, and (3) drug-active pathways. The definitions are as follows:

- Drug-affected genes are the candidates directly affected by the drug.
- Druggable genes regulate the drug-affected genes most strongly from upstream of the gene network.
- Drug-active pathways are regulatory pathways perturbed by the drug.

These three types of information play an essential role in pharmacogenomics and are unattainable from clustering methods, which are the standard for gene expression analysis. This information can only be reached by elucidating the gene network.

For identifying the drug-affected genes and druggable genes, **Fig. 1** shows the conceptual view of our strategy. From **Fig. 1**, it can be seen that the information of gene networks plays an essential role. To create this information, we use two types of microarray gene expression data: data obtained by single gene disruptions, and data obtained by biological experiments of several dose and time responses to the drug. We use two computational methods, virtual gene technique and Bayesian networks, for extracting network information from these data. The details are described in **Subheading 3**. The computer software we used for efficient exploration of the estimated gene networks is in **Note 1**.

Because the drug-affected genes are candidate genes that are directly affected by the drug, and the druggable genes potentially regulate the drug-affected genes, this information is very important for identifying and validating drug target genes. However, the drug-active pathway is not taken into account in **Fig. 1**. Because the drug-active pathway plays an important role in understanding the drug responses of the cellular system affected by the drug and the prediction of the side effects, it is also a very important pathway for identifying and validating drug target genes. The conceptual view of the computational strategy for identifying drug-active pathways from the same microarray data in **Fig. 1** is shown in **Fig. 2**.

Our approach for identifying drug-active pathways involves mainly two methods: (1) estimation of the gene network by using a Bayesian network model from microarray data obtained by single gene disruptions, and (2) identification of the drug-active pathways from time-course drug response microarray data by the probabilistic inference on the gene network estimated

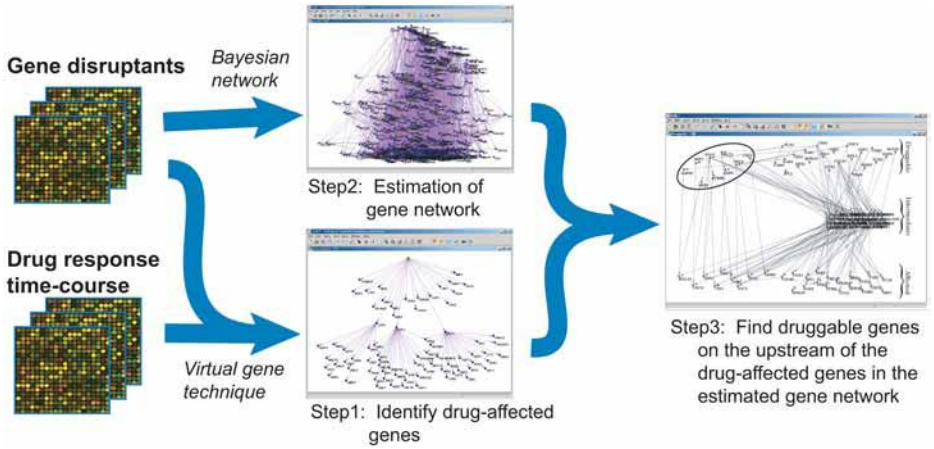


Fig. 1. Conceptual view of the identification of drug-affected and druggable genes ***(11)***.

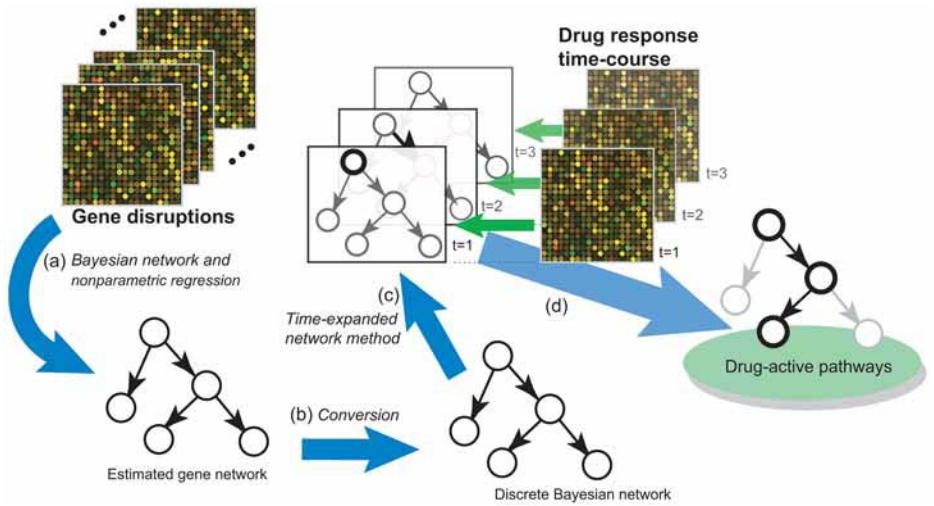


Fig. 2. Conceptual view of the identification of drug-active pathways ***(13)***.

in method 1. The drug-active pathways are identified as the subnetworks of the estimated gene network. The information of time dependency on drug response time-course data is used for reducing the number of falsely identified pathways. By highlighting the genes on the network for each time-point, we can observe the flow of the drug responses of genes in the network.

We demonstrate the whole process of our method for identifying the drug-affected genes, the druggable genes, and the drug-active pathways by using 120 disruptant microarray gene expression data and drug response time-course microarray data for an antifungal drug, griseofulvin, for *Saccharomyces cerevisiae*.

2. Materials

Two types of cDNA microarray measurements of *S. cerevisiae* have been prepared for our purpose. One type is the microarray data obtained by gene disruptions, and the other type is time-course data of responses to an antifungal medicine. The number of disruptants is 120 for the first type of data, and the time-course data for one dose consists of five microarrays.

2.1. Microarray Data of Single Gene Disruptions

The first microarray data are obtained by disrupting 120 genes, where mostly transcription factors are disrupted. The BY4741 (*MATa, HIS3D1, LEU2D0, MET15D0, URA3D0*) is used as the wild-type strain, and purchased gene disruptions were from Research Genetics (Huntsville, AL). The disrupted genes are listed in **Table 1**. To monitor the gene expression profile, cells were inoculated and grown in YPD medium (1% yeast extract, 1% bacto-peptone, 2% glucose) at 30ºC until optical density$_{600}$ reached 1.0 in the logarithmic growth phase, and then cells were harvested to isolate mRNA for assay of gene expression. The respective parental strain was the control used for each disruptant strain. *See* Aburatani et al. ***(14)*** for details. The total RNAs are extracted from these experiments and labeled by Cy5. These Cy5-labeled RNAs are hybridized with Cy3-labeled RNAs from nontreated cells. A microarray then measures 5871 gene expression levels at once.

2.2. Drug Response Time-Course Microarray Data

To determine the previously unknown underlying molecular affects of the popular generic antifungal agent griseofulvin, we created time-course drug response microarray data. Griseofulvin is a widely prescribed oral antifungal agent that is indicated primarily for severe fungal infections of the hair and nails. Although griseofulvin's molecular action is unknown, it is known that the drug disrupts mitotic structure in fungi, leading to metaphase arrest.

We incubated yeast cultures in dosages of 10, 25, 50, and 100 mg/mL antifungal medication in culture and took aliquots of the culture at five time-points (0, 15, 30, 45, and 60 min) after addition of the agent. Here, time 0 means the start point of this observation and just after exposure to the drug. We then extracted the total RNAs from these experiments, labeled the RNAs with Cy5, hybridized them with Cy3-labeled RNAs from nontreated cells, and applied them to full genome cDNA microarrays, thereby creating a data set of 20 microarrays for drug response data.

2.3. Data Normalization

It is known that expression profile data measured by microarrays have several systematic biases, such as spotting position-specific bias, intensity dependent

Table 1
List of 120 Disrupted Genes

ACE2, ACH1, ADR1, ARGR2, ASG7, ASH1, BAR1, BEM4, BGL2, BOI1, BTT1, BUD4, BUD6, BUD8, BUD9, CAD1, CAD1, CAT8, CIN5, CNB1, CPR6, CUP9, CWP1, DAL82, DOT6, DST1, ECM22, EGD1, EGD2, FIR1, FZF1, GAL4, GAL80, GAT3, GCN4, GCS1, GZF3, HAC1, HAP2, HAP3, HAT2, HMS2, HMS2, HPA2, HPA3, HST2, HST4, HTA2, HTB2, INO2, INO4, KRE1, LEU3, LSM1, LYS14, MAK31, MBR1, MET28, MIG2, MRS1, MUC1, NRG1, OPI1, OSH1, PIP2, RCK2, RGM1, RGT1, RIC1, RLM1, RME1, RMS1, RPD3, RSC1, RTG1, SCW10, SCW11, SCW4, SDS3, SFL1, SIP2, SKI8, SKN7, SMP1, SPS18, SPT23, SPT8, STB1, STB3, STB4, STB6, STD1, SUM1, SWI5, SWI6, SWR1, TEA1, THI2, THI20, TPK1, TSP1, TUB3, UME1, UME6, YAP7, YAR003W, YBL036C, YBL054W, YDR340W, YER028C, YER130C, YFL052W, YLL054C, YLR266C, YML076C, YML081W, YNR063W, YPL060W, YPR125W, YPR196W

bias, and so on. For extracting reliable information from microarray data, we need to perform procedures called normalization on such biased data to remove any systematic and artificial biases. We performed the lowess normalization method to expression values in each pen group and then also globally normalized the microarray-transformed intensities by the lowess method. The former method aims at removing the position-specific bias, and the latter method aims at removing the intensity-dependent bias. For microarray data normalization, Quackenbush ***(15)*** gives an excellent overview.

3. Methods

3.1. Virtual Gene Technique for Identifying Drug-Affected Genes

3.1.1. Virtual Gene Technique

The first task is to find genes directly affected by the drug (**Fig. 1,** step 1). For this purpose, the most straightforward approach might be the fold-change analysis of the drug response microarray data. However, to perform more accurate screening of the candidate drug-affected genes, we use another method, virtual gene technique, which is more suited for inferring qualitative relations between genes. We regard the drug as a "virtual gene," and we consider that the state of this virtual gene is 1 (ON) if the drug is dosed; otherwise, it is 0 (OFF). The idealistic approach for identifying the genes directly affected by the virtual gene may be to use the Boolean network model and to apply the method developed by Akutsu et al. ***(1)*** for inferring Boolean network model that can suggest a series of mutants for identifying the network. However, it requires multiple disruptions and overexpressions for one mutant, and the number of mutants required for identifying the network is not realistic, even for a small network. Therefore, this approach is not in our scope, because we can deal with

	D.gene A	D.gene B	D.gene C	D.gene D	Drug 1
gene A		Not affected	Not affected	Not affected	Affected
gene B	Affected		Not affected	Not affected	Affected
gene C	Affected	Not affected		Affected	Affected
gene D	Affected	Not affected	Affected		Affected

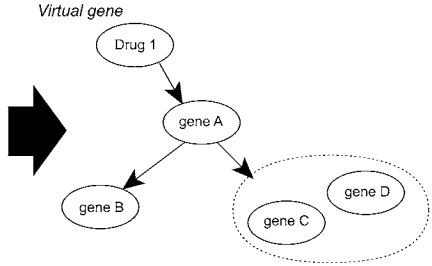


Fig. 3. Graphical view of the virtual gene technique. "D.gene A" means this microarray is observed by disrupting gene A. The dotted circle shows the equivalence class ***(11)***.

only single gene disruptants in our measurement experiment. Instead, we focus on a simpler network model whose structure is a multilevel directed acyclic graph (dag). Maki et al. ***(16)*** also have proposed a naive algorithm for constructing a multilevel dag based on information on how a single gene disruption affects other genes. We use this method for finding genes directly affected by the drug by regarding the drug as a gene, and we call this method the virtual gene technique. Here, we briefly explain the method.

Let $V=\{g_1,\ldots, g_p\}$ be the set of all genes and $D_m=\{d_1,\ldots, d_m\} \subseteq V$ be the set of genes to be disrupted. We assume D_m contains the virtual genes corresponding to the drug. Our cDNA microarray compares the gene expression level of a mutant with that of a wild type for each gene. From a gene disruption experiment for gene d_i, we obtain a microarray data $E_{d_i}[g_j]$ $(1 \le j \le p)$. Then, by setting a threshold, η, we define a relation R as follows:

$$R(a, b) = \begin{cases} 1 \text{ if } a \in D_m \text{ and } E_a[b] > \eta \text{ or } E_a[b] < 1/\eta. \\ 0 \text{ otherwise.} \end{cases}$$

By using the transitive closure R^* of R, we define an equivalence relation $\equiv_{R^*}$ on V by $a \equiv_{R^*} b$ if and only if $a, b \in D_m$ and $\{R^*(a, b) = 1\} \wedge \{R^*(b, a) = 1\}$. Then, a new relation $\hat{R}$ on the equivalence classes of $\equiv_{R^*}$ is defined by $R([a], [b]) = R^*(a, b)$ for $a, b \in V$. Note that $\hat{R}$ is well defined by the definition of $\equiv_{R^*}$. $\hat{R}$ defines a dag $\hat{G}$ on the set of equivalence classes, where self-loop edges are ignored. An indirect edge $([a], [b])$ of $\hat{G}$ is an edge such that there is another path from $[a]$ to $[b]$ in $\hat{G}$. By removing all indirect edges from $\hat{G}$, we obtain a multilevel dag. *See* Maki et al. ***(16)*** and Aburatani et al. ***(14)*** for more details about the dag construction. **Figure 3** shows an example of the resulting network based on the virtual gene technique. There are four genes, and their states for each disruptant and a drug are summarized as the matrix. We can find gene C $\equiv_{R^*}$ gene D and remove indirect edges (drug 1, gene B), (drug 1, gene C), and so on, and then a network having the virtual drug node as the root is obtained.

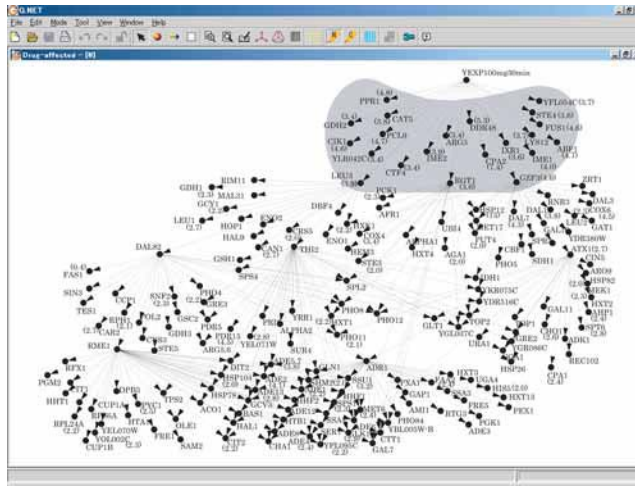


Fig. 4. Downstream pathway of the virtual gene YEXP100mg30min. The shadowed genes are affected by the drug ***(11)***.

Finally, by considering the dag whose root is the virtual gene, the children of this virtual gene would be the candidate genes directly affected by the drug. **Note 2** gives more comments.

3.1.2. Application of Virtual Gene Technique

Figure 4 is the result of the virtual gene technique for the drug corresponding to "100mg and 30min." The root node is the virtual gene corresponding to the drug, and the genes in the shadowed area are the candidates of the drug-affected genes. By applying the virtual gene technique to each drug response data and disruptant microarray data, we can obtain the candidates of the drug-affected genes. **Figure 5** is the superimposed graph of virtual gene technique generated for five separate time and dosage griseofulvin exposure experiments. **Figure 5** suggests that the *CIK1* (Gene Ontology [GO] annotation [function]:microtubule motor activity) expression is significantly affected under exposure to griseofulvin at each dosage and time point and can be viewed as a drug-affected genes with high reliability (**Notes 3** and **4**).

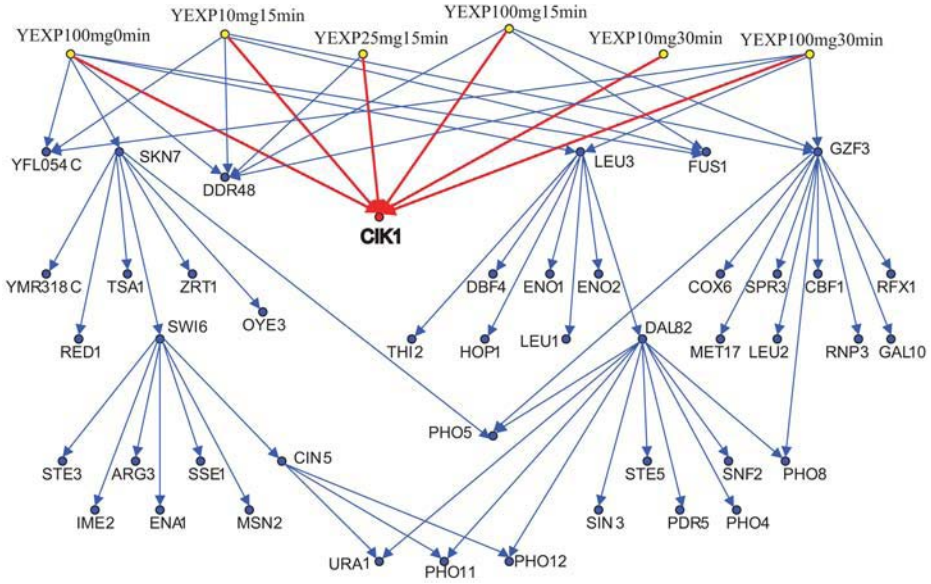


Fig. 5. Result of the identification of the drug-affected genes ***(12)***.

3.2. Bayesian Networks for Exploring Druggable Genes

3.2.1. Introduction of Bayesian Networks

Bayesian networks are a mathematical model for representing complex phenomena among a large number of random variables by using the joint probability. In Bayesian networks, the dependency of the random variables is represented by the dag encoding the Markov assumption between nodes. That is, the state of a node only depends on its direct parents. Mathematically, we have a set of random variables $\{X_1,\ldots,X_p\}$, and we consider a dag G as the dependency underling among random variables. The joint probability of the random variables can be decomposed as follows:

$$\Pr(X_1,\ldots,X_p)=\prod_{j=1}^{p}\Pr(X_j \mid Pa(X_j)), \tag{1}$$

where $Pa(X_j)$ is the set of random variables corresponding to the direct parents of X_j in G. **Figure 6** is an example of a dag of the random variables $\{X_1,X_2,X_3,X_4\}$. In this case, we have $Pa(X_1)=\{X_2,X_3\}$, $Pa(X_2)=\varnothing$, $Pa(X_3)=X_4$, and $Pa(X_4)=\varnothing$. Then, the joint probability of these four random variables is decomposed as:

$$\Pr(X_1,\ldots,X_4)=\Pr(X_1|X_2,X_3)\Pr(X_2)\Pr(X_3|X_4)\Pr(X_4).$$

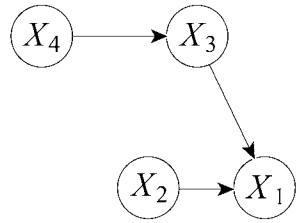


Fig. 6. Directed acyclic graph.

In gene network inference by Bayesian networks, a gene is regarded as a random variable and represented by a node in a dag, and the relationship between a gene and its direct parents is represented as a conditional probability $\Pr(X_j|Pa(X_j))$. Boolean networks are a deterministic system that explains a state of a gene by the Boolean function of the states of its direct parents. On the contrary, the Bayesian networks use a probabilistic nature to model the dependencies among genes. Because gene expression data measured by microarrays contain various kinds of noise, the probabilistic formulation in the Bayesian networks is advocated to draw more reliable information from microarray data.

The decomposition in **Eq. 1** holds when the structure of the directed acyclic graph G is given. However, for gene network estimation problem from microarray data, many parts of the true structure of the gene network are still unknown, and we need to estimate them from microarray data. This structural learning problem can be viewed as a statistical model selection problem ***(17)*** and can be solved by using an information criterion, such as the Akaike information criterion, the Bayesian information criterion, and so on. In this chapter, we describe a method to select a network structure based on the Bayes approach. In the Bayes statistics, the optimal graph can be selected as the maximizer of the posterior probability of the graph conditional on the data.

Suppose we have n independent observations $D = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}$ for $\{X_1,\ldots,X_p\}$, where the ith observation $\boldsymbol{x}_i = (x_{i1},\ldots,x_{ip})^t$ is the p-dimensional vector. Here $\boldsymbol{a}^t$ indicates the transpose of the vector $\boldsymbol{a}$. In our case, the posterior probability of the graph can be written as follows:

$$\pi_{post}(G \mid D) = \frac{\pi_{prior}(G)\, p(D \mid G)}{p(D)},$$

where $\pi_{prior}(G)$ is the prior probability of the graph, $p(D)$ is the normalizing constant given by $\sum_G \pi_{prior}(G)\, p(D \mid G)$ and $p(D \mid G)$ is the marginal likelihood obtained by

$$p(D \mid G) = \int p(D,\theta \mid G)d\theta = \int p(D \mid \theta,G)p(\theta \mid G)d\theta. \quad (2)$$

Here, $p(D \mid \boldsymbol{\theta}, G)$ is the likelihood function, and $p(\boldsymbol{\theta} \mid G)$ is the prior distribution on the parameter $\boldsymbol{\theta}$. Because the normalizing constant $p(D)$ does not depend on G, we can find

$$\pi_{post}(G|D) \propto \pi_{prior}(G)p(D|G). \quad (3)$$

Hence, for computing the posterior probability for structural learning of the Bayesian networks, we need to compute the high-dimensional integral in **Eq. 2**. For learning Bayesian networks, **Note 5** gives some information.

3.2.2. Bayesian Networks for Discrete Data

Suppose that we have a set of random variables $\{X_1,\ldots,X_p\}$ and that a random variable takes one of m values $\{u_1,\ldots,u_m\}$. Then, we put the conditional probability as

$$\theta_{jkl} = \Pr(X_j = u_k \mid Pa(X_j) = \boldsymbol{u}_{jl})$$

for $j = 1,\ldots,p$; $k = 1,\ldots,m$; $l = 1,\ldots,m^{|Pa(X_j)|}$. Note that $\sum_{k=1}^{m}\theta_{jkl} = 1$ holds. Here, $\boldsymbol{u}_{jl}$ is the lth entry of the state table of $Pa(X_j)$. For example, in **Fig. 4**, $Pa(X_1) = \{X_2, X_3\}$ contains m^2 entries: $\boldsymbol{u}_{11} = (u_1, u_1)$, $\boldsymbol{u}_{12} = (u_1, u_2),\ldots,\boldsymbol{u}_{1m^2} = (u_m, u_m)$. In this case, we can assume that $X_j \mid Pa(X_j) = \boldsymbol{u}_{jl}$ follows the multinomial distribution with probabilities $\theta_{j1l},\ldots,\theta_{jml}$.

Suppose we have n independent observations $D = \{\boldsymbol{x}_1,\ldots,\boldsymbol{x}_n\}$ for $\{X_1,\ldots,X_p\}$, where x_{ij} is one of the m values $\{u_1,\ldots,u_m\}$. Based on these observations, the likelihood function of $\theta = (\theta_{jkl})_{j,k,l}$ is written by

$$p(D \mid \theta, G) = \prod_j \prod_k \prod_l \theta_{jkl}^{N_{jkl}},$$

where N_{jkl} is the number of pairs satisfying $(X_j, Pa(X_j)) = (u_k, \boldsymbol{u}_{jl})$ *in* D. In the computation of the marginal likelihood $p(D|G)$ given in **Eq. 2** for the Bayesian networks for discrete data, the Dirichlet distribution is usually used as the prior distribution on the parameter $p(\boldsymbol{\theta}|G)$. Suppose that the prior distribution on $\boldsymbol{\theta}$ can be decomposed as $p(\theta \mid G) = \prod_j \prod_l p(\theta_{jl} \mid \alpha_{jl})$, where $\boldsymbol{\theta}_{jl} = (\theta_{j1l},\ldots,\theta_{jml})^t$ and $\boldsymbol{\alpha}_{jl} = (\alpha_{j1l},\ldots,\alpha_{jml})^t$. Here, $\boldsymbol{\alpha}_{jl}$ is called a hyperparameter vector. We assume the Dirichlet prior for $p(\boldsymbol{\theta}_{jl}|\boldsymbol{\alpha}_{jl})$ as

$$p(\theta_{jl}|\alpha_{jl}) = \frac{\Gamma(\alpha_{jkl})}{\prod_k \Gamma(\alpha_{jk'l})} \prod_k \theta_{jkl}^{\alpha_{jkl}-1},$$

where Γ (·) is the gamma function. Then, the marginal likelihood $p(D|G)$ can be analytically computed as

$$p(D \mid G) = \prod_j \prod_l \frac{\Gamma(\alpha_{j\cdot l})}{\Gamma(\alpha_{j\cdot l} + N_{j\cdot l})} \prod_k \frac{\Gamma(\alpha_{jkl} + N_{jkl})}{\Gamma(\alpha_{jkl})}, \quad (4)$$

where $\alpha_{j\cdot l} = \sum_k \alpha_{jkl}$ and $N_{j\cdot l} = \sum_k N_{jkl}$. By replacing $P(D \mid G)$ in **Eqs. 3** by **4**, we obtain $\pi_{post}(G \mid D)$ for the multinomial distribution and the Dirichlet prior to perform the structural learning of the Bayesian networks for discrete data. This criterion is called the BDe metric originally derived by Cooper and Herskovits ***(18)***.

3.2.3. Bayesian Networks for Continuous Data

Let X_j be a continuous type random variable. This situation is more realistic for gene network inference problem from microarray data, because microarray data essentially take continuous variables. In this sense, when we apply the Bayesian network for discrete data to microarray data, we need to discretize expression data into several categories. The number of categories is usually set to be three, i.e., "suppressed," "unchanged," and "overexpressed." However, actually the number of categories is a parameter to be estimated, and we should notice that the discretization leads to information loss. Therefore, in microarray data analysis, a computational method that can handle microarray data as continuous data is advocated. As for discretization of gene expression data, several methods are listed in **Note 6**.

The decomposition of the joint probability can be hold for continuous variables by using densities instead of the probabilistic measure and we have

$$f(x_{i1}, \ldots, x_{ip} \mid \theta) = \prod_j f_j\ (x_{ij} \mid pa(X_j)_i,\ \theta_j),$$

where $f, f_1, \ldots, f_p$ are densities, $\theta = (\theta_1^t, \ldots \theta_p^t)^t$ is the parameter vector, and $pa(X_j)_i$ is the set of observations of $Pa(X_j)$ measured by ith microarray. The likelihood function is then given by

$$p(D \mid \theta, G) = \prod_i \prod_j f_j(x_{ij} \mid pa(X_j)_i,\ \theta_j).$$

Therefore, in the Bayesian networks for continuous data, the construction of the conditional densities $f_j(x_{ij} \mid pa(X_j)_i,\ \theta_j)$ is crucial, and this is essentially the same as the regression problem.

The linear regression model can be used to construct the conditional density ***(6)*** and is written as

$$x_{ij} = \beta_0 + \beta_1 p_{i1}^{(j)} + \cdots + \beta_{q_j} p_{iq_j}^{(j)} + \varepsilon_{ij}, \text{ where } pa(X_j)_i = (p_{i1}^{(j)}, \ldots, p_{iq_j}^{(j)})^t,$$

β_k (k = 1,…,q_j) are parameters and ε_{ij} (i = 1,…,n) are independently and normally distributed with mean 0 and variance σ_j^2. This model assumes the relationships between genes are linear, and it is unsuitable to extract effective information from the data with complex structure. To capture even nonlinear dependencies, Imoto et al. ***(9)*** proposed the use of the nonparametric additive regression model ***(19)*** of the form

$$x_{ij} = m_{j1}(p_{i1}^{(j)}) + \cdots + m_{jq_j}(p_{iq_j}^{(j)}) + \varepsilon_{ij},$$

where $m_{jk}(\cdot)$ (k =1,…,q_j) are smooth functions from R to R. We construct $m_{jk}(\cdot)$ by the basis function expansion method with B-splines ***(20,21)***:

$$m_{jk}(p) = \sum_{s=1}^{M_{jk}} \gamma_{sk}^{(j)} b_{sk}^{(j)}(p),$$

where $\gamma_{sk}^{(j)}$ ($s = 1,\ldots,M_{jk}$) are parameters, $\{b_{1k}^{(j)}(\cdot),\cdots\cdot b_{M_{jk}k}^{(j)}(\cdot)\}$ is the prescribed set of B-splines, and M_{jk} is the number of B-splines. We then have the Bayesian network and nonparametric regression model

$$f_j(x_{ij} \mid pa(X_j)_i, \theta_j) = \frac{1}{\sqrt{2\pi\sigma_j^2}} \exp\left[-\frac{\left\{ x_{ij} - \sum_k \sum_s \gamma_{sk}^{(j)} b_{sk}^{(j)}(p_{ik}^{(j)}) \right\}^2}{2\sigma_j^2} \right].$$

Note that the Bayesian network model based on the linear regression is included in this model as a special case. To extend from the additive regression model to a general regression, see **Note 7**.

Let the prior distribution on the parameter $\boldsymbol{\theta}$ be specified by the hyperparameter $\boldsymbol{\lambda}$ and let $\log p(\boldsymbol{\theta} \mid \boldsymbol{\lambda}) = O(n)$, then the Laplace approximation for integrals ***(22–24)*** gives an analytical solution for computing $p(D|G)$ as follows:

$$p(D \mid G) = (2\pi / n)^{r/2} \mid J_\lambda(\hat{\theta} \mid D) \mid^{-1/2} \exp\{n l_\lambda(\hat{\theta} \mid D)\}\{1 + O_p(n^{-1})\},$$

where

$$l_\lambda(\theta \mid D) = \frac{1}{n} \{\log p(D \mid \theta,\ G) + \log p(\theta \mid \lambda)\},$$

$$J_\lambda(\theta \mid D) = -\frac{\partial^2}{\partial\theta\partial\theta^t} l_\lambda(\theta \mid D),$$

r is the dimension of $\boldsymbol{\theta}$, and $\hat{\boldsymbol{\theta}}$ is the mode of $l_\lambda(\boldsymbol{\theta} \mid D)$. Based on the result of the Laplace approximation, Imoto et al. ***(9)*** derived a criterion named Bayesian network and nonparametric regression criterion (BNRC) for choosing the optimal graph:

$$BNRC(G) = -2\log \pi_{prior}(G) - r \log(2\pi / n) + \log |J_\lambda(\hat{\theta} \mid D)| - 2nl_\lambda(\hat{\theta} \mid D). \quad (5)$$

The optimal graph $\hat{G}$ is chosen such that the criterion defined in **Eq. 5** is minimal. Imoto et al. ***(10)*** also extended the results of Imoto et al. ***(9)*** to handle the nonparametric heteroscedastic regression.

Recently, research is underway to estimate gene networks from multisource biological information (*see* **Note 8)**. Also, the Bayesian networks force us to consider gene networks as dags, whereas the real biological systems have cyclic regulation. Some research has dealt with cyclic regulations based on the dynamic models (*see* **Note 9)**.

3.2.4. Application of Bayesian Networks for Identifying Druggable Genes

We selected 735 genes from the yeast genome for identifying drug targets based on the 120 gene disruptant data and drug response time-course data. These genes were selected based on the following strategy. We collected 314 genes that are known as transcription factors. Ninety-eight of these 314 genes have already been studied for their control mechanisms. The expression data for 735 genes chosen for our analysis includes the genes controlled by these 98 transcription factors from 5871 genes in addition to nuclear receptor-like genes, which have a pivotal role in gene expression regulation and are popular drug targets. We have constructed the Bayesian network models described in **Subheading 3.2.3.** of these 735 genes from 120 gene disruption conditions. The druggable genes are the drug targets related to these drug-affected genes, which we want to identify for the development of novel leads. We can explore the druggable genes upstream of the drug-affected genes in the estimated gene network by the Bayesian network method. Here, we focus on the nuclear receptor-like genes as the druggable genes for two reasons. (1) In general, nuclear receptor proteins are known to be useful drug targets, and together they represent more than 20% of the targets for medications presently on the market. (2) Nuclear receptors are involved in transcription regulatory affects that are directly measured in cDNA microarray experiments.

Figure 7 shows a partial resulting network. The druggable genes (top) that are the nuclear receptor-like genes in the circle connect directly to the drug-affected genes, and other druggable genes have one intermediary gene per

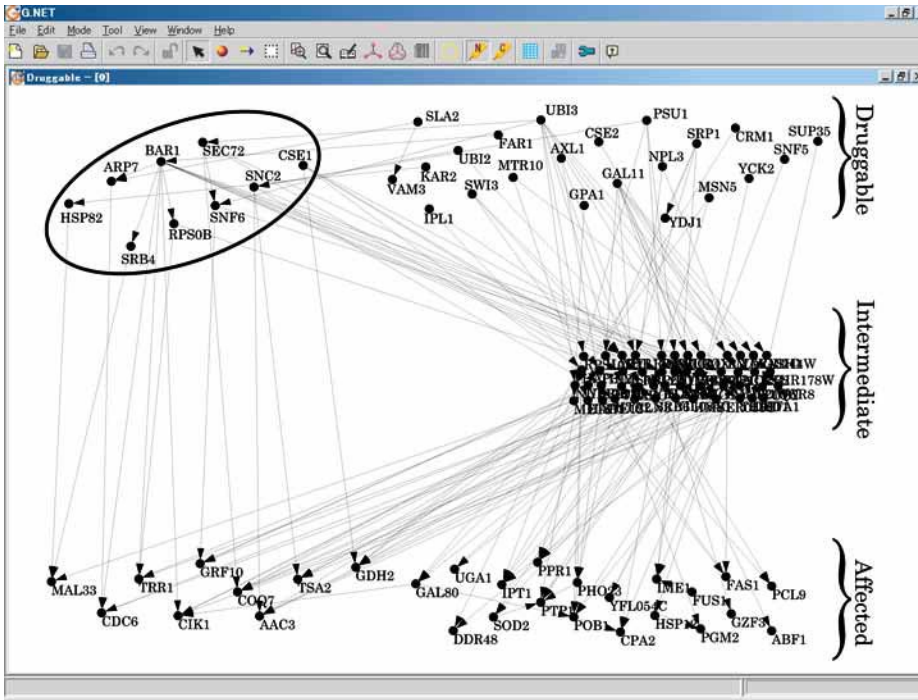


Fig. 7. Partial resulting network among the druggable (top), drug-affected (bottom), and intermediary genes (middle) ***(11)***.

one druggable gene. But, we can find more pathways from the druggable genes to the drug-affected genes if we admit more intermediary genes. By using the Bayesian network model, we can determine the reliability of each edge and select a more reliable pathway, an advantage of the Bayesian network model in searching for ideal druggable targets. For information on measuring reliability of an edge, see **Note 10**.

We can find the druggable genes for each drug-affected gene, e.g., we can find the druggable genes for *MAL33* and *CDC6* shown in **Table 2**. The drug-affected gene *CDC6* in **Table 2** is a protein that regulates the initiation of DNA replication. It binds to origins of replication at the end of mitosis, directing the assembly and disassembly of minichromosome maintenance proteins and the prereplication complex. It is a member of the AAA+ family of ATPases. The genetic mechanism and effectiveness of this antifungal medication was made clear by this result. It was determined that the localization of CDC6p is nuclear; therefore, any other extracellular molecule, such as a drug, cannot affect it directly. Thus, our result through gene network analysis indicates that *CDC6* is influenced by the extracellular molecule.

Table 2
Druggable Genes of MAL33 and CDC6

	Druggable	
Drug-affected	Parents[a]	Grandparents[a]
MAL33 (YBR297W): Maltose fermentation regulatory protein	HSP82 (YPL240C): heat shock protein SRB4 (YER022W): DNA-directed RNA polymerase II holoenzyme and Kornberg's mediator (SRB) subcomplex subunit	BAR1 (YIL015W): barrierpepsin precursor GPA1 (YHR005C): GTP-binding protein α subunit of the pheromone pathway KAR2 (YJL034W): nuclear fusion protein
CDC6 (YJL194W): cell division control protein	ARP7 (YPR034W): component of SWI-SNF global transcription activator complex and RSC chromatin remodeling complex BAR1 (YIL015W): barrierpepsin precursor	GAL11 (YOL051W): DNA-directed RNA polymerase II holoenzyme and Kornberg's mediator (SRB) subcomplex subunit FAR1 (YJL157C): cyclin-dependent kinase inhibitor (CKI) SLA2 (YNL243W): cytoskeleton assembly control protein

[a]"Parents" means these genes connected directly to the drug-affected genes. "Grandparents" means there is one intermediary gene between these genes and the drug-affected genes ***(11)***.

3.3. Time-Expanded Network Method for Identifying Drug-Active Pathways

3.3.1. Time-Expanded Network Method

Figure 2 shows the conceptual view of our strategy. As the first step **(Fig. 2a)**, we use the Bayesian network and nonparametric regression model for estimating a gene network from disruptant microarray data. This model can even capture nonlinear relationships between genes and is appropriate to estimate the structure of the gene network. However, the probabilistic inference using the nonparametric regression models has several problems. One problem is that we need to consider the extrapolation for handling the drug response data in the estimated gene network. The other problem is that we need to consider the calibration that is a probabilistic inference from the value of the target gene to the distribution of its parent genes. The nonparametric regression in our Bayesian network model may not be suitable to consider the extrapolation problem and the calibration.

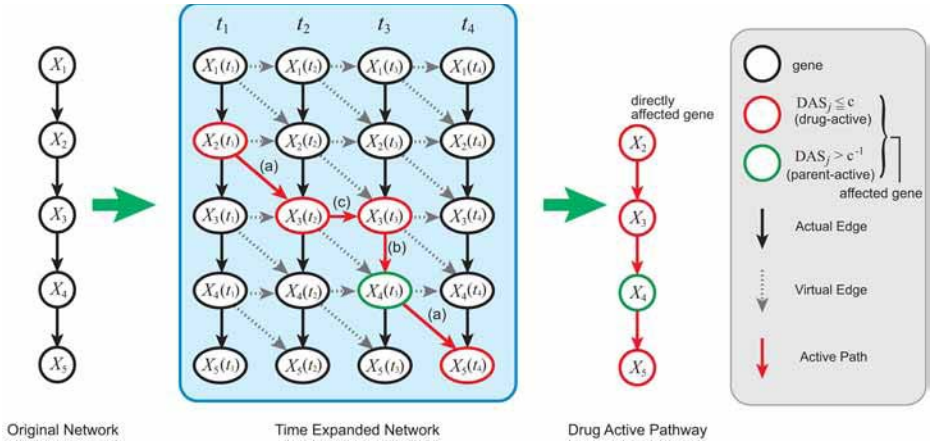


Fig. 8. Strategy to identify the drug active pathways. The red nodes represent drug-active genes considered to be independent on their parent genes. The green nodes are parent-active genes considered to be affected by their parent genes. The red arrows represent the active paths ***(13)***.

To avoid these problems, we use Bayesian networks for discrete data for the probabilistic inference after estimating the structure of the gene network. We convert the estimated Bayesian network and nonparametric regression model into the Bayesian network for discrete data **(Fig. 2b)**. The drug-active pathways are identified for every time-point **(Fig. 2c)**. This step is based on the identification of affected genes directly and indirectly by the drug. However, it is possible that several genes might still be falsely determined as the affected genes. Therefore, we define time-expanded network **(Fig. 8)** and consider the time-dependent relationships among the estimated affected genes to reduce the number of false positives. The definition of the time-dependent network is given below (*see* **Subheading 3.2.2.**). The connected components in the time-expanded network are selected as the estimated drug-active pathways **(Fig. 2d)**. This step also can reduce the number of falsely identified pathways.

The identification of drug-active pathways is done based on the probabilistic inference on the Bayesian networks for discrete data. For this purpose, we convert the Bayesian network and nonparametric regression model that was used for structural learning of the gene network into the Bayesian network for discrete data.

At first, we need to discretize the expression values into three categories. We regard expression values that are greater than a certain threshold h (or less than h^{-1}) as "overexpreesed" (or "suppressed") and otherwise as "unchanged." The threshold h is determined so that the network estimated by the Bayesian network for discrete data with h becomes the closest to the network from the Bayesian

network and nonparametric regression model. After determining the threshold, we compute the conditional probability table of the multinomial distribution. Let X_j be a discrete type random variable. The posterior probability of $X_j = u_k | Pa(X_j) = \boldsymbol{u}_{jl}$ conditional on the n sets of data $D = \{x_i,\ldots,x_n\}$ can be written as follows:

$$\pi_{post}(X_j = u_k | Pa(X_j) = \boldsymbol{u}_{jl}) = \frac{\alpha_{jkl} + N_{jkl}}{\sum_{k'}(\alpha_{jk'l} + N_{jk'l})}.$$

If a gene is directly affected by a drug, the expression level of the gene may be overexpressed or suppressed independently on the expression level of its parent genes. In contrast, if a gene is not directly affected by a drug, the expression value should be determined by the expression values of its parent genes. Based on this assumption, we define a score function called drag-affected score (DAS)j that determines whether jth gene is affected by a drug or its parents,

$$DAS_j(x, \boldsymbol{u}) = \frac{\pi_{post}(X_j = x | Pa(X_j) = \boldsymbol{u})}{\Pr(X_j = x)},$$

where x and $\boldsymbol{u}$ are the observed expression values of jth gene and its parents, and $\Pr(X_j)$ is the marginal probability given by $\Pr(X_j) = \Sigma_{X_j : i \neq j} \Pr(X_1,\ldots, X_p)$. This marginalization is often referred as the probabilistic inference and takes exponential time for the number of genes and is infeasible for the large network. To compute this interference efficiently, several approximation algorithms have been proposed. In this chapter, we use the loopy belief propagation algorithm. For additiontal comments on the probabilistic inference based on the Bayesian networks, see **Note 11**.

Let $x_j(t)$ and $\boldsymbol{p}_j(t)$ is the expression values of jth gene and its direct parents at time t. The procedure for finding drug-active pathways can be performed by the following steps.

Step 1. Construct the time-expanded network from the estimated gene network.

Step 2. Compute $\mathrm{DAS}_j(x_j(t), \boldsymbol{p}_j(t))$ and put the constant c (c>0),
if $\mathrm{DAS}_j(x_j(t), \boldsymbol{p}_j(t)) \leq c$, jth gene at time t is called a *drug-active gene*,
if $\mathrm{DAS}_j(x_j(t), \boldsymbol{p}_j(t)) > c^{-1}$, jth gene at time t is called a *parent-active gene*.
We call both drug-active and parent-active genes *affected genes*.

Step 3. If a direct edge, in the time-expanded network, from $X_u(t)$ to $X_v(s)$, where $X_u(t)$ is a node corresponding to uth gene at time t, satisfies with a condition out of three, this direct edge is called a *drug-active path*.

(i) $X_v(s)$ is a drug-active gene, where $s = t + 1$ **(Fig. 8a)**.
(ii) $X_v(s)$ is a parent-active gene, where $s = t$ **(Fig. 8b)**.
(iii) $X_v(s)$ is an affected gene, where $s = t + 1$ and $v = u$ **(Fig. 8c)**.

Step 4. Extract connected components of the drug-active paths from the time-expanded network.

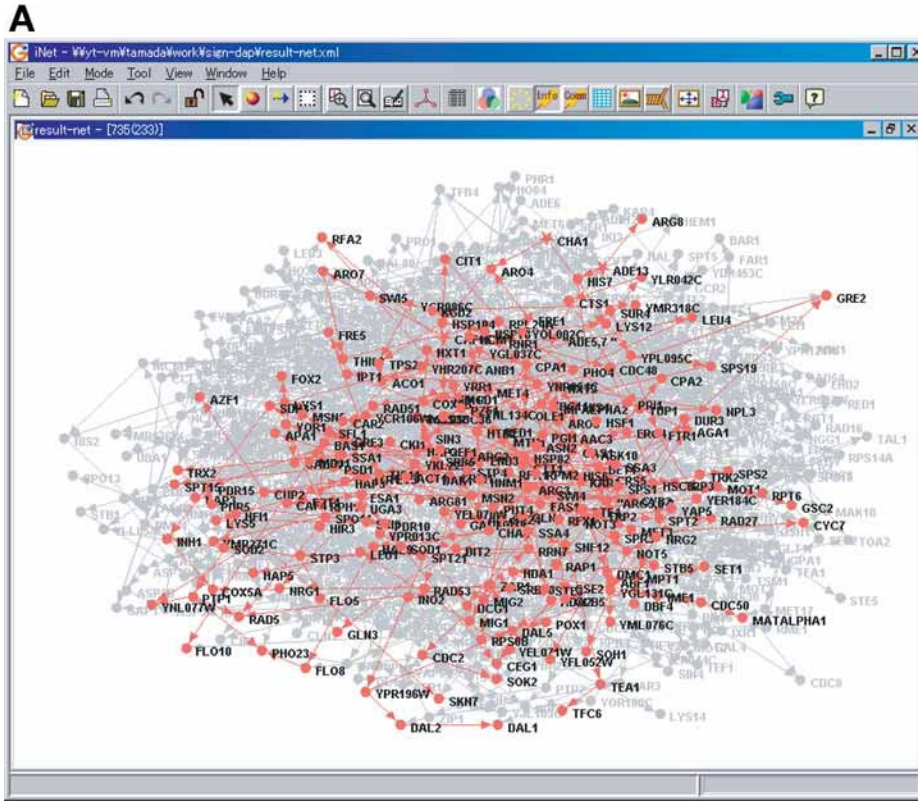


Fig. 9. *(continued)*

From the Monte Carlo simulations in Tamada et al. ***(13)***, this method succeeded in identifying the target drug-affected genes and pathways with the high coverage greater than 80%.

3.3.2. Application of Time-Expanded Network Method

From the gene–gene relationships of 735 genes estimated in **Subheading 3.2.4.**, we performed the time-expanded network method to extract drug-active pathways of the drug griseofulvin. We used the time-course microarray data of dosing 50 mg/mL, i.e., we used five microarrays corresponding to 0, 15, 30, 45, and 60 min. **Figure 9A** shows the identified drug-active pathways in the gene network estimated in **Subheading 3.2.4.** The identified drug-active pathways are colored red.

The identified drug-active pathways are extracted as subnetworks shown in **Fig. 9B**. Note that we focus on the subnetworks from the root genes up to the third descendants. In these drug-active pathways, only nine distinct genes listed

B

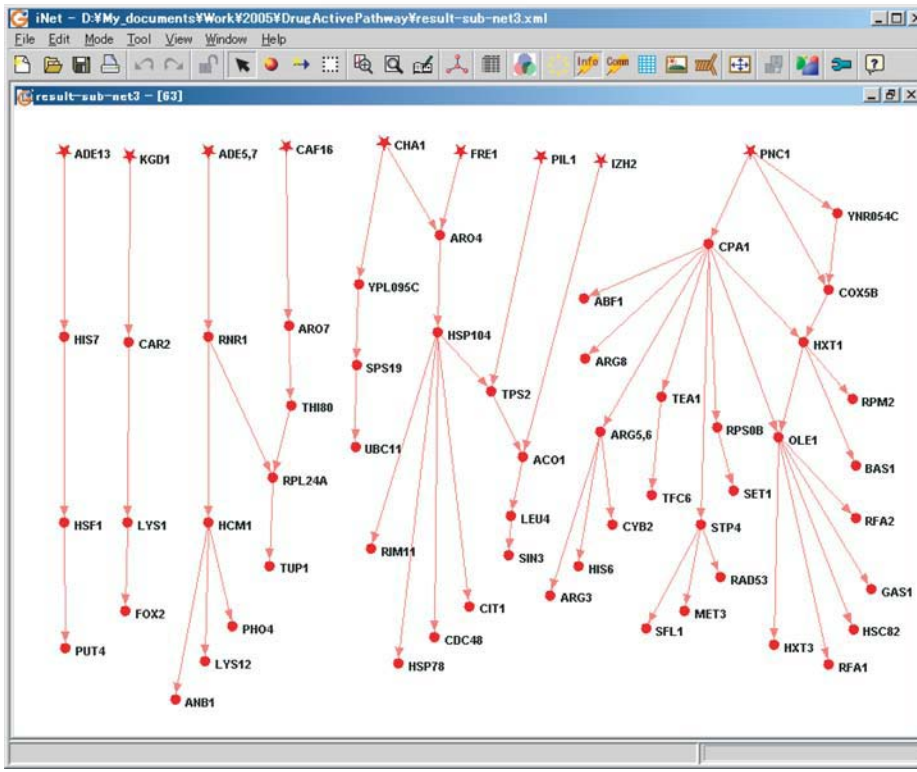


Fig. 9. **(A)** Identified drug-active pathways in the 735 gene network estimated in **Subheading 3.2.4.**

in **Table 3** were the root genes. These genes can be considered as potential genes that are directly affected by the antifungal drug griseofulvin. Interestingly, most genes in **Table 3** have GO annotations related to cytoplasm or membranes. The identified drug-affected pathways may elcidate the system of the affection of griseofulvin in cell and help to discover new drug targets.

4. Notes

1. For the gene network exploration, it is very difficult to efficiently extract information from a large gene network, thus a visualization tool for gene network analysis plays a crucial role. We therefore also have developed a gene network analysis tool called i.NET that provides a computational environment for various path searches among genes with annotated gene network visualization. This total system constitutes our method.
2. The advantages of the use of the multilevel dag model are as follows: (1) The model is very simple and can be easily understood, and (2) the model shows the parent–child

Table 3
List of Root Genes of Identified Drug-Active Pathways *(13)*

Gene name	GO annotations
CHA1	Serine family amino acid catabolism, mitochondrial nucleoid
CAF16	Regulation of transcription, cytoplasm
PNC1	Chromatin silencing, cytoplasm
ADE5,7	Purine base metabolism, cytoplasm
PIL1	Response to heat, cytoplasm
KGD1	Tricarboxylic acid cycle, mitochondrial matrix
FRE1	Iron ion transport, plasma membrane
ADE13	Purine base metabolism
IZH2	Lipid metabolism, integral to membrane

GO, gene ontology.

relations correctly, when the data have sufficient accuracy and information. However, this method requires discretization of the expression values into two levels (0 or 1), and the quantization probably causes information loss. By choosing the threshold η appropriately in a heuristic manner, it is reported that this naive algorithm works well in practice, although the method is not theoretically well founded.

3. In practice, biological experiments that disrupt the target druggable genes are needed for confirming the results of the analysis.
4. For identifying drug-affected genes by the virtual gene technique, we simply consider genes that have five or more virtual genes as the parent genes as the putative drug-affected genes. That is, genes are under direct influence of the virtual genes. However, a gene that has only one virtual gene as its parent may be the primary drug-affected gene, depending on the mode of action for a given drug, and this possibility must be analyzed case by case.
5. In Bayesian network literature, for example, Neapolitan ***(25)***, it is shown that determining the optimal network is an NP-hard problem. However, recent development of supercomputer technology enables us to optimize the small gene network, including 40 or fewer genes, by using a suitable algorithm ***(26)***. However, for larger numbers of genes, we use a heuristic strategy such as a greedy hill-climbing algorithm to discipher graph structure.
6. The simplest way for discretizing microarray gene expression data is the use of some thresholds: let $\{u_1, \ldots, u_m\}$ be a finite set of discrete values and $\{I_1,\ldots, I_m\}$ be regions satisfying $\cup_j I_j = R$ and $I_i \cap I_j = \varnothing$ for $i \neq j$. An expression value x_{ij} is then transformed to u_j when $x_{ij} \in I_j$ (*see* Friedman et al. ***[6]***). An alternative way for discretization is the use of *k*-means clustering ***(8)***. Also, Shmulevich and Zhang ***(27)*** proposed a method for binalizing gene expression data.
7. The nonparametric regression model in the Bayesian networks of Imoto et al. ***(9,10)*** is an additive regression model. However, in reality, there are several known transcriptional relationships that are more additive. Imoto et al. ***(28)*** proposed a

Bayesian network with moving boxcel median filtering to capture such interactions.

8. Imoto et al. ***(29)*** provided a general framework to combine microarray data and various types of biological data aimed at estimating gene networks. Tamada et al. ***(30)*** used promoter elements detection method together with the Bayesian networks for simultaneous estimation of gene networks and transcription factor binding sites. Nariai et al. ***(31,32)*** used protein–protein interaction data for refining gene networks estimated from microarray data. Tamada et al. ***(33)*** used evolutionary information to estimate gene networks of two distinct organisms.
9. The dynamic Bayesian networks ***(34–36)*** can estimate gene networks with cyclic relations from time-course microarray data. Also, Yoshida et al. ***(37)*** proposed a dynamic linear model with Markov switching to estimate time-dependent gene networks that can allow structural change of the estimated gene network depending on time.
10. For measuring reliability of an estimated edge, the bootstrap ***(38)*** is a useful method. As an advanced method of the bootstrap method, Kamimura et al. ***(39)*** used the multiscale bootstrap method ***(40)*** for computing more acculate bootstrap probability for the estimated edge in the Bayesian network.
11. Belief propagation ***(41)*** is an algorithm to infer the posterior marginal probability of a variable in the Bayesian networks. The exact computation of this algorithm is limited for singly connected networks; however, it takes only linear time in practice. Generally, calculating the probability in an arbitrary Bayesian network is known to be NP-hard ***(42)***. Therefore, several approximation algorithms have been proposed such as the loopy belief propagation ***(41)***, the junction tree algorithm ***(43)***, and sampling based methods such as Markov chain Monte Carlo ***(44)***.

Acknowledgments

We thank our colleagues and collaborators Hideo Bannai, Michiel de Hoon, Ryo Yoshida, Takao Goto, Sunyong Kim, Naoki Nariai, Sascha Ott, Tomoyuki Higuchi, Hidetoshi Shimodaira, Sachiyo Aburatani, Kousuke Tashiro, and Satoru Kuhara.

References

1. Akutsu, T., Kuhara, S., Maruyama, O., and Miyano, S. (1998) A system for identifying genetic networks from gene expression patterns produced by gene disruptions and overexpressions. *Genome Inform.* **9,** 151–160.
2. Akutsu, T., Miyano, S., and Kuhara, S. (1999) Identification of genetic networks from a small number of gene expression patterns under the Boolean network model. *Pac. Symp. Biocomput.* **4,** 17–28.
3. Shmulevich, I., Dougherty, E. R., Kim, S., and Zhang, W. (2002) Probabilistic Boolean networks: a rule-based uncertainty model for gene regulatory networks. *Bioinformatics* **18,** 261–274.
4. Chen, T., He, H. L., and Church, G. M. (1999) Modeling gene expression with differential equations. *Pac. Symp. Biocomput.* **4,** 29–40.

5. De Hoon, M. J. L., Imoto, S., Kobayashi, K., Ogasawara, N., and Miyano, S. (2003) Inferring gene regulatory networks from time-ordered gene expression data of *Bacillus subtilis* using differential equations. *Pac. Symp. Biocomput.* **8,** 17–28.
6. Friedman, N., Linial, M., Nachman, I., and Pe'er, D. (2000) Using Bayesian network to analyze expression data. *J. Comp. Biol.* **7,** 601–620.
7. Hartemink, A. J., Gifford, D. K., Jaakkola, T. S., and Young, R. A. (2001) Using graphical models and genomic expression data to statistically validate models of genetic regulatory networks. *Pac. Symp. Biocomput.* **6,** 422–433.
8. Pe'er, D., Regev, A., Elidan, G., and Friedman, N. (2001) Inferring subnetworks from perturbed expression profiles. *Bioinformatics* **17 (Suppl. 1),** S215–S224.
9. Imoto, S., Goto, T., and Miyano, S. (2002) Estimation of genetic networks and functional structures between genes by using Bayesian networks and nonparametric regression. *Pac. Symp. Biocomput.* **7,** 175–186.
10. Imoto, S., Kim, S., Goto, T., et al. (2003) Bayesian network and nonparametric heteroscedastic regression for non-linear modeling of genetic network. *J. Bioinform. Comp. Biol.* **1,** 231–252.
11. Imoto, S., Savoie, C. J., Aburatani, S., et al. (2003) Use of gene networks for identifying and validating drug targets. *J. Bioinform. Comput. Biol.* **1**, 459–474.
12. Savoie, C. J., Aburatani, S., Watanabe, S., et al. (2003) Use of gene networks from full genome microarray libraries to identify functionally relevant drug-affected genes and gene regulation cascades. *DNA Res.* **10,** 19–25.
13. Tamada, Y., Imoto, S., Tashiro, K., Kuhara, S., and Miyano, S. (2005) Identifying drug active pathways from gene networks estimated by gene expression data. Genome Inform Ser Workshop Genome Inform **16,** 182–191.
14. Aburatani, S., Tashiro, K., Savoie, C. J., et al. (2003) Discovery of novel transcription control relationships with gene regulatory networks generated from multiple-disruption full genome expression libraries. *DNA Res.* **10,** 1–8.
15. Quackenbush, J. (2002) Microarray data normalization and transformation. *Nat. Genet.* **32 (Suppl.),** 496–501.
16. Maki, Y., Tominaga, D., Okamoto, M., Watanabe, S., and Eguchi, Y. (2001) Development of a system for the inference of large scale genetic networks. *Pac. Symp. Biocomput.* **6,** 446–458.
17. Konishi, S. (1999) Statistical model evaluation and information criteria, in *Multivariate Analysis, Design of Experiments and Survey Sampling* (Ghosh, S., ed.), Marcel Dekker, New York, pp. 369–399.
18. Cooper, G. and Herskovits, E. (1992) A Bayesian method for the induction of probabilistic networks from data. *Mach. Learn.* **9,** 309–347.
19. Hastie, T. and Tibshirani, R. J. (1990) *Generalized Additive Models*. Chapman & Hall, London.
20. De Boor, C. (1978) *A Practical Guide to Splines*. Springer-Verlag, Berlin.
21. Imoto, S. and Konishi, S. (2003) Selection of smoothing parameters in *B*-spline nonparametric regression models using information criteria. *Ann. Inst. Stat. Math.* **55,** 671–687.
22. Davison, A. C. (1986) Approximate predictive likelihood. *Biometrika* **73,** 323–332.

23. Tinerey, L. and Kadane, J. B. (1986) Accurate approximations for posterior moments and marginal densities. *J. Am. Stat. Assoc.* **81,** 82–86.
24. Konishi, S., Ando, T., and Imoto, S. (2004) Bayesian information criteria and smoothing parameter selection in radial basis function networks. *Biometrika* **91,** 27–43.
25. Neapolitan, R. E. (2004) *Learning Bayesian Networks*. Peirson Prentice Hall, Upper Saddle River, NJ.
26. Ott, S., Imoto, S., and Miyano, S. (2004) Finding optimal models for small gene networks. *Pac. Symp. Biocomput.* **9,** 557–567.
27. Shmulevich, I. and Zhang, W. (2002) Binary analysis and optimization-based normalization of gene expression data. *Bioinformatics* **18,** 555–565.
28. Imoto, S., Higuchi, T., Kim, S., Jeong, E., and Miyano, S. (2005) Residual bootstrapping and median filtering for robust estimation of gene networks from microarray data, in *Proceedings of the 2nd Computational Methods in Systems Biology*, Lecture Note in Bioinformatics 3082, Springer-Verlag, Berlin, 149–160.
29. Imoto, S., Higuchi, T., Goto, T., Tashiro, K., Kuhara, S., and Miyano, S. (2004) Combining microarrays and biological knowledge for estimating gene networks via Bayesian networks. *J. Bioinform. Comput. Biol.* **2,** 77–98.
30. Tamada, Y., Kim, S., Bannai, H., et al. (2003) Estimating gene networks from gene expression data by combining Bayesian network model with promoter element detection. *Bioinformatics* **19 (Suppl. 2),** II227–II236.
31. Nariai, N., Kim, S., Imoto, S., and Miyano, S. (2004) Using protein–protein interactions for refining gene networks estimated from microarray data by Bayesian networks. *Pac. Symp. Biocomput.* **9,** 336–347.
32. Nariai, N., Tamada, Y., Imoto, S., and Miyano, S. (2005) Estimating gene regulatory networks and protein-protein interactions of *Saccharomyces cerevisiae* from multiple genome-wide data. *Bioinformatics* **21 (Suppl.),** II206–II212.
33. Tamada, Y., Bannai, H., Imoto, S., Katayama, T., Kanehisa, M., and Miyano, S. (2005) Utilizing evolutionary information and gene expression data for estimating gene regulations with Bayesian network models. *J. Bioinform. Comput. Biol.* **3,** 1295–1313.
34. Murphy, K. and Mian, S. (1999) Modelling gene expression data using dynamic Bayesian networks. Technical Report, Computer Science Division, University of California, Berkeley, CA.
35. Kim, S., Imoto, S., and Miyano, S. (2003) Inferring gene networks from time series microarray data using dynamic Bayesian networks. *Brief. Bioinform.* **4,** 228–235.
36. Kim, S., Imoto, S., and Miyano, S. (2004) Dynamic Bayesian network and nonparametric regression for nonlinear modeling of gene networks from time series gene expression data. *Biosystems* **75,** 57–65.
37. Yoshida, R., Imoto, S., and Higuchi, T. (2005) Estimating time-dependent gene networks from time series microarray data by dynamic linear models with Markov switching. *Proc. IEEE 4th Comput. Syst. Bioinform. Conf.*, 289–298.
38. Efron, B. and Tibshirani, R. J. (1993) *An Introduction to the Bootstrap*. Chapman & Hall, New York.

39. Kamimura, T., Shimodaira, H., Imoto, S., et al. (2003) Multiscale bootstrap analysis of gene networks based on Bayesian networks and nonparametric regression. *Genome Inform.* **14,** 350–351.
40. Shimodaira, H. (2004) Approximately unbiased tests of regions using multistep-multiscale bootstrap resampling. *Ann. Stat.* **32,** 2616–2641.
41. Pearl, J. (1988) *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Francisco, CA.
42. Cooper, G. (1990) The computational complexity of probabilistic inference using bayesian belief networks. *Artif. Intell.* **42,** 393–405.
43. Lauritzen, S. L. (1996) *Graphical Models*. Oxford University Press, New York.
44. Neal, R. M. (1993) Probabilistic inference using Markov chain Monte Carlo methods, Technical Report CRG-TR-93-1, Department of Computer Science, University of Toronto, Toronto, Ontario, Canada.

4

Target Discovery and Validation in Pancreatic Cancer

Robert M. Beaty, Mads Gronborg, Jonathan R. Pollack, and Anirban Maitra

Summary

Pancreatic cancer is a lethal disease and rational strategies for early detection and targeted therapies are urgently required to alleviate the dismal prognosis of this neoplasm. The use of global RNA and protein expression-profiling technologies, such as DNA microarrays, serial analysis of gene expression, and mass spectrometric analysis of proteins, have led to identification of cellular targets with considerable potential for clinical application and patient care. These studies underscore the importance of pursuing large-scale profiling of human cancers not only for furthering our understanding of the pathogenesis of these malignancies but also for developing strategies to improve patient outcomes.

Key Words: Bioinformatics; DNA microarrays; pancreatic cancer; proteomics; target discovery; serial analysis of gene expression; SAGE.

1. Introduction

1.1. Pancreatic Cancer

Infiltrating ductal adenocarcinoma of the pancreas, more commonly known as "pancreatic cancer" is the fourth leading cause of cancer deaths in the United States ***(1)***. It is estimated that in 2006, approx 213,000 individuals worldwide will be diagnosed with pancreatic cancer and approx 213,000 will die from this cancer ***(2)***. It is an almost uniformly fatal disease, with an overall 5-yr survival rate of less than 5%. The majority (~80%) of patients present with locally advanced or with metastatic malignancies, rendering the cancers surgically unresectable. Unfortunately, even patients who undergo surgical resection eventually succumb to metastatic disease, including those cases where the primary lesion was extremely small at diagnosis ***(3)***. The implication is that pancreatic cancer must metastasize distantly, before clinical presentation, in virtually all cases. A small number of

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarker Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

cytotoxic drugs and radiation therapy are reported to have modest clinical benefit, measured as time to progression or clinical improvement, in the treatment of pancreatic cancer. The most common current regimen is gemcitabine, used as either as a single agent, or increasingly, in combination with a platinum compound (e.g., cisplatin or oxaliplatin) ***(4)***. Despite these interventions, however, pancreatic cancer continues to be, in essence, a disease of near-uniform mortality.

The last 15 yr have seen an exponential increase in our understanding of the pathogenesis of pancreatic cancer, and it is now fairly evident that pancreas cancer is a disorder of the genome, caused by progressive accumulation of inherited and acquired mutations ***(5)***. These alterations include activating point mutations in the k-*ras* gene, overexpression of *HER-2/neu*, and inactivation of the *p16*, *p53*, *DPC4*, and *BRCA2* tumor suppressor genes (***1***,***3***,***6–11***). Other mechanisms may contribute to carcinogenesis of the pancreas, such as overexpression of growth factors and their receptors or changes in activity of signal transduction pathways (***12***,***13***). Although these gene-by-gene approaches have contributed toward a better understanding of pancreatic cancer pathogenesis, the advent of global profiling technologies have led to quantum advances in identifying molecular targets that can form the substrate for developing rational early detection and therapeutic strategies for pancreatic cancer ***(14–22)***. Unraveling the transcriptome and proteome of human pancreatic cancers by using large-scale approaches such as DNA microarrays ***(23)***, serial analysis of gene expression (SAGE) ***(24)***, and mass spectrometry ***(25)*** have led to an unbiased elucidation of cancer biomarkers, many of which have are being translated into direct patient care, whereas others have the potential to do so in the near future.

In this review, we discuss three approaches. The first two approaches, cDNA microarrays and SAGE, are used for analyzing the pancreatic cancer transcriptome; and the third approach, liquid chromatography and tandem mass spectrometry (LC-MS/MS), are used for analyzing its proteome. All three of these global approaches have been used with considerable success for identification of molecular targets in human pancreatic cancer. The last portion of this review focuses on candidate molecular targets identified through such large-scale profiling studies and how these relate to care of patients with pancreatic cancer.

1.2. DNA Microarrays

DNA microarrays consist of hundreds or thousands of PCR-amplified cDNAs or synthetic oligonucleotides spotted onto a glass microscope slide in a high-density pattern of rows and columns ***(23)***. DNA microarrays were first used widely to quantify gene expression across hundreds or thousands of genes simultaneously ***(26***,***27)***. To measure gene expression, mRNAs from two different samples are differentially fluorescently labeled and cohybridized to a DNA microarray, which is then scanned in dual wavelengths. For each DNA element

on the microarray, the ratio of fluorescence intensities reflects the relative abundance for that mRNA between the two samples.

The two-color fluorescence format provides robust measurements of gene expression ratios, effectively mollifying variations in amounts of spotted DNA and of hybridization kinetics. In some experimental designs, the selection of the second-color "reference" specimen is natural, for example, the zero time-point in a time-course, or the untreated sample in a pharmacological treatment. In other cases, an arbitrary "universal" RNA reference, comprising a pool of mRNA from cell lines or tissues, is more appropriate ***(28)***. Because each sample is hybridized against the same reference, samples can be readily compared with each other. In such cases, gene expression ratios are typically "mean-centered," i.e., reported in relation to the average of the specimens assayed (rather than to the arbitrary reference).

For small specimens (e.g., tissue biopsies or microdissected samples), insufficient input RNA may be available for the labeling protocol detailed in **Subheading 3.1.3**. In such cases, the reader is directed to linear RNA amplification protocols based on T7 RNA polymerase-mediated in vitro transcription ***(29)***.

1.3. Serial Analysis of Gene Expression

SAGE is a comprehensive expression profiling technology for quantitative gene expression in samples that does not depend on the prior availability of transcript information ***(24)***. In this aspect, it varies from other expression profiling technologies such as cDNA microarrays (*see* **Subheading 2.1.2.** and Chapter 5, Volume 1) where analyses is limited to a known repertoire of arrayed sequences. The premise of SAGE rests on the concept that a short sequence of nucleotides (~11 base pairs [bp], known as a "tag") is sufficient for uniquely identifying a transcript ***(24,30)***. The "serial" aspect of this platform comes from the generation of "concatemers" by ligating large numbers of SAGE tags within a "library" that are then serially analyzed by sequencing for gauging relative transcript levels.

The sequential steps involved in generation of a SAGE library are illustrated in **Fig. 1** ***(30)***. Briefly, double-stranded cDNA is generated from mRNA isolated from a sample of interest and immobilized on magnetic beads. The cDNA is then digested with a frequent cutting restriction enzyme such as NlaIII and ligated to a linker with a restriction recognition site and a PCR primer site. Subsequent digestion by another restriction enzyme (e.g., BsmfI) generates tags that contain the linker and a short nucleotide sequence (~11 bp) downstream of BsmfI recognition site that is specific for the transcript from which it was generated. Pairs of tags are then ligated to generate so-called "ditags," which are PCR amplified, digested by NlaIII to remove the linkers

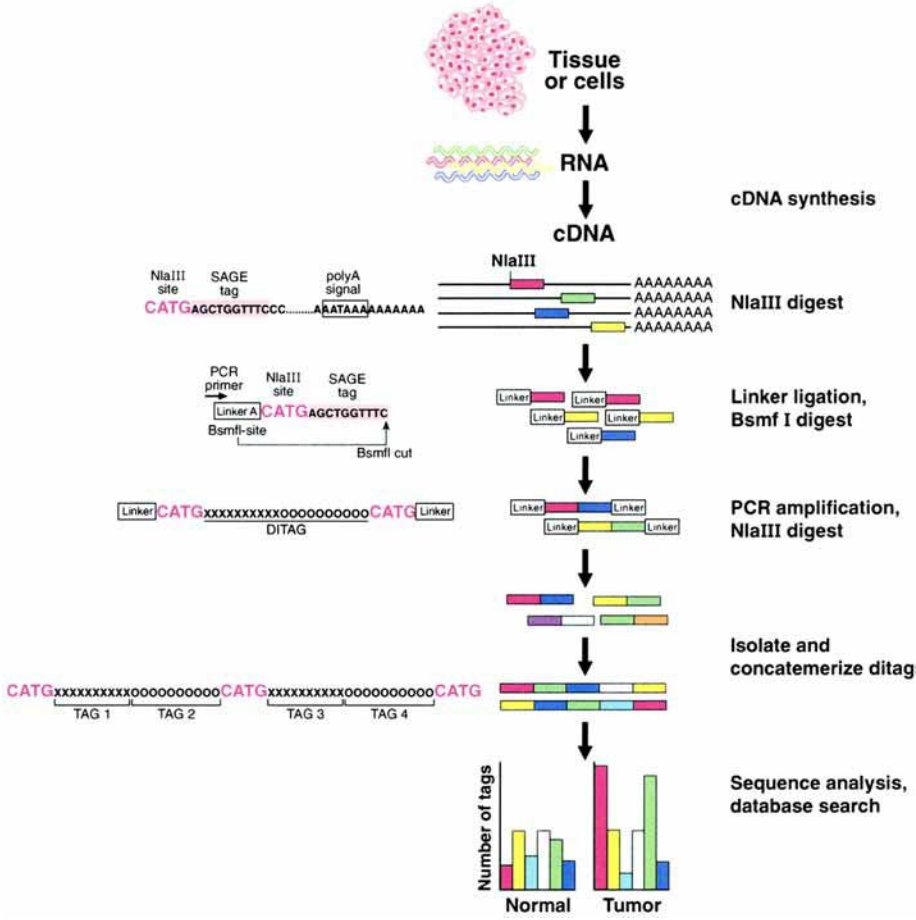


Fig. 1. Principles of serial analysis of gene expression (SAGE) analysis. (From **ref. *30***.)

and concatenated. The concatemers are subcloned into an appropriate vector, and the colonies resulting thereof are sequenced. The sequence data from a given SAGE library can be analyzed using a variety of online publicly as well as commercially available resources (*see* **Subheading 3.2.14.**), and any two or more such libraries (e.g., pancreatic cancer and normal pancreas) can be compared for differential gene expression. Since its original description, numerous modifications have been made to the SAGE protocol ***(31,32)***, including a "microSAGE" ***(33)*** (separately described as SAGE-lite ***[34]***) technique that permits generation of libraries from minute quantities of starting template (nanograms of mRNA), such as from microdissected material.

1.4. Mass Spectrometric Analysis of Proteins

Initial sample collection and preparation are crucial for obtaining optimal results for most analytical techniques, including mass spectrometry analysis of proteins. Despite the high sensitivity of current mass spectrometers (routine analysis at femtomole level), great care should be taken to avoid degradation of the sample because of the high abundance of naturally occurring proteases present in pancreatic juice and pancreatic tissue. General guidelines and protocols regarding sample handling, initial preparation, and fractionation by one-dimensional (1D) electrophoresis, staining, tryptic digestion, nano-LC-MS/MS analysis, database searches, and validation of data are provided in **Subheading 3.3.8**. It should be emphasized that these protocols are general protocols that should be modified if more specific analysis is needed. It is very important not to contaminate the sample with, e.g., keratin (hair and skin). In addition, universal precautions must be followed during handling of the specimens.

1.5. Molecular Targets Identified by Large-Scale Transcriptomic and Proteomic Analysis of Human Pancreatic Cancers

The large-scale analyses of human pancreatic cancers by using techniques described above (**Subheadings 1.2.–1.4.**) has led to the identification of nearly 200 differentially overexpressed transcripts and a similar number of differentially overexpressed proteins ***(15–18)***. These molecules have immediate translational relevance in terms of patient care, harboring the potential for application in imaging studies, targeted therapies, and early detection. As the ensuing discussion highlights, such translational work is actively underway. For example, the transcripts prostate stem cell antigen (PSCA) and mesothelin were identified by Argani et al. ***(19,20)*** as being differentially overexpressed in human pancreatic cancers by SAGE analysis ***(19,20)***. Because both PSCA and mesothelin are membranous proteins, they can be used for specific targeting of imaging agents to pancreatic cancers in vivo ***(35)*** or as a target for monoclonal antibody therapy by using humanized antibodies (*see* http://www.agensys.com). Mesothelin also has been used as a target for immunotoxin therapy with a conjugated anti-mesothelin antibody ***(36)***. Most recently, it also has been reported that mesothelin is a target antigen against which a host immunologic reaction is mounted by the cytotoxic T-cell repertoire in human pancreatic cancer patients ***(37)***; thus, it is currently being used in the preparation of pancreatic cancer vaccines. In addition to therapeutic applications for pancreatic cancer, PSCA and mesothelin also are secreted proteins; therefore, one can envision quantitative assessment of these molecules in pancreatic juice samples for early detection of pancreatic cancer in at-risk patients ***(38)***. Claudin 4 is an overexpressed cell surface protein in human pancreatic cancers identified using an

oligonucleotide microarray approach ***(39)***. This tight junction molecule also has been used for targeting immunotoxins to pancreatic cancer cells ***(40,41)*** as well for imaging purposes ***(35)***. In transcriptomic studies, proteomic analysis of pancreatic juice samples has led to the identification of regenerating islet-derived 3 alpha (REG3A), also known as pancreatitis-associated protein, as overexpressed in cancer samples compared with nonneoplastic pancreatic juice ***(21)***. The selective overexpression of REG3A in neoplastic juice samples implies that this protein can be used as an early detection biomarker for pancreatic cancer.

2. Materials

2.1. DNA Microarrays

1. 20X Standard saline citrate (SSC) solution.
2. 10% Sodium dodecyl sulfate (SDS) solution.
3. Bovine Serum Albumin (BSA) (cat. no. A7888, Sigma); prepare a 10 mg/mL BSA stock solution, filtered.
4. Ethanol, 95% (Gold Shield Chemical Co.).
5. Diamond scribe (cat. no. 52865-005, VWR).
6. Slide humidifying chamber (cat. no. H6644, Sigma).
7. Heating block (cat. no. 13259-050, VWR), invert so solid metal surface faces up.
8. Ultraviolet source (Stratalinker 2400, Stratagene; or equivalent).
9. Pyrex dish for denaturing spotted cDNA on arrays (cat. no. 08-741F, Fisher).
10. Slide rack and glass staining dish (Wheaton).
11. Desktop centrifuge (Beckman Allegra 6R, or equivalent), with microtiter plate carriers.

2.1.1. Labeling mRNA

1. RNase-free H_2O (cat. no. 9935, Ambion; or equivalent).
2. Anchored oligo(dT) reverse transcriptase (RT) primer ($T_{20}VN$, custom primer from Operon; or equivalent; HPLC purify); prepare 2.5 µg/µL stock in $TE_{8.0}$ buffer.
3. Superscript II reverse transcriptase (cat. no. 18064-014, Invitrogen, Carlsbad, CA).
4. 5X First-Strand buffer (cat. no. Y00146, Invitrogen).
5. 0.1 *M* Dithiothreitol DTT (Y00147, Invitrogen).
6. RNasin (cat. no. N2511, Promega).
7. 50X dNTP mix: 10 m*M* dTTP, 25 m*M* each dGTP, dATP, dCTP, in $TE_{8.0}$ buffer); prepare from ultrapure dNTP solutions; GE Healthcare).
8. Cy5-dUTP and Cy3-dUTP (cat. nos. PA55022 and PA53022, GE Healthcare).
9. 0.5 *M* EDTA.
10. Water bath(s) (70, 42, and 37°C).

2.1.2. Hybridization of Labeled Sample

1. RNase-free H_2O (cat. no. 9935, Ambion; or equivalent).
2. PolyA RNA (cat. no. P9403, Sigma). Prepare 10 µg/µL stock solution in $TE_{8.0}$.

3. Yeast tRNA (cat. no. 15401-011, Invitrogen). Prepare 5 μg/μL stock solution in $TE_{8.0}$.
4. Human Cot-1 DNA (cat. no. 15279-011, Invitrogen). Supplied as 1 μg/μL stock solution.
5. 20X SSC.
6. 10% SDS.
7. Microcon-30 filters (Amicon).
8. Hybridization chambers (Corning; Monterey Industries, http://www.montereyindustries.com; or equivalent).
9. Glass microscope slide cover slips, 22 × 60 mm (Fisher).
10. Slide rack and three glass staining dishes (Wheaton).
11. Desktop centrifuge (Beckman Allegra 6R, or equivalent), with microtiter plate carriers.
12. Water bath (65°C).

2.1.3. Data Reduction and Analysis

1. GenePix 4000B scanner (Molecular Devices), or equivalent, for imaging microarrays after hybridization.
2. GenePix software (Molecular Devices), or equivalent, for data extraction (i.e., matching pixels to DNA spots; calculating fluorescence ratios).
3. Microsoft Excel, or more specialized microarray databases (e.g., AMAD, http://www.microarrays.org/software.html), for data storage, retrieval, and analysis.

2.2. Serial Analysis of Gene Expression

The materials required for SAGE can be divided into 13 discrete steps that are listed stepwise. Note that some reagents are required in multiple steps.

2.2.1. Kinasing Reaction for Linkers (Step 1)

1. Linker sequences.
   a. Linker 1A (obtain gel-purified).
      5′-TTT GGA TTT GCT GGT GCA GTA CAA CTA GGC TTA ATA GGG ACA TG-3′.
   b. Linker 1B (obtain gel-purified).
      5′-TCC CTA TTA AGC CTA GTT GTA CTG CAC CAG CAA ATC C[amino mod. C7]-3′.
   c. Linker 2A (obtain gel-purified).
      5′-TTT CTG CTC GAA TTC AAG CTT CTA ACG ATG TAC GGG GAC ATG-3′.
   d. Linker 2B (obtain gel-purified).
      5′-TCC CCG TAC ATC GTT AGA AGC TTG AAT TCG AGC AG[amino mod. C7]-3′.
2. NEB T4 polynucleotide kinase (10 U/μL; cat. no. 201S, New England Biolabs).

3. The 12% polyacrylamide gel electrophoresis-Tris acetate-EDTA (PAGE-TAE) gel may be made from the following:
   a. 40% Polyacrylamide: 19:1 acrylamide:bis (cat. no. 161-0144, Bio-Rad).
   b. 50X TAE buffer (cat. no. 330-008-161, Quality Biological) (*see* **Note 1**).

2.2.2. First- and Second-Strand cDNA Synthesis (Step 2)

1. Dynabeads oligo(dT)25 mRNA direct kit (cat. no. 610.11, Dynal).
2. Dynal magnetic particle concentrators (MPC-S cat. no. 120.20, Dynal).
3. Superscript Choice System cDNA synthesis kit (cat. no. 18090-019, Invitrogen).
4. 0.5 *M* EDTA (cat. no. E7889, Sigma).
5. SDS (cat. no. L4390, Sigma).

2.2.3. Cleavage of cDNA With Anchoring Enzyme (NlaIII)(Step 3)

1. *Nla*III (cat. no. 125S, New England Biolabs).

2.2.4. Ligating Linkers top cDNA (Step 4)

1. T4 Ligase (high concentration 5 U/µL; cat. no. 15224-041, Invitrogen).
2. BSA (cat. no. B9001S, New England Biolabs).

2.2.5. Release Tags Using Tagging Enzyme BsmfI of cDNA (Step 5)

1. *Bms*fI (cat. no. 572S, New England Biolabs).
2. Phenol-chloroform.
3. 0.5 *M* EDTA (cat. no. P2069, Sigma).
4. LoTE: 3 m*M* Tris-HCl, pH 7.5, 0.2 m*M* EDTA, pH 7.5.

2.2.6. Blunt-Ending Release Tags (Step 6)

1. 2 U/L Klenow (cat. no. 27-0929-01, Pharmacia).
2. 7.5 *M* Ammonium acetate (cat. no. A2706, Sigma).

2.2.7. Ligate to Form Ditags (Step 7)

1. T4 Ligase (high concentration 5 U/µL; cat. no. 15224-041, Invitrogen).

2.2.8. PCR Amplification of Ditags (Step 8)

1. Primer 1: 5′-GGA TTT GCT GGT GCA GTA CA-3′.
2. Primer 2: 5′-CTG CTC GAA TTC AAG CTT CT-3′.
3. 40% Polyacrylamide: 19:1 acrylamide:*bis* (cat. no. 161-0144, Bio-Rad) for making 12% PAGE-TAE gel.
4. 96-Well plates for PCR (cat. no. Z37,490-3, Sigma).
5. Taq polymerase (cat. no. 10996-034, Invitrogen).

2.2.9. Isolation of Ditags (Step 9)

1. SYBR Green I (cat. no. S-7567, Invitrogen).
2. SpinX filtration microcentrifuge tubes (Costar, Cambridge, MA).

2.2.10. Purification of Ditags (Step 10)

1. TE: 10 m*M* Tris-HCl, pH 7.5, 1 m*M* EDTA, pH 8.0.

2.2.11. Ligation of Ditag s to Form Concatemers (Step 11)

1. 40% Polyacrylamide: 37.5:1 acrylamide:bis (cat. no. 161-0148, Bio-Rad).

2.2.12. Ligating Concatemers in pZero (Step 12)

1. *Sph*I restriction enzyme (cat. no. 182S, New England Biolabs).
2. PZero plasmid (cat. no. K2500-01, Invitrogen).

2.2.13. Electroporation of Ligation Products and Colony PCR (Step 13)

1. ElectroMAX DH10B competent cells.
2. Zeocin (cat. no. R250-01, Invitrogen).
3. Glass beads.
4. Agarose (Select Agar cat. no. 30391-023, Invitrogen).

2.3. Mass Spectrometric Analysis of Proteins

2.3.1. Preparation of Pancreatic Juice and Pancreatic Tissue

1. Protease inhibitor tablets (cat. no. 1697498, Roche Diagnostics).
2. DC Protein Assay kit I (cat. no. 500-0111, Bio-Rad).
3. Scalpel handle (cat. no. S7772-1EA, Sigma).
4. Scalpel blade (cat. no. S2646-100EA, Sigma).
5. Phosphate-buffered saline (PBS) tablets (cat. no. P4417-100TAB, Sigma).
6. Modified radioimmunoprecipitation assay (RIPA) buffer (150 m*M* NaCl, 50 m*M* Tris-HCl, pH 7.4, 1% Nonidet P-40 (NP-40), 0.25% sodium deoxycholate, 1 m*M* EDTA).
7. Sodium chloride (cat. no. S9888, Sigma).
8. Deoxycholic acid (sodium deoxycholate) (cat. no. D6750, Sigma).
9. NP-40 (cat. no. 492015, Calbiochem).
10. Trizma base (cat. no. T6066, Sigma).
11. EDTA (cat. no. E4884, Sigma).
12. Sonicator (Sonifier 250, Branson; or similar).

2.3.2. Fractionation by 1D Gel Electrophoresis

1. XCell SureLock (cat. no. EI0001, Invitrogen).
2. 10% NuPAGE, Bis-Tris gel, 1.5 mm, 10-well (cat. no. NP0315BOX, Invitrogen).
3. 4–12% NuPAGE, Bis-Tris gel, 1.5 mm, 10-well (cat. no. NP0335BOX, Invitrogen).
4. 10X NuPAGE sample reducing agent (cat. no. NP0009, Invitrogen).
5. 4X NuPAGE LDS sample buffer (cat. no. NP0007, Invitrogen).
6. NuPAGE antioxidant (cat. no. NP0005, Invitrogen).

2.3.3. Protein Staining (Silver Staining and Coomassie Staining)

1. Colloidal blue staining kit (cat. no. LC6025, Invitrogen).
2. Acetic acid, glacial (cat. no. A38-500, FIsher).
3. Methanol (cat. no. MX0475P-4, EMD Chemicals Inc.).
4. Sodium carbonate (cat. no. S-2127, Sigma).
5. Silver nitrate (cat. no. SX0205-5, EMD Chemicals Inc.).
6. Sodium thiosulfate (SX0820-1, EMD Chemicals Inc.).
7. Formaldehyde (cat. no. 2106-01, J. T. Baker).

2.3.4. Tryptic Digestion

1. Scalpel handle (cat. no. S7772-1EA, Sigma).
2. Scalpel blade (cat. no. S2646-100EA, Sigma).
3. 10 m*M* DTT (cat. no. 43815, Sigma).
4. 55 m*M* Iodoacetamide (cat. no. I-6125, Sigma).
5. 50 and 100 m*M* NH_4HCO_3 (cat. no. 09830, Sigma).
6. Acetonitrile (cat. no. 9017-03, J. T. Baker).
7. 12.5 ng/µL Modified trypsin, sequencing grade (cat. no. V5111, Promega).
8. 5% Formic acid (cat. no. 06450, Sigma).

2.3.5. Reverse Phase Columns

1. Fused-silica capillary tubing for liquid chromatography 75 µm id/365 µm od, (cat. no. TSP075375, Polymicro Technologies).
2. Kasil 1™ (potassium/silicate solution) (cat. no. 25110, ASTRO PQ Corporation).
3. Formamide (super pure grade) (cat. no. BP228-100, Fisher).
4. Reverse phase material for liquid chromatography SB-C18, 5–15 µm (Zorbax).
5. Reverse phase material for liquid chromatography MS218, 5 µm (Vydac, Columbia, MD).
6. Methanol (MX0475P-4, EMD Chemicals Inc.).
7. High-pressure vessel for nano-LC column packing (cat. no. SP035, Proxeon A/S).

2.3.6. LC-MS/MS

1. Emitters for liquid chromatography FS360-20-10-N-20; tip, 10 ± 1 µm (cat. no. FS3602010N20, New Objective).
2. Mobile phase A (H_2O with 0.4% acetic acid and 0.005% heptafluorobutyric [v/v]).
3. Mobile phase B (90% acetonitrile, 0.4% acetic acid, 0.005% heptafluorobutyric acid in water).
4. Acetonitrile (cat. no. 9017-03, J. T. Baker).
5. Heptafluorobutyric acid (cat. no. H-7133, Sigma).
6. Glacial acetic acid (cat. no. A38-500, Fisher).

3. Methods

3.1. Microarays

Many universities, medical centers, and companies have established "core facilities" dedicated to printing DNA microarrays. Detailed protocols for printing spotted DNA microarrays have been published ***(42)***. DNA microarrays also can be purchased from various commercial vendors (e.g., Agilent Technologies). The protocols that follow are geared specifically to spotted cDNA microarrays. Where distinct methods are required for spotted oligonucleotide arrays, alternative protocols are detailed in **Subheading 4**.

3.1.1. Processing and Prehybridization of DNA Microarrays

Before use, and unless already processed by the manufacturer, DNA microarrays should be processed and prehybridized to enhance specific hybridization to DNA spots and to block nonspecific hybridization to the glass surface. The following protocol applies to microarrays of cDNAs spotted onto aminosilane–coated slides (e.g., GAPSII, Corning), a widely used surface for nucleic acid immobilization.

1. Using a diamond scribe, etch the underside (nonprinted) surface of the array to demark the boundaries of the printed area (*see* **Note 1**). Optionally, if arrayed DNA spots look small, rehydrate arrays over double distilled H_2O (ddH_2O) preheated to 50°C by using a slide humidifying chamber (printed surface facing down) for 20 s.
2. Quickly snap-dry each array by placing on a preheated 75°C heating block (printed surface facing up) for 5 s (*see* **Note 2**).
3. UV-crosslink printed DNA onto glass substrate with 60 mJ of energy by using Stratalinker (place slides printed side up; use Energy mode: 600 × 100 μJ) and then transfer arrays to a slide rack.
4. Denature arrayed cDNA spots by immersing arrays in near-boiling (*see* **Note 3**) ddH_2O in Pyrex dish and agitate for 2 min.
5. Prehybridize arrays in 3X SSC, 0.1 mg/mL BSA, 0.1% SDS in a glass staining dish (in a water bath) at 50°C for 60 min.
6. Rinse arrays by gentle shaking in ddH_2O at room temperature for 5 min; repeat twice.
7. Subsequently, plunge arrays into 95% ethanol and agitate for 2 min, spin-dry arrays in a tabletop centrifuge at 500*g* at room temperature for 5 min, and use arrays the same day (for oligonucleotide arrays, *see* **Note 4**).

3.1.2. Preparation of mRNA

Both total RNA and mRNA are suitable substrates for fluorescence labeling and expression profiling. Total RNA can be isolated from specimens by using TRIzol reagent (Invitrogen) or by anion exchange methods (e.g., QIAGEN RNeasy). mRNA can be isolated by oligo(dT) affinity purification (e.g., Invitrogen FastTrack 2.0).

3.1.3. Labeling of mRNA

1. For each test or reference sample, in a 1.5-mL Eppendorf tube on ice add the following:
   - 50 µg of total RNA (or 2 µg of mRNA).
   - 5 µg of RT primer, and RNase-free H_2O to a total volume of 15 µL.
2. Heat samples at 70°C for 10 min and then cool on ice 3 min, quick spin at 4°C and hold on ice.
3. On ice, add 3 µL of Cy5 dUTP or Cy3 dUTP to test and reference samples, respectively (*see* **Note 5**). Tap gently to mix then quick spin.
4. On ice, add 12.1 µL of RT Mastermix (6 µL of 5X First-Strand buffer, 3 µL of 0.1 *M* DTT, 0.6 µL of 50X dNTPs, 0.5 µL of RNasin, and 2 µL of Superscript II). Tap gently to mix and then quick spin and incubate 42°C for 1 h.
5. Thereafter, add an additional 1 µL of Superscript II, mix gently, quick-spin, and incubate at 42°C an additional 1 h.
6. Cool on ice and add 3 µL of 0.5 *M* EDTA stop solution, mix gently, and quick spin.

3.1.4. Hybridization of Labeled Sample

1. Combine Cy5-labeled test and Cy3-labeled reference samples, together with 450 µL of $TE_{7.4}$, into a Microcon-30 filter, and spin 12,000*g* for 9–12 min at room temperature in a microcentrifuge; retained volume should be approx 20 µL (*see* **Note 6**); discard the flow-through.
2. Add an additional 500 µL of $TE_{7.4}$, spin at 12,000*g* for 9–12 min, and discard the flow-through (*see* **Note 7**).
3. Add an additional 450 µL of $TE_{7.4}$, 2 µL of polyA RNA stock solution, 2 µL of yeast tRNA stock solution, and 20 µL of human Cot-1 stock solution (*see* **Note 8**).
4. Spin 12,000*g* for 9–12 min and discard the flow-through. The retained volume should be less than 32 µL.
5. Invert Microcon-30 filter into a new Eppendorf tube, and spin at 12,000*g* for 1 min to recover labeled cDNA sample.
6. Transfer the labeled cDNA mixture to a screw-top microcentrifuge tube, determine the sample volume by using a pipet, and increase the volume to 32 µL by adding ddH_2O.
7. Add 6.8 µL of 20X SSC and mix. Add 1.2 µL of 10X SDS, mix gently to avoid forming bubbles.
8. Boil the hybridization mixture for 2 min, incubate at 37°C for 20 min, and then quick spin.
9. Carefully pipet hybridization mixture onto DNA microarray and overlay with a 22- × 60-mm glass cover slip (*see* **Note 9**).
10. Overlay cover slip with several small droplets of 3X SSC (totaling 20 µL) to provide hydration and then enclose and seal slide in hybridization chamber. Incubate at 65°C for 16–18 h (*see* **Note 10**).

3.1.5. Washing Microarrays After Hybridization

After hybridization, wash the DNA microarrays to remove unbound labeled sample. The following three sequential wash steps should be performed using a slide rack, transferring among three separate glass staining dishes, each containing volumes of 350–400 mL, with gentle agitation:

- Wash 1: 2X SSC/0.03% SDS (see Note 11), at 65°C (see Note 12), 5 min.
- Wash 2: 1X SSC, room temperature, 5 min.
- Wash 3: 0.2X SSC, room temperature, 5 min.

After the third wash, DNA microarrays should be spun dry by using a desktop centrifuge, at 500*g* at room temperature for 5 min (*see* **Note 13**).

3.1.6. Microarray Imaging, Data Reduction, and Analysis

After hybridization, DNA microarrays should be scanned in dual wavelengths by using a GenePix 4000B (Molecular Devices) scanner, or equivalent. GenePix software, or equivalent, should be used to extract fluorescence intensity ratios for each DNA element on the DNA microarray. To compensate for variable labeling efficiencies, fluorescence ratios can be normalized such that the average fluorescence ratio for all DNA elements on the array is set to 1 (i.e., "global" normalization). Microsoft Excel, or any of several more sophisticated commercial or academic microarray databases ***(43)***, can be used to manipulate and analyze microarray data.

3.2. Serial Analysis of Gene Expression

3.2.1. Kinasing Reaction for Linkers (See ***Note 14***)

The linkers need to be "kinased" as described in the NEB T4 polynucleotide kinase instructions (New England Biolabs). Briefly,

1. Dilute Linker 1B, 2B, 1A, and 2A to 350 ng/µL.
2. Incubate kinase reaction solution per NEB instructions at 37°C for 30 min and heat inactivate at 65°C for 10 min.
3. Mix 9 µL of Linker 1A (350 ng/mL) into the 20 µL of kinased Linker 1B prepared above (final conc. 200 ng/µL) to generate "1AB."
4. Then, mix 9 µL of Linker 2A (350 ng/µL) into the 20 µL of kinased Linker 2B prepared in **Step 2** (final conc. 200 ng/µL) to generate "2AB."
5. Anneal both tubes of linkers with the following incubations: 1) 95°C for 2 min, 2) 65°C for 10 min, 3) 37°C for 10 min, 4) room temperature for 20 min, 5) Store at –20°C. Store at –20°C.
6. Test the anneal mixes. Kinasing should be tested by self-ligating both mixes and incubating at 37°C for 30 min.
7. Analyze on a 1.5-mm 12% PAGE-TAE gel. The ligated dimers should run at 80 bp, and the monomers should be run at 40 bp. Load all of the self-ligated linkers, and

1 µL of the annealed mixes as negative controls. Kinased linkers should allow linker–linker dimers (80–100 bp) to form after ligation, while unkinased linkers will prevent self-ligation. Only linker pairs that self-ligate greater than 70% should be used in further steps.

3.2.2. First- and Second-Strand cDNA Synthesis

Use Dynal Dynabeads oligo(dT)25 mRNA direct kit and magnets with the Superscript Choice system cDNA synthesis kit to accomplish first- and second-strand synthesis.

1. Thoroughly resuspend Dynabeads oligo(dT)25 beads. Transfer 100–120 µL to an RNase-free, 1.5-mL Eppendorf tube, and place it on the magnetic Eppendorf tube holder.
2. After the mixture of beads and buffer has cleared, remove the supernatant. Then, wash the beads in 500 µL of lysis/binding buffer, and mix 10 µg of total RNA with 1 mL of lysis/binding buffer (*see* **Note 15**).
3. Remove the 500 µL of lysis/binding buffer supernatant from the Dynabeads and add the RNA–buffer mixture to these prewashed Dynabeads.
4. Shake at room temperature for 5 min by hand, place the tube on the magnet for 2 min, and remove supernatant.
5. Wash the beads twice with 1 mL of buffer A/washing buffer.
6. Wash the beads once with 1 mL of buffer B/washing buffer containing 20 µg/mL glycogen and subsequently wash the beads four times with 100 µL of 1X First-Strand buffer.
7. Remove the supernatant and resuspend beads in the First-Strand synthesis mix from the cDNA kit.
8. Place the tube at 42°C for 2 min and then add 3 µL of SuperScript II RT.
9. Incubate at 42°C for 1 h. Mix the beads every 10 min by vortexing or tapping the tube.
10. After incubation, place the tube on ice to terminate the reaction.
11. Directly to the first-strand reaction, and on ice, add the second-strand synthesis components for a total of 500 µL.
12. Incubate at 16°C for 2 h; mix the beads every 10 min.
13. After incubation, place the tubes on ice and terminate the reaction by adding 40 µL of 0.5 *M* EDTA, pH 8.0.
14. Place the beads on the magnet and remove the supernatant.
15. Wash the beads once with 500 µL of 1X BW containing 1% SDS.
16. Resuspend the beads in 200 µL of 1X BW/1% SDS and heat at 75°C for 20 min.
17. Wash four times with 200 µL of 1X BW containing 2X BSA.
18. Wash twice with 200 µL of 1X NEB T4 containing 2X BSA.

3.2.3. Cleavage of cDNA With Anchoring Enzyme (NlaIII)

Resuspend the beads by setting up a restriction digest with NlaIII. Mix and incubate at 37°C for 1 h (mix every 15 min).

3.2.4. Ligating Linkers to cDNA

1. After the NlaIII restriction digest, remove the supernatant and wash the beads twice with 200 µL of 1X BW/0.1% SDS containing 2X BSA.
2. Wash four times with 200 µL of 1X BW containing 2X BSA (freshly made).
3. Wash twice with 200 µL of 1X ligase buffer.
4. Divide the beads into two tubes and place them on a magnet.
5. Remove the last wash and set up a 10-µL ligation with linker 1AB (1 µL) and linker 2AB (1 µL) in separate tubes. The amount of linkers used needs to be adjusted according to the number of cells (*see* **Note 16**). Two different linkers are used to prevent looping of the template in the PCR reaction. The linkers encode for the BsmfI restriction enzyme site.
6. Heat tubes at 50°C for 2 min and then let sit at room temperature for 15 min.
7. Add 1 µL of T4 high concentration ligase (5 U/µL) to each tube (10-µL reaction).
8. Incubate at 16°C for 2 h and mix beads intermittently by tapping the tube.

3.2.5. Release Tags by Using Tagging Enzyme (BsmfI) of cDNA

1. After ligation, wash each sample twice with 200 L of 1X BW/0.1% SDS containing 2X BSA (freshly made), by adding the 200 µL of buffer to the 10-µL ligation reaction.
2. Pool tube 1 and tube 2 after the first wash to minimize losses in subsequent steps.
3. Wash four times with 200 µL of 1X BW/2X BSA (freshly made).
4. Wash twice with 200 µL of 1X NEB T4 with 2X BSA (freshly made).
5. Set up a restriction digest with BmsfI. BsmfI produces a 5′ sticky end of four bases (10 bp + 4 bp into unknown sequence).
6. Incubate at 65°C for 1 h, mix intermittently, and then centrifuge at 14,000*g* for 2 min at 4°C.
7. Transfer supernatant to new tube (*see* **Note 17**).
8. Wash beads once with 40 µL of LoTE; thereafter, pool the 200-µL supernatant and the 40-µL wash (240 µL final volume).
9. Perform phenol:chloroform extraction by adding 240 µL of phenol:chloroform, pH 8.0 (“PC8”).
10. Vortex, then centrifuge for 5 min at 11,000*g* at 4°C and transfer the upper aqueous phase to a new tube.
11. Then, precipitate with 100% ethanol (EtOH) at –80°C. Centrifuge for 30 min at 21,000*g* at 4°C. Wash twice with 200 µL of 75% EtOH.
12. Resuspend pellet in 10 µL of LoTE by pipeting up and down.

3.2.6. Blunt Ending Released Tags

The tags now have a protruding end that needs to be filled in for blunt end ligation to form the ditags.

1. Set up blunt end ligation by using Klenow (2 U/µL).
2. Incubate at 37°C for 30 min.
3. Then, add 190 µL of LoTE (240 µL final volume).

4. Phenol:cholorform extract with 240 µL of PC8 (480 µL total volume).
5. Vortex, centrifuge for 5 min at 21,000*g* at 4°C.
6. Remove 200 µL of the upper aqueous phase containing the nucleic acids and aliquot into the "ligase-plus" tube; aliquot the rest of the aqueous upper phase into "ligase-minus" tube.
7. Equalize the volume of the plus and minus ligase samples by adding 160 µL of LoTE to the ligase-minus sample. EtOH precipitate both the plus and minus tubes with 7.5 *M* ammonium acetate.
8. Centrifuge for 30 min at 11,000*g* at 4°C.
9. Remove the supernatant and wash twice with 200 µL of 75% EtOH.
10. Resuspend the ligase-plus reaction in 5 µL of LoTE.
11. Resuspend the ligase minus reaction in 3 µL of LoTE.

3.2.7. Ligate to Form Ditags

Using high concentration T4 ligase (5 U/µL), set up a ligation to form ditags.

1. Add 5 µL of 2X ligase-plus mix to the ligase-plus tube.
2. Add 3 µL of 2X ligase-minus mix to the ligase-minus tube. Therefore, the total volume of ligase-plus is 10 µL and of ligase-minus is 6 µL.
3. Incubate overnight at 16°C.
4. Add 10 µL of LoTE to the ligase-plus tube (20 µL total volume).
5. Add 14 µL of LoTE to the ligase-minus tube (20 µL total volume).

3.2.8. PCR Amplification of Ditags

Use the P1 and P2 primers (350 ng/µL each) to amplify ditags for a test PCR to determine the optimal conditions for the large-scale PCR that follows.

1. Optimize amplification by using different dilutions of the ditag template (*see* **Note 18**). Cycle conditions are as follows: (A) 1 cycle: 94°C for 1 min; (B) 27 cycles (ligase-plus) or 35 cycles (ligase-minus): 94°C for 30 s; 55°C for 1 min; 70°C for 30 s; and (C) 1 cycle: 7°C for 5 min.
2. Use 27 cycles for the ligase-plus reactions, and 35 cycles for the ligase-minus reactions (*see* **Note 19**).
3. Remove 10 µL from each reaction and mix with 1 µL of loading dye. Electrophorese the samples on a 1-mm 12% polyacrylamide gel. Use 10- and 100-bp ladder as a marker. Amplified ditags should be 102 bp. A background band of equal or lower intensity occurs around 80 bp. All other background bands should be of substantially lower intensity. The ligase-minus samples should not contain any amplified product of the size of the ditags even at 35 cycles.

After the PCR test to determine the appropriate dilution, perform a large-scale PCR. About 300 reactions of 50 µL each of the optimal dilution is needed for the pooled PCR products. Three 96-well plates with a 50-L reaction per well is adequate. Use one less cycle for the large-scale than for the small-scale PCR.

3.2.9. Isolation of Ditags

1. After the large-scale PCR is complete, centrifuge the 96-well microplates in a swinging bucket centrifuge for 10 min to spin down the condensation.
2. Collect the PCR products into a 50-mL plastic conical tube and perform a phenol:cholorform extraction with an equal volume of PC8, and then centrifuge in a swinging bucket rotor for 10 min.
3. Transfer the upper aqueous phase to a new 50-mL tube and EtOH precipitate with 100% EtOH and 7.5 *M* ammonium acetate at –80°C.
4. Centrifuge at 11,000*g* 4°C for 30 min. Wash with 5 mL of 75% EtOH and centrifuge for 5 min at 4°C, followed by air-drying the pellet.
5. Resuspend each pellet in 200 µL of LoTE, pool the two tubes into one 400-µL sample, and then place it on ice. Add 40 µL of 10X loading dye.
6. Apply 110 µL to each of four 1.5-mm 12% acrylamide TAE gels by using one large well comb.
7. Electrophorese the gel for approximately 1.5 h at 160 V until the blue xylene band is near the bottom of the gel, and the purple bromophenol blue band has run off the gel.
8. Stain the gel using SYBR Green I stain, at a 1:10,000 dilution. Let the gel soak in the stain for 15 min on a shaker.
9. Visualize the gel on a UV box by using the yellow SYBR Green filter. Protect the DNA by putting an ethanol-cleaned glass plate on the UV box and then put the gel on the glass plate.
10. Cut out only the amplified ditag 102-bp band from the gel. Be sure to remove the markers and do not take the 80-bp background band.
11. Place one-third of each band in a 0.5-mL microcentrifuge tube (12 tubes total) whose bottom has been pierced two times with a 21-gage needle to form small holes of approx 0.5 mm in diameter.
12. Pierce the tubes with a syringe needle from the inside out for safety.
13. Place the 0.5-mL microcentrifuge tubes in 2-mL round-bottomed microcentrifuge tubes and centrifuge in microfuge at 11,000*g* for 5 min at room temperature. This process breaks up the cut-out bands into small fragments at the bottom of the 2-mL microcentrifuge tubes. If most of the gel is not contained in the 2-mL microfuge tube.
14. Centrifuge for an additional 5 min.
15. Discard 0.5-mL tubes and add 250 µL of LoTE and 50 µl of 7.5 *M* ammonium acetate to each 2-mL tube. Tubes can remain at 4°C overnight at this point (*see* **Note 20**).
16. Vortex each tube, place at 65°C for 15 min, and pulse centrifuge at 9000*g* to collect the condensation at the bottom of the tube.
17. Mix the contents well and the transfer the contents of each tube to a SpinX filtration tube. Transfer as much of the gel remnants and viscous solution as possible.
18. Centrifuge each SpinX tube for 5 min at 11,000*g* at room temperature. EtOH precipitate the eluates in new 1.5-mL tubes with 7.5 *M* ammonium acetate by centrifuging at 11,000*g* for 30 min at 4°C.

19. Wash twice with 200 µL of 75% EtOH for 5 min at 4°C for each wash.
20. Resuspend DNA in 10 µL of LoTE in each tube.
21. Pool samples into one tube, 120 µL total.
22. Remove 1 µL for quantitation of DNA concentration. The total amount of DNA at this stage should be 10 to 20 µg. If the concentration is not this high, EtOH precipitate and redo the large-scale preparation, and then combine the two 102-bp band samples. If the DNA concentration is too high, then the sample can be split in half, and the two digestion reactions can be electrophoresed. If the sample is split in half, then the reaction should be brought to volume with LoTE.
23. Digest the 102-bp DNA with NlaIII by incubating 1 h at 37°C (*see* **Note 21**).

3.2.10. Purification of Ditags

1. Extract with an equal volume PC8 (400 µL).
2. Vortex and then centrifuge for 5 min at 11,000*g* at 4°C. **Be very careful** in the following steps because the small 20-bp DNA segments are unstable. A low temperature and high salt concentration are needed.
3. Transfer the upper aqueous phase into two tubes (200 µL each) and then EtOH precipitate at –80°C.
4. Centrifuge the tubes at 4°C for 30 min at 11,000*g*.
5. Remove the supernatant and wash once with 200 µL of cold 75% EtOH. The EtOH is cold to protect the 26-bp ditags from denaturation. The melting point of 26-bp DNA is below room temperature.
6. Remove EtOH traces by air-drying on ice.
7. Resuspend the pellet in each tube in 10 µL of cold TE buffer. This higher concentration Tris-HCL buffer (compared with LoTE) is needed to protect the 26-bp DNA ditags.
8. Pool the resuspended DNA into one tube (20 µL total). On ice, add 2 µL of 10X loading dye (22 µL total volume).
9. Load 5.5 µL of this sample into four lanes of a 1.5-mm 12% polyacrylamide-TAE gel (10 well) and electrophorese at 100 V to 140 V (*see* **Note 22**). Load markers on both sides of the four lanes of sample leaving one open lane between the sample lanes and the marker lanes.
10. Stain the gel using SYBR Green I stain, at a 1:10,000 dilution.
11. Cut out the 24- to 26-bp band from the four lanes by using a glass plate, precleaned with EtOH, on top on the UV box.
12. Place two cut-out bands in each 0.5-mL microcentrifuge tube (two tubes total). More 0.5-mL microcentrifuge tubes can be used if more lanes are loaded with 26-bp ditags or if the bands are large. Pierce the bottom of 0.5-mL tube two times with a 21-gage needle and place the tubes in 2.0-mL round-bottomed microcentrifuge tubes and centrifuge in the microfuge at 11,000*g* for 5 min at room temperature. Continue to spin the tube until all of the gel is collected in the 2.0-mL tube.
13. Discard the 0.5-mL tubes and add 50 µL of 7.5 *M* ammonium acetate and 250 µL of LoTE (in this order) to the 2.0-mL tubes.

14. Vortex the tubes and place at 37°C for 20 min. Use four SpinX tubes to isolate the eluate.
15. Spin for 5 min at 10,000*g* at room temperature. EtOH precipitate in three tubes at –80°C. Centrifuge at 4°C at 10,000*g* for 30 min.
16. Wash twice with 200 µL of cold 75% EtOH. Air-dry to remove all residual EtOH.
17. Resuspend each DNA sample in 2.5 µL of cold (4°C) LoTE, making sure that the total volume is 8 µL after resuspending the pellet.
18. Remove 1 µL of the purified ditags for quantification.

3.2.11. Ligation of Ditags to Form Concatemers

Length of ligation time depends on quantity and purity of ditags. Typically, several hundred nanograms of ditags are isolated and produce large concatemers when the ligation reaction is carried for 1 to 3 h at 16°C. Lower quantities, or less pure ditags, will require longer ligations.

1. Set up a ligation by using the pooled purified ditags (7 µL), 5X ligation buffer (2 µL), and the high concentration (5 U/µL) T4 ligase (1 µL). If the volume of pooled purified ditags is high, the reaction volume can be increased.
2. Incubate for 1 h and 10 min at 16°C if you started with 10 µg of total RNA; fewer ditags require a shorter incubation.
3. Heat the sample at 65°C for 5 min.
4. Place it on ice for 10 min and then add loading dye to the ligation reactions. Use a 1-mm 8% polyacrylamide-TAE gel.
5. Load the concatemers in the center well, skip a lane on either side of the concatemers, and then use 100- and 1-kb ladders for molecular markers. Samples are electrophoresed at 130 V until the purple bromophenol blue dye reaches the bottom of the gel.
6. Stain the gel with SYBR Green I 1 at 10,000 dilution (*see* **Note 23**).
7. Visualize on UV box by using the yellow SYBR Green filter. Concatemers will form a smear with a range from about 100 bp to several kilobases.
8. Isolate the 800- to 1500-bp region and the 1500- to 3000-bp region. As before, place each of these gel pieces into a 0.5-mL microcentrifuge tube with a needle-pierced bottom (two tubes total).
9. Place the tubes in a 2.0-mL round-bottomed microcentrifuge tube and centrifuge at 11,000*g* for 5 min at room temperature. Discard the 0.5-mL tubes and add 300 µL of LoTE to the 2.0-mL tubes.
10. Vortex the tubes and place at 65°C for 20 min.
11. Transfer the contents of each tube into one SpinX microcentrifuge tube and centrifuge for 5 min at 11,000*g* at 4°C. EtOH precipitate both the high- and low-weight concatemers and place at –80°C overnight.

3.2.12. Ligating Concatemers Into pZero

1. Prepare the pZero cloning vector by digesting with the SpHI restriction enzyme, by incubating for 20 min at 37°C.

The Following Volumes Are in Microliters

Fraction	High mol. wt. fraction	Low mol. wt. fraction	Vector only	No ligase added
Vector	1	1	1	1
5X Ligase buffer	2	2	2	2
T4 Ligase (5 U/μL)	1	1	1	0
dH_2O	0	0	6	7
Concatemers	6	6	0	0

2. Remove 1 μL of the digestion to electrophorese on a 1% agarose TAE gel with markers, and uncut pZero. pZero should migrate in the gel at approx 2.5 kbp.
3. Add 30 μL LoTE to the rest of the sample and heat it at 70°C for 10 min to inactivate the enzyme. The concentration is now 25 ng/L, and 1 μL can be used per ligation.
4. Centrifuge the concatemers and pZero at 11,000*g* for 30 min at 4°C.
5. Wash the concatemers and pZero twice with 200 μL of 75% EtOH.
6. Resuspend each tube of the purified concatemer DNA in 6 μL of LoTE.
7. Resuspend the pZero in 30 μL of LoTE (~30 ng/L).
8. Incubate overnight at 16°C in 1.5-mL Eppendorf tubes. Before putting the ligation in, be sure that you can continue the next day.

3.2.13. Electroporation of Ligation Products and Colony PCR

1. Add 190 μL of LoTE to the ligation mix to bring the sample volume to 200 μL. Phenol:chloroform extract and precipitate at –80°C, and then centrifuge.
2. Transfer the upper aqueous upper layer to a new tube. EtOH precipitate the aqueous phase in 1.5-mL tubes at –80°C.
3. Centrifuge at 11,000*g* at 4°C for 30 min.
4. Wash four times with 200 μL of 70% EtOH.
5. Centrifuge at 11,000*g* for 5 min at 4°C.
6. Remove EtOH and air-dry the pellets.
7. Resuspend each pellet in 10 μL LoTE: **these mixtures are your ligation mixtures.**
8. Place ligation mixtures on ice.
9. Electroporate the 1 μL of ligation into 25 μL of ElectroMAX DH10Bs cells.
10. Plate the transformation mixture onto fresh LB-Zeocin plates (100 μg/mL Zeocin). One hundred microliters of liquid must be dispersed on each plate. Various amounts of both the high and low concatemer fractions are plated separately. The cells are spread by using glass beads with sterile technique. Two dilutions of each fraction are used: (A) 100 μL of undiluted electroporated competent cell culture and (B) 1:10 dilution of electroporated competent cell culture (diluted with LB media). (C) 100 μL of the two negative controls also are plated appropriately.

11. Incubate overnight at 37°C. Keep the rest of the transformation mixtures at 4°C. Plasmids with insert should have hundreds to thousands of colonies, whereas control plates should have zero to tens of colonies.
12. Check the insert sizes by performing colony-PCR. Check at least 36 colonies of each ligation to calculate the average insert size. Set up 25-µL PCR reactions by using the M13-forward and M13-reverse primers. Use a sterile pipet tip, or a toothpick, to pick colonies. Suggested parameters for PCR are as follows: (A) 1 cycle: 95°C for 2 min, (B) 30 cycles: 95°C for 30 s, 56°C for 1 min, 70°C for 1 min, (C) 1 cycle: 70°C for 5 min.
13. Electrophorese 5 µL of sample on a 1.5% agarose TAE gel at 100 V until the purple bromophenol blue band is near the bottom of the gel. Load a 1-kb and 100-bp ladder to determine the insert size. Place 1 µL of 10X loading buffer in the wells of a 96-well PCR plate. Transfer 5 µL of the PCR reactions to the dye by using the multichannel pipet and load into the gel.

3.2.14. Bioinformatics Analyses of SAGE Tags

The last step of SAGE is to sequence the colonies and to perform appropriate bioinformatics analyses on the sequenced products to elucidate SAGE tag sequences and to calculate relative expression levels of various tags in the sample (i.e., creation of a SAGE "library"). The frequency of each SAGE tag in the SAGE library directly correlates with transcript abundance. Sequencing of concatemers can be done either in-house in a core facility or at any commercial sequencing laboratory. The bioinformatics associated with analysis of SAGE tags is outside the scope of this chapter, but the reader is referred to excellent reviews on this subject ***(44–46)*** as well as to relevant websites (SAGENet, http://www.sagenet.org/; SAGEMap, http://www.ncbi.nlm.nih.gov/projects/SAGE/; TagMapper, http://tagmapper.ibioinformatics.org/index_html; and SAGE Genie, http://cgap.nci.nih.gov/SAGE). Of note, many of the databases house publicly available SAGE libraries of a variety of human tissues, cell types, and diseased specimens, including pancreatic cancer, and form a readily available resource for generating differentially expressed transcripts in an entity of interest.

3.3. Mass Spectrometric Analysis of Proteins

3.3.1. Sample Collection and Preparation of Pancreatic Juice

Human pancreatic juice is normally collected during surgery from the pancreatic duct of patients undergoing pancreatectomy. A total volume of 20–500 µL is usually collected and immediately stored at –80°C with or without any protease inhibitors. The sample should be kept at on ice (4°C) all time when not stored at –80°C. To remove potential debris the pancreatic juice is centrifuged at 4°C for 30 min at 16,000*g*. The protein concentration should be determined

using a protein assay kit. Several different protein assay kits can be used, including, e.g., Modified Lowry and Bradford. Modified Lowry (absorbance at 750 nm) is very sensitive, but it is a two-step procedure and requires an incubation time (~15–20 min). In addition, strong acids and ammonium sulfate can interfere with the measurement in Modified Lowry. The Bradford (absorbance at 590 nm) method is even more sensitive and can be read within 5 min. However, proteins with low arginine content will be underestimated when using the Bradford method. The protein concentration in pancreatic juice can vary from approx 2–15 mg/mL depending on the specific sample.

3.3.2. Sample Collection and Preparation Pancreatic Tissue

The amount of tissue needed depends on the type of analysis performed. This protocol is based on ~50 mg of starting material (tissue), which translates into approx 10–25 mg of protein after extraction. Note these numbers are only rough numbers and can fluctuate from sample to sample.

1. After surgical removal the pancreatic tissue is snap-frozen and stored at –80°C. During handling (cutting), the tissue is kept on a clean glass plate placed on top on an ice bucket to minimize degradation of the proteins. The tissue is first cut into small pieces by a sterile scalpel and transferred to a 1.5-mL tube. The tissue is then gently washed (by inverting the tube) in ice-cold PBS buffer to remove excess blood from the sample.
2. Remove the PBS buffer and replaced by 1 mL of ice-cold modified RIPA buffer containing protease inhibitors (150 m*M* NaCl, 50 m*M* Tris, pH 7.4, 1% NP-40, 0.25% sodium deoxycholate, 1 m*M* EDTA, and one protease inhibitor tablet per 50 mL).
3. The tissue is subsequently sonicated (output control, 3 to 4; duty-cycle, 30–40%; time, 30 s per cycle) three to four times. If possible the sample should be kept on ice during sonication but otherwise placed on ice for 5 min after every cycle of sonication. Alternatively, a Polytron homogenizer can be used followed by sonication for protein extraction (on ice).
4. After protein extraction, the sample is centrifuged for 30 min at 4°C, and the supernatant is transferred to a new 1.5-mL tube.

3.3.3. Fractionation of Pancreatic Juice and Tissue Proteins by 1D Gel Electrophoresis

Because of the relatively high complexity of pancreatic juice and tissue, it is recommended that the sample is fractionated by, e.g., 1D or two-dimensional (2D) gel electrophoresis before nano-LC-MS/MS analysis. Alternatively, one can perform in-solution digestion combined with 2D LC-MS/MS (not discussed in this chapter) ***(47,48)***. For automated nano-LC-MS/MS analysis, 20–30 µg of sample should be loaded on the gel (higher or lower amounts of sample can be loaded on the gel depending on the specific analysis).

Approximately 20–30 µg of protein is transferred to a 1.5-mL tube and mixed (1:5) with 6X loading buffer containing 10% β-mercaptoethanol. Vortex sample shortly (5 s) and boil for 5–8 min to denature the proteins before loading on 1D gel. For most purposes, a 10% gel is recommended, but alternatively a gradient gel can be used (4–12%) for higher resolution in a broad molecular mass range. The parameter for running the gel depends on the gel apparatus system and gel size and has to be done according to the individual manufacturer's instructions.

3.3.4. Staining of Proteins After 1D Gel Electrophoresis

After resolving the proteins by 1D gel electrophoresis, the proteins are visualized by either silver staining ***(49)*** or colloidal Coomassie ***(50)***. Silver staining can sensitize proteins down to 1–2 ng of protein ***(51)***, whereas colloidal Coomassie staining can detect levels down to approx 10–20 ng of protein. Both methods are equally compatible with mass spectrometry analysis.

3.3.5. Visualization of Proteins by Silver Staining

1. After resolving the proteins by 1D gel electrophoresis, fix the gel in destaining solution (50% methanol, 5% acetic acid, 45% water [Milli-Q], [v/v]) for 20–30 min at room temperature (on shaker).
2. Rinse in Milli-Q water for 30–60 min to remove acid (change water three to four times). However, the gel can be left overnight in water to reduce background (more transparent in areas with no protein staining).
3. Incubate the gel for 1 to 2 min in 0.02% sodium thiosulfate ($Na_2S_2O_3$).
4. Wash two times with Milli-Q water (1 min each time).
5. Then, incubate the gel in ice-cold (4°C) 0.1% silver-nitrate ($AgNO_3$) solution at 4°C for 20–40 min (on shaker).
6. Subsequently, wash the gel two times in Milli-Q water (two times for 1 min each) and then develop in developing solution (0.04% formaldehyde/formalin and 2% Na_2CO_3). The development is quenched by discharging the developing solution and replacing it with 1% acetic acid (*see* **Note 24**). After staining (silver or colloidal Coomassie), the gel is stored at 4°C in 1% acetic acid (container with tight lid). The gel can be stored for several months under these conditions.

3.3.6. Visualization of Proteins by Colloidal Coomassie Staining

1. After gel electrophoresis, fix the gel in destaining solution (50% methanol, 5% acetic acid, and 45% water [Milli-Q] [v/v]) for 10 min at room temperature (on shaker).
2. Discharge the destaining solution and replace it by staining solution without stainer B (55 mL of Milli-Q water, 20 mL of methanol, 20 mL of stainer A).
3. Incubate at room temperature for 10 min (on shaker).
4. Add 5 mL of stainer B to the existing solution and leave for 3–12 h.
5. When protein bands become visible, decant staining solution and replace by 200–300 mL of Milli-Q water.

6. Shake the gel for at least 7 h or until the background becomes clear. The water should be changed several times during destaining (*see* **Note 25**).

3.3.7. In-Gel Tryptic Digestion of Proteins Resolved by 1D Gel Electrophoresis

After visualization of the proteins by either silver staining or colloidal Coomassie staining, the resolved proteins are digested by trypsin before mass spectrometry analysis ***(17)***.

1. Excise the gel lane into approx 20–30 bands (depending on the size of the gel).
2. Cut each band into smaller pieces (~2 × 2 mm).
3. Wash the gel pieces from each band with Milli-Q water and water:acetonitrile 1:1 (v/v) two times for 15 min at room temperature. Solvent volumes in the washing step roughly equaled five times the gel volume.
4. After washing, remove the liquid and cover gel pieces in 100% acetonitrile to shrink the gel pieces.
5. After approx 5 min, remove acetonitrile and incubate the gel pieces it with 10 m*M* DDT in 100 m*M* ammonium bicarbonate (NH_4HCO_3) for 45 min at 56°C to reduce the cysteines.
6. Subsequently, remove the liquid and replace by 55 m*M* iodoacetamide in 100 m*M* NH_4HCO_3 and then incubate for 30 min at room temperature in the dark to alkylate the cysteines.
7. Remove the iodoacetamide solution surplus and wash the gel pieces two times in water and acetonitrile as described above and subsequently dehydrate by adding 100% acetonitrile.
8. Rehydrate the gel pieces by incubation in a digestion buffer containing 12.5 ng/µL trypsin (modified sequencing grade, Promega) in 50 m*M* NH_4HCO_3 and incubate for 45 min on ice. Enough digestion buffer is added to cover the gel pieces.
9. After 45 min, replace the digestion buffer by approximately the same volume 50 m*M* NH_4HCO_3 but without trypsin to keep the gel pieces wet during enzymatic digestion.
10. Incubate the samples at 37°C overnight.
11. After digestion, remove the remaining supernatant and save it in a 1.5-mL tube.
12. Extract the remaining peptides from the gel pieces by incubating in 5% formic acid (enough to cover the pieces) for 15 min and then adding the same volume of 100% acetonitrile for further 15-min incubation.
13. Repeat the extraction twice. Collect all three fractions in one tube and dry down in a vacuum centrifuge.
14. Resuspend in 10–20 µL of 5% formic acid.

3.3.8. Liquid Chromatography-Tandem Mass Spectrometry Analysis

Depending on the type of sample analyzed and available hardware (e.g., HPLC system and mass spectrometer), the LC-MS/MS setup can be modified

in many ways. However, two different strategies are frequently used for separation of peptides. One setup uses two columns in tandem (a precolumn followed by an analytical column), whereas the second setup only uses a single analytical column. The tandem column setup is preferred when large volumes (10–40 µL) have to be analyzed or if the samples need to be desalted or washed extensively. This setup is very robust and can be used for most types of samples. One drawback is that one might loose sensitivity because of broader chromatographic peaks. Several column materials can be used for precolumns (e.g., YMC ODS-A, 5- to 15-µm beads; Kanamatzu USA Inc., New York, NY) and analytical columns (e.g., MS218, 5-µm beads; Vydac) and it is therefore strongly suggested that several trial runs including different reverse phase materials are tested for obtaining optimal separation and resolution during the LC-MS/MS analysis.

The following sections describe a general strategy for building and packing reverse phase columns used for nano-LC-MS/MS analysis in addition to some general chromatographic guidelines (LC program, length of gradient, and context of mobile phase).

*3.3.9. Preparation of Reverse Phase Column (**Fig. 2**)*

1. The column is packed in a fused silica capillary tubing (375 µm od and 75 µm id) ***(52)***. A "frit" restrictor has to be generated inside the fused silica to block the reverse phase material during packing (the silica tubing is open in both ends). Mix 44 µL of Kvasil 1 with 8 µL of formamide and vortex for 1–3 min (the solution becomes viscous).
2. One end of a 20- to 25-cm fused silica is dipped in the solution for 1 to 2 s (the solution will move upward into the fused silica by capillary action). Excess solution is wiped off and the fused silica is left at room temperature o.n. to polymerize. The frit is checked under a microscope and should be approx 2–5 mm in lengths and has to present a V-shaped cone pointing outward. The frit can be cut if it is too long.
3. Insert the fused silica into a pressure vessel and wash it with 100% methanol. This process ensures that solvent can flow through the frit and that when pressure is applied, the frit is still affixed to the capillary. The reverse phase material (~4 mg) is placed in a glass microvial and resuspended in 500 µL of methanol.
4. To keep the column material in suspension a small magnetic stirring rod is added to the glass microvial before placing it in the pressure vessel. Place the pressure vessel on a stir plate and turn the magnetic stirrer at low speed. Turn the gas (He) pressure to 50 bar (800 psi), which initiates the packing of the column. The column should be 5–10 cm for precolumns and 15–20 cm for analytical columns.
5. The fused silica is connected to the LC system and flushed with mobile phase B (90% acetonitrile, 0.4% acetic acid, and 0.005% heptafluorobutyric acid in water) for 30 min and subsequently 30 min with mobile phase A (H_2O with 0.4% acetic acid and 0.005% heptafluorobutyric [v/v]) to equilibrate the column.

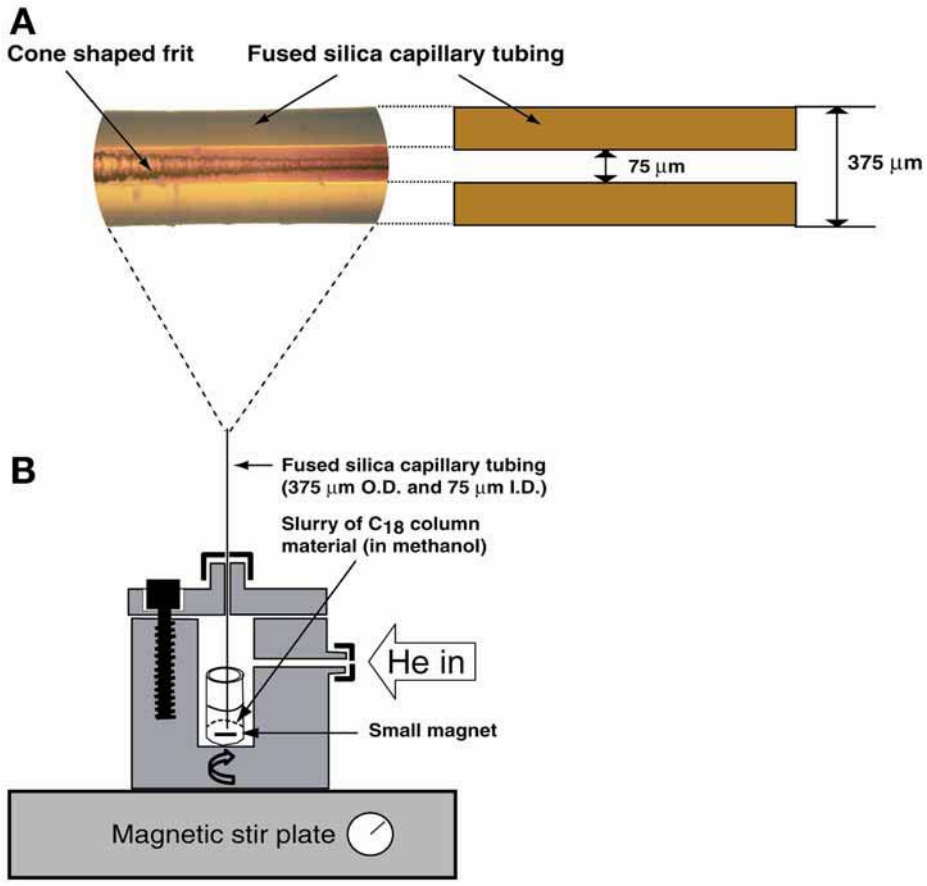


Fig. 2. Preparation of a reverse phase column for nano-LC-MS/MS analysis. A frit restrictor composed of a mix of Kvasil 1 and formamide is generated in one end of the fused silica capillary tubing (375 µm od and 75 µm id) to block the reverse phase material during packing. The frit restrictor should present a V-shaped cone and be 2–5 mm in length **(A)**. The fused silica tubing with the frit restrictor is subsequently placed in small pressure vessel (frit restrictor pointing outward from the pressure vessel) containing a slurry of reverse phase material **(B)**. The column is packed by applying a pressure (He) of 50 bar (800 psi) in the pressure vessel. The column should be approx 15–20 cm in length.

6. The performance of the generated reverse phase column is tested with several trial runs by a tryptic digest of a standard protein (e.g., 100–200 fmol of BSA). Elution time for the individual peaks in the chromatogram should be approx 15–20 s with 100 fmol of BSA and produce sequence coverage between 50 and 60% with a Mascot protein score of approx 1500 (*see* **Subheading 3.3.11.**).

3.3.10. Sample Analysis (LC-MS/MS)

The sample is automatically loaded onto the column by the autosampler in the HPLC system. The peptides are loaded onto the reverse phase column by using a linear gradient of 5–10% mobile phase B (90% acetonitrile, 0.4% acetic acid, and 0.005% heptafluorobutyric acid in water) during 6–8 min. No peptides will elute during this time interval. If needed, the loading time can be extended for additional washing or desalting of the sample. The peptides are subsequently separated and eluted from the column by 10–45% mobile phase B during 35–40 min, followed by 45–90% mobile phase B for 3 min. The column is finally rinsed flushed with 90% mobile phase B for 1 to 2 min and equilibrated in 5% mobile phase B for 3 min. The flow rate delivered from the pumps to the column(s) depends on the LC setup. In a tandem column setup, the flow rate should be 4 to 5 μL/min during loading and 1–1.5 μL/min for the single column setup. During peptide separation and elution, the flow rate is decreased to 250–300 nL/min for both setups. During equilibration of the column, the flow is again increased to 4 to 5 μL/min for tandem column setup and 1–1.5 μL/min for single column setup. Note that this program is only a rough guide and should be modified as required.

3.3.11. Database Search and Validation

The final step of the analysis is searching the generated LC-MS/MS data by using a search engine. Several search engines are available (e.g., Mascot, SEQUEST, XTandem, and Spectrum Mill) that use their own individual search algorithms and scoring systems. Here, the Mascot search engine (version 2.0) is used as an example. Before the search can be performed, peak-list files have to be generated. Peak-list files are made from the files initially recorded during the LC-MS/MS analysis (named, e.g. RAW-files on a Micromass instrument) and contain information about the mass of the precursor selected for MS/MS fragmentation and the masses and relative intensity of the fragment ions. Depending on the type of instrument used for the LC-MS/MS analysis, the search parameters have to be adjusted accordingly. The following Mascot search parameters can be used for high-resolution type instruments (e.g., Micromass QTOF and Sciex Q-STAR). Several search parameters have to be specified on the Mascot MS/MS ion-search, including peptide tolerance (set at 0.2 Daltons), MS/MS tolerance (set at 0.2 Daltons), enzyme (used in digestion, trypsin), missed cleavages (set at 1 or 2), fixed modifications (set carbamidomethylation on cysteins), variable modifications (set oxidation on methionine), peptide charge (set +1, +2, +3), and monoi-sotopic mass. One also has to specify the type of database (e.g., NCBInr, SwissProt, or RefSeq) and taxonomy (e.g. human, mouse, bacteria, or fungi) the peak list files has to be

search in. A widely used nonredundant database is RefSeq ***(53)*** (http://www.ncbi.nlm.nih.gov/RefSeq/) from National Center for Biotechnology Information, which contains 1,310.800 entries from 2780 organisms. One very important aspect of the final analysis is verification of the data obtained from the database searches. Far from all proteins retrieved from the database search are correct. Peptides identified by the Mascot search engine with a peptide score under 30–40 are usually false positives and have to be discarded from the data set. In cases where protein identification is based on a single peptide, special caution has to be taken. The amino acid sequence identified by the Mascot search engine has to be verified manually to make sure that the MS/MS spectrum truly corresponds to the peptide sequence predicted by Mascot. Manual validation is therefore essential to avoid a large number of false positives in the list of identified proteins.

4. Notes

1. After processing, the array spots will no longer be visible, so demarking the boundaries is important for later positioning the labeled sample and cover slip correctly.
2. Rehydration should produce spot sizes that are enlarged by approx 20%, with the DNA distributed more uniformly within spots.
3. Bring ddH_2O to a boil in the Pyrex dish and then remove from heat. Add microarrays in slide rack immediately after the boiling (bubbling) has subsided.
4. For processing oligonucleotide arrays, UV-crosslink by using 70 mJ, wash arrays in 0.2% SDS at room temperature for 10 min, wash three times in ddH_2O at room temperature for 5 min each, rinse in 95% EtOH for 1 min, and then spin-dry in centrifuge at 500 rpm.
5. Often by convention, the test sample is labeled with Cy5, whereas the reference sample is labeled with Cy3.
6. Centrifuge times are estimates. If necessary, here and in subsequent Microcon steps, spin in additional 1-min increments until volume of retained solution is approx 20 µL.
7. This additional wash step serves to further remove unincorporated fluorescent nucleotides. If labeling is successful, the retained labeled cDNA mixture should be light blue.
8. PolyA and human Cot-1 preannealing serves to block undesirable hybridization to polyA tails and highly repetitive DNA, respectively, contained in a subset of cDNA microarray elements. Yeast tRNA functions to block nonspecific hybridization.
9. A total volume of 40 µL for the hybridization solution is appropriate when using a 22- × 60-mm cover slip. If using a different-sized cDNA microarray and cover slip, adjust the total volume of hybridization solution accordingly, while maintaining final SSC and SDS concentrations.
10. The described hybridization protocol is optimized for cDNA microarrays. For oligonucleotide microarrays, improved hybridization sensitivity and specificity

can be achieved by performing the hybridization with 35% formamide (with the same concentration of SSC and SDS) and at 42°C for at least 24 h.
11. Add SDS last to avoid precipitation.
12. Performing the first wash above room temperature (65°C) provides higher stringency to increase specific to nonspecific hybridization signal. For oligonucleotide arrays, perform the first wash at 42°C.
13. Work quickly to avoid air-drying (which will leave salt residue) between the last wash and spin-dry steps.
14. High-quality linkers are crucial to several steps in the SAGE method. Linkers 1A, 1B, 2A, and 2B should be obtained gel-purified from the oligonucleotide synthesis company. These oligonucleotides can be ordered from Integrated DNA Technologies or other oligonucleotide synthesis companies.
15. The RNA volume should not exceed 100 µL.
16. A 1:10 dilution is often a good start.
17. Keep the supernatant; it contains your actual tags!
18. Use 1 µL of 1/50, 1/100, and 1/200 dilutions of ligation product per PCR reaction. This step indicates the best dilution for the reactions. If the starting RNA is poor quality, start out with a 1/25 dilution.
19. The appropriate cycle number is critical for isolating an adequate amount of ditag DNA for SAGE. Too few cycles result in a low yield and may cause problems with subsequent steps. Too many cycles give erratic results and also may result in low yields. Therefore, trying various cycle numbers (e.g., 26, 28, or 30) to determine the optimal number is recommended.
20. The DNA seeps out of the gel overnight. Do not leave for more than one night, or the acrylamide can break down substantially. Do not put at –80°C.
21. Freeze/thaw cycles decrease the effectiveness of NlaIII. Freeze/thaw NlaIII only once to insure optimal activity.
22. Do not run the voltage too high or your 26-bp product could melt.
23. This gel is thin and prone to rip. Be careful!
24. Developing of the gel should take approx 2 to 3 min and care should be taken not to overexpose the gel (have the 1% acetic acid solution ready for quenching the reaction).
25. Stainer A and B are provided in the colloidal Coomassie kit provided by Invitrogen.

Acknowledgments

A.M. is supported by an American Association for Cancer Research-PanCAN Career Development Award, the Lustgarten Foundation for Pancreatic Cancer Research, and National Cancer Institute R01CA113669.

References

1. Yeo, T. P., Hruban. R. H., Leach, S. D., et al. (2002) Pancreatic cancer. *Curr. Probl. Cancer* **26,** 176–275.
2. Takaori, K., Hruban, R. H., Maitra, A., and Tanigawa, N. (2004) Pancreatic intraepithelial neoplasia. *Pancreas* **28,** 257–262.

3. Kern, S. E., Hruban, R. H., Hidalgo, M., and Yeo, C. J. (2002) An introduction to pancreatic adenocarcinoma genetics, pathology and therapy. *Cancer Biol. Ther.* **1,** 607–613.
4. Louvet, C., Andre, T., Lledo, G., et al. (2002) Gemcitabine combined with oxaliplatin in advanced pancreatic adenocarcinoma: final results of a GERCOR multicenter phase II study. *J. Clin. Oncol.* **20,** 1512–1518.
5. Hruban, R. H., Iacobuzio-Donahue, C., Wilentz, R. E., Goggins, M., and Kern, S. E. (2001) Molecular pathology of pancreatic cancer. *Cancer J.* **7,** 251–258.
6. Li, D., Xie, K., Wolff, R., and Abbruzzese, J. L. (2004) Pancreatic cancer. *Lancet* **363,** 1049–1057.
7. Goggins, M., Hruban, R. H., and Kern, S. E. (2000) BRCA2 is inactivated late in the development of pancreatic intraepithelial neoplasia: evidence and implications. *Am. J. Pathol*. **156,** 1767–1771.
8. Day, J. D., Digiuseppe, J. A., Yeo, C., et al. (1996) Immunohistochemical evaluation of HER-2/neu expression in pancreatic adenocarcinoma and pancreatic intraepithelial neoplasms. *Hum. Pathol*. **27,** 119–124.
9. Caldas, C., Hahn, S. A., da Costa, L. T., et al. (1994) Frequent somatic mutations and homozygous deletions of the p16 (MTS1) gene in pancreatic adenocarcinoma. *Nat. Genet.* **8,** 27–32.
10. Hahn, S. A., Schutte, M., Hoque, A. T., et al. (1996) DPC4, a candidate tumor suppressor gene at human chromosome 18q21.1. *Science* **271,** 350–353.
11. Schutte, M., Hruban, R. H., Geradts, J., et al. (1997) Abrogation of the Rb/p16 tumor-suppressive pathway in virtually all pancreatic carcinomas. *Cancer Res.* **57,** 3126–3130.
12. Berman, D. M., Karhadkar, S. S., Maitra, A., et al. (2003) Widespread requirement for Hedgehog ligand stimulation in growth of digestive tract tumours. *Nature* **425,** 846–851.
13. Miyamoto, Y., Maitra, A., Ghosh, B., et al. (2003) Notch mediates TGF alpha-induced changes in epithelial differentiation during pancreatic tumorigenesis. *Cancer Cell* **3,** 565–576.
14. Iacobuzio-Donahue, C. A., Maitra, A., Shen-Ong, G. L., et al. (2002) Discovery of novel tumor markers of pancreatic cancer using global gene expression technology. *Am. J. Pathol.* **160,** 1239–1249.
15. Iacobuzio-Donahue, C. A., Maitra, A., Olsen, M., et al. (2003) Exploration of global gene expression patterns in pancreatic adenocarcinoma using cDNA microarrays. *Am. J. Pathol.* **162,** 1151–1162.
16. Iacobuzio-Donahue, C. A., Ashfaq, R., Maitra, A., et al. (2003) Highly expressed genes in pancreatic ductal adenocarcinomas: a comprehensive characterization and comparison of the transcription profiles obtained from three major technologies. *Cancer Res.* **63,** 8614–8622.
17. Gronborg, M., Bunkenborg, J., Kristiansen, T. Z., et al. (2004) Comprehensive proteomic analysis of human pancreatic juice. *J. Proteome Res.* **3,** 1042–1055.
18. Ryu, B., Jones, J., Blades, N. J., et al. (2002) Relationships and differentially expressed genes among pancreatic cancers examined by large-scale serial analysis of gene expression. *Cancer Res.* **62,** 819–826.

19. Argani, P., Rosty, C., Reiter, R. E., et al. (2001) Discovery of new markers of cancer through serial analysis of gene expression: prostate stem cell antigen is overexpressed in pancreatic adenocarcinoma. *Cancer Res.* **61,** 4320–4324.
20. Argani, P., Iacobuzio-Donahue. C., Ryu. B., et al. (2001) Mesothelin is overexpressed in the vast majority of ductal adenocarcinomas of the pancreas: identification of a new pancreatic cancer marker by serial analysis of gene expression (SAGE). *Clin. Cancer Res.* **7,** 3862–3868.
21. Rosty, C., Christa, L., Kuzdzal, S., et al. (2002) Identification of hepatocarcinoma-intestine-pancreas/pancreatitis-associated protein I as a biomarker for pancreatic ductal adenocarcinoma by protein biochip technology. *Cancer Res.* **62,** 1868–1875.
22. Koopmann, J., Fedarko, N. S., Jain, A., et al. (2004) Evaluation of osteopontin as biomarker for pancreatic adenocarcinoma. *Cancer Epidemiol. Biomarkers Prev.* **13,** 487–491.
23. Shalon, D., Smith, S. J., and Brown, P. O. (1996) A DNA microarray system for analyzing complex DNA samples using two-color fluorescent probe hybridization. *Genome Res.* **6,** 639–645.
24. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995) Serial analysis of gene expression. *Science* **270,** 484–487.
25. Pandey, A. and Mann, M. (2000) Proteomics to study genes and genomes. *Nature* **405,** 837–846.
26. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270,** 467–470.
27. DeRisi, J., Penland, L., Brown, P. O., et al. (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat. Genet.* **14,** 457–460.
28. Pollack, J. R. (2002) Comparative genomic hybridization using cDNA microarrays. In: *DNA microarrays: A Molecular Cloning Manual, Sambrook,* J., ed. Cold Spring Harbor Press, pp. 363–369.
29. Van Gelder, R. N., von Zastrow, M. E., Yool, A., Dement, W. C., Barchas, J. D., and Eberwine, J. H. (1990) Amplified RNA synthesized from limited quantities of heterogeneous cDNA. *Proc. Natl. Acad. Sci. USA* **87,** 1663–1667.
30. Polyak, K. and Riggins, G. J. (2001) Gene discovery using the serial analysis of gene expression technique: implications for cancer research. *J. Clin. Oncol.* **19,** 2948–2958.
31. Kenzelmann, M. and Muhlemann, K. (1999) Substantially enhanced cloning efficiency of SAGE (serial analysis of gene expression) by adding a heating step to the original protocol. *Nucleic Acids Res.* **27,** 917–918.
32. Angelastro, J. M., Klimaschewski, L. P., and Vitolo, O. V. (2000) Improved NlaIII digestion of PAGE-purified 102 bp ditags by addition of a single purification step in both the SAGE and microSAGE protocols. *Nucleic Acids Res.* **28,** e62.
33. Datson, N. A., van der Perk-de Jong, J., van den Berg, M. P., de Kloet, E. R., and Vreugdenhil, E. (1999) MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Res.* **27,** 1300–1377.

34. Peters, D. G., Kassam, A. B., Yonas, H., O'Hare, E. H., Ferrell, R. E., and Brufsky, A. M. (1999) Comprehensive transcript analysis in small quantities of mRNA by SAGE-lite. *Nucleic Acids Res.* **27,** e39.
35. Foss, C. A., Maitra, A., Iacobuzio-Donahue, C., Kern, S., snf Pomper, M. G. (2004) [125I]anti-Claudin 4 and [125I]anti-PSCA monoclonal antibodies as novel imaging agents for human pancreatic cancer in xenograft-bearing mice, In Proceedings of the Third Annual Meeting of the Society for Molecular Imaging, St. Louis, MO.
36. Hassan, R., Bera, T., and Pastan, I. (2004) Mesothelin: a new target for immunotherapy. *Clin. Cancer Res*. **10,** 3937–3942.
37. Thomas, A. M., Santarsiero, L. M., Lutz, E. R., et al. (2004) Mesothelin-specific CD8(+) T cell responses provide evidence of in vivo cross-priming by antigen-presenting cells in vaccinated pancreatic cancer patients. *J. Exp. Med*. **200,** 297–306.
38. Rosty, C. and Goggins, M. (2002) Early detection of pancreatic carcinoma. *Hematol. Oncol. Clin. North Am.* **16,** 37–52.
39. Nichols, L. S., Ashfaq, R., and Iacobuzio-Donahue, C. A. (2004) Claudin 4 protein expression in primary and metastatic pancreatic cancer: support for use as a therapeutic target. *Am. J. Clin. Pathol.* **121,** 226–230.
40. Michl, P., Buchholz, M., Rolke, M., et al. (2001) Claudin-4: a new target for pancreatic cancer treatment using *Clostridium perfringens* enterotoxin. *Gastroenterology* **121,** 678–684.
41. Offner, S., Hekele, A., Teichmann, U., et al. (2005) Epithelial tight junction proteins as potential antibody targets for pancarcinoma therapy. *Cancer Immunol. Immunother.* **54,** 431–445.
42. Tuteja, R. and Tuteja, N. (2004) Serial analysis of gene expression (SAGE): unraveling the bioinformatics tools. *Bioessays* **26,** 916–922.
43. Lash, A. E., Tolstoshev, C. M., Wagner, L., et al. (2000) SAGEmap: a public gene expression resource. *Genome Res*. **10,** 1051–1060.
44. Lal, A., Lash, A. E., Altschul, S. F., et al. (1999) A public database for gene expression in human cancers. *Cancer Res*. **59,** 5403–5407.
45. Wolters, D. A., Washburn, M. P., and Yates, J. R., 3rd. (2001) An automated multidimensional protein identification technology for shotgun proteomics. *Anal. Chem.* **73,** 5683–5690.
46. Washburn, M. P., Wolters, D., and Yates, J. R., 3rd. (2001) Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nat. Biotechnol*. **19,** 242–247.
47. Switzer, R. C., 3rd, Merril, C. R., and Shifrin, S. (1979) A highly sensitive silver stain for detecting proteins and peptides in polyacrylamide gels. *Anal. Biochem.* **98,** 231–237.
48. Neuhoff, V., Arold, N., Taube, D., and Ehrhardt, W. (1988) Improved staining of proteins in polyacrylamide gels including isoelectric focusing gels with clear background at nanogram sensitivity using Coomassie Brilliant Blue G-250 and R-250. *Electrophoresis* **9,** 255–262.

49. Rabilloud, T., Carpentier, G., and Tarroux, P. (1988) Improvement and simplification of low-background silver staining of proteins by using sodium dithionite. *Electrophoresis* **9,** 288–291.
50. Ishihama, Y., Rappsilber, J., Andersen, J. S., and Mann, M. (2002) Microcolumns with self-assembled particle frits for proteomics. *J. Chromatogr. A* **979,** 233–239.
51. Pruitt, K. D., Tatusova, T., and Maglott, D. R. (2005) NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Res.* **33,** D501–D504.

5

Molecular Classification of Breast Tumors

Toward Improved Diagnostics and Treatments

Therese Sørlie

Summary

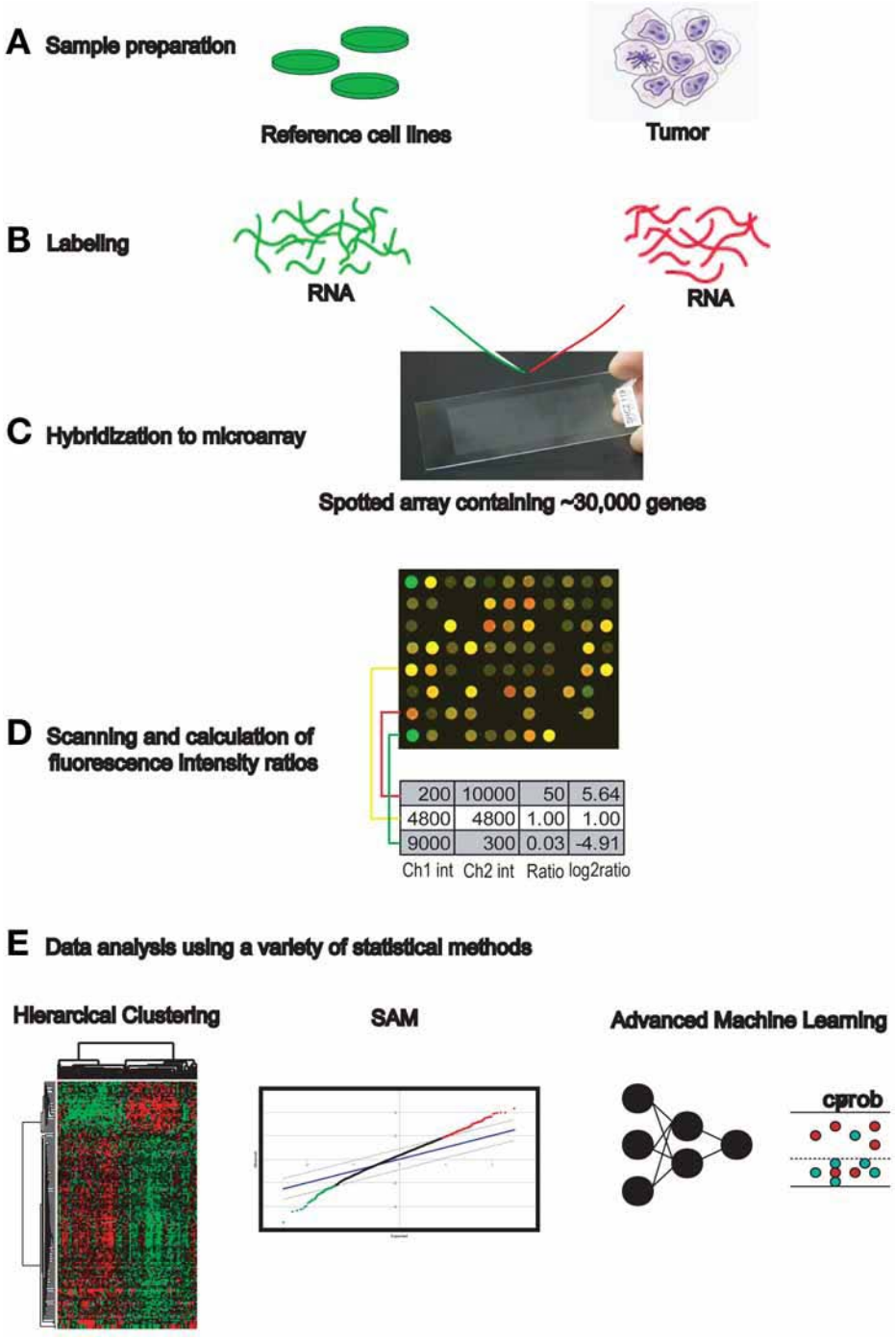
Recent advances in gene expression profiling and other "omics" technologies have revolutionized cancer research and hold the potential of also revolutionizing clinical practice. These high-throughout approaches have radically changed our ability to study cells and tissues in a more comprehensive way. Combined with advanced bioinformatics and the possibility to simulate biological processes in computers, this field of "systems biology" allows us to study the organism as a whole entity. This chapter describes the molecular classification and characterization of breast tumors into distinct subtypes by using DNA microarrays and discusses the statistical relationships of the subgroups with clinical features of the disease.

Key Words: Breast cancer, DNA microarrays, gene expression patterns; prediction; prognosis.

1. Introduction

Cancer diseases result from the accumulation of mutations, chromosomal instabilities and epigenetic changes that together facilitate an increased rate of cellular evolution and damage that progressively impairs the cell's detailed and complex system of regulation of cell growth and death. Changes in gene activities are further influenced by the microenvironment within and in the vicinity of tumor cells as well as by exogenous factors. When one combines all of these aspects with inborn genetic variations among individuals, there is every kind of reason to expect tumors to display prodigiously diverse phenotypes. Breast tumors, like most solid cancers, are heterogeneous and consist of several pathological subtypes with different histological appearances of the malignant cells and different clinical presentations and outcomes, and the patients show a diverse range of responses to a given treatment. Furthermore, breast tumor tissue shows

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarker Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ


A Sample preparation
Reference cell lines
Tumor
B Labeling
RNA
RNA
C Hybridization to microarray
Spotted array containing ~30,000 genes
D Scanning and calculation of fluorescence intensity ratios
200 10000 50 5.64
4800 4800 1.00 1.00
9000 300 0.03 -4.91
Ch1 int Ch2 int Ratio log2ratio
E Data analysis using a variety of statistical methods
Hierarcical Clustering
SAM
Advanced Machine Learning
cprob

heterogeneity with respect to its microenvironment, including specifically the types and numbers of infiltrating lymphocytes, adipocytes, and stromal and endothelial cells. The cellular composition of tumors is a central determinant of both the biological and clinical features of an individual's disease.

Microarray technologies, applied to the study of DNA, RNA, and protein profiles as well as to the genome-wide distribution of epigenetic changes such as DNA methylation, can be used to portray a tumor's detailed phenotype in its unique context (*see* Chapters 1 and 4, this volume). Systematic and detailed characterization of tumors on a genomic scale can be correlated with clinical information and greatly enhance our understanding of the causes and progression of cancer, ability to discover new molecular markers, and possibilities for therapeutic intervention. Eventually, advances in tumor portraiture will lead to improved and individualized treatments.

In recent years, the use of DNA microarrays in breast cancer research has led to important discoveries: first, individual tumors arising in the same organ may be grouped into distinct subclasses based on their gene expression patterns, independent on stage and grade; and second, the biological relevance of such classification is corroborated by significant prognostic impact ***(1)***.

2. Microarray Procedures and Handling of Data

Most published studies have used spotted cDNA arrays that were originally introduced by Schena and colleagues in 1995 ***(2)***; however, commercially manufactured oligonucleotide-based arrays are increasingly gaining market also among academic laboratories. Most of the platforms allow the use of two different fluorescent labels to distinguish, on the same spots, the abundance of gene-specific transcripts from two different samples. The data presented herein

---

Fig. 1. *(Opposite page)* Strategy for a typical reference-based DNA microarray experiment. **(A)** Tumor biopsies are obtained and immediately frozen for isolation of high-quality RNA. In parallel, RNA from a reference sample, usually a collection of different cell lines is isolated for use in a two-color hybridization protocol. **(B)** The two samples are differentially labeled with fluorescent nucleotides (by convention Cy5 for tumor RNA and Cy3 for reference RNA). **(C)** The two samples are combined and allowed to hybridize to a DNA microarray in which each gene is represented as a distinct spot. **(D)** A laser scanner is used to excite the hybridized array at the appropriate wavelengths, and the relative abundance of the two transcripts is visualized in a pseudo-colored image by the ratio of the "red" to "green" fluorescence intensities at each spot. The ratios are log transformed and placed in a table in which each row corresponds to a gene and each column corresponds to a single hybridization experiment. **(E)** For further data analysis and interpretation of multiple experiments, a wide range of statistical methods exist, both supervised and unsupervised: average-linkage hierarchical clustering; SAM, and machine learning algorithms. Adapted from **ref. *1***.

are based on such a platform. We have routinely used a reference strategy in which transcripts (or genomic DNA for copy number detection) extracted from the tumor sample were labeled with one fluorescent dye (Cy5), whereas transcripts from a standard reference *(3)* were labeled with a different dye (Cy3). Relative levels of the two transcripts were calculated by log transforming the ratio between the two fluorescent intensities (Cy5/Cy3; red/green) (**Fig. 1**).

Appropriate treatment and handling of surgical specimens is crucial to obtain high-quality, intact RNA, which is a fundamental requirement for a successful and reliable microarray experiment. Immediate snap-freezing of clinical samples after surgery or storage in appropriate buffers that permeate the cells and stabilize the RNA reduce the chances of jeopardizing both the quality and the quantity of RNA available for analysis.

The advantage of microarray experiments when applied to the study of cancer is most clear when a large number of samples is analyzed similarly and combined so that variation in gene expression patterns across a large number of tumors can be investigated. For the analysis and interpretation of microarray data, several sophisticated computational tools are available *(4,5)*, and constantly being developed. In general, they can be divided into unsupervised approaches that entail searching for patterns in the data with no prior assumptions, and supervised approaches in which predictions are generated based on existing knowledge of the data. As more primary data are compiled on specific tissues and tumor types, integrative analyses that explore the data in the context of other data sources may aid in a deeper biological understanding *(6)*. In our analyses of breast tumors, hierarchical clustering algorithms have been applied to organize both genes and samples into meaningful groups based on similarity in their overall expression patterns *(7)*. In addition, different supervised methods to analyze the gene expression profiles in relation to existing knowledge of the tumor samples have been successfully used *(8–11)*. Another great challenge of microarray experiments is the storage of the massive amounts of data that are generated in these experiments. Common standards and ontologies for the management and sharing of data are essential and have been put forward by the microarray community *(12)*. At present, the challenges of microarray technology include the use of different platforms, issues relating to consistent reproducibility, sample variability, and high cost. Despite these issues, more data comparing different data sets and various microarray platforms illustrate that the underlying biology is still the largest contributing factor *(13,13a)*.

3. Molecular Portraits of Breast Tumors

3.1. Whole-Genome Profiling

The phenotypic diversity of tumors is accompanied by a corresponding diversity in gene expression patterns that can be captured by DNA microarrays.

Our initial studies of 65 surgical specimens of human breast tissue from 42 individuals by using microarrays representing 8102 genes showed that there was great molecular heterogeneity among the tumors, with multidimensional variation in the patterns of gene expression ***(14)***. To help provide a framework for interpreting the variation in expression patterns observed in the tumor samples, 17 cultured cell lines representing models for many of the cell types encountered in these tissue samples also were characterized. A subset of 1753 genes was selected, those genes whose transcripts varied in abundance by at least fourfold from their median abundance in this sample set in at least three samples. Hierarchical clustering was then used to group both genes and tissue samples together based on overall similarity in their expression patterns (**Fig. 2**). Clusters of genes with coherent expression patterns could be related to specific features of biological variation among the samples, for example, variation in proliferation rates: the largest distinct cluster of genes was the "proliferation cluster," which is a group of genes whose levels of expression correlate with cellular proliferation rates ***(15–17)***. Another large cluster containing genes regulated by the type I and type II interferon signal transduction pathways showed substantial variation in expression among the tumors. Furthermore, clusters of coexpressed genes were identified whose expression patterns derived from different cell types within the grossly dissected tumors, including luminal and basal epithelium, stromal fibroblasts, and lymphocytes **(Fig. 2C–J)**. These findings were corroborated by immunohistochemistry ***(14)***.

A unique quality of this study was the availability of tumor pairs; 20 tumors were sampled twice as part of a larger prospective study on locally advanced breast cancer (T3/T4, N2 tumors, or both) ***(18–20)***. After an open surgical biopsy to obtain the "before" sample, each of these patients was treated with doxorubicin for an average of 15 wk, followed by resection of the remaining tumor, when the "after" sample was obtained. In addition, primary tumors from two patients were paired with a lymph node metastasis from the same individual. A striking result observed in the cluster dendrogram was that most of the tumor pairs (15 of the 20 before and after pairs and both primary tumor/lymph node metastasis pairs) clustered together on terminal branches in the accompanying dendrogram **(Fig. 2B)**. That is, despite the potential confounding effects of an interval of 15 wk, cytotoxic chemotherapy, and different sample preparations, independent samples taken from the same tumor were in most cases recognizably more similar to each other than either was to any of the other samples. This finding implied that every tumor is unique and has a distinctive gene expression signature or portrait. It also implied that the type and numbers of nonepithelial cells in tumors is a remarkably consistent and enduring feature of each individual tumor. Breast tumors thus seem to be very diverse, but they

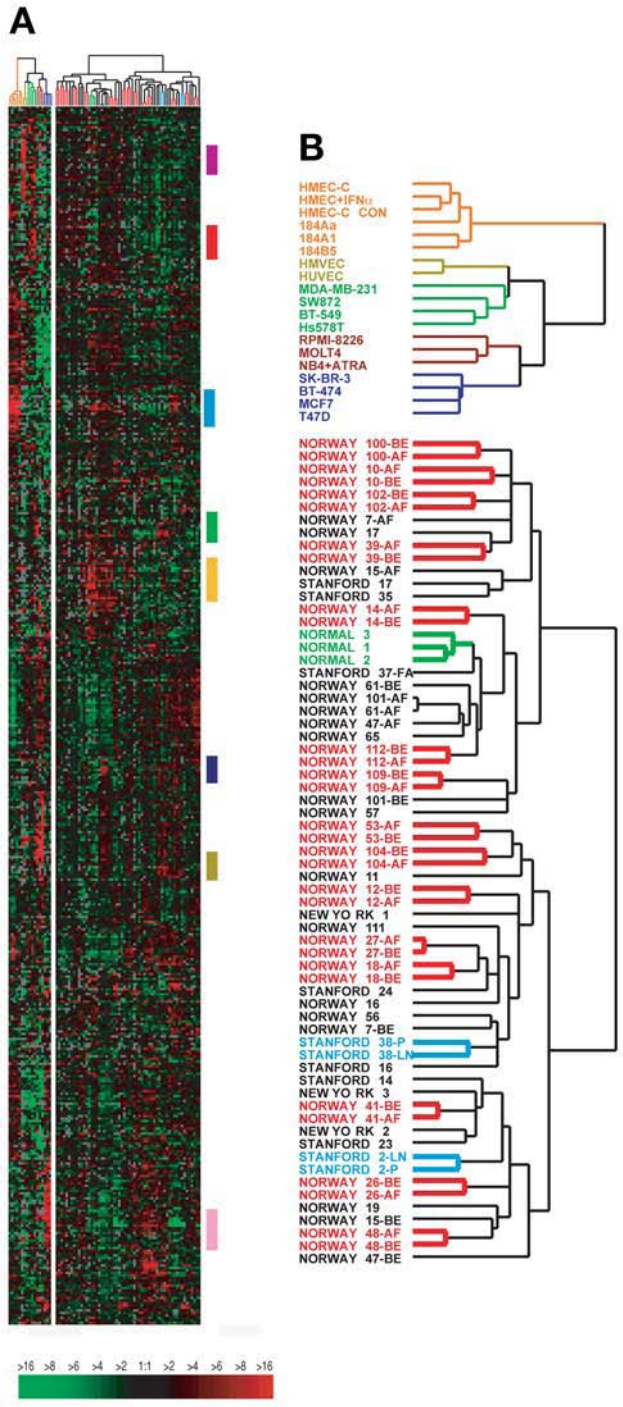


Fig. 2. Cluster analysis of a subset of genes in 84 experimental samples. Median-centered hierarchical clustering analysis of cell lines (left) and tissue samples (right). **(A)** Scaled-down representation of the 1753 gene cluster diagram; the colored bars to

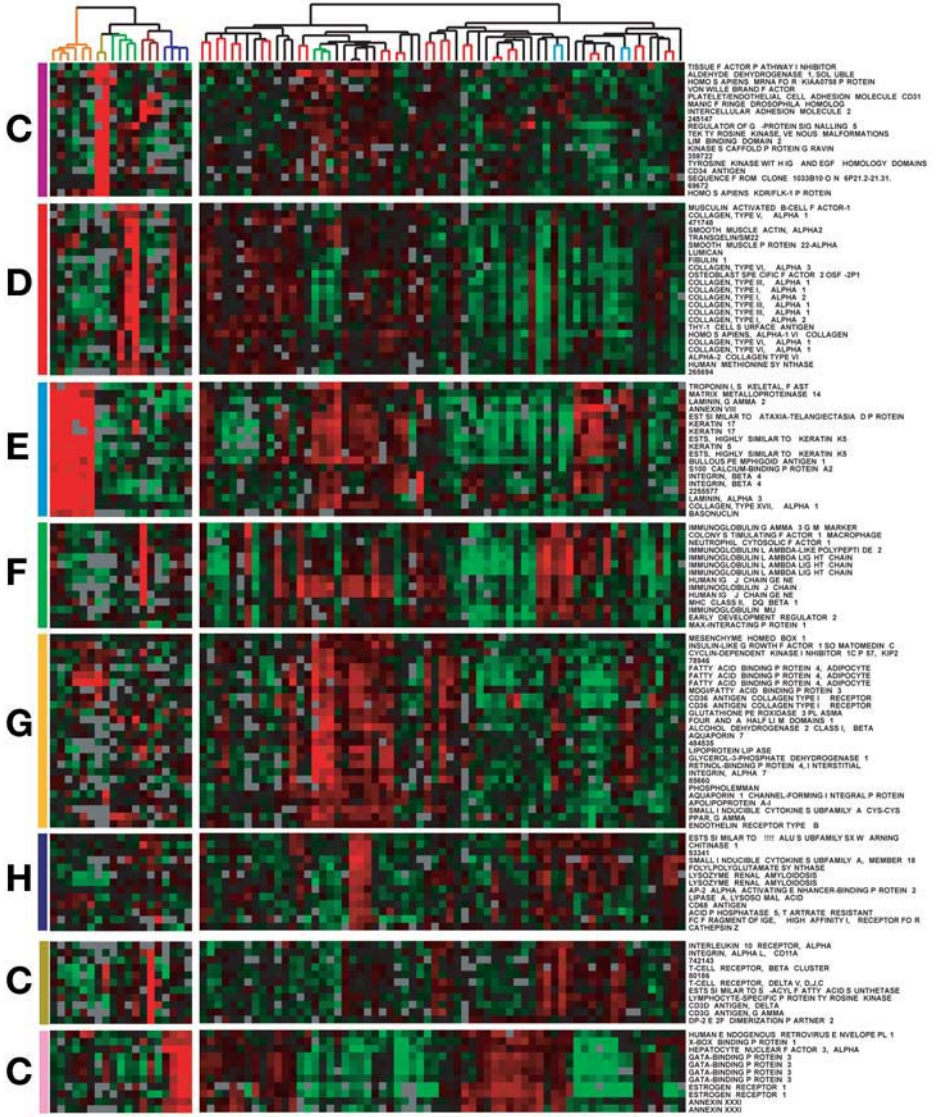


Fig. 2. *(Continued)* the right identify the locations of the inserts displayed in 1c–j. **(B)** Dendrogram representing similarities in the expression patterns between experimental samples. All tumor pairs that clustered together on terminal branches are highlighted in red, the two primary tumor/lymph node metastasis pairs are highlighted in light blue, and the three clustered normal breast samples are highlighted in light green. The branches representing the four breast luminal epithelial cell lines are displayed in dark blue, breast basal epithelial cell lines are displayed in orange, the endothelial cell lines in dark yellow, the mesynchemal-like cell lines in dark green, and the lymphocyte-derived cell lines in brown. **(C)** Endothelial cell gene expression cluster. **(D)** Stromal/fibroblast cluster. **(E)** Breast basal epithelial cluster. **(F)** B-cell cluster. **(G)** Adipose-enriched/normal breast. **(H)** Macrophage. **(I)** T-cell. **(J)** Breast luminal epithelial cell. Adapted from **ref.** ***14***.

are not internally heterogeneous, when millions of cells are sampled. The finding that the two primary tumors and the corresponding metastases were similar in their overall pattern of gene expression suggests that the molecular program of a primary tumor may be retained in its metastases. This idea has been explored by other investigators and suggests that metastatic capacity is an intrinsic feature of the primary tumor and that the fate of the tumor might have been programmed at an early stage of the disease ***(21)***. Hence, specific expression patterns in the primary tumors may help identifying genes, pathways, or both that are important for the metastatic process and for tumor progression in general.

3.2. Subclassification

To explore the possibilities for further refining these distinctions among subtypes of breast tumors, we took advantage of the paired tumor samples. The specific features of a gene expression pattern that are to be used as the basis for classifying tumors should be similar in any sample taken from the same tumor, and they should vary among different tumors. The paired samples therefore provided a unique opportunity for a deliberate and systematic search for genes whose expression levels reflected such intrinsic characteristics of the tumors. Using well measured expression data from the paired samples, a subset of genes termed "intrinsic gene set" was selected that consisted of genes with significantly greater variation in expression between different tumors than between paired samples from the same tumor ***(14)***. The rationale behind this list of genes is that it is enriched for those genes whose expression patterns were characteristic of each tumor as opposed to those that varied as a function of tissue sampling; hence, they would be ideally suited for classification.

Our extended analyses have since included an intrinsic gene set of 540 genes selected from expression data of 45 tumor pairs (including two primary lymph node pairs) and approx 8000 common genes for all samples ***(22,23)***. Together, 122 tissue samples were included in the clustering analysis: 115 carcinomas and 7 nonmalignant tissues. Most of the tumors were sampled as part of two independent studies evaluating response to chemotherapy of locally advanced breast cancer in an adjuvant setting. From the first cohort of patients treated with doxorubicin monotherapy ***(19)***, 55 tumor samples were analyzed, and from the second cohort of patients treated with 5-fluorouracil and mitomycin C ***(20)***, 34 tumor samples were analyzed. The remaining 26 samples were primary tumor specimens collected either at Stanford or in Norway.

The overall expression patterns showed that the tumors could be classified into five distinct subtypes, and the main distinction was between tumors that expressed genes characteristic of luminal epithelial cells (two groups), including the estrogen receptor (ER) and those that were negative for these genes (three groups) **(Fig. 3)**. Among the luminal epithelial-type tumors, the largest group,

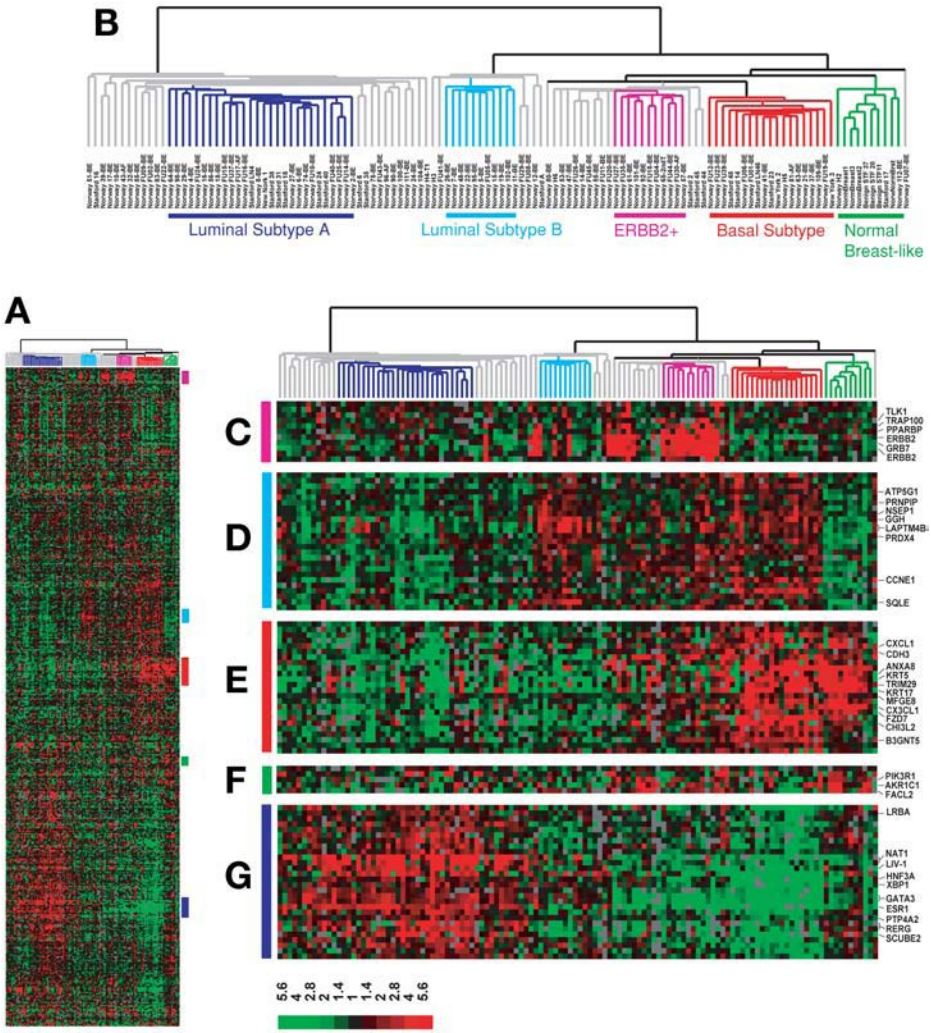


Fig. 3. Hierarchical clustering of 115 tumor tissues and seven nonmalignant tissues by using the intrinsic gene set. **(A)** A scaled-down representation of the entire cluster of 540 genes and 122 tissue samples based on similarities in gene expression. **(B)** Experimental dendrogram showing the clustering of the tumors into five subgroups. Branches corresponding to tumors with low correlation to any subtype are shown in gray. **(C)** Gene cluster showing the *ERBB2* oncogene and other coexpressed genes. **(D)** Gene cluster associated with luminal subtype B. **(E)** Gene cluster associated with the basal subtype. **(F)** A gene cluster relevant for the normal breast-like group. **(G)** Cluster of genes including the ER (*ESR1*) highly expressed in luminal subtype A tumors. Scale bar represents -fold change for any given gene relative to the median level of expression across all samples. Adapted from **ref.** ***22***.

termed luminal subtype A (dark blue branches), demonstrated the highest expression of ER, estrogen-regulated protein LIV-1, and the transcription factors hepatocyte nuclear factor 3, alpha (HNF3A), X-box binding protein 1, and GATA-binding protein 3 (GATA3) **(Fig. 3G)**. The second, smaller group of tumors with luminal epithelial characteristics, termed luminal subtype B (light blue branches), showed low expression of the ER cluster, but it was further distinguished from luminal subtype A by the high expression of a novel set of genes such as *GGH*, *LAPTMB4*, *NSEP1*, and *CCNE1*, whose coordinated function is unknown **(Fig. 3D)**. Among the three ER-negative groups, the basal epithelial-like subtype (red branches) was characterized by high expression of *KRT5* and *KRT17*, annexin 8, *CX3CL1*, and *TRIM29* and was completely negative for the luminal/ER cluster of genes **(Fig. 3E)**, whereas the ERBB2+ subtype (pink) was characterized by high expression of several genes in the ERBB2 amplicon at 17q22.24, including *ERBB2*, *GRB7*, and *TRAP100* **(Fig. 3C)**, suggesting that much of the genetic influences are because of gene amplification. A normal breast tissue-like group (green branches) was identified that showed the highest expression of many genes known to be expressed by adipose tissue and other nonepithelial cell types **(Fig. 3F)**. These tumors also showed strong expression of basal epithelial genes and low expression of luminal epithelial genes. However, it is unclear whether these tumors represent poorly sampled tumor tissue or a distinct, clinically important group. In conclusion, the molecular portraits defined by gene expression patterns seem to be stable and homogenous and represent the tumor itself, not merely the particular tumor sample, because we could recognize the distinctive expression profile of a tumor in independent samples. The inherent properties of the tumors thus seem to be sustained throughout chemotherapy as well as between a primary tumor and its lymph node metastasis, and these properties could be represented by a relatively small number of genes whose variation in expression formed a platform for classification.

3.3. Validation

The robustness of the tumor subtypes has been validated in independent data sets by us ***(22)*** as well as by others ***(24–29)***. When conducting a similar analysis of a breast tumor data set published by van't Veer et al. ***(30)***, to display the expression pattern of 460 of the intrinsic genes previously mentioned across their 97 tumor samples, the same subtypes were noted. In particular, the basal-like subtype was the most homogenous group of tumors, tightly clustered together with high correlation to the intrinsic expression centroid representing this phenotype ***(22)***. This relationship has been true for all data sets analyzed this far, even though they have comprised different patient populations, indicating that the substantial differences in the characteristics of

the patients are less important determinants of tumor expression phenotypes than intrinsic biology. Our studies also show that although there is variability between microarray platforms, the gene expression profiles emerging from using different technologies are highly correlated to the biological variation in the data, and the same tumor subtype pattern is identified ***(13a)***.

4. Clinical Implications

A major goal in the field of oncogenomics is to try to answer the clinically important questions about which tumors will behave aggressively, which tumors will remain dormant, which patients do and do not require systemic therapy, and what type of drugs should be used.

4.1. Prognosis

To investigate whether the five different tumor subtypes identified by gene expression profiling may represent clinically distinct and relevant groups of patients, univariate survival analyses comparing the subtypes with respect to overall survival (OS) and relapse-free survival (RFS) were performed. (For these analyses, only the patients who were enrolled in the two prospective studies on locally advanced disease were included.) The Kaplan–Meier curves based upon four subclasses (excluding the normal-like group) showed a highly significant difference in OS between the patients belonging to the different subclasses ($p < 0.01$) **(Fig. 4)**. Specifically, the basal-like and ERBB2+ subtypes were associated with the shortest survival time. Overexpression of the ERBB2 oncoprotein is a well-known prognostic factor associated with poor survival in breast cancer ***(31)***, which also was found for the ERBB2+ group defined in this study.

Perhaps the most intriguing result was the considerable difference in outcome observed between tumors classified as luminal A versus luminal B; the luminal B group showed a much worse prognosis compared with the luminal A group. Notably, each patient included in the survival analysis and harboring ER-α positive tumors were exposed to tamoxifen subsequent to preoperative chemotherapy and surgery ***(19,20)***. The luminal subtype B tumors may represent a clinically distinct group with a different and worse disease course, in particular with respect to relapse, and that may have little effect of tamoxifen treatment. The potential clinical significance of this molecular subtype is further highlighted by the similarities in expression of some of the luminal B genes with the ER-negative tumors in the basal-like and ERBB2+ subtypes, which suggests that high level of expression of these genes is associated with poor disease outcomes.

Most microarray studies of breast tumors to date have investigated the prognostic value of the identified gene signatures ***(26,30,32–37)***. Together, they reveal not only interesting information but also contradictions with respect to prognosis. Hence, the real value of the prognosis based on gene expression

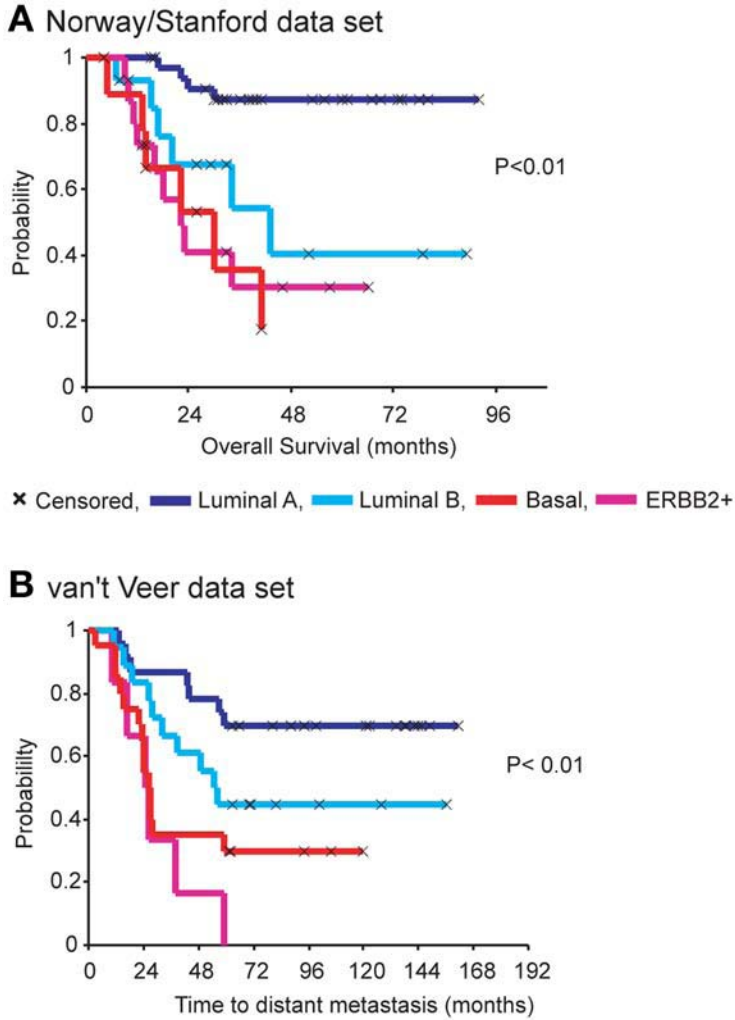


Fig. 4. Kaplan–Meier analysis of disease outcome in two patient cohorts. **(A)** Overall survival for 72 patients with locally advanced breast cancer in the Norway cohort. The normal-like tumor subgroups were omitted from both data sets in this analysis. **(B)** Time to development of distant metastasis in the 97 sporadic cases from **ref.** ***30.*** (Adapted from **ref.** ***22.***)

profiling needs to be further addressed. In a recent study, Eden et al. ***(38)*** reported that an optimal combination of "conventional" prognostic factors was of similar value as published gene signatures. The great advantage of gene profiles as opposed to conventional prognostic factors is the ability of the biological causes, beyond the mere empirical observations, to occur, and to be able to dissect the molecular biology discriminating different tumor forms ***(1)***.

4.2. Prediction

Despite reduced mortality because of earlier diagnosis and implementation of adjuvant chemo- and hormone therapies, breast cancer is still the most common cause of cancer death in women worldwide ***(39)***. The patients selected for adjuvant therapy (based on current criteria) experience a reduction is death hazard ratio only in the range of 35–30% ***(40)***. Even with moderate improvement in outcome with use of novel drug combinations ***(41)***, we are left with the truth that adjuvant therapy may cure only a minority of the patients. To select patients for adjuvant treatment, both prognostic and predictive factors are used. In brief, a prognostic factor defines outcome with respect to survival or relapse-free survival in a group of patients (ideally) not exposed to systemic therapy. In contrast, a predictive factor should define sensitivity of a tumor to a distinct therapeutic agent. With the exception of expression of ER-α and progesterone receptor for endocrine therapy and ERBB2+ amplification for the response to anti-ERBB2 treatment (Trastuzumab®), no predictive factor has unequivocally been accepted. Thus, the selection of patients for adjuvant therapy is mainly based on prognosis rather than sensitivity to treatment. Very few microarray studies have addressed the question of prediction of treatment response by using gene expression patterns ***(42–45)***. However, the predictive values reported in these studies were not robust enough for use in clinical practice. A key problem when characterizing "biological systems" is the high degree of genetic redundancy. For example, manipulating *TP53* in *TP53*-defective cell lines may result in altered expression of several hundred genes ***(46,47)***. It is most likely that only a few of these genes are associated with the execution of growth arrest or apoptosis. Similarly, the observation that prognosis or response to drug therapy varies between tumors that are characterized by different gene expression patterns is most likely because there are small "subgroups" of genes within the clusters whose function determines the outcome. The question remains how to identify these targets. The mechanisms involved in sensitivity, resistance, or both to therapy are most likely complex and multifactorial; thus, prognosis, or the risk for a particular outcome, should be assessed as a vector composed of several parameters: metastatic potential, growth rate of the tumor, and effect of therapy ***(1)***. Different subgroups of tumors from the same organ may respond differently to therapy: hence, molecular stratification of tumors is needed to be able to uncover biologically and statistically valid relationships to treatment response.

5. Molecular Characteristics of the Subtypes

5.1. Proliferation Cluster

The previously defined proliferation cluster is a group of genes whose levels of expression correlate with cellular proliferation rates ***(15,16)***. Expression of

this cluster of genes varied widely among the tumor samples analyzed in our study and was generally well correlated with the mitotic index. Genes encoding two generally used immunohistochemical markers of cell proliferation, Ki-67 and proliferating cell nuclear antigen, also were included in this cluster. The proliferation cluster has been observed in virtually every tumor type examined, including ovarian carcinomas ***(48)***, lymphomas ***(49)***, liver ***(50)***, and prostate cancer ***(51)***. More than half of the genes in the proliferation cluster were shown to be cell cycle regulated when the patterns of expression for these genes were analyzed in synchronized HeLa cell cultures ***(17)***. To investigate the expression of these genes in relation to the five subtypes, expression data were extracted, and the genes were clustered using average-linkage clustering, but the samples were ordered according to the subtypes as presented in **ref. *22***. What could be learned from this analysis was that the basal-like and the luminal subtype B both highly expressed these proliferation-associated genes, whereas luminal A, the normal-like and to some extent, the ERBB2+ subtype, were mostly negative for the expression of these genes. This finding may indicate large differences in the amount of cycling cells among the tumors. That the ERBB2+ tumors were less proliferative, despite being characterized by overexpression of an oncogene and being a poor prognostic group, was somewhat surprising. However, this characteristic underlines the very strong influence of the ERBB2 amplicon on the expression patterns of these tumors; no other distinct molecular signature protrudes at the transcriptional or at the genomic level (assessed by array comparative genomic hybridization [aCGH]).

5.1.1. Effects of BRCA1 Mutations on Global Gene Expression

Germline mutations in the *BRCA1* gene predispose carriers to early onset breast and ovarian cancer ***(52)***. A previous microarray study reported significantly different gene expression profiles of tumors from BRCA1 vs. BRCA2 carriers ***(53)***. In the data set produced by van't Veer and colleagues ***(30)***, breast tumors from 18 carriers of *BRCA1* mutations and two carriers of *BRCA2* mutations, also were analyzed. When we included these 20 tumors along with the 97 sporadic tumor samples in the clustering analysis, we saw little difference in the overall pattern, except for the striking result that all the tumors from patients carrying *BRCA1* mutations fell within the basal-like tumor subgroup ***(22)***. This finding indicates that a mutation in the *BRCA1* gene predisposes for the basal-like tumor subtype, which is associated with lack of expression of the ER receptor and poor prognosis. As also reported previously, *BRCA1*-associated breast cancers are usually highly proliferative, *TP53* mutated, and lack expression of *ESR1* and *ERBB2* ***(54,55)***. Recent studies have found that BRCA1 interacts with and regulates the activity of the ER and the androgen receptor ***(56)***, thereby providing a link between BRCA1 and hormone-related cancers. Furthermore, wild-type,

but not mutant BRCA1, was able to inhibit estradiol-induced activation of extracellular signal-regulated kinase as well as the synthesis of cyclins D1 and B1, the activity of cyclin-dependent kinases Cdk4 and CDK1, and G_1/S and G_2/M cell cycle progression ***(57)***.

Differences in mutation frequencies of cancer genes, such as those described here, highlight the important roles for these genes and associated pathways in determining the gene expression patterns of the various tumors. Such data allow us to unveil previously unknown genes that may be involved in tumorigenesis. Similar variation in expression of a set of genes across a set of samples indicates similar means of regulation and function; hence, it provides a powerful way of identifying novel biologically important genes that could be used as markers and targets for therapy. One such example is ras-related and estrogen-regulated growth inhibitor (*RERG*) that was identified by its coexpression with the ER ***(58)***. *RERG* overexpression inhibits growth of tumor cells, and its high expression is correlated with a favorable prognosis for patients.

5.1.2. Effects of TP53 Mutations on Global Gene Expression

The profound differences (clinical and molecular) among the tumor subtypes most likely reflect different alterations in molecular pathways within the tumor cells. TP53 plays an important role in directing cellular responses to genotoxic damage and regulates the activation of downstream genes that are involved in apoptosis, cell cycle arrest, and DNA repair ***(59–61)***. Previous studies have shown that mutations in the *TP53* gene predict poor prognosis and are associated with poor response to systemic therapy ***(62–64)***. Even though *TP53* itself is not differentially expressed at a detectable level across tumors, it is likely that TP53 plays a significant role in shaping the gene expression patterns in the various tumor subtypes. The coding region of the *TP53* gene (exons 2–11) was screened for mutations in all but eight tumor samples (not including benign tumors or normal breast tissue samples) by using temporal temperature gel electrophoresis ***(65)***. The frequencies of *TP53* mutations among the different subclasses was significantly different ($p < 0.001$). Luminal subtype A contained only 16% mutated tumors, whereas the luminal B, ERBB2+, and basal-like subclasses had 71, 86, and 75% *TP53*-mutated tumors, respectively **(Fig. 5)**. The finding of *TP53* mutations in tumors that simultaneously expressed the *ERBB2* gene at high levels supports previous observations of an interdependent role for *TP53* and *ERBB2* ***(20,66)***.

To more directly investigate the effect of *TP53* mutations on the expression patterns in these tumors, we searched for genes whose expression was consistently different between *TP53*-mutated and *TP53*-wild type tumors (two-class significance analysis of microarrays [SAM]); ***[11]***). A list of 158 genes significantly correlated with *TP53* status was selected that showed a median false discovery

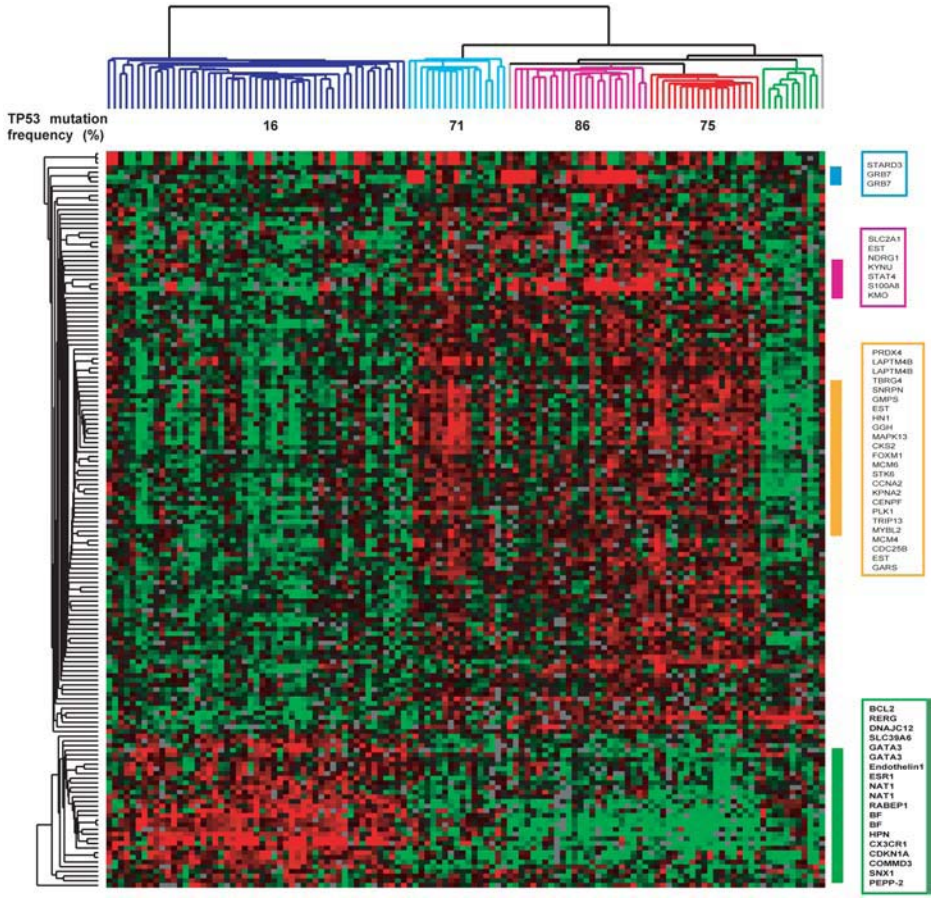


Fig. 5. Cluster of genes associated with *TP53* mutational status. Hierarchical clustering of 158 genes whose expression levels were significantly correlated to the mutation status of the *TP53* gene (SAM, false discovery rate < 1). Only tumor tissues from the two prospective studies from Norway were included in the cluster analysis. The tumor samples are ordered according to the clustering groups from **Fig. 3**. Frequencies of TP53 mutations are listed. Adapted from **ref.** ***67***.

rate of less than 1% ***(67)***. The genes were clustered, whereas the samples were ordered according to the subtypes **(Fig. 5)**. As expected, many of the genes that were highly expressed in the mutated tumors (mostly basal-like and luminal B type tumors) are cell cycle regulated genes such as *PLK1*, *CCNA2*, *STK6*, and *MAPK13* ***(17)***. Two of the genes coexpressed with *ERBB2* oncogene, *GRB7* and *STARD3*, also were found in this gene list, but not the *ERBB2* gene itself. Interestingly, many of the genes whose high expression and specific pattern

distinguished luminal subtype B from subtype A, such as *GGH*, *LAPTMB4*, and *MYBL2*, were highly expressed in *TP53*-mutated tumors. Among the genes highly expressed in tumors with a wild-type *TP53* gene, most of the genes in the luminal/ER+ cluster (*ER*, *GATA*-3, *RERG*, and so on) were found, which recapitulates the low mutation rate of the luminal subtype A. Further studies are needed to determine which of the genes are direct targets of *TP53* and which are present only by association with a particular expression phenotype.

6. Genome-Wide Copy Number Changes

The power of microarrays also is illustrated in their broad range of utility; the same types of DNA microarrays can be used to investigate both the structural and the expressed genome ***(68)***. Variations in gene expression measured by analyzing the tumor RNA may be caused by underlying genomic DNA copy number alterations, which are key genetic events in the development and progression of human cancers. A genome-wide aCGH analysis of a subset of the tumors described here was carried out using similar cDNA microarrays as for the gene expression studies. In total, a profile was generated across 6691 mapped genes in 45 primary breast tumors and 10 breast cancer cell lines ***(69)***. Numerous DNA copy number alterations were detected in the tumors despite the presence of euploid nontumor cell types; the magnitudes of the observed changes were generally lower in the tumor samples than the cell lines. DNA copy number alterations were identified in all cancer cell lines and nearly all (40/45) tumors, and on every human chromosome in at least one sample. Recurrent regions of DNA copy number gain and loss were readily identifiable; gains within 1q, 8q 17q, and 20q were observed in all breast cancer cell lines, and a large proportion of tumor samples (75, 48, 50, and 34%, respectively), consistent with published cytogenetic studies ***(70–72)***. The parallel measurements of mRNA levels and copy number alterations revealed a global impact of widespread DNA copy number alteration on gene expression in tumor cells. The overall patterns of gene amplification and elevated gene expression are concordant, i.e., a significant fraction of highly amplified genes seem to be correspondingly highly expressed ***(69,73)***. The concordance between high-level amplification and increased gene expression is best illustrated for genes on chromosome 17 **(Fig. 6)**, where the amplicon containing the oncogene *ERBB2* is located. A comprehensive analysis of nearly 150 breast tumors by using aCGH is underway with a specific emphasis on the distribution of genomic alterations across the subtypes. Preliminary data indicate specific genomic aberrations to be present in the different tumor subtypes, further supporting our results that these molecular subtypes of breast cancer represent biologically distinct disease entities.

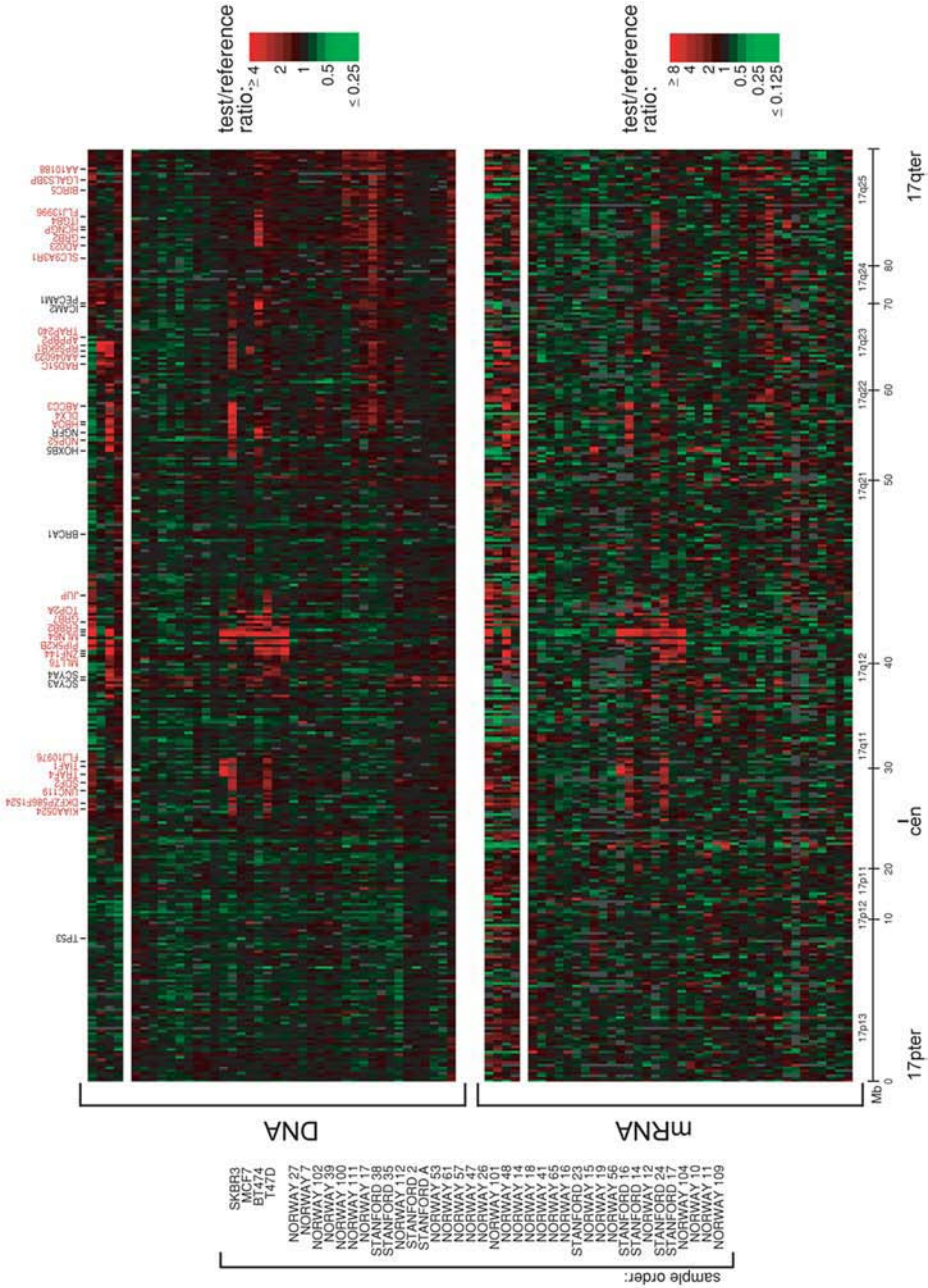


Fig. 6. Concordance between DNA copy number and gene expression across chromosome 17. DNA copy number alteration (top) and mRNA levels (bottom) are illustrated for breast cancer cell lines and tumors. Breast cancer cell lines and tumors are separately ordered by hierarchical clustering in the top panel, and the identical sample order is

7. Identification of Novel Markers

Gene signatures identified by whole-genome profiling that are significantly linked to clinically important parameters are valuable sources for discovering novel prognostic factors and targets for therapy. Many such markers have been suggested for breast cancer in the literature, in particular for predicting survival ***(23,26,30,33,34,74)***, with varying degree of success. A striking discovery was the lack of overlap of the predictive genes in most of these studies. This lack of overlap could be because of different microarray technologies, patient populations, study design, and statistical methods, or it could reflect the redundancy of valid correlations in such data. Recently, this idea was investigated in a study of one single breast cancer data set and by one statistical method ***(75)***, where the researchers showed that the resulting gene set was not unique and was strongly influenced by the subset of patients selected for the analysis. Many equally prognostic gene sets could be produced. This illustrates the consequence of molecular heterogeneity even within tumor groups that share the same histomorphology.

Recently, we have identified a subset of 54 marker genes that specifically capture the luminal A and the basal-like subtypes ***(13a)***, which represent two clinically very different groups of patients. These markers define a set of highly validated and promising potential prognostic molecular markers for breast cancer, and further confirmation in larger patient cohorts will be conducted.

8. Conclusion and Future Perspectives

Cancer genomics, as illustrated here by DNA microarray techniques, has challenged the traditionally used tumor classification schemes by identifying genetic alterations that can be used to group tumors into novel and clinically useful categories. In addition to leading to a improved cancer taxonomy, comprehensive gene profiling of tumors promises to enhance our understanding of the underlying biological processes and their biological effects. It also offers the greatest challenge to how clinical trials should be conducted. Designing the appropriate clinical trials in combination with genomics that use optimal laboratory methods and advanced statistics that incorporate biological knowledge may offer the opportunity to explore and understand the mechanisms controlling tumor behavior in vivo. In time, this approach will likely lead to improved prognostication and aid in therapy selection.

Fig. 6. *(Coninued)* maintained in the bottom panel. The 354 genes present on the microarrays and mapping to chromosome 17, and for which both DNA copy number and mRNA levels were determined, are ordered by position along the chromosome; selected genes are indicated in color-coded text. Fluorescence ratios (test/reference) are depicted by separate $\log_2$ pseudocolor scales (indicated). Adapted from **ref. *69***.

Acknowledgments

All colleagues who have contributed to these projects over the years are acknowledged.

References

1. Lønning, P. E., Sørlie, T., and Børresen-Dale, A. -L. (2005) Genomics in breast cancer—therapeutic implications. *Nat. Clin. Prac. Oncol.* **2,** 26–33.
2. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray [see comments]. *Science* **270,** 467–470.
3. Novoradovskaya, N., Whitfield, M. L., Basehore, L. S., et al. (2004) Universal Reference RNA as a standard for microarray experiments. *BMC Genomics* **5,** 20.
4. Dudoit, S., Gentleman, R. C., and Quackenbush, J. (2003) Open source software for the analysis of microarray data. *Biotechniques* **Suppl,** 45–51.
5. Quackenbush, J. (2001) Computational analysis of microarray data. *Nat. Rev. Genet.* **2,** 418–427.
6. Rhodes, D. R., and Chinnaiyan, A. M. (2005) Integrative analysis of the cancer transcriptome. *Nat. Genet.* **37 Suppl,** S31–S37.
7. Eisen, M. B., Spellman, P. T., Brown, P. O., and Botstein, D. (1998) Cluster analysis, and display of genome-wide expression patterns. *Proc. Natl. Acad. Sci. USA* **95,** 14,863–14,868.
8. Brown, M. P., Grundy, W. N., Lin, D., et al. (2000) Knowledge-based analysis of microarray gene expression data by using support vector machines. *Proc. Natl. Acad. Sci. USA* **97,** 262–267.
9. Khan, J., Wei, J. S., Ringner, M., et al. (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat. Med.* **7,** 673–679.
10. Tibshirani, R., Hastie, T., Narasimhan, B., and Chu, G. (2002) Diagnosis of multiple cancer types by shrunken centroids of gene expression. *Proc. Natl. Acad. Sci. USA* **99,** 6567–6572.
11. Tusher, V. G., Tibshirani, R., and Chu, G. (2001) Significance analysis of microarrays applied to the ionizing radiation response. *Proc. Natl. Acad. Sci. USA* **98,** 5116–5121.
12. Ball, C. A., Sherlock, G., Parkinson, H., et al. (2002) Standards for microarray data. *Science* **298,** 539.
13. Yauk, C. L., Berndt, M. L., Williams, A., and Douglas, G. R. (2004) Comprehensive comparison of six microarray technologies. *Nucleic Acids Res.* **32,** e124.

13a. Sørlie, T., Wang, Y., Xiao, C., et al. (2006) Distinct molecular mechanisms underlying clinically relevant subtypes of breast cancer: gene expression analyses across three different platforms. *BMC Genomics* **7,** 127.

14. Perou, C. M., Sorlie, T., Eisen, M. B., et al. (2000) Molecular portraits of human breast tumours. *Nature* **406,** 747–752.

15. Perou, C. M., Jeffrey, S. S., van de Rijn, M., et al. (1999) Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc. Natl. Acad. Sci. USA* **96,** 9212–9217.
16. Ross, D. T., Scherf, U., Eisen, M. B., et al. (2000) Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24,** 227–235.
17. Whitfield, M. L., Sherlock, G., Saldanha, A. J., et al. (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* **13,** 1977–2000.
18. Aas, T., Borresen, A. L., Geisler, S., et al. (1996) Specific P53 mutations are associated with de novo resistance to doxorubicin in breast cancer patients. *Nat. Med.* **2,** 811–814.
19. Geisler, SI., Lonning, P. E., Aas, T., et al. (2001) Influence of TP53 gene alterations and c-erbB-2 expression on the response to treatment with doxorubicin in locally advanced breast cancer. *Cancer Res.* **61,** 2505–2512.
20. Geisler, S., Borresen-Dale, A. L., Johnsen, H., et al. (2003) TP53 Gene Mutations Predict the Response to Neoadjuvant Treatment with 5-Fluorouracil and Mitomycin in Locally Advanced Breast Cancer. *Clin. Cancer Res.* **9,** 5582–5588.
21. Weigelt, B., Glas, A. M., Wessels, L. F., Witteveen, A. T., Peterse, J. L., and van't Veer, L. J. (2003) Gene expression profiles of primary breast tumors maintained in distant metastases. *Proc. Natl. Acad. Sci. USA* **100,** 15, 901–15, 905.
22. Sorlie, T., Tibshirani, R., Parker, J., et al. (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc. Natl. Acad. Sci. USA* **100,** 8418–8423.
23. Sorlie, T., Perou, C. M., Tibshirani, R., et al. (2001) Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc. Natl. Acad. Sci. USA* **98,** 10, 869–10, 874.
24. Bertucci, F., Finetti, P., Rougemont, J., et al. (2005) Gene expression profiling identifies molecular subtypes of inflammatory breast cancer. *Cancer Res.* **65,** 2170–2178.
25. Chang, H. Y., Nuyten, D. S., Sneddon, J. B., et al. (2005) Robustness, scalability, and integration of a wound-response gene expression signature in predicting breast cancer survival. *Proc. Natl. Acad. Sci. USA* **102,** 3738–3743.
26. Sotiriou, C., Neo, S. Y., McShane, L. M., et al. (2003) Breast cancer classification and prognosis based on gene expression profiles from a population-based study. *Proc. Natl. Acad. Sci. USA* **100,** 10,393–10,398.
27. Wang, Z. C., Lin, M., Wei, L. J., et al. (2004) Loss of heterozygosity and its correlation with expression profiles in subclasses of invasive breast cancers. *Cancer Res.* **64,** 64–71.
28. Yu, K., Lee, C. H., Tan, P. H., and Tan, P. (2004) Conservation of breast cancer molecular subtypes and transcriptional patterns of tumor progression across distinct ethnic populations. *Clin. Cancer Res.* **10,** 5508–5517.
29. Zhao, H., Langerod, A., Ji, Y., et al. (2004) Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. *Mol. Biol. Cell* **15,** 2523–2536.
30. van't Veer, L. J., Dai, H., van de Vijver, M. J., et al. (2002) Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415,** 530–536.

31. Slamon, D. J., Clark, G. M., Wong, S. G., Levin, W. J., Ullrich, A., and McGuire, W. L. (1987) Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science* **235,** 177–182.
32. Ahr, A., Karn, T., Solbach, C., et al. (2002) Identification of high risk breast-cancer patients by gene expression profiling. *Lancet* **359,** 131–132.
33. Huang, E., Cheng, S. H., Dressman, H., et al. (2003) Gene expression predictors of breast cancer outcomes. *Lancet* **361,** 1590–1596.
34. Ma, X. J., Wang, Z., Ryan, P. D., et al. (2004) A two-gene expression ratio predicts clinical outcome in breast cancer patients treated with tamoxifen. *Cancer Cell* **5,** 607–616.
35. van de Vijver, M. J., He, Y. D., van't Veer, L. J., et al. (2002) A gene-expression signature as a predictor of survival in breast cancer. *N. Engl. J. Med.* **347,** 1999–2009.
36. Onda, M., Emi, M., Nagai, H., et al. (2004) Gene expression patterns as marker for 5–year postoperative prognosis of primary breast cancers. *J. Cancer Res. Clin. Oncol.* **130,** 537–545.
37. Nagahata, T., Onda, M., Emi, M., et al. (2004) Expression profiling to predict postoperative prognosis for estrogen receptor-negative breast cancers by analysis of 25, 344 genes on a cDNA microarray. *Cancer Sci.* **95,** 218–225.
38. Eden, P., Ritz, C., Rose, C., Ferno, M., and Peterson, C. (2004) "Good Old" clinical markers have similar power in breast cancer prognosis as microarray gene expression profilers. *Eur. J. Cancer* **40,** 1837–1841.
39. Sasco, A. J. (2003) Breast cancer and the environment. *Horm. Res.* **60 Suppl 3,** 50.
40. Early Breast Cancer Trialists' Collaborative Group (1998) Tamoxifen for early breast cancer: an overview of the randomised trials. Early Breast Cancer Trialists' Collaborative Group. *Lancet* **351,** 1451–1467.
41. Citron, M. L., Berry, D. A., Cirrincione, C., et al. (2003) Randomized trial of dose-dense versus conventionally scheduled and sequential versus concurrent combination chemotherapy as postoperative adjuvant treatment of node-positive primary breast cancer: first report of Intergroup Trial C9741/Cancer and Leukemia Group B Trial 9741. *J. Clin. Oncol.* **21,** 1431–1439.
42. Ayers, M., Symmans, W. F., Stec, J., et al. (2004) Gene Expression Profiles Predict Complete Pathologic Response to Neoadjuvant Paclitaxel and Fluorouracil, Doxorubicin, and Cyclophosphamide Chemotherapy in Breast Cancer. *J. Clin. Oncol.* **22,** 2284–2293.
43. Chang, J. C., Wooten, E. C., Tsimelzon, A., et al. (2003) Gene expression profiling for the prediction of therapeutic response to docetaxel in patients with breast cancer. *Lancet* **362,** 362–369.
44. Iwao-Koizumi, K., Matoba, R., Ueno, N., et al. (2005) Prediction of docetaxel response in human breast cancer by gene expression profiling. *J. Clin. Oncol.* **23,** 422–431.
45. Jansen, M. P., Foekens, J. A., van Staveren, I. L., et al. (2005) Molecular classification of tamoxifen-resistant breast carcinomas by gene expression profiling. *J. Clin. Oncol.* **23,** 732–740.

46. Kannan, K., Kaminski, N., Rechavi, G., et al. (2001) DNA microarray analysis of genes involved in p53 mediated apoptosis: activation of Apaf-1. *Oncogene* **20,** 3449–3455.
47. Yoon, H., Liyanarachchi, S., Wright, F. A., Davuluri, R., Lockman, J. C., de la, C. A., and Pellegata, N. S. (2002) Gene expression profiling of isogenic cells with different TP53 gene dosage reveals numerous genes that are affected by TP53 dosage and identifies CSPG2 as a direct target of p53. *Proc. Natl. Acad. Sci. USA* **99,** 15,632–15,637.
48. Schaner, M. E., Ross, D. T., Ciaravino, G., et al.. (2003) Gene expression patterns in ovarian carcinomas. *Mol. Biol. Cell* **14,** 4376–4386.
49. Alizadeh, A. A., Eisen, M. B., Davis, R. E., et al. (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling [see comments]. *Nature* **403,** 503–511.
50. Chen, X., Cheung, S. T., So, S., et al. (2002) Gene expression patterns in human liver cancers. *Mol. Biol. Cell* **13,** 1929–1939.
51. Lapointe, J., Li, C., Higgins, J. P., Van de, R. M., et al. (2004) Gene expression profiling identifies clinically relevant subtypes of prostate cancer. *Proc. Natl. Acad. Sci. USA* **101,** 811–816.
52. Ford, D., Easton, D. F., Stratton, M., et al. (1998) Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. The Breast Cancer Linkage Consortium. *Am. J. Hum. Genet.*, **62,** 676–689.
53. Hedenfalk, I., Duggan, D., Chen, Y., et al. (2001) Gene-expression profiles in hereditary breast cancer. *N. Engl. J. Med.* **344,** 539–548.
54. Armes, J. E., Trute, L., White, D., et al. (1999) Distinct molecular pathogeneses of early-onset breast cancers in BRCA1 and BRCA2 mutation carriers: a population-based study. *Cancer Res.* **59,** 2011–2017.
55. Lakhani, S. R., Reis-Filho, J. S., Fulford, L., et al. (2005) Prediction of BRCA1 status in patients with breast cancer using estrogen receptor and basal phenotype. *Clin. Cancer Res.* **11,** 5175–5180.
56. Fan, S., Wang, J., Yuan, R., Ma, Y., et al. (1999) BRCA1 inhibition of estrogen receptor signaling in transfected cells. *Science*, **284,** 1354–1356.
57. Razandi, M., Pedram, A., Rosen, E. M., and Levin, E. R. (2004) BRCA1 inhibits membrane estrogen and growth factor receptor signaling to cell proliferation in breast cancer. *Mol. Cell Biol.* **24,** 5900–5913.
58. Finlin, B. S., Gau, C. L., Murphy, G. A., et al. (2001) RERG is a novel ras-related, estrogen-regulated and growth-inhibitory gene in breast cancer. *J. Biol. Chem.* **276,** 42,259–42,267.
59. Balint, E. E. and Vousden, K. H. (2001) Activation and activities of the p53 tumour suppressor protein. *Br. J. Cancer* **85,** 1813–1823.
60. Guimaraes, D. P. and Hainaut, P. (2002) TP53: a key gene in human cancer. *Biochimie*, **84,** 83–93.
61. Vogelstein, B., Lane, D., and Levine, A. J. (2000) Surfing the p53 network. *Nature* **408,** 307–310.

62. Bergh, J., Norberg, T., Sjogren, S., Lindgren, A., and Holmberg, L. (1995) Complete sequencing of the p53 gene provides prognostic information in breast cancer patients, particularly in relation to adjuvant systemic therapy and radiotherapy. *Nat. Med.* **1,** 1029–1034.
63. Berns, E. M., Foekens, J. A., Vossen, R., et al. (2000) Complete sequencing of TP53 predicts poor response to systemic therapy of advanced breast cancer. *Cancer Res.* **60,** 2155–2162.
64. Borresen, A. L., Andersen, T. I., Eyfjord, J. E., et al. (1995) TP53 mutations and breast cancer prognosis: particularly poor survival rates for cases with mutations in the zinc-binding domains. *Genes Chromosomes Cancer* **14,** 71–75.
65. Sorlie, T., Johnsen, H., Vu, P., Lind, G. E., Lothe, R., and Borresen-Dale, A. L. (2004) Mutation Screening of the TP53 Gene by Temporal Temperature Gradient Gel Electrophoresis. *Methods Mol. Biol.* **291,** 207–216.
66. Nakopoulou, L. L., Alexiadou, A., Theodoropoulos, G. E., Lazaris, A. C., Tzonou, A., and Keramopoulos, A. (1996) Prognostic significance of the co-expression of p53 and c-erbB-2 proteins in breast cancer. *J. Pathol.* **179,** 31–38.
67. Sorlie, T. (2004) Molecular portraits of breast cancer: tumour subtypes as distinct disease entities. *Eur. J. Cancer* **40,** 2667–2675.
68. Pollack, J. R., Perou, C. M., Alizadeh, A. A., et al. (1999) Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat. Genet.* **23,** 41–46.
69. Pollack, J. R., Sorlie, T., Perou, C. M., et al. (2002) Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc. Natl. Acad. Sci. USA* **99,** 12, 963–12, 968.
70. Forozan, F., Mahlamaki, E. H., Monni, O., et al. (2000) Comparative genomic hybridization analysis of 38 breast cancer cell lines: a basis for interpreting complementary DNA microarray data. *Cancer Res.* **60,** 4519–4525.
71. Kallioniemi, A., Kallioniemi, O. P., Piper, J., et al. (1994) Detection and mapping of amplified DNA sequences in breast cancer by comparative genomic hybridization. *Proc. Natl. Acad. Sci. USA* **91,** 2156–2160.
72. Tirkkonen, M., Tanner, M., Karhu, R., Kallioniemi, A., Isola, J., and Kallioniemi, O. P. (1998) Molecular cytogenetics of primary breast cancer by CGH. *Genes Chromosomes Cancer* **21,** 177–184.
73. Hyman, E., Kauraniemi, P., Hautaniemi, et al. (2002) Impact of DNA amplification on gene expression patterns in breast cancer. *Cancer Res.* **62,** 6240–6245.
74. Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T., R. (2003) Amolecular signature of metastasis in primary solid tumors. *Nat. Genet.* **33,** 49–54.
75. Ein-Dor, L., Kela, I., Getz, G., Givol, D., and Domany, E. (2005) Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* **21,** 171–178.

6

Discovery of Differentially Expressed Genes

Technical Considerations

Øystein Røsok and Mouldy Sioud

Summary

Identification and characterization of differentially expressed genes may be an important first step toward the understanding of both normal physiology and disease. A multitude of techniques belonging to two main categories have been developed to identify the differences in gene expression between samples from different biological origin: selection techniques and global techniques. Whereas the selection techniques strive to identify specific differentially expressed genes, the global techniques analyze the total transcriptome or a major part of the RNA population in a defined biological material. By exploiting the known sequences of the adaptors used in suppressive subtraction hybridization technique, a strategy named novel rescue–suppression-subtractive hybridization was developed. It should facilitate the discovery of differentially expressed genes.

Key Words: Differential display; microarray; representational difference analysis (RDA); serial analysis of gene expression (SAGE); target discovery; target identification.

1. Introduction

There are two eras of selection techniques: before and after the development of PCR ***(1)***. Traditionally, selection was achieved by subtractive hybridization and differential screening of cDNA libraries. However, the introduction of PCR made it possible to design several new techniques that sped up gene discovery. The new techniques also allowed for analysis of small amounts of biological material ***(2,3)***.

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols*
Volume I, Emerging Strategies for Targets and Biomarker Discovery
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

1.1. Selection Techniques

1.1.1. Differential Screening

In differential hybridization, a cDNA library is constructed from one cell population, and then duplicate membranes of colony/plaque lifts are hybridized separately with probes made from the same cell population and from a cell population to which it is supposed to be compared. Colonies/plaques hybridizing only to the probe from the library-derived population have expression restricted to this cell type. Because of high-complexity cDNA probes, the method works well only for the small subclass of genes that are expressed in abundance in the cell type from which the library is made ***(4)***. Other disadvantages of this method are the requirement for large quantities of RNA and the ability to identify interesting clones from only one population.

1.1.2. Subtractive Hybridization (Subtractive Cloning)

The basic idea of subtraction cloning is to remove nucleic acids representing common genes expressed in two different sources of biological material, leaving behind for analysis nucleic acids representing genes that are uniquely or abundantly expressed in the biological material of interest.

Nucleic acids (first strand cDNA or mRNA) from the material of interest (the tester or tester population) are hybridized to complementary nucleic acids (mRNA or first strand cDNA) from the biological material to which it is supposed to be compared (the driver population). The driver population is present in excess to optimize the probability and speed of the reannealing process of complementary tester and driver duplexes. Only sequences common to the two populations can form hybrids. After the hybridization, the duplexes and single-stranded driver sequences are removed, and the remaining nucleic acid population is enriched for tester-specific nucleic acids. The process usually must be performed repeatedly to remove all common sequences. Remaining nucleic acid can either be used to prepare a cDNA library enriched for tester-specific clones or to generate an absorbed probe that can be used to screen a cDNA library for tester-specific clones. cDNA clones corresponding to rare mRNAs (0.005–0.01% of total mRNA) may thereby be detected ***(4)***. In the original subtraction techniques, first strand cDNA was used as the tester and mRNA as the driver ***(5)***. RNA generally makes a poor tester because it is easily degraded. Conversely, it makes a good driver because driver molecules not removed during the hybrid removal step can easily be degraded enzymatically or by using alkali treatment. A problem with mRNA as driver, is that large amounts of starting material are needed, and only two rounds of subtraction can be performed before the amounts of remaining tester becomes too small. Thus, the usefulness of absorbed probes may be limited by the amount of material that is available.

In the first protocols describing subtraction cloning, double-stranded (ds) driver–tester was removed by hydroxyapatite (HAP) column chromatography, and single-stranded driver mRNA was removed by RNase or alkali degradation ***(5–7)***. HAP binds double-stranded molecules more tightly than single-stranded molecules, enabling separation of driver–tester and driver–driver duplexes from unhybridized tester. Because this method is cumbersome, alternative techniques for removing driver and tester duplexes have been developed, and they involve both methods for physical removal, enzymatic degradation, immobilization, and chemical cross-linking (reviewed in **ref. *8***).

1.1.3. PCR-Based Subtraction Cloning

Using polyA+ mRNA as the driver is possible only when one is able to obtain large amounts of starting material and when the biological sources to be compared are closely related. When material is limited, or the tissue is complex, amplification must be performed before subtraction. Multiple rounds of subtraction also may require amplification of the resulting population of tester. A drawback is that the representation of individual transcripts may be altered during multiple rounds of amplification. The transcript from both driver and tester may be amplified from cDNA libraries, by in vitro transcription, or by PCR.

Single-stranded driver is more efficient than double-stranded driver because of competition between driver–driver duplex formation and the desired driver–tester duplex formation. This competition necessitates more rounds of subtraction to ensure optimal subtraction. Amplified single-stranded driver may be obtained from libraries as single-stranded phagemids ***(9)***, by in vitro transcription with T3 or T7 RNA polymerase ***(10)***, or by asymmetric PCR on single-stranded template ***(11)***. However, subtraction performed using double-stranded driver also can be effective ***(12)***.

1.1.4. Positive Selection

In positive selection, nucleic acids representing genes uniquely or predominantly expressed in one source of biological material compared with another are selected and isolated. Undesired nucleic acids representing genes expressed in both sources are left behind.

For subtraction cloning, nucleic acids from the interesting material (tester) are hybridized with nucleic acids from the compared material (driver). In contrast to subtraction cloning, double-stranded nucleic acids are used both for driver and tester. The principle is to isolate tester–tester duplexes, representing genes uniquely expressed in the material of interest. Tester–driver and driver–driver duplexes as well as unhybridized driver and tester are left behind. No physical separation of single-stranded and double-stranded nucleic acids is included.

Compared with subtraction cloning, positive selection can be more efficient for isolating differentially expressed transcripts because subtractive hybridization rarely goes to completion. Therefore, a background of common unsubtracted clones may persist. In contrast, because tester concentration is low, the desired tester–tester duplexes take longer to form than driver–driver duplexes. Thus, rare clones that have not reannealed are likely to be missed in positive selection. A solution to this problem is to use reagents, which decreases the aqueous volume, and thus increases the hybridization rate. Polyethylene glycol and phenol may serve this purpose ***(13)***. Using a normalized tester population also makes positive selection more effective. The positive selection may be done without including a PCR step. More recent methods include PCR, which is done in representational difference analysis (RDA) and in subtractive suppression hybridization (SSH).

1.1.5. Representational Difference Analysis (RDA)

In RDA, target cDNA fragments from one of two cell populations are sequentially enriched by favorable hybridization kinetics and subsequently amplified by PCR, whereas material common to both populations is eliminated by selective degradation ***(14)***.

RDA, originally developed as a method for isolation of differences between two complex genomes ***(15)***, has further been developed to cDNA RDA. This technique is a PCR-based subtractive hybridization technique for identifying differentially expressed genes. The mRNA from both populations is isolated and converted into dscDNA. By cutting this cDNA with a four-base pair (bp) restriction enzyme, a large proportion of the transcripts will be represented as amplifiable fragments (150 bp–1.2 kilobases [kb]) ***(16)***. Amplified fragments, known as "representations," are hybridized at high driver:tester ratio. By using a combination of linker ligation, restriction cutting, and PCR, exponential amplification of only tester–tester hybrids is achieved. Single-stranded DNA is enzymatically degraded.

To amplify only true differences, a second and often a third round of subtraction is performed, in which the ratio of driver to tester is increased. Difference products are visible in agarose gels and can easily be subcloned. Transcripts expressed in less than one copy per cell can be detected with low rate of false positives ***(14)***. In circumstances where many differences are expected, RDA enriches the products that amplify most effectively, and not necessarily all the interesting differences. Thus, RDA is not the method of choice when investigating tissues with abundant differences ***(14)***. Because no normalization step is included, the method is not efficient for the detection of rare transcripts.

1.1.6. Suppression Subtractive Hybridization (SSH)

In SSH, normalization and subtraction are combined in a single step to equalize the abundance of cDNA within the target population and to exclude the

sequences common to both populations ***(17)***. No physical separation or enzymatic degradation of undesired sequences is performed.

As in RDA, dscDNAs representing transcripts from the driver and the tester populations are digested with a four-base pair restriction enzyme. The tester cDNA is then divided into two portions, each ligated to separate sets of adaptors. Subtractive hybridization is then performed in two steps. First, an excess of driver is hybridized to each sample of tester, generating dscDNA hybrid molecules common between driver and tester molecules as well as dscDNA between single-stranded (ss) cDNA present only in tester or in excess in the driver population. In this hybridization, the sscDNA tester fraction is normalized; meaning the concentration of high- and low-abundance sscDNAs become roughly equalized. This equalization is because the reannealing process is much faster for the more abundant molecules. Furthermore, ssDNAs representing differentially expressed mRNAs in the tester fraction are significantly enriched because common cDNAs form hybrids with the driver. In the second hybridization step, the two separate hybridizations, containing tester with different adaptors, are mixed without denaturating, and freshly denaturated driver is added. This process increases the fraction of common tester-generated cDNAs hybridized to driver-generated cDNA molecules, and enriches the fraction of tester-unique dscDNA molecules flanked by different adaptors. The ends of the dscDNA molecules are then filled by DNA polymerase, producing different annealing sites on the differently expressed tester dscDNA molecules for primers in the following PCR step. Because of the suppression PCR effect, only dscDNA molecules with two different adaptors can be exponentially expressed. dscDNA flanked by the same primer anneal to themselves, forming pan-like structures that prevents exponential amplification ***(18)***. In this way, cDNAs for differentially expressed genes of both high and low abundance are highly enriched. The SSH procedure also can be modified to identify transcripts moderately (e.g., two- to fourfold) regulated between the tester and driver populations ***(19)***.

The main advantage with SSH is the ability to enrich both for differentially expressed transcripts with high and low expression levels. A drawback is the relatively high amounts of starting material needed; 2–4 µg of polyA+ RNA is normally suggested. Alternatively an amplification step can be incorporated, or total RNA may be used as starting material ***(20)***. Problems associated with several subtraction strategies are relevant for SSH, such as lack of restriction sites in some transcripts, and generation of false positives. To reduce the numbers of false positives, we have developed a strategy that includes the confirmation step in the selection protocol ***(21)***. We present a protocol for this strategy in this chapter, and name it “rescue suppression subtractive hybridization.”

1.1.7. Differential Display (DD)

In message display techniques, PCR-amplified fragments representing the transcripts are visualized on a gel, revealing the differences in gene expression between populations of two or more sources. No subtraction is included.

Differential display was developed as a quicker alternative to subtractive hybridization techniques to obtain cDNAs representing differentially expressed transcripts ***(22)***. The basic principle of the technique is to isolate total RNA or mRNA from the cell types to be compared and then to reverse transcribe the RNA by using three different "anchor primers" containing a poly(dT) region that targets the polyA region of eukaryotic mRNA. The primers differ from each other in the last 3′ non-T base, fixing the priming site at the beginning at the polyA tail of three different RNA populations and limiting the number of RNA species transcribed in each reaction. PCR amplification is performed using the anchor primers in combination with a short arbitrary primer. The primers originally used for DD were 14-mer anchor primers with two- ***(22)*** and then one-base–anchored primers ***(23)*** to ensure specific priming in the 3′ end. Since then, the selection of primers used for DD has evolved ***(24)***. A related method known as RNA arbitrarily primed PCR (RAP-PCR) that uses only primers of random sequence was developed in 1992 ***(25)***. The objective of both strategies is to amplify fragments of sufficient length to allow unique identification of the mRNA species they represent. As a rule, fragments from 150 to 500 bp can be resolved by size on a DNA sequencing gel ***(22)***. PCR primers and reaction conditions must be chosen to such that a sufficiently small number of fragments, typically 50–100 fragments are amplified. This amplification can be achieved by initially running PCR at low stringency to generate mismatching at a controlled frequency ***(24)***. To obtain a comprehensive representation of the expressed genes of a given cell type, many combinations of anchor and arbitrary primer must be used. The amplified cDNAs from the different cell types are separated by size in adjacent lanes of sequencing gels for comparison. To identify putative differentially expressed genes, the amplified fragments must be visualized with either isotopes or fluorescent dyes. Putative differentially expressed fragments can be excised and eluted from the dried gel and then reamplified.

Different strategies for analyzing the candidate fragments have been reported. All strategies involve cloning, sequencing, identification, and confirmation of differential expression; however, these steps occur in a variety of orders. Some approaches streamline the process by incorporating direct sequencing of DD fragments ***(26)***. However, this strategy is limited to fragments generated by different primers on each end. Database matching ultimately identifies readable sequences.

There is no doubt that a band pattern showing potentially differentially expressed transcripts may be obtained in a short time. However, the most time-consuming and problematic step in the overall DD technique is the confirmation that the isolated cDNA clone represents a differentially expressed gene. This problem has led to a multitude of publications reporting improvements and simplification of the basic protocol to reduce the high frequency of false positives and to confirm the differential expression pattern of selected candidates ***(27–30)***. A high rate of false positives and time-consuming effort to find the true positives have been the main drawbacks with this method ***(31,32)***. In addition, rare transcripts are difficult to detect.

The advances are the experimental simplicity by which candidates are obtained, the ability to use as little as a few nanograms of starting material ***(33)***, and the possibility to identify novel transcripts.

1.2. Global Techniques

Global cDNA techniques analyze the total or large parts of the mRNA populations in a single assay. The aim may be to quantitatively estimate the number of expressed genes in a defined biological material, estimate their absolute or relative abundance levels, survey global changes in gene expression patterns in response to extrinsic and intrinsic factors, identify specific genes whose expression is altered because of such factors, or simply to identify as many as possible of the genes expressed in a defined biological material. This process may or may not include differentially expressed, novel as well as known genes. Because of the unstable nature of RNA, the mRNA populations are usually reverse transcribed into DNA complementary to the RNA strand (cDNA), which in theory has the same distribution as the original mRNA populations. It is the populations of cDNA that normally are the subjects of investigation.

1.2.1. Large-Scale Sequencing of Expressed Sequence Tag (EST) Sequences

ESTs are segments of sequences from cDNA clones that correspond to mRNA. The tags can thus be used to identify transcripts in a specific tissue. Adams and co-workers coined the term EST in 1991, when the first high-throughput sequencing of a cDNA library was reported ***(34)***. The clones are randomly picked and either randomly single-pass sequenced, or more commonly, sequenced from the 5′ or 3′ ends or both, to generate on average a 200- to 700-bp sequence. Even much smaller segments are sufficient for rapid identification of expressed genes by sequence analysis. The initial objective of EST sequencing was to speed up the process of gene identification and possible splicing alternative events ***(35)***. In addition, large-scale sequencing of EST sequences has proven to be

a powerful approach to discover novel genes and for transcript profiling ***(36)***. By comparing the transcription profiles from cDNA libraries representing different tissues, differently expressed genes may be identified. However, because several tissues are represented by low numbers of clones (<10.000), differences in transcript abundances between EST data sets can serve only as indications of differential expression. Picking a large set of clones from all major organs, several developmental stages and disease states, and many individual cell types provides a large-scale view of differential gene expression that encompasses human anatomy and development ***(36)***. The sequences obtained from several large-scale EST sequencing projects are available in public databases, such as dbEST (http://www.ncbi.nlm.nih.gov./dbEST/index.html), which currently (2006) contains 2 million human ESTs.

As early as in 1976, Hastie and Bishop ***(37)*** found that mRNA of the prevalent and intermediate frequency classes comprised more than 50% of the mRNA mass in a typical cell, but not more than 10% of the different mRNA species. Later, this finding was supported by Zhang and co-workers ***(38)*** who found that 86% of all transcripts are present at less than five copies per cell. The low-abundance class of mRNA is thus poorly represented by using primary, unsubtracted libraries ***(39)***. Thus, the use of normalized libraries, in which the frequency of all clones is within a narrow range ***(40)***, has been shown to be beneficial for large-scale sequencing ***(41,42)***. To avoid redundant identification of already identified genes, subtracted cDNA libraries enriched for genes not yet identified have been constructed ***(39)***. By using suppression PCR, Diatchenko and co-workers ***(17)*** combined normalization and subtraction in a single procedure to make a subtracted library.

Compared with other global approaches, the advantages with EST sequencing is that novel transcripts may be discovered, specific identification of expressed transcripts may be obtained, and sequencing of large numbers of clones may give a good representation of the transcripts present in the investigated cell type. The major drawback is the laborious effort needed to sequence a large number of clones.

1.2.2. Serial Analysis of Gene Expression (SAGE)

SAGE allows effective quantitative and qualitative analysis of a large number of transcripts, including identification of novel expressed genes, thereby generating a gene expression profile for the investigated cell type (*see* Chapter 4, Volume 1). The technology is based on sequencing cloned concatenated tags, representing fragments of transcripts sufficiently long to be uniquely identified in sequence databases.

SAGE is another tag sequencing strategy, described in 1995 ***(43)***. In SAGE, short tags (~10 bases) are generated from the 3′ end of mRNA by restriction

digestion, ligation, and amplification of cDNA. Ligated concatemers of up to 50 tags are cloned into plasmid vectors and sequenced. Compared with conventional EST sequencing, SAGE is 30–50 times more effective per plasmid. The abundance of a particular tag relates directly to the expression level of the gene from which it is derived, resulting in a statistical representation of the transcripts within the investigated cell type. The technology enables high throughput and has the ability to compare data from numerous sources. Since the development of SAGE in 1995, it has been applied in a large number of studies involving different cell types and organisms (reviewed in **ref.** ***44***). The principle of SAGE is as follows: dscDNA synthesized using biotinylated oligo(dT) primers, is cleaved with a so-called "anchoring enzyme," a four-base restriction enzyme. The resulting 3′ ends are recovered on beads via biotin-streptavidin binding, by using biotinylated polyT primer. These cDNA fragments are then ligated to linkers containing the recognition site for a restriction enzyme that cuts approx 14 bp downstream of the recognition site. The small fragments, each with a 9- to 17-bp gene-specific tag are concatenated into long sequences and cloned into plasmids that are sequenced.

The greatest advantage with the methodology is the high throughput it allows. It also provides an absolute quantitation of gene expression, enabling the combination of results from different laboratories deposited in the public SAGEmap database (http://www.ncbi.nlm.nih.gov/SAGE) and the omnibus database (GEO) (http://www.ncbi.nlm.gov/projects/geo/). In addition, SAGE has the power to measure very rare transcripts and to discover novel transcript not detectable with EST sequencing (*see* Chapter 4, Volume 1).

The main concern with the method is the short length of the specific tags, ~10 bp. This length may be too short to be unique for a specific transcript; especially in cases were the tag recognizes conserved domains and gene families. Thus, multiple genes may share the same tag. In addition, these short tags are very sensitive to sequencing errors and single nucleotide polymorphisms ***(45)***. These effects may be compensated for by introducing alternative restriction enzymes, which generate tags with up to 17 bp ***(46)***. For a transcript to be included in the analysis, its gene must have a restriction site for the chosen restriction enzyme. More transcripts would be present in the tag library by generating libraries by using several enzymes ***(47)***. Whereas a limitation with the original SAGE was the need of large amounts of starting material (2.5–5 µg of mRNA), modified strategies such as MicroSAGE (requiring < 10 cells) ***(48)***, SAGE-lite (100 ng of total RNA) ***(49)***, and miniSAGE (1 µg of total RNA without PCR amplification ***(50)*** have been developed.

A more recently developed method based on a similar principle to SAGE is massively parallel signature sequencing ***(51)***. In this method, millions of transcripts are captured and identified by bead-based EST signature sequencing. The level of expression is analyzed by counting the number of mRNA molecules that

represent each gene. The individual mRNAs are identified through the generation of 17- to 20-base signature sequences.

1.2.3. DNA Microarray

DNA microarray methods provide simultaneously hybridization-based monitoring of relative expression levels of genes on a genomic scale, from two sources.

Important advances have led to the development of DNA microarray technology. First, large-scale sequencing projects have made it possible to assemble collections of DNA corresponding to large fractions of genes in many organisms. Second, development of chip-printing technology has made it possible to generate microscopic arrays of large sets of immobilized DNA sequences. Third, advances in fluorescent labeling and detection of nucleic acids have made the use of DNA microarray simpler and more accurate. The term "microarray" is normally used for two independently developed technologies: cDNA microarrays ***(52)*** and GeneChip arrays, developed by Affymetrix ***(53)***.

The basic principle of the techniques is that cDNA representing expressed genes from two different sources (normally test and control preparation), and labeled with two different fluorophores, is hybridized simultaneously in competition to gene-specific probes immobilized at predetermined positions at high density on two-dimensional solid microarrays. These solid supports, often called chips, may contain several thousand sequences, and they represent the total transcriptome in the investigated cell type. Capturing of the target molecules by the probes on the array through sequence complementation enables determination of the relative abundance in the two samples of mRNA corresponding to each arrayed gene. Detection of the hybridization signal is performed by a laser scanner that measures ratio of fluorescence intensities of the two flours emitted by each gene spot. The two most common platforms for gene expression analysis use *in situ* synthesis of oligonucleotide probes or robotic deposition of cDNA probes to the solid support. For discussions about aspects of the array production, sample labeling, hybridization, data analysis, and so on, we refer the readers to comprehensive reviews ***(54–57)***.

Advantages with microscale assays are reduction of reagent consumption; minimized reaction volumes, which increases sample concentration; and accelerated reaction kinetics. This assay enables analyzing expression of all the genes in the organism of interest in one reaction. Because each spot on the array represent a known sequence, the identity of a gene is known once a signal is detected. Microchips allow true parallelism and automation, which increases speed of biochemical research.

Problems concerning limiting amounts of mRNA may be overcome by using amplification methods that will not distort the representation of the various mRNAs ***(2,3)***. Monitoring of relative expression levels is restricted to genes

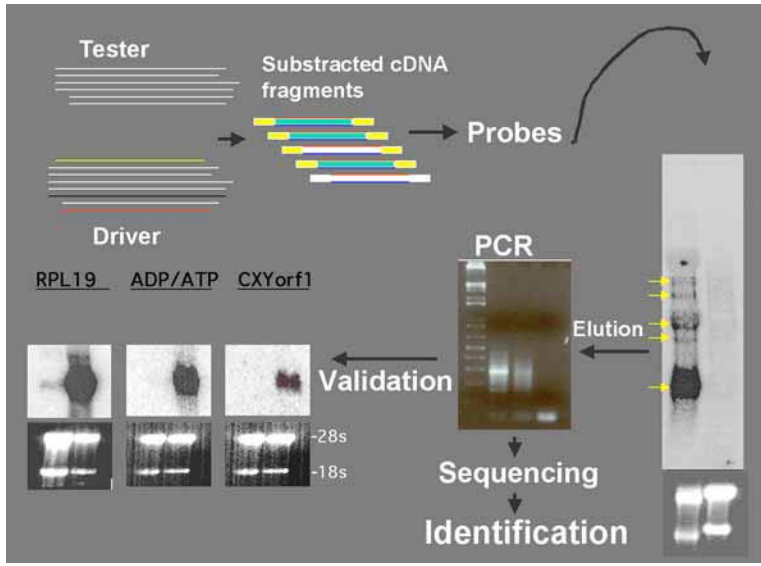


Fig. 1. Processing steps in rescue suppression subtractive hybridization (SSH). First, SSH was carried out using the PCR-selected cDNA subtraction kit (Clontech). PolyA+ RNA (2 µg) was reverse transcribed according to manufacturer's instructions and then digested with RsaI restriction enzyme. The adapters were ligated to the tester cDNAs and the ligation efficiency was investigated. Two rounds of hybridization with an excess of driver cDNA and one PCR amplification were performed with the driver cDNA. Subsequently, two rounds of PCR were performed to obtain differentially expressed genes from which a radioactive probe was prepared. Second, total RNA (20 µg/lane) from either the tester or the driver was electrophoresed on 1% agarose/formaldehyde gels, transferred to Hybond N+ membrane, and UV-cross-linked. Membranes were hybridized with ^{32}P-labelled probe representing the entire subtracted PCR fragments, which are expected to contain genes overexpressed in SKBR3 (tester, left), but not in HMEC cells (driver, right). Third, the nitrocellulose portion containing differential expressed genes were marked, cut out, and incubated in 100 µL of water for 10 min at 95°C. Eluted probe molecules (10 µL) were reamplified in a 50-µL PCR reaction by using the secondary PCR conditions as described in the PCR-selected cDNA subtraction kit. The PCR products were resolved on 2% agarose gels and then purified using the QIAGEN gel extraction kit as described by the manufacturer. Fourth, the PCR-reamplified probe molecules were cloned into the pGEM-T vector (Promega) according to the manufacturer's instructions, and positives clones were sequenced using either T7 or SP6 primers by automatic sequencing. The identity of the sequenced clones was identified by Blast searches. Fifth, individual gene probes were used in Northern blots to confirm their differential expression ***(21)***.

represented by sequences on the array, i.e. known genes. Significant transcriptional changes of excluded genes that may be of importance thus are missed. However, by arraying amplified open reading frames from genomic DNA, unknown genes can be analyzed for differential gene expression, and alternative

splicing may be detected effectively as well ***(58)***. Problems concerning limiting number of known transcripts will in a short time be negligible for important model organisms because genome projects for several species are completed. Identification of all transcribed genes of model organisms initiates the production of arrays harboring every gene for the specific organism and thereby increases the potential of the method.

1.3. Rescue–Suppression-Subtractive Hybridization

As discussed in **Subheading 1**, several techniques have been developed to identify the differences in gene expression between samples from different biological materials. Suppression-subtractive hybridization is one of the techniques that enables the enrichment of nucleic acids representing genes that are uniquely or abundantly expressed in cells of interest ***(7)***. And as explained above **(Subheading 1.1., item 2)**, due to the suppression PCR effect, only dscDNA molecules with two different adaptors can be exponentially expressed; thus, transcripts differentially expressed in one cell population are enriched. Despite the identification of differentially expressed genes by using SSH, a high number of false positives are reported ***(30)***. False positives are transcripts that seem to be differentially expressed by using one technique, but that are unable to be confirmed by using other independent methods, e.g., Northern, the gold standard for verification of differential expression. Sometimes, these transcripts may represent as much as 50–90% of the identified candidates ***(31)***, especially when two closely related RNA populations are being compared. Thus, the most time-consuming step is the confirmation that the isolated cDNA clones represent differentially expressed genes. By combining PCR-based cDNA subtraction and Northern blotting, truly differentially genes can be quickly identified **(Fig. 1)**. Notably, this novel strategy enables integration of the time-consuming confirmation of candidates for differential expressed transcripts in the selection stage ***(21)***.

References

1. Saiki, R. K., Scharf, S., Faloona, F., et al. (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230,** 1350–1354.
2. Kacharmina, J. E., Crino, P. B., and Eberwine, J. (1999) Preparation of cDNA from single cells and subcellular regions. *Methods Enzymol*. **303,** 3–18.
3. Liu, M., Subramanyam, Y. V., and Baskaran, N. (1999) Preparation and analysis of cDNA from a small number of hematopoietic cells. *Methods Enzymol*. **303,** 45–55.
4. Sambrook, J., Fritsch, E., and Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual*, Cold Spring Harbor Laboratory, Cold Spring, NY.

5. Sargent, T. D. and Dawid, I. B. (1983) Differential gene expression in the gastrula of *Xenopus laevis*. *Science* **222,** 135–139.
6. Timberlake, W. E. (1980) Developmental gene regulation in *Aspergillus nidulans*. *Dev. Biol*. **78,** 497–510.
7. Hedrick, S. M., Cohen, D. I., Nielsen, E. A., and Davis, M. M. (1984) Isolation of cDNA clones encoding T cell-specific membrane-associated proteins. *Nature* **308,** 149–153.
8. Sagerstrom, C. G., Sun, B. I., and Sive, H. L. (1997) Subtractive cloning: past, present, and future. *Annu. Rev. Biochem*. **66,** 751–783.
9. Rubenstein, J. L., Brice, A. E., Ciaranello, R. D., Denney, D., Porteus, M. H., and Usdin, T. B. (1990) Subtractive hybridization system using single-stranded phagemids with directional inserts. *Nucleic Acids Res*. **18,** 4833–4842.
10. Rothstein, J. L., Johnson, D., Jessee, J., et al. (1993) Construction of primary and subtracted cDNA libraries from early embryos. *Methods Enzymol*. **225,** 587–610.
11. Houge, G. (1993) Simplified construction of a subtracted cDNA library using asymmetric PCR. *PCR Methods Appl*. **2,** 204–209.
12. Wang, Z. and Brown, D. D. (1991) A gene expression screen. *Proc. Natl. Acad. Sci. USA* **88,** 11,505–11,509.
13. Yokota, H. and Oishi, M. (1990) Differential cloning of genomic DNA: cloning of DNA with an altered primary structure by in-gel competitive reassociation. *Proc. Natl. Acad. Sci. USA* **87,** 6398–63402.
14. Hubank, M. and Schatz, D. G. (1999) cDNA representational difference analysis: a sensitive and flexible method for identification of differentially expressed genes. *Methods Enzymol*. **303,** 325–349.
15. Lisitsyn, N., Lisitsyn, N., and Wigler, M. (1993) Cloning the differences between two complex genomes. *Science* **259,** 946–951.
16. Hubank, M. and Schatz, D. G. (1994) Identifying differences in mRNA expression by representational difference analysis of cDNA. *Nucleic Acids Res*. **22,** 5640–5648.
17. Diatchenko, L., Lau, Y. F., Campbell, A. P., et al. (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc. Natl. Acad. Sci. USA* **93,** 6025–6030.
18. Siebert, P. D., Chenchik, A., Kellogg, D. E., Lukyanov, K. A., and Lukyanov, S. A. (1995) An improved PCR method for walking in uncloned genomic DANN. *Nucleic Acids Res*. **23,** 1087, 1088.
19. Gurskaya, N. G., Diatchenko, L., Chenchik, A., et al. (1996) Equalizing cDNA subtraction based on selective suppression of polymerase chain reaction: cloning of Jurkat cell transcripts induced by phytohemaglutinin and phorbol 12-myristate 13-acetate. *Anal. Biochem*. **240,** 90–97.
20. Lukyanov, K., Diatchenko, L., Chenchik, A., et al. (1997) Construction of cDNA libraries from small amounts of total RNA using the suppression PCR effect. *Biochem. Biophys. Res. Commun*. **230,** 285–288.
21. Leirdal, M., Shadidy, M., Rosok, O., and Sioud, M. (2004) Identification of genes differentially expressed in breast cancer cell line SKBR3: potential identification of new prognostic biomarcers. *Int. J. Mol. Med*. **14,** 217–222.

22. Liang, P. and Pardee, A. B. (1992) Differential display of eukaryotic messenger RNA by means of the polymerase chain reaction. *Science* **257,** 957–971.
23. Liang, P., Zhu, W., Zhang, X., et al. (1994) Differential display using one-base anchored oligo-dT primers. *Nucleic Acids Res.* **22,** 5763, 5764.
24. Martin, K. J. and Pardee, A. B. (1999) Principles of differential display. *Methods Enzymol*. **303,** 234–258.
25. Welsh, J., Chada, K., Dalal, S. S., Cheng, R., Ralph, D., and McClelland, M. (1992) Arbitrarily primed PCR fingerprinting of RNA. *Nucleic Acids Res*. **20,** 4965–4970.
26. Wang, X. and Feuerstein, G. Z. (1995) Direct sequencing of DNA isolated from mRNA differential display. *BioTechniques* **18,** 448–453.
27. Liang, P. and Pardee, A. B. (1995) Recent advances in differential display. *Curr. Opin. Immunol*. **7,** 274–280.
28. Liang, P. (1998) Factors ensuring successful use of differential display. *Methods* **16,** 361–364.
29. Cho, Y.-J., Prezioso, V. R., and Liang, P. (2002) Systematic analysis of intrinsic factors affecting differential display. *BioTechniques* **32,** 762–766.
30. Rosok, O., Odeberg, J., Rode, M., et al. (1996) Solid-phase method for differential display of genes expressed in hematopoietic stem cells. *BioTechniques* **21,** 114–121.
31. Sompayrac, L., Jane, S., Burn, T. C., Tenen, D. G., and Danna, K. J. (1995) Overcomming limitations of the mRNA differential display technique. *Nucleic Acids Res*. **23,** 4738– 4739.
32. Luce, M. J. and Burrows, P. D. (1998) Minimizing false positives in differential display. *BioTechniques* **24,** 766–768.
33. Bosch, I., Melichar, H., and Pardee, A. B. (2000) Identification of differentially expressed genes from limited amounts of RNA. *Nucleic Acids Res*. **28,** E27.
34. Adams, M. D., Kelley, J. M., Gocayne, J. D., et al. (1991) Complementary DNA sequencing: expressed sequence tags and human genome project. *Science* **252,** 1651–1656.
35. Wilcox, A. S., Khan, A. S., Hopkins, J. A., and Sikela, J. M. (1991) Use of 3′ untranslated sequences of human cDNAs for rapid chromosome assignment and conversion to STSs: implications for an expression map of the genome. *Nucleic Acids Res*. **19,** 1837–1843.
36. Adams, M. D., Kerlavage, A. R., Fleischmann, R. D., et al. (1995) Initial assessment of human gene diversity and expression patterns based upon 83 million nucleotides of cDNA sequence. *Nature* **377(Suppl.),** 3–174.
37. Hastie, N. D. and Bishop, J. O. (1976) The expression of three abundance classes of messenger RNA in mouse tissues. *Cell* **9,** 761–764.
38. Zhang, L., Zhou, W., Velculescu, et al. (1997) Gene expression profiles in normal and cancer cells. *Science* **276,** 1268–1272.
39. Bonaldo, M. F., Lennon, G., and Soares, M. B. (1996) Normalization and subtraction: two approaches to facilitate gene discovery. *Genome Res*. **6,** 791–806.
40. Soares, M. B., Bonaldo, M. F., Jelene, P., Su, L., Lawton, L., and Efstratiadis, A. (1994) Construction and characterization of a normalized cDNA library. *Proc. Natl. Acad. Sci. USA* **91,** 9228–9232.

41. Berry, R., Stevens, T. J., Walter, N. A., et al. (1995) Gene-based sequence-tagged-sites (STSs) as the basis for a human gene map. *Nat. Genet.* **10,** 415–423.
42. Houlgatte, R., Mariage-Samson, R., Duprat, S., et al. (1995) The Genexpress index: a resource for gene discovery and the genic map of the human genome. *Genome Res*. **5,** 272–304.
43. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995) Serial analysis of gene expression. *Science* **270,** 484–487.
44. Yamamoto, M., Wakatsuki, T., Hada, A., and Ryo, A. (2001) Use of serial analysis of gene expression (SAGE) technology. *J. Immunol. Methods* **250,** 45–66.
45. Silva, A. P., De Souza, J. E., Galante, P. A., Riggins, G. J., De Souza, S. J., and Camargo, A. A. (2004) The impact of SNPs on the interpretation of SAGE and MPSS experimental data. *Nucleic Acids Res*. **32,** 6104–6110.
46. Saha, S., Sparks, A. B., Rago, C., et al. (2002) Using the transcriptome to annotate the genome. *Nat. Biotechnol*. **20,** 508–512.
47. Unneberg, P., Wennborg, A., and Larsson, M. (2003) Transcript identification by analysis of short sequence tags—influence on tag length, restriction site and transcript database. *Nucleic Acid Res*. **31,** 2217–2226.
48. Datson, N. A., van der Perk-de Jong, J., van den Berg, M. P., de Kloet, E. R., and Vreugdenhil, E. (1999) MicroSAGE: a modified procedure for serial analysis of gene expression in limited amounts of tissue. *Nucleic Acids Res*. **27,** 1300–1307.
49. Peters, D. G., Kassam, A. B., Yonas, H., O'Hare, E. H., Ferrell, R. E., and Brufsky, A. M. (1999) Comprehensive transcript analysis in small quantities of mRNA by SAGE-lite. *Nucleic Acids Res*. **27,** e39.
50. Ye, S. Q., Zhang, L. Q., Zheng, F., Virgil, D., and Kwiterovich, P. O. (2000) miniSAGE: gene expression profiling using serial analysis of gene expression from 1 microg total RNA. *Anal. Biochem* **287,** 1444–1452.
51. Brenner, S., Johnson, M., Bridgham, J., et al. (2000) Gene expression analysis by massively parallell signature sequencing (MPSS) on microbead arrays. *Nat. Biotechnol*. **18,** 630–634.
52. Schena, M., Shalon, D., Davis, R. W., and Brown, P. O. (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* **270,** 467–470.
53. Chee, M., Yang, R., Hubbell, E., et al. (1996) Accessing genetic information with high-density DNA arrays. *Science* **274,** 610–614.
54. Duggan, D. J., Brittner, M., Chen, Y., Meltzer, P., and Trent, J. H. (1999) Expression profiling using cDNA microarrays. *Nat. Genet.* **21 (Suppl. 1),** 10–14.
55. Schulze, A. and Downward, J. (2001) Navigating gene expression using microarrays—a technical review. *Nat. Cell Biol*. **3,** E190–E195.
56. Chung, C. H., Bernard, P. S., and Peron, C. M. (2002) Molecular portraits and the family tree of cancer. *Nat. Genet*. **32 (Suppl. 2),** 533–540.
57. Weeraratna, A. T., Nagel, J. E., de Mello-Coelho, V., and Taub, D. D. (2004) Gene expression profiling: from microarrays to medicine. *J. Clin. Immunol*. **24,** 213–224.
58. Penn, S. G., Rank, D. R., Hanzel, D. K., and Barker, D. L. (2000) Mining the human genome using microarrays of open reading frames. *Nat. Genet*. **26,** 315–318.

7

Genome-Wide Screening by Using Small-Interfering RNA Expression Libraries

Sahohime Matsumoto, Makoto Miyagishi, and Kazunari Taira

Summary

RNA interference (RNAi) is an evolutionarily conserved phenomenon in which gene expression is silenced by double-stranded RNA (dsRNA) in a sequence-specific manner. This technology has the potential to affect all aspects of target discovery and validation. With the completion of the human genome, it is now possible to design small-interfering RNA (siRNA) libraries targeting every human gene. Specific siRNAs, libraries containing a pathway, gene family, or gene set of interest, are expected to unsecure new targets in pathways of therapeutic interest. Here, we highlight the potential of siRNA screens for target identification by using cell-based assays.

Key Words: RNA interference; small-interfering RNA (siRNA) library; target discovery; apoptosis; genomics.

1. Introduction

RNA interference (RNAi) is a phenomenon whereby double-stranded RNAs (dsRNAs) induce the degradation of corresponding target mRNAs in a sequence-specific manner ***(1,2)***. In this process, the dsRNAs are processed by Dicer into approx 21–23 nucleotides (nt) with 2-nt 3′ overhangs, known as small-interfering RNAs (siRNAs) ***(3)***. Subsequently, these siRNAs are incorporated into a silencing complex called RNA-induced silencing complex (RISC) ***(4)***, and the siRNA–RISC complexes recognize and cleave the target mRNA ***(5)*** **(Fig. 1)**. In somatic mammalian cells, long dsRNAs also induce nonspecific effects, which are known as the interferon response ***(6)***. However, transfection of siRNAs in mammalian cells can circumvent the interferon response and

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarker Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

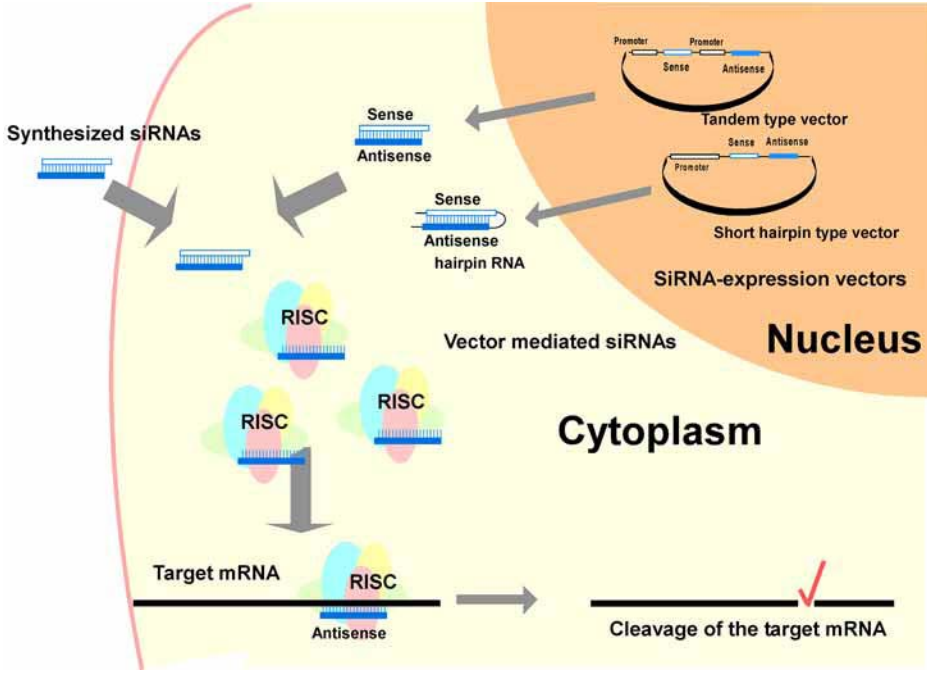


Fig. 1. Schematic of the mechanism of RNA interference.

efficiently induce the cleavage of corresponding mRNAs ***(7)***. Since then, RNAi has received attention as a breakthrough technology for silencing specific genes in mammalian cells. However, the efficiency of transfection of synthesized siRNAs into human cells depends on cell types. Moreover, the RNAi effect seems to be sustained only for short time ***(8)***. To overcome this limitation, several groups have developed vectors encoding hairpin RNAs, leading to the expression of siRNA in cells and sustained gene silencing ***(9–14)***. Applications of RNAi-based screens were initially applied to *Caenorhabditis elegans* ***(15,16)***. Similarly, some recent articles highlight the potential of siRNA libraries for target discovery ***(17–34)***. We have screened siRNA library to identify components of the dsRNA-induced apoptotic pathway. The siRNA library contains 700 vectors targeting 241 genes directly or indirectly involved in apoptosis.

2. Materials

2.1. Construction of siRNA Expression Vectors

1. pcPUR hU6 vector (iGENE Therapeutic, Inc., Tsukuba, Japan, http://www.iGENE-therapeutics.co.jp). This plasmid contains the human U6 promoter and BspMI cloning sites (**Fig. 2**; ***[11]***).

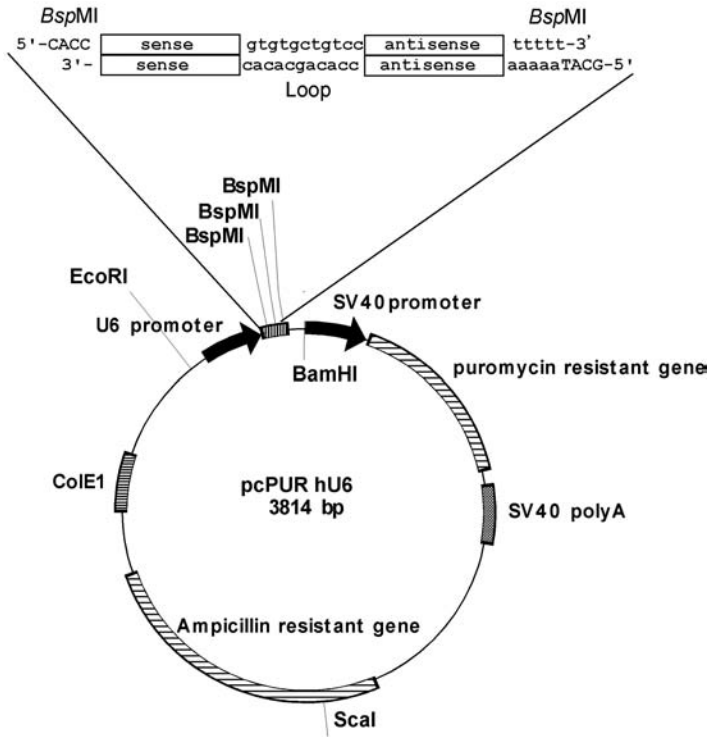


Fig. 2. Schematic of the pcPUR U6i cassette.

2. Custom-synthesized oligonucleotides (oligos): primer: 5′-cac c [sense sequence] gt gtg ctg tcc [antisense sequence] tt ttt-3′; 5′-gca taa aaa [sense sequence] gg aca gca cac gt gtg ctg tcc [antisense sequence]-3′. For each, atleast two favorable sites *(35)* should be selected.
3. BspMI (New England Biolabs, Beverly, MA).
4. TE buffer: 10 m*M* Tris-HCl, pH 8.0, 1 m*M* EDTA.
5. QIAquick® gel extraction kit (QIAGEN, Valencia, CA).
6. Ligation high (Toyobo Engineering, Osaka, Japan).
7. Competent *Escherichia coli* cells (DH-5α).
8. Luria–Bertani (LB) medium: 10 g tryptone, 5 g yeast extract, 10 g NaCl in 1 L of distilled water.
9. LB plates with 100 μg/mL ampicillin.
10. Ampicillin: diluted in distilled water to 100 mg/mL; stock concentration, filter-sterilized (Wako Pure Chemicals, Osaka, Japan).

2.2. Cell Culture and Transient Transfection

1. HeLa S3 cells (Riken Cell Bank).

2. 96-Well microtiter plates (IWAKI, Tokyo, Japan).
3. Dulbecco's modified Eagle's medium (DMEM) (Sigma-Aldrich, St. Louis, MO).
4. Fetal bovine serum (Invitrogen, Carlsbad, CA), heat-inactivated at 56°C for 30 min.
5. 0.25% Trypsin solution with 1 m*M* EDTA (Invitrogen).
6. Antibiotic-antimyoctic (Invitrogen).
7. Lipofectamine 2000 (Invitrogen).
8. Opti-MEM I (serum-free medium, Invitrogen).
9. PBS buffer: 8 g NaCl, 0.2 g KCl, 1.15 g Na_2HPO_4, 0.2 g Na_2HPO_4 in 1 L distilled water. Alternatively, 10X PBS can be purchased from Invitrogen.
10. 0.2% Crystal violet stain: 0.2 g crystal violet dye, 2 mL ethanol, 98 mL distilled water.

2.3. Library Screening

1. Puromycin (Wako Pure Chemicals): prepare stock solution at 1 mg/mL concentration in PBS and then filter-sterilize.
2. Poly(I:C) (GE Healthcare, Little Chalfont, Buckinghamshire U.K.): dissolve in PBS and anneal as follows: heat to 95°C for 2 min with thermal cycler, cool to 72°C, and then cool to 4°C for 2 h. Adjust to the final concentration 1 µg/µL.
3. FuGENE 6 (Roche Diagnostics, Indianapolis, IN).
4. 4% Paraformaldehyde (Wako Pure Chemicals).
5. Crystal violet (Wako Pure Chemicals): prepare 2% stock solution in PBS.

2.4. Western Blotting

1. Lysis buffer: 1% Triton X-100, 150 mM NaCl, 20 mM Tris-HCl, pH 7.6, protease inhibitors (Protease Inhibitor cocktail, Roche Diagnostics, Mannheim, Germany).
2. DC Protein Assay kit (Bio-Rad, Hercules, CA).
3. Sample buffer: 100 mM Tris-HCl, pH 6.5, 20% glycerol, 4% SDS, 4% mercaptoethanol, 0.2% BPB (bromo phenol blue).
4. Bio-Rad Mini-Protein II Electrophoresis and blotting system.
5. Immobilon™-PSQ membrane (Millipore, Billerica, MA), skim milk (BD Biosciences, San Jose, CA).
6. Tris-buffered saline (TBS)-Tween: 100 mM Tris-HCl pH 7.5, 0.9% NaCl, 0.1% Tween-20 (MP Biomedicals, Irvine, CA).
7. Chemiluminescence system (GE Healthcare).

3. Methods

As a first experiment to apply siRNA screen for target identification, we have constructed a limited siRNA library of 241 genes to identify genes that modulate the ability of dsRNA-induced apoptotic pathway ***(28,30)***. We found that the dsRNA-induced apoptotic pathway is composed of at least three independent

pathways. Thus, several signaling proteins might work in concert to affect apoptosis in HeLa S3 cells ***(30)***.

3.1. Construction of siRNA Expression Vectors

The siRNA library was constructed in pcPUR hU6 vector (iGENE Therapeutics, Inc.). The vector contains a human U6 promoter, a puromycin-resistance gene, and BspMI cloning sites (Fig. 2).

3.1.1. Preparation of Vector DNA

1. Digest 3–5 μg of pcPUR hU6 plasmid with restriction enzyme BspMI.
2. Purify the cleaved vector from an agarose gel (0.8%) by using QIAquick gel extraction kit.
3. Elute DNA in 30 μL of TE buffer and use 1 μL for ligation reaction.

3.1.2. Preparation of siRNA Oligonucleotides

1. Design and synthesize the required DNA oligonucleotides for targeting the human genes that are selected to be included in the library (*see* **Note 1**). Once the target sites are selected, for each site design two complementary oligonucleotides: 5′-cac c [sense sequence] gt gtg ctg tcc [antisense sequence] tt ttt-3′ and 5′-gca taa aaa [sense sequence] gg aca gca cac [antisense sequence]-3′ (*see* **Notes 2** and **3**).
2. Each pair of oligonucleotides should be annealed using the following conditions: combine 5 μL of each oligonucleotide (100 μ*M*), heat to 95°C for 2 min with thermal cycler, cool to 72°C and then cool to 4°C for 2 h.
3. Dilute the annealed oligonucleotide to 200-fold with TE buffer.

3.1.3. Ligation and Transformation

1. Mix 1 μL of the annealed oligonucleotides, 1 μL of the purified DNA vector, and 1 μL of ligation high (the DNA ligase is included).
2. Incubate the mixture for at least 30 min at 16°C.
3. Transform the ligated DNA into *E. coli* host cells (DH5α) by using the following protocol: mix the ligated plasmid and competent *E. coli* host cells (DH5α), chill the mixture on ice for 5 min, heat shock the cells in hand for 30 s, and then incubate on ice for 2 min. Plate the cells directly on preheated LB plates with 100 μg/mL ampicillin. More than 90% of colonies should be positive clones.

3.2. Cell Culture and Transient Transfection

1. One day before the transfection, plate HeLa S3 cells at 1×10^4 cells/100 μL in DMEM supplemented with 10% fetal bovine serum and 1% antibiotic-antimyoctic (*see* **Note 4**).
2. Wash the wells twice with PBS buffer.

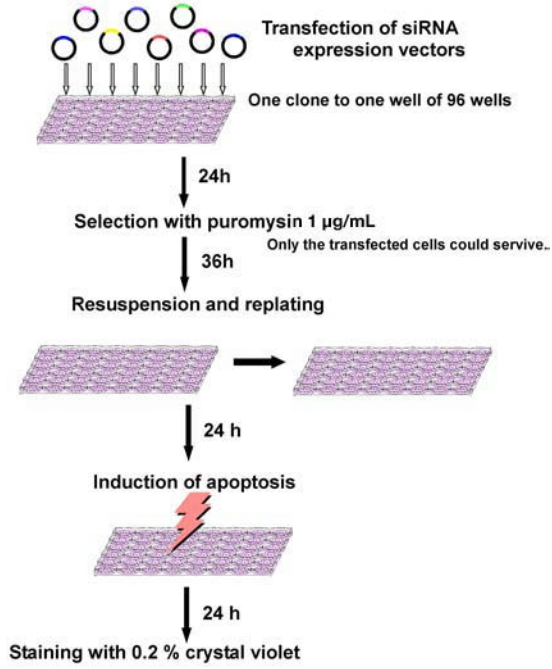


Fig. 3. Schematic screen for proapoptotic genes using siRNA library.

3. Add 100 μL of DMEM into each well.
4. For each well, mix 200 ng of siRNA expression vector and 25 μL of Opti-MEM I. In separate tube, mix 45 μL of LipofectAMINE 2000 and 25 μL of Opti-MEM I (*see* **Note 5**).
5. Add the mixed solutions to cells. Each well should be transfected with only one single siRNA construct.
6. Culture the cells at 37°C for additional 24 h.

3.3. Target Identification Using siRNA Screen

Target identification by using siRNA library involves several steps that are illustrated in **Fig. 3**.

3.3.1. Puromysin Selection

1. Twenty-four hours after transfection, replate the transfecetd cells in a new 96-well plate and add 1 μg/mL (final concentration) puromycin to each well (*see* **Note 6**).
2. Incubate the cell at 37°C for 36 h.
3. Wash the wells twice with PBS and add fresh DMEM, transfer the cells into new 96-well plate, and incubate at 37°C for additional 24 h.

3.3.2. Induction of Apoptosis by Transfection With Poly(I:C)

1. For each well, mix 25 μL of Opti-MEM I, 0.125 μg of poly(I:C), and 0.375 μL FuGENE 6 (*see* **Note 7**).
2. After incubation for 15 min at room temperature, add the mixture to the cells.
3. Culture these cells for 24 h.

3.3.3. Assay

1. After 24-h incubation, wash the wells twice with PBS and fix with 4% paraformaldehyde for 15 min at room temperature.
2. Incubate the fixed cells in 0.2% crystal violet liquid (prepared in PBS) for 15 min and then wash the cells twice with PBS.

In this assay, nonapoptotic cells would be stained with crystal violet. In these cells, genes that are involved in the induction of apoptosis are more likely inhibited.

3.4. Validation of Functional siRNA Expression Vectors by Western Blot Analysis

After the first screening, positive clones should be confirmed with more stringent conditions, which aims to remove the false positives. Check the levels of protein or mRNA knockdown.

1. Plate HeLa S3 cells at 2×10^5 cells/well in a six-well plate and grow for 24 h.
2. For each well, mix 2 μg of siRNA expression vector and 250 μL of Opti-MEM I.
3. For each well, mix 5 μL of LipofectAMINE 2000 reagent and 250 μL of Opti-MEM I medium.
4. Mix both solutions and incubate at room temperature for 15 min.
5. Add the mixture to the cells and incubate at 37°C for 24 h.
6. After incubation, split the cells and transfer approx 1/2 to 1/4 of the split cells to a new six-well plate and add 2 mg/mL puromysin (final concentration) and culture for 36 h. The 96-well plate is assayed for apoptosis as described in **Subheading 3.3**.
7. Remove selection medium, wash the cells twice with PBS, harvest the cells, and centrifuge at 1000*g* for 5 min.
8. Remove the supernatant, add lysis buffer to the pellet, and then sonicate the cells.
9. After sonication, centrifuge the lysates at 10,000*g* for 15 min and determine the protein concentration in the supernatants by using the DC Protein Assay kit.

10. Add equal amount of 4X sample buffer (final concentration 2X) to the proteins and boil at 99°C for 3 min.
11. Size fractionate the proteins (30 µg/well) on 12% sodium dodecyl sulfate-polyacrylamide gel electrophoresis (200 V; 1 h at room temperature).
12. After electrophoresis, transfer proteins to an Immobilon-PSQ membrane (15 V; overnight at 4°C) using Bio-Rad Mini-Protein II blotting system.
13. Block membranes with 5% skim milk for 1 h.
14. Wash the membrane with TBS-Tween three times.
15. Incubate with primary antibody in 0.5% skim milk/TBS-Tween for 1 h at room temperature. Here, we suppose that antibodies against the identified targeted genes are commercially available.
16. Wash the membranes with TBS-Tween three times.
17. Incubate with secondary antibody in 0.5% skim milk/TBS-Tween for 1 h at room temperature.
18. Chemiluminescence reagents detect reactive immunoreactive complexes.

To validate the experimental design and selection assays, include some positive and negative controls, in particular when the signaling proteins involved in the investigated pathway are known (*see* **Note 8**). Here, we have targeted dsRNA-dependent protein kinase R (PKR). The sequences of siRNA expression vectors targeted against two different target sites of PKR transcripts were as follows: site 1, 5′-TAA TGA ATC AAT CAA TTC ATA-3′ and site 2, 5′-AAG ACT AAC TGT AAA TTA TGA-3′. Moreover, two negative controls also were included: NC1, targeting firefly luciferase (5′-ATG AAA CGA TAT GGG CTG-3′) and NC2 targeting *Renilla* luciferase (5′-GTA GCG CGG TGT ATT ATA CCA-3′). As shown in **Fig. 4A**, both siRNA expression vectors targeting PKR suppressed expression of endogenous PKR. Notably, inhibition of PKR inhibited dsRNA-induced apoptosis **(Fig. 4B, C)**. Compared with the positive control, S1 to S4 clones also inhibited dsRNA-induced apoptosis **(Fig. 4D)** (*see* **Note 9**).

Notably, the construction of large siRNA expression libraries against all human genes is a big task. These libraries should enable us to identify a number of known and novel functional genes. Scientists are making a great first step in functional genome analysis, and people will soon enjoy the unraveled mystery of cells ***(17–35)***.

4. Notes

1. To identify potential off-target effects ***(33)*** and nonspecific suppressive effects of siRNA expression vectors, or both, for each gene, at least two sites should be targeted. Two active siRNAs against the same gene should show the same phenotype.
2. In designing the oligonucleotides, we recommend introducing multiple C-to-T (or A-to-G) mutations in the sense strand, avoiding near the 5′-sense terminal end. When constructing hairpin-type siRNA expression vectors, some technical problems may occur, such as sequence difficulty and high rate of mutation in hairpin

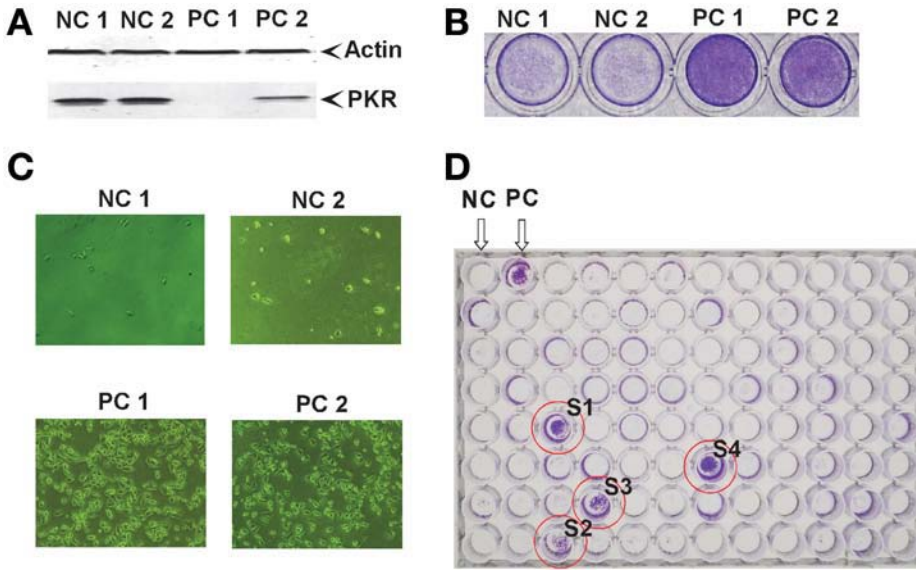


Fig. 4. Example of analysis of dsRNA-induced apoptosis by using siRNA expression vector library. HeLa S3 cells were transfected with siRNA expression vectors directed against the gene for PKR at two different sites (site 1 and site 2) or with negative control vectors (NC1 and NC2). After selection with puromycin, the cells were transfected with poly(I:C). **(A)** Western blotting analysis of extracts of cells transfected with siRNA expression vectors directed against the gene for PKR (site 1 and site 2; PC1 and PC2) and negative controls (NC1 and NC2). **(B)** Plates after staining with crystal violet. **(C)** Microscopic images of cells. **(D)** The wells surrounded by the circle (S1–S4) contain cells that survived under the induction of apoptosis.

region of siRNA sequences. Introducing mutations only into the sense strand of the hairpin RNAs could reduce these problems. Loop sequence derived from natural microRNA sequence was found to improve silencing efficacy.

3. The followings are the guidelines for choosing the gene targets: avoid stop sequences of pol III (UUUU and UUAUU), regions with GC content <30% or >60%, repeat sequences, and sites containing single nucleotide polymorphisms. However, the siRNA should have low internal stability at the 5′-end antisense strand, high internal stability at the 5′ sense strand (but if there is low internal stability at the 5′ end antisense strand, this criterion is not so critical), presence of A (position 19) and U (positions 10 and 19) of the sense strand, absence of G or C (position 19) of the sense strand and absence of close homology to other gene sequences.
4. This protocol was designed for HeLa S3. When other cell lines are used, the following conditions need to be optimized: type of transfection reagents, amount of DNA, used for transfection optimal cell density before transfection and drug

selection, and puromycin concentration. Initial cell number and puromycin concentration affect screening threshold because selection by antibiotics also affects knockdown efficiency.

5. MasterMix for one 96-well plate should contain 2.5 mL of Opti-MEM I + 25 µL of LipofectAMINE 2000 reagent.
6. Puromycin selection is critical because the transfection efficiency is not always 100%.
7. MasterMix for one plate of 96-well plate: 2.5 mL of Opti-MEM I medium, 12.5 µg of poly(I:C) and 37.5 µL of FuGENE6.
8. To select the positive clones correctly, appropriate negative controls should be included. As negative controls, we use a pcPUR hU6 vector for firefly luciferase gene and a pcPUR U6 vector for *Renilla* luciferase gene ***(30)***.
9. Some proteins may have long half-lives and therefore phenotypic changes may require longer incubation. The cells could be analyzed approx 5 d after the transfection with siRNA expression library.

References

1. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391,** 806–811.
2. Zamore, P. D. (2001) RNA interference: listening to the sound of silence. *Nat. Struct. Biol.* **8,** 746–750.
3. Bernstein, E., Caudy, A. A., Hammond, S. M., and Hannon, G. J. (2001) Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409,** 363–366.
4. Hammond, S. M., Boettcher, S., Caudy, A. A., Kobayashi, R., and Hannon, G. J. (2001) Argonaute2, a link between genetic and biochemical analyses of RNAi. *Science* **293,** 1146–1150.
5. Nykanen, A., Haley, B., and Zamore, P. D. (2001) ATP requirements and small interfering RNA structure in the RNA interference pathway. *Cell* **107,** 309–321.
6. Williams, B. R. (1997) Role of the double-stranded RNA-activated protein kinase (PKR) in cell regulation. *Biochem. Soc. Trans.* **25,** 509–513.
7. Elbashir, S. M., Harborth, J., Lendeckel, W., Yalcin, A., Weber, K., and Tuschl, T. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* **411,** 494–498.
8. Yang, S., Tutton, S., Pierce, E., and Yoon, K. (2001) Specific double-stranded RNA interference in undifferentiated mouse embryonic stem cells. *Mol. Cell Biol.* **21,** 7807–7816.
9. Brummelkamp, T. R., Bernards, R., and Agami, R. (2002) A system for stable expression of short interfering RNAs in mammalian cells. *Science* **296,** 550–553.
10. Lee, N. S., Dohjima, T., Bauer, G., et al. (2002) Expression of small interfering RNAs targeted against HIV-1 rev transcripts in human cells. *Nat. Biotechnol.* **20,** 500–505.
11. Miyagishi, M. and Taira, K. (2002) U6 promoter-driven siRNAs with four uridine 3′ overhangs efficiently suppress targeted gene expression in mammalian cells. *Nat. Biotechnol.* **20,** 497–500.

12. Paddison, P. J., Caudy, A. A., Bernstein, E., Hannon, G. J., and Conklin, D. S. (2002) Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev.* **16,** 948–958.
13. Paul, C. P., Good, P. D., Winer, I., and Engelke, D. R. (2002) Effective expression of small interfering RNA in human cells. *Nat. Biotechnol.* **20,** 505–508.
14. Sui, G., Soohoo, C., Affar el, B., Gay, F., Shi, Y., and Forrester, W. C. (2002) A DNA vector-based RNAi technology to suppress gene expression in mammalian cells. *Proc. Natl. Acad. Sci. USA* **99,** 5515–5520.
15. Fraser, A. G., Kamath, R. S., Zipperlen, P., Martinez-Campos, M., Sohrmann, M., and Ahringer, J. (2000) Functional genomic analysis of *C. elegans* chromosome I by systematic RNA interference. *Nature* **408,** 325–330.
16. Gonczy, P., Echeverri, C., Oegema, K., et al. (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* **408,** 331–336.
17. Aza-Blanc, P., Cooper, C. L., Wagner, K., Batalov, S., Deveraux, Q. L., and Cooke, M. P. (2003) Identification of modulators of TRAIL-induced apoptosis via RNAi-based phenotypic screening. *Mol. Cell* **12,** 627–637.
18. Miyagishi, M. and Taira, K. (2003) Strategies for generation of an siRNA expression library directed against the human genome. *Oligonucleotides* **13,** 325–333.
19. Berns, K., Hijmans, E. M., Mullenders, J., et al. (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* **428,** 431–437.
20. Paddison, P. J., Silva, J. M., Conklin, D. S., et al. (2004) A resource for large-scale RNA-interference-based screens in mammals. *Nature* **428,** 427–431
21. Sen, G., Wehrman, T. S., Myers, J. W., and Blau, H. M. (2004) Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nat. Genet.* **36,** 183–189.
22. Shirane, D., Sugao, K., Namiki, S., Tanabe, M., Iino, M., and Hirose, K. (2004) Enzymatic production of RNAi libraries from cDNAs. *Nat. Genet.* **36,** 190–196.
23. Hsieh, A. C., Bo, R., Manola, J., et al. (2004) A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res.* **32,** 893–901.
24. Kaykas, A. and Moon, R. T. (2004) A plasmid-based system for expressing small interfering RNA libraries in mammalian cells. *BMC Cell Biol.* **5,** 16.
25. Colland, F., Jacq, X., Trouplin, V., et al. (2004) Functional proteomics mapping of a human signaling pathway. *Genome Res.* **14,** 1324–1332.
26. Luo, B., Heard, A. D., and Lodish, H. F. (2004) Small interfering RNA production by enzymatic engineering of DNA (SPEED) *Proc. Natl. Acad. Sci. USA* **101,** 5494–5499.
27. Zheng, L., Liu, J., Batalov, S., et al. (2004) An approach to genomewide screens of expressed small interfering RNAs in mammalian cells. *Proc. Natl. Acad. Sci. USA* **101,** 135–140.
28. Futami, T., Miyagishi, M., and Taira, K. (2004) Identification of a network involved in thapsigargin-induced apoptosis using a library of siRNA-expression vectors. *J. Biol. Chem.* **280,** 826–831.

29. Miyagishi, M., Matsumoto, S., and Taira, K. (2004) Generation of an shRNAi expression library against the whole human transcripts. *Virus Res.* **102,** 117–124.
30. Matsumoto, S., Miyagishi, M., Akashi, H., Nagai, R., and Taira, K. (2005) Analysis of double-stranded RNA-induced apoptosis pathways using interferon-response noninducible small interfering RNA expression vector library. *J. Biol. Chem* **280,** 25,687–25,696.
31. Westbrook, T. F., Martin, E. S., Schlabach, M. R., et al. (2005) A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121,** 837–848.
32. Kolfschoten, I. G., van Leeuwen, B., Berns, K., et al. (2005) A genetic screen identifies PITX1 as a suppressor of RAS activity and tumorigenicity. *Cell* **121,** 849–858.
33. Jackson, A. L., Bartz, S. R., Schelter, J., et al. (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat. Biotechnol.* **21,** 635–637.
34. Curran, S. P., Leverich, E. P., Koehler, C. M., and Larsen, P. L. (2004) Defective mitochondrial protein translocation precludes normal Caenorhabditis elegans development. *J. Biol. Chem*. **279,** 52,655–52,662.
35. Miyagishi, M. and Taira, K. (2005) siRNA becomes smart and intelligent. *Nat. Biotechnol.* **23,** (8): 946–947.

8

Hammerhead Ribozyme-Based Target Discovery

Masayuki Sano and Kazunari Taira

Summary

With the accumulation of vast amounts of data as a result of the sequencing of the human genome, it is necessary to identify human genes that are involved in various cellular, developmental, and disease-related processes and to clarify their functions and potential utility as targets in the treatment of disease. Identification methods based on the use of hammerhead and hairpin ribozymes have received increasing attention as possible tools for the rapid identification of key genes involved in biological processes. This chapter describes the method known as gene-discovery by a hammerhead ribozyme library for elucidation of the gene function. Use of this technology has already revealed new insights into several important biological phenomena.

1. Introduction

A hammerhead ribozyme is a small RNA molecule with catalytic activity and it can be engineered to optimize its activity in the intracellular environment. The hammerhead ribozyme recognizes a target gene via interactions between the binding arms of the ribozyme and the gene transcript, and it can cleave any RNA substrate that contains an NUH triple, where N is any nucleotide and H is A, C, or U. To date, numerous studies have been directed towards applications of ribozymes in vivo, and many successful experiments suppressing genes of interest have been reported ***(1,2)***.

In our laboratory, successful inactivation of genes by hammerhead ribozymes in living cells has been achieved by using an appropriate expression cassette **(Fig. 1A)**. The use of the human $tRNA^{Val}$ promoter ensures a high level of expression of the ribozymes, and an appropriate design of the expression cassette leads to the intracellular stability and cytoplasmic transport of the ribozyme transcripts ***(3–5)***. These features would apparently be advantageous for the exploitation of ribozymes in vivo.

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols*
Volume I, Emerging Strategies for Targets and Biomarker Discovery
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

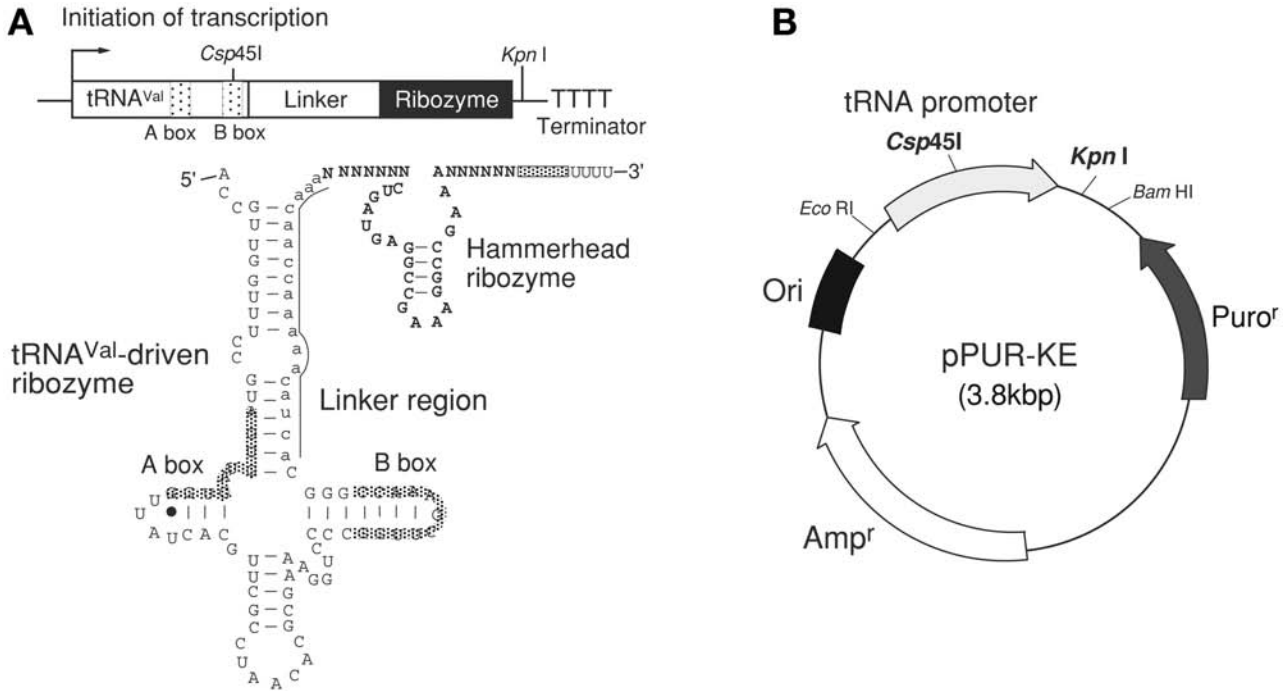


Fig. 1. The ribozyme-expression system. **(A)** The $tRNA^{Val}$ portion is attached to the hammerhead ribozyme (indicated by the bold letters) because the promoters of the gene, namely A and B boxes, are internal. **(B)** A plasmid for cloning of randomized ribozymes. The promoter of a human gene for $tRNA^{Val}$ is inserted into the pPUR plasmid. The restriction enzyme sites, *Csp45*I and *Kpn*I, are used for cloning.

Libraries of randomized hammerhead ribozymes provide a novel tool for the identification of functional genes ***(6–10)***. While the binding arms can be designed to include sequences complementary to the target RNA, randomization of binding arms generates a large pool of ribozymes that are capable of cleaving many mRNA substrates. Such a pool of randomized ribozymes is referred to as a library.

In a ribozyme-based gene-discovery system, a library of randomized ribozymes is introduced into cells, where it causes phenotypic changes as a result of cleaving transcripts that are critically involved in a particular phenotype. The ribozymes can be recovered from the cells and subsequent sequence analysis, which involves searching for sequences of a target site in databases with, for example, the BLAST search program, leads to the identification of the targets of the ribozymes ***(6–16)***. Alternatively, a partial cDNA fragment that subsequently facilitates the cloning of full-length cDNA can be obtained by 5′ and 3′ rapid amplification of cDNA ends (RACE), a PCR-based cloning technique ***(11–15)***. Once target genes have been identified, they can be characterized individually with respect to the specific involvement of each in the phenotype of interest. The process can then be refined and repeated.

The system is simple and direct, as compared to other "high-throughput" gene-screening methods, and no prior sequence information is required. Moreover, because the system allows the identification of direct effectors of a change in phenotype, it reduces the level of background and false-positive results and produces, for the most part, interpretable results. Thus, this gene discovery system based on the ribozyme library is indeed useful for the rapid identification of functional genes in the post-genome era.

2. Materials

2.1. Construction of the Randomized Hammerhead Ribozyme-Expression Vector

1. $tRNA^{Val}$-expression vector: pPUR-KE (a plasmid that contains the promoter of a human gene for $tRNA^{Val}$ between the *Eco*RI and *Bam*HI sites of a plasmid pPUR [Clontech, Palo Alto, CA]; **Fig. 1B**).
2. Custom-synthesized oligonucleotides.
   a. Template DNA: 5′-TCC CCG GTT CGA AAC CGG GCA CTA CAA AAA CCA ACT TTN NNN NNN CTG ATG AGG CCG AAA GGC CGA AAN NNN NNG GTA CCC CGG ATA TCT TTT TTT-3′.
   b. Primer A: 5′-TCC CCG GTT CGA AAC CGG GCA-3′.
   c. Primer B: 5′-GCT TGC ATG CCT GCA GGT CGA CGC GAT AGA AAA AAA GAT ATC CGG GGT-3′.
3. 10X *Taq* polymerase buffer: 100 m*M* Tris-HCl, pH 8.3, 15 m*M* $MgCl_2$, 500 m*M* KCl.
4. dNTP mixture: dATP, dCTP, dGTP, and dTTP, each at 2.5 m*M*.

5. 10X L (low-salt) buffer: 100 m*M* Tris-HCl, pH 7.5, 100 m*M* $MgCl_2$, 10 m*M* dithiothreitol (DTT).
6. 5 U/µL *Taq* DNA polymerase.
7. Restriction enzymes: 8 U/µL *Csp*45I; 10 U/µL *Kpn*I.
8. 50X TAE buffer: 2 *M* Tris-acetate, 0.05 *M* EDTA, pH 8.0.
9. Ethanol: both 100 and 70% in distilled water.
10. Phenol and chloroform.
11. Agarose.
12. SOC medium: 10 g bacto-tryptone, 5 g bacto-yeast extract, 10 g NaCl in 1 L distilled water.
13. *Escherichia coli* (JM109, DH-5α or other) competent cells.
14. Kit for purification of DNA fragments from agarose gels (e.g., QIAquick Gel Extraction Kit; Qiagen, Hilden, Germany).
15. Kit for purification of PCR products (e.g., QIAquick PCR purification kit; Qiagen).
16. Kit for ligation of DNA fragments to the vector (e.g., DNA Ligation kit Ver.2; Takara Shuzo Co., Kyoto, Japan).

2.2. Introduction of a Ribozyme Library Into Cells and Recovery of the Ribozymes

1. Cell lines (e.g., HeLa S3, HT1080, B16-BL6 mouse melanoma cells or other).
2. Cell growth medium (e.g., Dulbecco's modified Eagle's medium [DMEM]; RPMI 1640, Sigma Chemical, St. Louis, MO).
3. Serum-free medium (e.g., Opti-MEM I, Life Technologies).
4. Fetal bovine serum.
5. Antibiotic mixture (e.g., Antibiotic-Antimycotic, Life Technologies).
6. Transfection reagents (e.g., TransIT-LT1, Pan Vera, Madison, WI; LipofectAMINE™ 2000, Invitrogen, Carlsbad, CA).
7. PBS buffer: 8000 mg NaCl, 200 mg KCl, 1150 mg Na_2HPO_4 (anhydrous), 200 mg KH_2PO_4 (anhydrous) in 1000 mL distilled water.
8. Trypsine-EDTA.

3. Methods

3.1. Construction of a Library Expressing Randomized Hammerhead Ribozymes

The method for constructing a hammerhead ribozyme library includes three steps. At first, DNA templates encoding ribozyme catalytic core flanked by a substrate-binding region that is completely randomized should be chemically synthesized and amplified by polymerase chain reaction (PCR). After restriction digestion of PCR fragments, they are ligated into the plasmid. Finally, the plasmids are used for transformation of competent *E. coli* cells. Importantly, we must care and maintain the diversity of a ribozyme library in all steps.

3.1.1. Preparation of the Insert DNA

1. Synthesize or purchase the DNA template described here;
   5′-TCC CCG GTT CGA AAC CGG GCA CTA CAA AAA CCA ACT TT**N NNN NNN** CTG ATG AGG CCG AAA GGC CGA A**AN NNN NN**G GTA CCC CGG ATA TCT TTT TTT-3′
   Bold letters correspond to the sequence of the ribozyme's binding arms, and N stands for any nucleotides (*see* **Note 1**).
2. Amplify the DNA template by PCR. For PCR, the reaction mixture contains (final concentration): 0.02 μ*M* template DNA, 0.5 μ*M* primer A, 0.5 μ*M* primer B, 1 X *Taq* buffer, dNTP mix (0.25 m*M* each), and 10 U *Taq* polymerase. Place the reaction mixture in a thermal cycler. Program the cycler to excute 6–8 cycles with the following program: 95°C, 30 s; 55°C, 30 s; 74°C, 30 s.
3. Purify the PCR fragments by using a column (e.g., QIAquick PCR purification kit) or phenol-chloroform extraction, and ethanol-precipitation (*see* **Note 2**).
4. Digest the PCR fragments with the restriction enzymes, *Csp45*I and *Kpn*I, in 1X L (low salt) buffer for >2 h at 37°C. After the reaction, purify the fragments as described in **step 3**.

3.1.2. Preparation of Plasmids for Expressing Randomized Ribozymes

1. Digest the vector plasmid "pPUR-KE" with *Csp45*I and *Kpn*I. After digestion, load the reaction mixture onto a 1.0% agarose gel and electrophorese the sample in 1X TAE buffer.
2. Excise gel pieces that contain the fragment of the vector and purify them using a column (e.g. QIAquick Gel Extraction Kit).
3. Mix 20 μg of the linerized vector and 550 ng of the insert DNA. The ligation reaction is performed by using DNA Ligation kit Ver.2 at 16°C for >4 h (*see* **Note 3**).
4. For transformation, mix 20 mL of *E. coli* host cells and 1 mL of the ligation mixture and incubate it on ice for at least 30 min. After heating the mixture at 42°C in a hot water, add preheated 180 mL of SOC medium at 37°C.
5. Incubate the *E. coli* mixture at 37°C for 1 h and then plate a part of the mixtures (10 μL and 100 μL) on LB agar plates that contain ampicillin at 100 μg/mL (*see* **Note 4**). Residual solution is subjected to the plasmid preparation.
6. After colonies have become apparent (12–16 h after plating), count the number of colonies to statistically check the diversity of the library. For example, if 1000 colonies are found on a plate to which 10 μL of *E. coli* solution in 200 mL has been plated, the diversity is estimated as 2×10^7=1000 (colonies) × 200 (mL)/ 10 (μL). To confirm the diversity, about 50 independent clones should be picked up and sequenced.

3.2. Transfection of the Library Expressing Randomized Hammerhead Ribozymes and Induction of the Selection Pressure

In this system, genes responsible for a specific phenotype of interest can be isolated using a ribozyme library. When the library is introduced into cells,

ribozymes may cleave messenger RNAs, and as a result, may change the cell phenotype (**Fig. 2**). Thus, the identification of the target of the ribozyme allows to the identification of the genes responsible for the specific phenotype.

Probably, one of the most important steps in this system is to isolate cells with a phenotypic change caused by introducing the library. So far, several studies have identified genes of interest that are responsible for many specific phenotypes (***6–16***). A major advantage of the use of the ribozyme library is that no prior sequence information is required and, thus, it allows the identification of previously unidentified genes when large numbers of ribozymes are introduced into cells.

For successful experiment, we must carefully discriminate between cells with phenotypic change and unchanged cells, because the recovery of plasmids from unchanged cells increases the number of false positives. Thus, we must fully consider the strategy for reducing such false positives prior to the introduction of the library into cells.

For example, we identified genes responsible for cell migration by a library of randomized hammerhead ribozymes (***6***). A line of highly invasive cancer cells, HT1080 fibrosarcoma cells, was transfected with a plasmid vector harboring a randomized hammerhead ribozyme library and subjected to a chemotaxis assay in 12-μm-pore Transwell inserts with fibronectin as a chemo-attractant. Although HT1080 cells normally migrate toward a chemo-attractant, some cells did not migrate toward the chemo-attractant after treatment with the randomized ribozyme library. Ribozymes were recovered from cells that remained in the top chamber and the randomized regions of the ribozymes were sequenced in order to search for the relevant sequences in databases (**Fig. 3**).

At the first step, thus, we must design a sophisticated strategy and optimize the condition for separating positive cells. Here, we describe typical transfection protocol for introducing a ribozyme library into cells. After transfection, cells should be separated by induction of the selection pressure that one chooses.

1. The day before transfection, plate cells at 70–90% confluence in a 10-cm dish on the day of transfection.
2. Mix 15 μL of LipofectAMINE™ 2000 with 750 μL of OptiMEM I and incubate at room temperature for 5 min. Appropriate transfection regents can be used, depending on the type of cells used.
3. Mix 15 μg of the plasmid DNA and 750 μL of OptiMEM I.
4. Combine the diluted DNA solution with the diluted LipofectAMINE™ 2000 reagent and incubate for 15 min. While complex formation takes place, wash cells with phosphate-buffered saline (PBS) and replace with 8.5 mL of fresh cell-growth medium.
5. Apply the complexes to cells and incubate at 37°C for appropriate time (typically 24–48 h) in a CO_2 incubator.
6. Induce the specific selection-pressure to isolate the cells with the phenotypic change.

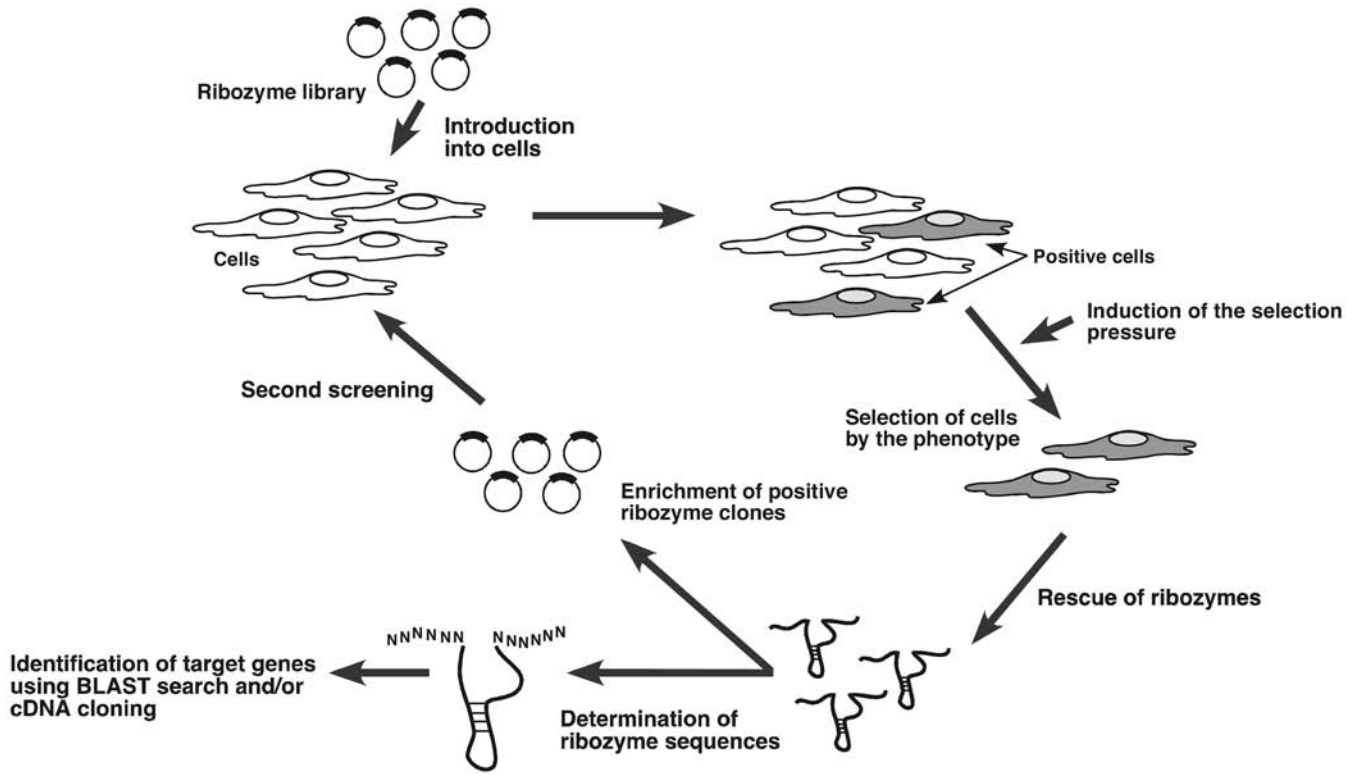


Fig. 2. Discovery of genes involved in the particular phenotype using a randomized ribozyme library. The use of the randomized ribozyme library as a tool for gene discovery involves the introduction of the library into cells, selecting cells that have a change in phenotype, isolating cells that harbored positive ribozymes, recovering ribozymes from isolated cells, re-introducing positive ribozymes into fresh cells, and performing the assay again (if needed), sequencing positive ribozymes, and identifying the target genes involved in the particular phenotype.

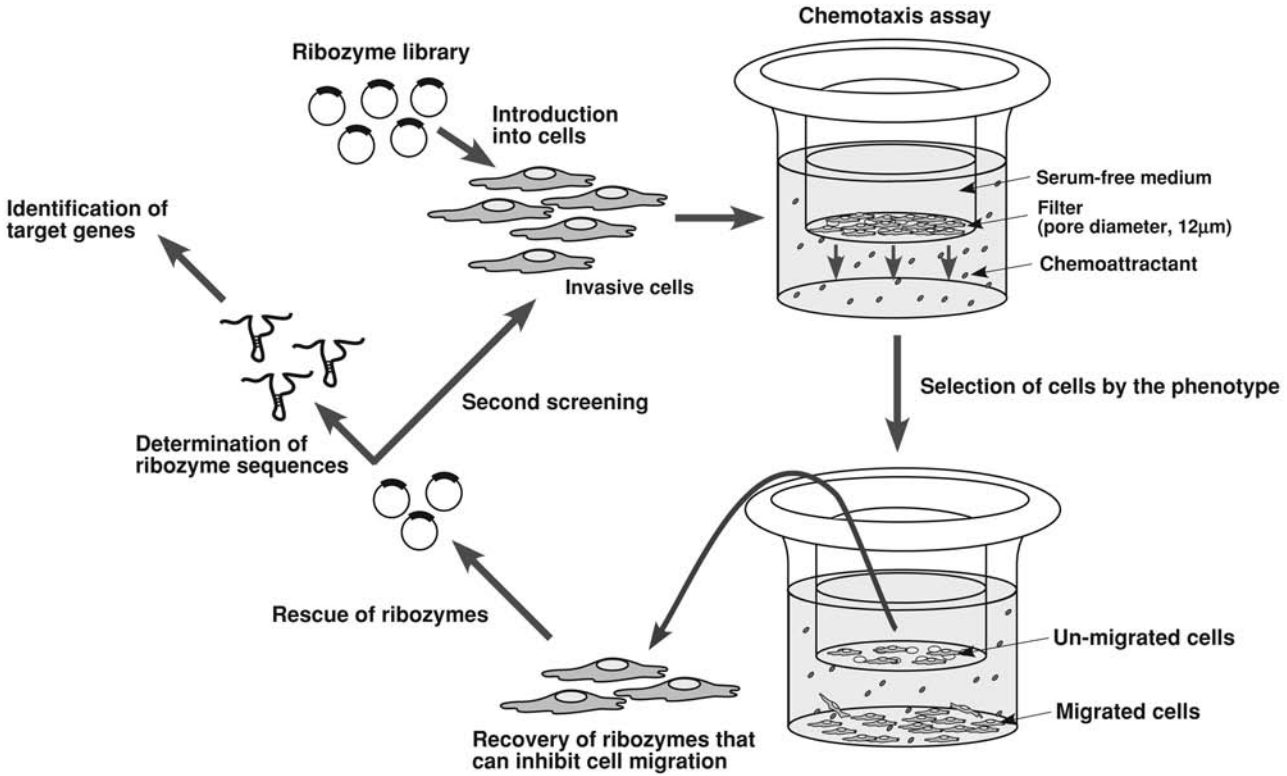


Fig. 3. A system for the identification of genes involved in metastasis. A hammerhead ribozyme library is used to identify genes that promote cell migration. After introducing it into invasive cells, chemotaxis assay is performed using Transwells with a chemo-attractant. Cells with reduced migration are isolated from the top chamber, and positive ribozymes are rescued and sequenced to identify genes that can promote cell migration.

3.3. Recovery of Active Ribozymes

To isolate plasmids from cells, Hirt's DNA isolation method is often used. This method enables the recovery of plasmids without nicking, but it is time consuming and requires many steps. Thus, instead of it, commercially available Miniprep kit also can be use, such as QIAGEN's Miniprep kit, to isolate plasmids from mammalian cells.

When we introduce the ribozyme library into cells, we must consider that one transfected cell may have more than one type of plasmid. In such case, if one of the ribozymes should influence the phenotype, other plasmids are also isolated from the positive cells. Such unrelated plasmids may be regarded as the false-positives. To reduce the false-positives, in many cases, we need to repeat the same screening at least three to four times. To do this, the isolated plasmids from first-cycle selection are used to transform competent *E. coli* cells. The resultant library should be introduced into cells as second-cycle selection. To determine the nature of plasmids, we should sequence individual colonies in each cycle.

3.4. Identification of the Candidate Genes Responsible for Phenotypic Changes

With the sequence information derived from substrate-binding arms of the ribozyme, we can identify candidate genes responsible for the phenotypic changes either by 5′ and 3′ rapid amplification of cDNA ends (RACE) or simple bioinformatics, searching for homology using the entire sequence as query. The latter method easily leads to the identification of the candidates. In our laboratory, we screen targets by mining the publicly accessible gene database with the BLAST program at the NCBI website (http://www.ncbi.nlm.nih.gov/blast/).

1. Access the NCBI website; Click "search for short nearly exact matches."
2. Fill in the sequence in the search box. The obtained sequence probably is 15 nt with one gap such as NNNNNNTHNNNNNNN, where H can be filled in this format (*see* **Notes 5** and **6**).
3. Click the [BLAST!] button.
4. Click the [Format!] button. The result will be ready within 5 min.
5. Sequences and the names of genes with score (bits) and E value are listed. A higher score and lower E value indicate higher homology.

If candidates of interest are found, they must be validated by using other ribozymes that are capable of cleaving a specific-target mRNA at other sites. Inactive ribozymes with point mutations within the catalytic site are convenient controls in such analysis. Alternatively, inactivation by small interfering RNAs (siRNAs) can be also used to validate candidate genes. Once target genes have been identified, they can be characterized individually with respect to the specific involvement of each in the phenotype of interest.

4. Notes

1. We recommend that the length of ribozyme's binding arms should be 7 nt each. A shorter length of the binding arms may reduce the binding ability to a substrate and longer length may cause nonspecific binding.
2. The PCR purification kit removes enzymes and salts as well as excess primers.
3. To check the quality of plasmids, linearized plasmids also should be performed the ligation reaction in the absent of the insert DNA. Then, we must confirm that no or very few transformed colonies appear.
4. Do not incubate *E. coli* cells for more than 1 h. Since each transformed cell replicates at different rates, longer incubation may vary the original population of each cell.
5. In the case of a hammerhead ribozyme, a gap nucleotide described as H (A, C, or U) does not form a base pairing.
6. Optionally, species restriction may reduce the unrelated sequences. In addition, ribozymes often target EST (expression sequencing tag), which is the identified sequences expressing in cells.

Acknowledgment

The authors thank Dr. Renu Wadhwa for a critical reading of the manuscript.

References

1. Eckstein, F. and Lilley, D. M. J., eds. (1996) Nucleic Acids and Molecular Biology, vol. 10, Springer-Verlag, Berlin
2. Sioud, M., ed. (2004) *Ribozymes and siRNA Protocols, Second Edition*. Vol. 252, Methods in Molecular Biology, Humana Press, Totowa NJ.
3. Koseki, S., Tanabe, T., Tani, K., et al. (1999) Factors governing the activity in vivo of ribozymes transcribed by RNA polymerase III. *J. Virol*. **73,** 1868–1877.
4. Kato, Y., Kuwabara, T., Warashina, M., Toda, H., and Taira, K. (2001) Relationships between the activities in vitro and in vivo of various kinds of ribozyme and their intracellular localization in mammalian cells. *J. Biol. Chem.* **276,** 15,378–15,385.
5. Kuwabara, T., Warashina, M., Koseki, S., et al. (2001) Significantly higher activity of a cytoplasmic hammerhead ribozyme than a corresponding nuclear counterpart: engineered tRNAs with an extended 3' end can be exported efficiently and specifically to the cytoplasm in mammalian cells. *Nucleic Acids Res*. **29,** 2780–2788.
6. Suyama, E., Kawasaki, H., Kasaoka, T., and Taira, K. (2003) Identification of genes responsible for cell migration by a library of randomized ribozymes. *Cancer Res*. **63,** 119–124.
7. Suyama, E., Kawasaki, H., Nakajima, M., and Taira, K. (2003) Identification of genes involved in cell invasion by using a library of randomized hybrid ribozymes. *Proc. Natl. Acad. Sci. USA* **100,** 5616–5621.
8. Suyama, E., Wadhwa, R., Kaur, K., et al. (2004) Identification of Metastasis-related Genes in a Mouse Model Using a Library of Randomized Ribozymes. *J. Biol. Chem*. **279,** 38,083–38,086.

9. Onuki, R., Bando, Y., Suyama, E., et al. (2004) An RNA-dependent protein kinase is involved in tunicamycin-induced apoptosis and Alzheimer's disease. *EMBO J.* **23,** 959–968.
10. Wadhwa, R., Yaguchi, T., Kaur, K., et al. (2004) Use of a randomized hybrid ribozyme library for identification of genes involved in muscle differentiation. *J. Biol. Chem.* **279,** 51,622–51,629.
11. Kruger, M., Beger, C., Li, Q. X., et al. (2000) Identification of eIF2Bgamma and eIF2gamma as cofactors of hepatitis C virus internal ribosome entry site-mediated translation using a functional genomics approach. *Proc. Natl. Acad. Sci. USA* **97,** 8566–8571.
12. Li, Q. X., Robbins, J. M., Welch, P. J., Wong-Staal, F., and Barber, J. R. (2000) A novel functional genomics approach identifies mTERT as a suppressor of fibroblast transformation. *Nucleic Acids Res.* **28,** 2605–2612.
13. Welch, P. J., Marcusson, E. G., Li, Q. X., et al. (2000). Identification and validation of a gene involved in anchorage-independent cell growth control using a library of randomized hairpin ribozymes. *Genomics* **66,** 274–283.
14. Beger, C., Pierce, L. N., Kruger, M., et al. (2001) Identification of Id4 as a regulator of BRCA1 expression by using a ribozyme-library-based inverse genomics approach. *Proc. Natl. Acad. Sci. USA* **98,** 130–135.
15. Kruger, M., Beger, C., Welch, P. J., Barber, J. R., Manns, M. P., and Wong-Staal, F. (2001) Involvement of proteasome alpha-subunit PSMA7 in hepatitis C virus internal ribosome entry site-mediated translation. *Mol. Cell Biol.* **21,** 8357–8364.
16. Waninger, S., Kuhen, K., Hu, X., Chatterton, J. E., Wong-Staal, F., and Tang, H. (2004) Identification of cellular cofactors for human immunodeficiency virus replication via a ribozyme-based genomics approach. *J. Virol.* **78,** 12,829–12,837.

9

Production of siRNA- and cDNA-Transfected Cell Arrays on Noncoated Chambered Coverglass for High-Content Screening Microscopy in Living Cells

Holger Erfle and Rainer Pepperkok

Summary

In this chapter, we provide a protocol for the production of transfected cell arrays in living mammalian cells on noncoated chambered coverglass for the systematic functional analyses of human genes by high-content screening microscopy. This method should facilitate drug target validation by small-interfering RNAs.

Key Words: Cell arrays; genomics; high-content screening microscopy; small-interfering RNAs; siRNA.

1. Introduction

The information available through sequencing of several genomes together with high-throughput techniques such as protein analysis by mass spectrometry or expression and transcription profiling by protein or DNA microarrays has the potential to help analyze the complexity of biological processes on a comprehensive scale (*see* Chapters 2, 4, 5, Volume 1). An elegant high-throughput method allowing parallel analysis of gene function in intact living cells has been introduced recently by Ziauddin and Sabatini ***(1)***. In this method, plasmid DNAs or small-interfering RNAs (siRNAs) ***(2)*** are printed together with the appropriate transfection reagents in a gelatin matrix at defined locations on glass slides. Overlaying these arrays with tissue culture cells results in clusters of living cells transfected with siRNAs or expressing the respective cDNAs at each location. In combination with automated fluorescence scanning microscopy and image processing ***(3,4)***, this method allows the rapid analysis of gene function on a large scale in intact cells.

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarker Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

2. Materials

1. siRNA oligonucleotides (QIAGEN, Valencia, CA).
2. Cy3-labeled DNA marker oligomer (BioSpring, Frankfurt, Germany).
3. 384-Well plates (Nalge Nunc International, Rochester, NY).
4. Lab-Tek chambered coverglass (cat. no. 155361, Nalge Nunc International).
5. Transfection reagent Effectene (QIAGEN, Hilden, Germany) or LipofectAMINE 2000 (Invitrogen, Carlsbad, CA).
6. Sucrose (USB, Cleveland, OH).
7. Gelatin (cat. no. G-9391, Sigma, St. Louis, MO).
8. Fibronectin (Sigma).

3. Methods

The method involves five major steps **(Fig. 1)**, including the preparation of the transfection solutions, followed by their spotting onto a cell substrate (e.g., Lab-Tek culture dishes, Nalge Nunc International), plating of the cells onto the arrays of spotted transfection solutions, preparation of the transfected cells for functional analysis, and the analysis of transfected cells by high-content screening microscopy.

3.1. Sources of Tagged cDNAs and siRNAs and Preparation of Transfection Solutions

3.1.1. Sources of Tagged cDNAs and siRNAs

Information on the collection of novel human cDNAs fused to spectral variants of the green fluorescent protein (GFP) is available at http://gfp-cdna.embl.de/ ***(5)***.

Synthesized siRNAs targeting human proteins can be obtained from QIAGEN (http://www1.qiagen.com), Ambion (http://www.ambion.com/), or Dharmacon (http://www.dharmacon.com).

3.1.2. Preparation of Transfection Solutions

The siRNA (plasmid cDNA) gelatin transfection solutions are prepared in 384-well plates (Nalge Nunc International). As transfection reagent, we use Effectene (QIAGEN) or LipofectAMINE 2000 (Invitrogen), giving optimal transfection efficiencies for both siRNAs and plasmid cDNAs in MCF7, HeLa, COS7L, or human embryonic kidney (HEK)293 cells. The presence of sucrose in the spotting solution facilitates the storage of the dried arrays without loss in transfection efficiencies. Additionally, it is essential for the successful transfer of the siRNA (cDNA)–gelatin transfection solutions to uncoated substrates during the spotting procedure. The presence of fibronectin in the gelatin solution increases cell adherence, resulting in a reduced migration of transfected

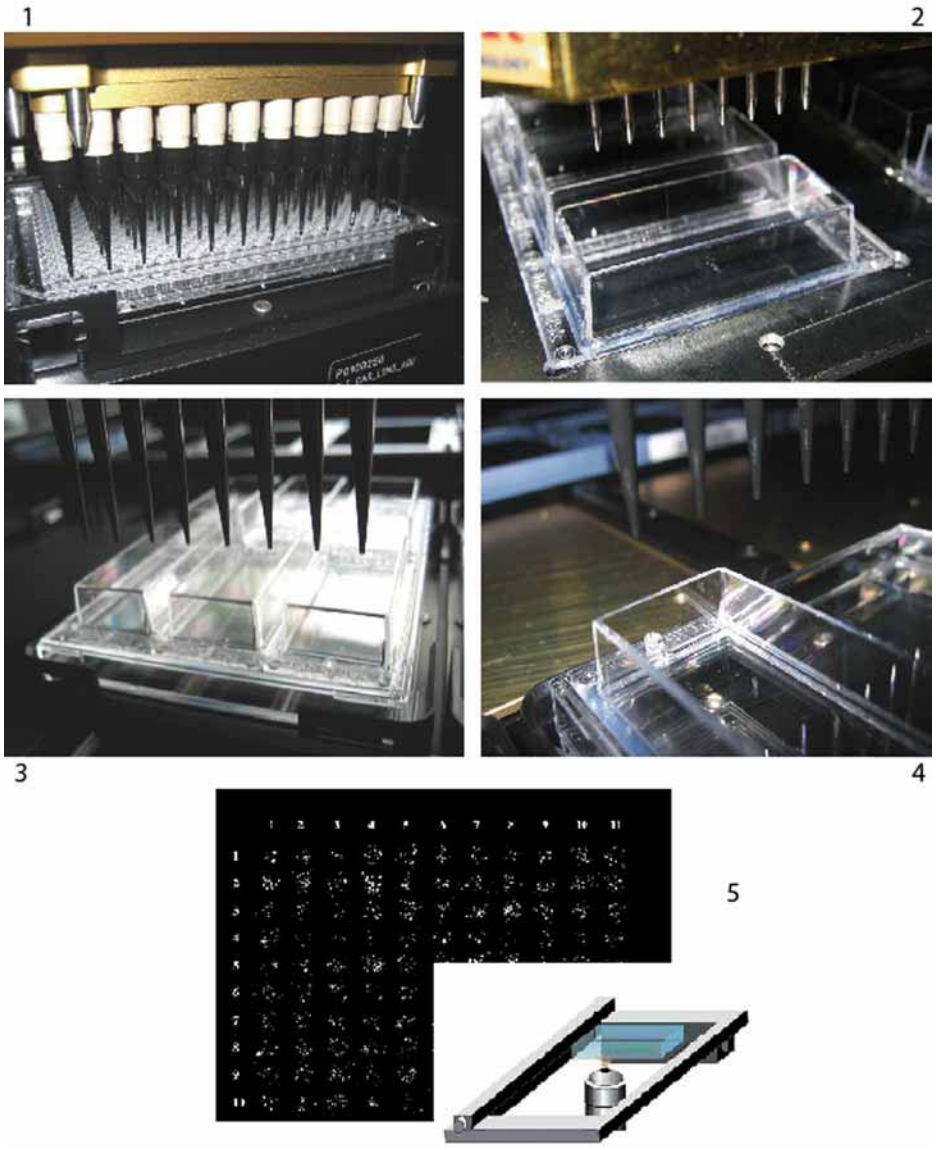


Fig. 1. Five basal steps of the method: (1) Preparation of the transfection solutions on an automated liquid handler with a 96-pipet head. (2) Spotting of the transfection solutions with a spotting Robot, e.g. ChipWriter Compact, on Lab-Tek chambered coverglass. (3) Plating of the cells on dishes with the dried transfection solutions. (4) Preparation of samples (manually or automatically) for functional analysis, e.g., immunostaining. (5) Analysis of samples by high-content screening microscopy.

cells away from the spot region. To retrieve the spot regions and to highlight successfully transfected cells of siRNA experiments, a Cy3-labeled DNA oligonucleotide is used as cotransfection marker. In this case, 0.5 µL of a 40 µ*M* marker solution is included in step 1 of the protocol, resulting in a total oligonucleotide volume of 1.5 µL (1 µL of siRNA plus 0.5 µL of Cy3-labeled oligonucleotide). With this protocol, it is possible to cotransfect plasmid cDNA and siRNA. In this case, 1 µL of siRNA plus 1 µL of plasmid cDNA are added in step 1, resulting in a total volume of 2 µL.

1. Add 1 µL of the respective siRNA stock solution (20 µ*M* in RNA dilution buffer as supplied by the manufacturer) to each well. For plasmid transfections, 1 µL of plasmid DNA at a stock concentration of 500 ng/µL is added.
2. Add 7.5 µL EC buffer (part of the Effectene transfection kit, QIAGEN) containing 0.2 *M* sucrose and mix thoroughly by pipeting three times up and down. The EC buffer can be replaced by water, also containing 0.2 *M* sucrose for LipofectAMINE 2000 (Invitrogen).
3. Incubate the mixture for 10 min at room temperature.
4. Add 4.5 µL of Effectene transfection reagent (QIAGEN) or 3.5 µL of LipofectAMINE 2000 (Invitrogen).
5. Incubate for 10 min at room temperature.
6. Add 7.25 µL of 0.08% gelatin (G-9391, Sigma) containing 3.5×10^{-4}% fibronectin (Sigma).
7. Spin down the 384-well plate for 1 min at 216*g*. The solution is now ready for the spotting process.

3.2. Spotting the Transfection Solution on Uncoated Lab-Tek Dishes

As a substrate, we use untreated chambered coverglass dishes, allowing live cell imaging and easy-to-perform immunostaining. We use a ChipWriter Compact robot (Bio-Rad, Hercules, CA) equipped with PTS 600 (Point Technologies) solid pins. The distance between the spots is 1125 µm, resulting in spot diameters of approx 450 µm. The spotted solutions on the Lab-Tek chamber are dried at room temperature for at least 12 h after printing before cells are plated. The solutions of one 384-well plate (*see* **Subheading 3.1.**) are sufficient to spot at least 50 identical Lab-Tek chambers, which can be stored for later use in a dry environment for several months without obvious loss in transfection efficiencies. The outer dimensions of a Lab-Tek chamber are 57.2 × 25.6 mm (length × width). Of this area we use, 34.875 mm in length and 12.375 mm in width to spot 384 samples. Spotting 384 samples on 48 Lab-Tek chambers, by using eight solid pins in the *x*-direction mode typically lasts 6 h.

1. We adjust the temperature of the 384-well plate with an in-house–built water-cooled plate to 12°C.

2. The dwell time the spotter pins remain in the source plate is set to 0.5 s.
3. The dwell time the spotter pins remain on the Lab-Tek chamber is set to 0.1 s.
4. The washing procedures between the spotting of distinct samples are set as follows:
   a. Wash container: 10 s.
   b. Sonication container: 10 s.
   c. Vacuum drying: 10 s.

3.3. Plating of Cells on Lab-Tek Dishes With Dried Transfection Solution

The density of the cells plated on the dried spots of a Lab-Tek chamber is always a compromise between the improved statistics that can be achieved with high cell densities and the quality of the microscopic analyses that is best at low cell densities.

1. Typically, we plate 1.25×10^5 actively growing HeLa, MCF7, COS7L, or HEK293 cells in 2.5 mL of culture medium (Dulbecco's modified Eagle's medium containing 10% heat-inactivated fetal calf serum, 2 m*M* glutamine, 100 U/ml penicillin, and 100 µg/mL streptomycin) on the dried spots of one Lab-Tek culture dish. This process results in approx 100 to 200 cells residing on one spot.
2. The incubation time (at 37°C and 5% CO_2) for the successful expression of plasmid cDNAs varies between 12 and 24 h.
3. The incubation time for RNA interference (RNAi) experiments varies between 20 to 50 h and strongly depends on the stability of the proteins targeted by the siRNAs spotted. For long-term experiments lasting several days, e.g., RNAi experiments targeting stable proteins, the cell density needs to be lowered compared with the density typically used (*see* **step 1**), because cells may stop growing because of contact inhibition at later time-points of the experiment (e.g., 60 h), which makes experiments addressing cell cycle- or signal transduction-related questions difficult to interpret.

3.4. Preparation of Samples for Functional Analysis

For functional analysis involving high-content screening microscopy, we frequently use immunofluorescence ***(6)*** to monitor molecule-specific morphological and biochemical parameters. The immunostaining procedure in Lab-Tek chambered glass cover slips is very effective and cheap because it can be performed with the same antibody for hundreds of target genes in parallel.

1. For the staining of Lab-Tek chambers, we apply 250 µL of the corresponding-antibody by carefully distributing the fluid over the spotted area.
2. Incubate for 10 min with the lid closed to avoid rapid evaporation.
3. Wash two times with 2 mL of phosphate-buffered saline (PBS) (30 min each).
4. Stain cell nuclei with 0.2 µg/mL Hoechst B2261 (Sigma). The strong nuclear staining achieved in this way facilitates automated focusing during image acquisition ***(3)*** (*see* **Subheading 3.5.**).

5. The stained samples are stored at 4°C either embedded in Mowiol or in PBS solution containing 0.1% azide after a brief poststaining fixation of the samples with paraformaldehyde for 2 min.

3.5. Analysis of Samples by High-Content Screening

In principle, images of the cells on the spots can be acquired with any commercially available inverted microscope. We use a ScanR (Olympus Biosystems, Munich, Germany; ***[3]***) scanning microscope, with automated focus, allowing time-lapse data acquisition. The microscope is equipped with standard filters to detect DAPI, GFP, and Cy3. A 10 × /0.4 air or a 40 × /0.95 air PlanApo objective (Olympus) is used for image acquisition.

A key step of the whole image acquisition process is to find the first spot of the array. We use the fluorescent signal of the cotransfected Cy3 DNA oligonucleotide. In addition, we mark the first spot manually with a thin and water-resistant black marker pen on the opposite side of the coverglass where the spots are located, before cell seeding.

1. Assign the spot-to-spot distance, the number of samples, and the array dimensions (number of samples in *x*-direction and number of samples in *y*-direction).
2. For image processing the images are background corrected by subtracting the average pixel value in a blank region of the image. The locations of the cells are determined by Hoechst nuclear stain. A possible fluorescence statistics can be based on the mean fluorescence values in a dilated region around the nucleus (including or not including nucleus, depending on the assay). Average values obtained should be normalized to the siRNA control (1.0) (nonsilencing) spotted on the same slide.

4. Notes

1. The spotting robot used has to be able to pass the walls of the Lab-Tek chamber.
2. To achieve the required accuracy of the spot-to-spot distance, high-resolution spotting robots are necessary.

References

1. Ziauddin, J. and Sabatini, D. M. (2001) Microarrays of cells expressing defined cDNAs. *Nature* **411,** 107–110.
2. Erfle, H., Simpson, J. C., Bastiaens, P. I., and Pepperkok, R. (2004) siRNA cell arrays for high-content screening microscopy. *BioTechniques* **37,** 454–462.
3. Liebel, U., Starkuviene, V., Erfle, H., et al. (2003) A microscope-based screening platform for large scale functional analysis in intact cells. *FEBS Lett.* **554,** 394–398.
4. Starkuviene, V., Liebel, U., Simpson, J. C., et al. (2004) High-content screening microscopy identifies novel proteins with a putative role in secretory membrane traffic. *Genome Res.* **14,** 1948–1956.

5. Simpson, J. C., Wellenreuther, R., Poustka, A., Pepperkok, R., and Wiemann, S. (2000) Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* **1,** 287–292.
6. Pepperkok, R., Girod, A., Simpson, J. C., and Rietdorf, J. (2000) Imunofluorescence microscopy, in Shepherd, P. and Dean, C., eds., *Monoclonal Antibodies: A Practical Approach.* Oxford University Press, New York, pp. 355–370.

10

Transgenic Animal Models in Biomedical Research

Louis-Marie Houdebine

Summary

Transgenic animals have become a key tool in functional genomics to generate models for human diseases and validate new drugs. Transgenesis includes the addition of foreign genetic information to animals and specific inhibition of endogenous gene expression. Recently, animal models provided novel insight and significantly improved our understanding of the initiation and perpetuation of human diseases. Moreover, they are an invaluable tool for target discovery, validation, and production of therapeutic proteins. However, despite the generation of several transgenic and knockout models, obtaining relevant models still faces several theoretical and technical challenges. Indeed, genes of interest are not always available and gene addition or inactivation sometimes does not allow clear conclusions because of the intrinsic complexity of living organisms or the redundancy of some metabolic pathways. In addition to homologous recombination, endogenous gene expression can be specifically inhibited using several mechanisms such as RNA interference. Here, some animal models are described to illustrate their importance in biomedical research. Moreover, guidelines for generation of these animals are presented.

Key Words: Diseases; models; pharmaceuticals; transgenic animals.

1. Introduction

One of the main aims of modern biology is to identify and characterize the function of vertebrate genes. Approximately 30 yr ago, it became possible to clone and investigate the function of most abundantly expressed genes. The discovery of microsatellite sequences and the complete sequencing of the genome of some species, namely, mice, have facilitated the identification of genes involved in human diseases. For example, correlations between the sequence of microsatellites and other genetic markers in the same family members with cystic fibrosis led to the identification of the cystic fibrosis transmembrane conductance regulator (CFTR) gene and mutations that are causing the disease in question. Moreover,

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols*
Volume I, Emerging Strategies for Targets and Biomarker Discovery
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

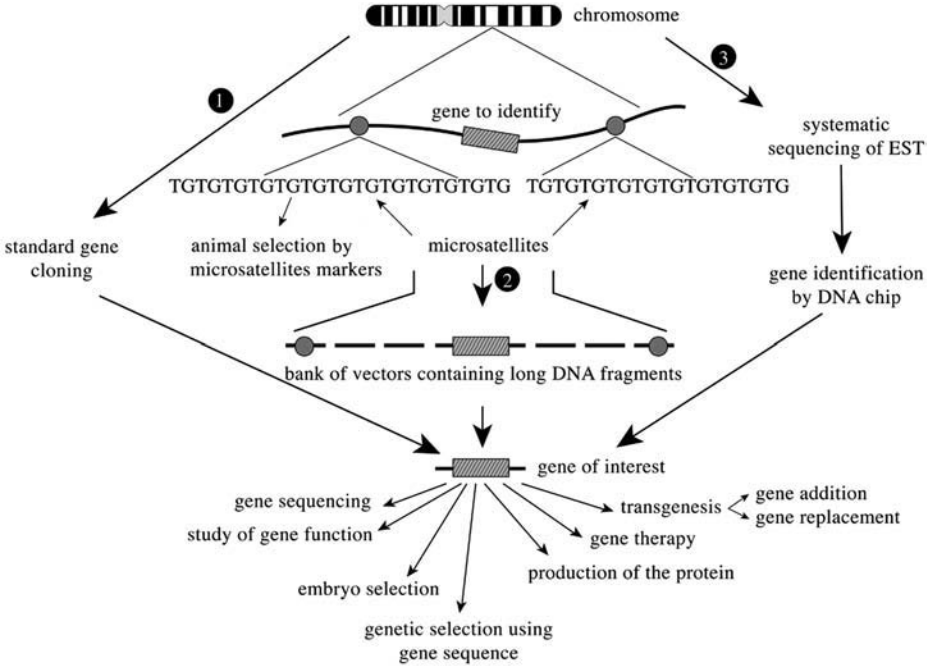


Fig. 1. Different methods to isolate genes and the contribution of transgenesis in gene study and use.

forwards genetics using the chemical mutagen ethylnitrosourea to produce random mouse mutations identified several genes, some of which have been correlated with the obtained phenotype. Then, the mutated genes were identified and characterized ***(1)***. Recently, the systematic sequencing of genomes and expressed sequence tags has facilitated the identification of most genes **(Fig. 1)**. However, the functions of most of the predicted human genes are not yet known.

Once a gene is identified, it can be used for several applications **(Fig. 1)**: for example, discovery of alleles and mutants, study of gene expression mechanisms, embryo selection in humans, animal selection, production of the corresponding protein, gene therapy, and biotechnological applications implementing transgenesis (*see* Chapter 11, Volume 1). These postgenomic approaches, which reconcile physiology and molecular biology, are one of the major challenges for biologists. Transgenesis is one of the major tools of postgenomics. This technique is being used to produce recombinant protein in milk ***(2)*** to adapt pig organs to human recipients ***(3)***, and to improve animal production ***(4)***. Yet, at least 95% of transgenic animals are being generated to study gene function and mechanisms as well as to create models for the study of human diseases and the validation of new drugs.

The first transgenic animals, mice, were obtained in 1980 by using DNA microinjection into embryo pronuclei. Two years later, it was shown with the

giant mice expressing the rat growth hormone gene that a transgene not only can be expressed but also can have a phenotypic effect. The same microinjection technique was adapted to generate transgenic rabbits, sheep, and pigs (1985) as well as fish (1986), rats (1990), and cows (1991). The microinjection technique remains laborious, but it is still the most commonly used technique. Another major problem of transgenesis is the expression of transgenes. Generally, experiments reach their goal by microinjecting poorly sophisticated vectors. Indeed, the generation of some mice expressing no more than 10^4 or 10^5 copies per cell of the protein encoded by the transgene is generally possible if five to 10 lines of animals are prepared. This level of expression is usually sufficient to obtain the expected cellular effect. However, it is not sufficient for the preparation of large amounts of recombinant proteins, such those expressed in milk. The generation of transgenic farm animals is also laborious and costly. Here, it is preferable to use efficient expression vectors to reduce the number of transgenic founders. The use of reliable vectors is also essential to create relevant transgenic models. Substantial progress has been made in the past decade to design expression vectors to express transgenes in a satisfactory manner.

Notably, adding foreign genetic information is an important issue and may provide crucial information about the function of the added gene in whole animals. Inhibiting an endogenous gene expression also can provide useful information about gene function. Using different techniques, several transgenic and knockout animal models are being generated and used for various applications such as target validation.

2. Generation of Transgenic Animals

2.1. Random DNA Integration

Several animal species are used to generate transgenic models: mice, rats, rabbits, pigs, and more exceptionally sheep, cows, and monkeys. In addition to these animals, *Caenorabditis elgans*, *Drosophila*, *Xenopus*, and the fish medaka are taken as models. One of the major advantages of these species is that a large number of mutants are available. Gene transfer cannot be achieved with the same techniques for these different species ***(5)***.

2.1.1. DNA Microinjection

The microinjection of a linear DNA fragment into a pronucleus of a one-cell embryo, which was successfully used for the first time in mice in 1980, is still the most frequently used technique ***(6)***. Its efficiency is limited and lower for rabbits, rats, and pigs than for mice. It is very low in ruminants. This difference does not seem to be because of microinjection problems but instead to the natural capacity of the different genomes to integrate foreign DNA. The optimal concentration of DNA is 1 to 2 ng/µL for injection into mouse embryos. Higher

concentrations become toxic. In rabbits and pigs, the transgenesis yield is increased but not linearly between 0, 5, and 7 to 8 ng/µL. These concentrations correspond to approx 1000–5000 copies of the genes. When long DNA fragments are used, the same amount of DNA as for short fragments must be injected irrespective of the copy number. In ruminants, two reasons are convergent to reduce microinjection efficiency: (1) reproduction is slow with a limited number of available embryos, and (2) integration of foreign DNA occurs less frequently than in other mammals.

DNA microinjection in cytoplasm is used in some nonmammalian species in which pronuclei are not visible. Very high amount of DNA (1–20 million copies in up to 20 nL) must be injected for success. For unknown reasons, this method does not lead to an integration of foreign DNA with a significant yield in several species, such as chicken, *Xenopus,* and medaka. The purity of DNA to be microinjected must be as high as possible. Usually, fragments from plasmids or bacterial artificial chromosome (BAC) vectors have to be isolated using agarose gene electrophoresis. It is recommended to not add ethidium bromide to DNA. The band of interest may be visualized in a separate and independent lane of the gel, by using ethidium bromide. It is also preferable not to illuminate the gel with UV light, which breaks DNA. Several kits including columns to isolate the DNA fragments from the gel are available. They generally give satisfactory results.

When long DNA fragments have to be isolated from BAC, electrophoresis must be followed by a digestion of agarose with agarase ***(7)***. The activity of commercial agarase may be variable and leads to incomplete agarose digestion. Batches of agarase must preferably be tested before use. The fraction must be dialyzed on a filter to eliminate agarose digestion products and to add polyamines (spermidine and spermine) to stabilize the long DNA fragments.

For long DNA fragments, but also for plasmid fragments, electroelution from the agarose gels may be used. The eluted material is trapped in a filter and finally eluted from the filter. Electrocution is simple, and it gives concentrated and highly purified DNA fragments. The concentration of DNA must preferably be evaluated by using both a spectrophotometer and an agarose gel electrophoresis. Isolated plasmid fragments may be kept frozen for months before microinjection. Long DNA fragments must be kept at 4°C and used within a few weeks. Integrity of long DNA fragments must be checked after 2 or 3 wk of storage. Any fraction that contains impurities must be eliminated and replaced by a new preparation.

2.1.2. DNA Transfer by Using Electroporation

Electroporation of medaka embryos generated a large number of transgenic animals ***(8)***. Many groups have not adopted this method, and its efficiency has not been proved in other species.

2.1.3. Use of Transposons

Transposons are mobile DNA fragments that integrate genomes with high efficiency. The gene of interest can replace the transposase gene of the transposons. This replacement prevents the recombinant transposons from being disseminated in an uncontrolled manner. This implies that transposase or a gene coding for this enzyme is coinjected with the transposon to allow its integration. Transposon P is currently used to generate transgenic *Drosophila.* Different transposons have been designed to be able to transfer genes in a variety of species. The transposon Mariner proved efficient to generate transgenic chickens and medaka. The transposon Sleeping Beauty works in mammals as in medaka ***(9)***. In silkworm, a specific transposon, Piggy Bac, is being used successfully ***(10)***. These transposons are not significantly complemented by endogenous transposons, and they do not disseminate in host genomes. One of the limitations of this technique is that transposons can harbor only 2 or 3 kilobases (kb) of foreign DNA and that cellular mechanisms recognize these foreign sequences and may silence the transgenes. Research for improving the use of transposons are underway ***(11,12)***.

2.1.4. Use of Viral Vectors

Retroviral vectors have been extensively studied to become a tool for gene therapy in humans. However, the success with this tool is limited. Yet, retroviral vectors allowed several immunodeficient children to leave their protective bubbles. The efficiency of retroviral vectors has been greatly improved by using lentiviral genomes and vesicular somatitis virus (VSV) envelope ***(13)***. Lentiviral vectors have the advantage of transferring their DNA to the nucleus at any stage of the cell cycle. The VSV envelope recognizes any animal cell type, including oocytes and embryos with high efficiency, and it allows the preparation of concentrated viral vectors **(Fig. 2)**.

Retroviral vectors were originally used to generate transgenic chickens. Transgenic monkeys also were obtained with conventional retroviral vectors ***(14)***. Lentiviral vectors have greatly improved this tool. Transgenic chickens ***(15)***, pigs ***(16)***, sheep ***(17)***, and cows ***(18)*** have been generated in this way much more easily than with DNA microinjection or with conventional retroviral vectors ***(19,20)***. For unknown reasons, attempts to generate transgene monkeys have not been successful so far ***(21)***. Commercial kits are available to generate efficient and safe recombinant lentiviral particles, but care must be taken to manipulate the recombinant particles, especially when the foreign genes are oncogenes or are involved in cell multiplication, apoptosis, or signal transduction to cells. Interestingly, foreign genes transferred by lentiviral vectors are generally well expressed, as opposed to those added to conventional retroviral vectors. This

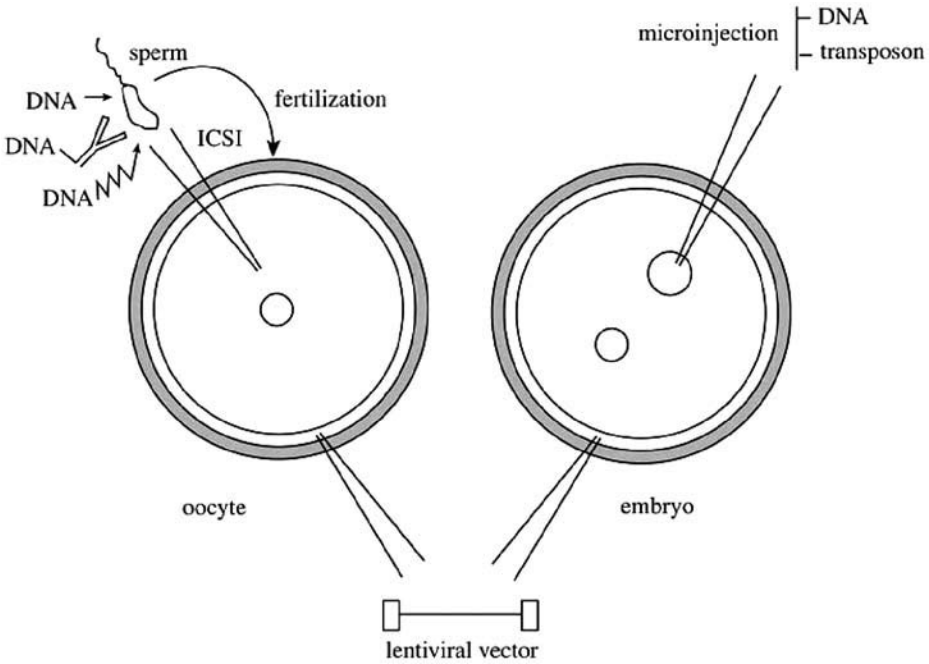


Fig. 2. Major methods to introduce genes into oocytes and embryos.

high expression may be because lentiviral vectors integrate preferentially in genome regions containing genes and often within transcribed regions. It remains that lentiviral vectors cannot harbor more than 8 or 9 kb of foreign DNA and that the regulatory regions present in the viral long terminal repeat (LTRs) may interfere with the promoters associated to the genes of interest. This interference may reduce the cell specificity of transgene expression. Sophisticated vectors inactivate LTR enhancers, but their manipulation is more complex.

In some circumstances, it may be interesting to obtain a transient expression of foreign genes in the whole cells of the organism or in a few cell types only. This end can be achieved by using adenoviral vectors. These vectors are extensively studied for gene therapy in humans. Methods to introduce foreign genes in the adenoviral genome have been standardized. Highly efficiency and safe vectors can be prepared. They remain stable for a few days without being integrated into host genome. Transient expression experiments should determine whether a particular gene has an effect. In a positive effect, classical transgenic animals may be generated to study the function of the transgene. This approach may enhance the chance of creating relevant models. Adenoviral vectors were used to express human growth hormone gene transiently and at a high rate in goat milk ***(22)***. In this case, the method seems suitable to predict whether

a recombinant protein of pharmaceutical interest will be correctly posttranscriptionally modified in transgenic animals.

2.1.5. Use of Episomal Vectors

To avoid position effect of chromatin or transgenes, one theoretical possibility consists of using vectors, which remain autonomous as plasmids or minichromosomes. Several viruses, namely, herpes viruses, are maintained for long periods in organisms as independent genomes. Chromosome fragments also can replicate autonomously and be stably maintained. These natural mechanisms are being used to generate autonomous vectors. Vectors containing the origin of replication and Epstein–Barr nuclear antigen-1 gene from Eptein–Barr virus are spontaneously maintained in human cells. Interestingly, plasmids containing a matrix attached region (MAR) can autoreplicate in a number of animal cell types ***(23)***. It is not known whether this very flexible vector can be used to generate transgenic animals. Chromosomes fragments containing foreign genes also can be transmitted to daughter cells in vitro ***(24)*** and even maintained in transgenic mice ***(25)***. Additional studies are required before these tools can be used easily to generate transgenic animals.

2.1.6. Use of Male Gametes

Experiments carried out more than one decade ago showed that mouse sperm incubated with DNA could generate transgenic animals after in vivo or in vitro fertilization. This technique proved poorly reproducible in part because sperm contain DNAse, which degrades the foreign DNA **(Fig. 2)**. The technique has been greatly improved, and it allowed the generation of transgenic pigs with a high yield ***(26,27)*** and also of transgenic sheep. This approach, which allowed gene transfer in rabbit embryos ***(28,29)***, is attractive only for animals in which microinjection is difficult, inefficient, or costly. A new method was defined several years ago to generate transgenic *Xenopus*. Indeed, DNA microinjection leads to the maintenance of DNA for some stages during embryo development but not to its integration. This method consists of degrading the membrane of isolated sperm to facilitate DNA uptake, incubating this damaged sperm in the presence of DNA, and fertilizing oocytes by using intracytoplasmic sperm injection (ICSI) ***(30)***. This method is currently used in *Xenopus*, and it has been extended to mice ***(31,32)***. Interestingly, it was recently shown that long DNA fragments from BAC or yeast artificial chromosome (YAC) vectors could be efficiently transferred to mouse oocytes ***(32)***. For unknown reasons, ICSI did not allow the generation of transgenic monkeys ***(33)***.

Transgenesis in *Xenopus* has been recently improved after an interesting observation originally done in medaka. DNA fragments released from plasmids by the action of the meganuclease I-Sce 1 (I-Sce) showed a much higher capacity to

integrate into the medaka genome ***(34)***. The mechanism of action of I-Sce 1 is not known. It is postulated that the enzyme remains bound to the DNA fragment, allowing its protection against exonucleases and its unspecific targeting to host DNA. This approach proved successful in *Xenopus* but not in mice so far. To improve DNA binding to sperm, a group recently used a monoclonal capable of recognizing a surface antigen present in sperm of a number of vertebrates. This antibody contains a basic amino acid-rich region in its C-terminal end, allowing a spontaneous binding of DNA and its transfer to oocytes during in vitro or in vivo fertilization (***[35,36]***; **Fig. 2**). Direct DNA transfection into sperm precursors in seminiferous tubules can be achieved in mice ***(37)***. Alternatively, sperm precursors may be collected, transfected in vitro, and reimplanted into recipient testis ***(38–40)***. These methods are promising but require additional studies before being used.

2.1.7. Use of Pluripotent or Somatic Cells

Foreign DNA can be transfected into pluripotent cells or somatic cells further used to generate embryos harboring the transgene. Pluripotent cells, also known as embryonic stem (ES) cells, can be introduced in developing embryos at the morula or the blastocytes stage and give birth to transgenic animals capable of transmitting their transgenes to progeny. In practice, this end can be achieved with only two mouse strains from pluripotent cells that can be cultured. Recent studies showed that ES cells also could be obtained in medaka. Animal cloning by nuclear transfer was rapidly implemented to generate transgenic sheep ***(41)*** and cows ***(42)***. Although laborious, the method is more efficient than DNA microinjection in ruminant pronuclei, and it has not been adopted for other animals such as rats, rabbits, and pigs. The use of cloning in these species is therefore restricted to targeted gene transfer.

2.2. Targeted DNA Integration

For several reasons, it is important to target integration of foreign genes. This approach may avoid uncontrolled position effect of chromatin on transgene expression. This approach is also the major method to specifically inhibit the expression of a cellular gene, also known as gene knockout.

2.2.1. Gene Knockout

In all species, although very weakly in plants, a foreign DNA fragment can replace a genomic DNA by a homologous recombination mechanism. This replacement implies that relatively long regions of the endogenous and the foreign DNA have identical sequences **(Fig. 3)**. Homologous recombination is a rare event, and it cannot be achieved at a sufficient rate in one-cell embryos after microinjection of the gene constructs. Gene replacement must therefore be

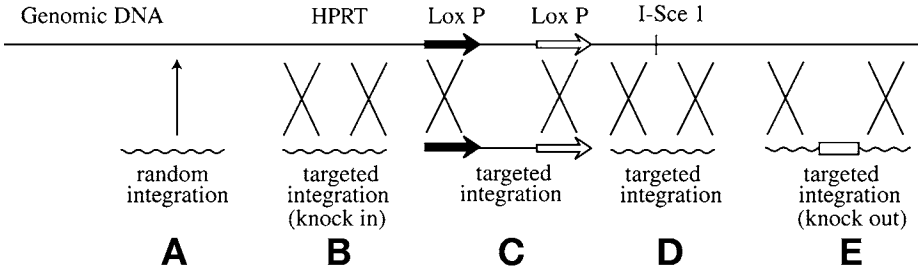


Fig. 3. Different methods to integrate foreign genes randomly or in a targeted manner. **(A)** Random integration after DNA microinjection, **(B)** targeted integration in the HPRT locus, **(C)** targeted integration in LoxP sites, **(D)** enhanced targeted integration by genomic DNA cleavage I-Sce1 site, and **(E)** classical targeted integration.

carried out in cells, which can participate to embryo development, including the formation of gametes. Pluripotent ES cells are currently used for this purpose. In two lines of mice, pluripotent cell lines keep their capacity to transfer their genetic modification to progeny. For this purpose, these ES cells are selected in vitro and implanted into a morula or a blastocyst ***(43)***. This approach is presently possible only in two lines of mice and in medaka. The other lines and species allow the generation of chimera after ES cell implantation but no transmission of the mutation to progeny. Although laborious, this technique allows the knockout of approx 5000 mouse genes. ES cells in which the knockout of essentially all the mouse genes has been achieved are available, allowing the potential inhibition of all the genes in the animals.

2.2.2. Use of Meganucleases

The yield of homologous recombination can be greatly increased by inducing the cleavage of the double strand of genomic DNA at the targeted site. This increase can be achieved by introducing the site of a meganuclease like I-Sce 1 **(Fig. 3)**. This double cleavage is a strong stimulation of the reparation system, which induces the specific gene replacement ***(44)***. This induction implies that an I-Sce 1 has been previously introduced in the animal, either randomly by DNA microinjection or in a targeted manner by homologous recombination. This technique allows the introduction of foreign genes always at the same site. This phenomenon may not suppress the position effect on the transgene, but the phenomenon is the same for each gene transfer. Recent studies have shown that some meganucleases may be engineered to recognize specific sites in genomes rather than their natural restriction sites. This recognition offers the interesting possibility to target foreign genes in a variety of sites in the genome. This promising approach might facilitate gene therapy and favor the generation of relevant transgenic animal models ***(45)***.

2.2.3. Knock-In Into Hypoxanthine–Guanosine Phosphoribosyl Transferase (HPRT) Locus

Targeted DNA integration is mainly used to knock out genes. It may be implemented as well to introduce an active gene in a given site of a genome. This operation known as gene knock-in allows reproducible expression of transgenes always subjected to the same position effect. The HPRT locus is one the sites in which foreign genes are successfully targeted in mice **(Fig. 3)**. This gene is expressed in all cell types. Interestingly, the HPRT locus allows not only a satisfactory expression of foreign genes designed to be expressed in all cell types but also proves to direct the expression in animals of a variety of gene constructs containing cell specific promoters ***(46)***. A company is extensively using this technique to express successfully foreign genes for academic or industrial laboratories.

2.2.4. Use of Cre-Lox P and Flp-FRT Systems

Two systems allow the integration of foreign genes into selected sites of genomes. One system known as Cre-LoxP is of bacterial origin and the other system, Flp-FRT, comes from yeast. In these two systems, the short LoxP and FRT DNA sequences are first introduced into a genome, either randomly or in a targeted manner. Foreign genes containing either LoxP or FRT sequences can integrate specifically in the homologous sites previously introduced into genomes under the action of the specific recombinases Cre and Flp, respectively. The LoxP and FRT sites in the genome may be retained after the demonstration that they favor transgene expression. Alternatively, the two LoxP and FRT sites may have been introduced previously in vectors containing elements able to favor transgene expression such as insulators.

These two systems require the use of two different sites added on both sides of the foreign gene. This technique known as recombinase-mediated cassette exchange allows integration of the foreign gene into the genome but not its deletion, which spontaneously occurs with a single site (**Fig. 3**; ***[47,48]***).

Notably, the Cre-LoxP and the Flp-FRT systems allow the conditional deletion of a gene. For this purpose, two LoxP or FRT sites must be introduced in the periphery of the gene to be deleted in the genome. This approach implies two homologous recombinations. The expression of the Cre or Flp recombinases in given tissues induces the specific knockout of the targeted gene. This knockout can be achieved by crossing mice harboring the LoxP or FRT sites with other mice expressing the recombinase genes in the chosen tissues. Alternatively, adenoviral vectors harboring the recombinase genes may be injected into the tissues in which gene knockout is desired.

The Cre-LoxP and the Flp-FRT systems allow the deletion of long DNA fragments from a genome and chromosome crossover. In the first case, LoxP or

FRT sites are added on the same chromosome at a chosen distance. The action of the recombinases deletes the region between the LoxP or FRT sites. In the second case, the LoxP or FRT sites may be introduced in two different chromosomes. The recombinases induce crossover of the two chromosomes. This operation has not been achieved frequently because the recombination yield is very low. These techniques are summarized in **ref. *49***.

3. Design of Transgene Expression Vectors

In early 1980s, the first transgenesis experiments revealed that human globin gene could be expressed in a satisfactory manner in cultured red blood cells but only very poorly in transgenic mice. It took more than a decade to start understanding the molecular mechanisms of transgene expression. In practice, many of transgenes are not expressed in an appropriate manner. Most of them are subjected to position effects. Sometimes, enhancers located in the vicinity of the transgene alter the cellular specificity of its expression. More frequently, transgenes remain silent. This silence was attributed to the presence of silencers in the vicinity of transgenes. It is now clear that the composition of the transgene itself induces or does not induce its own silencing.

3.1. Vectors to Enhance Frequency and Level of Transgene Expression

The major reason why a transgene is not expressed seems to be because it is poorly transcribed, first suggested approx 15 yr ago. This poor transcription suggested that some distal regulatory elements were absent in gene constructs. The demonstration was given by simultaneous studies carried out in transgene *Drosophila* and mice ***(50)***. In some people suffering from β-thalassemia, their β-globin gene promoter and proximal enhancers are not mutated. However, approx 30 kb upstream of the gene, these patients are lacking a genomic DNA sequence that contains several hypersensitive sites (HSs) to DNAseI. The addition of this genomic sequence to the β-globin gene allowed it to be highly and specifically expressed in transgenic mice. This region known as locus control region (LCR) is located between the cluster of the four globin genes of the β-globin locus and unrelated genes. This finding supported the idea that the LCR acts as a barrier insulating unrelated genes. The β-globin LCR contains several elements. One element containing a CTCF binding site plays the role of enhancer blocker, preventing cross-talk between the enhancers and the promoter of the globin genes and the neighbor genes. Another element seems to be a conventional distal enhancer. A third element binds transcription factors having a histoneacetylase activity. This binding maintains the β-globin locus in an open configuration, which allows the transcription machinery to express the genes of the locus ***(51)***.

Regions having properties more or less similar to β-globin LCR have been found in an increasing number of loci. These regulatory regions are also known

as insulators. The concept of barrier seems not able to account for all the mechanisms of gene regulation. Indeed, essential regulatory elements are present 63 kb upstream of the β-globin gene locus. These elements act by forming loops, which put remote factors in proximity. This looping forms a hub, which concentrates all the regulatory factors needed to express a gene. The loops often contain unrelated genes ***(52)***. The concept of barrier bordered by LCR seems more idealistic than realistic. These observations fit with the long genomic DNA fragments being needed to obtain a highly reproducible, specific, and copy number-dependent expression of transgenes. Sometimes, a few elements were sufficient to improve transgene expression. The 5′HS4 region from the chicken β-globin locus allow genes to be expressed in all cell types ***(53)*** or in specific cells ***(54)***. Recent data have been summarized in a review ***(55)***.

In practice, the addition of two copies of 5′HS4 upstream of a promoter favors expression of a transgene without altering its specificity. In the future, it seems conceivable that many genomic fragments allowing reliable and specific transgene expression will be available. Their use implies the construction of vectors containing long fragments of DNA. This end can be achieved essentially by homologous recombination in bacteria by using BAC vectors ***(56)***. Also, the major regulatory elements present in the long DNA fragments will have to be identified. They will be used to create compact and efficient vectors for transgene expression. MAR is generally composed of AT-rich sequences found in LCRs and insulators. They participate to the insulating effect, but their capacity to stimulate, alone, transgene expression has rarely been observed.

Other factors control transgene expression. These points have been discussed in a previous article ***(57)***. The recommendations to obtain frequently satisfactory expression of transgenes are summarized in **Fig. 4**. The transgene must not contain long GC-rich stretches of DNA ***(58)***. Some promoters that are efficiently used in cultured cells (simian virus 40, cytomegalovirus [CMV], phosphoglycerate kinase [PGK], thymidine kinase [TK], and elongation factor [EF]1α) often do not direct a high expression of transgenes. The transgene is often better expressed when only one copy is integrated. The elimination of extra copies may be achieved using the Cre-LoxP system, but this method is laborious ***(59)***. At least one intron must be present in the vector. Efficient signals for exon splicing, for mRNA transfer from the nucleus to the cytoplasm, and for the stabilization of mRNA must be preferably used. To avoid the nonsense-mediated decay effect, the termination codon of the cDNA or gene must be located no more than 50 nucleotides from the following donor-splicing site. Otherwise, a quality control mechanism destroys the mRNA. Enhancers can be valuably added to the constructs in the promoter region or within the transcribed region. The presence of these enhancers may concentrate transcription factors having histoneacetylase activity, which contributes to maintain chromatin in an open configuration ***(60)***.

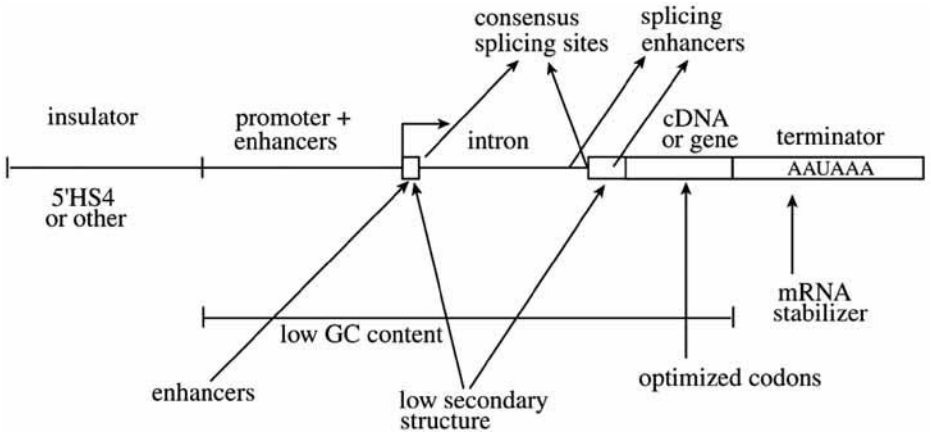


Fig. 4. Different elements to be added in a vector to favor transgene expression.

Adaptation of the codons to the organism in which the transgene is introduced may greatly enhance its expression. The chemical synthesis of cDNAs or genes is an efficient and not too costly way to obtain these optimized coding regions. Moreover, the 5′ untranslated region (UTR) must contain as little as possible of secondary structure.

It is sometimes necessary to generate transgenic animals expressing simultaneously several genes. The simple way consists of coinjecting two independent constructs, which cointegrate in 70–80% of the cases. Alternatively, the two independent constructs may be linked in the same vector. This linkage complicates vector constructions, but it allows a better control of the transgene activity.

Another possibility consists of adding sequences known as internal ribosome entry site (IRES) between two cistrons. It is thought that an IRES recruits ribosomes directly, independently of the cap. A study carried out in our laboratory suggests that this mechanism may be not right for all IRESs ***(61)***. In practice, it must be known that biscistronic mRNAs containing IRES between the two cistrons express, most of the time, less efficiently their corresponding proteins. This lower expression is particularly true for the second cistron. The cistron to be expressed at the lowest level should preferably be put after the IRES. It also was shown that the second cistron is efficiently expressed if the distance between the termination codon of the first cistron and the beginning of the IRES is of approx 80 nucleotides ***(61)***.

3.2. Vectors to Inhibit Specific Gene Expression

The understanding of gene function and the generation of animal models rely in part on the specific inhibition of cellular genes. It is also important to inhibit the expression of viral genes to study their function or to generate lines of animals

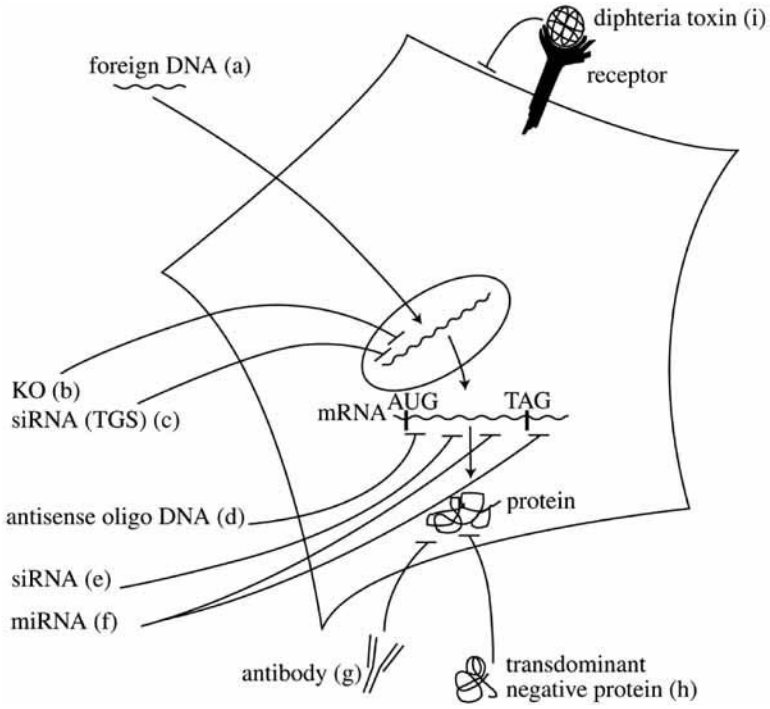


Fig. 5. Different methods to inhibit specifically the expression of a gene. **(A)** Classical gene addition, **(B)** gene knockout, **(C)** gene knockdown by targeted gene silencing (TGS), **(D)** inactivation of a mRNA by hybridizing with an antisense oligonucleotide, **(E)** knockdown by mRNA inactivation by using siRNA, **(F)** knockdown by inhibition of mRNA translation by using microRNA, **(G)** inhibition of a cellular protein by using an intrabody, **(H)** competitive inhibition of a cellular protein by using overexpression of a transdominant negative protein, and **(I)** specific destruction of a given cell type in vivo (genomic ablation).

resistant to diseases. In **Fig. 5**, the different levels in which the expression of a gene can be blocked are shown. The levels are essentially genes themselves, mRNAs, and proteins. Indeed, the expected effect is most of the time finally at the protein level. Genes can be specifically inactivated after having been knocked out. In addition, gene function can be inhibited by several mechanisms, such the use of small-interfering RNAs (siRNAs), and synthetic antisense siRNAs and microRNAs can inactivate specifically mRNAs by inducing their degradation or the inhibition of their translation, respectively (*see* Chapters 4–12, Volume 2). Recombinant intracellular antibodies known as intrabodies can inactivate proteins. An overexpression of transdominant negative proteins playing the role of decoys can abrogate specifically protein function as well. To generate relevant

models, the blocking molecules (siRNAs, mRNAs, transdominant negative proteins) should be encoded by transgenes, which can be regulated.

3.2.1. Roles of Small RNAs

Recent data show that cells contain a large number of small RNAs, which have multiple functions: protein synthesis (tRNA and 5S rRNA), control of chromatin structure, degradation of mRNA, inhibition of translation, control of mRNA splicing, and other actions yet to be discovered ***(62)***.

3.2.2. Use of siRNAs to Inactivate mRNAs

siRNAs were discovered fortuitously a few years ago ***(63)***. Long double-stranded RNAs are cleaved in 21- to 23-base pair (bp) fragments by an enzymatic complex (Dicer). Another enzymatic complex, RNA-induced silencing complex (RISC), targets the small RNAs to complementary mRNAs and induces their degradation ***(64)***. In plants, this mechanism, known as posttranscriptional gene silencing, acts as an antiviral system. Indeed, a number of viruses are in double strand at some steps of their replication cycle. siRNAs are not encoded by genomes. Using siRNAs implies that they are introduced into cells, which may be achieved by adding synthetic siRNA to cells or whole organisms. Alternatively, they can be endogenously expressed from appropriate vectors.

Biosynthesis of long double-stranded RNAs can be easily achieved with conventional vectors by using RNA polymerase II promoters. Long double-stranded RNAs induce interferon secretion, inhibition of protein synthesis, and cell death in higher organisms. Long double-stranded RNAs can thus be used in plants but not in most animals. They are used successfully in early animal embryos because the interferon mechanism is functional at this stage. RNA polymerase III vectors can synthesize small double-stranded RNAs at a high rate in all cell types. Promoters from U6 and H1 genes are being used successfully for this purpose. Synthetic DNA can direct the synthesis of a double-stranded RNA containing a short loop between the two complementary strands and ending in five uridines, acting as a termination signal for RNA polymerase III. Partial degradation of the five Us generates RNAs ending with two overhanging nucleotides required for siRNA action ***(64)***.

Gene constructs containing RNA polymerase III promoter followed by a short sequence such as siRNA are highly expressed in cultured cells, even in stable clones, but they are essentially silent in transgenic animals. The mechanism of action of RNA polymerase III promoters is not fully understood, and some important elements for in vivo expression may be lacking in the vectors currently used. Alternative vectors are being used and studied. However, before generating transgenic animals, it is preferable to determine which siRNAs or microRNAs induce the most intense knockdown of the targeted gene in cell

systems. Moreover, stable clones expressing the siRNA and tested individually should be tested to evaluate properly the siRNA efficacy. Alternatively, transfection of siRNA synthesized in vitro or chemically may be transfected with high efficiency into cells and give a good prediction of the siRNA potency.

The U6- or H1-siRNA constructs are expressed in transgenic mice when they are integrated into lentiviral vectors ***(65)***. A vector designed to be devoid of introns to loose its cap and its polyA tail and containing a RNA polymerase II promoter has been constructed. The RNA synthesized by this vector cannot leave the nucleus. The RNA which can be a long double RNA molecule is degraded in the nucleus and the 21–23 bp fragments migrate to the cytoplasm where they can act as siRNAs. No long double-stranded RNA is in cytoplasm, and the interferon response is not induced. This vector made it possible the generation of mice in which ski gene was knocked down. Mice showed essentially the same biological properties as those obtained after ski gene knockout ***(66)***. A vector containing a murine CMV early gene promoter, no intron, a minimal transcription terminator, and a region coding for a siRNA was able to generate an active siRNA ***(67)***. A vector containing the ecdysone regulatory system proved able to synthesize siRNA in a controlled manner ***(68)***. Vectors based on the use of the tetracycline-dependent system may be used as well. Ideally, others vectors dependent on RNA polymerase II promoters remain to be found. They would offer the most flexible system to control siRNA synthesis in transgenic animals. These vectors must express siRNA at a high level. Indeed, in plants and in lower invertebrates such as *C. elegans*, siRNAs are autoamplified, whereas this mechanism is lost in other species, such as *Drosophila*, and in vertebrates ***(64)***.

Experimental data have shown that many different siRNAs can be designed to target an mRNA, but only some of them seem active. The active siRNA sequence can be searched empirically, by using siRNA libraries ***(69,70)***. Such siRNA screening should be performed with appropriate vectors. The vector shown in **Fig. 6** contains the target sequence followed by IRES and a luciferase reporter gene. The target sequence may or my not be an open reading frame. The IRES enables ribosomes to start translation of the following cistron irrespective of the sequence of the proceeding cistron ***(61)***. This vector can be cotransfected plasmid coding for hairpin siRNAs. The level of luciferase in cells should reflect the inhibitory effect of the designed siRNA. The same type of vector may be used for screening siRNA libraries.

A systematic comparison of the active and inactive siRNAs showed that those siRNAs are the most frequently efficient are rich in GC and rich in AU in the 5′P and 3′OH end, respectively, of the sequence similar to that which is in the target. A U must preferably be present in position 10 of the siRNA ***(71–76)***. Both strands of a siRNA may target mRNAs. The choice of the strand by the RISC

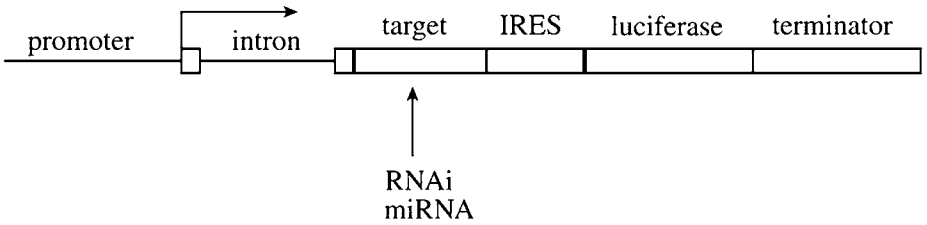


Fig. 6. Vector to identify specific siRNAs. The targeted sequence is introduced before IRES, which is followed by luciferase gene. The siRNA cleaves the target, which suppresses expression of the luciferase gene.

complex depends on the structure and the stability of the siRNA *(77–79)*. Recent studies have shown that the Dicer complex generates 21- to 23-bp siRNA, but that these active RNAs are more efficiently generated when the precursors contains 26 to 27 bp *(76,80,81)*. The structure of efficient siRNAs cannot be completely predicted. Moreover, the efficiency of siRNAs is dependent not only on their structures but also of the local structure of the targeted mRNA *(82)*. siRNAs may generate side effects. Their sequence may recognize several sequences unspecifically *(83)*. The systematic search of possible multiple target sequences in a genome must be preferably achieved to design a specific siRNA. However, in some cases, siRNAs may induce interferon and inflammatory cytokines *(84)*.

Some siRNA may be derived from genomic sequences. Regions of genomes that are transcribed in the sense and antisense orientation are more numerous than imagined thus far. This phenomenon may generate multiple siRNAs having important functions in the control of gene expression *(85,86)*. Transgenes that are integrated in an uncontrolled manner also may be transcribed in both orientations and generate functional siRNAs having unpredictable side effects. Interestingly, recent studies showed that a RNA fragments resulting from intron degradation might generate functional siRNAs and microRNAs *(87,88)*. This finding indicates that single- or double-stranded sequences targeting an mRNA may be added to introns of RNA polymerase II-dependent vectors. This approach offers additional possibilities to use siRNA in transgenic models. Finally, it should be mentioned that siRNA could be generated from genomic microRNAs.

3.2.3. Use of siRNA to Inhibit Transcription

Studies carried in yeast, plants, and vertebrates indicate that siRNAs induce targeted modifications of histones, leading to a local DNA methylation, an inactivation of the promoters, and silencing of the corresponding genes. This phenomenon known as transcriptional gene silencing (TGS) also has been observed in human cells *(89,90)*. Interestingly, these epigenetic mechanisms, which

contribute to generate heterochromatin and to inactivate transposons and retrovirus integrated into genomes, induce a gene silencing that is transmitted to daughter cells. TGS should therefore be a potent way to silence endogenous or viral genes in transgenic models.

3.2.4. Use of MicroRNAs

Plant and animal genomes contain several hundred genes coding for 120-bp RNAs known as microRNAs. These RNAs are processed by an enzymatic complex (Drosha). The resulting short-hairpin RNAs (shRNAs) are processed by Dicer and act as siRNAs to degrade the targeted mRNAs or as potent translation inhibitors if they recognize sequences in the 3′ UTR of mRNAs **(Fig. 5)**. It should be noted that these small RNAs act as siRNA if they perfectly match with the targeted mRNAs and as translation inhibitors if they form several mismatches with their target ***(64,91)***. The microRNAs play an essential role in the control of gene expression especially during development ***(92,93)***. They are transcribed by RNA polymerase II-dependent promoters ***(94)***, and they can be engineered to express recombinant small RNAs activity as siRNAs or microRNAs ***(95)*.** These observations offer many possibilities to generate transgenic models expressing siRNAs or microRNAs. Interestingly, natural microRNAs are sometimes present in the same transcription unit. A single RNA polymerase II-dependent vector may thus express several microRNAs targeting different mRNAs. It is interesting to note that several companies are providing researchers with chemical compounds or kits optimized to synthesize siRNA and to transfect them into cells ***(96,97)***.

3.3. Inhibition of Gene Expression at the Protein Level

Antibodies theoretically offer many possibilities to inhibit protein action. Monoclonal antibodies can be expressed in transgenic animals to create lines resistant to pathogens ***(98)***, and intrabodies can be expressed in transgenic animals. Their targeting in cell compartments may be necessary to optimize their action ***(99)***. This approach, although attractive, has not been often used. The overexpression of a transdominant negative protein acting as a decoy or as a competitive inhibitor is a natural mechanism of protein action. This concept has been used in some cases to create models or to generate animals resistant to diseases. Overexpressed mutated insulin receptor traps the hormone, which cannot bind anymore to the wild receptor. This binding allows the generation of transgenic mice suffering from type II diabetes ***(100)***. Similarly, transgenic mice overexpressing the soluble part of the receptor for the Aujeszky disease virus are protected against the infection ***(101)***. Transgenic pigs could be protected as well, and this protection could reduce the frequency of the disease. The use of transdominant negative proteins may be easy and efficient. Conventional

vectors can direct synthesis of these of these proteins. The major limit is the availability of the proteins having a potent transdominant negative action without any side effect.

3.4 Genetic Ablation

Inducing the specific destruction of a cell type in a given tissue at a well defined step of development may give interesting information on the function of these cells. This technique, known as genetic ablation, has been used for more than 15 yr, although not at a high frequency. The limit of this technique is to obtain a potent and specific destruction of the targeted cells. Genes coding for toxins are being used for this purpose. Ideally, the action of the cytotoxin should be induced in a precise manner. Two models can exemplify this approach. One aims at destroying chondrocytes at different stages of development. For this purpose, a first line of mice harbored a transgene containing PGK promoter active in all cell types associated to a mutated version of TK gene no longer able to induce infertility in mice. A neo^r gene and a transcription terminator separated the PGK promoter and the TK gene. The TK gene was not expressed in these mice. The first line of mice was crossed with another line expressing the Cre recombinase specifically in chondrocytes. The neo^r gene was deleted in chondrocytes of hybrid mice. Fialuridine (FIAU), a drug that is transformed in a cytotoxic compound in cells expressing the TK gene, was injected. The injection of FIAU induced chondrocyte destruction at different steps of development ***(102)***. Another model is based on the use of diphtheria toxin. This protein is extremely potent and its presence in targeted cells must be strictly restricted to the period of genetic ablation. Transgenic mice expressing specifically the toxin receptor in a chosen cell type were generated. Injection of the toxin at any step of the life of the animals induced the destruction of the cells ***(103)***.

3.5. Vectors for Conditional Transgene Expression

Ideally, in a number of situations, inducers not acting on host genes should selectively and reversibly activate transgenes. This process avoids that a possibility that a natural transgene inducer stimulates or inhibits simultaneously some host genes, lowering the relevance of the model.

Several systems allowing specific activation of transgenes are available. They are all based on the use of chimeric transcription factors containing one region sensitive to the inducer and another region able to stimulate the associated gene. The most frequently used system is based on the use of tetracycline and derivatives as inducers and tetracycline repressor fused to an animal transcription factor ***(104,105)***. These systems use murine promoters modulated by the fusion transcription factor. One of the limits of these systems is that the promoter

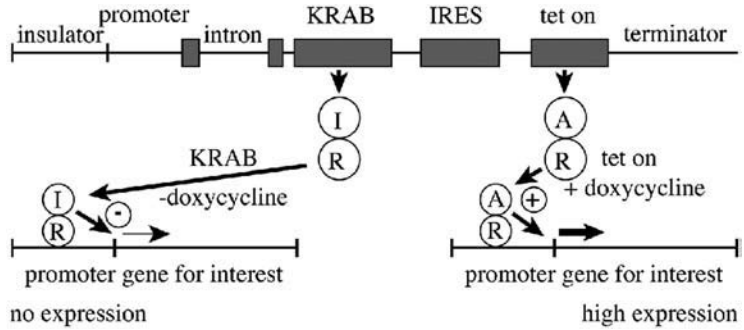


Fig.7. Example of a vector expressing a foreign gene under the control of tetracycline or derivatives with a low background of expression.

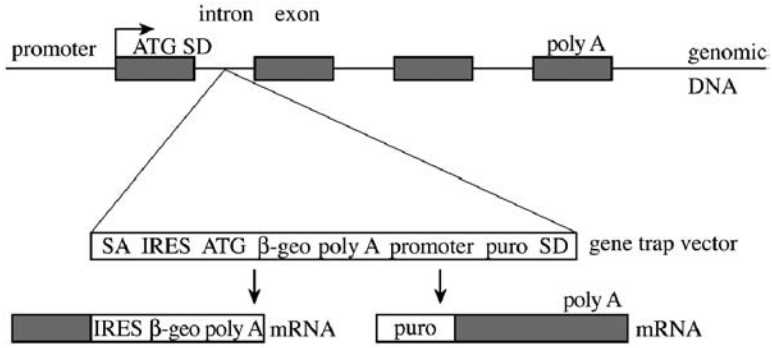


Fig. 8. Vector for gene trapping. The random integration of the vector is performed in ES cells and further used to develop embryos. The expression of β-galactosidase-neo^r genes reveals that the vector was integrated into a gene. The inhibition of the interrupted gene may be correlated with a phenotypic effect.

retains a slight basal activity in the absence of the inducer. To reduce the background expression, factors acting as enhancers and silencers can be alternatively expressed under the action of an inducer such as tetracycline. Such a system constructed in our laboratory is described in **Fig. 7**. Transgene expression also can be controlled at the translation level ***(106)***.

3.6. Vectors for Gene Trap

The systematic identification of genes by genome sequencing and the establishment of gene expression pattern by DNA chips may be insufficient to determine the role of a gene in a given biological function. Gene trapping is currently being used to identify genes of interest. For this purpose, vectors containing a marker gene, an intron splicing site, and a transcription terminator, but devoid of promoter, are introduced by electroporation in ES cells **(Fig. 8)**. The cells are

then used to generate chimeric mice. The expression of the marker gene (often a hybrid containing β-galactosidase and neomycin-resistant genes) indicates that the vector was randomly integrated within a gene. This gene can easily be identified. Interesting models can be obtained in this way if a correlation can be established between the expression of the gene trap vector in a given tissue and particular health problems in the animals. A series of gene trap vectors have been designed to identify genes expressed in ES cells or in differentiated tissues, coding for secreted proteins or activated by specific inducers such as hormones or growth factors ***(107–110)***. Banks of ES cells in which gene trap vectors have been introduced are available ***(111)***.

4. Transgenic Models: Case Studies

Very diverse models of transgenic animals are currently generated. Established lines of transgenic animals are available as conventional lines, particularly for transgenic mice used to study some human diseases.

Several databases are available. Information on one database recently launched can be accessed at http://www.biomedels.net/, http://www.ebi.ac.uk/biomedels, and http://research.bmn.com/mkmdj. Animals are available in The Jackson Laboratories (Bar Harbor, ME) and EMMA (European Mouse Mutant Archive) as well as in private companies. The number of models is very large and some of them are described in the present review. They have been chosen to exemplify the potency and the flexibility of transgenesis in this field.

4.1. Preparation of Recombinant Proteins

Recombinant proteins are prepared to study their biological properties and use as pharmaceuticals. Several production systems are being used or are in development: bacteria, yeast, animal cells, and transgenic plants and animals. Each system offers advantages and disadvantages on a case-by-case basis ***(112)***. Milk from transgenic animals is one the most mature systems ***(2)***. Before being transferred to farm animals (rabbits, goats, sheep, pigs, or cows) for the large production of proteins, gene constructs are evaluated in model transgenic mice. This approach indicates whether the construct expresses efficiently the transgene in milk and whether the foreign protein in milk in properly posttranscriptionally modified.

Some recombinant proteins prepared in Chinese hamster ovary (CHO) cells or in milk of transgenic animals are not properly glycosylated, γ-carboxylated, or cleaved to generate mature molecules. This limited glycosylation may be attributable to the low level of maturation enzymes in producing cells. CHO cells overexpressing some glycosylating enzymes secrete more completely glycosylated proteins. Transgenic mice overexpressing the furin gene in their milk can cleave human protein C precursors

more extensively. Transgenic mice are valuable models to define genes to be added to farm animals ***(112)***.

Pharmaceutical recombinant proteins sometimes induce formation of antibodies in patients. This antibody production may take place because the recombinant proteins have slightly different structure, namely, in the carbohydrate content. Transgenic animals and essentially mice expressing a recombinant protein in any of their tissues are models to study the immunogenicity of the recombinant protein. Indeed, the recombinant protein belongs to the self of the animals, which mimics the situation of patients treated by the protein.

4.2. Xenografting

Using animal, essentially pig, organs for humans would save the lives of many patients but will not be possible until the strong rejection mechanisms are not controlled. In the 1990s, an understanding of some of the major rejection mechanisms urged the generation of transgenic animals harboring genes having an anticomplement effect. Mice, rats, and rabbits were used for this purpose before transgenic pigs were prepared ***(113)***. Kidneys from transgenic pigs expressing human DAF and CD59 genes could be kept intact for a few days after having been grafted to experimental monkeys. This finding validated the transgenic approach to study rejection mechanisms and to obtain pig organs to be potentially grafted to humans.

Other transgenic models are regularly obtained by different academic groups and by private companies to study rejection mechanisms. The most advanced project is based on transgenic pigs in which the α-galactosyl transferase gene has been knocked out. This knock out was achieved using the homologous recombination and cloning techniques described above ***(114,115)***. Homozygous pigs are healthy, and their kidneys grafted to monkeys immunosuppressed by classical methods could survive more than 2 mo without any degradation, whereas control organs were destroyed after at most a few days. The expression of nonpathogen retrovirus in pigs must be controlled to prevent any transmission to patients ***(116)***. Interestingly, strains of pigs not expressing the retrovirus have been found ***(117)***.

4.3. Models for Embryo Development and Organogenesis

Transgenesis is extensively used to study development mechanisms. Gene addition, knockout, and knockdown are essential tools to determine gene function in development. Gene coding for proteins giving colors to cells are currently used. β-Galactosidase is one of these genes. The blue color resulting from the enzymatic activity of this protein allows the positioning of cells in which a given promoter is active. β-Galactosidase suffers from several limitations. Cells must

be fixed and are thus dead before proceeding to the enzymatic reaction. This death prevents repeated observation on the same organ. However, the β-galactosidase gene is GC rich and frequently silenced in transgenic animals. A redesigned β-galactosidase gene in which most of the codons rich in GC sequences have been replaced by isocodons is available. It does not induce gene silencing. Genes for green fluorescent protein (GFP) and other related proteins having different colors under UV light are more and more frequently used. They offer the advantage of being easily and repeatedly observed in the same tissues without using invasive techniques if the organ is in the periphery of the body ***(118,119)***. Moreover, the observation does not require cell death.

Interestingly, a device, Cell Vizio, has been created to observe GFP-labeled cells individually in living animals ***(119a)***.

Transgenic rabbits expressing GFP in all their cells ***(120)*** are currently used to follow cells after grafting of organs to other rabbits. Because of its size, the rabbit makes this approach easier than in mice. Essentially, in all the major animal species in which transgenesis has been achieved, GFP animals have been generated to be used as models.

One limitation of GFP may be that this protein becomes cytotoxic at high concentration ***(120)***.

Other specific examples are also worth consideration. Homozygous goat lines without horns are subfertile. It has been shown that this infertility is because of a mutation responsible for an abnormal differentiation of fetal gonads ***(121)***. This abnormality is similar to some human genetic diseases. Mice are not relevant models for this study ***(122)***. Transgenic goats obtained by cloning are currently being used for this purpose (E. Pailhoux, personal communication, 2005).

Several hormone receptor genes have been overexpressed or knocked out to determine their role in tissue growth and differentiation. The mammary gland is composed of two major categories of cells. The parenchyma contains the secretory epithelial and the myoepithelial cells. The stroma, which surrounds the parenchyma, is formed by fibroblasts, adipocytes, and connective tissue. Progesterone is one of the hormones favoring mammary gland growth in vivo and preventing milk secretion during pregnancy until parturition. The progesterone receptor gene is particularly present in the parenchyma. To determine its role, the progesterone receptor gene was knocked out. The development of the mammary gland was impaired in these animals. The grafting of the fetal parenchyma from knocked out mice into the stroma of normal mice did not lead to a development of the mammary gland. Conversely, normal parenchyma grafted into the stroma of knocked out mice developed to become a normal mammary gland. These experiments confirm the role of progesterone in mammary gland development, and they point out the different roles of parenchyma

and stroma. Several other gene modifications in mice have been carried out to decipher the different mechanisms in the development and the differentiation of the mammary gland ***(123)***.

4.4. Models for Cystic Fibrosis

Cystic fibrosis results from mutations of the CFTR gene. However, several hundred mutations have been found in this gene, but only some of them cause a severe pathogenicity. In contrast, it is known that the most potent mutations, namely, the ΔF 508 mutation, have not the same effect in all patients. Thus, the cooperation with other genes may enhance or attenuate the effect of the mutation. Transgenic mice harboring the ΔF 508 mutation in their CFTR gene have been independently generated by several groups. In no single case did the models seem fully relevant. CFTR is a chloride channel that is less active and not properly transported to plasma membrane in mutants. Mice have other chloride channels, and their lungs are only weakly affected by the CFTR gene knockout. Yet, these animals show intestinal disorders typical of cystic fibrosis. Rabbit and sheep are expected to be better models for cystic fibrosis because their CFTR genes are more similar to the human CFTR gene. Experiments are underway to generate rabbits and sheep with the mutated CFTR gene, but the major hurdle is that the cloning technique by nuclear transfer must be implemented.

The siRNA approach might be an alternative requiring less laborious techniques. However, inactivation of mRNAs by siRNAs is often not complete, and the residual expression of mouse CFTR might diminish the relevance of the model.

4.5. Models Requiring Several Transgenes

In many cases, a mutated gene increases the probability of a disease developing but is insufficient alone. This situation is seen with the breast cancer *BRCA1* and *BRCA2* genes in human breast cancer. Several transgenes must then be expressed simultaneously to generate the pathology. The cooperative effect of two genes is sometimes unexpected. The transgenic mice used to study amyotrophic lateral sclerosis exemplify this point. The crossing of two lines of mice, one harboring the superoxide dismutase-1 gene and the other gene for a neurofilament subunit gave much more relevant models than each line separately ***(124)***.

Occasionally, the effect of a knockout gene may be compensated by the action of a very different gene, revealing a nonanticipated role of the second gene. Mice in which the LAM2 gene has been knocked out suffer from congenital muscular dystrophy. The addition of the agrin gene known to be involved in the formation of neuromuscular junctions restored muscular function ***(125)***. Alzheimer's disease (AD) is complex, and it occurs after a long

period of incubation. AD is characterized by the extracellular accumulation of polymerized tau protein. β-Amyloid is cleaved from amyloid precursor protein (ADP), the cleavage being performed by three secretases: -α, -β, and -γ ***(126)***. Transgenic mice, in which the ADP gene had been knocked out, kept their sensitivity to the disease ***(127)*** because of the presence of another gene called ADP-like gene. The knockout of the gene is lethal. Transgenic mice with knocked-out ApoE and IL1 genes are less sensitive to the disease, indicating that these two genes contribute to AD development.

Different models have been obtained. Some of the models overexpress normal APD and prenisilin genes, whereas other models express mutated forms of the genes. None of these models reflect all the aspects of AD. Fewer neurons in generals are destroyed in the models than in patients. Correlation between behavioral disorders and accumulation of β-amyloid is not always strict. Yet, the models show a similar disease pattern, and they may therefore help to identify active molecules and to design analogs capable of reaching the brain ***(128)***.

4.6. Models for Degenerative Diseases

Degenerative diseases generally have slow development, and transgenic models may be helpful to decipher their mechanisms, not only for Alzheimer's disease but also for prion diseases and other pathologies. Prion diseases would be much less well understood without the contribution of transgenic mice ***(129,130)***. Creutzfeldt–Jakob and related diseases are typically pathologies, which incubate for very long periods and require the cooperation of several cell types, especially when infection occurs by the oral route. It was essential to show that the PrP–/– mice were insensitive to any prion infection, which then demonstrated that the PrP gene has a major role in prion diseases.

Transgenic mice overexpressing the PrP genes from different species become more sensitive to the prion disease of the same species. This phenomenon is amplified when the endogenous mouse PrP gene is knocked out. For example, PrP–/– mice expressing the bovine PrP gene are sensitive to bovine spongiform encephalopathy ***(131)***. PrP–/– mice expressing the hamster PrP gene specifically in astrocytes are sensitive to hamster scrapie ***(132)***. Transgenic mice also revealed that some epitopes of the PrP protein are essential for the transmission of the disease ***(133)***, that some PrP alleles are more favorable for injection than others, and that genetic environmental factors modify the incubation period of bovine spongiform encephalopathy ***(134)***.

The prion disease manifests more rapidly when the PrP gene of the mouse and the cell extract used for infection come from more closely related species. It was also surprising to observe that the cell extracts become increasingly infectious after repeated passages to animals. This adaptation has not yet been

explained at the molecular level. All these studies in vivo have led to the definition of more flexible in vitro systems in which cultured cells can transform the Prp^c into PrP^{sc} ***(135)***. Various transgenic mice are invaluable models for the evaluation of the therapeutic effects of chemical compounds. Encouraging results showing that different drugs can delay or even partly cure scrapie have recently been obtained. Other brain degenerative diseases are being studied by using transgenic models. Transgenic mice, which overexpress the α-synuclein gene in their brain, are used to study Parkinson's disease ***(136)***.

Huntington's disease also is studied in transgenic mice overexpressing the Huntington gene variants having oligoglutamines of variable length in their N terminus ***(137)***. Apoptosis plays a major role in many biological events and is highly conserved in evolution. It is indispensable for development and organogenesis that imply specific cell elimination. Defaults in apoptosis may lead to autoimmune diseases, tumor formation, and neurodegeneracy. At least 25 genes involved in apoptosis have been knocked out ***(138)***, including genes for caspases, adaptors, regulators, bcl2 family numbers, and mitochondrial proteins. Aging is a complex phenomenon that has only been partially described. Defects in genomes seem to be a major cause of aging. A growing number of transgenic models are being used to study aging. Mice in which the XPD gene has been knocked out are more sensitive to oxidative DNA damage. This sensitivity was increased further when the XPA gene also was knocked out ***(139)***. These models reflect some of the agent syndromes in human. Arthrosclerosis results from complex disorders in lipid metabolism. Many genes are involved in this process. The disease has a genetic component, and its development is favored by cholesterol-rich diets.

Numerous genes coding for different apolipoproteins, lipases, and other factors have been added or knocked out in mice ***(140)***. These studies have given important information as to the role of the different genes. However, lipid metabolism is very different in mouse and human. Mice must be submitted to diets very rich in cholesterol to mimic even partially atherosclerosis. The rabbit is much closer to humans for this biological function, and this animal is used extensively in addition to mice for the study of atherosclerosis ***(141)***. Some transgenic rabbits also are used to define new drugs capable of protecting humans against arteriosclerosis. Cancer occurs under multiple forms, and it is the result of several steps occurring sequentially. All these steps result from alterations of oncogenes and antioncogene expression and activity. Mutations, chromosomal translocations, and environmental factors are thus responsible for tumor formation.

Transgenic mice were soon used to try to generate models for cancer study. The first oncomouse expressed c-myc gene in the mammary gland. Further studies made it possible to identify additional genes involved in mammary cancer

(142). Genes whose expression is amplified in mammary tumors (c-myc, crbB2, and cyclin D1) have an oncogenic effect when used as transgenes. Crossing mice harboring different oncogenes and having knocked-out genes has made it possible to determine the cooperative actions of some of these genes.

Mice having knocked-out cyclin D1 gene are insensitive to *neu* and *ras* oncogenes, whereas Wnt-1 and c-myc genes have kept their capacity to trigger mammary tumors. These data help to search for new drugs capable of inhibiting neu, ras, and cyclin D1 genes ***(143,144)***.

The stability of mammary epithelial cells is highly dependent on the extracellular matrix. Development of mammary tumors and metastasis are correlated with local degradation of some of the extracellular matrix components. Degradation of the extracellular matrix is achieved by metalloproteinases, which are controlled by specific inhibitors. The role of different genes involved in extracellular matrix synthesis and degradation is being studied using transgenic mouse models. A recent study has pointed out the role of Akt gene, which delays mammary gland involution. This effect is mediated by the prolonged presence of one of the metalloproteinase inhibitors ***(145)***. A better understanding of these phenomena may lead to the identification of new drugs preventing metastasis.

It is agreed that tumor cells derive from a single cell in which all the mutations required for cancer development occurred, thereby explaining in part why cancer development is a slow process at the early steps and becomes rapid in time. The classical transgenic models do not take into account the fundamental clonal origin of cancers. Specific gene constructs capable of sporadically activating oncogenes have been designed ***(146)***. The technique makes it possible to activate k-*ras* oncogene randomly and at a low frequency ***(147)***. This new approach is expected to reflect more precisely the process of tumor formation under natural conditions.

4.7. Models for Infectious Diseases

Many pathogens have species specificity, and conventional laboratory animals may not be relevant models. The limitation is often because of the lack of appropriate receptors for the pathogens in the animal cells. Receptor genes can then be transferred to animals to be preferentially expressed in tissues, which are the normal target of the pathogens.

Listeria monocytogenes is responsible for severe foodborne infections. A surface protein of these bacteria, called internalin, has been generated that recognizes the intestinal host receptor E-cadherin gene exclusively in the enterocytes of the small intestine. This recognition was achieved by using the promoter from the rat *FABP-1* gene. These mice were sensitive to infection by *L. monocytogenes* and constitute an excellent model for the study of this infectious pathogen. Interestingly, E-cadherin is not expressed in the tight junction

of the intestinal epithelium, which is accessible only when enterocytes divide. The mouse model correctly reproduced this phenomenon. They are also good candidates for internalizing foreign molecules as soon as they are associated with internalin ***(148)***. *L. monocytogenes* is known to impact cells other than enterocytes in the organism. This type of tool could study these mechanisms.

Mice sensitive to measles have been obtained by transferring to them the viral receptor CD46 gene ***(149)***. Similarly, mice harboring the gene for the polymyelitis receptor are used as models to study viral injection ***(150)***.

Hepatitis C infection is more complex. This virus does not infect cells in vitro, and in vivo models are currently strictly needed. Severe combined immunodeficient mice harboring human hepatocytes can be infected by serum from patients suffering from hepatitis C. This model has been greatly improved by using homozygous transgenic mice expressing the U-plasminogen activator gene in their liver under the control of an albumin promoter. The grafted human hepatocytes survived a longer time in mice ***(151)***.

The HIV-1 virus is known to have two major receptors in human cells, CD4 and CCR5. Transgenic rabbits expressing the human CD4 gene were shown to become seropositive after HIV-1 injection ***(152)***. Although the virus replicated in rabbit cells, it was unable to generate disease. The CCR5 receptor could have been added to the CD4 rabbits. Because of the difficulty in maintaining rabbits in appropriate confinement, and for other reasons, many potential mouse models were generated ***(153)***. It finally seemed that transgenic rats expressing all the HIV-1 genes except gag and pol showed pathology having many similarities to human AIDS ***(154)***. Transgenic mice overexpressing the soluble domain of the pseudorabies virus receptor are highly resistant to infection by the virus. This model strongly suggests that pigs expressing the same transgene could be protected against the Aujeszky disease ***(101)***.

5. Conclusions and Perspectives

Transgenesis is increasingly used not only to study gene regulation and function but also to create models for the study of human diseases. The trend is to use a lower number of more sophisticated models. It is indeed difficult to generate relevant models because the function of many genes remains unknown or the genes have not been clearly identified. Genetic background is an important consideration ***(155)***. Transgenesis is more efficient in some strains of mice such as FVB/N ***(156)***. Yet, the genetic background of this strain is not the best for all the models. Backcrossing is then needed to transfer the transgene to more appropriate genetic background. That ES cells can be used only in two strains of mice is a clear limitation to the rapid generation of relevant models. Recent studies have identified some of the genes that are required for the maintenance of the pluripotency of ES cells. The transfer of these genes in embryonic cells

from different mouse strains is currently being performed to tentatively obtain additional and functional ES cell lines. This approach could be extended to other species such as rats, rabbits, or pigs to obtain ES cell lines.

Approximately 300,000 lines of transgenic mice are expected to be generated in the coming years. All the mouse genes will be knocked out or knocked down. Knockout requires the systematic use of ES cells. ES cell clones in which all the mouse genes will have been mutated are expected to be available within 5 yr ***(157)***. A systematic targeting of chromatin regions in mouse ES cells can be achieved more easily with BAC vectors containing long DNA genome fragments the recombinant BAC vectors are obtained by homologous recombination in bacteria ***(158)***. Animal cloning by nuclear transfer offers a substitute to the use of ES cells. This technique has been implemented to inactivate a gene in pigs ***(114,115)*** and both alleles of two genes (PrP and immunoglobulin-μ) in the same cow ***(159)***. This model is expected to occur in rats and rabbits because techniques to clone these species have been described recently ***(160,161)***. No model using this technique has been reported so far.

The quality of the transgenic models is dependent on the availability of genes to be studied. It also depends on the precision with which transgenes are expressed. Approximately 5–20% of the transgenic mice obtained after DNA microinjection have insertional mutations, which are revealed most of the time when the animals are homozygous for the transgene. These uncontrolled mutations may alter the relevance of the models. Paradoxically, these numerous mutants are not frequently used to identify the genes that are mutated by the integration of the transgene and that induce phenotypic effects. The transgenes are essentially expressed at a constant rate within a given line. Some individual variations may occur because of the genetic background of the animals or because of some epigenetic mechanisms. Large variability of expression may result from two independently integrated genes expressed at different rates.

The different tools described here to generate transgenic animals allow for more regularly obtaining foreign gene integration and expression. The knowledge of the distal gene regulatory elements should help to prepare reliable constructs. The systematic search of enhancer blockers in human genome offers multiple elements to improve gene construct efficiency ***(162)***. The use of shRNA (siRNA and microRNA) for knocking down genes should have a strong impact on gene studies and replace, in several cases, the laborious knockout. Efficient and precise vectors able to express shRNA in transgenic animals should be available soon. Models based on the use of gene knockout are sometimes disappointing. In up to 30% of the knockouts, no phenotypic effects are observed, perhaps because of insufficient observations of the animals or because redundant mechanisms mask the knockout effects. In other cases, the

interference of the gene, which was knocked out with other genes, is too complex and cannot be analyzed easily, leading no clear conclusion ***(163)***.

Gene knockdown by using shRNA also may face problems. The shRNA may not be strictly specific of the target and affect other genes. Moreover, gene knockout by shRNAs is generally not complete. Sometimes, the leaked expression of the targeted gene may reduce considerably the relevance of the model. Expectedly, the simultaneous use of several shRNAs targeting the same mRNA should reduce the leaky expression. It has not clearly demonstrated whether the simultaneous use of multiple shRNAs is justified. Increasingly sophisticated models are being prepared to mimic the human diseases as much as possible. Mouse lines harboring different alleles of the same human gene are thus prepared by knock-in to evaluate their involvement in the efficiency of new pharmaceutical molecules. This process reduces the number of phase III assays to be performed in humans ***(164)***. Sometimes, the Cre recombinase used to trigger a conditional gene knockout is more precisely expressed when its gene is inserted into a long genomic DNA fragment in BAC vectors.

Notably, many gene knockouts are lethal in the early stages of embryo development, e.g., the Rb gene. Chimeric embryos formed by tetraploid Rb+/+ placenta cells and Rb–/– inner cell mass can develop and allow study of Rb gene inactivation in adults.

Gene knockout may be lethal because one essential organ has become no longer functional, whereas other interesting effects may take place in other organs. In these situations, expressing specifically the normal gene in this organ of transgenic mice may restore the function of the organ responsible for animal death. The effects of the knockout gene can then but studied in the other organs of the animals ***(165)***.

Mice are a good model to study many but not all the human diseases. Although mice as humans are mammals, some functions are too different in the two species, e.g., lipid metabolism and arteriosclerosis. Thus, transgenic rabbits are extensively used to study human diseases resulting from disorders of lipid metabolism ***(141)***. Rabbits are not intensively used but they are regularly used to generate models ***(141)***. This species has been selected mainly for breeding. The available lines of rabbits show relatively high heterogeneity in their genetic background. This characteristic significantly reduces the relevance of the models, particularly when lipid metabolism is being studied. The establishment of one or two well characterized lines of rabbits would be helpful for experimenters. This operation is relatively costly, and it could result from an international cooperation. This cost becomes more justified because the rabbit genome is currently being sequenced. Gene addition in pigs is used to generate specific models that cannot be obtained with other species ***(166)***. Importantly, experiments in animals should be guided with care and when it is possible, the number of animals should be reduced ***(167)***.

Acknowledgment

I thank Annie Paglino for her help in the preparation of the manuscript.

References

1. Brown, S. D. and Balling, R. (2001) Systematic approaches to mouse mutagenesis. *Curr. Opin. Genet. Dev.* **11,** 268–273.
2. Houdebine, L. M. (2004) Preparation of recombinant proteins in milk. *Methods Mol. Biol*. **267,** 485–494.
3. Houdebine, L. M. and Weill, B. (1999) The impact of transgenesis and cloning on cell and organ xenotransplantation to humans, in *Focus on Biotechnology* (Van Brockhoven, A., Shapiro, F., and Anné, J., eds.), Kluwer Academic Publishers, pp. 351–361.
4. Houdebine, L. M. (2002) Transgenesis to improve animal production. *Livest. Prod. Sci*. **74,** 255–268.
5. Houdebine, L. M. (ed.) (1997) *Transgenic Animals. Generation and Use*. Harwood Academic Publishers, Amsterdam.
6. Pinkert, C. A. (2002) *Transgenic Animal Technology*. Academic, Orlando, FL.
7. Schedl, A., Larin, Z., Montoliu, L., et al. (1993) A method for the generation of YAC transgenic mice by pronuclear microinjection. *Nucleic Acids Res*. **21,** 4783–4787.
8. Hostetler, H. A., Peck, S. L., and Muir, W. M. (2003) High efficiency production of germ-line transgenic Japanese medaka (*Oryzias latipes*) by electroporation with direct current-shifted radio frequency pulses. *Transgenic Res*. **12,** 413–424.
9. Dupuy, A. J., Clark, K., Carlson, C. M., et al. (2002) Mammalian germ-line transgenesis by transposition. *Proc. Natl. Acad. Sci. USA* **99,** 4495–4499.
10. Tamura, T., Thibert, C., Royer, C., et al. (1999) Germiline transformation of the silkworm *Bombyx mori* L. using a piggyBac transposon derived vector. *Nat. Biotechnol.* **18,** 81–84.
11. Mikkelsen, J. G., Yant, S. R., Meuse, L., Huang, Z., Xu, H., and Kay, M. A. (2003) Helper-independent Sleeping Beauty transposon-transposase vectors for efficient nonviral gene delivery and persistent gene expression *in vivo*. *Mol. Ther*. **8,** 654–665.
12. Masuda, K., Yamamoto, S., Endoh, M., and Kaneda, Y. (2004) Transposon-independent increase of transcription by the Sleeping Beauty transposase. *Biochem. Biophys. Res. Commun.* **317,** 796–800.
13. Lois, C., Hong, E. J., Pease, S., Brown, E. J., and Baltimore, D. (2002) Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295,** 868–872.
14. Chan, A. W., Chong, K. Y., Martinovich, C., Simerly, C., and Schatten, G. (2001) Transgenic monkeys produced by retroviral gene transfer into mature oocytes. *Science* **291,** 309–312.
15. McGrew, M. J., Sherman, A., Ellard, F. M., et al. (2004) Efficient production of germline transgenic chickens using lentiviral vectors. *EMBO Rep*. **5,** 728–733.
16. Hofmann, A., Kessler, B., Ewerling, S., et al. (2003) Efficient transgenesis in farm animals by lentiviral vectors. *EMBO Rep*. **4,** 1054–1060.

17. Whitelaw, C. B. A. (2004) Transgenic livestock made easy. *Trends Biotechnol*. **22,** 157–159.
18. Hofmann, A., Zakhartchenko, V., Weppert, M., et al. (2004) Generation of transgenic cattle by lentiviral gene transfer into oocytes. *Biol. Reprod*. **71,** 405–409.
19. Fassler, R. (2004) Lentiviral transgene vectors. *EMBO Rep*. **5,** 28–29.
20. Pfeifer, A., Hofmann, A., Kessler, B., and Wolf, E. (2004) Response to Whitelaw: Lentiviral transgenesis in livestock. *Trends Biotechnol.* **22,** 159–160.
21. Wolfgang, M. J., Eisele, S. G., Browne, M. A., et al. (2001) Rhesus monkey placental transgene expression after lentiviral gene transfer into preimplantation embryos. *Proc. Natl. Acad. Sci. USA* **98,** 10,728–10,732.
22. Sanchez, O., Toledo, J. R., Rodriguez, M. P., and Castro, F. O. (2004) Adenoviral vector mediates high expression levels of human growth hormone in the milk of mice and goats. *J. Biotechnol.* **114,** 89–97.
23. Lipps, H. J., Jenke, A. C., Nehlsen, K., Scinteie, M. F., Stehle, I. M., and Bode, J. (2003) Chromosome-based vectors for gene therapy. *Gene* **304,** 23–33.
24. Lindenbaum, M., Perkins, E., Csonka, E., et al. (2004) A mammalian artificial chromosome engineering system (ACE System) applicable to biopharmaceutical protein production, transgenesis and gene-based cell therapy. *Nucleic Acids Res*. **32,** e172.
25. Kuroiwa, Y., Kasinathan, P., Choi, Y. J., et al. (2002) Cloned transchromosomic calves producing human immunoglobulin. *Nat. Biotechnol*. **20,** 889–894.
26. Lavitrano, M., Bacci, M. L., Forni, M., et al. (2002) Efficient production by sperm-mediated gene transfer of human decay accelerating factor (hDAF) transgenic pigs for xenotransplantation. *Proc. Natl. Acad. Sci. USA* **99,** 14,230–14,235.
27. Lavitrano, M., Forni, M., Bacci, M. L., et al. (2003) Sperm mediated gene transfer in pig: Selection of donor boars and optimization of DNA uptake. *Mol. Reprod. Dev*. **64,** 284–291.
28. Wang, H. J., Lin, A. X., Zhang, Z. C., and Chen, Y. F. (2001) Expression of porcine growth hormone gene in transgenic rabbits as reported by green fluorescent protein. *Anim. Biotechnol.* **12,** 101–110.
29. Wang, K. (2003) Improving sperm mediated transgenesis: linker based sperm gene transfer: application to multiple species with a high success rate, in Proceedings of the Transgenic Animal Research Conference IV, Tahoe City, CA.
30. Marsh-Armstrong, N., Huang, H., Berry, D. L., and Brown, D. D. (1999) Germline transmission of transgenes in *Xenopus laevis*. *Proc. Natl. Acad. Sci. USA* **96,** 14,389–14,393.
31. Kato, M., Ishikawa, A., Kaneko, R., Yagi, T., Hochi, S., and Hirabayashi, M. (2004) Production of transgenic rats by ooplasmic injection of spermatogenic cells exposed to exogenous DNA: a preli Carl A. Pinkert Ed. (Ed.)minary study. *Mol. Reprod. Dev*. **69,** 153–158.
32. Moreira, P. N., Giraldo, P., Cozar, P., et al. (2004) Efficient generation of transgenic mice with intact yeast artificial chromosomes by intracytoplasmic sperm injection. *Biol. Reprod*. **71,** 1943–1947.
33. Chan, A. W., Luetjens, C. M., Dominko, T., et al. (2000) Transgene ICSI reviewed: Foreign DNA transmission by intracytoplasmic sperm injection in rhesus monkey. *Mol. Reprod. Dev*. **56,** 325–328.

34. Thermes, V., Grabher, C., Ristoratore, F., et al. (2002) I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech. Dev*. **118,** 91–98.
35. Chang, K., Qian, J., Jiang, M., et al. (2002) Effective generation of transgenic pigs and mice by linker based sperm-mediated gene transfer. *BMC Biotechnol*. **2,** 5.
36. Wang, K. (2003) Improving sperm mediated transgenesis: linker based sperm gene transfer: application to multiple species with a high success rate, in Proceedings of the Transgenic Animal Research Conference IV, Tahoe City, CA.
37. Celebi, C., Auvray, P., Benvegnu, T., Plusquellec, D., Jegou, B., and Guillaudeux, T. (2002) Transient transmission of a transgene in mouse offspring following *in vivo* transfection of male germ cells. *Mol. Reprod. Dev*. **62,** 477–482.
38. Honaramooz, A., Behboodi, E., Blash, S., Megee, S. O., and Dobrinski, I. (2003) Germ cell transplantation in goats. *Mol. Reprod. Dev*. **64,** 422–428.
39. Readhead, C., Jarvis, S., Morgan, D., and Winston, R. (2003) Male germ cells: manipulating their genome, in Proceedings of the Transgenic Animal Research Conference IV, Tahoe City, CA.
40. Oatley, J. M., de Avila, D. M., Reeves, J. J., and McLean, D. J. (2004) Spermatogenesis and germ cell transgene expression in xenografted bovine testicular tissue. *Biol. Reprod*. **71,** 494–501.
41. Schnieke, A. E., Kind, A. J., Ritchie, W. A., et al. (1997) Human factor IX transgenic sheep produced by transfer of nuclei from transfected fetal fibroblasts. *Science* **278,** 2130–2133.
42. Cibelli, J. B., Stice, S. L., Golueke, P. J., et al. (1998) Transgenic bovine chimeric offspring produced from somatic cell-derived stem-like cells. *Nat. Biotechnol*. **16,** 642–646.
43. Capecchi, M. R. (1989) Altering the genome by homologous recombination. *Science* **244,** 1288–1292.
44. Cohen-Tannoudji, M., Robine, S., Choulika, A., et al. (1998) I-SceI-induced gene replacement at a natural locus in embryonic stem cells. *Mol. Cell Biol*. **18,** 1444–1448.
45. Epinat, J. C., Arnould, S., Chames, P., et al. (2003) A novel engineered meganuclease induces homologous recombination in yeast and mammalian cells. *Nucleic Acids Res*. **31,** 2952–2962.
46. Farhadi, H. F., Lepage, P., Forghani, R., et al. (2003) A combinatorial network of evolutionarily conserved myelin basic protein regulatory sequences confers distinct glial-specific phenotypes. *J. Neurosci*. **23,** 10,214–10,223.
47. Bode, J., Schlake, T., Iber, M., et al. (2000) The transgeneticist's toolbox: novel methods for the targeted modification of eukaryotic genomes. *Biol. Chem*. **381,** 801–813.
48. Baer, A. and Bode, J. (2001) Coping with kinetic and thermodynamic barriers: RMCE, an efficient strategy for the targeted integration of transgenes. *Curr. Opin. Biotechnol*. **12,** 473–480.
49. Houdebine, L. M. (2003) *Animal Trangenesis and Cloning*. Wiley, Chichester, U.K.
50. West, A. G., Gaszner, M., and Felsenfeld, G. (2002) Insulators; many functions, many mechanisms. *Genes Dev.* **16,** 271–288.

51. Bell, A. C., West, A. G., and Felsenfeld, G. (2001) Insulators and boundaries: versatile regulatory elements in the eukaryotic genome. *Science* **291,** 447–450.
52. De Laat, W. and Grosveld, F. (2003) Spatial organization of gene expression: the active chromatin hub. *Chromosome Res*. **11,** 447–459.
53. Taboit-Dameron, F., Malassagne, B., Viglietta, C., et al. (1999) Association of the 5′HS4 sequence of the chicken beta-globin locus control region with human EF1 alpha gene promoter induces ubiquitous and high expression of human CD55 and CD59 cDNAs in transgenic rabbits. *Transgenic Res*. **8,** 223–235.
54. Rival-Gervier, S., Viglietta, C., Maeder, C., Attal, J., and Houdebine, L. M. (2002) Position-independent and tissue-specific expression of porcine whey acidic protein gene from a bacterial artificial chromosome in transgenic mice. *Mol. Reprod. Dev*. **63,** 161–167.
55. Giraldo, P., Rival-Gervier, S., Houdebine, L. M., and Montoliu, L. (2003) The potential benefits of insulators on heterologous constructs in transgenic animals. *Transgenic Res*. **12,** 751–755.
56. Zhang, Y., Muyrers, J. P., Testa, G., and Stewart, A. F. (2000) DNA cloning by homologous recombination in *Escherichia coli*. *Nat. Biotechnol*. **18,** 1314–1317.
57. Houdebine, L., Attal, J., and Vilotte, J. L. (2002) Vector design for transgene expression, in *Transgenic Animal Technology*, 2nd ed., Pinkert, C. A. (ed.), pp. 419–458.
58. Cohen-Tannoudji, M., Vandormael-Pournin, S., Drezen, J., Mercier, P., Babinet, C., and Morello, D. (2000) lacZ sequences prevent regulated expression of housekeeping genes. *Mech. Dev*. **90,** 29–39.
59. Whitelaw, E. and Martin, D. I. (2001) Retrotransposons as epigenetic mediators of phenotypic variation in mammals. *Nat. Genet*. **27,** 361–365.
60. Kwaks, T. H., Sewalt, R. G., van Blokland, R., et al. (2005) Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. *J. Biotechnol*. **115,** 35–46.
61. Houdebine, L. M. and Attal, J. (1999) Internal ribosome entry sites (IRESs): reality and use. *Transgenic Res*. **8,** 157–177.
62. Mattick, J. S. and Makunin, I. V. (2005) Small regulatory RNAs in mammals. *Hum. Mol. Genet*. **14 (Special no. R),** 121–132.
63. Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* **391,** 806–811.
64. Novina, C. D. and Sharp, P. A. (2004) The RNAi revolution. *Nature* **430,** 161–164.
65. Unwalla, H. J., Li, M. J., Kim, J. D., Li, et al. (2004) Negative feedback inhibition of HIV-1 by TAT-inducible expression of siRNA. *Nat. Biotechnol.* **22,** 1573–1578.
66. Shinagawa, T. and Ishii, S. (2003) Generation of Ski-knockdown mice by expressing a long double-strand RNA from an RNA polymerase II promoter. *Genes Dev*. **17,** 1340–1345.
67. Xia, H., Mao, Q., Paulson, H. L., and Davidson, B. L. (2002) siRNA-mediated gene silencing *in vitro* and *in vivo*. *Nat. Biotechnol*. **20,** 1006–1010.

68. Gupta, S., Schoer, R. A., Egan, J. E., Hannon, G. J., and Mittal, V. (2004) Inducible, reversible, and stable RNA interference in mammalian cells. *Proc. Natl. Acad. Sci. USA* **101,** 1927–1932.
69. Sen, G., Wehrman, T. S., Myers, J. W., and Blau, H. M. (2004) Restriction enzyme-generated siRNA (REGS) vectors and libraries. *Nat. Genet.* **36,** 183–189.
70. Shirane, D., Sugao, K., Namiki, S., Tanabe, M., Iino, M., and Hirose, K. (2004) Enzymatic production of RNAi libraries from cDNAs. *Nat. Genet.* **36,** 190–196.
71. Hohjoh, H. (2004) Enhancement of RNAi activity by improved siRNA duplexes. *FEBS Lett.* **557,** 193–198.
72. Mittal, V. (2004) Improving the efficiency of RNA interference in mammals. *Nat. Rev. Genet.* **5,** 355–365.
73. Reynolds, A., Leake, D., Boese, Q., Scaringe, S., Marshall, W. S., and Khvorova, A. (2004) Rational siRNA design for RNA interference. *Nat. Biotechnol.* **22,** 326–330.
74. Ui-Tei, K., Naito, Y., Takahashi, F., et al. (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res.* **32,** 936–948.
75. Yoshinari, K., Miyagishi, M., and Taira, K. (2004) Effects on RNAi of the tight structure, sequence and position of the targeted region. *Nucleic Acids Res.* **32,** 691–699.
76. Williams, B. R. (2005) Dicing with siRNA. *Nat. Biotechnol.* **23,** 181–182.
77. Chalk, A. M., Wahlestedt, C., and Sonnhammer, E. L. (2004) Improved and automated prediction of effective siRNA. *Biochem. Biophys. Res. Commun* **319,** 264–274.
78. Schwarz, D. S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P. D. (2003) Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115,** 199–208.
79. Khvorova, A., Reynolds, A., and Jayasena, S. D. (2003) Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115,** 209–216.
80. Kim, D. H., Behlke, M. A., Rose, S. D., Chang, M. S., Choi, S., and Rossi, J. J. (2005) Synthetic dsRNA Dicer substrates enhance RNAi potency and efficacy. *Nat. Biotechnol.* **23,** 222–226.
81. Siolas, D., Lerner, C., Burchard, J., et al. (2005) Synthetic sh RNA as potent RNAi triggers. *Nat. Biotechnol.* **23,** 227–231.
82. Luo, K. Q. and Chang, D. C. (2004) The gene-silencing efficiency of siRNA is strongly dependent on the local structure of mRNA at the targeted region. *Biochem. Biophys. Res. Commun.* **318,** 303–310.
83. Snove, O., Jr., and Holen, T. (2004) Many commonly used siRNAs risk off-target activity. *Biochem. Biophys. Res. Commun.* **319,** 256–263.
84. Judge, A. D., Sood, V., Shaw, J. R., Fang, D., McClintock, K., and Maclachlan, I. (2005) Sequence-dependent stimulation of the mammalian innate immune response by synthetic siRNA. *Nat. Biotechnol.* **23,** 457–462.
85. Yelin, R., Dahary, D., Sorek, R., et al. (2003) Widespread occurrence of antisense transcription in the human genome. *Nat. Biotechnol.* **21,** 379–386.

86. Carmichael, G. G. (2003) Antisense starts making more sense. *Nat. Biotechnol*. **21,** 371–372.
87. Ying, S. Y. and Lin, S. L. (2004) Intron-derived microRNAs—fine tuning of gene functions. *Gene* **342,** 25–28.
88. Ying, S. Y. and Lin, S. L. (2005) Intronic microRNAs. *Biochem. Biophys. Res. Commun* **326,** 515–520.
89. Kawasaki, H. and Taira, K. (2004) Induction of DNA methylation and gene silencing by short interfering RNAs in human cells. *Nature* **431,** 211–217.
90. Morris, K. V., Chan, S. W., Jacobsen, S. E., and Looney, D. J. (2004) Small interfering RNA-induced transcriptional gene silencing in human cells. *Science* **305,** 1289–1292.
91. Zeng, Y., Yi, R., and Cullen, B. R. (2003) MicroRNAs and small interfering RNAs can inhibit mRNA expression by similar mechanisms. *Proc. Natl. Acad. Sci. USA* **100,** 9779–9784.
92. He, L. and Hannon, G. J. (2004) MicroRNAs: small RNAs with a big role in gene regulation. *Nat. Rev. Genet*. **5,** 522–531.
93. Xie, X., Lu, J., Kulbokas, E. J., et al. (2005) Systematic discovery of regulatory motifs in human promoters and 3′ UTRs by comparison of several mammals. *Nature* **434,** 338–345.
94. Lee, Y., Kim, M., Han, J., et al. (2004) MicroRNA genes are transcribed by RNA polymerase II. *EMBO J*. **23,** 4051–4060.
95. Zeng, Y. and Cullen, B. R. (2003) Sequence requirements for micro RNA processing and function in human cells. *RNA* **9,** 112–123.
96. Ecclestin, A. and Eggleston, A. K. (2004) RNA interference. *Nature* **431,** 337–378.
97. Clayton, J. (2004) RNA interference: the silent treatment. *Nature* **431,** 599–605.
98. Müller, M. (2000) Increasing disease resistance in transgenic domestic, in *Molecular Farming* (Toutant, J. P. and Balazs, E., eds.), INRA Editions, Paris, pp. 87–98.
99. Jones, S. D. and Marasco, W. A. (1997) Intracellular antibodies (intrabodies): potential applications in transgenic animal research and engineered resistance to pathogens, in *Transgenic Animal Generation and Use* (Houdebine, L. M., ed.), Harwood Academic Publishers, Amsterdam, pp. 501–506.
100. Chang, P. Y., Benecke, H., Le Marchand-Brustel, Y., Lawitts, J., and Moller, D. E. (1994) Expression of a dominant-negative mutant human insulin receptor in the muscle of transgenic mice. *J. Biol. Chem*. **269,** 16,034–16,040.
101. Ono, E., Amagai, K., Taharaguchi, S., et al. (2004) Transgenic mice expressing a soluble form of porcine nectin-1/herpesvirus entry mediator C as a model for pseudorabies-resistant livestock. *Proc. Natl. Acad. Sci. USA* **101,** 16,150–16,155.
102. Chen, Y. T., Levasseur, R., Vaishnav, S., Karsenty, G., and Bradley, A. (2004) Bigenic Cre/loxP, puDeltatk conditional genetic ablation. *Nucleic Acids Res*. **32,** e161.
103. Saito, M., Iwawaki, T., Taya, C., et al. (2001) Diphtheria toxin receptor-mediated conditional and targeted cell ablation in transgenic mice. *Nat. Biotechnol*. **19,** 746–750.
104. Jiang, W., Zhou, L., Breyer, B., et al. (2001) Tetracycline-regulated gene expression mediated by a novel chimeric repressor that recruits histone deacetylases in mammalian cells. *J. Biol. Chem*. **276,** 45,168–45,174.

105. Weber, W. and Fussenegger, M. (2004) Approaches for trigger-inducible viral transgene regulation in gene-based tissue engineering. *Curr. Opin. Biotechnol*. **15,** 383–391.
106. Boutonnet, C., Boijoux, O., Bernat, S., et al. (2004) Pharmacological-based translational induction of transgene expression in mammalian cells. *EMBO Rep*. **5,** 721–727.
107. Cecconi, F. and Meyer, B. I. (2000) Gene trap: a way to identify novel genes and unravel their biological function. *FEBS Lett.* **480,** 63–71.
108. Jackson, I. J. (2001) Mouse mutagenesis on target. *Nat. Genet.* **28,** 198–200.
109. Medico, E., Gambarotta, G., Gentile, A., Comoglio, P. M., and Soriano, P. (2001) A gene trap vector system for identifying transcriptionally responsive genes. *Nat. Biotechnol.* **19,** 579–582.
110. Mitchell, K. J., Pinson, K. I., Kelly, O. G., et al. (2001) Functional analysis of secreted and transmembrane proteins critical to mouse development. *Nat. Genet.* **28,** 241–249.
111. Goodwin, N. C., Ishida, Y., Hartford, S., et al. (2001) DelBank: a mouse ES-cell resource for generating deletions. *Nat. Genet*. **28,** 310–311.
112. Houdebine, L. M. (2002) Antibody manufacture in transgenic animals and comparisons with other systems. *Curr. Opin. Biotechnol*. **13,** 625–629.
113. Houdebine, L. M. and Weill, B. (1999) The impact of transgenesis and cloning on cell and organ xenotransplantation to humans, in *Focus on Biotechnology* (Van Brockhoven, A., Shapiro, F., and Anné, J., eds.), Kluwer Academic Publishers, pp. 351–361.
114. Lai, L., Kolber-Simonds, D., Park, K. W., et al. (2002) Production of alpha-1, 3-galactosyltransferase knockout pigs by nuclear transfer cloning. *Science* **295,** 1089–1092.
115. Dai, Y., Vaught, T. D., Boone, J., et al. (2002) Targeted disruption of the alpha1, 3-galactosyltransferase gene in cloned pigs. *Nat. Biotechnol*. **20,** 251–255.
116. Switzer, W. M., Michler, R. E., Shanmugam, V., et al. (2001) Lack of cross-species transmission of porcine endogenous retrovirus infection to nonhuman primate recipients of porcine cells, tissues, or organs. *Transplantation* **71,** 959–965.
117. Oldmixon, B. A., Wood, J. C., Ericsson, T. A., et al. (2002) Porcine endogenous retrovirus transmission characteristics of an inbred herd of miniature swine. *J. Virol.* **76,** 3045–3048.
118. Chan, F., Bradley, A., Wensel, T. G., and Wilson, J. H. (2004) Knock-in human rhodopsin-GFP fusions as mouse models for human disease and targets for gene therapy. *Proc. Natl. Acad. Sci. USA* **101,** 9109–9114.
119. Boulanger, L., Mallet, S., Chensé, P., et al. (2002) Advantages and limits of using the ubiquitous expressed EF1alpha promoter for transgenesis *in vivo* and *in vitro* in rabbit. *Transgenic Res*. **11,** 88.
119a. al Gubarg, K. and Houdebine, L. M. In vivo imaging of green fluorescent protein-expressing cells in transgenic animals using fibered confocal fluorescience microscopy. *Eur. J. Cell Biol.*, in press.
120. Devgan, V., Rao, M. R., and Seshagiri, P. B. (2004) Impact of embryonic expression of enhanced green fluorescent protein on early mouse development. *Biochem. Biophys. Res. Commun.* **313,** 1030–1036.

121. Pailhoux, E., Vigier, B., Chaffaux, S., et al. (2001) A 11.7-kb deletion triggers intersexuality and polledness in goats. *Nat. Genet*. **29,** 453–458.
122. Vaiman, D. (2003) Sexy transgenes: the impact of gene transfer and gene inactivation technologies on the understanding of mammalian sex determination. *Transgenic Res*. **12,** 255–269.
123. Shillingford, J. M. and Henneighausen, L. (2001) Experimental mouse geneticsanswering fundamental questions about mamary gland biology. *Trends Endocrinol Metab* **12,** 402–408.
124. Kong, J. and Xu, Z. (2000) Overexpression of neurofilament subunit NF-L and NF-H extends survival of a mouse model for amyotrophic lateral sclerosis. *Neurosci Lett*. **281,** 72–74.
125. Moll, J., Barzaghi, P., Lin, S., et al. (2001) An agrin minigene rescues dystrophic symptoms in a mouse model for congenital muscular dystrophy. *Nature* **413,** 302–307.
126. Esler, W. P. and Wolfe, M. S. (2001) A portrait of Alzheimer secretases—new features and familiar faces. *Science* **293,** 1449–1454.
127. Lewis, J., Dickson, D. W., Lin, W. L., et al. (2001) Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP. *Science* **293,** 1487–1491.
128. Chapman, P. F., Falinska, A. M., Knevett, S. G., and Ramsay, M. F. (2001) Genes, models and Alzheimer's disease. *Trends Genet*. **17,** 254–261.
129. Moore, R. C. and Melton, D. W. (1997) Transgenic analysis of prion diseases. Mol. Hum. Reprod. **3,** 529–544.
130. Prusiner, S. B., Scott, M. R., DeArmond, S. J., and Cohen, F. E. (1998) Prion protein biology. *Cell* **93,** 337–348.
131. Scott, M. R., Will, R., Ironside, J., et al. (1999) Compelling transgenetic evidence for transmission of bovine spongiform encephalopathy prions to humans. *Proc. Natl. Acad. Sci. USA* **96,** 15,137–15,142.
132. Raeber, A. J., Race, R. E., Brandner, S., et al. (1997) Astrocyte-specific expression of hamster prion protein (PrP) renders PrP knockout mice susceptible to hamster scrapie. *EMBO J.* **16,** 6057–6065.
133. Scott, M. R., Safar, J., Telling, G., et al. (1997) Identification of a prion protein epitope modulating transmission of bovine spongiform encephalopathy prions to transgenic mice. *Proc. Natl. Acad. Sci. USA* **94,** 14,279–14,284.
134. Manolakou, K., Beaton, J., McConnell, I., et al. (2001) Genetic and environmental factors modify bovine spongiform encephalopathy incubation period in mice. *Proc. Natl. Acad. Sci. USA* **98,** 7402–7407.
135. Vilette, D., Andreoletti, O., Archer, F., et al. (2001) Ex vivo propagation of infectious sheep scrapie agent in heterologous epithelial cells expressing ovine prion protein. *Proc. Natl. Acad. Sci. USA* **98,** 4055–4059.
136. Betarbet, R., Sherer, T. B., and Greenamyre, J. T. (2002) Animal models of Parkinson's disease. *Bioessays* **24,** 308–318.
137. Rubinsztein, D. C. (2002) Lessons from animal models of Huntington's disease. *Trends Genet.* **18,** 202–209.

138. Ranger, A. M., Malynn, B. A., and Korsmeyer, S. J. (2001) Mouse models of cell death. *Nat. Genet*. **28,** 113–118.
139. De Boer, J., Andressoo, J. O., De Wit, J., et al. (2002) Premature aging in mice deficient in DNA repair and transcription. *Science* **296,** 1276–1279.
140. Miller, M. W. and Rubin, E. M. (1997) Transgenic animals in atherosclerosis research, in *Transgenic Animal and Generation and Use* (L. M. Houdebine, ed.), Harwood Academic Publishers, Amsterdam, pp. 445–448.
141. Fan, J. and Watanabe, T. (2003) Transgenic rabbits as therapeutic protein bioreactors and human disease models. *Pharmacol. Ther*. **99,** 261–282.
142. Siegel, P. M., Hardy, W. R., and Muller, W. J. (2000) Mammary gland neoplasia: insights from transgenic mouse models. *Bioessays* **22,** 554–563.
143. Bartek, J. and Lukas, J. (2001) Are all cancer genes equal? *Nature* **411,** 1001–1002.
144. Yu, Q., Geng, Y., and Sicinski, P. (2001) Specific protection against breast cancers by cyclin D1 ablation. *Nature* **411,** 1017–1021.
145. Schwertfeger, K. L., Richert, M. M., and Anderson, S. M. (2001) Mammary gland involution is delayed by activated Akt in transgenic mice. *Mol. Endocrinol*. **15,** 867–881.
146. Berns, A. (2001) Cancer. Improved mouse models. *Nature* **410,** 1043–1044.
147. Johnson, L., Mercer, K., Greenbaum, D., et al. (2001) Somatic activation of the K-ras oncogene causes early onset lung cancer in mice. *Nature* **410,** 1111–1116.
148. Lecuit, M., Vandormael-Pournin, S., Lefort, J., et al. (2001) A transgenic model for listeriosis: role of internalin in crossing the intestinal barrier. *Science* **292,** 1722–1725.
149. Oldstone, M. B., Lewicki, H., Thomas, D., et al. (1999) Measles virus infection in a transgenic model: virus-induced immunosuppression and central nervous system disease. *Cell* **98,** 629–640.
150. Ren, R. B., Costantini, F., Gorgacz, E. J., Lee, J. J., and Racaniello, V. R. (1990) Transgenic mice expressing a human poliovirus receptor: a new model for poliomyelitis. *Cell* **63,** 353–362.
151. Fausto, N. (2001) A mouse model for hepatitis C virus infection? *Nat. Med*. **7,** 890–891.
152. Dunn, C. S., Mehtali, M., Houdebine, L. M., Gut, J. P., Kirn, A., and Aubertin, A. M. (1995) Human immunodeficiency virus type 1 infection of human CD4-transgenic rabbits. *J. Gene. Virol*. **76,** 1327–1336.
153. Cohen, J. (2001) Building a small-animal model for AIDS, block by block. *Science* **293,** 1034–1036.
154. Reid, W., Sadowska, M., Denaro, F., et al. (2001) An HIV-1 transgenic rat that develops HIV-related pathology and immunologic dysfunction. *Proc. Natl. Acad. Sci. USA* **98,** 9271–9276.
155. Carvallo, G., Canard, G., and Tucker, D. (1997) Standardization of transgenic lines: from founder to an established animal model, in *Transgenic Animal Generation and Use* (Houdebine, L. M., ed.), Harwood Academic Publishers, Amsterdam, pp. 403–410.

156. Auerbach, A. B., Norinsky, R., Ho, W., et al. (2003) Strain-dependent differences in the efficiency of transgenic mouse production. *Transgenic Res.* **12,** 59–69.
157. Abbott, A. (2004) Geneticists prepare for deluge of mutant mice. *Nature* **432,** 541.
158. Valenzuela, D. M., Murphy, A. J., Frendewey, D., et al. (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat. Biotechnol.* **21,** 652–659.
159. Kuroiwa, Y., Kasinathan, P., Matsushita, H., et al. (2004) Sequential targeting of the genes encoding immunoglobulin-micro and prion protein in cattle. *Nat. Genet.* **36,** 671–672.
160. Zhou, Q., Renard, J. P., Le Friec, G., et al. (2003) Generation of fertile cloned rats by regulating oocyte activation. *Science* **302,** 1179.
161. Chesne, P., Adenot, P. G., Viglietta, C., Baratte, M., Boulanger, L., and Renard, J. P. (2002) Cloned rabbits produced by nuclear transfer from adult somatic cells. *Nat. Biotechnol.* **20,** 366–369.
162. Kwaks, T. H., Barnett, P., Hemrika, W., et al. (2003) Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. *Nat. Biotechnol*. **21,** 553–558.
163. Szathmary, E., Jordan, F., and Pal, C. (2001) Molecular biology and evolution. Can genes explain biological complexity? *Science* **292,** 1315–1316.
164. Liggett, S. B. (2004) Genetically modified mouse models for pharmacogenomic research. *Nat. Rev. Genet*. **5,** 657–663.
165. Lee, D. and Threadgill, D. W. (2004) Investigating gene function using mouse models. *Curr. Opin. Genet. Dev.* **14,** 246–252.
166. Kues, W. A. and Niemann, H. (2004) The contribution of farm animals to human health. *Trends Biotechnol*. **22,** 286–294.
167. Moore, A. (2001) Of mice and Mendel. The predicted rise in the use of knock-out and transgenic mice should cause us to reflect on our justification for the use of animals in research. *EMBO Rep*. **2,** 554–558.

11

Keratin Transgenic and Knockout Mice

Functional Analysis and Validation of Disease-Causing Mutations

Preethi Vijayaraj, Goran Söhl, and Thomas M. Magin

Summary

The intermediate filament (IF) cytoskeleton of mammalian epithelia is generated from pairs of type I and type II keratins that are encoded by two large gene families, made up of 54 genes in humans and the mouse. These genes are expressed in a spatiotemporal and tissue-specific manner from the blastocyst stage onward. Since the discovery of keratin mutations leading to epidermolysis bullosa simplex, mutations in at least 18 keratin genes have been identified that result in keratinopathies of the epidermis and its appendages. Recently, noncanonical mutations in simple epithelial keratins were associated with pancreatic, liver, and intestinal disorders, demonstrating that keratins protect epithelia against mechanical and other forms of stress. In recent years, animal models provided novel insight and significantly improved understanding of IF function in tissue homeostasis and its role in disease. Pathological phenotypes detected in mutant mice generated so far range from embryonic lethality to tissue fragility to subtlety, which often depends on their genetic background. This range implies at least a partial influence of yet unidentified modifier genes on the phenotype after the ablation of the respective keratin. To date, nearly all available keratin mouse models were generated by taking advantage of conventional gene-targeting strategies. To reveal their cell type-specific functions and the mechanisms by which mutations lead to disease, it will be necessary to use conditional gene-targeting strategies and the introduction of point-mutated gene copies. Furthermore, conditional strategies offer the possibility to overcome embryonic or neonatal lethality in some of the keratin-deficient mice.

Key Words: Blastocysts injection; Cre/LoxP system; epidermolysis bullosa simplex (EBS); epidermolytic hyperkeratosis (EHK); pachyonychia congenita; embryonic stem (ES) cell culture; Flp/FRT system; genomic cluster knockout; intermediate filaments; knockout mice; tetraploids; transgenic mice.

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarker Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

1. Introduction

1.1. Characterization of Genetically Altered Keratin Mice

1.1.1. Two Families of Keratin Genes

Of the currently known and characterized intermediate filament (IF) genes ***(1)***, keratin genes establish the most complex and conserved gene family in mammals. They are encoded by a total of 54 members in the human and mouse genome, located in two different arrays and designated as type I and type II clusters ***(2,3)*** (*see* **Fig. 1**).

All type I keratin genes, except for K18, are clustered on human chromosome 17q21 in synteny to mouse chromosome 11D, whereas the type II cluster, localized on human chromosome 12q13, is in synteny to mouse chromosome 15F ***(4)*** (The International Human Genome Sequencing Consortium).

The human keratin I gene cluster comprises totally 977 kilobases (kb), consists of 27 functional and 4 pseudokeratin type I genes as well as high and ultrahigh sulfur keratin-associated proteins (KPs) ***(5)***. In the mouse, the keratin I gene cluster is approx 1,034 kb and highly orthologous in terms of number as well as orientation of the orthologous keratin genes and KPs ***(3)***. The central KP subdomain (362 kb in human and 466 kb in mouse) divides this cluster into a centromeric and telomeric portion and contains in human 29 genes coding for high and ultrahigh sulfur (depending on their cysteine content) hair KPs, being transcribed in the upper cortex of the hair shaft ***(5)***. This KP subdomain is surrounded by keratin genes of the inner root sheath of hair follicles ***(6)*** and several hair keratin genes ***(3)***. Thus, the human and mouse keratin I gene cluster can be considered as highly organized in subdomains containing cytokeratins (K9–K24), hair keratins (K25–K36), and inner root sheath keratins (K38–K41) **(Fig. 1)**.

The human keratin II gene cluster is located on chromosome 12q13.13, comprises 783 kb and contains 26 functional type II genes, 5 type II pseudogenes, and 3 keratin unrelated pseudogenes. In the mouse type II cluster, however, there are 27 functional type II genes, mostly arranged in the same orientation. Interestingly, both type II clusters (mouse and human) end up with a type I keratin gene, the K18 gene. However, the type II clusters of both species diverge more than their type I clusters, which could be considered as highly identical. The human type II cluster contains shortly upstream of the K18 gene three nonkeratin genes as well as the cornea-specific keratin gene K3, all of which lack a mouse ortholog. This difference, most likely, resulted from gene duplication in human and other species ***(7)*** or vice versa from a K3 gene lost in the mouse genome.

1.1.2. Keratin Assembly, Structure, and Modifications

Similar to other intermediate filament protein, the secondary structure of keratin proteins can be subdivided into three subdomains: a central, largely α-helical

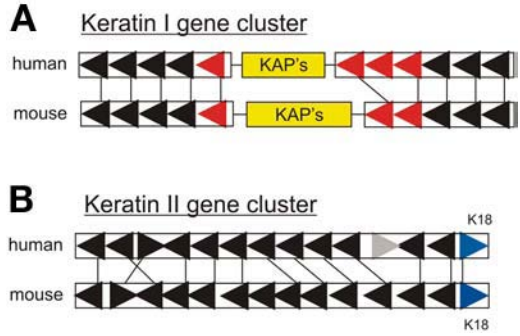


Fig. 1. Schematic alignment of both mouse and human keratin type I and type II gene cluster. One black arrowhead represents approximately two to three keratin genes orientated in the direction of the tip. Gray arrowheads in **(A)** delineate keratins of the inner root sheet of hair follicles ***(6)***. The light gray arrowhead in **(B)** indicates an array of processed pseudogenes unrelated to keratins. The dark gray arrowheads in **(B)** highlight the only type I keratin (K18) at the end of both mouse and human type II cluster. KAP's is the abbreviation of a cluster composed of keratin associated proteins (*see* **Note 3**).

"rod" domain flanked by non-α helical amino-terminal and carboxy-terminal domains, which guide two monomers into a double-stranded and parallel coiled-coil formation. The rod domain itself is composed of four highly conserved domains, termed 1A and B and 2A and B connected by non-α-helical linkers (*see* **Fig. 2**). An individual α-helical segment exhibits a heptad substructure $(abcdefg)_n$, in which positions a and d are occupied by hydrophobic amino acids, thus establishing a seam around the axis of a single right-handed α-helix that contributes to the formation of the coiled-coil structure. Only a short sequence in the middle of segment 2B interrupts this regular phasing of heptades that does not form a coiled-coil structure but rather runs in parallel with the corresponding part of the second molecule of the adjacent coiled-coil dimmer ***(8,9)***.

Special motifs within the 1B and 2B coil might trigger the self-assembly of an obligatory type I and II keratin into a parallel heterodimer that further assembles into antiparallel, half-staggered tetramers that laterally associate to the so-called unit-length filaments (ULFs). Finally, these structures anneal longitudinally into nonpolar filaments compacting to IFs of approx 11 nm in diameter ***(10,11)***. There is mounting evidence from in vitro as well from in vivo studies that at least some keratins might replace each other, supporting the idea of functional redundancy ***(8)***.

Motifs within the head domain have been shown to be essential for the formation of filaments, whereas tail sequences are suspected to regulate filament diameter or to interact with associated proteins. However, a side contribution of the tail to filament bundling, at least in some keratins, cannot be ruled out ***(12)***.

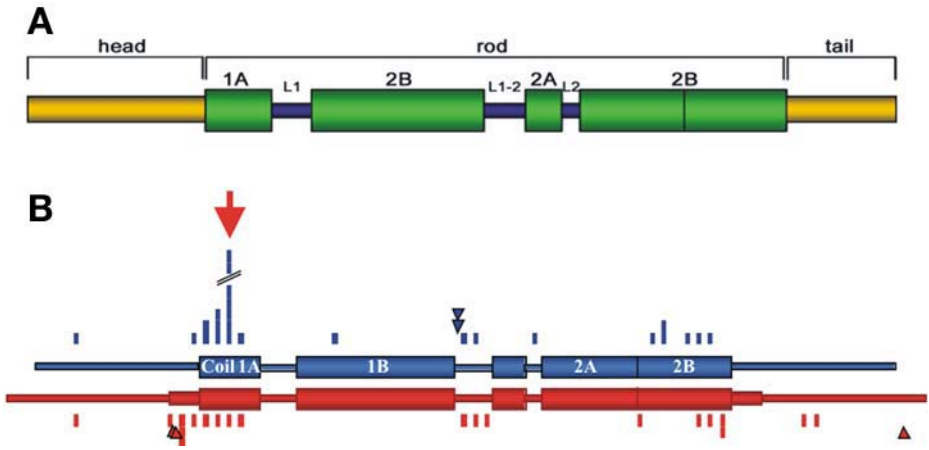


Fig. 2. **(A)** General structure of an intermediate filament (IF) protein divided into three different sections: coiled-coil rod domain (green); head and tail domain (yellow); keratin type I cluster genes **(B**, above**)**. The latter are very similar to keratin type II cluster genes (below) except that they are missing a traditional head and tail domain. Vertical bars indicate the frequency of already detected human pathological mutations at a certain position.

Mutations identified in the tail sequences of K1 and K5 have been shown to be responsible for ichthyosis hystrix and epidermolysis bullosa simplex (EBS) with migrating erythema. However, the most severe mutations found in keratin genes are located in the evolutionarily conserved end domains of the α helical rod ***(13)*** rather than in both head and tail domains ***(14,15)***. Generally, point mutations, such as those found in K5 and K14 that cause different forms of EBS ***(16,17)*** (*see* **Fig. 3A**), provide insight into the process of IF formation. Interestingly, the K14R125C/H mutation has no effect on filament assembly or length in vitro ***(10)***, but it reveals a decreased resilience of filaments against large deformations ***(18)*** and leads to filament aggregation depending on the molar ratio of mutant to the wild-type protein ***(19)***. Filament formation, *per se*, seems to initiate close to the plasma membrane. The role of desmosomes, which might play a role in IF assembly is still unclear ***(20)***.

Keratin-deficient animals helped answer the question of whether there is a functionally redundant keratin substituting an ablated keratin. The K5 null mice (*see* **Fig. 3F**) die shortly after birth, lack keratin filaments in the basal epidermis, and are more severely affected than K14(–/–) mice. However, the unaffected formation of intact K1- and K10-containing IFs in the suprabasal layers indicates that K5 is not needed as a “chaperone” or “scaffold” for K1 and K10 ***(21)***. Additionally, many studies have been undertaken to determine the influence of posttranslational modifications such as phosphorylation, glycosylation, transglutamination, deimination, and ubiquitination on keratin assembly and

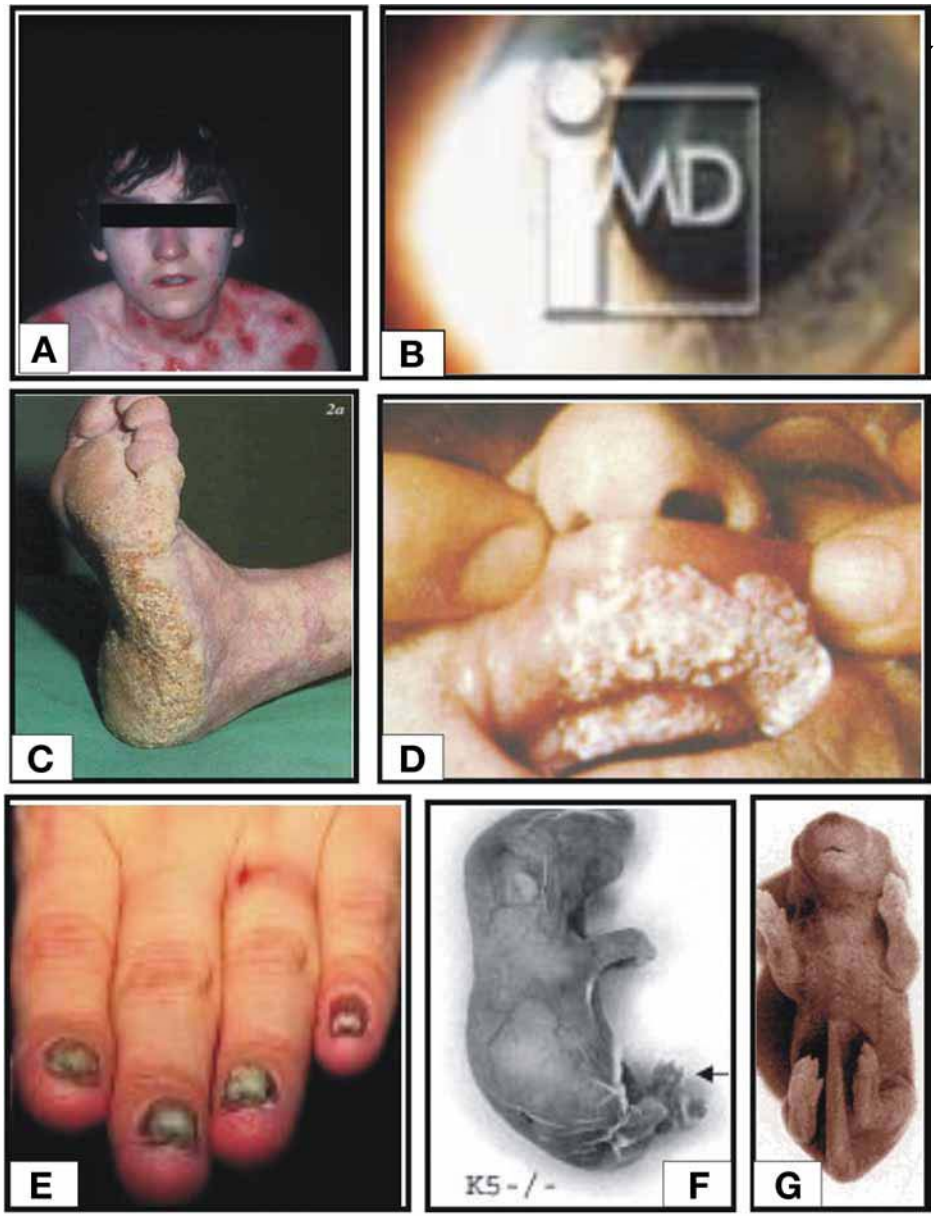


Fig. 3. **(A)** Epidermolysis bullosa simplex (EBS) evoked by mutations in K5 and also K14. **(B)** Meesmann's corneal dystrophy elicited by mutations in K12. **(C)** Mutations in the keratin K6b and also other keratins are likely to trigger epidermolytic hyperkeratosis (EHK). **(D)** K4 mutations are reported to be responsible for white sponge nevus. **(E)** Pachyonychia congenita caused by mutations in K6a, K6b, and also K17. **(F)** Ablation of K5 in mouse causes EBS and neonatal death ***(21)***. **(G)** Total loss of K10 in the suprabasal epidermis in neonatal mice is without signs of fragility or wound healing response ***(75)***. There is free access to the pictures A to E, showing patients suffering from the mentioned syndromes on the World Wide Web. Pictures F and G are derived from the corresponding publications and slightly modified.

filament organization, which we cannot refer to because of space limitations. For these modifications, more detailed literature is recommended ***(22–24)***.

1.1.3. Keratin Gene Expression and Regulation

Today, it is commonly accepted that in simple as well as stratified epithelia, keratin genes are expressed in a "pairwise" manner. An acid type I keratin protein always attaches to a basic type II keratin to assemble a dimeric structure ***(8)***. Although redundant expression of further type I and type II keratins in different epithelia readily occurs, expression also takes place in this already mentioned pairwise manner. But what regulates this distinct pairwise turning-on and -off of different type I and II keratin genes?

The human K18 keratin gene (~10 kb) ***(25)*** serves as a prime example of what might influence the transcriptional control of a keratin gene. Regulatory elements (Alu sequences and four putative SP1 binding sites) surrounding the seven coding exons as well as an AP-1 consensus site in exon 6 interacting with Ets-1 elements in intron 1 have been demonstrated to be involved in the copy number-dependent and position-independent transgenic expression of hK18. Additionally, two differentially methylated CpGs islands have been found in the first intron enhancer ***(26–28)***. Recently, taking this knowledge into consideration, Oshima and colleagues verified the coincidence of both hK18 transgene expression and that of different reporter genes in addition to the gene coding for the Cre-recombinase subcloned together with an internal ribosome entry side (IRES) immediately downstream of exon 7 ***(29)***.

Results from the paradigmatic expression of the human K18 transgene imply that putative transcription factor binding sites merely identified in the (core)-promoter might be not sufficient to explain the specific and pairwise expression pattern of keratin genes ***(30–32)***.

In internal so-called "simple" epithelia, K8 and K18 form the pair that is predominantly expressed from embryonic day (ED)2.5 on ***(32)*** and that is complemented by K7, K19, K20, and K23 in most epithelia starting at ED3.5 ***(32)***. Generally, these keratins seem to be able to replace for K8 and K18, provided that they are expressed in the same cell type ***(33,34)***.

In stratified epithelia, the prominent keratins are K5 and K14, which are complemented by different amounts of K15 and K17 in their basal layer ***(21,35,36)***. The upper parts of the epidermis maintain K5 and K14 as proteins, probably because of their long half-lives, but when keratinocytes start to differentiate they express highly abundant K1 and K10, accompanied by K1b, K2e, and K9 present in upper spinous and granular keratinocytes, particularly exposed to stress ***(37,38)***. Under conditions of proliferation or differentiation such as keratinocyte migration and wound healing, however, the so-called reinforcement keratins K6a, K6b, K6e, and K6hf as well as K16 and K17 often replace either transiently

or constitutively K1, K2e, K9, and K10 in mouse and human ***(39–41)***. More related to epidermal appendages, the novel keratin 17 (K17n) gene is mainly expressed in nail tissue as well as in fili- and fungiform papillae of the oral mucosa ***(36)***.

Nevertheless, the aforementioned concept of "keratin expression pairs" is proven by the exception that there is no mouse ortholog to the human K3 gene ***(3)***. In the differentiated layers of the human cornea, both K3 and K12 are abundantly expressed; the former keratin has to be replaced by K5 in the mouse ***(42)***. During embryonic development, keratin protein expression starts very early with K8 and K18 at ED2.5 ***(32)*** being detectable in the surface ectoderm ***(43,44)*** and is followed by K5 (ED9.25) and K14 (ED9.75) ***(32)***. Shortly afterward, at ED11.5, keratins K6a, K6b, K16, and K17 appear, even before the onset of stratification at approx ED13.5 in the mouse ***(45)***. In this respect, it is intriguing that the mouse K14 gene might be a direct target of the p53 tumor suppressor homolog p63 because K14 was expressed at very low levels in the single layer of p63-deficient epidermis of mouse ***(46)***.

1.1.4. Keratin Function in Cell Signaling and Keratinopathies

Mutations within keratin genes lead to various epidermal diseases. The diseases have different pathologies mainly caused by cell fragility, which could be accompanied by a disturbance of the normal cell signaling ***(1,47,48)***. An electronic database including various keratin mutations is available: at www.interfil.org.

The most prominent K14R125C mutation not only leads to EBS **(Fig. 3A)** because of keratin (IF) aggregate formation coinciding with decreased resilience to mechanical forces but also to a weakening of the K14–tumor necrosis factor (TNF)α receptor-associated death domain (TRADD). This weakening leaves transfected keratinocytes more susceptible for caspase-8–mediated apoptosis caused by an alleviated cytoprotection mediated by K14–TRADD interactions ***(49)***. Furthermore, cell culture studies with these transfected keratinocytes revealed that keratin aggregates are in a dynamic equilibrium with soluble subunits, suggesting that dominant negative mutations act by altering cytoskeletal dynamics and solubility largely depending on the "mutant to wild-type" ratio ***(19)***. Keratin mutations render keratinocytes more susceptible to osmotic shock and activate the stress-activated protein kinase/c-Jun NH_2-terminal kinase (JNK) pathway ***(50)***.

Generally, point mutations in or truncation of epidermal keratins result in visible monogenic keratinopathies, whereas phenotypes of mutated embryonic and internal epithelial keratins are much more subtle. Depending on the population, K8 as well as K18 gene mutations (*see* **Tables 1** and **2**) have been associated with cryptogenic cirrhosis, liver disorders, and chronic pancreatitis as well as

Table 1
Orthologous Human and Mouse Keratin Genes of Type I Cluster

Human keratin	Mouse keratin	Major expression	Knockout phenotype	Human pathology	Reference
K9	K9		n.d.	Unknown	
K10	K10	Suprabasal epidermis	Hyperkeratosis, acanthosis; K14 compensation	Non-epidermolytic hyperkeratosis	***75,77***
	K11P		n.d.	Unknown	
K12	K12	Cornea	Cell fragility	Meesmann's corneal dystrophy	Ko et al., 1996 (*see* **Fig. 3B**)
K13	K13		n.d.	Unknown	
K14	K14	Basal epidermis	Postnatal death at P4, skin blistering cytolysis of basal cells	Epidermolysis bullosa simplex (EBS)	***35*** (*see* **Fig. 3A**)
K15	K15		n.d.	Unknown	
K16	K16		n.d.	Pachyonychia congenita	***93*** (*see* **Fig. 3E**)

K17	K17	Hair follicles, nail bed; esophagus, activated keratinocytes	Alopecia, compensation by K14 and K16	Pachyonychia congenita	***93*** (*see* **Fig. 3E**)
K18	K18	Simple epithelia	Mild liver pathology	Ulcerative colitis	***34***
K19	K19	Simple epithelia	None, compensation by K18 and/or K20	Unknown	***63,64***
K8 K19	K8 K19	Simple epithelia	Embryonic death at ED10; trophoblast/ placental defect	Ulcerative colitis	***63,51***
K18 K19	K18 K19	Simple epithelia	Embryonic death at ED9.5; trophoblast/giant cell fragility	Ulcerative colitis	***64***

Table 2
Orthologous Human and Mouse Keratin Genes of Type II Cluster[a]

Human keratin	Mouse keratin	Major expression	Knockout phenotype	Human pathology	Reference
K1	K1		n.d.	Unknown	
K2	K2		n.d.	Unknown	
K3				Unknown	
K4	K4	Stratified epithelia	Cell fragility, acanthosis; K6 partially compensates	White sponge nevus	Ness et al. (1998) (**Fig. 3D**)
K5	K5	Basal epidermis	Neonatal death; skin blistering; cytolysis of basal cells	Epidermolysis bullosa simplex (EBS)	***21*** (*see* **Fig. 3A**)
K6a	K6a	Hair follicles; activated keratinocytes; nail bed epithelum	None; compensation by K6b	Pachyonychia congenita	***96*** (*see* **Fig. 3E**)
K6a K6b	K6a K6b	Hair follicles; activated keratinocytes; nail bed epithelum	Tongue lesions; postnatal death P10; Cell lysis upon migration during wound healing; 25% survival because of compensation by K18	Pachyonychia congenita	***93,96*** (**Fig. 3E**)

K6a K6b K17 K7	K6a K6b K17 K7	Hair follicles; activated keratinocytes	Lysis of the nail bed epithelium; neonatal death around P3	Pachyonychia congenita-like nail lesions Unknown	***97*** (*see* **Fig. 3E**)
K8	K8	Simple epithelia	a) Embryonal death ED12.5; trophoblast giant cell fragility (C57BL/6x129X1SvJ) b) Colorectal hyperplasia, colitis, rectal prolapse, predisposition for liver injury (FVB/N)	Ulcerative colitis	***53–55***

n.d; not done.

[a]Orthologous human and mouse keratin genes of each clusters I and II are compared in a separate table **Tables 1** and **2** follow the new and open nomenclature based on the principles of the HUGO Gene Nomenclature Committee. A comparative table aligning the old nomenclature to the novel system is available as a guideline in Magin et al. ***(4)***. The major expression and knockout phenotype is briefly outlined in all cases where keratin-deficient mice are generated and characterized. Double- and triple-keratin knock outs are shaded in gray. The human pathology that could be deduced from a mutated keratin gene is briefly outlined but does not inevitably coincide with the phenotype of mice with a targeted deletion in the orthologuous keratin gene. (C57BL/6x129X1SvJ) and (FVB/N) are mice strains.

ulcerative colitis and Crohn's disease ***(51,52)***. Crohn's disease and ulcerative colitis, together known as inflammatory bowel disease, are chronic illnesses of unknown origin. The inflammation within the intestinal tract (within the colon in ulcerative colitis and anywhere from the mouth to the anus in Crohn's disease) leads to some or all of the following clinical symptoms: diarrhea (with or without blood), abdominal pain, fever, and fatigue.

1.1.5. Phenotypes of Keratin Knockout and Transgenic Mice

1.1.5.1. Simple and Embryonic Epithelia

Ablation of keratins expressed in simple epithelia and during embryogenesis only partially helped to identify their functional significance. Side effects because of the strain background apparently influence the phenotypical outcome of generated knockouts. Keratin K8-deficient mice, initially thought to die from hepatocyte fragility leading to fetal liver bleeding at ED12.5 in the C57BL/6 strain ***(53,54)***, have been later demonstrated to collapse from placental malfunctions ***(55)***. However, keratin K8 deficiency crossed into FVB/N strain results in viable offspring featuring colonic hyperplasia, diarrhea, and colitis ***(53,54)***. Mistargeting of apical H^+/K^+-ATPase-β, anion exchanger AE1/2 and the Na^+ transporter might lead to the observed phenotype, possibly because of the net Na^+ absorption associated with the Cl^- secretion in jejunal epithelia ***(33,56)***. The colonic hyperplasia and colitis were later determined as chronic spontaneous helper T-cell 2 colitis that probably arises from colonic bacterial infection ***(57)***. Either distinct modifier genes or the second type II keratin, K7, expressed early during development, might compensate for the loss of K8 in FVB/N K8(–/–) mice **(Tables 1** and **2)**.

Hepatocytes from K8(–/–) mice are more susceptible to mechanical stress and hepatotoxic drugs and secrete less bile acids and ecto-ATPase, a phenomenon that is mechanistically still unexplored ***(58)***. Concanavalin A (ConA)-induced liver damage left K8(–/–) [and also K18(–/–)] epithelial cells approx 100 times more sensitive to TNF-induced cell death and increased both downstream targeted JNK- and nuclear factor-κB-protein levels. ConA-induced immunoresponse through activated T-cells was therefore thought to release TNFα, acting on the TNF receptor (TNFR)2 ***(59)***. Moreover, glutathione *S*-transferase-fusion protein binding assays suggested an interaction of the N-terminal parts of K8 and K18 with the TNFR2 ***(59)***. Primary hepatocytes from K8(–/–) mice being more sensitive to Fas-mediated apoptosis fall into the same category ***(60)***. Although transported by microtubules, Fas receptor abundance at the cell surface was increased in K8(–/–) mice, thus implying K8 to contribute to Fas receptor distribution ***(60)***. Concerning other keratins, a direct interaction between the TNFR1-associated death domain protein (TRADD) and the coil 1a domain of K18 and K14 was

demonstrated by coimmunoprecipitation ***(61)***. TRADD, the adaptor protein linking FADD and RIP to the TNFR1, is thought to be sequestered by K8 and K18 IFs, thereby precluding TRADD to operate as a mediator of apoptosis and thus modulating TNFR1 signaling. This type of explanation also was used as a paradigm for the interpretation of results after targeted deletion of keratins suspected to be essential for embryonic development. The K8 deficiency and subsequent embryonic lethality at ED12.5 in the sensitive C57BL/6 and 129Sv background is caused by trophoblast giant cell layer failure, suggested to result from TNF-sensitive failure of trophoblast giant cell barrier function ***(55)***. Maternal TNFα was thought to trigger an apoptotic response in extraembryonic trophoblast giant cells becasue after the loss of K8/K18 IF network, cells can no longer moderate apoptosis by sequestering TNFR1 and TRADD ***(55)***.

Targeted deletion of K18 in C57BL/6 mice was less dramatic than that of K8. Keratin K18(–/–) animals are viable, fertile, and show a normal life-span. Old K18 null mice, however, developed a distinctive liver pathology with abnormal hepatocytes containing K8-positive aggregates, identified as Mallory bodies, a hallmark of human alcoholic hepatitis ***(34)***. Absence of K8 and K18 in hepatocytes furthermore disturbs the function of at least some cell cycle regulators, such as the signaling integrator 14-3-3, and causes disturbances in cell cycle regulation and aberrant cytokinesis, which might drive cells into the S-to-G_2 transition ***(62)***. Except for hepatocytes, in which K19 is not expressed, perfect IFs between K8 and K19 have been found in these mice, underlining that functional redundancy is common in keratin biology ***(34)***. The obvious lack of defects in K19 deficient mice also falls into this category ***(63)***.

To identify the role of keratins during embryogenesis, double-deficient animals for K8 and K19 ***(63)*** as well as K18 and K19 ***(64)*** have been generated **(Tables 1** and **2)** to get rid of nearly all K8/K19 or all K18/K19 IFs [Note that in K8(–/–)/K19(–/–) mice, a compensatory IF network composed of K7 and K18 is formed.] Mice from both lines die concomitantly between ED9.5 and 10 but from different phenotypes. In K8(–/–)/K19(–/–)-embryos, a decreased number of disorganized labyrinthine trophoblast and spongiotrophoblast cells will lead to the penetration of maternal as well as embryonic blood into lesions of the placenta ***(63)***. In K18(–/–)/K19(–/–)-embryos, a rupture of giant trophoblast cells because of mechanical fragility, provoked by the maternal blood pressure, lead to hematomas and deformations of the yolk sac, which might disturb the nutrition of the embryo ***(64)***. Very recently, this defect could be rescued by aggregation of double-deficient embryos with tetraploid wild-type embryos up to ED11.5 that underscored the necessity of keratin filaments for the proper function of extraembryonic (like trophoblastic and placental tissue) rather than embryonic epithelia ***(65)***.

A possible explanation for the early embryonic death of K8 (–/–)/K19(–/–) embryos was in line with that of the single K8 knockout ***(55)***. An increment

of the TNF- or Fas-mediated apoptosis caused by a putative loss of interaction with the K8-mediated IF network might lead to an atrophy and degradation of the placenta ***(59,61)***. This effect possibly could be proven after rescuing of the embryonal lethality in a TNFα-deficient mice strain.

1.1.5.2. Basal Stratified Epithelia

In basal keratinocytes, three keratins (K5, K14, and K15) are coexpressed ***(8)***. They are linked to hemidesmosomes via plectin and bullous pemphigoid antigen-1 ***(66,67)*** and to desmosomes via desmoplakin ***(68)*** and plakophilin ***(69)***. To investigate the function of these keratins in the basal epidermis, mouse models have been established by gene targeting: K14-deficient mice, K14(–/–), of a mixed 129Sv/C57BL/6 background generally die within 48 h after birth (although there are some exceptions surviving up to 3 mo). Abundance of K15, however, is normally compensating for K14, but it is too low in neonatal mice ***(35)***. An age- or tissue-dependent upregulation of K15 in a few K14(–/–) mice may permit their survival up to 3 mo ***(35)***.

Subsequently, deletion of the partner keratin K5, being the only type II keratin protein in most basal keratinocytes, should abolish the formation of Ifs, which was done in a mixed 129Ola/BALB/c strain. Indeed, all mice die immediately after birth **(Fig. 3F)** containing an extensive cytolysis in almost all basal keratinocytes ***(21)***. Residual type I keratins K14, K15, and K17 (now without partner keratin) aggregated along hemidesmosomes in K5(–/–) mice. Because K5(–/–) pups are more severely affected than K14(–/–) mice, it is likely that K5 null mutations may be lethal in human EBS patients **(Fig. 3F)**. In contrast to K14-deficient mice, K5(–/–)-mice showed a strong induction of the wound-healing keratins K6 and K10 in the suprabasal epidermis of the cytolyzed compartments ***(21)***. Very early during embryogenesis of K5(–/–)-mice, K8 is able to compensate for K5 until its expression is downregulated at ED14.5 ***(32)***. Thus, an embryonic keratin (K8) is at least partially able to compensate for an epidermal keratin, provided that it is expressed in the same spatiotemporal manner. Confirmation for this hypothesis can be drawn from transgenic rescue experiments, in which the hK18 transgene was expressed in the basal epidermis of K14(–/–) mice. Although filaments between K5 and hK18 have been generated, they were unable to resist against mechanical stress ***(70)***.

Another, more sophisticated transgenic experiment imitated the inducible expression of the most severe Dowling–Meara-related K14 mutation of EBS [(R125C) in mouse (R131C) ***(71)***, 1995, OMIM 131760] in mice derived from a mixed 129Sv/C57BL/6 strain background. Heterozygous mice seemed normal with no gross skin phenotype and no microscopic blisters. Homozygous mice, however, were normal at birth but developed extensive blisters on the front legs, paws, and chest and died within 2 d ***(72)***. An inducible transgenic mouse model

was established after mating these mice with a strain expressing a truncated progesterone receptor-1–Cre-recombinase fusion protein under the control of regulatory elements of either the K5 or K14 gene promoter ***(73)***. Upon induction with antiprogestin or RU486 only in those cells expressing either K5 or K14, the Cre-recombinase will be induced to excise the floxed neo-cassette inserted after the first exon of the mutated K14 transgene ***(72)***, thus activating the transgene. Correspondingly, blisters filled with fluid on the front legs and paws became visible but healed after 2 wk. The unexpected healing process was explained with the partial activation of the Cre-recombinase in the nonstem cell population of basal keratinocytes. The migration and reepithelialization by nonphenotypic stem cells (containing the floxed neo-cassette) or transiently amplifying cells derived from these stem cells of the untreated surrounding areas was finally considered as the working hypothesis ***(72)***.

1.1.5.3. Suprabasal Stratified Epithelia

In the upper epidermis, 60% of the total protein content consists of keratins K1 and K10 ***(74)***. Surprisingly, epidermal integrity was maintained in neonates after targeted disruption of K10 **(Fig. 3G)**. Adult K10(–/–) animals showed a fivefold increase of basal cell proliferation and hyperkeratosis ***(75)*** In the suprabasal cell layer of the neonatal skin, an IF network composed of K1, K5, and K14 keratins compensates for the loss of K10 **(Fig. 3G)**, whereas within the adult skin, an induced upregulation of the cytokeratins K6a, K6b, K16, and K17 might take over at least some of the functions of K10-containing intermediate filaments ***(75)***. Moreover, an induction of cyclin D1 and of c-myc in basal cells as well as that of the cell cycle regulator 14-3-3 in postmitotic keratinocytes was noted ***(76)***, implying an involvement of K10 in the regulation of epidermal cell proliferation ***(77)***. Cell transfection experiments indicated the K10 head domain to sequester both the Akt as well as the atypical protein kinase Cζ kinases avoiding their translocation to the plasma membrane, thus representing a negative regulator of epidermal cell cycle ***(78)***. Therefore, K10 was proposed to suppress cell proliferation; however, the K10(–/–) strain contradicts this view because a general increase of suprabasal cell proliferation could not be observed ***(79)*** except for an increase in sebocyte proliferation and differentiation. Sebaceous gland cells of K10(–/–) mice showed an accelerated turnover and secreted more sebum, including wax esters, triglycerides, and cholesterol esters. It was concluded that the suprabasal intermediate filament cytoskeleton devoid of keratin K10 increased the differentiation of epidermal stem cells toward the sebocyte lineage ***(79)***. Furthermore, K10(–/–) mice treated with 7,12-dimethylbenz[a]anthracene/12-*O*-tetradecanoylphorbol-13-acetate developed far less papillomas than wild-type mice. 5-Bromo-2′-deoxyuridine-labeling revealed a strongly accelerated keratinocyte turnover in K10(–/–) epidermis,

suggesting an increased elimination of initiated keratinocytes at early stages of developing tumors. The concomitant increase in K6a, K16, and K17 in K10-deficient epidermis and the increased motility of keratinocytes is in agreement with the pliability versus resilience hypothesis, proclaiming that K10 and K1 render cells more stable and static. The K10(–/–) knockout represent the first mouse model showing that loss of a cytoskeletal keratin reduces tumor formation. This response might be achieved by an accelerated turnover of keratinocytes, possibly mediated by activation of mitogen-activated protein kinase pathways ***(80)***.

Expression of a truncated and dominant negative form of K10 ***(81,82)*** produced skin lesions comparable with severe forms of epidermolytic hyperkeratosis (EHK) (*see* **Fig. 3C**). Moreover, homozygous mice (K10T/K10T) revealed changes in the composition of epidermal lipids and barrier function ***(83,84)***. Transgenic mice, generated as described for K14 ***(72)***, harboring a K10 point mutation (R154C, equivalent to R156C in the human), produced local blisters and scaling at the sites of induction that persisted later in life together with unaffected epidermis. Apparently, epidermal stem cells have been targeted in these mice. In agreement with mosaic form of EHK in humans, which is because of postzygotic somatic mutations, this finding shows that in the absence of selective pressure—the mutant K10 is silent in presumptive stem cells—mutant and wild-type stem cells can coexist. This idea has to be considered in therapy approaches of EHK; they must include either ablation or correction of mutant stem cells ***(85)***.

Generally, extensively bundled IFs consisting of K1/K10 parallel the surface of flatted keratinocytes, being covalently cross-linked to cornfield envelope (CE) proteins ***(86)***. However, K10 mutations, resulting in EHK, do not affect CE integrity ***(87)***, which is in contrast at least to some mutations found in K1. K1 frameshift mutations resulting in the shortening of its glycine-rich V2 tail domain generate the dominant inherited skin disorder ichthyosis hysterix ***(88,89)***, most likely caused by the inability of the occurring IFs to form bundles. Loricrin, but not involucrin and filaggrin, was found irregularly distributed throughout the cytoplasm of keratinocytes, thus pointing to a specific role of the K1–loricrin interaction during CE formation ***(90)***. Although the targeted ablation of keratin K1 is not published yet (Meier-Bornheim et al., in preparation), an interesting aspect to clarify might be if the distribution or assembly of proteins directed to the CE is also disturbed after the loss of K1 within IFs.

The ectodermal dysplasia called pachyonychia congenita (PC) affects epithelial appendages and results in profoundly dyskeratotic nails **(Fig. 3E)**. It can be divided into type 1 (Jadahsson–Lewandowski) PC characterized by oral leukoplakia and palmar–plantar keratoderma and type 2 (Jackson–Lawler) PC typified by subepidermal cysts, natal teeth, but generally milder nail lesions ***(91)***. Type 1 PC is

caused by keratin K6a, K6b, or keratin K16 mutations, whereas type 2 is generally evoked by keratin K17 mutations (*see* www.interfil.org).

Initially, it seemed easy to mimic the corresponding pathology in mice after targeted disruption of the corresponding keratin orthologs. However, K6a-deficient mice do not show any spontaneous skin or oral mucosa lesions ***(92)***, and K17-deficient mice only develop a strain-dependent (C57BL/6) and reversible alopecia, but they histologically fail abnormalities in oral mucosa, glands, footpad epidermis, and nails ***(36,93)***. Even mice being double deficient for both K6 gene isoforms a and b contain histologically normal skin epithelial appendages ***(94)***. This obvious discrepancy between the mentioned null mouse models and the PC disease led to the idea that other keratins, such as K5, K6hf, K14, K16, or K17n known to be expressed in nail bed epithelium ***(36,95,96)***, might compensate their functional loss. Support for this concept of functional redundancy was provided after targeted deletion of three keratin genes (K6a, K6b, and K17) by intercrossing the existing keratin-deficient strains to generate the corresponding K6a/K6b/K17 triple-null mouse model ***(97)***. These mice also died, as expected from the double-deficient K6a and K6b mice, shortly after birth but they exhibited severe lysis to the nail bed epithelium, where all the three genes are abundantly expressed. Moreover, a decrease in the level of a fourth keratin protein, K16, was observed ***(97)***. Conclusively, the loss of these three keratin genes strongly suggests that the nail bed epithelium is initially targeted in PC (*see* **Fig. 3E**).

Wound healing, proliferation, and migration of keratinocytes harvested from K6a(–/–) animals seemed unaffected compared with wild-type controls ***(92)***. Targeted deletion of both keratin genes K6a and K6b in mice with different genetic background produced conflicting but interesting results. All K6a(–/–)/K6b(–/–)-double deficient mice with a mixed 129Sv-Ck35/C57BL/6 background die about 1 wk after birth, showing growth delay and weight loss associated with reduced milk intake, white plaques, and extensive cell blistering in the posterior region of the dorsal tongue and the upper palate epithelia combined with traumatic filiform papillae in the anterior compartment ***(94)***. The anterior compartment of filiform papillae, normally expressing K6a and K6b, is devoid of keratin IFs in these double knockouts ***(94)***. In contrast, 25% of K6a(–/–)/K6b(–/–) double-deficient mice with a mixed AB2.2/129SvEv-C57BL/6N background survived to adulthood with normal hair and nails (*see* **Tables 1** and **2**). Nevertheless, 75% of these double-deficient mice die within the first 2 wk of life because of disintegration of the dorsal tongue epithelium causing a plaque of cell debris that severely impairs feeding ***(96)***. These authors cloned a third, and at that time unidentified, murine keratin 6 gene (designated K6hf) likely to be the ortholog to human K6hf expressed in hair follicles. Redundancy of this novel K6hf gene in K6a/K6b-deficient follicles and nails is an attractive explanation for the

absence of hair and nail defects and also for the, at least partial survival, of K6a(–/–)/K6b(–/–) double-deficient mice ***(96)***. However, other modifier genes deviating in both genetic backgrounds could influence either directly or indirectly the abundance and expression pattern of yet unknown genes or the novel K6hf gene itself.

The contribution to wound healing of K6a and K6b also was investigated by keratin explant assays, indicating that the lack of these keratins renders keratinocytes more motile, including changes in the actin filament organization that might support either an increase in cell fragility or a more rapid wound healing ***(98)***. In contrast, mice overexpressing K6a proteins with different C-terminal truncations were largely found with severe blistering and died shortly after birth. Those mice that survived showed lesions in the interfollicular epidermis and hair follicles and developed severe alopecia followed by the destruction of the outer root sheath (ORS) cells. The latter finding indicates that the innermost ORS cells are uniquely sensitive to expression of even slightly altered K6a proteins, suggesting that mutations affecting the human K6a ortholog expressed in this cell layer could result in alopecia in humans as well ***(99)***.

Although the precise mechanism is far from being understood, the relative amounts between K6a, K6b, K16, and K17 seem to be important to establish a less stable but functional keratin cytoskeleton. It is also likely that these keratins, but especially K6a and K6b, provide attachment sites for associated proteins ***(40,96)***. The transgenic expression of human K16 in the basal epidermis of K14(–/–) mice led to a rearrangement of the keratin Ifs and normalized the epidermis in neonatal animals, but these animals later developed extensive alopecia and chronic epidermal ulcers in areas of frequent physical contact, and alterations in other stratified epithelia ***(100)***. As control, mice expressing a K14 cDNA under regulatory elements of the K16 gene also rescue the blistering phenotype of the K14 null mice with only a small percentage exhibiting minor alopecia. Despite their high sequence similarity, K16 and K14 are not functionally equivalent in the epidermis and other stratified epithelia, and it is primarily a carboxy-terminal segment of approx 105 amino acids of K16 that defines these differences ***(100)***. Moreover, comparative data from skin explant assay and wound healing studies revealed a delay in wound healing as a result from hair shaft fragility and apoptosis in hair matrix cells. This pathology reverted to normalcy as a consequence of the K14 transgene expression, implying that the composition of the keratin cytoskeleton can influence cell migration ***(101)***.

K17 predominantly occurs in epidermal appendages ***(36,41)***, but it also occurs in basal cells of stratified epithelia ***(4)***. It forms IFs with K5 and any of the different K6 proteins currently identified ***(93)***. In humans, K17 mutations

are responsible for the nail disorder PC type 2 and for the sebaceous gland disorder steatocystoma multiplex ***(102)***. K17(–/–) mice do not develop a proper fur coat during the first postnatal week. The alopecia is likely to be caused by hair shaft fragility and apoptosis in the hair matrix cells being reverted within the first postnatal hair cycle at 3 wk of age, coinciding with the presumably compensatory upregulation of K16 ***(93)***. The alopecia detected in K17 null animals was only featured in the mouse strain C57BL/6. The unexpected lack of a nail pathology in K17 null mice was most recently explained by the expression of K22 ***(3,36)***, which is a pseudogene in humans ***(3)***.

1.1.6. Lessons From Loss of Keratin Genes and Redundant Expression of Other Family Members

The disruption of keratin IFs in the K5(–/–) and K18(–/–)/K19(–/–) mice ***(21,63,64)***, their changed composition in K6a(–/–)/K6b(–/–) and K14(–/–) mice ***(35,94,96)*** as well as the expression of dominant negative mutations in K10 and K14 ***(72,81,85)*** generated phenotypes that were comparable. However, the molecular mechanisms that might lead to cell and tissue fragility are either diverging or not clarified yet. The absence, i.e., of the keratin K10 protein ***(75)***, has only subtle effects compared with the dramatic outcome of the presence of a mutated (in this case, truncated) form of keratin K10 ***(81,85)***. Often, an inserted mutation more readily leads to a disorganized and aggregated keratin IFs than merely the loss of the corresponding keratin. Within internal epithelia of the liver or pancreas, keratins are thought to be nonessential cytoskeletal proteins but rather function as scaffolds, providing proper topology for other accessory proteins ***(59,61)***. In some cases, certain keratins can compensate for each other; in other cases they cannot. Keratin K19 seems to fulfill the function of keratin K18 and vice versa, because they have the same partner keratin K8 ***(34,61)***. In other cases, the substitution of K14 by K18 has only partially rescued the phenotypical outcome of the K14 deficiency in mice ***(70)***, probably because of an inappropriate expression level compared with its partner keratin K5. The replacement of K14 by a K16/K14 hybrid protein seemed to achieve and expression level of the same abundance as the partner keratin K5, which rescued the K14 phenotype to a large extent ***(100)***. With respect to the genetic background in different mouse strains, there are many deviations found in the resulting phenotypes of the keratin-deficient mice. The K8 deficiency within the C57BL/6 background is 100% lethal, whereas the loss of this keratin protein in FVB/N mice allows survival of ~30% of the affected animals ***(53)***. Several different modifier genes might help to compensate for the loss of a certain keratin in one strain, or there is a beneficial expression profile of redundant keratins that help to overcome the often devastating consequences of a vanished keratin.

1.2. Methodological Considerations

1.2.1. Design of Targeting Vectors

Targeted gene manipulation can be accomplished by the use of either replacement or insertion vectors ***(103)***. Replacement vectors are used to disrupt target gene function by deleting specific gene sequences and by replacing them with heterologous DNA, usually a drug selection gene or a marker gene to analyze target gene expression. Insertion vectors are used to disrupt gene function by inserting heterologous DNA, but they also allow for the introduction of more subtle genetic alterations such as point mutations. Here, we focus first on replacement vectors and describe approaches to modify a gene or gene fragment constitutively or conditionally. Then, we describe how large-scale genomic deletions are carried out using insertion vectors. All gene-targeting vectors should be constructed from isogenic DNA (i.e., from the same mouse strain as the embryonic stem [ES] cell line to be used). Deviation from this rule will decrease targeting frequencies by 10- to 1000-fold.

1.2.2. Constitutive and Conditional Alterations

Most IF genes have a gene structure well suited for gene targeting, because they span approx 10 kb of genomic DNA. To generate a null allele or to introduce consecutively several independent mutations, it is desirable to delete the coding sequences completely from the genome. As long as no more than 20 kb of genomic DNA is replaced, the efficiency of gene targeting remains high. A functional null allele also can be generated by deleting the core promoter and the first coding exon. In such a case, alternative start codons should be considered that can be removed by site-directed mutagenesis. **Figure 4A** illustrates typical replacement-type targeting vectors carrying two segments of sequences homologous to the gene of interest. A short arm of ~1.5 kb to allow for easy PCR detection and a long arm of 4–6 kb is required. For PCR detection, one PCR primer resides in the selectable marker gene, and the other primer resides outside the region of homology. Homologous arms can either be isolated from genomic libraries, or, more conveniently, by using long-range PCR from genomic DNA preparations or bacterial artificial chromosome clones carrying the gene of interest. Vector arms flank the part of the gene to be manipulated and carry between them a positive selectable marker gene. Upon homologous recombination between vector arms and the corresponding genomic DNA, the selectable marker gene replaces the genomic sequences and creates a modified gene. In our hands, this approach has resulted in targeting frequencies between 1:5 and 1:35 positive ES cell clones ***(21,34,75,81,104)***. The selectable marker genes that code for resistance against neomycin, hygromycin, puromycin, or hypoxanthine phosphoribosyl transferase (HPRT) are most commonly used. In

the latter case, HPRT-deficient ES cells are required. Only ES cells that harbor an HPRT-minigene can grow in hypoxanthine, aminopterin, and thymidine (HAT)-containing selective medium. The major advantage of the HPRT system is that it allows for negative selection in medium containing 6-thioguanine (6-TG) ***(103)***. In our experience, we have not observed any advantage for the use of negative selectable marker genes outside the homology arms.

In most cases, selectable marker genes with their own promoters, usually PGK-1, are used. In rare instances, the phosphoglycerate kinase (PGK)-1 promoter becomes silenced because it carries a partial pCpG island and may have to be replaced by other promoters ***(105)***. To minimize transcriptional interference between the gene's own promoter and the minigene promoter, we recommend placing the selectable marker gene opposite the target gene's transcriptional orientation. To minimize potential effects of the selectable marker gene on neighboring genes, and on the phenotype of mice, it is desirable to remove it. This removal can be done by placing Lox P or FRT recognition sequences at both ends of the selectable marker cassette. After transient expression of Cre or Flp-recombinase in ES cells, the marker gene is removed and ES cells become sensitive to the selective agent. This selectivity allows the use of the original targeting vector to create ES cells homozygous at the targeting locus ***(106,107)***. If conditional gene targeting is planned, it is advisable to use FRT sites to tag the selectable marker gene **(Fig. 4B)**.

For those IF genes that are transcribed in ES cells, very high gene targeting frequencies can be achieved using promoterless selectable marker genes that are placed in the first coding exon, either in frame with the endogenous ATG codon or by positioning the initiation codon of the marker gene downstream from the major transcription start site. If insertion of a promoterless marker gene is required in an exon further 3′ into the gene, an IRES ***(108)*** can be incorporated to enable translation of the marker gene. A slight modification of the aforementioned design can be used to introduce a promoterless reporter gene along with the selectable marker to follow the cells carrying the targeted allele. The enhanced green fluorescent protein- or the more sensitive lacZ-reporter genes are convenient but do not necessarily provide quantitative readouts ***(109,110)***.

There are several ways to introduce subtle alterations, i.e., point mutations, into genes. The mutation can be positioned in one of the homology arms of the targeting vector and be cotransferred along with the selectable marker, provided recombination does not separate the mutation from the selectable marker **(Fig. 4B)**. The selectable marker gene has to be positioned either in a noncoding exon or in an intron, from where it can be removed using Flp or Cre-recombinases, leaving one recombinase recognition site in the genome. In general, cotransfer of mutations work well, provided the distance between the marker

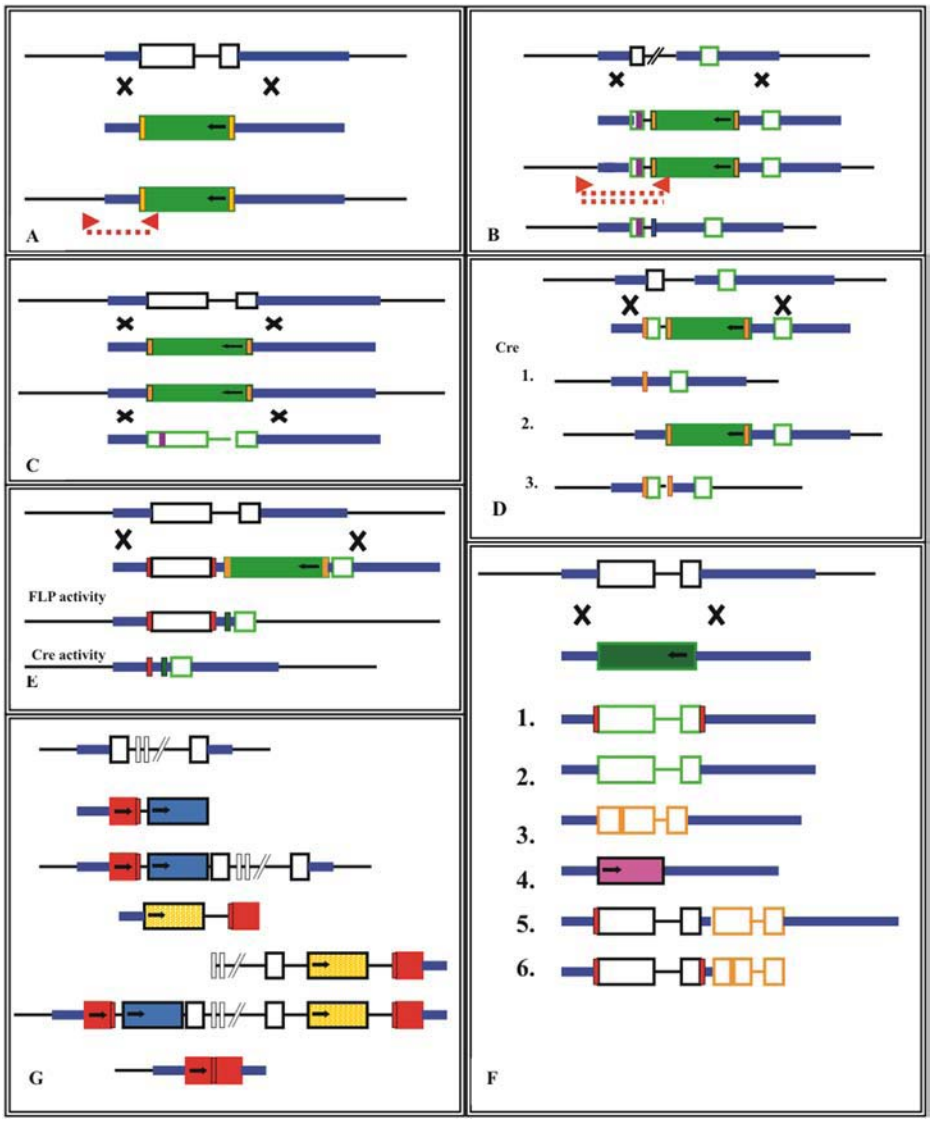


Fig. 4. **(A)** Conventional gene-targeting experiment for creation of null allele. Top, genomic locus with two exemplary exons (open boxes). Sequences homologous to replacement-type gene-targeting vector are in black. Middle, targeting vector with short and long arms of homology, flanking a selectable marker gene (gray). The arrow indicates the marker gene promoter and its transcription direction. X indicates regions of crossover. Bottom, targeted locus at which the marker gene has replaced two exons. Arrowheads, primers used for PCR detection of homologous recombinants. Dotted line, predicted PCR product. **(B)** Introduction of subtle alteration by co-transfer. Top, genomic locus. Middle,

Fig. 4. *(Continued)* targeting vector. The short arm of homology carries the desired alteration (upright dark gray bar in the 5′ open box) created by mutagenesis. If in addition, a new restriction site has been generated without altering the properties of this sequence, presence of the desired mutation can be directly screened by restriction analysis of PCR products. Middle, target locus after homologous recombination. Presence or absence of desired alteration is indicated by dotted lines before and after restriction of PCR products. Bottom, target locus after deletion of marker gene by site-specific recombinase. One recognition site resides in the genome. **(C)** Double replacement gene-targeting procedure by using positive/negative selection. Top, target locus. Middle, gene-targeting vector. Middle, after the first gene-targeting step, the gene or gene fragment has been replaced by an HPRT-minigene (gray box), providing resistance against HAT-containing media. The target locus can be replaced by modified gene sequences. Below, second step gene targeting to introduce a subtle alteration (upright bar) in the left exon. Selection in media containing 6-TG (thioguanine) indicates loss of HPRT-marker gene. **(D)** Conditional gene targeting by using three Lox P recognition sites. Top, target locus. Middle, gene-targeting vector with lox P site flanking the exon to be removed upon Cre activation (upright gray boxes, lox P recognition sites). The lox P-flanked marker gene is placed in an intron. After Cre activity, several types of deletions can occur. Type I, complete excision creating a null allele (type I), Type II undesired backwardness of the marker gene. Type III, excision creating a conditional allele. The outcome is determined by various parameters, e.g., the relative position of the lox P sites and the extent of Cre activity. **(E)** Conditional gene targeting by using FRT and lox P recognition sites. Top, target locus. Gene-targeting vector with FRT sites (light gray upright boxes) flanking the marker gene (gray) and two lox P sites (gray upright boxes) flanking the exon open box to be deleted. After homologous recombination and FLP activity, the marker gene has been removed with one FRT site left behind (black). Upon Cre activity, the "floxed" exon is deleted with one lox P site left behind. **(F)** Gene modifications by using double replacement gene targeting. Top, target locus. Bottom, gene-targeting vector. Below, after the first gene-targeting step, the gene or gene fragment has been replaced by an HPRT-minigene, providing resistance against HAT-containing media. The target locus can be replaced by modified gene sequences as indicated below. (1) Replacement with a conditional floxed allele. (2) Replacement with a related gene (open box). (3) Replacement with a mutant gene copy (gray box). (4) Replacement with an inducible, promoterless Cre- or FLP-recombinase to be driven by the target gene promoter. (5) Replacement with a conditional, silent gene copy, followed by a wild-type copy of a related gene (gray box). (6) As before, but followed by a mutant (bar) copy of a related gene. **(G)** Generation of large-scale genomic deletions, based on a HPRT-deficient ES cell line. Top, target locus. Below, targeting vector for 5′ flank of targeting vector, consisting of a nonfunctional half-HPRT-minigene (small gray box, arrow indicates minigene promoter), followed by a lox P site (upright dark gray box) and a neomycin resistance gene (dark gray framed box, arrow marking promoter). After homologous recombination, G418-resistant ES cell clones are identified by PCR. Below, targeting vector for 3′ end, consisting of a puromycin resistance gene (light gray box) and the 3′ half of a nonfunctional HPRT-minigene (dark gray), preceded by a lox P site (dark gray upright box). Below, after homologous recombination, puromycin-resistant ES cell clones are identified by PCR. Complex rearrangements can

gene and the alteration is shorter than between the alteration and the end of the homology arm. One potential drawback of this strategy is that the expression of the target gene can be severely altered by placing the marker gene in regulatory introns ***(85,111,112)***.

An alternative strategy follows the so-called double replacement procedure ***(113–115)*** **(Fig. 4C)**. In the first round of gene targeting, an HPRT-minigene replaces the (part of the) gene to be modified. After the first round of correct targeting, the minigene is replaced by a modified version of the original gene or by another sequence. In the second step, selection with 6-TG allows cell survival only upon loss of the HPRT-minigene. If HPRT-deficient ES cells are not available, a combination of two selectable marker genes for positive and negative selection has been described previously ***(113,115)***. In principle, this procedure allows the introduction of many point mutations by using the same second-round gene-targeting vector. In contrast to Cre-Lox or FRT/Flp procedures, no foreign DNA remains in the target locus **(Fig. 4D)**.

With the use of the site-specific recombinases, conditional cell type and temporal gene alterations have become feasible ***(116–120)***. The original vector design used a Lox P-flanked marker gene and a third Lox P site placed in an intron such that the Lox P sites flank an exon to be deleted ***(117)***. After Cre-recombinase activity, three types of excisions can occur **(Fig. 4D)**, of which complete and type II excisions are the desired excisions and are selectable by the loss of the marker gene. The type II excision provides a conditional allele that can be deleted either in ES cells or in vivo, after mating with a strain of mice expressing the Cre-gene under the control of an appropriate promoter. Because type II excisions can be difficult to obtain, an improved strategy relies on the combined use of Lox P and FRT sites. FRT sites are used to flank the selectable marker gene and allow its removal by the Flpe gene ***(121)***, either in ES cells or in vivo. Two Lox P sites are placed at intronic or noncoding exon positions to delete (parts of) the gene of interest **(Fig. 4E)**. Conditional alleles created in this way can be controlled in a temporal manner by using an inducible Cre-recombinase. At present, tamoxifen, RU 486, and tetracycline-controlled Cre-recombinases, in combination with various promoters, are available. Currently, there are two limitations of these inducible transgenes, namely, leakiness at the uninduced stage and the time to remove the target gene product after Cre expression. It can take several days for this process, especially for IF proteins with their long half-lives ***(122)***. The use

Fig. 4. *(Continued)* occur after Cre activity, depending on the cell cycle phase in which Cre is active and the chromosomal integration of the two targeting vectors ***(123)***. Below, after Cre activity, the marker genes and the genomic sequences between the two lox P sites are deleted and surviving, correctly targeted ES cell clones are resistant against HAT and sensitive against G418 and puromycin.

of Cre- and Flp-recombinases can be combined with double replacement to allow a variety of options **(Fig. 4D,E,F)**.

1.2.3. Large-Scale Genome Engineering

If it becomes necessary to delete larger genes or gene fragments from the genome, the approaches described above become inefficient as the distance between Lox P sites increases. To create deletions of a large segment of chromosomal DNA, Bradley and colleagues developed the following scheme, based on two consecutive homologous recombination events, one at each end of the target locus. The 5′ vector consists of a neomycin resistance gene for gene targeting, a Lox P site, a 5′ half HPRT minigene (nonfunctional HPRT fragment), and approx 6- to 10-kb of genomic DNA homologous to one end of the target locus **(Fig. 4G)**. Similarly, the 3′ vector contains a puromycin resistance gene, a Lox P site, a 3′ HPRT-minigene (nonfunctional HPRT fragment), and approx 6- to 10 kb of genomic DNA homologous to the other end of the target locus ***(123)***.The Lox P site has been placed in an intron of the HPRT-minigene. These vectors can be generated conventionally by sequential insertion of various genetic components into a plasmid construct ***(123)***, or they can be isolated directly from genomic libraries of premade targeting vectors (*see* http://www.ensembl.org/ and http://www.ensembl.org/Mus_musculus/). Depending on the orientation of the LoxP sites, the location of the Lox P site on the two alleles, and whether the recombination takes place between sister chromatids or between the nonsister chromatids in the G_2 phase of cell cycle (*see* **ref.** ***123***), various recombination events, such as deletion, inversion, and duplication of the region between the LoxP sites, can occur. Upon Cre activity, recombination between the Lox P sites reconstitutes a functional HPRT-minigene. This HRRT-minigene provides a positive selection for this event when the ES cells are cultured in HAT medium.

1.3. Culture of ES Cells

ES cells are available from several laboratories and from commercial suppliers. The majority of presently available lines has been derived from one of several mouse 129 inbred strains and maintain good germline transmission capabilities. C57BL/6 ES lines are also available, but they seem more susceptible with respect to germline transmission. Recently, ES cells from F1 or mixed hybrid genetic background strains were isolated, and they show very high germline transmission frequencies because of hybrid vigor ***(124–126)***. These cells were developed to shorten the time required for the generation of gene-targeted mice considerably. The concept is based on the use of tetraploid embryos, generated by electrofusion of ES cells, for injection of genetically engineered diploid ES cells. Tetraploid cells have been shown to give rise exclusively to extraembryonic tissues, whereas diploid ES cells contribute to

the embryo proper ***(127)***. In consequence, this procedure leads to germline chimeras that are completely derived from ES cells, saving one breeding step, equaling 3 mo, along the route of establishing a line of genetically altered mice. This procedure can be shortened even further by carrying out two rounds of gene targeting to conditionally target both alleles of the gene of interest ***(126)***. If this gene targeting is performed in ES cells that express an inducible Cree gene, it is possible to generate mice with two conditionally modified alleles in 6 mo compared with 14 mo.

Many ES cell lines are grown on feeder cells, either primary embryonic (mouse embryonic fibroblast [MEF]) or immortalized STO cells (available from the The Jackson Laboratories). If feeder cells are used during drug selection, keep in mind that they must express the appropriate resistance gene. Alternatively, ES cells can be grown on gelatinized dishes. Several ES cell lines are available that maintain an excellent germline potential under these conditions, including the HPRT-deficient line HM-1, which is in extensive use in our laboratory ***(123,128–131)***. To maintain the germline transmission potential of ES cells, great care has to be taken with respect to their culture conditions, including rigorous testing of serum and growth factors and the use of reagents tested for ES cell work (*see* **Note 1**). Recently, the major signaling pathways responsible for the maintenance of ES pluripotency have been identified and have lead to the formulation of serum-free media with defined growth factors ***(130,131)***. In our hands, these conditions are favorable (*see* **Fig. 5**).

In **Subheading 2.**, culture of HM-1 cells in standard and serum-free media is described, both on gelatin-coated dishes and on feeder cells. Culture conditions for other ES cell lines are very similar; however, they may react differently on the same serum batches. It is recommended to follow the advice of the donating laboratory on the source of serum. In our hands, ES cell-tested serum batches of American, Australian, and New Zealand origin, offered by several companies, are preferable. Changing to serum-free conditions puts cells through a crisis that lasts three to four passages, after which the line should be expanded and frozen in aliquots ***(130,131)***.

2. Materials

2.1. Reagents and Solution Preparations

Self-renewal of ES cells is dependent on signals from the cytokine leukemia inhibitory factor (LIF) and from either serum or bone morphogenetic proteins (BMPs). Hence, two alternative cell media formulations can be used to culture ES cells. One formulation uses fetal bovine serum and the other uses the BMP-2. Here, we focus on the serum-free conditions to culture ES cells.

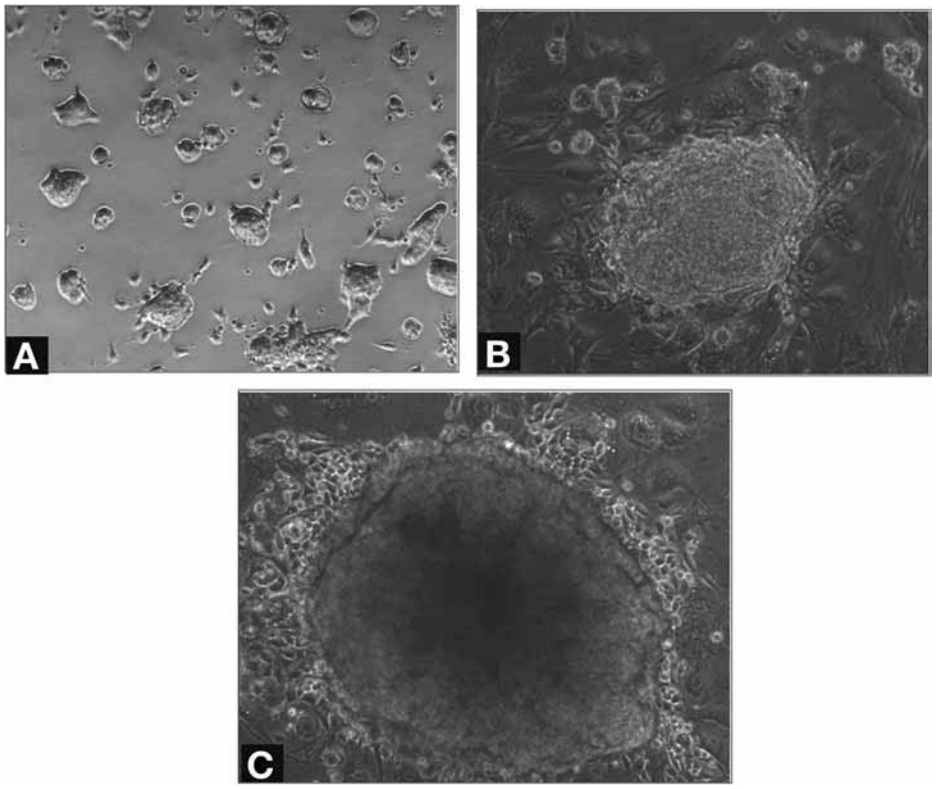


Fig. 5. **(A)** HM-1 ES cells plated at a low density (10×) grown on a gelatin-coated dish. Cells require another passage to increase individual colony numbers. **(B)** HM1 ES (original magnification ×10) colony after a first step gene targeting ready to be passaged and screened for homologous recombination event by using PCR. **(C)** Highly differentiated ES cells (original magnification ×25). Note the loss of discrete ES cell border, the formation of cobblestone-like cells, and ES cells differentiating into fibroblasts that extend outward from the colony. It is possible to attempt to rescue this colony by passaging several times; however, it is not usually recommended.

1. Components and preparation of 50X stock solutions:
   a. Dulbecco's modified Eagle's medium (DMEM)-Ham's F12 (cat. no. 11320-033, Invitrogen).
   b. Neurobasal medium (cat. no. 21103-049, Invitrogen).
   c. Supplement B27 (cat. no. 17504-044, Invitrogen) is 50X, so aliquot as 2 mL and store at –20°C.
   d. Bovine serum albumin (BSA) fraction V (cat. no. 15260-037, Invitrogen)-13.3 mL/100 mL medium (1:1). Aliquot as 10 mL and store at –20°C.
   e. Apo-transferrin (cat. no.11108-016, Invitrogen) 1 g/10 mL medium (1:1). Aliquot as 0.5 mL and store at –20°C.

f. Human insulin (cat. no. 0030110 SA, Invitrogen) stock solution at 4 mg/mL. Aliquot as 2 mL and store at –20°C.
g. Progesterone (cat. no. P6149, Sigma): dissolve 1 mg in 1 mL ethanol, add 49 mL of medium (1:1), and filter via 0.1-µm filter. Aliquot as 1 mL and store at –20°C.
h. Putrescin (cat. no. P6024, Sigma), dissolve in 6.2 mL of medium (1:1).
i. Na-selenite (cat. no. S 5261, Sigma) 10 m*M* stock I solution (0.865g/50 mL cell culture grade H_2O). To obtain stock II (100 µ*M*), take 1 mL of stock I solution and then complete to 100 mL with medium (1:1). Filter sterile (0.1-µm filter) aliquot as 2 mL and store at –20°C.
j. BMP-4 (cat. no. 314-BP, R&D Systems): take 23,30 µL of 25% HCl solution and complete to 100 mL with cell culture grade H_2O (4 m*M*), and then filter sterile (0.1-µm filter). To each BMP-4 vial add 1 mL of this solution and 1 µL of BSA fraction V. Aliquot as 100 µL and store at 80°C.

For 500 mL medium:

225 mL DMEM-F12, 225 mL neurobasal medium, 10 mL Supplement B27, 5 mL BSA fraction V, 500 µL APO-transferrin, 3.125 mL human insulin, 300 µL progesterone, 50 mL putrescin, 300 µL Na-selenite, 910 µL glutamine, monothioglycerol, 500 µL BMP-4, and 1 U/mL of final medium LIF. Store complete medium at 4°C and use for up to no more than 1 wk.

2. Gelatin solution: use Sigma G2500 swine skin type I gelatin. Prepare a 1% solution in cell culture grade water and autoclave. Leave overnight and autoclave again the next day and store at 4°C. Warm up before preparing 0.1% working dilution.
3. Trypsin: for 500 mL of trypsin solution, mix the following: 400 mL of cell culture grade sterile phosphate-buffered saline (PBS), 100 mL of sterile EDTA (1.85 g/L), and 5 mL of trypsin (cat. no. 25090, Invitrogen). Mix and store as 100-mL aliquots at –20°C. Once unfrozen, keep aliquots at 4°C.
4. ES cell freezing medium: although the ES cells are cultured in the absence of serum, in our experience it is advisable to freeze ES cells in the presence of serum *(see* **item 5**). On thawing, the cells are washed twice in serum-free medium before plating.
5. Serum-containing culture medium: Glasgow modification of Eagle's medium (BHK-21 medium) without tryptose phosphate broth with L-glutamine (cat. no. 041-01710, Invitrogen), 5% 100 m*M* sodium pyruvate (Invitrogen), 5% Nonessential amino acids (100X, Invitrogen), 5% Fetal calf serum (FCS) (Sigma), 15% Monothioglycerol (Sigma), 1:50,000 LIF, 1 U/mL medium. For homemade LIF, compare individual batches to commercial sources in a functional assay.

KSOM medium complemented with amino acids: 95 m*M* NaCl, 25 m*M* $NaHCO_3$, 2.5 m*M* KCl, 10 m*M* lactate, 0.35 m*M* KH_2PO_4, 0.2 m*M* pyruvate, 0.2 m*M* $MgSO_4$, 0.2 m*M* glucose, 1.71 m*M* $CaCl_2(2H_2O)$, 0.2 m*M* L-glutamine, 0.01 m*M* EDTA, 1 mg/mL BSA; phenol red; 100X nonessential amino acids (1:50); 50X essential amino acids (1:100).

6. 20 mL 2X Freezing medium: mix 10 mL of complete ES medium, 4 mL of ultrapure dimethyl sulfoxide, and 6 mL of FCS. Store at –20°C.
7. 10X HEPES-buffered saline (HBS) for electroporation: mix 16 g of NaCl, 0.74 g of KCl, 0.252 g of Na_2HPO_4, 2 g of D-glucose (dextrose), and 10 g of HEPES. Dissolve all in 180 mL of cell culture grade water, adjust pH to 7.2, bring up to 200 mL, and then filter through 0.1-μm pore filter. Store at –20°C.

To prepare 1X HBS, dilute with sterile water, which brings pH to 7.05. HAT and HT media are bought as 50X stocks from Sigma and used as if they were 100X concentrated. To prepare 100X HAT, dissolve hypoxanthine in 2 mL of 1 *N* NaOH, add 8 mL of cell culture grade water, and mix well. Add another 10 mL of water and mix until most of the hypoxanthine is dissolved. Add 1 mL of 1 *N* NaOH and mix until hypoxanthin is dissolved. Subsequently, add 75 mL of water, thymidine, and then 100 mL of water. Add aminopterin and mix again (aminopterin is a poison, so consult material safety data sheet). Filter through 0.1-μm filter and then store as 20-mL aliquots at –20°C.

8. 1 m*M* Aminopterin: weigh out 100 mg and dissolve in 45 mL of cell culture grade water. Add about 1 mL of 1 *N* NaOH to dissolve aminopterin, sterilize as in **item 7**, and then store at –20°C.
9. Additional reagents:
   a. 1X Trypsin inhibitor from *Glycine max* (soybean; cat. no. 074K1366, Sigma). Store as 10-mL aliquots at –20°C.
   b. 200X Pen/strep (Invitrogen) and stored as 1-mL aliquots at –20°C.
   c. G418 (Invitrogen or any other manufacturer). Prepare 40 mg/mL in PBS and filter sterilize. Store at –20°C. Final concentration used is 350–400 μg/mL.
   d. Puromycin dihydrochloride (Calbiochem). Reconstituted in water to get a stock concentration of 3 mg/mL. Final concentration used is 3 μg/mL.
   e. Hygromycin B solution (Calbiochem). The final concentration used is 150 μg/mL.
   f. Primocin (cat. no. Amaxa Biosystems) and used against mycoplasms as recommended by supplier. When using primocin, Pen/Strep is not additionally required in the medium.
   g. 6-TG is prepared as a 1000X stock (i.e., 2 m*M*). Dissolve 50 mg in small volume of 1 *N* NaOH. Add water up to 25 mL, dissolve completely, and filter sterilize. Store as 5-mL aliquots at –20°C.
   h. LIF is prepared according to protocol.
   i. 0.3 *M* Mannitol (cat. no. M4125, SIgma) is dissolved in ultrapure water with 0.3% BSA added (cat. no. A4378, Sigma); filter through 0.22-μm Millipore filter and store in aliquots at –20°C.

2.2. Websites, Tables, Vendors, and Vectors

2.2.1. Online Databases

1. http://www.mshri.on.ca/nagy/Cre-pub.html.
2. http://jaxmice.jax.org/models/cre_intro.html.

3. http://genetics.med.harvard.edu/~dymecki/.
4. http://www.ensembl.org/Mus_musculus/.

2.2.2. Tables

Conditional alleles in mice: tissue-specific knockouts, kkwan@mail.mdanderson.org.

2.2.3. Vendors

1. http://www.invitrogen.com/.
2. http://www.sigmaaldrich.com.
3. http://www.rndsystems.com/.
4. www.calbiochem.com.

2.2.4. Vectors

1. http://www.aldevron.com/vectors.php?vid=5.
2. http://www.stratagene.com/lit/vector.aspx.
3. http://www.genome.gov/10001852.
4. Mutant Lox P vectors, hiroshi@genetics.hpi-uni-hamburg.de.

3. Methods

3.1. Routine Culture, Generation, and Characterization of Genetically Altered Mouse ES Cells

Here, we describe culture of HM-1 cells, which are derived from the mouse strain 129/Ola and are HPRT-deficient. They are grown feeder-free and in gelatinized dishes. Other feeder-independent ES cell lines are grown essentially in the same way. For routine culture, cells should be passaged twice a week, splitting them between 1:6 and 1:12, depending on conditions. They should not be plated too thin because thin plating encourages excessive differentiation. Overconfluent cultures should be avoided for the same reason. Medium change should take place every other day as long as the culture is thin; later, medium should be changed every day. Any drastic change in cell doubling times or cell morphology is indicative of unwanted damage to the cells, e.g., chromosomal aberrations (*see* **Note 2**). Typical doubling time is approx 20–22 h (*see* **Fig. 5**).

3.2. Thawing and Plating Cells

1. Gelatinize flask by using enough gelatin to completely cover the flask or dish.
2. Leave at room temperature (RT) for at least 10 min. Aspirate off before plating cells.
3. If feeders are used, plate appropriate amount of cells together with ES cells or plate them 3 to 4 h earlier. Feeder-coated dishes can be used up to 7–10 d. Check for integrity of monolayer before using them to plate ES cells.
4. Thaw cells quickly in 37°C bath and transfer into 10-mL prewarmed medium. Spin for 5 min at 250*g*.

5. Remove medium, replace with 5 mL of fresh medium containing Pen/Strep, if necessary, and replate all cells on 25-cm^2 flask.
6. The next day, replace with fresh medium without antibiotics. Routine culture without antibiotics is strongly recommended.
7. To subculture, aspirate the medium and wash cells with 1 mL of prewarmed trypsin per 25-cm^2 flask (2 mL/75 cm^2).
8. Remove and replace with fresh trypsin and leave in 37°C incubator until cells are dislodged by gentle agitation. Dislodging happens almost instantly. Do not overtrypsinize or damage to the cells may result.
9. Add 1 mL of soybean trypsin inhibitor and 4 mL of medium per 1 mL of trypsin and pipet up and down against flask wall approx 10–20 times (you may vary the distance between pipet tip and flask wall between 0.5 and 2.5 cm). It is not essential to obtain a single cell suspension for routine culture, but large cell clumps must be avoided because they lead to differentiation.
10. Check suspension under microscope before transferring cells into a centrifuge tube. There should be a uniform size distribution of cells and cell aggregates.

In the meantime, gelatinize new flask(s). Spin cells for 5 min at 250*g* (at RT), aspirate off supernatant, and resuspend cell pellet gently in a few milliliters of medium. Replate at desired density and add 6–8 mL of medium to small flasks and 25–30 mL to large flasks. A 25-cm^2 flask yields approx 7–10 million cells, and a 75-cm^2 flask yields approx 25–30 million cells.

3.3. Preparation of Feeder Cells From STO or MEF Cells

Here, we describe routine culture and inactivation of feeder cells. When ES cells are grown in the presence of feeders, the culture medium containing serum should be used (refer to freezing of ES cells in **Subheading 2.** for medium components). There are two ways to inactivate feeder cells, mitomycin C treatment and γ-irradiation.

1. Mitomycin C treatment:
   a. Add 10 μg/mL mitomycin C to a confluent culture of feeder cells and place in incubator for 2 to 3 h.
   b. Wash the dishes five times with PBS and collect cells by trypsinization (1 to 2 min). Spin cells 3 min at 250*g* (at RT), discard supernatant, and then resuspend 1 times in medium and respin.
   c. Resuspend cells in ES cell medium and count them. Adjust at appropriate density, e.g., 10^6 cells/mL and use them. Mitomycin C-treated feeder cells can be frozen, following the same regimen as described for ES cell freezing.
2. γ-Irradiation:
   a. Grow cells to near confluence. Trypsinize cells from 10 to 15 75-cm^2 flasks, spin down 3 min at 250*g*. Discard supernatant. Resuspend cells in 10–12 mL of culture medium in a 50-mL plastic tube.

b. Irradiate cell suspension at 15 Gy for 7 to 8 min.
c. Replate irradiated cells on 10–15 75-cm^2 gelatinized flasks overnight. Trypsinize and freeze in aliquots or use in experiment.

For use in ES cell growth, plate feeders on gelatin-coated dishes 1 to 2 h before plating ES cells. A density of 50,000 cells/cm^2 is required. Feeder layers last approx 1 wk after treatment.

3.4. Electroporation of ES Cells (First Step of Gene Targeting)

The conditions recommended will work with a Bio-Rad gene pulser and cuvettes of 0.4-cm electrode gap. Setting: 800 V, 3 microfarads (standard). The conditions will work with any ES line and are relatively insensitive to changes in cell number and amount of DNA. They do not yield the highest possible number of transformed cells, but they are gentle to cells. Other settings that yield higher efficiencies are 250 V, 500 microfarads (optional). Under these conditions, use 25 µg of DNA and 10^6 cells per electroporation. Changes in DNA amounts and cell number alter efficiency and cell viability. Optimal results are currently achieved if the DNA is purified with Nucleobond or QIAGEN columns. EndoFree purified DNA is not necessary, but it may improve transfection efficiency.

1. For a standard targeting experiment, trypsinize approx 20–30 million cells.
2. Determine cell number in a counting chamber.
3. Resuspend ES cells in 0.8 mL of 1X HBS buffer.
4. Add cells to 200 µg of linearized DNA in 100 µL of sterile TE in Eppendorf tube (phenol/chloroform-extracted, ethanol-precipitated, remove 70% ethanol in clean bench).
5. Mix by pipeting up and down using a 1-mL cell culture pipet.
6. Transfer into electroporation cuvet.
7. Apply one pulse (standard or optional) and allow to stand for 10 min at RT.
8. Add cells to appropriate amount of medium and plate approx 1 to 2×10^6 cells/10-cm dish in 10 mL of medium.
9. Start positive selection the next day (refer to additional reagents in **Subheading 2.**). Leave one plate without selection to check cell recovery.
10. After 3–5 d, change medium again (electroporation is day 1). Colonies will grow after 8–10 d.
11. Isolate colonies as described in ES cell colony isolation protocol (**Subheading 3.6.**), use half for PCR and leave the remaining cells in 24-well plate with approx 1 mL of medium.
12. After a few days, multiple colonies should have grown.
13. Aspirate off medium, wash with 0.2 mL of trypsin, and aspirate off again. Add 2 to 3 drops of trypsin and incubate for 3–5 min at 37°C.
14. Add 0.5 mL of medium and break up cell clumps by using a blue-tip Gilson pipet.
15. Transfer cells to a six-well plate and add 5 mL of medium.

16. Grow to near confluence.
17. Trypsinize and plate onto two 25-cm^2 flasks adding two-thirds on one flask and one-third to the other flask. When confluent, use the two-thirds flask for freezing cells (two to three vials) and the other flask for preparation of genomic DNA.
18. Maintain selection until targeted ES cells are expanded and characterized by Southern blotting.
19. To remove cells from HAT selection, replace HAT with HT for at least 2 d before growing cells in normal medium. Cells grow slightly faster in the absence of selection. For HPRT negative selection, the final concentration of 6-TG is 5 µg/mL.
20. For correctly targeted ES cells, thaw an aliquot on 25-cm^2 flask; grow until confluent, and passage on 1X 75-cm^2 flask. Prepare six to eight freezing aliquots for blastocyst injections.

3.5. Double Replacement by Using HPRT-Minigenes (Second Step of Gene Targeting)

1. Keep passage number between first and second gene-targeting step as low as possible to minimize the accumulation of 6-TG–sensitive cells before electroporation.
2. Electroporate approx 20 million cells as in **Subheading 3.4.** and plate onto six 75-cm^2 flasks.
3. Apply 6-TG selection at day 6. Most likely, it will be necessary to split cells before selection. Split in such a way that you end up with six near confluent flasks at day 6. Then, trypsinize cells.
4. From each flask, plate 6 × 1–1.5 million cells onto 10-cm dishes in medium containing 6-TG.
5. Discard remaining cells.
6. Change medium after 2 and 4 d, respectively.
7. Colonies grow approx on day 10 (there will only be one to five colonies per dish). Process resistant colonies as described in **Subheading 3.4**.

3.6. Isolation of ES Cell DNA for PCR-Based Genotyping

With good vector design and the use of isogenic DNA for vector construction, high gene-targeting frequencies can be expected. For a range of different genes, we noted frequencies of 1:5 to1:35 PCR-positive colonies. Therefore, it should be sufficient to analyze between 100 and 200 colonies per experiment. The procedure detailed here is based on using approximately half of a colony for DNA isolation and the other half for further growth. In this way, positive colonies can be scored the next day. For analysis, small, slower growing ES cell colonies with a good morphology should be used.

1. Label colony position with a marker pen and isolate them with a yellow Gilson tip under the clean bench.
2. Coat 24-well plate with gelatin and remove by aspiration after 10–20 min.
3. Pick colonies along with approx 150 µL of medium by using a P 200 Gilson pipet.

4. Transfer colony to microtiter plate, disperse by pipetting up and down three to four times.
5. Leave half of cells in well and transfer other half into an Eppendorf tube.
6. Add 1 mL of selective medium and grow cells until 80% confluent.
7. Spin down cells in Eppendorf tube for 2 min at full speed.
8. Aspirate off supernatant and add 50 µL of 1X PCR buffer. Choose the buffer corresponding to Taq polymerase you are going to use. It must contain approx 0.5% of Nonidet P-40 or Triton X-100 for cell lysis, including 200 µg/mL proteinase K.
9. Vortex briefly and incubate 1 h at 65°C, preferably at slow motion (rotating wheel).
10. Heat 10 min at 95°C to inactivate proteinase K.
11. Spin briefly and use 2–5 µL of cell lysate for 25-µL PCR reaction. You must carry out PCR immediately after having the extract ready.

3.7. Freezing of ES Cells

Freeze cells from near confluent flasks. One aliquot should contain 2–4 millions cells. Cells (i.e., 2–3 million per small and 5–8 million from a large flask). It is beneficial to freeze cells as concentrated as possible, e.g., in 0.5–1 mL per vial. Freeze at approx 1°C/min, e.g., by using a Nalgene® freezing container.

1. To freeze, trypsinize cells as usual.
2. Resuspend cells gently in X mL of serum containing medium (*see* **Subheading 2.**).
3. Add X mL of 2X freezing medium (dropwise).
4. Mix gently and transfer instantly to Nalgene freezing container and place into –80°C overnight (or at least 4 h).
5. Thereafter, transfer cells to liquid nitrogen container.

3.8. Karyotyping of ES Cells

Before using a clone for a second electroporation or for blastocyst injections, the number of chromosomes should be determined by kryotype analysis. A high degree of euploidy (greater than 70%) is necessary but does not guarantee germline competence. Therefore, it is preferable to characterize a sufficient number of euploid clones to increase the probability of germline transmission.

1. Starting with confluent 25-cm^2 flasks, the cells are arrested in metaphase by incubating them for 1 h at 37°C, 5% CO_2 with 1 mL of fresh medium containing 10 µL of demecolcine solution (10 µg/mL in Hanks' balanced salt solution [cat no. D-1925, Sigma]; toxic-handle and dispose according to safety guidelines).
2. Remove medium and wash twice with PBS.
3. Add 1 mL of trypsin solution and incubate for 4 min at 37°C.
4. Stop trypsinization by adding 5 mL of medium, spin down cells.
5. Remove medium and add 1 mL of 0.56% KCl dropwise to the cell pellet.
6. Resuspend cells by flicking and then add another milliliter of 0.56% KCl and mix by flicking.
7. Incubate at RT for 8–10 min.

8. Spin down gently (90*g*; 5 min) and remove supernatant.
9. Add dropwise 2 to 3 mL of chilled fixation solution (3:1 methanol/acetic acid, freshly prepared) onto the pellet, resuspend by flicking, and incubate 5 min at RT.
10. Spin down gently for 5 min at 90*g*; and remove supernatant.
11. Repeat fixation procedure once or twice.
12. Finally, resuspend pellet in 0.5–1 mL of fixation solution, drop the solution onto clean slides from a distance of 15–30 cm, and air-dry slides.
13. Check under the microscope if the broken nuclei are sufficiently separated from each other to ensure that every chromosome can be assigned to its metaphase spread.
14. Incubate slides for 1 min in Giemsa staining solution (0.4% modified Giemsa stain, Sigma).
15. Wash twice with distilled water, air-dry, and mount the slides.
16. Count the number of chromosomes in metaphase spreads of the nuclei. Because direct counting under the microscope is tedious and error prone, it is recommended that pictures be taken on which the counted chromosomes can be marked.

3.9. Preparation of HM-1 Cells for Blastocyst Injections

A subconfluent 25-cm^2 flask of cells is sufficient. Ideally, thaw cells 4 d before injection. Approximately 24 h before microinjection, they should be near confluent.

1. Trypsinize and replate 70–100% of cells.
2. Change medium early in the morning of the injection. Start trypsinization only when everything at the microinjection facility is ready to go. Keep in mind that cells become fragile after 1.5–2 h.
3. Trypsinization: aspirate medium, add 2 mL of trypsin, mix gently, and remove. Add 3 mL of trypsin and incubate 5–10 min at 37°C to receive single-cell suspension (check under microscope and continue incubation if necessary).
4. Add 5 mL of medium and pipet up and down gently.
5. Spin 5 min at 250*g*, aspirate medium, and resuspend cells in 5 mL of injection medium (GMEM, 10% FCS, 20 m*M* HEPES, 1X Pen/Strep).
6. Before adding cells to microinjection droplet, add 1 µl of ultrapure DNase I (20–40 U/µL). Keep cells on ice during injection.

3.10. Production of Chimeric Mice

The setup for the injection of blastocysts or the production of aggregation chimeras has been extensively described previously ***(127)***.

3.10.1. Preparation of Hormones for Superovulating Female Mice

Prepare stock solution in sterile 0.9% NaCl. Keep solution frozen in convenient aliquots at –20°C until use (hormone stocks are stable in the freezer for more than 1 yr). Thaw before use and do not refreeze. If aliquots are not used in full, discard them after injection.

1. Gonadotropin: gonadotropin from pregnant mare's serum (PMS) (1000 or 5000 U/vial; cat. no. 367222, Calbiochem). Store powder at 4°C. For administration, the PMS is resuspended at 50 U/mL in sterile 0.9% NaCl and divided into aliquots.
2. Human chorionic gonadotropin (hCG): chorionic gonadotropin, human urine, standard grade (1 mg; cat. no. 230734, Calbiochem). hCG is resuspended at final concentration of 50 U/mL in sterile 0.9% NaCl and divided into aliquots.

3.10.2. Media

KSOM medium complemented with amino acids is recommended to overcome the inability of two-cell embryos to divide further in culture (two-cell block).

3.11. Superovulating of Females for Preparation of Donor Blastocysts

The consecutive action of two hormones is used to superovulate females. Gonadotropin from PMS (used to mimic follicle-stimulating hormone) is the first hormone; hCG (used to mimic luteinizing hormone) is the second hormone. Strain response to such induction is widely variable. The doses used depend upon the particular strain and age (or weight) of female and are worked out empirically. For isolation of blastocysts, we use 4- to 5-wk-old C57BL/6J females; for collecting oocytes, we use 8- to 10-wk-old FVB females. The times that the PMS and the hCG are administered are relative to each other and to the light/dark cycle of the mouse room. For a light/dark cycle of 12 h (6 AM–6 PM), the optimal isolation protocol is as follows:

1. Injection of the first hormone, PMS, at 12 PM, on day 1:5 U for 4- to 5-wk old C57BL/6J and 7.5 U for 8- to 10-wk-old FVB.
2. Injection of the second hormone, hCG, 48 h later at 12 PM, on day 3:5 U for 4- to 5-wk-old C57BL/6J and 7.5 U for 8- to 10-wk-old FVB strain mice.
3. Females are mated with normal males after the second injection, one female per male. A period of 1 or 2 h is normally given for the females to recover from the stress of injection before mating. Plugs are checked the next morning.

3.11.1. Production of Chimeras by Aggregation of Diploid With Tetraploid Embryos

The generation of aggregation chimeras has been proven as a powerful tool for the analysis of genes essential for embryonic development. This technique can be used to produce a mosaic of two different mice or to restrict the contribution of one mouse to the extraembryonic or embryonic compartments ***(132)***. Because of the limited developmental potential of tetraploid embryos, their cells contribute preferentially to the extraembryonic tissues of the embryo, whereas ES cells are unable to form those compartments. By aggregation of ES cells and tetraploid embryos, it is possible to generate embryos that consist of a diploid embryo and a tetraploid extraembryonic tissue. This system has the

power to overcome an embryonic lethal phenotype that is caused by an extraembryonic defect, if tetraploid wild-type embryos are aggregated with diploid knockout ES cells ***(55)***. The principle of this experiment is outlined in Hesse et al., 2005 ***(32)***. If a homozygous ES cell line of the knockout of interest is not available, knockout embryos of the eight-cell stage can be used for aggregation. The efficiency is significantly lower, compared with aggregation with ES cells, because the embryos can still contribute to the extra embryonic compartments.

3.11.2. Recovery of Two- and Eight-Cell Embryos

Two-cell embryos are flushed out on day 1.5 p.c., whereas eight-cell embryos are flushed out 1 d later at day 2.5 p.c. To perform these experiments, the following additional materials are needed: dissecting microscope, flushing needle (the sharp tip of no. 30-guage 1/2 needle is cut off and then rounded using sharpening stone), 1-mL syringe, dissecting instruments (fine-pointed scissors, fine forceps); 5 forceps (Dumont), mouth pipet (aspirator mouth piece, latex tubing, blue tip, micropipette), sterile tissue culture dishes (35 × 10 mm), and M2 and KSOM-AA media.

1. Superovulate an appropriate number of CD-1 females (*see* **Subheading 3.10.1.**), mate with CD-1 males, and check for copulation plugs.
2. On day 1.5 or 2.5 post coitus, sacrifice the plugged females by cervical dislocation to obtain two- or 8-cell embryos, respectively.
3. Dissect the oviducts by cutting the upper part of the uterus and underneath the ovary.
4. Place oviducts in a 35-mm tissue culture dish filled with M2 medium at RT.
5. Transfer an oviduct into a tissue culture dish with M2 medium under a dissecting microscope.
6. Connect the flushing needle to a 1-mL syringe and fill it with M2 medium.
7. Grab the infundibulum with a fine forceps and insert flushing needle.
8. After successful insertion, hold the needle with the forceps.
9. Flush out oviduct with 100–150 µL of M2 medium by applying constant pressure.
10. Repeat all steps for each oviduct dissected.
11. Spin the tissue culture dish containing the embryos manually with fast circular movements. This motion brings the embryos to the middle of the dish.
12. Collect all embryos with the mouth pipet and rinse them trough 3 drops of M2 medium and 3 drops of equilibrated KSOM-AA medium.
13. Store the embryos in KSOM-AA drop under mineral oil in an incubator (37°C; 5% CO_2).

3.11.3. Generation of Tetraploid Embryos

Fusion of the blastomeres of two-cell embryos occurs when a square pulse is applied perpendicular to the plane of contact of the two cells. Adjustable

alternating current (AC) field is applied to allow the correct orientation of embryos (enable, 1 or 2 V on the display). AC fields that are too high can cause lysis. The following supplemental materials are needed for these experiments: cell fusion instrument, two dissecting microscopes, M2 and KSOM-AA media, 0.3 *M* mannitol and tissue culture dish with KSOM-AA microdrops covered with oil (cat. no. M8410, Sigma), and mouth- or finger-controlled pipet.

1. Place slide with electrodes onto a 100-mm Petri dish by "fixing" it with two drops of water.
2. Disperse a drop of M2 medium and mannitol over the slide with the electrodes and 1 drop of M2 medium under it.
3. Place a large drop of mannitol between the electrodes.
4. Connect the electrodes with the pulse generator and make sure that all settings are correct according to the manufacturer.
5. Place all recovered two-cell embryos into the M2 microdrop.
6. Wash a group of approx 30 embryos in the mannitol drop and transfer them between the electrodes into the mannitol.
7. Switch on the ac current and watch the embryos orientating through a microscope, until the contact area of the two blastomeres is parallel to the electrodes. Wait until all embryos are in the proper orientation and use a mouth pipet to abut nonorientated embryos.
8. Apply the pulse to the embryos and immediately switch of the AC field.
9. Wash the embryos twice in KMSO media and transfer them into a KMSO drop overlaid by light mineral oil and put them into the incubator at 37°C, 5% CO_2.
10. Exchange mannitol solution to prevent water evaporation and mannitol crystallization 45 min after applying the pulse.
11. Check for perfectly fused embryos and put them into a drop of KSOM-AA with oil overlay.
12. Repeat the pulse with unfused two-cell embryos. Culture tetraploid embryos overnight at 37°C, 5% CO_2.
13. After culturing for 24 h, use four-cell tetraploid embryos for the aggregation experiment, and discard all other embryos.

3.11.4. Preparation of Aggregation Plate

The following supplemental materials are needed for these experiments: dissecting microscope, sterile tissue culture dishes (35 × 10-mm Easy Grip; cat. no. 3001-3, Falcon); 1-mL syringe with 26-gage 1/2 needle of KSOM-AA medium, light mineral oil (embryo tested), and aggregation (darning) needle (DN-09).

1. Dispense small drops of KSOM-AA medium (~15 µL each) on a 100 × 20-mm tissue culture dish. Flood dish with light mineral oil, until all drops are covered completely.

2. Wash aggregation needle briefly in ethanol.
3. Make about six aggregation wells into each of the drops by pressing and turning the needle into the ground of the tissue culture dish.
4. Place aggregation plate in an incubator at 37°C, 5%CO_2 for at least 2 h.

3.11.5. Removal of Zona Pellucida With Tyrode's Solution

The following supplemental materials are needed for these experiments: dissecting microscope, acid Tyrode's solution (cat. no. T1788, Sigma), tissue culture dishes (100 × 15 mm), M2 and KSOM-AA medium, and mouth pipet.

1. Add embryos to the first M2 drop. Rinse 20 embryos in the first drop of Tyrode's solution and transfer them into the second drop.
2. Observe dissolving of the zona pellucida with the microscope and instantly transfer the embryos to the M2 drop containing 0.3% BSA to prevent embryos from sticking to the plate or to each other.
3. Wash embryos by transferring them through 3 drops of M2 and 3 drops of KSOM-AA.
4. Transfer embryos into a KSOM-AA drop without depressions on the aggregation plate.
5. Trypsinize ES cells for aggregation if ES cell/tetraploid embryo aggregation is planned.

3.11.6. ES Cells or Diploid Embryo/Tetraploid Embryo "Sandwich" Aggregation (**Table 3**)

The following supplemental materials are needed for these experiments: dissecting microscope, prepared aggregation plate with depressions and embryos with removed zona, trypsinized ES cells or eight-cell embryos, and mouth-controlled pipet.

1. Transfer tetraploid embryos in 1 microdrop of the aggregation plate without depressions.
2. Also transfer diploid eight-cell knockout embryos or trypsinized ES cells into another microdrop.
3. Place a tetraploid embryo in each of the aggregation wells.
4. Place an eight-cell embryo or a clump of approx 10 ES cells on the tetraploid embryo in each depression.
5. Place a second tetraploid embryo on top of the eight-cell embryo or ES cell clumps, respectively, thus forming a sandwich.
6. Repeat steps 2 to 4 for all aggregations and culture the plate overnight in an incubator at 37°C and 5% CO_2.
7. On the afternoon of the next day check aggregates for blastocyst formation. Transfer blastocysts into the uteri of day 2.5 p.c. pseudopregnant foster females.

Table 3
Time Schedule For Aggregation of Tetraploid–Diploid Embryos

Day 1	Day 2	Day 3	Day 4	Day 5	Day 6	Day 7
Superovulation of CD-1 and knockout		Superovulation 5 U of hCG	Check for copulation plugs	Flushing of two-cell embryos	Flushing of eight-cell embryos	Transfer into uterus of day 2.5 pseudopregnant recipients
				Electrofusion	**Setting up aggregation**	
5 U of PMSG		Setting up matings of CD-1 and knockout 1:1	Mating of recipients	**Setting up aggregation plate**		
					Incubation o/n, 5% CO_2 at 37°C	
				Incubation o/n, 5% CO_2 at 37°C		

4. Notes

1. It is important to keep ES cell passage number as low as possible to maximize germline competence and to minimize genetic and epigenetic alterations. It is recommended to prepare a sufficiently large stock of frozen vials and to expand a vial every time for a new experiment. In general, ES cells should be split at rates between 1:6 and 1:12. High- and low-density plating carries the risk of differentiating the cells. Their growth rate (typical doubling time is 20–22 h) should be carefully monitored; significant changes in growth rate and morphological appearance signal chromosomal alterations and preclude further use of such cells for gene-targeting experiments.
2. To retain their excellent germline transmission potential, it is essential to keep passage number of HM-1 cells as low as possible. Good germline transmission rates depend on (1) media, plasticware, and so on; (2) limiting passage number; and (3) experimental skills. We use cell culture plastic from BD Biosciences, cell culture grade water, and reagents.
3. More details about human and mouse keratin clusters can be found in **ref.** *3*.

Acknowledgments

We thank Ursula Reuter, Silke Loch, and Claudia Wohlenberg for excellent technical assistance. Work by the authors was supported by the Deutsche Forschungsgemeinschaft, the Bundesministerium für Bildung und Forschung, and the Thyssen foundation. We apologize to those colleagues whose work has not been cited because of space restrictions.

References

1. Omary, M. B., Coulombe, P. A., and McLean, W. H. (2004) Intermediate filament proteins and their associated diseases. *N Engl J Med* **351,** 2087–2100.
2. Hesse, M., Magin, T. M., and Weber, K. (2001) Genes for intermediate filament proteins and the draft sequence of the human genome: novel keratin genes and a surprisingly high number of pseudogenes related to keratin genes 8 and 18. *J. Cell Sci.* **114,** 2569–2575.
3. Hesse, M., Zimek, A., Weber, K., and Magin, T. M. (2004) Comprehensive analysis of keratin gene clusters in humans and rodents. *Eur. J. Cell. Biol.* **83,** 19–26.
4. Magin, T. M., Reichelt, J., and J. C. (2005) The role of keratins in epithelial homeostasis, Elias review, submitted.
5. Rogers, M. A., Langbein, L., Winter, H., et al. (2001) Characterization of a cluster of human high/ultrahigh sulfur keratin-associated protein genes embedded in the type I keratin gene domain on chromosome 17q12-21. *J. Biol. Chem.* **276,** 19,440–19,451.
6. Bawden, C. S., McLaughlan, C., Nesci, A., and Rogers, G. (2001) A unique type I keratin intermediate filament gene family is abundantly expressed in the inner root sheaths of sheep and human hair follicles. *J. Invest. Dermatol.* **116,** 157–166.

7. Zimek, A. and Weber, K. (2005) Terrestrial vertebrates have two keratin gene clusters; striking differences in teleost fish. *Eur. J. Cell Biol.* **84,** 623–635.
8. Herrmann, H., Hesse, M., Reichenzeller, M., Aebi, U., and Magin, T. M. (2003) Functional complexity of intermediate filament cytoskeletons: from structure to assembly to gene ablation. *Int. Rev. Cytol.* **223,** 83–175.
9. Herrmann, H., Strelkov, S. V., Feja, B., et al. (2000) The intermediate filament protein consensus motif of helix 2B: its atomic structure and contribution to assembly. *J. Mol. Biol.* **298,** 817–832.
10. Herrmann, H., Wedig, T., Porter, R. M., Lane, E. B., and Aebi, U. (2002) Characterization of early assembly intermediates of recombinant human keratins. *J. Struct. Biol.* **137,** 82–96.
11. Strelkov, S. V., Herrmann, H., Geisler, N., et al. (2002) Conserved segments 1A and 2B of the intermediate filament dimer: their atomic structures and role in filament assembly. *EMBO J.* **21,** 1255–1266.
12. Herrmann, H. and Aebi, U. (2004) Intermediate filaments: molecular structure, assembly mechanism, and integration into functionally distinct intracellular Scaffolds. *Annu. Rev. Biochem.* **73,** 749–789.
13. Irvine, A. D. and McLean, W. H. (1999) Human keratin diseases: the increasing spectrum of disease and subtlety of the phenotype-genotype correlation. *Br. J. Dermatol.* **140,** 815–828.
14. Whittock, N. V., Ashton, G. H., Griffiths, W. A., Eady, R. A., and McGrath, J. A. (2001) New mutations in keratin 1 that cause bullous congenital ichthyosiform erythroderma and keratin 2e that cause ichthyosis bullosa of Siemens. *Br. J. Dermatol.* **145,** 330–335.
15. Whittock, N. V., Wan, H., Morley, S. M., et al. (2002) Compound heterozygosity for non-sense and mis-sense mutations in desmoplakin underlies skin fragility/woolly hair syndrome. *J. Invest. Dermatol.* **118,** 232–238.
16. Coulombe, P. A., Hutton, M. E., Letai, A., Hebert, A., Paller, A. S., and Fuchs, E. (1991) Point mutations in human keratin 14 genes of epidermolysis bullosa simplex patients: genetic and functional analyses. *Cell* **66,** 1301–1311.
17. Lane, E. B., Rugg, E. L., Navsaria, H., et al. (1992) A mutation in the conserved helix termination peptide of keratin 5 in hereditary skin blistering. *Nature* **356,** 244–246.
18. Ma, L., Yamada, S., Wirtz, D., and Coulombe, P. A. (2001) A 'hot-spot' mutation alters the mechanical properties of keratin filament networks. *Nat. Cell Biol.* **3,** 503–506.
19. Werner, N. S., Windoffer, R., Strnad, P., Grund, C., Leube, R. E., and Magin, T. M. (2004) Epidermolysis bullosa simplex-type mutations alter the dynamics of the keratin cytoskeleton and reveal a contribution of actin to the transport of keratin subunits. *Mol. Biol. Cell* **15,** 990–1002.
20. Windoffer, R., Woll, S., Strnad, P., and Leube, R. E. (2004) Identification of novel principles of keratin filament network turnover in living cells. *Mol. Biol. Cell* **15,** 2436–2448.
21. Peters, B., Kirfel, J., Bussow, H., Vidal, M., and Magin, T. M. (2001) Complete cytolysis and neonatal lethality in keratin 5 knockout mice reveal its fundamental

role in skin integrity and in epidermolysis bullosa simplex. *Mol. Biol. Cell* **12,** 1775–1789.
22. Coulombe, P. A. and Omary, M. B. (2002) 'Hard' and 'soft' principles defining the structure, function and regulation of keratin intermediate filaments. *Curr. Opin. Cell Biol.* **14,** 110–122.
23. Ku, N. O. and Omary, M. B. (2000) Keratins turn over by ubiquitination in a phosphorylation-modulated fashion. *J. Cell Biol.* **149,** 547–552.
24. Steinert, P. M. and Marekov, L. N. (1995) The proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and small proline-rich proteins 1 and 2 are isodipeptide cross-linked components of the human epidermal cornified cell envelope. *J. Biol. Chem.* **270,** 17,702–17,711.
25. Abe, M. and Oshima, R. G. (1990) A single human keratin 18 gene is expressed in diverse epithelial cells of transgenic mice. *J. Cell Biol.* **111,** 1197–1206.
26. Neznanov, N., Umezawa, A., and Oshima, R. G. (1997) A regulatory element within a coding exon modulates keratin 18 gene expression in transgenic mice. *J. Biol. Chem*. **272,** 27,549–27,557.
27. Rhodes, K. and Oshima, R. G. (1998) A regulatory element of the human keratin 18 gene with AP-1-dependent promoter activity. *J. Biol. Chem*. **273,** 26,534–26,542.
28. Willoughby, D. A., Vilalta, A., and Oshima, R. G. (2000) An Alu element from the K18 gene confers position-independent expression in transgenic mice. *J. Biol. Chem.* **275,** 759–768.
29. Wen, F., Cecena, G., Munoz-Ritchie, V., Fuchs, E., Chambon, P., and Oshima, R. G. (2003) Expression of conditional cre recombinase in epithelial tissues of transgenic mice. *Genesis* **35,** 100–106.
30. Sinha, S., Degenstein, L., Copenhaver, C., and Fuchs, E. (2000) Defining the regulatory factors required for epidermal gene expression. *Mol. Cell Biol.* **20,** 2543–2555.
31. Sugihara, T. M., Kudryavtseva, E. I., Kumar, V., Horridge, J. J., and Andersen, B. (2001) The POU domain factor Skin-1a represses the keratin 14 promoter independent of DNA binding. A possible role for interactions between Skn-1a and CREB-binding protein/p300. *J. Biol. Chem.* **276,** 33,036–33,044.
32. Lu, H., Hesse, M., Peters, B., and Magin, T. M. (2005) Type II keratins precede type I keratins during early embryonic development. *Eur. J. Cell Biol*. **84,** 709–718.
33. Ameen, N. A., Figueroa, Y., and Salas, P. J. (2001) Anomalous apical plasma membrane phenotype in CK8-deficient mice indicates a novel role for intermediate filaments in the polarization of simple epithelia. *J. Cell Sci.* **114,** 563–575.
34. Magin, T. M. (1998) Lessons from keratin transgenic and knockout mice. *Subcell. Biochem*. **31,** 141–172.
35. Lloyd, C., Yu, Q. C., Cheng, J., et al. (1995) The basal keratin network of stratified squamous epithelia: defining K15 function in the absence of K14. *J. Cell Biol.* **129,** 1329–1344.
36. Tong, X. and Coulombe, P. A. (2004) A novel mouse type I intermediate filament gene, keratin 17n (K17n), exhibits preferred expression in nail tissue. *J. Invest. Dermatol.* **122,** 965–970.

37. Herzog, F., Winter, H., and Schweizer, J. (1994) The large type II 70-kDa keratin of mouse epidermis is the ortholog of human keratin K2e. *J. Invest. Dermatol.* **102,** 165–170.
38. Swensson, O., Langbein, L., McMillan, J. R., et al. (1998) Specialized keratin expression pattern in human ridged skin as an adaptation to high physical stress. *Br. J. Dermatol.* **139,** 767–775.
39. Freedberg, I. M., Tomic-Canic, M., Komine, M., and Blumenberg, M. (2001) Keratins and the keratinocyte activation cycle. *J. Invest. Dermatol.* **116,** 633–640.
40. McGowan, K. and Coulombe, P. A. (1998a) The wound repair-associated keratins 6, 16, and 17. Insights into the role of intermediate filaments in specifying keratinocyte cytoarchitecture. *Subcell. Biochem.* **31,** 173–204.
41. McGowan, K. M. and Coulombe, P. A. (1998b) Onset of keratin 17 expression coincides with the definition of major epithelial lineages during skin development. *J. Cell Biol.* **143,** 469–486.
42. Kasper, M. (1992) Patterns of cytokeratins and vimentin in guinea pig and mouse eye tissue: evidence for regional variations in intermediate filament expression in limbal epithelium. *Acta Histochem.* **93,** 319–332.
43. Brock, J., McCluskey, J., Baribault, H., and Martin, P. (1996) Perfect wound healing in the keratin 8 deficient mouse embryo. *Cell Motil. Cytoskeleton* **35,** 358–366.
44. Jackson, B. W., Grund, C., Schmid, E., Burki, K., Franke, W. W., and Illmensee, K. (1980) Formation of cytoskeletal elements during mouse embryogenesis. Intermediate filaments of the cytokeratin type and desmosomes in preimplantation embryos. *Differentiation* **17,** 161–179.
45. Mazzalupo, S., Wong, P., Martin, P., and Coulombe, P. A. (2003) Role for keratins 6 and 17 during wound closure in embryonic mouse skin. *Dev. Dyn.* **226,** 356–365.
46. Mills, A. A., Zheng, B., Wang, X. J., Vogel, H., Roop, D. R., and Bradley, A. (1999) p63 is a p53 homologue required for limb and epidermal morphogenesis. *Nature* **398,** 708–713.
47. Lane, E. B. and McLean, W. H. (2004) Keratins and skin disorders. *J. Pathol.* **204,** 355–366.
48. Porter, R. M. and Lane, E. B. (2003) Phenotypes, genotypes and their contribution to understanding keratin function. *Trends Genet.* **19,** 278–285.
49. Yoneda, K., Furukawa, T., Zheng, Y. J., et al. (2004) An autocrine/paracrine loop linking keratin 14 aggregates to tumor necrosis factor alpha-mediated cytotoxicity in a keratinocyte model of epidermolysis bullosa simplex. *J. Biol. Chem.* **279,** 7296–7303.
50. D'Alessandro, M., Russell, D., Morley, S. M., Davies, A. M., and Lane, E. B. (2002) Keratin mutations of epidermolysis bullosa simplex alter the kinetics of stress response to osmotic shock. *J. Cell Sci.* **115,** 4341–4351.
51. Owens, D. W. and Lane, E. B. (2004) Keratin mutations and intestinal pathology. *J. Pathol.* **204,** 377–385.

52. Zatloukal, K., Stumptner, C., Fuchsbichler, A., et al. (2004) The keratin cytoskeleton in liver diseases. *J. Pathol.* **204,** 367–376.
53. Baribault, H., Penner, J., Iozzo, R. V., and Wilson-Heiner, M. (1994) Colorectal hyperplasia and inflammation in keratin 8-deficient FVB/N mice. *Genes Dev.* **8,** 2964–2973.
54. Baribault, H., Price, J., Miyai, K., and Oshima, R. G. (1993) Mid-gestational lethality in mice lacking keratin 8. *Genes Dev.* **7,** 1191–1202.
55. Jaquemar, D., Kupriyanov, S., Wankell, M., et al. (2003) Keratin 8 protection of placental barrier function. *J. Cell Biol.* **161,** 749–756.
56. Toivola, D. M., Krishnan, S., Binder, H. J., Singh, S. K., and Omary, M. B. (2004) Keratins modulate colonocyte electrolyte transport via protein mistargeting. *J. Cell Biol.* **164,** 911–921.
57. Habtezion, A., Toivola, D. M., Butcher, E. C., and Omary, M. B. (2005) Keratin-8-deficient mice develop chronic spontaneous Th2 colitis amenable to antibiotic treatment. *J. Cell Sci.* **118,** 1971–1980.
58. Satoh, M. I., Hovington, H., and Cadrin, M. (1999) Reduction of cytochemical ecto-ATPase activities in keratin 8-deficient FVB/N mouse livers. *Med. Electron Microsc.* **32,** 209–212.
59. Caulin, C., Ware, C. F., Magin, T. M., and Oshima, R. G. (2000) Keratin-dependent, epithelial resistance to tumor necrosis factor-induced apoptosis. *J. Cell Biol.* **149,** 17–22.
60. Gilbert, S., Loranger, A., Daigle, N., and Marceau, N. (2001) Simple epithelium keratins 8 and 18 provide resistance to Fas-mediated apoptosis. The protection occurs through a receptor-targeting modulation. *J. Cell Biol.* **154,** 763–773.
61. Inada, H., Izawa, I., Nishizawa, M., et al. (2001) Keratin attenuates tumor necrosis factor-induced cytotoxicity through association with TRADD. *J. Cell Biol.* **155,** 415–426.
62. Toivola, D. M., Nieminen, M. I., Hesse, M., et al. (2001) Disturbances in hepatic cell-cycle regulation in mice with assembly-deficient keratins 8/18. *Hepatology* **34,** 1174–1183.
63. Tamai, Y., Ishikawa, T., Bosl, M. R., et al. (2000) Cytokeratins 8 and 19 in the mouse placental development. *J. Cell Biol.* **151,** 563–572.
64. Hesse, M., Franz, T., Tamai, Y., Taketo, M. M., and Magin, T. M. (2000) Targeted deletion of keratins 18 and 19 leads to trophoblast fragility and early embryonic lethality. *EMBO J.* **19**, 5060–5070.
65. Hesse, M., Watson, E. D., Schwaluk, T., and Magin, T. M. (2005) Rescue of keratin 18/19 doubly deficient mice using aggregation with tetraploid embryos. *Eur. J. Cell Biol.* **84,** 355–361.
66. Andra, K., Lassmann, H., Bittner, R., et al. (1997) Targeted inactivation of plectin reveals essential function in maintaining the integrity of skin, muscle, and heart cytoarchitecture. *Genes Dev.* **11,** 3143–3156.
67. Guo, L., Degenstein, L., Dowling, J., et al. (1995) Gene targeting of BPAG1: abnormalities in mechanical strength and cell migration in stratified epithelia and neurologic degeneration. *Cell* **81,** 233–243.

68. Kowalczyk, A. P., Bornslaeger, E. A., Norvell, S. M., Palka, H. L., and Green, K. J. (1999) Desmosomes: intercellular adhesive junctions specialized for attachment of intermediate filaments. *Int. Rev. Cytol.* **185,** 237–302.
69. Smith, E. A. and Fuchs, E. (1998) Defining the interactions between intermediate filaments and desmosomes. *J. Cell Biol.* **141,** 1229–1241.
70. Hutton, E., Paladini, R. D., Yu, Q. C., Yen, M., Coulombe, P. A., and Fuchs, E. (1998) Functional differences between keratins of stratified and simple epithelia. *J. Cell Biol.* **143,** 487–499.
71. Fuchs, E. (1995) Keratins and the skin. *Annu. Rev. Cell Dev. Biol.* **11,** 123–153.
72. Cao, T., Longley, M. A., Wang, X. J., and Roop, D. R. (2001) An inducible mouse model for epidermolysis bullosa simplex: implications for gene therapy. *J. Cell Biol.* **152,** 651–656.
73. Berton, T. R., Wang, X. J., Zhou, Z., et al. (2000) Characterization of an inducible, epidermal-specific knockout system: differential expression of lacZ in different Cre reporter mouse strains. *Genesis* **26,** 160–161.
74. Fuchs, E. and Green, H. (1980) Changes in keratin gene expression during terminal differentiation of the keratinocyte. *Cell* **19,** 1033–1042.
75. Reichelt, J., Bussow, H., Grund, C., and Magin, T. M. (2001) Formation of a normal epidermis supported by increased stability of keratins 5 and 14 in keratin 10 null mice. *Mol. Biol. Cell* **12,** 1557–1568.
76. van Hemert, M. J., Steensma, H. Y., and van Heusden, G. P. (2001) 14-3-3 proteins: key regulators of cell division, signalling and apoptosis. *Bioessays* **23,** 936–946.
77. Reichelt, J. and Magin, T. M. (2002) Hyperproliferation, induction of c-Myc and 14-3-3sigma, but no cell fragility in keratin-10-null mice. *J. Cell Sci.* **115,** 2639–2650.
78. Paramio, J. M. and Jorcano, J. L. (2002) Beyond structure: do intermediate filaments modulate cell signalling? *Bioessays* **24,** 836–844.
79. Reichelt, J., Breiden, B., Sandhoff, K., and Magin, T. M. (2004a) Loss of keratin 10 is accompanied by increased sebocyte proliferation and differentiation. *Eur. J. Cell Biol.* **83,** 747–759.
80. Reichelt, J., Furstenberger, G., and Magin, T. M. (2004b) Loss of keratin 10 leads to mitogen-activated protein kinase (MAPK) activation, increased keratinocyte turnover, and decreased tumor formation in mice. *J. Invest. Dermatol.* **123,** 973–981.
81. Porter, R. M., Leitgeb, S., Melton, D. W., Swensson, O., Eady, R. A., and Magin, T. M. (1996) Gene targeting at the mouse cytokeratin 10 locus: severe skin fragility and changes of cytokeratin expression in the epidermis. *J. Cell Biol.* **132,** 925–936.
82. Reichelt, J., Bauer, C., Porter, R., Lane, E., and Magin, V. (1997) Out of balance: consequences of a partial keratin 10 knockout. *J. Cell Sci.* **110 (18),** 2175–2186.
83. Elias, P., Man, M. Q., Williams, M. L., Feingold, K. R., and Magin, T. (2000) Barrier function in K-10 heterozygote knockout mice. *J. Invest. Dermatol.* **114,** 396, 397.
84. Reichelt, J., Doering, T., Schnetz, E., Fartasch, M., Sandhoff, K., and Magin, A. M. (1999) Normal ultrastructure, but altered stratum corneum lipid and protein composition in a mouse model for epidermolytic hyperkeratosis. *J. Investig. Dermatol.* **113,** 329–334.

85. Arin, M. J., Longley, M. A., Wang, X. J., and Roop, D. R. (2001) Focal activation of a mutant allele defines the role of stem cells in mosaic skin disorders. *J. Cell Biol.* **152,** 645–649.
86. Steinert, P. M. and Marekov, L. N. (1999) Initiation of assembly of the cell envelope barrier structure of stratified squamous epithelia. *Mol. Biol. Cell* **10,** 4247–4261.
87. Akiyama, M., Takizawa, Y., Sawamura, D., Matsuo, I., and Shimizu, H. (2003) Disruption of the suprabasal keratin network by mutation M150T in the helix initiation motif of keratin 10 does not affect cornified cell envelope formation in human epidermis. *Exp. Dermatol.* **12,** 638–645.
88. Sprecher, E., Ishida-Yamamoto, A., Becker, O. M., et al. (2001) Evidence for novel functions of the keratin tail emerging from a mutation causing ichthyosis hystrix. *J. Invest. Dermatol.* **116,** 511–519.
89. Sprecher, E., Yosipovitch, G., Bergman, R., et al. (2003) Epidermolytic hyperkeratosis and epidermolysis bullosa simplex caused by frameshift mutations altering the v2 tail domains of keratin 1 and keratin 5. *J. Invest. Dermatol.* **120,** 623–626.
90. Kalinin, A., Marekov, L. N., and Steinert, P. M. (2001) Assembly of the epidermal cornified cell envelope. *J. Cell Sci.* **114,** 3069–3070.
91. McLean, W. H., Rugg, E. L., Lunny, D. P., et al. (1995) Keratin 16 and keratin 17 mutations cause pachyonychia congenita. *Nat. Genet.* **9,** 273–278.
92. Wojcik, S. M., Bundman, D. S., and Roop, D. R. (2000) Delayed wound healing in keratin 6a knockout mice. *Mol. Cell Biol.* **20,** 5248–5255.
93. McGowan, K. M., Tong, X., Colucci-Guyon, E., Langa, F., Babinet, C., and Coulombe, P. A. (2002) Keratin 17 null mice exhibit age- and strain-dependent alopecia. *Genes Dev.* **16,** 1412–1422.
94. Wong, P., Colucci-Guyon, E., Takahashi, K., Gu, C., Babinet, C., and Coulombe, P. A. (2000) Introducing a null mutation in the mouse K6alpha and K6beta genes reveals their essential structural role in the oral mucosa. *J. Cell Biol.* **150,** 921–928.
95. Wang, Z., Wong, P., Langbein, L., Schweizer, J., and Coulombe, P. A. (2003) Type II epithelial keratin 6hf (K6hf) is expressed in the companion layer, matrix, and medulla in anagen-stage hair follicles. *J. Invest. Dermatol.* **121,** 1276–1282.
96. Wojcik, S. M., Longley, M. A., and Roop, D. R. (2001) Discovery of a novel murine keratin 6 (K6) isoform explains the absence of hair and nail defects in mice deficient for K6a and K6b. *J. Cell Biol.* **154,** 619–630.
97. Wong, P., Domergue, R., and Coulombe, P. A. (2005) Overcoming functional redundancy to elicit pachyonychia congenita-like nail lesions in transgenic mice. *Mol. Cell Biol.* **25,** 197–205.
98. Wong, P. and Coulombe, P. A. (2003) Loss of keratin 6 (K6) proteins reveals a function for intermediate filaments during wound repair. *J. Cell Biol.* **163,** 327–337.
99. Wojcik, S. M., Imakado, S., Seki, T., et al. (1999) Expression of MK6a dominant-negative and C-terminal mutant transgenes in mice has distinct phenotypic consequences in the epidermis and hair follicle. *Differentiation* **65,** 97–112.
100. Paladini, R. D. and Coulombe, P. A. (1999) The functional diversity of epidermal keratins revealed by the partial rescue of the keratin 14 null phenotype by keratin 16. *J. Cell Biol.* **146,** 1185–1201.

101. Wawersik, M. J., Mazzalupo, S., Nguyen, D., and Coulombe, P. A. (2001) Increased levels of keratin 16 alter epithelialization potential of mouse skin keratinocytes in vivo and ex vivo. *Mol. Biol. Cell* **12,** 3439–3450.
102. Smith, F. (2003) The molecular genetics of keratin disorders. *Am. J. Clin. Dermatol.* **4,** 347–364.
103. Melton, D. W. (1994) Gene targeting in the mouse. *Bioessays* **16,** 633–638.
104. Plum, A., Hallas, G., Magin, T., et al. (2000) Unique and shared functions of different connexins in mice. *Curr. Biol.* **10,** 1083–1091.
105. Melton, D. W., Ketchen, A. M., and Selfridge, J. (1997) Stability of HPRT marker gene expression at different gene-targeted loci: observing and overcoming a position effect. *Nucleic Acids Res.* **25,** 3937–3943.
106. Abuin, A. and Bradley, A. (1996) Recycling selectable markers in mouse embryonic stem cells. *Mol. Cell Biol,* **16,** 1851–1856.
107. Rossant, J. and Nagy, A. (1995) Genome engineering: the new mouse genetics. *Nat. Med.* **1,** 592–594.
108. Jeannotte, L., Ruiz, J. C., and Robertson, E. J. (1991) Low level of Hox1.3 gene expression does not preclude the use of promoterless vectors to generate a targeted gene disruption. *Mol. Cell Biol.* **11,** 5578–5585.
109. Godwin, A. R., Stadler, H. S., Nakamura, K., and Capecchi, M. R. (1998) Detection of targeted GFP-Hox gene fusions during mouse embryogenesis. *Proc. Natl. Acad. Sci. USA* **95,** 13,042–13,047.
110. Le Mouellic, H., Lallemand, Y., and Brulet, P. (1990) Targeted replacement of the homeobox gene Hox-3.1 by the *Escherichia coli* lacZ in mouse chimeric embryos. *Proc. Natl. Acad. Sci. USA* **87,** 4712–4716.
111. Meyers, E. N., Lewandoski, M., and Martin, G. R. (1998) An Fgf8 mutant allelic series generated by Cre- and Flp-mediated recombination. *Nat. Genet.* **18,** 136–411.
112. Nagy, A., Moens, C., Ivanyi, E., et al. (1998) Dissecting the role of N-myc in development using a single targeting vector to generate a series of alleles. *Curr. Biol.* **8,** 661–664.
113. Askew, G. R., Doetschman, T., and Lingrel, J. B. (1993) Site-directed point mutations in embryonic stem cells: a gene-targeting tag-and-exchange strategy. *Mol. Cell Biol.* **13,** 4115–4124.
114. Stacey, A., Schnieke, A., McWhir, J., Cooper, J., Colman, A., and Melton, D. W. (1994) Use of double-replacement gene targeting to replace the murine alpha-lactalbumin gene with its human counterpart in embryonic stem cells and mice. *Mol. Cell Biol.* **14,** 1009–1016.
115. Wu, H., Liu, X., and Jaenisch, R. (1994) Double replacement: strategy for efficient introduction of subtle mutations into the murine Col1a-1 gene by homologous recombination in embryonic stem cells. *Proc. Natl. Acad. Sci. USA* **91,** 2819–2823.
116. Danielian, P. S., Muccino, D., Rowitch, D. H., Michael, S. K., and McMahon, A. P. (1998) Modification of gene activity in mouse embryos in utero by a tamoxifen-inducible form of Cre recombinase. *Curr. Biol.* **8,** 1323–1326.

117. Gu, H., Marth, J. D., Orban, P. C., Mossmann, H., and Rajewsky, K. (1994) Deletion of a DNA polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* **265,** 103–106.
118. Guo, C., Yang, W., and Lobe, C. G. (2002) A Cre recombinase transgene with mosaic, widespread tamoxifen-inducible action. *Genesis* **32,** 8–18.
119. Utomo, A. R., Nikitin, A. Y., and Lee, W. H. (1999) Temporal, spatial, and cell type-specific control of Cre-mediated DNA recombination in transgenic mice. *Nat. Biotechnol.* **17,** 1091–1096.
120. Wunderlich, F. T., Wildner, H., Rajewsky, K., and Edenhofer, F. (2001) New variants of inducible Cre recombinase: a novel mutant of Cre-PR fusion protein exhibits enhanced sensitivity and an expanded range of inducibility. *Nucleic Acids Res.* **29,** E47.
121. Schaft, J., Ashery-Padan, R., van der Hoeven, F., Gruss, P., and Stewart, A. F. (2001) Efficient FLP recombination in mouse ES cells and oocytes. *Genesis* **31,** 6–10.
122. Denk, H., Lackinger, E., Zatloukal, K., and Franke, W. W. (1987) Turnover of cytokeratin polypeptides in mouse hepatocytes. *Exp. Cell Res*. **173,** 137–143.
123. Yu, Y. and Bradley, A. (2001) Engineering chromosomal rearrangements in mice. *Nat. Rev. Genet.* **2,** 780–790.
124. Eggan, K., Akutsu, H., Loring, J., et al. (2001) Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc. Natl. Acad. Sci. USA* **98,** 6209–6214.
125. Eggan, K., Rode, A., Jentsch, I., et al. (2002) Male and female mice derived from the same embryonic stem cell clone by tetraploid embryo complementation. *Nat. Biotechnol.* **20,** 455–459.
126. Seibler, J., Zevnik, B., Kuter-Luks, B., et al. (2003) Rapid generation of inducible mouse mutants. *Nucleic Acids Res.* **31,** e12.
127. Nagy, A. G. M., Vintersten, K., and Behringer, R. (2003) *Manipulating the Mouse Embryo: A Laboratory Manual*, 3rd, Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
128. Magin, T. M., McEwan, C., Milne, M., Pow, A. M., Selfridge, J., and Melton, D. W. (1992a) A position- and orientation-dependent element in the first intron is required for expression of the mouse hprt gene in embryonic stem cells. *Gene* **122,** 289–296.
129. Magin, T. M., McWhir, J., and Melton, D. W. (1992b) A new mouse embryonic stem cell line with good germ line contribution and gene targeting frequency. *Nucleic Acids Res.* **20,** 3795–3796.
130. Ying, Q. L., Nichols, J., Chambers, I., and Smith, A. (2003) BMP induction of Id proteins suppresses differentiation and sustains embryonic stem cell self-renewal in collaboration with STAT3. *Cell* **115,** 281–292.
131. Ying, Q. L. and Smith, A. G. (2003) Defined conditions for neural commitment and differentiation. *Methods Enzymol.* **365,** 327–341.
132. Kupriyanov, S. and Baribault, H. (1998) Genetic control of extraembryonic cell lineages studied with tetraploid<—>diploid chimeric concepti. *Biochem. Cell Biol.* **76,** 1017–1027.

12

The HUVEC/Matrigel Assay

An In Vivo Assay of Human Angiogenesis Suitable for Drug Validation

Dag K. Skovseth, Axel M. Küchler, and Guttorm Haraldsen

Summary

The future ability to manipulate the growth of new blood vessels (angiogenesis) holds great promise for treating ischemic disease and cancer. Several models of human in vivo angiogenesis have been described, but they seem to depend on transgenic support and have not been validated in a therapeutic context. Here, we describe an in vivo assay that mimics human angiogenesis in which native human umbilical vein-derived endothelial cells are suspended in a liquid laminin/collagen gel (Matrigel), injected into immunodeficient mice, and develop into mature, functional vessels that vascularize the Matrigel plug in the course of 30 d. Moreover, we demonstrate how to target this process therapeutically by sustained delivery of the angiogenesis inhibitor endostatin from subcutaneously implanted microosmotic pumps.

Key Words: Adoptive transfer; angiogenesis; endostatin; human umbilical vein-derived endothelial cells; HUVECs; Matrigel; RAG2 mice.

1. Introduction

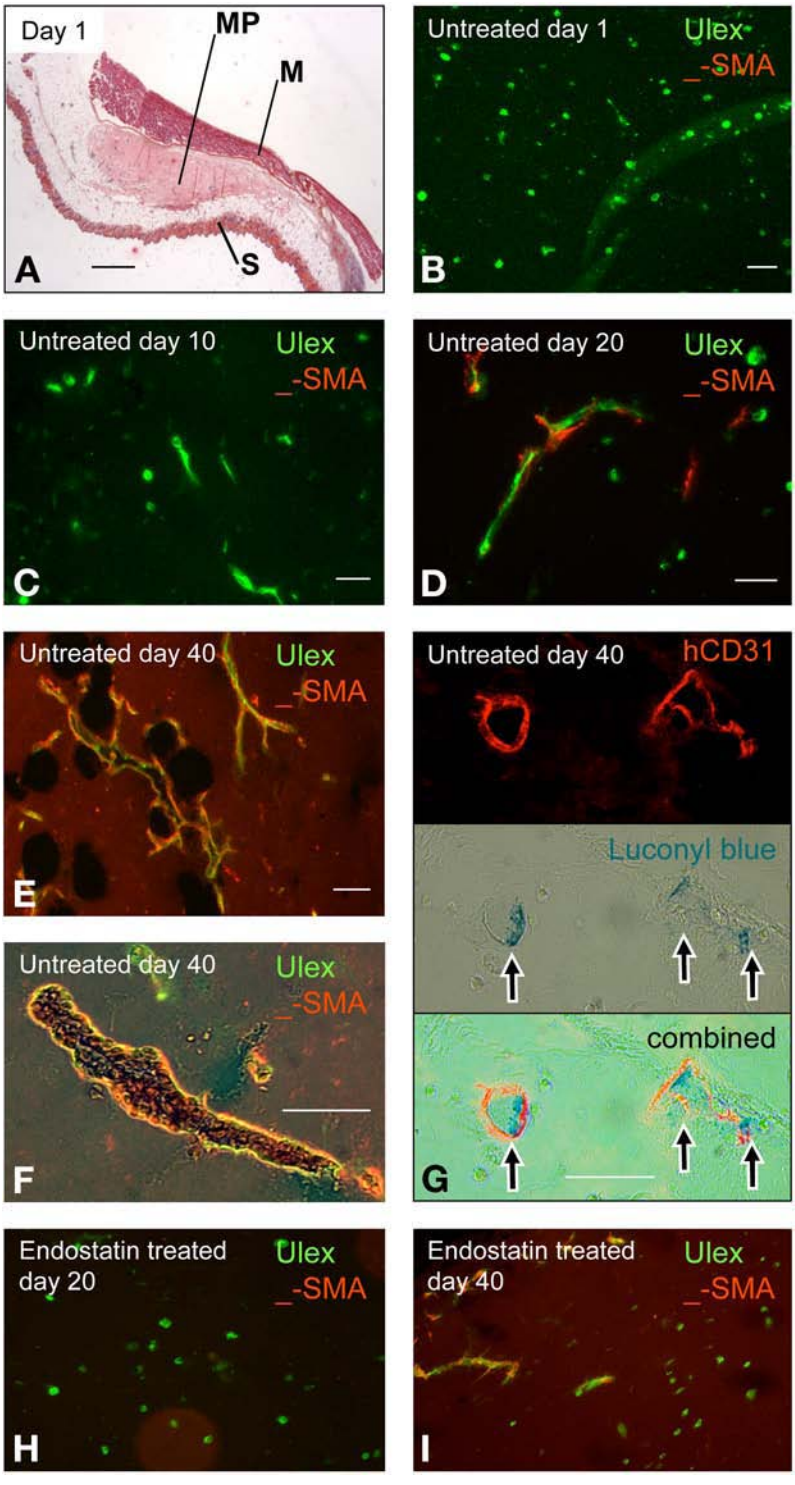
Angiogenesis is the growth of new capillaries from preexisting blood vessels ***(1,2)***. This multistep process begins with the activation of quiescent endothelial cells (ECs) leading to vessel dilation and permeabilization accompanied by degradation of the basement membrane. ECs then migrate toward the angiogenic stimulus and proliferate to form new patent tubules surrounded by a new basement membrane. The ECs of the sprouting, new vessels adapt to the local demand of the tissue by differentiation into capillaries, veins, or arteries, and they are subsequently stabilized by recruitment of perivascular cells (reviewed in **ref.** ***3***).

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarker Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

In the healthy adult, the vasculature remains quiescent, except in response to exercise-induced muscle hypertrophy and during the transient neovascularization of the uterus. However, angiogenesis is a prominent feature of wound repair and has the potential to become an important therapeutic modality in ischemic disease ***(4)***. Angiogenesis is also an attractive target of neoplastic tumor growth, in particular because ECs are genetically stable and therefore less likely to accumulate mutations that allow them to develop drug resistance ***(5)***. Indeed, the antiangiogenic approach to treat cancer has made substantial progress reflected in the recent approval of Avastin (a humanized antibody to vascular endothelial growth factor-A) for support therapy of colorectal cancer ***(6)***. However, it should be noted that Avastin's effect is restricted to prolonging progression-free survival times with an average of few months ***(7)***. Thus, although this achievement is a milestone toward antiangiogenic treatment of human cancer, it is still far from copying the full potential that Avastin® displayed in animal experiments. In our view, this discrepancy reflects the lack of both standardized tools and models that reliably mirror the complexity of tumor angiogenesis mechanisms in humans.

Characterization of candidates for treatment of angiogenesis-dependent diseases has traditionally been performed using a variety of in vitro and in vivo models (reviewed in **ref. *8***). Here, we briefly describe the most commonly used in vivo models based on host vessel growth. One classical in vivo approach has been the chick chorioallantoic membrane (CAM) assay ***(9)*** where angiogenesis can be monitored in response to implanted slow-release polymer pellets or organ grafts onto the developing CAM. Another assay exploits the avascular cornea in which slow release pellets induce vascularization from the limbus and offer another elegant way to study and manipulate angiogenesis in vivo ***(10)***. A third, frequently used in vivo assay is based on the injection of Matrigel into the abdominal wall of mice where it polymerizes at body temperature. Matrigel is a commercially available, solubilized basement membrane preparation extracted from the Engelbreth–Holm–Swarm mouse sarcoma-derived cell line. It contains laminin, collagen IV, heparan sulfate proteoglycans, entactin, and nidogen, and it gels rapidly at 22–35°C, forming a solid mass upon subcutaneous injection

Fig. 1. *(Opposite page)* Development of functional human vessels in Matrigel and inhibition of angiogenesis after endostatin treatment. **(A)** H&E staining of Matrigel plug (MP) positioned between the skin (S) and abdominal muscle layer (M). (**B–F, H,** and **I**) Immunofluorescence staining for α-smooth muscle actin (red) and *Ulex* lectin staining (green) of Matrigel sections from control and endostatin-treated animals on days as indicated. **(G)** Immunofluorescence staining for hCD31 with luconyl blue inside the vessels indicating functional vessels (phase contrast and immunofluorescent image of identical fields). Bars = 500 μm (A), 50 μm (B–I). Images (A, E, G, and G).


Day 1
MP
M
S
A
Untreated day 1
Ulex
_-SMA
B
Untreated day 10
Ulex
_-SMA
C
Untreated day 20
Ulex
_-SMA
D
Untreated day 40
Ulex
_-SMA
E
Untreated day 40
hCD31
Luconyl blue
combined
G
Untreated day 40
Ulex
_-SMA
F
Endostatin treated
day 20
Ulex
_-SMA
H
Endostatin treated
day 40
Ulex
_-SMA
I

and retaining its shape during the experimental period. Mouse ECs invade and vascularize the Matrigel plug in response to the level of angiogenic agents dissolved in the gel ***(11)***.

Although these model systems have certainly proven useful for the study of the angiogenic processes, they rely on the extrapolation of results obtained in nonhuman systems to the understanding of human disease. Recently, two experimental models of surgical implantation of human ECs have been published. In the first bioassay, developed by Nör and co-workers, microvascular ECs derived from the skin are suspended in a 1:1 mixture of Matrigel and EC growth medium and absorbed into poly(L-lactic acid) sponges before subcutaneous implantation to immunodeficient SCID mice ***(12,13)***. The ECs organize into functional vessels that carry mouse blood and are covered by α-smooth muscle actin-positive cells in the course of 21 d. The fraction of human ECs in the vessels dropped from 90–100% at day 14 to 50–65% at day 28, perhaps (as suggested by the authors) caused by mouse ECs replacing apoptotic human ECs. Furthermore, analysis of the inflammation markers intercellular adhesion molecule (ICAM)-1, vascular cell adhesion molecule (VCAM)-1, and E-selectin revealed their upregulation in the human endothelium peaking at day 7, possibly indicating proinflammatory activation of the microenvironment because of traumatization of the tissue. The second assay for transplantation of human ECs was reported by Schechner and colleagues ***(14,15)***. In this assay, Bcl-2–transduced human umbilical vein-derived ECs, HUVECs are suspended in a collagen/fibronectin gel and incubated for 20 h in vitro to obtain tubular structures before subcutaneous implantation into immunodeficient SCID/beige mice. Analysis of day 60 gels revealed functional vessels that had the characteristics of capillaries, venules, and arterioles, carried mouse blood, and were invested by α-smooth muscle actin-positive cells. The authors observed that the use of Bcl-2–transduced ECs enhanced the initial formation of vascular structures fourfold and that the ECs were dependent on Bcl-2 transduction to recruit vascular smooth muscle cells. Furthermore, transplantation of polymerized collagen/fibronectin gels where the 20-h initial tubular formation had not taken place in vitro resulted in avascular gels in the mice ***(14,15)***.

Here, we describe in detail a novel bioassay of human angiogenesis in which HUVECs suspended in Matrigel are transferred to immunodeficient RAG2 knockout mice by means of a single subcutaneous injection **(Fig. 1A)** and develop into mature vessels in the course of 30 days ***(16)***. Differentiated ECs form a functional vascular network by adapting sprouting, migrating, and proliferating phenotypes to vascularize the Matrigel plug. This process very closely resembles angiogenesis as it is described in the literature ***(3,17,18)***. Furthermore, we demonstrate the controlled, sustained delivery of endostatin, a powerful inhibitor of the angiogenic process ***(19)***.

To follow the fate of injected HUVECs and to relate them to the contribution of preexisting murine vessels, we double stained tissue sections with antibodies recognizing either human or mouse CD31, which in both species is a pan-endothelial cell marker. In another protocol, we detected human ECs and invading smooth muscle cells by means of *Ulex* lectin staining and an antibody to α-smooth muscle actin, respectively. Thus, specimens excised 1 or 3 d after Matrigel injection revealed single HUVECs uniformly distributed in the gel **(Fig. 1B)**. After 10 d, we observed single ECs with elongated cytoplasmic protrusions resembling pseudopods and also ECs that were lining up and in contact with other ECs **(Fig. 1C)**, particularly at the border of the Matrigel plug and close to preexisting mouse vessels. After 20 d, we observed HUVECs forming tubular structures in all areas of the gels and 50% of the vessels were covered by α-smooth muscle actin-positive cells **(Fig. 1D)**. Some of the vessel structures observed after 20 d also contained erythrocytes, indicative of functional vessels. On day 40, nearly all human vessels were fully mature, containing erythrocytes and being surrounded by α-smooth muscle actin-expressing cells (94.5%) **(Fig. 1E,F)**. Mature vessels can still be detected in the mice 100 d postinjection, and the human vessels are composed almost exclusively by human ECs and α-smooth muscle actin-expressing cells. To further investigate the maturity and functionality of the vessels, we injected intracardially a water-insoluble dye that does not leave the vascular circulation (luconyl blue) before termination of the experiment. Because the dye could be observed in the human vessels together with the erythrocytes, it nicely demonstrated that the vessels were circulated at the point of sacrifice **(Fig. 1G)**. Furthermore, no signs of infection or inflammation could be observed after injection and during the experimental period, as assessed by general histology and the finding that the proinflammatory markers ICAM-1, VCAM-1, and E-selectin were negative at all time-points.

Several tumor-derived, circulating angiogenesis inhibitors generated in vivo by proteolytic degradation have been recently identified (reviewed in **refs.** ***20*** and ***21***). In particular, a 20-kDa C-terminal proteolytic fragment of collagen XVIII, termed endostatin, inhibits tumor growth in several animal models ***(5,22–26)***. Thus, it is not only considered a promising anticancer drug but also has so far failed to reproduce its properties in human clinical trials ***(27–29)***. Although the mechanisms of action are incompletely understood, endostatin is thought to interact with several endothelial cell surface receptors that are critically involved in angiogenesis (reviewed in **ref.** ***30***).

In addition to the model itself, we also show how to target adoptively transferred human endothelial cells by using sustained delivery of the angiogenesis

inhibitor endostatin as an example. To this end, we subcutaneously implanted microosmotic pumps that released rhendostatin at a constant rate of 2 µg/h throughout the experimental period (*see* **Note 1**). To allow activation of the osmotic pumps, Matrigel injections were performed 2 d after pump implantation. In the presence of endostatin, single HUVECs almost completely failed to migrate or accumulate into vessels. The fraction of cells with a migratory phenotype (cells expressing F-actin and extending pseudopods) was at day 10 strongly reduced from 50% in untreated animals to 13% in the presence of endostatin. Furthermore, the number of human vessels was reduced by 95% at day 20 **(Fig. 1H)** and by 99.5% at days 30 and 40 **(Fig. 1I)** after treatment with endostatin based on a vessel:single cell ratio. We also assessed the maturation of the vessels and found that the ratio of mature vessels (vessels that were covered by α-smooth muscle actin-expressing cells) was reduced from 64.3% in untreated to 28.6% in endostatin-treated animals at day 30. To investigate the possible mechanisms that may have reduced the ratio of vessels covered with α-smooth muscle actin-expressing cells because of endostatin treatment, we analyzed endothelial cell-expressed platelet-derived growth factor (PDGF)B (currently considered the most important recruiter of pericytes during angiogenesis; ***[31]***) by means of *in situ* hybridization with a probe to human (h)PDGFB in costaining with *Ulex* lectin. Whereas PDGFB transcripts were found in most vascular constructs in both groups, the remaining single cells that dominate the gels of endostatin-treated mice were negative. To more accurately quantify the mRNA levels of hPDGFB at days 10 and 20, we also performed quantitative PCR based on mRNA from immunopurified human endothelial cells of enzyme-digested Matrigel plugs and human transcript-specific primers. In line with our *in situ* observations, we found that endostatin reduced the copy numbers of hPDGFB by more than 90 and 70% at days 10 and 20, respectively, compared with untreated littermates.

The in vivo HUVEC/Matrigel assay is a convenient and a powerful tool for the study of vascular morphogenesis, physiological and pathological angiogenic responses, and functional studies on mature vessels such as inflammatory responses. It offers several possibilities in terms of experimental design, i.e., a growth factor or inhibitor could be administered to the endothelium at different stages of vascular development. For example, serial injections of 3×300 U of recombinant human tumor necrosis factor-α over 24 h at the edge of day 60 Matrigel plugs revealed extravasation of mouse leukocytes through the human vessels (Skovseth and Haraldsen, unpublished data). Furthermore, incorporation of growth-arrested, transfected cell lines or tumor cell lines into the Matrigel plug is another possibility for manipulating endothelial cell phenotype regulation.

2. Materials

2.1. Cell Preparation

1. Collagenase (cat no. C8176, Sigma).
2. HUVECs isolated from umbilical cords obtained from Rikshospitalet University Hospital (Oslo, Norway).
3. MCDB-131 (Invitrogen).
4. Fetal calf serum (cat. no. F7524, Sigma).
5. Recombinant human basic fibroblast growth factor (rhbFGF; Sigma).
6. Recombinant human epidermal growth factor (rhEGF; Sigma).
7. Hydrocortisone (Sigma).
8. L-Glutamine (Sigma).
9. Amphotericine B (Invitrogen).
10. Gentamicin (Invitrogen).
11. Trypsin/EDTA (Cambrex Bio Science Copenhagen ApS).

2.2. Mice and Animal Experiments

1. RAG2 knockout mice (cat. no. 000601-M, Taconic Farms).
2. Matrigel (BD Matrigel basement membrane matrix, BD Biosciences).
3. Forceps.
4. Scissors.
5. Suture, nonabsorbable.
6. Hypnorm (fentanyl + fluanisone)/Dormicum (midazolam) solution for anesthesia.
7. Surgical forceps.
8. 1- to 5-mL Syringe depending on volume used.
9. 23-Gauge injection needle.
10. Alzet microosmotic pumps (models 1002 [0.25 µL/h, 14 d] or 2004 [0.25 µL/h, 28 d], Alzet).
11. Recombinant human endostatin derived from *Pichia pastoris* (gift from Entremed, available from Sigma).
12. Luconyl blue 9600 (BASF Corporation).

2.3. Staining of Endothelial Cells

2.3.1. Reagents and Equipment for Tissue Processing, Histology, and Immunohistochemistry

1. OCT (Tissue-Tek).
2. Cryomold for OCT embedding (Tissue-Tek).
3. Methanol.
4. Ethanol.
5. Xylene.
6. Cryostat.
7. Microtome.
8. Liquid nitrogen.
9. Acetone.

10. Nikon E-800 fluorescence microscope (Nikon).
11. F-view digital camera controlled by the AnalySIS 3.2 Software (Soft Imaging Systems GmbH).
12. Reagents for H&E staining (hematoxylin, hexamine, azophloxine, saffron, and ethanol).
13. Phosphate-buffered saline (PBS).
14. Bovine serum albumin to dilute antibodies (1.25% for primary reagents and 12.5% for secondary reagents).
15. SuperFrost or chromalun-coated glass slides.
16. 10% Formalin.

2.3.2. Primary Antibodies (Clone, Working Dilution, Specification, Source)

1. hCD31 (hec7, 0.4 μg/mL, mouse IgG2a, W. A. Muller, New York).
2. hICAM-1 (RRI/1.1.1, 10 μg/mL, mouse IgG1, R. Rothlein, Ridgefield, CT).
3. hVCAM-1 (hec7, 0.4 μg/mL, mouse IgG2a, W. A. Muller, New York).
4. hE-selectin (hec7, 0.4 μg/mL, mouse IgG2a, W. A. Muller, New York).
5. mCD31 (Mec 13.3, 1/100, rat IgG2a, A. Vecchi, Milan).
6. α-Smooth muscle actin, α-SMA (M0851, 10 μg/ml, mouse IgG2a, DakoCytomation www.dako.com).

2.3.3. Secondary Reagents

1. Alexa 594-conjugated goat anti-rat IgG (10 μg/mL, Invitrogen).
2. Alexa 546-conjugated goat anti-mouse IgG (1 μg/mL, Invitrogen).
3. Fluorescein isothiocyanate (FITC)-conjugated *Ulex europaeus* lectin (1/400, Vector Laboratories).
4. Rhodamine-conjugated phalloidin (1/200, Sigma).

2.4. Re-isolation of Endothelial Cells, mRNA Purification, and Analysis

1. Collagenase (cat. no. C8167, Sigma).
2. Dispase II (Roche Diagnostics).
3. DNaseI (Roche Diagnostics).
4. Complete MCDB-131 (containing 7.5% fetal calf serum, 10 ng/mL rhEGF, 1 ng/mL rhbFGF, 1 μg/mL hydrocortisone, 50 μg/mL gentamicin, and 0.25 μg/mL amphotericin B).
5. 40-μm Filter (Falcon).
6. Rock'n roller.
7. Antibody to hCD31 (hec7).
8. Dynabeads M-450 coated with sheep anti-mouse IgG (Dynal Biotech).
9. Dynal MPC-6 magnet (Dynal Biotech, Oslo).
10. RNeasy mini kit (QIAGEN GmbH).
11. Dithiothreitol (DDT), oligo(dT), and Superscript III reverse transcriptase (Invitrogen).
12. Taq polymerase and buffer, dNTPs.

13. PDGF-β LightCycler kit (Search-LC).
14. Human PDGFB-specific PCR primers 5′-CAT AGA CCG CAC CAA CGC CAA CTT C-3′ as forward primer and 5′-ATC TTT CTC ACC TGG ACA GGT CGC AG-3′ as reverse primer.
15. Human hydroxymethylbilane synthase-specific PCR primers 5′-GTA ACG GCA ATG CGG CTG-3′ as forward primer and 5′-GGT ACG AGG CTT TCA ATG-3′ as reverse primer.
16. QuantiTect SYBR Green PCR kit (QIAGEN GmbH).
17. LightCycler (Roche Diagnostics).

2.5. In Situ *Hybridization*

2.5.1. Probe Generation and Labeling

1. A chloramphenicol-resistant pBC Vector containing a 1-kb insert of hPDGFB was kindly provided by C. Betsholtz (Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden).
2. *Eco*RI and *Not*I restriction enzymes (New England Biolabs).
3. Phenol/chloroform extraction: phenol saturated TE8, phenol isoamyl alcohol, sodium acetate, 96% ethanol, and 70% ethanol.
4. DIG RNA labeling kit (Roche Diagnostics).
5. T7 and T3 RNA polymerases (Roche Diagnotics).
6. Ribohybe (Ventana Medical Systems).

2.5.2. In Situ *Hybridization*

1. Ventana Discovery automated histology robot (Ventana Medical Systems).
2. Ventana solutions as supplied by Ventana Medical Systems: mild alkaline protease 3, Ribo CC, standard saline citrate (SSC), and BlueMap kit for developing the nitroblue tetrazolium salt/5-bromo-4-chloro-3-indolyl phosphate alkaline phosphatase substrate.
3. Biotinylated anti-digoxin antibody (Jackson ImmunoResearch Laboratories).
4. DIG-labeled RNA probe.

2.5.3. Double Staining of In Situ *Hybridized Sections With* Ulex *Lectin*

1. 70, 96, and 100% ethanol baths.
2. Xylene bath.
3. FITC-labeled *Ulex* lectin and polyvinyl alcohol (PVA).

3. Methods

3.1. Cell Preparation

1. Isolate HUVECs from the umbilical vein with 0.2% collagenase as described by Jaffe et al. ***(32)***.
2. Culture HUVECs in complete MCDB-131 (containing 7.5% fetal calf serum, 10 ng/mL rhEGF, 1 ng/mL rhbFGF, 1 µg/mL hydrocortisone, 50 µg/mL gentamicin,

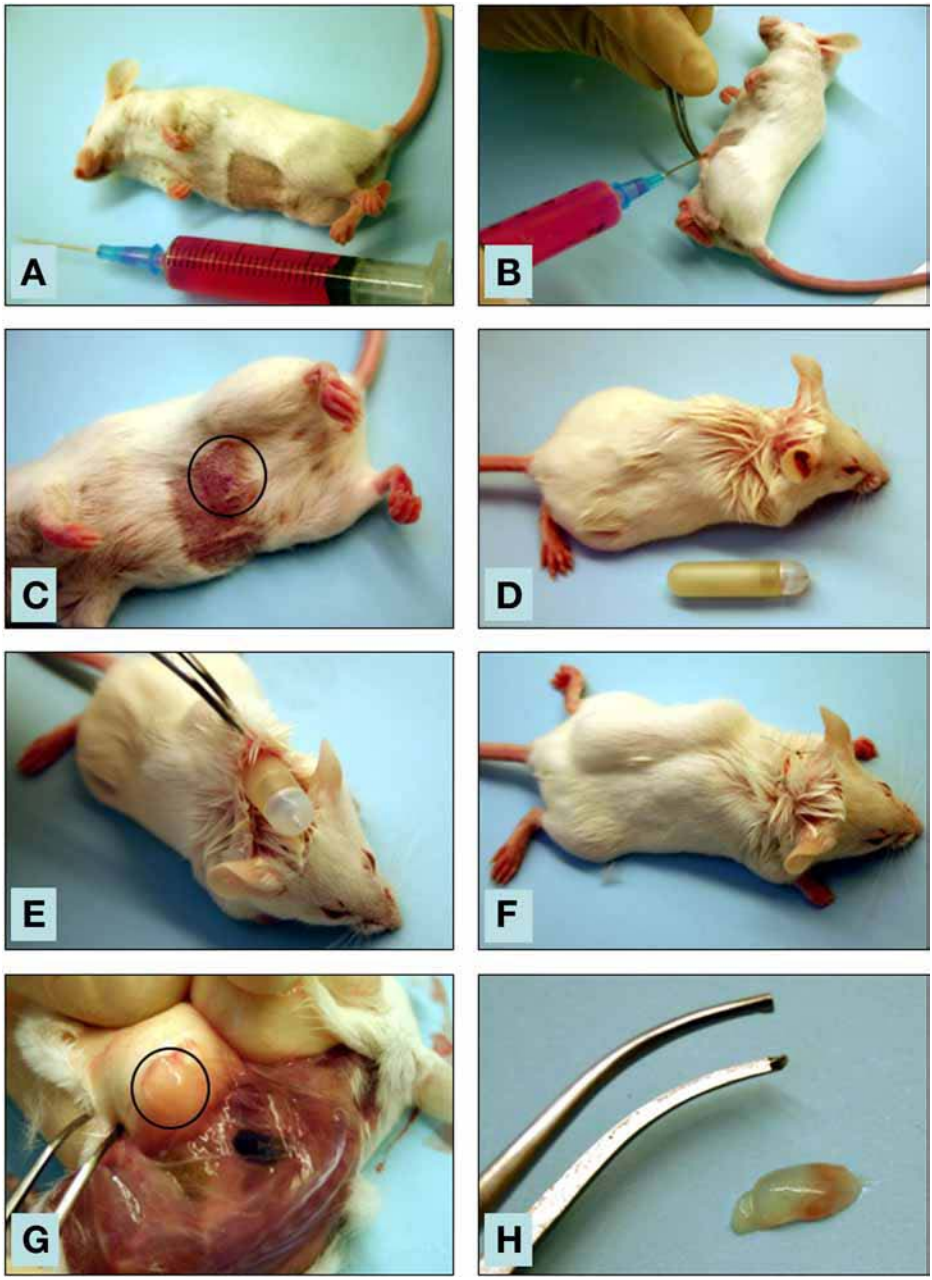


Fig. 2. Animal handling. **(A–C)** Injection of Matrigel. **(D–F)** Implantation of an endostatin-filled microosmotic pump. **(G)** Harvesting of a Matrigel plug for RNA analysis. **(H)** Vascularized Matrigel plug. Note the Matrigel plug after injection in C and the polymerized Matrigel under the skin in G (both marked with a circle).

and 0.25 µg/mL amphotericin B) at 37°C in a humid 95% air, 5% CO_2 atmosphere and split the cells at ratios 1:3.

3.2. Animal Experiments

3.2.1. Transfer of Human Endothelial Cells Into RAG2 Knockout Mice

1. Trypsinize subconfluent HUVECs, wash, and count. Pellet the cells (3×10^5 cells per plug) by means of another centrifugation and discard all extra medium from the tube (*see* **Note 2**).
2. Thaw Matrigel on ice and adjust it to a final protein concentration of 8 mg/mL with sterile PBS. Add 200 µL of Matrigel per plug to the cell pellet and mix gently to avoid foaming.
3. Anesthetize RAG2 knockout mice of 4–12 wk of age by injection of 150 µL of a saturated mixture of hypnorm + dormicum in the abdominal cavity.
4. Inject the Matrigel/HUVEC mix (200 µL per plug) subcutaneously on the abdominal midline of the anesthetized animal with a 23-gauge needle (**Fig. 2A,B**; *see* **Note 3**). Leave the mouse on its back to allow gelation of the Matrigel as it will rapidly gel at room temperature **(Fig. 2C)**. It is also possible to inject up to four gels into one mouse: one on either side of the abdominal midline and one laterally to the former in either flanks.
5. Sacrifice the mice by cervical dislocation upon termination of the experiment.

3.2.2. Implantation of Endostatin Sontaining Microosmotic Pumps

1. Fill Alzet microosmotic pump (model 1002 for experiments up to 14 d or model 2004 for up to 28 d) with a 7.9 mg/mL solution of recombinant endostatin.
2. Anesthetize mice with Hypnorm/Dormicum.
3. Wash the neck with 70% ethanol.
4. Make a small midline incision in the skin and carefully make a subcutaneous pocket by blunt dissection towards the tail, large enough to accommodate the pump.
5. Insert the pump and make sure the edges of the skin adapt without strain. If necessary, extend the pocket to relieve pressure from the implant **(Fig. 2D,E)**.
6. Close the wound with nonabsorbable suture **(Fig. 2F)**.
7. Carefully inspect the wound 1 d postinjection for signs of inflammation and infection before proceeding to injection of the Matrigel plugs.

3.3. Histology and Immunohistochemical Analysis of Matrigel Plugs

3.3.1. Methods for Processing and Staining

1. Sacrifice the mice and use scissors to gently cut out the abdominal wall around the injected Matrigel plug, making sure to include all tissue layers. Transfer the harvested sample to PBS on ice. Fix samples in methanol or 10% formalin (4°C; 24 h), or snap-freeze them in OCT in liquid nitrogen.
2. Embed the methanol- or formalin-fixed tissues into paraffin and cut 4-µm-thick sections. Deparaffinize and rehydrate the methanol-fixed sections in successive

baths of xylene (5 min; 20°C), 100% ethanol (1 min; 20°C), 96% ethanol (1 min; 20°C), 70% ethanol (1 min; 20°C), and finally PBS (2 min twice; 20°C) before staining.
3. Cut frozen sections on a cryotome (4 µm) and fix in acetone (20°C; 15 min) before staining.
4. Stain methanol-fixed and the acetone-fixed sections with H&E for histological analysis to facilitate identification of Marigel plug.
5. Stain the frozen sections with primary and secondary antibodies with each 1-h incubation step at room temperature for immunohistochemistry. Wash between these steps in PBS for 1 min.
6. Stain the deparaffinized methanol- or formalin-fixed sections with the primary antibody (overnight), followed by a 3–h incubation with a secondary fluorochrome-conjugated reagent for immunohistochemistry. Wash between these steps in PBS for 10 min.
7. For two-parameter analysis, add rhodamine-conjugated phalloidin or FITC-conjugated *Ulex* lectin to the primary antibody solution.

3.4. Harvesting of Human ECs, mRNA Isolation, and Expression Analysis of Human Endothelium

3.4.1. Isolation of Human Endothelium From Matrigel Plugs

1. Dissect Matrigel plugs from the skin and muscle layers **(Fig. 2G, H)** and cut into small pieces with a scalpel on ice.
2. Incubate the Matrigel fragments in a mixture of 0.1% collagenase, 0.2% dispase, and 20 µg/mL DNaseI in PBS at 37°C for 40 min on a magnet stirrer (*see* **Note 4**).
3. Wash the cell suspension in 40 mL of cold complete MCDB-131.
4. Filter the cell suspension through a 40-µm sieve filter to remove debris.
5. Incubate 1 mL of cell suspension with a primary mouse antibody to 0.3 µg/mL hCD31 (15 min on ice).
6. Wash the cell suspension in 15 mL of cold complete MCDB-131.
7. Incubate 1 mL of the cell suspension with sheep anti-mouse immunoglobulin-coated Dynabeads according to the recommendation of the manufacturer.
8. Dilute the cell suspension to 5 mL and extract the rosetting cells on a Dynal magnet. Wash the cells by repeating the extraction of rosetting cells five times.
9. Spin down the cells and isolate RNA.

3.4.2. Isolation of RNA

1. Isolate total RNA and perform on-column DNaseI treatment by using the RNeasy mini kit according to instructions of the manufacturer.
2. Reverse transcribe 200 ng of total RNA by using oligo(dT) and the Superscript III reverse transcriptase kit according to the manufacturer's instructions. In brief, mix RNA, dNTPs, buffer, DDT, oligo(dT), RNasin, and enzyme to a final volume of 20 µL and incubate for 1 h at 50°C and subsequently for 15 min at 70°C.

3.4.3. Analysis of hPDGFB RNA by Quantitative Reverse Transcription-PCR

1. Amplify human PDGFB-specific transcripts by using 5′-CAT AGA CCG CAC CAA CGC CAA CTT C-3′ as forward primer and 5′-ATC TTT CTC ACC TGG ACA GGT CGC AG-3′ as reverse primer at 3.0 m*M* $MgCl_2$ and 68°C annealing temperature.
2. In brief, for hPDGFB determination in the cDNA samples, mix the cDNA (0.5 µL of the step 1 reaction), dNTPs, primers, $MgCl_2$, and enzyme mix containing SYBR Green to a final volume of 20 µL, load into the glass capillaries, and subsequently perform quantitative real-time PCR by using LightCycler. Read the fluorescence intensities at 72°C after each PCR cycle.
3. Quantify hPDGFB copy numbers by using the standard delivered with the PDGF-β kit from Search-LC.
4. Quantify the human housekeeping gene hydroxymethylbilane synthase by using the QuantiTect SYBR Green PCR kit at 60°C annealing temperature, 3.5 m*M* $MgCl_2$ concentration with forward primer 5′-GTA ACG GCA ATG CGG CTG-3′ and reverse primer 5′-GGT ACG AGG CTT TCA ATG-3′ as internal control for the amount of human RNA input.
5. Determine the actual copy numbers of the transcripts in the samples by correlation between the crossing points of the respective PCRs on test samples and standard samples by using LightCycler Software 3.

3.5. In Situ *Hybridization of hPDGFB on Human Endothelium in Mice*

3.5.1. Probe Generation and Labeling

1. A chloramphenicol-resistant pBC vector containing a 1-kb insert for hPDGFB flanked by upstream EcoRI and downstream NotI was kindly provided by C. Betsholtz (Department of Medical Biochemistry and Biophysics, Karolinska Institute, Stockholm, Sweden).
2. Linearize the plasmid by overnight incubation with restriction enzymes EcoRI for the antisense and NotI for the sense probe. Purify the probes by phenol/chloroform extraction followed by ethanol precipitation.
3. Transcribe and label the riboprobes by using the DIG RNA labeling kit. Add transcription buffer, RNase inhibitor, NTP labeling mix, and RNA polymerases T7 or T3 for the antisense or sense probes, respectively, to the purified templates and incubate for 2 h at 37°C. Add DNaseI to the probes and incubate for 15 min (37°C).
4. Rinse the probes by overnight precipitation in ammonium acetate, followed by ethanol wash twice and resuspension in 10 µg/mL Ribohybe.
5. Prepare the probes for Ventana robot by adjusting the concentration to 0.1 µg/mL in Ribohybe.

3.5.2. In Situ *Hybridization*

1. Prepare freshly cut (4-µm) formalin-fixed tissue sections of Matrigel plugs on SuperFrost glass slides and air-dry at 37°C overnight.

2. Place slides in Ventana robot and activate the hybridization protocol. Slides will be deparaffinized and pretreated for enhanced detection. The probes that are described here were most efficiently visualized by choosing a mild protocol for target retrieval (alkaline protease 3 and Ribo CC).
3. Add 100 µL of the sense or antisense dilutions to the sections. The hybridization is performed at 60°C (6 h) with three stringency washes in 2X SSC (6 min; 60°C). Riboprobes are detected with biotinylated 2.2 µg/mL anti-digoxin (32 min; 37°C) and the BlueMap kit.
4. Retrieve the slides from the robot and wash the hybridized sections in hot water to remove the oil.

3.5.3. Double Staining With Ulex Lectin

1. Wash freshly hybridized sections after removal of oil in baths of 70% alcohol (1 min; 20°C), 96% alcohol (1 min; 20°C), 100% alcohol (1 min; 20°C), xylene (2 min; 20°C) (*see* **Note 5**), 100% alcohol (1 min; 20°C), 96% alcohol (1 min; 20°C), 70% alcohol (1 min; 20°C), and finally PBS (2 min twice; 20°C).
2. Incubate the sections with FITC-labeled *Ulex* lectin for 1 h at room temperature.
3. Wash the sections in PBS (2 min twice) and mount in PVA (*see* **Note 5**).

4. Notes

1. Endostatin released from the osmotic pumps can be detected throughout the experimental period as demonstrated by analysis of endostatin in serum by using an ELISA kit from CytImmune Sciences ***(19)***.
2. In our studies, we have used freshly isolated confluent HUVECs in low passage (passage 1–4 is preferable). A pool of three individual cell cultures in low passage gives reproducible results in the model system.
3. At all time-points before injection, keep the Matrigel and the Matrigel/HUVEC mix on ice. It is also important to gently mix the Matrigel/EC suspension in the syringe before injection to avoid sedimentation of ECs. This mixing can be done by letting a small air bubble travel through the gel in the syringe.
4. It is important to use sufficient amounts of the collagenase/dispase solution to completely dissolve the Matrigel. Use at least 6 mL of solution for one to five plugs and 10 mL of solution for six to 12 plugs.
5. Usually, *in situ*-hybridized sections would be mounted in Eukitt for greater stability of the 5-bromo-4-chloro-3′-indolyphosphate *p*-toluidine (BCIP) signal, but this is not favorable for the fluorescent FITC signal. We therefore chose to mount the stained sections with PVA, which is the standard mounting medium for fluorescent signals. Because the PVA does not stabilize the BCIP signal, we recommend analyzing the sections immediately.

References

1. Risau, W. (1997) Mechanisms of angiogenesis. *Nature* **386,** 671–674.
2. Carmeliet, P. (2000) Mechanisms of angiogenesis and arteriogenesis. *Nat. Med.* **6**, 389–395.

3. Folkman, J. (2003) Fundamental concepts of the angiogenic process. *Curr. Mol. Med.* **3,** 643–651.
4. Losordo, D. W. and Dimmeler, S. (2004) Therapeutic angiogenesis and vasculogenesis for ischemic disease. Part II: Cell-based therapies. *Circulation* **109,** 2692–2697.
5. Boehm, T., Folkman, J., Browder, T., and O'Reilly, M. S. (1997) Antiangiogenic therapy of experimental cancer does not induce acquired drug resistance. *Nature* **390**, 404–407.
6. Hurwitz, H., Fehrenbacher, L., Novotny, W., et al. (2004) Bevacizumab plus irinotecan, fluorouracil, and leucovorin for metastatic colorectal cancer. *N. Engl. J. Med*. **350,** 2335–2342.
7. Ferrara, N., Hillan, K. J., and Novotny, W. (2005) Bevacizumab (Avastin), a humanized anti-VEGF monoclonal antibody for cancer therapy. *Biochem Biophys. Res. Commun*. **333,** 328–335.
8. Auerbach, R., Lewis, R., Shinners, B., Kubai, L., and Akhtar, N. (2003) Angiogenesis assays: a critical overview. *Clin. Chem.* **49,** 32–40.
9. Folkman, J. (997) Proceedings: tumor angiogenesis factor. *Cancer Res*. **34,** 2109-2113.
10. Gimbrone, M. A., Jr., Leapman, S. B., Cotran, R. S., and Folkman, J. (1973) Tumor angiogenesis: iris neovascularization at a distance from experimental intraocular tumors. *J. Natl. Cancer Inst.* **50,** 219–228.
11. Passaniti, A., Taylor, R. M., Pili, R., et al. (1992) A simple, quantitative method for assessing angiogenesis and antiangiogenic agents using reconstituted basement membrane, heparin, and fibroblast growth *factor. Lab. Invest.* **4,** 519–528.
12. Nor, J. E., Christensen, J., Mooney, D. J., and Polverini, P. J. (1999) Vascular endothelial growth factor (VEGF)-mediated angiogenesis is associated with enhanced endothelial cell survival and induction of Bcl-2 expression. *Am. J. Pathol.* **154,** 375–384.
13. Nor, J. E., Peters, M. C., Christensen, J. B., et al. (2000) Engineering and characterization of functional human microvessels in immunodeficient mice. *Lab. Investig*. **81,** 453–463.
14. Schechner, J. S., Nath, A. K., Zheng, L., et al (2000) In vivo formation of complex microvessels lined by human endothelial cells in an immunodeficient mouse. *Proc. Natl. Acad. Sci. USA* **97,** 9191–9196.
15. Enis, D. R., Shepherd, B. R., Wang, Y., et al. (2005) Induction, differentiation, and remodeling of blood vessels after transplantation of Bcl-2-transduced endothelial cells. *Proc. Natl. Acad. Sci. USA* **102,** 425–430.
16. Skovseth, D. K., Yamanaka, T., Brandtzaeg, P., Butcher, E. C., and Haraldsen, G. (2002) Vascular morphogenesis and differentiation after adoptive transfer of human endothelial cells to immunodeficient mice. *Am. J. Pathol*. **160,** 1629–1637.
17. Folkman, J. (1995) Angiogenesis in cancer, vascular, rheumatoid and other disease. *Nat. Med*. **1,** 27–31.
18. Carmeliet, P. and Jain, R. K. (2000) Angiogenesis in cancer and other diseases. *Nature* **407,** 249–257.

19. Skovseth, D. K., Veuger, M. J., Sorensen, D. R., De Angelis, P.M., and Haraldsen, G. (2005) Endostatin dramatically inhibits endothelial cell migration, vascular morphogenesis, and perivascular cell recruitment in vivo. *Blood* **105,** 1044–1051.
20. Cao, Y. (2001) Endogenous angiogenesis inhibitors and their therapeutic implications. Int. J. Biochem. *Cell Biol*. **33,** 357–369.
21. Marneros, A. G. and Olsen, B. R. (2001) The role of collagen-derived proteolytic fragments in angiogenesis. *Matrix Biol.* **20,** 337–345.
22. O'Reilly, M. S., Boehm, T., Shing, Y., et al. (1997) Endostatin: an endogenous inhibitor of angiogenesis and tumor growth *Cell* **88,** 277–285.
23. Dhanabal, M., Ramchandran, R., Volk, R., et al. (1999) Endostatin: yeast production, mutants, and antitumor effect in renal cell carcinoma. *Cancer Res*. **59,** 189–197.
24. Read, T. A., Sorensen, D. R., Mahesparan, R., et al. (2001) Local endostatin treatment of gliomas administered by microencapsulated producer cells. *Nat. Biotechnol*. **19,** 29–34.
25. Sorensen, D. R., Read, T. A., Porwol, T., et al. (2002) Endostatin reduces vascularization, blood flow, and growth in a rat gliosarcoma. *Neurooncology* **4,** 1–8.
26. Sorensen, D. R., Leirdal, M., Iversen, P. O., and Sioud, M. (2002) Combination of endostatin and a protein kinase C alpha DNA enzyme improves the survival of rats with malignant glioma. *Neoplasia* **4,** 474–479.
27. Eder, J. P., Jr., Supko, J. G., Clark, et al. (2002) Phase I clinical trial of recombinant human endostatin administered as a short intravenous infusion repeated daily. *J. Clin. Oncol.* **23,** 772–784.
28. Herbst, R. S., Hess, K. R., Tran, H. T., et al. (2002) Phase I study of recombinant human endostatin in patients with advanced solid tumors. *J. Clin. Oncol.* **20**, 3792–3803.
29. Thomas, J. P., Arzoomanian, R. Z., Alberti, D., et al. (2003) Phase I pharmacokinetic and pharmacodynamic study of recombinant human endostatin in patients with advanced solid tumors. *J. Clin. Oncol.* **21,** 223–231.
30. Abdollahi, A., Hlatky, L., and Huber, P. E. (2005) Endostatin: the logic of antiangiogenic therapy. *Drug Resist. Updat.* **8,** 59–74.
31. Lindahl, P., Johansson, B. R., Leveen, P., and Betsholtz, C. (1997) Pericyte loss and microaneurysm formation in PDGF-B-deficient mice. *Science* **277,** 242–245.
32. Jaffe, E. A., Nachman, R. L., Becker, C. G., and Minick, C. R. (1973) Culture of human endothelial cells derived from umbilical veins. Identification by morphologic and immunologic criteria. *J. Clin. Invest.* **52,** 2745–2756.

13

A Murine Model for Studying Hematopoiesis and Immunity in Heart Failure

Per Ole Iversen and Dag R. Sørensen

Summary

Recent epidemiological research indicates that a coexistent anemia among patients with heart failure might worsen their prognosis. However, whether the reduced synthesis of red blood cells is a contributing factor to the development and progression to overt heart failure, or whether it simply is a mere consequence of a dysfunctional heart, remains to be elucidated. Studies in mice with experimentally induced acute myocardial infarction leading to subsequent development of a postinfarction congestive heart failure have shed some light on this problem. Careful analyses of the number and of the functions of various hematopoietic cells residing in either blood or bone marrow point to a possible inhibitory role of cytokines, such tumor necrosis factor α, on hematopoiesis. The present protocols will hopefully encourage further studies of hematopoiesis and immunity in heart failure by using a combination of animal models with state-of-the-art techniques in molecular biology to define and validate possible targets for therapy.

Key Words: Cytokines; heart failure; hematopoiesis; mouse; myocardial infarction.

1. Introduction

Heart failure has for a long time remained one of the most common cardiovascular disorders encountered in various Western populations ***(1,2)***. With the growing numbers of obese-related pathological conditions, heart failure also is expected to rise in developing countries in the years to come. The associated disease-burden amounts to considerable resources used in terms of money and health care.

In recent years, several important advances have been made in terms of pharmacology, including both new drugs (e.g., angiotensin II receptor blockers) as well as altered dosages of established drugs (e.g., beta-blockers). We also have

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols*
Volume I, Emerging Strategies for Targets and Biomarker Discovery
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

witnessed important new contributions in terms of revascularization regimens, e.g., coronary artery bypass surgery and angioplasty. However, in spite of these advances, the prognosis of the heart failure patient remains dismal with high morbidity and mortality ***(3)***.

Although the heart failure syndrome is generally characterized primarily as a hemodynamic disorder with impaired left ventricular function as the main hallmark, we are still short of a comprehensive understanding of the pathology of this debilitating disorder. Refinement of tools for measuring minute concentrations of biologically active molecules as well as the advent of genetically altered mice has changed our view of heart failure as a mere circulating disorder to that of a possible inflammatory condition. In line with this changed view, several cytokines seem to play important roles in the pathogenesis of heart failure, at least in experimental models. Hence, in particular, the proinflammatory cytokine tumor necrosis factor (TNF)α, and also the interleukins (IL)-1β and -18, seem to exert cardiodepressive effects in various in vitro and in vivo settings ***(4–6)***.

Paralleling this development is the recognition that patients with heart failure also suffer from conditions related to an impairment of bone marrow function, in particular, hematopoiesis. Although not well documented in the past, it has been a clinical experience that the hemoglobin concentrations of these patients usually are somewhat reduced. Based on autopsies, Bohm ***(7)*** concluded that a gelatin transformation of the bone marrow was more frequent among patients deceased from heart failure. Later, epidemiological studies concluded that anemia significantly predicted a worsened outcome among heart failure patients who were anemic compared with those with normal hemoglobin levels ***(8,9)***. The crucial question is whether the anemia has a true negative impact on the development and progression to overt heart failure or whether it simply is a result of a deteriorating bone marrow function because of the failing heart ***(10)***.

Although wild mice and rats can survive almost any climatic condition, their laboratory relatives have lost the genetic background for this adaptability. When animals are bred for experimental use, one major concern is to produce individuals or populations with a minimum genetic variance, which makes the animals sensible to even small changes in the environment that may influence the outcome of the experiments. Examples of factors that may influence on the experimental results are physical parameters, such as temperature, humidity, light cycle, and chemical factors (*see* **Notes 1–4**), and also social factors, such as human contact, cage design, and animal interactions. Control of the most important of these factors may reduce “the background noise” in animal experiments and make it possible to obtain statistically significant results with a limited number of animals.

We have used a mouse model of postinfarction congestive heart failure combined with various in vitro techniques to specifically study the contribution of cytokines and thereby try to delineate the mechanism whereby heart failure may hamper normal hematopoiesis ***(11,12)***. Because we are investigating various aspects of inflammation on heart failure and hematopoiesis, the careful selection and handling of the experimental animals are particularly important to avoidance of undesirable confounding effects. This approach relies on first inducing heart failure by ligation of the left coronary artery to produce an acute myocardial infarction. Heart failure also may be induced by aortic banding, atrial pacing or injection of the right ventricular depressant monocrotaline. After development of overt heart failure, which usually lasts approx 4–6 wk, the animals are sacrificed, and blood and bone marrow samples are collected. Then, in vitro methods are used to pinpoint the contribution of various molecules, e.g. cytokines, to growth and viability of hematopoietic progenitor cells. Importantly, the development of overt heart failure also can be manipulated in vivo by either direct injection of putative regulatory molecules or constant delivery via implanted osmotic minipumps. Here, we present important aspects of this strategy with particular emphasis on adequate handling of the animals and ways of identifying putative inhibitors of hematopoiesis in heart failure mice.

2. Materials

2.1. Experimental Animals

For our studies, we have used adult 6- to-8-wk old Balb/c mice (*see* **Note 5**). Consent from appropriate ethics committees before commencing the actual experiments is mandatory because these experiments involve induction of a serious and lethal disease.

2.1.1. Health Monitoring

Because most infections in laboratory animals are subclinical, the animals and biological material to be used must be tested before the experiments. It is therefore vital to have a sentinel animal program and test regularly for parasitical, bacterial, and viral agents according to the needs of the experiment in question or by following the recommendations of the Federation of European Laboratory Animal Associations (FELASA) ***(13)***.

2.1.2. Testing of Biological Material

Biological material, such as tumor cell lines, may be tested in antibody production tests, where the biological material is given via different ports of entry (per os, intraperitoneal, intravenous, subcutaneous, and intranasal) to clean animals placed in an isolator. The animals are kept in the isolator for a

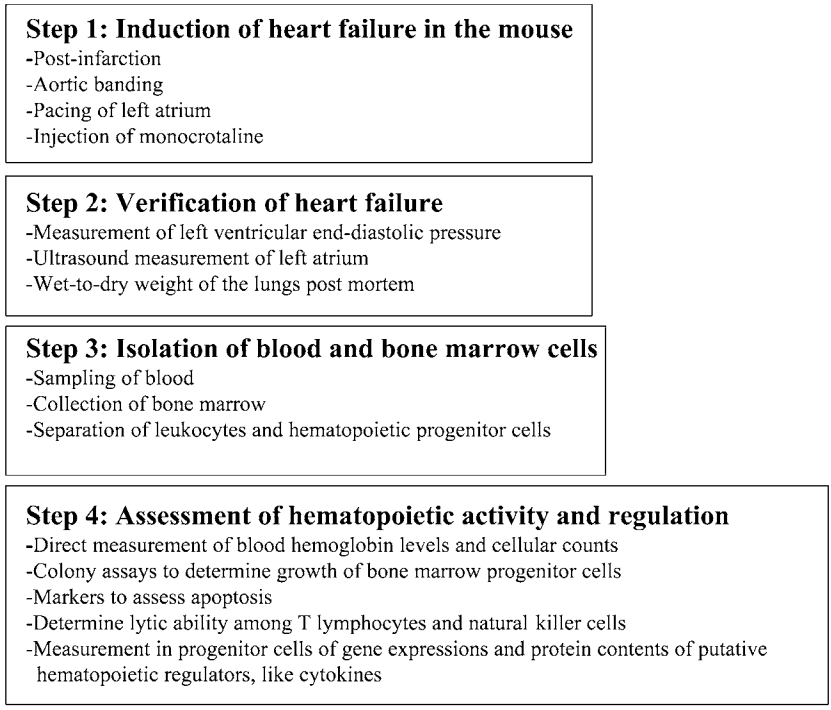


Fig. 1. Proposed sequence for studying hematopoiesis in mice with congestive heart failure.

minimum of 6 wk before they are euthanized, and serology is used to test for antibodies against the agents in question. The other option is to use PCR to ensure that biological material does not harbor viral contaminants that may influence the outcome of the experiments. It is currently under discussion whether PCR can replace antibody production ***(14)***.

2.2. Anesthetics and Postoperative Analgesics

Because the induction of myocardial infarction requires vascular surgery on a very small animal (body weight ~25 g), a stable general plane of anesthesia is important. After insertion of an endotracheal tube, we recommend a mixture of 2% isoflurane and 98% oxygen while keeping the animal on a ventilator. During the postoperative phase, repeated (every 8–12 h) subcutaneous injections with buprenorphine (0.3 mg/mL) is used as postoperative analgesia.

3. Methods

The experimental protocol is outlined in **Fig. 1**.

3.1. Surgery and Induction of Acute Myocardial Infarction

3.1.1. Ligation of Left Coronary Artery

In general anesthesia, a thoracotomy is made, and the left coronary artery is carefully dissected free of the surrounding tissue. Then, a ligature is placed around this artery and tightened (*see* **Note 6**). In the sham control animals, this artery is only exposed before the thoracic cavity is closed.

3.2. Secondary Operation and Verification of Heart Failure

By experience, a chronic congestive heart failure develops within the next 4 to 6 wk. In general anesthesia, a catheter is introduced into the left ventricular cavity to measure the end-diastolic pressure (LVEDP) (*see* **Note 7**). In this model, an elevated LVEDP is a hallmark of heart failure ***(15)***. An ultrasound examination of an enlarged left atrium also substantiates a failing heart. Postmortem, the wet-to-dry weight of the lungs is also assessed because this parameter is indicative of congestion (*see* **Note 8**).

3.3. Sampling of Blood and Bone Marrow Tissue

A blood sample is obtained by heart puncture in deep anesthesia immediately before the mouse is killed. Then, the femoral bone is isolated, and small holes are made in both ends. The femoral marrow is then flushed out of the bone by using a syringe containing either saline or air.

3.4. In Vitro Assays of Hematopoeitic Growth and Leukocyte Activity

Hematopoietic cells are cultured on semisolid media (methyl cellulose) for approx 2 wk before colonies are scored, and the numbers obtained are taken as the growth potential of these cells ***(11)***.

Neutrophilic activity of cells is measured as their oxygen consumption by using an O_2-sensitive electrode, whereas lytic activity among lymphocytes (T- and natural killer [NK]-cells) is measured with a conventional ^{51}Cr liberation assay ***(12)***.

3.5. Identifying Inhibitors of Hematopoiesis

The gene expression of possible inhibitors (e.g., TNFα) is done by an RNase protection assay on candidate cells (T- and NK-cells). The corresponding protein contents are measured with Western blotting ***(11)***.

4. Notes

1. The homeotherm animal is able to maintain a constant body temperature, within a range specific to each species, as a response to changes in the surrounding temperature. When subjected to surrounding temperatures outside this range, the body tries to remain in the thermoneutral zone by regulation of the flow of heath between

the core of the body and the body surface. This regulation may affect the metabolic rate of the animal and consequently result in different experimental results compared with an animal in its thermoneutral zone. It is important that the environmental temperature is precisely defined and kept within the defined range. The temperature must be recorded and published. Environmental parameters are listed in Appendix A in the Convention of the Council of Europe. Most rodents may be kept at room temperatures between 20 and 24°C. It is important that the temperature is kept constant, and it must be recorded and published.

2. Relative humidity should be kept in the range of 55±5%. Higher levels of relative humidity may lead to disturbances in the metabolic rate, and excessively high levels of relative humidity predispose for pathological conditions.
3. The light cycle has an important effect on the animal's physiology and behavior. Although most animals have their periods of highest activity during daylight hours, the majority of laboratory animals are nocturnal and sleep during day. In general 12 h of light and 12 h of dark is recommended for most laboratory animal species. The light intensity must be adjusted to the need of the species and should be constant in all parts of the room.
4. Exposure to chemicals may be a part of the procedure in an experiment or it may be unintentional in the laboratory animal facility. The most important sources of chemical contamination are animal excretions, bedding, cage or enclosure materials, feed, and household chemicals.
5. We prefer to study male mice because these are generally larger than the females. Size is a critical aspect because the induction of myocardial infarction requires surgery on the coronary vessels of the small murine heart.
6. The ligature is tightened so the myocardial tissue is blanched on visual inspection.
7. The use of microtipped transducer catheters is preferable to avoid pressure artifacts.
8. Any weight should be adjusted for possible edema, e.g., by expressing the weight relative to a bone length.

References

1. McKee, P. A., Castelli, W. P., McNamara, P. M., and Kannel, W. B. (1971) The natural history of congestive heart failure: the Framingham study. *N. Engl. J. Med.* **285,** 1441–1446.
2. Khand, A., Gemmel, I., Clark, A. L., and Cleland, J. G. (2000) Is the prognosis of heart failure improving? *J. Am. Coll. Cardiol.* **36,** 2284–2286.
3. Kloner, R. A. and Rezkalla, S. H. (2004) Cardiac protection during acute myocardial infarction: where do we stand in 2004. *J. Am. Coll. Cardiol.* **44,** 276–286.
4. Torre-Amione, G., Kapadia, S., Lee, J., et al. (1996) Tumor necrosis factor-alpha and tumor necrosis factor receptors in the failing human heart. *Circulation* **93,** 704–711.
5. Iversen, P. O., Nicolaysen, G., and Sioud, M. (2001) DNA enzyme targeting TNF-alpha mRNA improves hemodynamic performance in rats with postinfarction heart failure. *Am. J. Physiol*. **281,** H2211–H2217.

6. Woldbaek, P. R., Tonnessen, T., Henriksen, U. L., et al. (2003) Increased cardiac IL-18 mRNA, pro-IL-18 and plasma IL-18 after myocardial infarction in the mouse: a potential role in cardiac dysfunction. *Cardiovasc. Res.* **59,** 122–131.
7. Bohm, J. (2000) Gelatinous transformation of the bone marrow: the spectrum of underlying diseases. *Am. J. Surg. Pathol.* **24,** 56–65.
8. Ezekowitz, J. A., McAlister, F. A., and Armstrong, P. W. (2003) Anemia is common in heart failure and is associated with poor outcomes: insights from a cohort of 12 065 patients with new-onset heart failure. *Circulation* **107,** 223–225.
9. Mozaffarian, D., Nye, R., and Levy, W. C. (2003) Anemia predicts mortality in severe heart failure: the prospective randomized amlodipine survival evaluation (PRAISE). *J. Am. Coll. Cardiol.* **41,** 1933–1939.
10. Kalra, P. R., Collier, T., Cowie, M. R., et al. (2003) Haemoglobin concentration and prognosis in new cases of heart failure. *Lancet* **362,** 211–212.
11. Iversen, P. O., Woldbaek, P. R., Tonnessen, T., and Christensen, G. (2002) Decreased hematopoiesis in bone marrow of mice with congestive heart failure. *Am. J. Physiol.* **282,** R166–R172.
12. Iversen, P. O., Woldbaek, P. R., and Christensen, G. (2005) Reduced immune responses to an aseptic inflammation in mice with congestive heart failure. *Eur. J. Haematol.* **75,** 156–163.
13. Rehbinder, C., Baneux, P., Forbes, D., et al. (1996) FELASA recommendations for the health monitoring of mouse, rat, hamster, gerbil, guinea pig and rabbit experimental units. *Lab. Anim.* **30,** 193–208.
14. Mahabir, E., Jacobsen, K., Brielmeier, M., Peters, D., Needham, J., and Schmidt, J. (2004) Mouse antibody production test: can we do without it? *J. Virol. Methods* **120,** 239–245.
15. Finsen, A. V., Christensen, G., and Sjaastad, I. (2005) Echocardiographic parameters discriminating myocardial infarction with pulmonary congestion from myocardial infarction without congestion in the mouse. *J. Appl. Physiol.* **98,** 680–689.

14

An Overview of the Immune System and Technical Advances in Tumor Antigen Discovery and Validation

Mouldy Sioud

Summary

The ability of the immune system to distinguish between self- and nonself antigens is controlled by mechanisms of central and peripheral tolerance. Although the induction and maintenance of tolerance is important for preventing autoimmunity, breaking self-tolerance is a crucial constituent for combating cancers. Cancer patients are able to develop spontaneous immune responses to tumors that they bear, however these responses are not suboptimal for eradicating tumors. Moreover, none of the current immune strategies is able to activate the immune system to respond against tumor cells as it responds against infectious agents. These observations have raised the question of how to activate immunity in cancer patients to a threshold required for tumor rejection. Because tolerance is emerging as a central obstacle for immune recognition of human tumor antigens, this chapter describes how T- and B-cells are generated and activated in the periphery. It also outlines the technical advances in tumor antigen discovery and validation.

Key Words: B-cell development; B-cell receptor; dendritic cells; genomics; germinal center; immune tolerance; immunoediting; phage display; proteomics; serological analysis of recombinant tumor cDNA expression libraries (SEREX); T-cell development; T-cell receptor; tumor antigens.

1. Introduction

The primary function of the immune system is to defend the body against pathogenic microorganisms. The self–nonself (SNS) discrimination model assumes that the "foreign-ness" of a particular entity is what triggers the activation of innate and adaptive responses ***(1)***. Given the vast number of genetic and epigenetic changes associated with carcinogenesis, it becomes evident that tumors express many neoantigens that should be recognized by the immune system. In this respect, antibodies against a large repertoire of tumor-associated

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols*
Volume I, Emerging Strategies for Targets and Biomarker Discovery
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

gene products have been described in cancer patients ***(2,3)***. Moreover, we and others have shown that a high number of T-cells are present within tumors and that these cells are specifically recruited to tumor beds ***(4,5)***. In addition, T-cells reactive with tumor antigens are present in the peripheral blood and lymph nodes of diverse patients. Collectively, these results indicate that the immune system can recognize cancer cells. However, despite their potential immunogenicity, tumor antigens do not induce significant immune responses, which could destroy malignant cells; therefore, a central question is whether recognition of tumor antigens by immune cells would mainly lead to tolerance.

The different mechanisms for recognition and elimination of pathogens have been divided in the immune system into "innate immunity," which encompasses the presumably more primitive elements of the system, including macrophages, natural killer (NK), and antigen-presenting cells, and "adaptive immunity," which encompasses T- and B-lymphocytes. The activation of innate immunity is mainly mediated by the recognition of potential pathogens by Toll-like receptors (TLRs) that are mainly expressed on immune cells such as macrophages, dendritic cells (DCs), and monocytes ***(6,7)***. These receptors function as sensors of infection and induce the activation of innate and adaptive immune responses. Upon recognizing conserved structures, which are referred to as pathogen-associated molecular patterns (PAMPs), TLRs activate host defense responses through their intracellular signaling domain, the Toll/interleukin-1 receptor (TIR) domain, and the downstream adaptor protein MyD88. Although members of the TLRs all signal through MyD88, the signaling pathways induced by individual receptors differ ***(7)***. So far, 10 TLRs have been described in humans, and ligands have been defined for nine of these TLRs. TLR1 and -2 are activated by triacylated lipoproteins, TLR3 by double-stranded viral RNA, TLR4 by lipopolysaccharide (LPS), TLR5 by bacterial flagellin, TLR2 and -6 by diacylated lipopeptides, TLR7 and -8 by imidazoquinolines and single-stranded RNA, and TRL9 by unmethylated bacterial CpG DNA sequences ***(7)***. Upon activation, TLRs trigger the release of cytokines and the induction of co-stimulatory molecules that can influence the nature of the adaptive T- and or B-cell responses. Activation of TLRs also induces antimicrobial pathways that kill intracellular organisms ***(6–8)***. Although the innate response is rapid, the response sometimes damages normal tissues, leading to severe diseases, such a sepsis.

Adaptive immunity is the hallmark of the immune system in higher animals. It uses antigen-specific receptors expressed on T- and B-cells. The receptors on B-lymphocytes are Igs, which can be secreted by corresponding plasma cells after antigenic stimulation. Bound and secreted Igs can bind to a variety of soluble as well as cell-bound antigens, such as virus proteins. In contrast to B-cells, T-cells recognize only processed antigens that are presented as peptides by major histocompatibility complex (MHC) class I or II. Despite this

difference in antigen recognition, both B- and T-cell receptors have selected peptides as their major recognition targets. Cells of innate immunity, such as macrophages and DCs, produce considerable amounts of cytokines, which play an essential role in influencing the nature of the adaptive immune responses.

To explain immune tolerance, Polly Matzinger ***(9)*** presented the danger model, which suggests that a specific immune response develops as a result of danger detection rather than by discrimination between self- and nonself antigens. This model assumes that what really matters is whether the "antigen" causes damage. It does not matter if the damage is done by a pathogen or by self-proteins. Although this model refers mainly to immune tolerance, its assumptions are valid for tumors. Recently, Charlie Janeway ***(10)*** has suggested that antigen-presenting cells (APCs) and other cells of the innate immune system express a series of pattern recognition receptors (PRRs) specific for common motifs expressed by several pathogens. These motifs have been termed pathogen-associated molecular pattern (PAMP). PAMP recognition by PRRs, in particular TLRs, enables the innate immune system to distinguish self from nonself and is important not only for triggering innate immunity against microbial infection but also for priming the adaptive immune response. Triggering of PRRs activates the APCs and without this activation, no immune responses occur. This expanded SNS model assumes that the primary role of the immune system is to distinguish self-tissues from infectious nonself tissues, and this ability is necessary even for relatively lower organisms that have innate immune systems, but not adaptive systems capable of generating antigen-specific T- and B-lymphocytes. However, this model is not complete in that it does not account for responses against transplants or tumors. Because immune tolerance plays a crucial role in tumor immunology, this chapter describes how tolerant T- and B-cells to self-antigens are generated and activated in the periphery, and it outlines the technical advances in tumor antigen discovery and validation.

2. T-Cell Development and Maturation

The main goal of lymphocyte development is to generate a large repertoire of cells expressing diverse receptors that enable them to recognize and react to millions of foreign antigens ***(8)***. The key challenges of this specific immunity are to have as broad a repertoire of receptors as possible in the absence of potentially harmful autoreactive lymphocytes. In most adult vertebrates, the bone marrow contains hematopoietic stem cells (HSCs) that can differentiate into progenitors with more restricted lineage and that are capable for generating all blood cell types via commitment events **(Fig. 1)**.

Despite their different roles in the immune response, T-helper (Th) CD4+ cells and CD8+ cytotoxic T (Tc) cells originate from a common precursor ***(11)***.

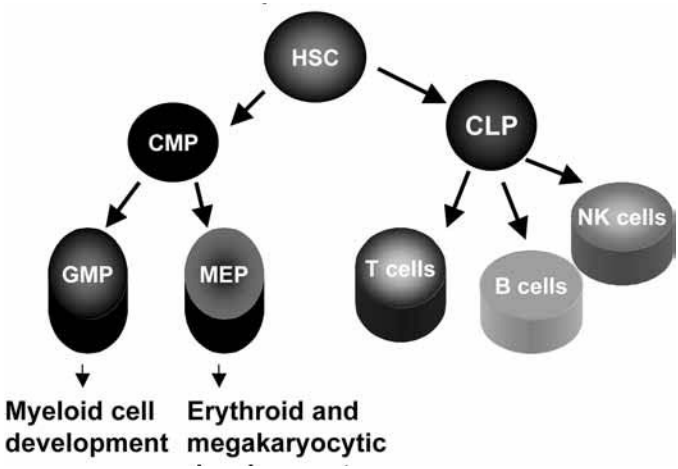


Fig. 1. Schematic of hematopoietic cell generation. Hematopoietic stem cells (HSCs) have self-renewal ability and differentiate through several intermediates into all the lineage of the hematopoietic system. In the adult, the lymphoid lineage differentiates via the common lymphoid progenitor (CLP), resulting in the generation of B-, T- and natural killer (NK) cells.

Activation of human CD34+ hematopoietic stem cells with R848, a specific ligand for TLR7 and TLR8, induced their differentiation into the myeloid cell lineage, including macrophages and dendritic cells (Johansen, L. F. and Sioud, M., unpublished data). This finding would indicate that microbial ligands for TLRs might shape the homeostasis of innate immune cells. As shown in **Fig. 1**, the adult lymphoid lineage, which includes NK, B-, and T-cells, is generated by the differentiation of HSCs via the common lymphoid progenitor (CLP), which is the earliest lineage-restricted lymphoid progenitor that has the potential to develop into these cells. Developing lymphocyte must go through several major developmental checkpoints before becoming a functional mature T-cell tolerant to self-antigens and having a huge potential to react with unlimited number of foreign antigens ***(11,12)***. The first checkpoint occurs when the CLP commits to the T-cell lineage. Recent studies indicated a crucial role for Notch-1 signaling in this process **(Fig. 2)**. Notch proteins were first identified in *Drosophila* and *Caenorhabditis elegans* by their ability to direct cell fate decisions as well as control survival and proliferation. These proteins are widely expressed in different tissues, including the hematopoietic system. In the absence of Notch-1, bone marrow progenitors enter the thymus and develop into B-cells. Thus, Notch-1 “instructs” an early lymphoid progenitor to adopt a T- versus B-cell fate ***(13,14)***. Notably, transcriptional activation by Notch requires its binding to CSL/RBP-J, leading to the formation of an active transcriptional activation complex. In this respect CSL-deficient bone marrow

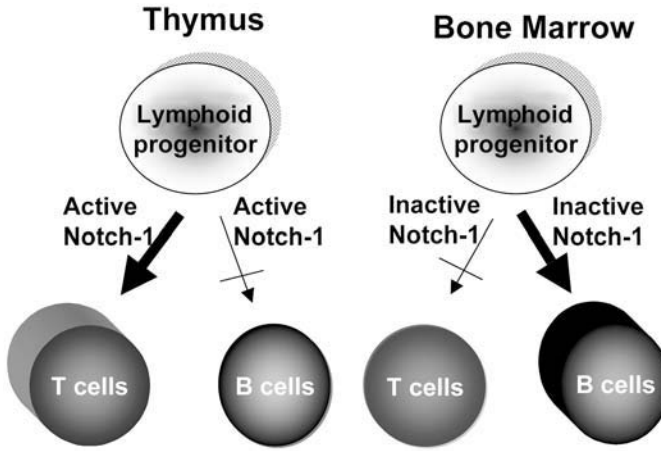


Fig. 2. Notch signaling is required for T-cell development. Notch-1 signaling promotes T-cell development more likely by activating T-cell transcriptional program. In absence of Notch-1 signaling, B-cell development is activated.

progenitors fail to develop into T-cells in the thymus and consequently develop ectopically into B-cells ***(13)***. Notch-1 signaling also may affect pre-T-cell receptor (TCR) expression and functions in immature thymocytes. Recent data indicate that Notch-1 expression is under the control of microRNA-181 (*see* Chapter 16, Volume 2).

In contrast to most genes, the genes encoding T- and B-cell antigen receptors are unusual in that they are composed of a series of V (variable), D (diversity), and J (joining) gene segments, which require assembly by a site-specific DNA recombination reaction known as V (D) J recombination. The ability of V, D, and J gene segments to combine randomly introduces a large element of combinatorial diversity into the Ig and TCR repertoires. Also, during the rearrangement process, additional nucleotides not encoded by either gene segments can be added at the junction between the joined gene segments. This joining process is random, and, on average, two out of three rearrangements will be nonproductive because the joining has occurred in different translational reading frames. In developing T-cells, rearrangement of TCRβ gene occurs before TCRα, whereas in developing B-cells, assembly of the Ig heavy chain (IgH) gene occurs before that of the Ig light chain (IgL) genes.

2.1. Pre-TCR Functions as a Biological Sensor

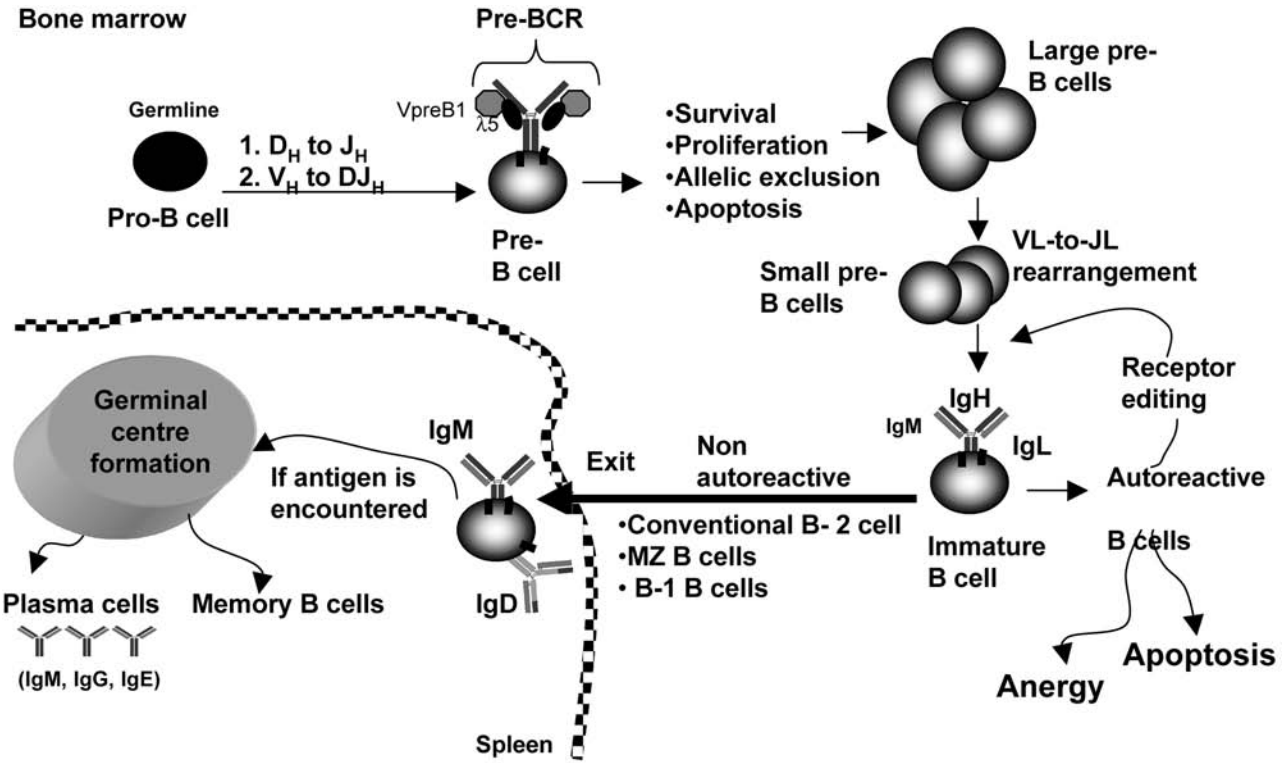
Subsequent to entry into the thymus, T-cell precursors go through a series of well defined marker changes and selective events **(Fig. 3)**. CLP and the immature T-cell precursor do not express CD4 and CD8 molecules and are referred to as CD4–CD8– (double negative [DN]) thymocytes ***(11)***. These thymocytes

Notch 1+
Germline
Pro-T cell
CD25-
C-Kit+
CD44
C-Kit+
CD25+
CD25+
CD3
TcR γδ
rearrangement
TcR β
Rearrangement
γδ T cells
Periphery
Pre-TCR
signaling
CD3
•Survival
• β allellic exclusion
•Apoptosis
DN thymocytes
DP thymocytes
TcRα
rearrangement
•Survival
•Apoptosis
CD3
TCR
TcRα/β
Positive
selection
SP
thymocytes
CD4+
CD3
TcRα/β
CD8+
CD3
TcRα/β
Maturation (e.g.
up regulation of
co-stimulatory
Molecules)
Periphery
CD3
CD25+
T reg
cells
Apoptosis
(negative selection)
CD3
NK1.1
NKT
cells

then express the recombinases (RAG-1 and RAG-2) and enter a major checkpoint in T-cell development, where the DN thymocyte commits to either the α/β or γ/δ T-cell lineage ***(11)***. Developmental control checkpoints seem to ensure that T-cells do not complete their intrathymic differentiation program in the absence of productive TCR gene rearrangements. As a result, T-cells do not mature with TCRαβ having inappropriate specificities. If the γ and δ genes are rearranged productively, the cells express the TCR γδ receptor at the surface and subsequently follow the γδ lineage ***(15)***. In these cells, the CD4 and CD8 expression remains off **(Fig. 3)**. Although successful rearrangement of either β or γδ genes may influence the lineage decision, the mechanism by which thymocytes choose between the αβ and γδ lineage remain unknown ***(16)***. However, the successful rearrangement of a β-chain gene must occur before successful rearrangement of both γ and δ. Notably, most thymocytes develop into the αβ lineage. In this case, the checkpoint transition is induced by the rearrangement of the TCR.

Productive rearrangement of TCRβ gene leads to the formation of the pre-TCR signaling complex, which contains a TCRβ chain, a nonpolymorphic surrogate α-chain (pTα), and the multisubunit CD3 signaling complex. Assembly of the pre-TCR of the cell surface of DN thymocytes promotes their survival, informs them to progress to the CD4+CD8+ (double-positive [DP]) stage of development and causes them to undergo significant clonal expansion through an Lck/Ras-dependent signaling pathway ***(17,18)***. The assembly of the pre-TCR establishes a key checkpoint in early thymocyte differentiation at which developing T-cells undergo "β-selection." Pre-TCR signaling triggers the phosphorylation and degradation of RAG-2, halting β-chain gene rearrangement, thereby enforcing strict allelic exclusion at the β-locus. This process is mainly controlled by the protein kinase C (PKC)α signaling pathway ***(16–18)***. We recently have found that PKCα also is involved in the differentiation of HSC into the myeloid lineage ***(19)***. Pre-TCR also induces rapid cell proliferation and eventually the

Fig. 3. *(Opposite page)* T-cell development. Various stages of T-cell development in the thymus are shown. Productive V to DJβ rearrangement in double negative (DN) pro-T-cells results in the assembly of the pre-TCR. These pre-T-cells become CD4/CD8 double positive (DP) thymocytes. After TCRβ locus allelic exclusion and activation of V to Jα rearrangement, DP thymocytes express mature αβ TCR. T-cell receptors that recognize self-peptide/major histocompatability complex (MHC) on thymic epithelial cells (TECs) with high affinity are deleted by apoptosis (negative selection), whereas T-cells reacting with self-peptide/MHC complexes with low affinity are positively selected, resulting in the generation of CD4 or CD8 single positive thymocytes. Also, the selection results in the progression into regulatory T-cells (Treg) cells and natural killer T-cells (NKT). Some specific markers are indicated. DN pro-T-cells that succeed to rearrange the T-cell receptor (TCR) γδ will lead to the generation of γδ T-cells.


Bone marrow
Pre-BCR
Germline
VpreB1
λ5
1. D_H to J_H
2. V_H to DJ_H
Pro-B cell
Pre-
B cell
•Survival
•Proliferation
•Allelic exclusion
•Apoptosis
Large pre-
B cells
VL-to-JL
rearrangement
Small pre-
B cells
Receptor
editing
IgH
IgM
IgL
Autoreactive
B cells
Immature
B cell
Non
autoreactive
Exit
•Conventional B- 2 cell
•MZ B cells
• B-1 B cells
Germinal
centre
formation
If antigen is
encountered
IgM
IgD
Plasma cells
(IgM, IgG, IgE)
Memory B cells
Spleen
Anergy
Apoptosis

expression of the coreceptor proteins CD4 and CD8 ***(18)***. During this proliferative phase, the expression of RAG-1 and RAG-2 genes is inhibited. Hence, no rearrangement of the α-chain locus occurs until the proliferative phase ends.

The TCRα chain genes are comparable with Igκ and λ light chain genes in that they do not have D gene segments and are rearranged only after their partner receptor chain gene has been expressed. Subsequent to the expression of productive Vβ gene rearrangement and few cycles of cell division, the developing thymocytes exit the cell cycle, reexpress RAG-1 and RAG-2, and eventually start rearranging the α-chain genes. Unlike TCRβ-chain, the expression of the TCRα-chain is not in itself sufficient to shut off gene rearrangement ***(20)***. As a result, up to one-third of mature T-cells harbor two productively rearranged TCRα alleles ***(21)***. Although the exact function of the second TCR is unclear, recent data indicated that the second TCRs are involved in extending TCR repertoires against microorganisms, and they also are involved in autoimmunity ***(22)***. Noteworthy, up to this stage, the development of the thymocytes has been independent of antigens.

2.2. TCR-Mediated Selection of Thymocytes: Central Tolerance

Productive Vα gene rearrangement allows the developing T-cell to assemble a complete TCR molecule, which is expressed on thymocyte surface in association with CD3 complex, forming active TCR. Most of the TCR-mediated selection processes begin at this stage of development that occurs primarily in the thymus. The selection process includes both positive and negative selection of immature CD4+ and CD8+ cells **(Fig. 3)**. Both processes are regulated through the interaction between the T-cell receptor expressed by a given DP thymocyte and peptide presented by thymic epithelial cells ***(8,11)***. Thymocytes with TCRs that bind strongly to MHC-associated self-peptide are eliminated (negative

---

Fig. 4. *(Opposite page)* B-cell development and selection. Various stages of B-cell development in the bone marrow and spleen are shown. In stem cells, Ig genes are in the germline configuration as found in all nonlympoid cells. In pro-B-cells D to JH rearrangement on both heavy chain alleles precedes V to DJH rearrangement. Cells that succeed to rearrange a functional heavy chain (pre-B-cells) assemble the pre-B-cell antigen receptor (BCR), which leads to heavy chain allelic exclusion and activation of Ig light chain (IgL) rearrangement (κ and λ). After successfully rearrangement of the IgL, immature B-cells express the BCR (IgM+). Immature B-cells that interact strongly with self-antigens are deleted or undergo another rearrangement at the IgL locus (receptor editing) to alter their specificity. Most of the autoreactive B-cells are deleted in the bone marrow. However, some of these autoreactive B-cells reacting with self-antigens can be rescued by undergoing receptor editing to generate new receptor specificity. Nonautoreactive B-cells, exit the bone marrow via sinusoidal endothelium as IgM-, IgD-positive cells. Notably, immature B-cells (IgM^{high} IgD^{low}) progressively acquire more IgD as they develop into mature B-cells.

selection), whereas thymocytes with TCR that bind with lower affinity to MHC-associated self-peptides receive prosurvival signals and are allowed to progress to the simple positive (SP) stage (positive selection). Thymocytes expressing TCRs that do not interact with the appropriate MHC die by neglect (immunological ignorance).

Recent studies indicated that the positively selected thymocytes are subjected to a maturation step, which increases the threshold for their activation by self-selected peptides ***(23)***. As a consequence, peripheral mature T-cells are properly activated by only foreign peptides of higher affinity (dominant epitopes). In conclusion, T-cell development in the thymus starts with CD4–CD8– DN thymocytes, which progress to CD4+CD8+ DP thymocytes, and finally to CD4+ or CD8+ SP thymocytes. Subsequent to maturation, these positively selected T-cells leave the thymus and eventually enter the circulation **(Fig. 3)**.

3. B-Cell Development and Maturation

3.1. Pre-B-Cell Antigen Receptor (BCR) Functions as a Biological Sensor

Similar to T-cells, early B-cell development is ordered by successive rearrangements of IgH and IgL gene segments during pro-B- and pre-B-cell development **(Fig. 4)**. This crucial developmental process leads to the generation of a diverse antibody repertoire for humoral defense against all possible foreign antigens ***(24,25)***. The first step is antigen independent and takes place in primary lymphoid organs, the bone marrow, and the fetal liver. In developing B-cells (pro-B-cells), assembly of the IgH gene occurs before that of the IgL genes. The pro-B-cells express surrogate light chain encoded by VpreB and λ5 genes and the rearrangement machinery encoded by the RAG-1, RAG-2, and TdT genes. In human, only one VpreB gene has been found, whereas several λ5-like genes have been identified. They are all located on chromosome 22, which also carries the λ light chain genes.

Within the IgH locus, rearrangement begins on both alleles with joining of a D_H to a J_H gene segment. Notably, most cells successfully rearrange two D-J sequences. These pro-B-cells then go on to recombine a V_H segment with one of the partially assembled DJ_H alleles, leading to the generation of pre-B-cell population. After productive VDJ_H rearrangement, the IgH allele is transcribed and its message is translated, resulting in the production of the transmembrane form of the heavy chain protein Igμ. Because there are two alleles for each immunoglobulin locus in the diploid genome, each of which can be activated, each cell must prevent both alleles from making productive chains. This is accomplished by checking for productive chain as soon as an allele has rearranged. Thus, μ heavy chain seems to be functionally equivalent to TCRβ-chain in the early states of lymphocyte development.

Once a functional heavy chain is made, it associates with other proteins into a signaling complex known as the pre-BCR **(Fig. 4)**. In addition to a homodimer of Igμ, the pre-BCR contains the products of the surrogate light chain (SLC) genes λ5 and *VpreB* and a heterodimeric signaling module made up of the transmembrane proteins Igα and Igβ (CD79α and CD79β) ***(17,24)***. A developing pro-B-cell that makes a functional Ig heavy chain is rescued from apoptosis and enters the pre-B-cell compartment, whereas nonproductive VDJH rearrangement on both alleles leads to programmed cell death ***(25)***. Cell surface expression of a functional pre-BCR induces two to five cell divisions, leading to a large pre-B-cell population that subsequently becomes small resting pre-B-cells ***(24,25)***. Notably, this process generates B-cell clones expressing a common μ heavy chain but with the potential to express a diverse array of light chains, thereby extending B-cell repertoire. Similarly to T-cells, the pre-BCR functions as a biological sensor that informs pro-B-cells that they have succeeded in rearranging a functional Ig heavy chain ***(17)***.

Once a functional/productive μ heavy chain is formed, the cell proceeds to the pre-B-stage, expresses the recombinase, and renders the light chain loci (λ and κ) accessible to the recombination machinery. The κ-chain locus is generally rearranged before the λ-chain locus; the latter chain only initiates rearrangement if the κ-locus rearrangement has failed to generate a productive light chain. In contrast to T-cells, the mechanism of allelic exclusion also occurs at the light chain loci.

As soon as a functional light chain has paired with the preexisting μ-heavy chain, a complete IgM molecule is displayed on the cell surface in association with Igα and Igβ, forming the BCR of immature B-cell **(Fig. 4)**. If the newly expressed receptor encounters a strong interaction with self-antigens, B-cell development is inhibited, and the cell is deleted by apoptosis. This process is the first negative selection process that B-cells undergo in the bone marrow. Negative selection at this stage would therefore promote elimination of unwanted autoreactive specificities while preserving immature B-cells reactive with foreign antigens. However, a large proportion of autoreactive B-cells escape clonal deletion by altering their BCRs specificity, so that they are no longer autoreactive, a process called “receptor editing” ***(26,27)*** **(Fig. 4)**. Thus, immature B-cells maintain the rearrangement machinery upregulated to allow for secondary rearrangements at the IgL chain gene loci. In addition to clonal deletion, autoreactive B-cells are inactivated via a second mechanism called anergy ***(25)***. Anergic B-cells express a unique surface phenotype in that they express much reduced sIgM (because of receptor modulation after antigenic exposure), normal IgD levels, and are desensitized to further receptor-induced signal transduction. The “choice” of an immature B-cells to undergo either clonal deletion or clonal anergy is mainly determined by the concentration

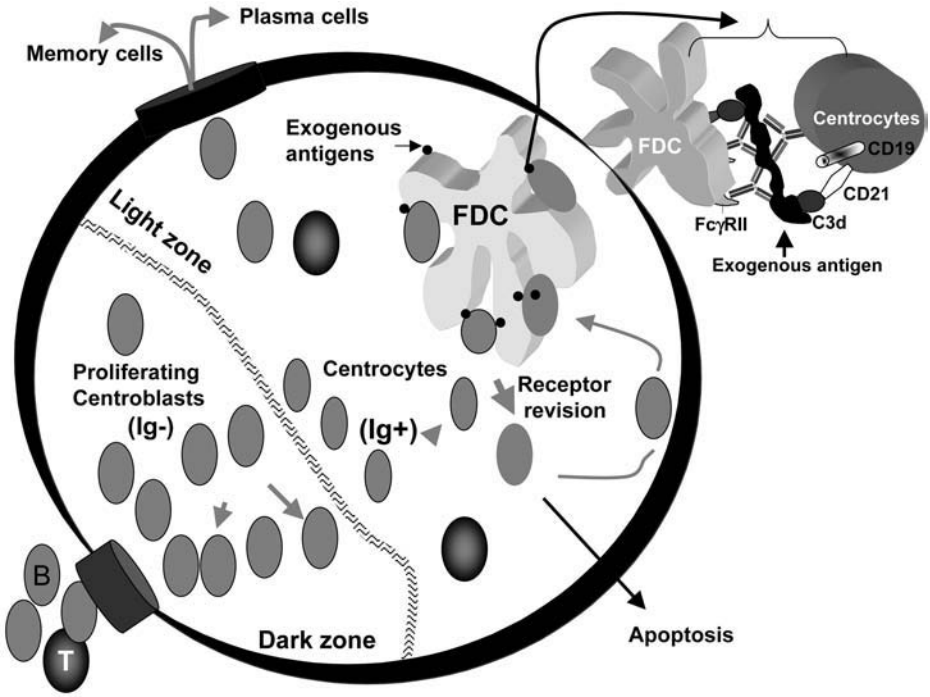


Fig. 5. Affinity maturation of B-cells within the germinal center. Antigen recognized by the surface IgM of the B-cells is internalized, processed, and presented on major histocompatability complex (MHC) class II molecule of the B-cells. The antigen can then be presented to a primed specific T-cell. The T-cells in turn produce cytokines, resulting in B-cell division and maturation to antibody-secreting cells. This activation occurs mainly within the germinal centers of lymph nodes, where B-cells proliferate in the dark zone as centroblasts. These cells do not express surface Ig (Ig–). During cell division they can accumulate V (D)J hypermutation. Subsequently, some cells express membrane antibodies with various affinities and migrate into the germinal center light zone as centrocytes with surface Ig (Ig+). Within the light zone, the centrocytes recover the exogenous antigens from follicular dendritic cells. Antibody–antigen complexes are bound to complement and Fc receptors on the follicular dendritic cell surface. Antigens that bind with high affinity to the antibody will lead to the positive selection of the centrocytes, resulting into the progression into plasma and memory B-cells. Antigens that bind weakly to the antibody would either activate RAG proteins and rearrangement at the Ig light chain (IgL) locus (receptor revision) or induce apoptosis in B-cells. Revised receptors may acquire high antigenic affinity, resulting in B-cell export to the periphery and further maturation.

and valency of the interacting self-antigens. In general, highly expressed membrane-bound antigens or antigens with multiple identical epitopes induce extensive BCR cross-linking, leading to immature B deletion. In contrast,

soluble, monovalent antigens mainly induce anergy. Anergic B-cells exhibit a shortened half-life in the periphery ***(24)***.

Once nonautoreactive B-cells are generated, they can leave the bone marrow for the spleen where they further mature into naive B-cells, which express both IgM and IgD molecules, by alternative splicing. However, it is not yet clear how, in the absence of binding to specific antigen in the bone marrow, an immature B-cell senses that a functional IgM has been made and subsequently receives signals for further maturation. There are three mature B-cell subsets: follicular (FO) or B-2 cells, marginal zone (MZ) B-cells, and B1 cells. FO B-cells contain the bulk of the peripheral B-cell pool and are the main cellular constituent of the spleenic and lymph node follicles and the bone marrow as recirculating cells. MZ B-cells constitute a small fraction (5%) of all peripheral B-cells and are located in the marginal zones of the spleen and are responsible for the early antibody response to blood-borne pathogens. B1 cells are located principally in the peritoneal and pleural cavities. They provide protection against antigen of the gut and peritoneum. Both MZ and B1 cells are often autoreactive, a property that enhance their survival and rapid responses to foreign antigens. Upon antigen encounter, FO naive B-cells differentiate into plasma cells or memory cells in secondary lympoid organs, in particular the spleen **(Fig. 4)**. It is worth noting that in the bone marrow, IgM+ immature B-cells that interact strongly with self-antigen either are eliminated by apoptosis (negative selection) or undergo further rearrangement at the IgL loci to alter their specificity, whereas stimulation via the IgM BCR induces proliferation of mature naive B-cells. Thus, both cell populations differ in their signaling downstream of the BCR. Indeed, the BCR is recruited to lipid rafts, an important structure in cell signaling, in mature B-cells but not in immature B-cells.

3.2. Somatic Hypermutation

B-cells, which express a self-tolerant BCR, exit the bone marrow to join the peripheral B-cell population. In contrast to TCRs, the BCRs mature antigenically. Upon encountering exogenous antigens, which they recognize specifically by means of the IgM BCR, mature naive B-lymphocytes are activated, proliferate, and form germinal centers. B-cells are first activated outside the follicles by antigen and T-cells **(Fig. 5)**. Then, they migrate into a primary lymphoid follicle where they continue to proliferate and subsequently form the germinal center (GC). Proliferating B-cells reduce the expression of surface immunoglobulin. These cells are termed centroblasts, which first proliferate in the dark zone of the GC **(Fig. 5)**. Subsequently, some cells reduce their rate of cell division and express higher levels of surface immunoglobulin. These cells are termed centrocytes, which are mostly found in the light zone of the GC. Within the light zone, centrocytes make

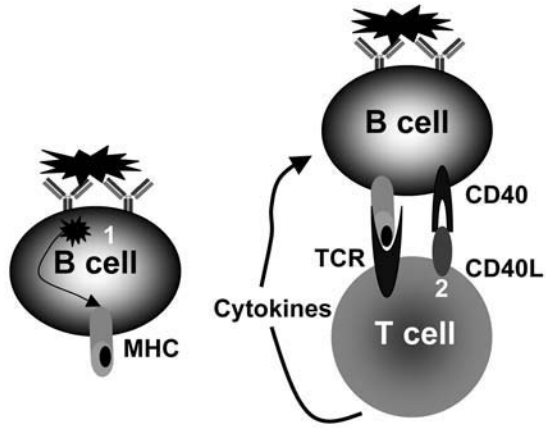


Fig. 6. A second signal is required for B-cell activation. The interaction of the B-cell antigen receptor (BCR) with the antigen constitutes the first signal. For thymus-dependent antigens, the second signal is delivered by the helper cells that recognize the antigen as peptides bound to major histocompatability complex (MHC) II molecules on B-cells. The interaction between CD40 ligand (CD40L) on the T-cell and CD40 on the B-cell is a part of the second signal. For thymus-independent antigens, the antigens or other cells of the immune system deliver the second signal.

contact with a dense network of follicular dendritic cells (FDCs). B-cells with mutations that improve affinity for exogenous antigen are positively selected. However, antigen receptors that fail to bind adequately to low levels of antigen-presented by FDCs can trigger receptor revision, apoptosis, or both. Notably, negative selection is an important force in the GC, most likely eliminating approximately one in every two cells.

The main purpose of the GC reaction is to enhance the later part of the primary immune response. Some GC cells differentiate first into plasmablasts and then into plasma cells. These nondividing, terminally differentiated plasma cells are specialized to secrete antibody at a high rate. Other germinal center cells differentiate into memory B-cells, which are long-lived descendants of cells that were once stimulated by antigen and had proliferated in the germinal center. These cells divide very slowly, if at all. They express surface Ig, but they do not secrete antibody at a high rate. Signals from both FDCs and T-cells are important in stimulating a B-cell to become a memory cell.

Notably, the production of highly efficient neutralizing antibodies requires antibody affinity maturation, which involves two main events ***(28)*** (1) immunoglobulin class-switch recombination (CSR), which directs antibody production towards the synthesis of IgG, IgA, and IgE; and (2) generation of somatic hypermutation (SHM) in Ig-variable region (V region), which can result in

amino acid replacements in immunoglobulin V regions that affect the fate of B-cells. Because not all IgGs or IgAs carry somatic hypermutations and some IgMs have such mutations ***(28)***, neither CSR nor SHM is a prerequisite of the other. However, both processes require close cooperation between T- and B-cells. The best known form of such cooperation is the interaction between CD40 ligand transiently expressed on activated Th cells and CD40 molecules constitutively expressed by B-cells **(Fig. 6)**. Defect in this interaction can lead to hyper-IgM syndromes ***(29)***. Because of somatic hypermutations, tolerance must be induced on B-cells both during their development and after antigenic stimulation in the secondary lymphoid tissues. Then, useful IgG molecules require a higher affinity binding to exogenous antigens, whereas useful TCRs must bind their MHC-self peptide with a lower affinity.

3.3. Receptor Revision

BCRs that fail to bind adequately to low levels of exogenous antigens presented by FDCs within the GC can trigger yet another mechanism for optimizing the B-cell affinity, receptor revision, which refers to secondary rearrangements in peripheral cells ***(30)***. This process is initiated by the absence of signaling through the BCR and results in secondary RAG-mediated rearrangement at the IgL locus, leading to the expression of a surface Ig molecule that binds antigen with affinity higher than that of the parent Ig molecule **(Fig. 5)**. Importantly, BCR signaling results in a variety of cellular events, including activation, cell cycling, survival, differentiation, apoptosis, and induction or suppression of Ig gene rearrangement, depending on the developmental state and interaction with their microenvironments.

Although antibody response to most protein antigens is dependent on Th cells, humans and mice with T-cell deficiencies nevertheless make antibodies to many bacterial antigens. Such antigens are known as thymus-independent antigens because they stimulate strong antibody responses in athymic individuals. These antigens are often called mitogens, a mitogen being a substance that induces cells to undergo mitosis (e.g., LPS). This immunity is an important component of the humoral immune response to nonprotein antigens that do not engage peptide-specific T-cell help.

4. Peripheral Tolerance

After development and selection in the thymus or the bone marrow, respectively, naive T- and B-cells migrate to the secondary lymphoid organs (e.g., lymph nodes, tonsils, and spleen). These lymphoid tissues, in addition to bringing T- and B-cells together, contain professional APCs and microenvironments (e.g., cytokines) that are crucial for T- and B-cell function and survival ***(8)***. APCs include DCs, B-cells, and macrophages.

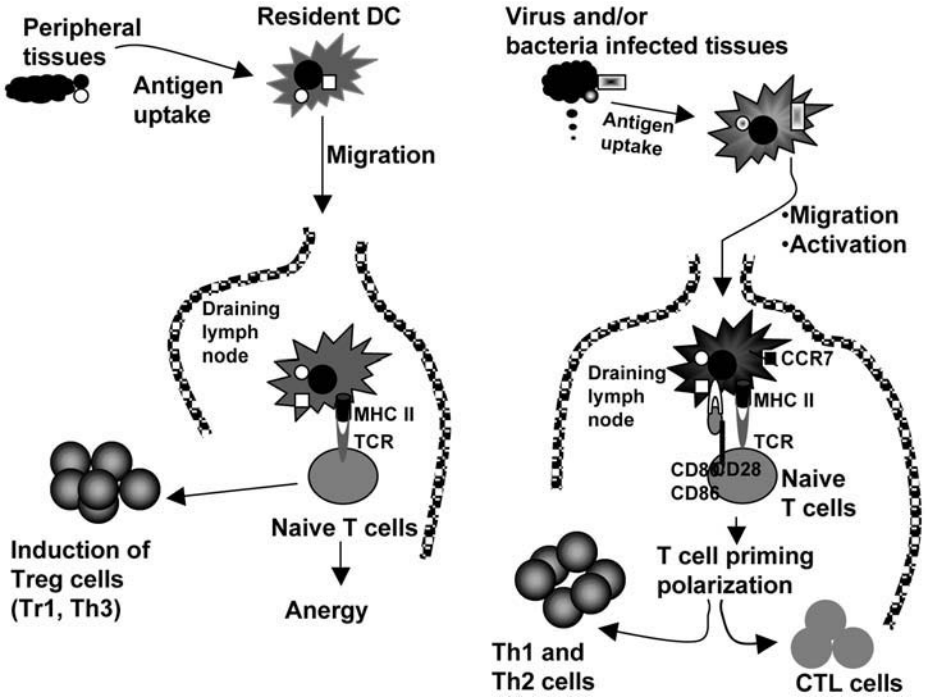


Fig. 7. Bridging between innate and acquired immunity. After antigen capturing, dendritic cells (DCs) migrate to draining lymph nodes. In the absence of danger signal (e.g., inflammation or exogenous antigens), the DCs remain in an immature state that present antigen to T-cells in the absence of co-stimulatory molecules (signal 2). This will lead to either the deletion of T-cells or the generation of inducible regulatory T-cells (Tr1, Th3). However, danger signals such tissue inflammations or viral or bacterial infections induce the maturation of DCs and the migration of a large number of DCs to draining lymph nodes. The expression of co-stimulatory molecules by DCs would lead to the activation of T-cells, activation of B-cells via T-cell–secreted cytokines, and the generation of effective adaptive immune response.

Despite the different central mechanisms operating on T- and B-positive selection, it is hard to believe that all autoreactive cells are deleted. Many self-antigens might not have access to the thymus, and others are only expressed later in life. Thus, not all self-reactive cells are deleted in central organs. Low-affinity self-reactive T-cells with specificity against antigens not represented in the primary lymphoid tissues mature and join the peripheral lymphocyte pool. So, there is a requirement for additional cellular and molecular mechanisms that prevent these autoreactive cells in initiating an autoimmune disease. Several mechanisms collectively referred to as “peripheral tolerance” normally operate to prevent autoreactivity. These mechanisms include the induction of anergy

(unresponsiveness); downregulation of cell surface expression of TCR, BCR, coreceptor molecules, or a combination; indifference; peripheral deletion; and immunosuppression ***(31–33)***. Noteworthy is that the adaptive immune system consists not only of immunostimulatory T-cells but also immunosuppressive T-cells. So far, three subsets of immunosuppressive CD4+ T-cells have been identified and the best-characterized population is defined by a constitutive expression of interleukin (IL)-2 receptor α-chain. Although the majority of these cells, known as regulatory T-cells (Treg cells) are generated in the thymus, they may be generated in peripheral lymphoid organs under certain conditions **(Fig. 7)**. The physical interaction of activated Treg cells with CD4+ T-cells is crucial in local and specific immunosuppression. However, the suppression can be antigen nonspecific and independent of inhibitory cytokines such as IL 10 and TGFβ. Recent results indicated that depletion of Treg cells could lead to autoimmunity in animal models. Additionally, dysfunction of these cells has been linked to autoimmunity in humans ***(33)***.

In addition of playing a major role in the activation of T-cells, DCs also can maintain both central and peripheral tolerance ***(34)***. Indeed, mature thymic DCs are essential for the deletion of newly generated T-cells that have receptors that recognize self-antigens with high affinity. By presenting tissue antigens to naive T-cells in the absence of co-stimulatory signals, immature DCs induce T-cell anergy/deletion or the development of inducible regulatory T-cells, which produce IL-10 (**Fig. 7**). Interestingly, some studies indicated that peripheral T-cell tolerance also could be maintained by receptor revision ***(35)***. Consistent with the important role of GC in this process, information so far available seems to indicate that TCR revision is restricted to CD4+ T-cell population. In addition to central tolerance, potential autoreactive B-cells also are eliminated in the periphery. However, if an autoreactive B-cell escapes clonal deletion or receptor editing, or it is generated in the GC after somatic mutations, it should be kept in check by the absence of the appropriate Th cell population, whose tolerance is maintained at several levels ***(31)***.

5. Recognition of Self-Antigens Is Crucial for Both T- and B-Cell Survival: How Strong Can the Interaction Be?

Recent studies indicated that naive peripheral T-cells bearing TCRs that fail to interact with MHC molecules plus self-peptides have a shortened life-span and do not undergo homeostatic proliferation to maintain the proper balance of lymphoid cell populations ***(36,37)***. To survive, peripheral naive T-cells require the recognition of self-peptide/MHC complexes. TCRs transduce survival signals upon binding to self-MHC molecules. These results underlie the important roles played by the selecting self-peptides. Notably, this self-recognition constitutes a permanent danger because under certain conditions (e.g., peptide concentration or

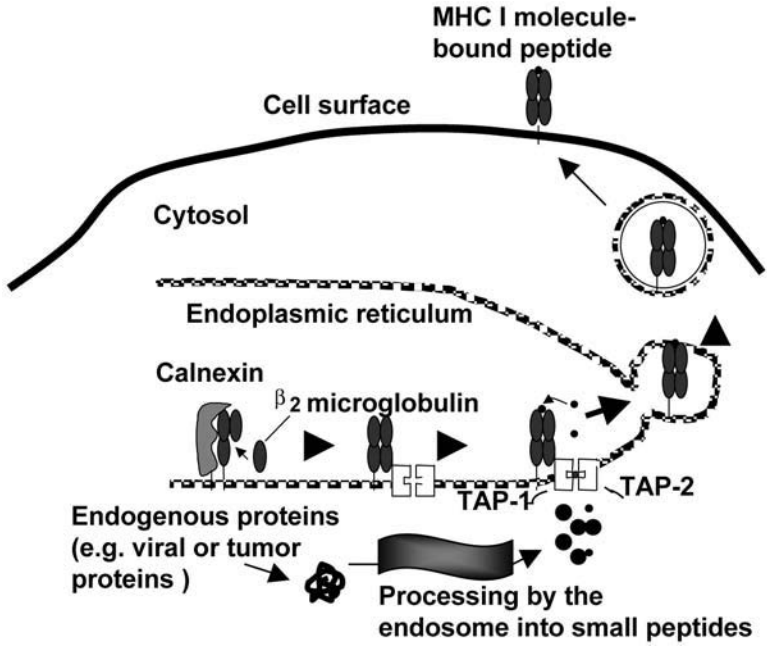


Fig. 8. Schematic of class I pathway of antigen processing. The proteasome digests endogenous cellular proteins into peptide fragments that are transported to the endoplasmic reticulum where they will be taken up by major histocompatability complex (MHC) I molecules. Subsequently, the complexes are delivered to the cell surface. TAP, transporter associated with antigen presentation.

posttranslational modifications), naive T-cell interaction with self-peptides may trigger their full activation, thereby inducing autoimmunity. This type of reactivity may benefit patients with cancers. Another question is whether B-cells, as T-cells, interact with self for survival. In this respect, all peripheral B-cells fail to be maintained when the BCR is deleted, indicating that BCR signaling also is required for B-cell survival ***(24)***. Again, the strength of BCR signaling seems to be a major player in peripheral B-cell fate. Therefore, the data predict that both T- and B-cells must interact with self-antigens to survive.

6. Antigen Processing

6.1. Class I Pathway of Antigen Processing

The TCR recognizes antigen only when it is displayed on the surface of the target cell as peptide fragment by the class I and class II molecules of the MHC. Class I molecules are present on virtually all nucleated cells and predominantly present peptides derived from endogenous proteins as illustrated in **Fig. 8**. These

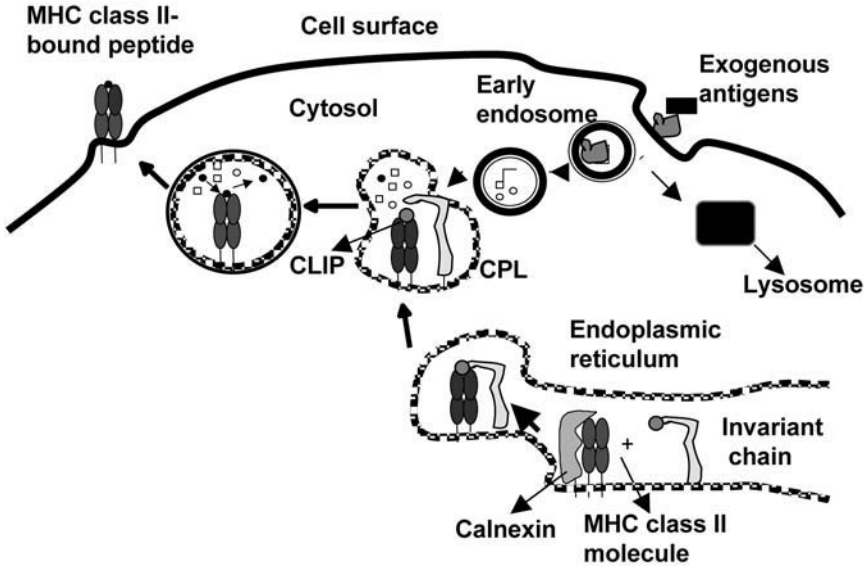


Fig. 9. Schematic of class II pathway of antigen processing. The invariant chain transports the major histocompatability complex (MHC) class II molecules from the endoplasmic reticulum to a specialized endosomal compartment where the invariant chain is cleaved leaving a small peptide CLIP. Exogenous antigens that are taken up by endocytosis are cleaved into small peptide fragments within the endosomes. Within the endosomes, the CLIP fragment is eventually replaced by the cleaved peptides and then complexes are delivered to the cell surface. CPL, compartment for peptide loading.

peptides are generated by the proteasomes, multicatalytic protease assemblies that are responsible for the major pathway of protein turnover in the cytoplasm ***(38)***. The peptides generated by the proteasome are subsequently translocated into the lumen of the endoplasmic reticulum (ER) by the transporter associated with antigen presentation (TAP) heterodimer, which may display a certain degree of specificity by selecting peptides with length and termini optimal for the class I binding ***(39)***. The binding between the peptide and class I molecule takes place inside the ER. Tumor cells therefore present peptides from tumor antigens in the context of class I MHC to CD8+ T-cells.

6.2. Class II Pathway of Antigen Processing

Class II molecules are present only on the surface of specialized APCs, including DCs, B-cells, and macrophages. These are predominantly loaded with peptides derived from exogenous proteins that enter the endocytic pathway of

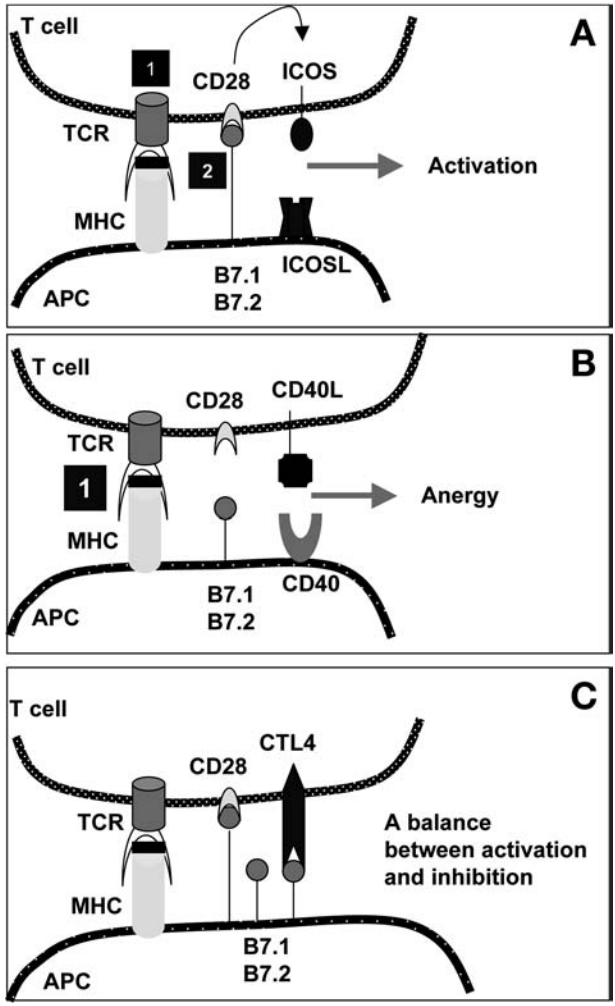


Fig. 10. Critical molecules involved in the interaction between the T-cells and the antigen-presenting cells. The first signal consists of the T-cell receptor (TCR)–peptide–MHC interaction. The second co-stimulatory signal resulting from the interaction between CD28 and B7 molecules, leading to the amplification of the signaling process in naive T-cells and protects activated cells from death. In absence of signal 2, T-cell activation will lead to anergy. Several inducible costimulatory molecules, in particular CTLA4, negatively modulate activation of T-cells. Notably, all of the constitutively expressed co-stimulators have positive amplifying effects, whereas the group of inducible costimulators contains positive (e.g., inducible co-stimulatory [ICOS]) and negative (e.g., CTL4 and PD-1) regulators. APC, antigen-presenting cell.

the cell by endocytosis **(Fig. 9)**. Thus, proteins released from tumor cells by secretion, shedding, or tumor lysis are captured by APCs. These antigens are processed, and peptides are presented to CD4+ cells by MHC class II. Notably, the majority of class II molecules acquire peptides in a specialized compartment, the so-called CPL. The synthesized class II molecules reach this compartment through the early endosomes. A special protein called invariant chain is responsible for the proper intracellular trafficking of class II molecules and peptides present in ER ***(40)***. Defects in either class I, II, or both types of antigen processing is one of the mechanisms by which bacteria, viruses, and tumor cells escape the immune system.

7. T-Cell Activation Requires at Least Two Signals

T-cell activation and differentiation require not only TCR recognition of the antigen–MHC complex but also co-stimulation through the interaction of accessory molecules on APCs and their corresponding receptors on T-lymphocytes ***(41)***. Therefore, class I and II presentation of peptide antigens to T-cells (signal 1) is not enough to induce their activation; coreceptor stimulation also is required (signal 2), which is delivered by the same APCs **(Fig. 10A)**. In the absence of signal 2, T-cells are inactivated via energy **(Fig. 10B)**. The outcome of T-cell responses after of recognition of specific antigens is modulated by costimulatory signals, which are required for both lymphocyte activation and development of adaptive immunity. The engagement of the CD28 molecule expressed on resting T-cells with the shared ligands B7.1 (CD80) and B7.2 (CD86) on activated APCs enhances the T-cell response. Other costimulatory molecules include lymphocyte function antigen-1 and -2 ***(42)***.

Sustained T-cell activation is expected to cause damage to the organism; thus, once T-cells are activated, they express additional molecules to modulate the immune response. Among the best studied of these counterregulatory pathways is the one initiated by engagement of the cytotoxic T-lymphocyte-associated antigen (CTLA) 4, which is a homolog to the CD28 and binds to B7.1 and B7.2 molecules with much higher affinity than CD28 **(Fig. 10C)**. CTLA4 is induced after T-cell activation and delivers inhibitory signals to T-cells that oppose the co-stimulatory signals delivered by CD28, leading to attenuation of T-cell activation late in the immune response ***(43)***. Inducible costimulatory (ICOS) is another member of the CD28 family, and it is not expressed by naive CD4+ cells, but it is induced after T-cell activation. Programmed death (PD) receptor ligand (PD-L) 1 (also called B7-H1) is a recently described B7 family member that binds to PD-1 expressed by T-cells ***(43)***. This signaling pathway has been reported to decrease TCR-induced proliferation and cytokine production.

In contrast to the expression of B7.1 and B7.2, PD-1 ligands are expressed not only on hematopoietic cells but also in nonlymphoid organs. The expression of PD-L1 and PD-L2 within nonlymphoid tissues suggests that these PD-1 ligands may modulate the autoreactivity of T- and B-cells in peripheral tissues. However, interaction of PD-L1 on tumor cells with PD-1 on tumor-specific T-cells may induce T-cell tolerance, leading to tumor escape. Together, these observations underlie the crucial role played by APCs in controlling T-cell activation and in determining the outcome of the immune responses whether directed against self or nonself.

8. T- and B-Cell Collaboration

Cross-linking of surface immunoglobulin by exogenous protein antigens can activate B-cells; however, this activation is not normally sufficient on its own to activate their proliferation and differentiation into plasma and memory B-cells. B-cell responses to these antigens are driven by interactions with a specialized subset of T-cells, namely, the helper T-cells ***(44)***. Thus, exogenous antigen, which is recognized by the surface IgM on the B-cell, is internalized, processed, and reexpressed on the MHC class II molecule of the B-cell **(Fig. 6)**. This B-cell can then present the antigen to a primed specific T-cell. When stimulated, the T-cell expresses CD40 ligand, which binds to CD40 on the B-cell, and the T-cell produces cytokines, leading to B-cell division and maturation to antibody-secreting plasma cells ***(45)***. T-cell interaction induces isotype switching from the initial IgM response. As soon as the isotype switch from IgM to another higher affinity antibody (IgG, IgA, and IgE) has occurred, some of the activated cells become long-lived memory cells. Memory is a characteristic hallmark of the adaptive immune response, thus making the host cells prepared to mount a more powerful immune response when they encounter the pathogen for the second time. Cooperation between both arms of the immune system is essential to produce optimized immune responses capable for eradicating infections. As mentioned above (*see* **Fig. 6**), B-cells also can respond to some antigens in a T-cell–independent manner by cross-linking of the receptors. However, these responses are limited to IgM, of poor affinity, and are short lived.

9. Th1 and Th2 Type Responses

Subsequent to T-cell development, two major types of effector T-cells are generated, Th cells bearing CD4 molecules on their surface and Tc cells bearing CD8 molecules **(Fig. 3)**. CD4+ T-cells only recognize antigen presented with MHC class II and CD8+ T-cells with MHC class I. CD4+ Th cells recognize foreign antigen and activate other parts of the cell-mediated immune response to eradicate the pathogen. They also play a major part in activation of B-cells

(44). Th cells can be further subdivided functionally by the pattern of cytokines they produce ***(46)***. Th1 cells produce IL-2 and interferon (IFN)γ, whereas Th2 cells produce IL-4, -5, and -6. Th1-type response plays a major role in cellular immunity, whereas Th2-type response is critical for cellular immunity. In mice, IgG2 B-cell response is associated with a Th1-like immune response and IgG1 with a Th2-like response ***(45)***. Human type 1 immunity is usually associated with IgG1 and IgG3 subclasses, and type 2 immunity is associated with the IgG4 isotype and IgE ***(47)***. There is one other Th subset, Th3, that secrets mainly the cytokine transforming growth factor (TGF)β, which can have suppressive activity on Th1 and Th2 cells ***(48,49)***. These T-cells may derive from either thymic Treg cells or peripheral T CD4+ Th cells under certain stimulation conditions **(Fig. 7)**.

CD8+ Tc cells are directly cytotoxic to cells bearing their specific antigen and are expected to play important role in tumor immunology through a process of cell-mediated cytotoxicity that involves Fas, perforin, or both. Perforin-mediated lysis requires a cognate interaction between the antigen-specific TCR on a T-lymphocyte and the specific antigen (usually a peptide) presented on an MHC molecule on the target cell's plasma membrane. Fas-mediated cytotoxicity involves the ligation of Fas on the target cell by Fas ligand on T-cells, but it does not require a cognate recognition interaction via the TCR. Cytotoxicity also may be signaled via the tumor necrosis factor (TNF) receptor on a target cell ***(8)***. It has been shown that the CD8+ T-cells also can be subdivided into IFNγ-producing Tc1 cells and IL-4–producing Tc2 cells similar to the Th subsets. Innate effector cells, such as DCs, NK, natural killer T-cells (NKT), and γδ T-cells, produce several cytokines, including Th1 and Th2 cytokines. Thus, several immunoregulatory cells in addition to T-cells control the outcome of an immune response. A type 1 immunity, which is regulated by IL-12 and IFNγ, plays a crucial role for eradication of tumors in vivo.

10. Immune Response to Tumors

Based upon the immunosurveillance model ***(1)***, the immune system can play a protective role in tumor development by recognizing and destroying nascent transformed cells. The first question for many of us was whether tumor-specific antigens exist, and, if so, whether there is immunological recognition of these antigens in humans. Studies in mice showed that the immune system can recognize and reject certain tumors and that immunodeficient mice, lacking INF-γ and RAG-2, have an increased incidence of cancer ***(50,51)***. In humans, the incidence of some cancers is increased in immunodeficient patients and is increased with age, owing to the immunosenescence. Evidence provided by autologous typing indicated the presence of patient-restricted antigens and shared antigens. Moreover, the presence of

tumor infiltrating lymphocyte supports the activation of these T-cells by tumor antigens ***(4,5)***. Indeed, specific homing of T-cells into tumor tissues compared with the surrounding normal tissues was found ***(5)***. The existence of immune response against tumors has now been validated in several studies (*see* Chapter 1–15, Volume 1). However, in contrast to the notion that the immune system should eliminate tumor cells, the developed responses failed to prevent tumor growth. According to the current view, the immune system also functions to select tumor variants with reduced immunogenicity capable of withstanding the tumor-suppressing activity of the immune response ***(52)***. This process, which is referred to as cancer "immunoediting," is a dynamic process composed of at least three phases: elimination, equilibrium, and escape. To become clinically detectable, cancer cells must circumvent both innate and adaptive immunity. Thus, the host immunological environment selects tumors during tumor development and therefore leads to the generation of the selected tumors that survive with reduced immunogenicity. This heartening truth, which arises because most of the tumor cells are under constant selection pressure for survival, has been underlined by several recent studies ***(52)***.

11. Technical Advances in Tumor Antigen Discovery

One aim of tumor immunology is to increase the antitumor immune response and to mediate killing of tumor cells by the host's immune system, in particular, Tc cells ***(53)***. Given the prevalence of autoantigens on human cancers and the ability of tumors to induce tolerance to neoantigens acquired during transformation, tolerance is emerging a central obstacle for immune recognition of human cancer antigens. Because tumor cells do not express costimulatory molecules, antigens presented by these cells will induce tolerance despite being tumor specific. Thus, one of the main challenges of this field is to identity tumor antigens capable of inducing protective antitumor immunity as well as to uncover the cellular and molecular mechanisms that lead to the strange behavior of the immune system toward tumor cells. To identify tumor antigens, here referred as tumor-associated antigens (TAAs) and tumor-specific antigens (TSAs), several techniques have been explored.

11.1. Autologous Typing

Autologous typing was an early approach of determining whether patients with cancer recognize tumor-specific antigens on the surface of their cancer cells ***(54)***. Sera from patients with cancer were tested for reactivity with surface antigens of cultured autologous cancer cells and normal cells. Positive reactivity simply signifies that the antigens being detected on the target cells. Because of the potential presence of alloantibodies directed against alloantigens not present on other autologous cell types, extensive absorption steps are needed to

demonstrate the specificity of binding. Using this method, a few antigens have been identified. A major impediment to progress in this approach was the general inability to grow primary human cells in tissue culture.

11.2. Biochemical Approach

This method relies on the purification and analysis of peptides eluted from MHC molecules of tumor cells. Acid elution of antigenic peptides bound to MHC class I molecules from tumor cells ***(55)***. Subsequently, tumor peptides are fractionated by high-performance liquid chromatography, and then the different fractions are tested for their ability to sensitize tumor cells for lysis by autologous tumor-specific cytotoxic CD8+ T-cell clones. Subsequently, peptide fractions conferring sensitivity to lysis by tumor-specific T-cells are sequenced. Although this method led to the identification of several human tumor antigens eliciting cellular immune responses ***(56)***, it also relies on the generation of autologous specific T-cell clones.

11.3. T-Cell-Based Identification of Recombinant Tumor Antigens

Several studies indicated that T-cells are a critical mediator of tumor immunity. Establishment of T-cell clones from tumor-infiltrated T-cells was important for this method. The first experimental approach is based on the transfection of recombinant cDNA libraries into cells expressing the MHC-presenting molecules ***(57,58)***. Autologous tumor-specific cytotoxic CD8+ T-cell clones (CTL) are then used to test for recognition of the transfected cells. In the case of positive response, cDNA fragments of the isolated gene are subcloned to define the region encoding the antigenic peptide. Subsequently, synthetic peptides are tested in a target cell sensitization assay for their ability to sensitize target cells (tumor cells) to lysis by the original tumor-specific CTLs. Using this approach, Boon and colleagues ***(57)*** reported the first successful cloning of human tumor antigen, termed melanoma antigen-1 or MAGE-1, which elicited a spontaneous CTL response in the autologous melanoma patient ***(57)***. Further studies showed MAGE-1 to be expressed exclusively in normal testis among normal tissues. The use of this "genetic" approach of T-cell epitope cloning also led to the identification of the BAGE and GAGE antigens, both of which are recognized by CTL of melanoma patients ***(58)***. Despite the identification of some tumor antigens, this approach is technically challenging in that it requires the establishment of autologous CTL lines/clones and tumor cell lines from the same patient, a task not achievable for most epithelial tumor types ***(58)***.

11.4. Serological Analysis of Autologous Tumor Antigens by Recombinant cDNA Expression Cloning (SEREX)

To overcome some of the technical problems related to the generation of T-cell clones, in 1995 Sahin and colleagues ***(59)*** developed the SEREX technique

(*see* Chapter 15, Volume 1). The method does not rely on the availability of T-cell clones or lines but rather screens antibody response of cancer patients against autologous (or allogeneic) tumor cDNA expression libraries. Briefly, cDNA libraries are constructed from tumor tissues or cell lines in a prokaryotic expression vector such as the lytic phage λ. The library is used to transduce *Escherichia coli*. Blotting to a nitrocellulose membrane transfers lytic plaques. The membranes are then incubated with diluted patient serum antibodies preabsorbed for removal of *E. coli* binding antibodies. Positive-reacting clones are detected by an enzyme-conjugated secondary antibody specific for human IgG. In their initial application of this method, they identified MAGE-1 and tyrosine, two antigens originally cloned as CTL targets, indicating that the strategy can identify tumor antigens that elicit a CTL-mediated immune response ***(59,60)***. Subsequent studies identified an array of TAAs in many human cancers, as demonstrated by more than 2000 different gene entries of potentially cancer related antigens (*see* Chapter 1–15, Volume 1). Notably, strong IgG antibody responses are considered to depend on the presence of T-cell help.

The SEREX antigens can be divided into several groups ***(61)***: (1) cancer-testis (CT) antigens (e.g., HOM-MEL-40) are selectively expressed in a variety of neoplasms, but not in normal tissues except for testis; (2) differentiation antigens show a lineage-specific expression in tumors and in normal cells, but not only at a defined stage of development (e.g., tyrosinase); (3) antigens encoded by mutated genes (e.g., p53); (4) splice variants (e.g., restin); (5) viral antigens (e.g., HERV-K10); (6) overexpressed genes (e.g., Galectin-9) are known to elicit immune response by overriding thresholds critical for the maintenance of tolerance; (7) overexpression because gene amplification (e.g., eIF-4γ); (8) cancer-related autoantigens (e.g., HOM-MEL-2.4); and (9) cancer-independent autoantigens (e.g., HOM-TES-11).

One of the technical problems with SEREX is that it is based upon a one-step screening. In contrast, the sensitivity and selectivity are expected to be high when the selection is performed through iterative, powerful enrichment steps, thus offering one of the major advantages of phage display method that we have established in 1993 for the purpose of profiling the immune response in patients with autoimmune diseases ***(62)***. The only prerequisite of this novel approach is the availability of patient sera. We have extended this method to patients with cancers, and several tumor antigens have been identified.

11.5. Phage Display

For the generation of peptides or protein collections for the discovery-oriented method, large numbers of synthetic peptides or recombinant proteins have to be prepared and purified, which is a major technical challenge. Therefore, recombinant peptides or proteins displayed on phages are highly recommended.

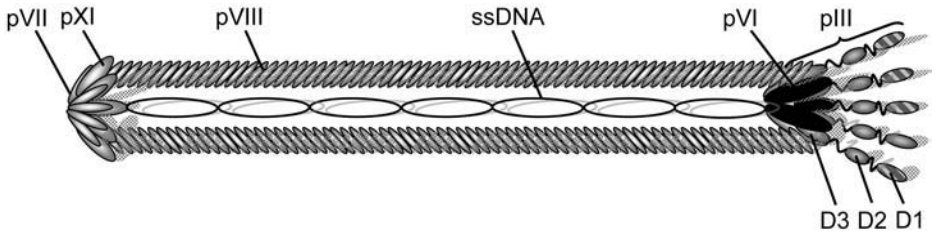


Fig. 11. Structure of the phage M13, showing minor and major coat proteins. Recombinant peptides and proteins can be expressed as fusion with either pIII, pVI, or pVIII coat proteins.

Peptides expressed on phages have a broad range of applications, which include drug and target discovery, protein evolution, and rational design. DNA can be cloned in frame at the amino terminus of the gene encoding for one of the two-capsid proteins pIII or pVIII, and the corresponding fusion product will be displayed on the surface of the virion **(refs. *63–65*; Fig. 11)**. Repertoires larger than 10^8 phage clones expressing different peptide sequences can be prepared from the host without lysis, giving very high titers of ligand-displaying phage. Notably, phage display offers the advantage of the displayed ligand being physically associated with the genetic information. Since the first reports, many different peptide phage libraries have been constructed and screened with several different ligates ***(65)***.

Because the integrity of the C terminus of pIII and pVIII is essential for efficient phage assembly, insertions of foreign peptides can only be tolerated at the N terminus. The use of phage for creating cDNA expression libraries has been hampered by the presence of stop codons in the 3′ untranslated regions of eukaryotic cDNA, which would prevent the fusion to the N terminus of the viral coat proteins. To overcome this problem, Crameri and Suter ***(66)*** explored the interaction between the Jun and Fos leucine zippers. Here, the Jun domain has been expressed as a fusion with pIII and Fos expressed as an N-terminal fusion peptide to a foreign protein. The resulting Fos-fusion protein associates with the Jun-expressing phage particles. Jesper et al. ***(67)*** have described a novel phagemid vector that allows the functional expression of full cDNA libraries as a fusion with the C terminus of pVI protein. The phage coat protein pVI is not known to be involved in infection, and its C terminus is thought to be surface exposed. Therefore, the presence of stop codons in full-length cDNA does not prevent display of cDNA repertoires fused to pVI. More importantly, the developed cloning system allows the display in all three reading frames. We have explored this expression system to identify tumor antigens ***(68)***.

A necessary step in filamentous phage assembly is the transport of fusion peptides or proteins to the periplasm. This transport is a major limitation to phage

display if the foreign protein cannot be secreted through the *E. coli* membrane. Additionally, the chemical characteristics of the periplasmic environment may affect the folding, processing, and stability of the protein to be displayed. In part because of such limitations, systems using lytic phage such as λ, T4, and T7 have been developed ***(69)***. Phage assembly takes place in the cytoplasm, and mature phages are released by cell lysis. This sequence implies that the protein included in the phage capsids is folded and assembled in the cytoplasm and does not require secretion through the membrane.

The T7 select TM display system exploits the product of gene 10 to display foreign peptides on the T7 capsid ***(70)***. The cDNA inserts are placed at the carboxy-terminal to the coat protein, so the expression of intact fusion protein is not affected by the presence of stop codon ***(71)***. The T7 bacteriophage has a very short life cycle and as a consequence, the time needed to perform multiple panning cycles is decreased compared with other types of display. Moreover, phage particles are extremely robust and resistant to harsh conditions that can be used for phage elution. We implemented this system in our laboratory and showed its feasibility in identifying tumor antigens ***(72)***.

11.6. Application of Phage Display to Study Immune Responses in Patients

The characterization of B- and T-cell specificities is crucial for the understanding of the immune response mechanism and its role in the prevention and cause of human diseases. However, the determination of B- and T-cell specificities requires access to the parental antigen(s) that initiate the immune response, a major limiting step if such antigens are not available. We have shown that random peptide libraries can identify serum antibody specificities whether or not the parental antigens are known ***(62,73,74)***, the rationale being that antibodies from patient sera will bind to the phage containing the peptide initiating the immune response against self- or nonself proteins. Thus, through this strategy, it is possible to probe any specific immune response in patients. Sera from patients, in addition to disease-specific antibodies, contains antibodies of irrelevant specificities. To enrich for the disease-specific epitopes, different approaches for subtraction or counterscreening have been developed ***(62,68)***.

An important application of the phage display technology will be the analysis of immune responses in patients with cancer, because it has been difficult to identify immunogenic antigens capable of inducing immune responses in patients to a level high enough for tumor rejection ***(68)***. Although, most of the studies carried out with phage display libraries have been applied to homogenous proteins, we have shown for the first time that peptide phage

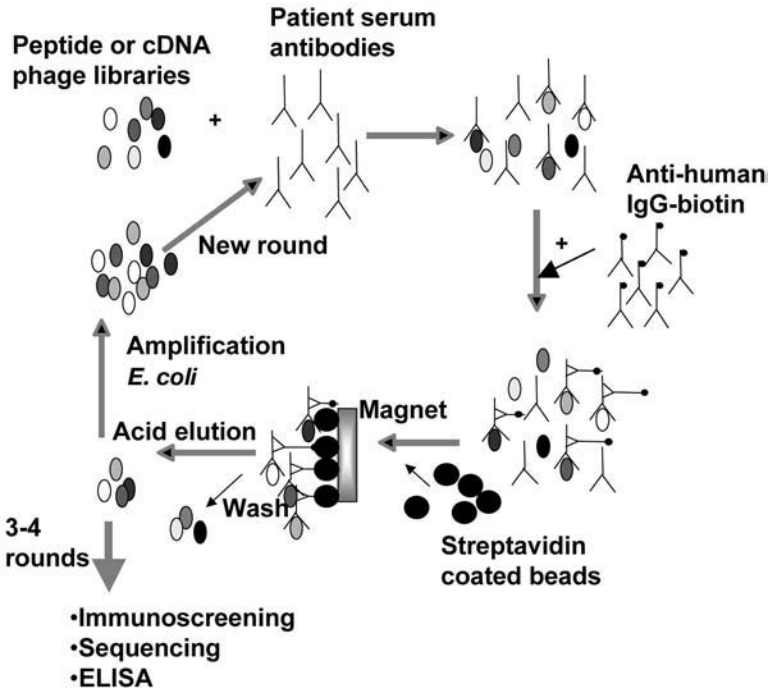


Fig. 12. Biopanning phage libraries on patient serum antibodies. In this protocol, the phage library (e.g., peptides and cDNAs) is incubated with patient serum antibodies. After incubation, biotin-conjugated anti-human IgG antibody was added to capture the reacting phages. After incubation, reactive phages are captured on streptavidin-coated magnetic beads, washed, acid eluted, and then amplified in *Escherichia coli*. Amplified phages are used for further rounds of selection.

libraries can be used to determine the specificities of antibodies present in whole sera of patients with breast cancers ***(75)***. Homology search with the selected peptides identified potential tumor antigens. Importantly, patients who responded to most of the selected peptides exhibited a good prognosis. The selected antibody ligands can be an important tool for diagnosis, prognosis, and immunotherapy. In the case of autoimmunity, the characterization of the autoimmune responses should facilitate the identification of new autoantigens that may act as tolerizing vaccines.

11.7. Biopanning and Immunoscreening Techniques

The phage display approach is far superior over one-step screening methods (e.g., SEREX) because sensitivity and selectivity are extremely high when selecting through iterative, powerful enrichment steps ***(68)***. To select phages

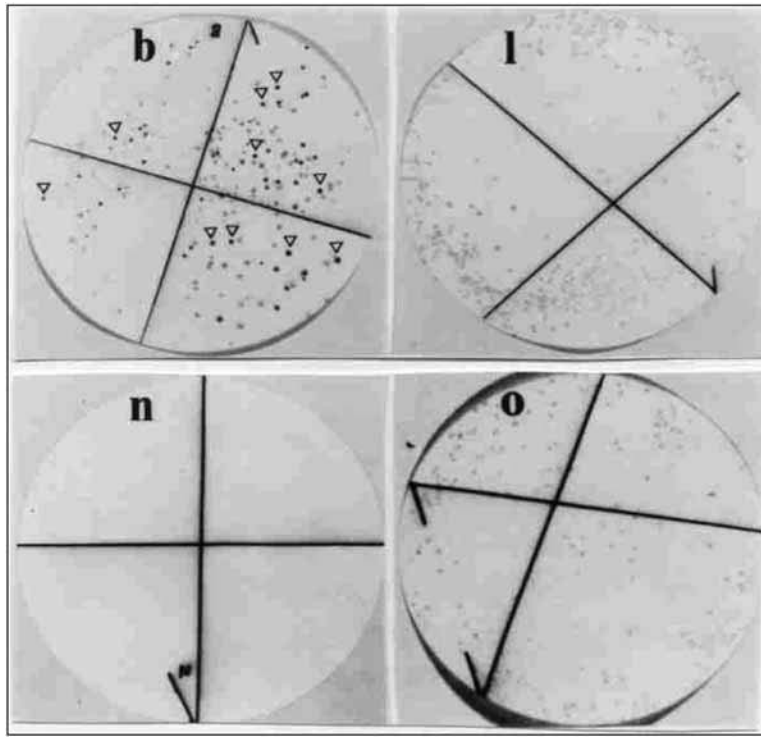


Fig. 13. Immunoscreening of phage-encoded peptides biopanned on serum antibodies from patients with breast cancer. After transfer to nitrocellulose and blocking, the membranes were incubated with 8 μg/mL pooled IgG fraction from breast cancer patients (b), lymphoma (l), osteosarcoma (o), or normal (n). Subsequent to incubation, alkaline phosphatase conjugated anti-human IgG antibodies detected the immunoreactive phages.

displaying a particular antigen (e.g., peptides, proteins, and cDNA repertoires), a biopanning technique is used **(Fig. 12)**. In this procedure, serum antibodies are incubated with the phage library and then biotin-conjugated anti-human IgG are added. Thus, serum antibody-binding phages can be immobilized on streptavidin-coated magnetic beads. After several washes to remove nonspecifically bound phages, the bound phages are acid eluted and amplified for the next round of selection. When using whole sera or purified IgG fraction, there are some important parameters to consider during the biopanning experiments. In contrast to monoclonal antibodies, whole serum affinity selection methods do not permit facile adjustment of the specific immune response in patient sera. In addition to tumor antigen-specific antibodies, patient sera contain a large proportion of physiological antibodies. This composition could result in enrichment of phage-displaying antigens for irrelevant antibodies ***(62,68)***. To overcome

some of these potential problems, we have introduced a subtraction step in which the undesired antibody-binding phages are subtracted by normal serum antibodies ***(62)***.

After the desired rounds of selection (usually three to four), individual phages should be tested for binding by, for example, ELISA. To simplify this time-consuming step, we have used a rapid immunological screening method ***(68,75,76)***. In this method, the immunoselected phage library is transferred to nitrocellulose membranes, and subsequently the filters are probed by either serum antibodies from patients or controls **(Fig. 13)**. Therefore, it is possible to screen many phages in a single step. The data shown in **Fig. 13** illustrate the power of the enrichment achieved by the phage display technology. A strong reactivity was found with pooled IgG antibodies from breast cancer patients, but no significant reactivity was found with IgG from normals, patients with lymphoma, or those with osterosarcoma. Positive clones are clearly distinguishable from negative clones, confirming the specificity of the immunoreaction.

11.8. Technical Limitation of SEREX and Phage Display Methodologies

The current bacterial expression systems are greatly limited in their ability to identify antigen that are posttranslationally modified. These modifications play an important role for proper folding and functioning of many proteins. Moreover, modifications such as glycosylations are known to affect the immunogenicity of proteins. As a consequence, several groups have established eukaryotic expression systems such as baculovirus ***(77)*** and retroviral surface display ***(78)***. Recently, the yeast *Saccharomyces cerevisiae* expression system was applied for the display and affinity maturation of an scFv antibody via the agglutinin mating adhesion receptor ***(79)***. Yeast display offers the advantage of using flow cytometry for rapid quantitative isolation of rare binders. The method was used for the isolation and characterization of tumor antigens recognized by autologous or allogeneic breast cancer serum ***(80)***.

11.9. Identification of Tumor Antigens by Expression Profiling of Cancer Cells

Another approach to identify tumor antigens is to search for potential T- and B-cell antigens by mRNA expression profiling, namely, by searching for genes that are expressed in cancer but not in other normal tissues ***(81–84)***. Several selection and global techniques have been used to monitor gene expression, including subtractive hybridization, PCR-based subtraction cloning, representational difference analysis, suppression subtractive hybridization, differential display, large-scale sequencing of expressed sequence tag (EST) sequences, serial analysis of gene expression, and DNA microarray array analysis (*see* Chapters 2–6, Volume 1). In addition to tumor antigen discovery, these

molecular techniques have proven to be a powerful approach for the discovery of novel therapeutic genes. So far, microararys are the most used methods for generating quantitative information about the expression of thousands of genes with relative rapidity. However, novel transcripts cannot be discovered by this method. To uncover new tumor antigens and therapeutic targets, we have combined PCR-based cDNA subtraction technique and Northern blotting ***(85)***. After cloning and sequencing, several potential known and new targets were identified, including ADP/ATP carrier protein, ErbB2, and ribosomal protein RPL19.

A key in tumor antigen discovery before proceeding to vaccine development is demonstrating the immunoreactivity of the new target antigens. A variety of in vitro and in vivo assays are used to identify peptide epitopes capable of inducing CD8+, CD4+, or both (*see* Chapter 1–15, Volume 1). Serum IgG antibody reactivity can be considered as a good indicator for activated humoral and cellular immunity. Significant antibody reactivity against several antigens was found in patients with breast cancer compared with controls ***(72,75,76,85)***.

Public access to the human genome sequences, cDNAs derived from normal and malignant tissues, together with the EST database has facilitated the discovery of tumor antigens through a bioinformatics approach (***[81,84]***; *see* Chapter 2, Volume 1). The data can be searched for overexpressed genes, which should be considered as candidate tumor antigens. However, as overexpression does not always imply that there will be an immune response against the protein in question; additional steps have to be integrated to direct the screening efforts toward identifying the best candidate antigen or therapeutic target for immunologic or therapeutic intervention. Differentially expressed genes need to be identified systematically (perhaps by the application of rapidly improving gene array technologies), and their immunogenicity needs to be determined. Ideally, candidate genes would only be expressed in tumor cells but not in any normal cells.

11.10. Proteomics and Tumor Antigen Identification

It is anticipated that RNA levels of many genes do not reflect protein contents; therefore, transcriptome analysis might be misleading for target discovery. Moreover, posttranscriptional modifications cannot be detected by this technique, and dynamic changes on protein level would not be revealed on the transcriptome level. Proteomics, the study of proteins and protein pathways, is a promising technique for tumor antigen discovery ***(86)***. It depends on a combination of methods, including two-dimensional gel electrophoresis, mass spectrometry, peptide sequencing, image analysis, and bioinformatics. The promise of this technology to identify novel antigen is described in Chapter 16, Volume 1. Using this method and

sera from patients with breast cancers, we have identified several potential tumor antigens, such as ADAM10, dihydrolipoamide *S*-acetyl transferase, and β-F1-ATPase.

12. Tumor Escape and Self-Tolerance

Cancer poses a difficult problem for immunotherapy because it arises from the host's own tissues. T-cells reactive with dominant determinants of tumor-associated antigens and perhaps subdominant epitopes are deleted in the thymus during negative selection. Thus, most of the tumor determinants are expected to be immunologically silent; hence, effective tumor immunity cannot be induced by tumor cells ***(87)***. Moreover, as tumors accumulate neoantigens during transformation, they also gradually induce tolerance in T-cells against these neoantigens ***(52)***. Thus, one of the major problems faced by tumor immunologists is the need to find strategies that induce immunity against self-antigens expressed by tumor cells. The goal of these strategies is to overcome one or more mechanisms of immune tolerance. During the last years, a number of approaches have been used by different groups, including the use of antigen expressed in xenogeneic cells, heteroclitic peptide strategies, xenogeneic DNA immunization, recombinant viruses expressing tumor antigens, whole cell vaccines (typically transfected with cytokines or costimulatory molecules), and exogenous use of cytokines ***(88)***.

The absence of efficient tumor-specific immunity can be related to several mechanisms, such as inadequate APC function and T-cell tolerance toward tumor antigens, which are seen as self-antigens ***(89)***. As mentioned above (*see* **Fig. 10**), initiation of the T-cell response requires stimulation of the TCR via its peptide/MHC ligand on the APCs as well as co-stimulatory signals ***(5)***. Several signaling events can prevent the initiation of T-cell activation against tumors, including the CTLA4 pathway ***(43)***. Moreover, recent reports indicate that several subsets of Treg cells are involved in controlling T-cell tolerance in the periphery. In addition to the naturally occurring CD4+CD25+ cell population, which has been shown to be continuously produced within the thymus, other T-cell subsets bearing suppressive ability have been described. Among these subsets, the most important are IL-10–producing type 1 regulatory (Tr1) cells and TGFβ-producing Th3 cells ***(89)***. These cells can be derived from the peripheral naive CD4+ T-cell population under particular conditions of antigenic stimulation (*see* **Fig. 7**). In addition to IL-10, CTLA4 plays an essential role in Treg function. It is constitutively expressed by Treg cells (CD4+CD25+) but not by other CD4+ T-cell subsets. Tumor cells also secrete immunosuppressive cytokines, such as TGFβ and IL-10, which can inhibit T-lymphocyte effector function. Downregulation of these suppressive cytokines by small-interfering RNAs (siRNAs) may increase tumor immunity ***(90)***.

Although much of the initial attention focused on the role of Treg cells in controlling autoreactivity, there is growing evidence that they significantly impact on the host response to cancer. In murine models, antibody-mediated depletion of CD25+ Treg cells enhanced the induction of tumor immunity *(90,91)*. However, specific depleting antibodies are limited; therefore, new strategies to modulate survival or apoptosis of Treg are warranted. Reduction of CD25 Treg cells triggered by certain drugs has been shown to enhance tumor immunity elicited by vaccination *(92)*.

An additional and highly significant form of immune escape is the ability of tumor cells to evolve mechanisms that impede antigen processing and presentation, expression of costimulatory molecules, or a combination *(89)*. Recent studies demonstrated that some tumor cells could overexpress inhibitors of T-cell activation, such as galectin. Thus, a tumor may directly inhibit antitumor responses by multiple mechanisms *(52)*.

The immunoediting model predicts that in all cases, clinically detectable tumors will develop and may result in death of the host. However, it cannot explain why tumors are sometimes spontaneously rejected *(93)*. In some patients, long-term remission was observed without treatment. Thus, surviving tumors that acquired insensitivity to immunologic detection, elimination, or both are destroyed! In my opinion, the analysis of the immune responses in this category of patients who spontaneously recovered from cancer should facilitate the design of better vaccines. More recently, we have found that some patients with breast cancer have developed an antitumor immune response with diversification of B- and perhaps T-cell epitopes *(94)*. This process, generally termed "epitope spreading," was found in all destructive autoimmune diseases *(94)*. Epitope spreading is more likely to depend on many factors, including the nature of the autoantigen(s) and the level of established immunologic tolerance. Thus, the identification of tumor antigens capable of triggering a strong immune response with epitope spreading should be considered as invaluable candidates for cancer vaccines.

13. Dendritic T-Cells and Cancer Vaccines

One of the immune cell types that link innate and adaptive immunity and play a major role for cancer immunotherapy is DCs *(95,96)*. These cells are able to affect immunity through several signals, leading to either activation or inhibition of T-cell response *(96)*. DCs, via the production of soluble cytokines, also activate other cell types such as NKT cells and NK cells. Immature DCs are highly phagocytic, but in the absence of co-stimulation, their activation can lead to tolerance *(97)*. Under normal conditions, phagocytosed self-antigens should lead to the induction of tolerance to self and thereby prevent autoimmunity **(Fig. 7)**. However, immature DCs that captured exogenous antigens undergo maturation,

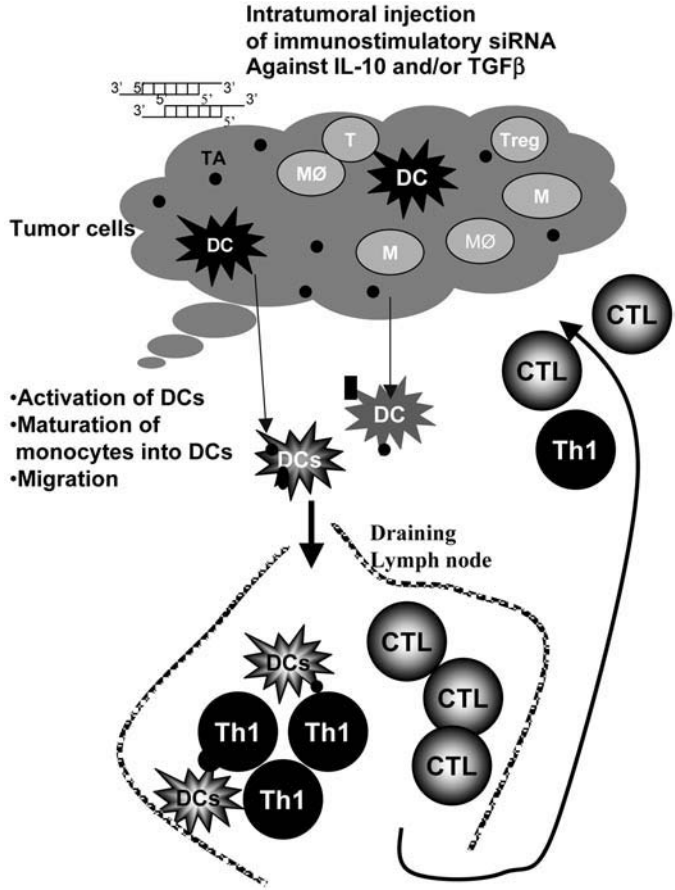


Fig. 14. Immunostimulatory small-interfering RNA (siRNA)-liposomes can be highly effective adjuvant for bridging between type 1 innate and acquired immunity. Administration of double-stranded siRNA or single-stranded siRNAs in liposomes induced the activation of Toll-like receptor 7- and 8-expressing monocytes and dendritic cells (DCs) to produce cytokines (e.g. tumor necrosis factor [TNF]α, IL-12) and interferons (IFNα), which can activate natural killer (NK) cells and natural killer T-cells (NKT) ***(97)***. The produced cytokines induce the activation DCs and potential maturation of tumor monocytes into dendritic cells, which will activate T- and B-cells, leading to T-helper (Th) type 1 response. Activated cytoxic T-cells (CTL) migrate into the tumor site from the draining lymph nodes and eventually will eradicate tumor cells.

which enables them to migrate to the lymph nodes and to present peptides derived from the acquired antigens to lymph node-residing T-cells, leading to a cellular immune response that involves both CD4+ and CD8+ T-cells **(Fig. 7)**.

Because of their important role in immunity, ex vivo-generated antigen-loaded DCs have now been used as vaccines to improve immunity. Although DCs can be obtained from several sources, for clinical trials monocyte-derived DCs are the most widely used DCs. Monocytes can be differentiated into either macrophages, which function as scavengers, or DCs, which induce specific immune responses. Different cytokines skew the in vivo differentiation of monocytes into DCs with different phenotypes and functions. Alteration of DC function can be achieved by transfection with immunomodulatory molecules. Endogenous signal also can be inhibited by several techniques, such as the use of antisense oligonucleotides and siRNAs.

It is worth noting that the inherent immunogenicity of DCs can be unregulated by stimulation of these cells through various receptors, such as the TLRs. More recently, we have found that certain siRNA sequences activated monocytes and DCs via endosomal TLRs, leading to the production of TNFα and IFNα ***(98)***. Internalization of siRNAs was a prerequisite to activate TLR7 and -8 signaling pathways, suggesting that encapsulation of immunostimulatory siRNA in liposomes may increase their recognition by the innate effector cells such DCs and macrophages. Besides the activation of DC maturation by immunostimulatory siRNAs, the inhibition of immunosuppressive cytokines and host factors by siRNAs should lead to successful cancer vaccines. Notably, IL-10–producing Treg cells and IL-10–producing regulatory DCs suppress cellular immune responses. By targeting IL-10 with an immunostimulatory siRNA, we hope to both downregulate the expression IL-10 and activate innate type 1 response against tumors, as illustrated in **Fig. 14.** Several recent studies have documented the utility of siRNA-mediated gene silencing in vitro and in vivo ***(90)***. Therapeutic vaccination combined with strategies that seek to inactivate the function of Treg cells via the use the RNA interference should lead to the design of better vaccines.

References

1. Burnet, F. M. (1971) Immunological surveillance in neoplasia. Transplant. Rev. **7,** 3–25.
2. Diefenbach, A. and Roulette, D. H. (2002) The innate immune response to tumors and its role in the induction of T-cell immunity. *Immunol. Rev*. **188,** 9–21.
3. Rosenberg, S. A. (1999) A new era for cancer immunotherapy based on the genes that encodes cancer antigens. *Immunity* **10,** 281–287.
4. Vose, B. M. and Moore, M. (1985) Human tumor-infiltrating lymphocytes: a marker of host response. *Semin. Hematol*. **22,** 27–40.
5. Østenstad, B., Sioud, M., Schlichting, E., Lea, T., and Harboe, M. (1995) Freshly isolated tumour-infiltrating T-lymphocytes have a high cytotoxic potential, as

measured by their ability to induce apoptosis in the target cell. *Scand. J. Immunol*. **41,** 42–48.
6. O'Neill, L. A. J. (2004) TLRs: Professor Mechnikov, sit on your hat. *Trends Immunol*. **25,** 687–693.
7. Janeway, C. A. (2001) How the immune system protects the host from infection. *Microbes Infect.* **3,** 1167–1171.
8. Janeway, C. A., Travers, P., Walport, M., and Shlomchik, M. (eds.) (2001) Immunobiology. The Immune System in Health and Disease. Garland Publishing, New York.
9. Matzinger, P. (1994) Tolerance, danger, and the extended family. *Annu. Rev. Immunol*. **12,** 991–1045.
10. Janeway, C. A. Jr. (1992) The immune system evolved to discriminate infectious nonself from noninfectious self. *Immunol. Today* **13,** 11–16.
11. Singer, A. (2002) New perspectives on a developmental dilemma: the kinetic signalling model and the importance of signal duration for the CD4/Cd8 lineage decision. *Curr. Opin. Immunol*. **14,** 207–215.
12. Orkin, S. H. (2000) Diversification of haematopoitic stem cells to specific lineages. *Nat. Rev. Genet*. **1,** 57–64.
13. Izon, D. J., Punt, A. J., and Pear, W. S. (2002) Deciphering the role of Notch signalling in lymphocytes. *Curr. Opin. Immunol*. **14,** 192–199.
14. Wilson, A., MacDonald, H. R., and Radtke, F. (2001) Notch 1-deficient common lymphoid precursors adopt a B cell fate in the thymus. *J. Exp. Med*. **194,** 1003–1012.
15. Aifantis, I., Azogui, O., Feinberg, J., Saint-Ruf, C., Buer, J., and von Boehmer, H. (1998) On the role of the pre-T-cell receptor in αβ versus γδ T lineage commitment. *Immunity* **9,** 649–655.
16. Iritani, B. M., Alberola-Il, J., Forbush, K. A., and Perimutter, R. M. (1999) Distinct signals mediate maturation and allelic exclusion in lymphocytes progenitors. *Immunity* **10,** 713–722.
17. Muljo, S. A. and Schlissel, M. S. (2000) Pre-B and pre-T-cell receptors: conservation of strategies in regulating early lymphocyte development. *Immunol. Rev.* **175,** 80–93.
18. Sleckman, B. and Khor, B. (2002) Allelic exclusion at the TCRb locus. *Curr. Opin. Immunol*. **14,** 230–324.
19. Myklebust, J. H., Smeland, S. M., Josefsen, D., and Sioud, M. (2000) Protein kinase C-alpha isoform is involved in erythropoietin-induced erythroid differentiation of CD34(+) progenitor cells from human bone marrow. *Blood* **95,** 510–518.
20. Padovan, E., Casorati, G., Dellabona, P., Meyer, S., Brockhaus, M., and Lanzavecchia, A. (1993) Expression of two T cell receptor alpha chains: dual receptor T cells. *Science* **262,** 422–424.
21. Heath, W. R., Carbone, F. R., Bertolino, P., Kelly, J., Cose, S., and Miller, J. F. (1995) Expression of two T cell receptor alpha chains on the surface of normal murine T cells. *Eur. J. Immunol*. **25,** 1617–1623.

22. He, X., Janeway, C. A. Jr., Levine, M., et al. (2002) Dual receptor T cells extend the immune repertoire for foreign antigens. Nat. Immunol. **3,** 127–134.
23. Kretz-Rommel, A. and Rubin, R. L. (2000) Disruption of positive selection of thymocytes causes autoimmunity. *Nature Med.* **6,** 298–305.
24. Nemazee, D. (2000) Receptor selection in B and T lymphocytes. *Annu. Rev. Immunol*. **18,** 19–51.
25. Hardy, R. R. and Hayakawa, K. (2001) B cell development pathways. *Annu. Rev. Immunol.* **19,** 521–621.
26. Gay, D., Saunders, T., Camper, S., and Weigert, M. (1993) Receptor editing: an approach by autoreactive B cells to escape tolerance. *J. Exp. Med*. **177,** 999.
27. Tiegs, S. L., Russell, D. M., and Nemazee, D. (1993) Receptor editing in self-reactive bone marrow B cells. *J. Exp. Med.* **177,** 1009.
28. Casali, P. and Diaz, M. (2006) Somatic immunoglobulin hypermutation. *Curr. Opin. Immunol*. **14,** 235–240.
29. Durandy, A. and Honjo, T. (2001) Human genetic defects in class-switch recomnination (hyper-IgM syndromes). *Curr. Opin. Immunol*. **13,** 543–548.
30. Nemazee, D. and Weigert, M. (2000) Revising B cell receptors. *J. Exp. Med.* **191,** 1813–1817.
31. Ferber, I., Schonrich, G., Schenkel, J., Mellor, A. L., Hammerling, G. J., and Arnold, B. (1994) Levels of peripheral T cell tolerance induced by different doses of tolerogen. *Science* **263,** 674–676.
32. Goodnow, C. C., Cryster, J. G., Hartley, S. B., et al. (1995) Self-tolerance checkpoints in B lymphocyte development. *Adv. Immunol*. **59,** 279–368.
33. Maloy, K. J. and Powrie, F. (2001) Regulatory T cells in the control of immune pathology. *Nat. Immunol*. **2,** 816–822.
34. Shortman, K. and Liu, Y. J. (2002) Mouse and human dendritic cell subtypes. *Nat. Rev. Immunol*. **2,** 151–161.
35. Cooper, C. J., Turk, G. L., Sun, M., Farr, A. G., and Fink, P. J. (2004) Cutting edge: TCR revisison occurs in germinal centers. *J. Immunol*. **173,** 6532–6536.
36. Ernst, B., Lee, D. S., Chang, J. M., Sprent, J., and Surch, C. D. (1999) The peptide ligands mediating positive selection in the thymus control T-cell survival and homeostatic proliferation in the periphery. *Immunity* **11,** 173–181.
37. Dorfman, J. R. and Germain, R. N. (2002) MHC-dependent survival of naïve T cells? A complicated answer to a simple question. *Microbes Infect*. **4,** 547–554.
38. Stoltze, L., Nussbaum, A. K., Sijts, A., Emmerich, N. P., Kloetzel, P. M., and Schild, H. (2000) The function of the proteasome system in MHC class I antigen processing. *Immunol. Today* **21,** 317–319.
39. Uebel, S. and Tampe, R. (1999) Specificity of the proteasome and the TAP transporter. *Curr. Opin. Immunol*. **11,** 203–208.
40. Bertolino, P. and Rabourdin-Combe, C. (1996) The MHC class II-associated invariant chain: a molecule with multiple roles in MHC class II biosynthesis and antigen presentation to CD4+ T cells. *Crit. Rev. Immunol*. **16,** 359–379.
41. Greenfield, E. A., Nguyen, K. A., and Kuchroo, V. K. (1998) CD28/B7 costimulation: a review. *Crit. Rev. Immunol*. **18,** 389–418.

42. Hogg, N. and Landis, R. C. (1993) Adhesion molecules in cell interactions. *Curr. Opin. Immunol.* **5,** 383–390.
43. Chen, L. (2004) Co-inhibitory molecules of the B7-CD28 family in the control of T-cell immunity. *Nat. Rev. Immunol*. **4,** 336–347.
44. Clark, E. A. and Ledbetter, J. A. (1994) How B and T cells talk to each other. *Nature* **367,** 425–428.
45. Stevens, T. L., Bossie, A., Sanders, V. M., et al. (1988) Regulation of antibody isotype secretion by subsets of antigen-specific helper T cells. *Nature* **334,** 255–258.
46. Swain, S. L., Bradley, L. M., Croft, M., et al. (1991) Helper T-cell subsets: phenotype, function and the role of lymphokines in regulating their development. *Immunol. Rev*. **123,** 115–144.
47. Bonifacio, E., Scirpoli, M., Kredel, K., Föchtenbusch, M., and Ziegler, A.-G. (1999) Early autoantibody responses in prediabetes are IgG1 dominated and suggest antigen-specific regulation. *J. Immunol*. **163,** 525–532.
48. Weiner, H. L. (2001) Induction and mechanism of action of transforming growth factor-beta-secreting Th3 regulatory cells. *Immunol. Rev*. **182,** 207–214.
49. Shevach, E. M. (2001) Certified professionals: CD4(+)CD25(+) suppressor T cells. *J. Exp. Med*. **193,** F41–F46.
50. Pardoll, D. (2003) Does the immune system see tumors as foreign or self? *Annu. Rev. Immunol*. **21,** 807–839.
51. Boon, T. and Kellermann, O. (1977) Rejection by syngeneic mice of cell variants obtained by mutagenesis of a malignant teratocarcinoma cell line. *Proc. Natl. Acad. Sci. USA* **74,** 272–275.
52. Dunn, G. P., Bruce, A. T., Ikeda, H., Old, L. J., and Schreiber, R. D. (2002) Cancer immunoediting: from immunosurveillance to tumor escape. *Nat. Immunol.* **3,** 991–998.
53. Rosenberg, S. A., Yang, J. C., and Restifo, N. P. (2004) Cancer immunotherapy: moving beyond current vaccines. *Nat. Med*. **10,** 909–915.
54. Pfreundschuh, M., Shiku, H., Takahashi, T., et al. (1978) Serological analysis of cell surface antigens of malignant human brain tumors. *Proc. Natl. Acad. Sci. USA* **75,** 5122–5126.
55. Van Bleek, G. M. and Natheson, S. G. (1990) Isolation of an endogenously processed immunodominant viral peptide from the class I H-2Kb molecule. *Nature* **348,** 213–216.
56. Falk, K., Rotzschke, O., Stevanovic, S., Jung, G., and Rammensee, H. G. (1991) Allele-specific motifs revealed by sequencing of self-peptides eluted from MHC molecules. *Nature* **351,** 290–296.
57. van der Bruggen, P., Traversari, C., Chomez, P., et al. (1991) A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* **254,** 1643–1647.
58. Sahin, U., Tureci, O., Chen, Y. T., et al. (1998) Expression of multiple cancer/testis (CT) antigens in breast cancer and melanoma: basis for polyvalent CT vaccine strategies. *Int. J. Cancer* **78,** 387–389.

59. Sahin, U., Tureci, O., Schmitt, H., et al. (1995) Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc. Natl. Acad. Sci. USA* **92,** 11,810–11,813.
60. Jäger, D., Stockert, E., Gure, A. O., et al. (2001) Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. *Cancer Res*. **61,** 6197–6204.
61. Jäger, D., Jäger, E., and Knuth, A. (2001) Immune responses to tumour antigens: implications for antigen specific immunotherapy of cancer. *J. Clin. Pathol*. **54,** 669–674.
62. Dybwad, A., Førre, Ø., Kjeldsen-Kragh, J., Natvig, J. B., and Sioud, M. (1993) Identification of new B cell epitopes in the sera of rheumatoid arthritis using a random nanopeptide phage library. *Eur. J. Immunol*. **23,** 3189–3193.
63. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228,** 1315–1317.
64. Smith, G. P. (1993) Surface display and peptide libraries. *Gene* **128,** 1–2.
65. Felici, F., Luzzago, A., Monaci, P., Nicosia, A., Sollazzo, M., and Traboni, C. (1995) Peptide and protein display on the surface of filamentous bacteriophage. *Biotechnol. Annu. Rev*. **1,** 149–183.
66. Crameri, R. and Suter, M. (1993) Display of biologically active proteins on the surface of filamentous phages: a cDNA cloning system for selection of functional gene products linked to the genetic information responsible for their production. *Gene* **137,** 69–75.
67. Jespers, L. S., Messens, J. H., De Keyser, A., et al. (1995) Surface expression and ligand-based selection of cDNAs fused to filamentous phage gene VI. *Biotechnology* **13,** 378–382.
68. Sioud, M., Hansen, M. H., and Dybwad, A. (2000) Profiling the immune responses in patient sera with peptide and cDNA display libraries. *Int. J. Mol. Med.* **6,** 123–128.
69. Castognoli, L., Zucconi, A., Quondam, M., et al. (2001) Alternative bacteriophage display systems. *Comb. Chem. High Throughput Screen*. **4,** 121–133.
70. Rosenberg, A., Griffin, K., Studier, F. W., et al. (1996) T7Select Phage Display System: A powerful new protein display system based on bacteriophage T7. *Innovations* **6,** 1–6.
71. Sche, P. P., McKenzie, K. M., White, J. D., and Austin, D. J. (1999) Display cloning: functional identification of natural product receptors using cDNA-phage display. *Chem. Biol*. **6,** 707–716.
72. Hansen, M. H., Østenstad, Ø., and Sioud, S. (2001) Identification of immunogenic antigens using a phage-displayed cDNA library from an invasive ductal breast carcinoma tumour. *Int. J. Oncol*. **19,** 1303–1309.
73. Sioud, M., Førre, Ø., and Dybwad, A. (1996) Selection of ligands for polyclonal antibodies from random peptide libraries: potential identification of (auto)antigens that may trigger B and T cell responses in autoimmune diseases. *Clin. Immunol. Immunopathol*. **79,** 105–114.

74. Dybwad, A., Bogen, A., Natvig, J. B., Førre, Ø., and Sioud, M. (1995) Peptide phage libraries can be an efficient tool for identifying antibody ligands for polyclonal antisera. *Clin. Exp. Immunol*. **102,** 438–442.
75. Hansen, M. H., Østenstad, B., and Sioud, M. (2001) Antigen-specific IgG antibodies in stage IV long-term survival breast cancer patients. *Mol. Med*. **7,** 230–239.
76. Sioud, M. and Hansen, M. H. (2001) Profiling the immune response in patients with breast cancer by phage-displayed cDNA libraries. *Eur. J. Immunol*. **31,** 716–725.
77. Boublik, Y., Di Bonito, P., and Jones, I. M. (1995) Eukaryotic virus display: engineering the major surface glycoprotein of the *Autographa californica* nuclear polyhedrosis virus (AcNPV) for the presentation of foreign proteins on the virus surface. *Biotechnology* **13,** 1079–1084.
78. Russel, S. J., Hawkins, R. E., and Winter, G. (1993) Retroviral vectors displaying functional antibody fragments. *Nucleic Acids Res*. **21,** 1081–1085.
79. Boder, E. T. and Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nat. Biotechnol*. **15,** 553–557.
80. Wadle, A., Mischo, A., Imig, J., et al. (2005) Serological identification of breast cancer-related antigens from a *Saccharomyces cerevisiae* surface display library. *Int. J. Cancer* **117,** 104–113.
81. Velculescu, V. E., Zhang, L., Vogelstein, B., and Kinzler, K. W. (1995) Serial analysis of gene expression. *Science* **270,** 484–487.
82. Lash, A. E., Tolstoshev, C. M., Wagner, L., et al. (2000) SAGEmap; a public gene expression resource. *Genome Res.* **10,** 1051–1060.
83. Boon, K., Osorio, E. C., Greenhut, S. F., et al. (2002) An anatomy of normal and malignant gene expression. *Proc Natl. Acad. Sci. USA* **99,** 11,287–11,292.
83a. Loging, W.T. et al. (2000) Identifying potential tumor markers and antigens by database mining and rapid expression screening. Genome Res. **10,** 1393–1402.
84. Segal, E., Fiedman, N., Kaminski, N., Regev, A., and Koller, D. (2005) From signatures to models: understanding cancer using microarrays. *Nat. Genet*. **37,** S38–S45.
85. Leirdal, M., Shadidy, M., Røsok, Ø., and Sioud, M. (2004) Identification of genes differentially expressed in breast cancer cell line SKBR3: Potential identification of new prognostic biomarkers. *Int. J. Mol. Med*. **14,** 217–222.
86. de Hoog, C. L. and Mann, M. (2004) Proteomics. *Annu. Rev. Genomics Hum. Genet*. **5,** 267–293.
87. Sioud, A. and Sørensen, D. (2003) Generation of an effective anti-tumor immunity after immunization with xenogeneic antigens. *Eur. J. Immunol*. **33,** 38–45.
88. Davis, I. D., Jefford, M., Parente, P., and Cebon, J. (2003) Rational approaches to human cancer immunotherapy. *J. Leukoc. Biol*. **73,** 3–29.
89. Zou, W. (2005) Immunosuppressive networks in the tumour environment and their therapeutic relevance. *Nat. Rev*. **5,** 263–274.
90. Sioud, M. (2004) Therapeutic siRNAs. *Trends Pharmacol. Sci*. **25,** 22–28.

91. Golgher, D., Jones, E., Powrie, F., Elliot, T., and Gallimore, A. (2002) Depletion of CD25 regulatory cells uncovers immune responses to shared murine tumour rejection antigens. *Eur. J. Immunol*. **32,** 3267–3275.
92. Dudley, M. E., Wunderlich, J. R., Robbins, P. F., et al. (2002) Cancer regression and autoimmunity in patients after clonal repopulation with antitumor lymphocytes. *Science* **298,** 850–854.
93. Wiernik, P. H. (1976) Spontaneous regression of hematologic cancers. *Natl. Cancer Inst. Monogr*. **44,** 35–38.
94. Sioud, M. (2002) How does autoimmunity cause tumor regression? A potential mechanism involving cross-reaction through epitope mimicry. *Mol. Med*. **8,** 115–119.
95. Banchereau, J. and Steinman, R. M. (1998) Dendritic cells and the control of immunity. *Nature* **392,** 245–252.
96. Langenkamp, A., Messi, M., Lanzavecchia, A., and Sallusto, F. (2000) Kinetics of dendritic cell activation: impact on priming of TH1, TH2 and nonpolarized T cells. *Nat. Immunol*. **1,** 311–316.
97. Steinman, R. M., Hawiger, D., and Nussenzweig, M. C. (2003) Tolerogenic dendritic cells. *Annu. Rev. Immunol*. **21,** 685–711.
98. Sioud, M. (2005) Induction of inflammatory cytokines and interferon responses by double-stranded and single-stranded siRNAs is sequence-dependent and requires endosomal localization. *J. Mol. Biol*. **348,** 1079–1090.

15

Potential Target Antigens for Immunotherapy Identified by Serological Expression Cloning (SEREX)

Dirk Jäger

Summary

Immunotherapy in cancer relies on the identification and characterization of potential target antigens that can be recognized by effector cells of the immune system. Several strategies have been developed to identify such antigens, which then can be used for immunization strategies. Serological analysis of recombinant tumor cDNA expression libraries (SEREX) identifies tumor antigens based on a spontaneous humoral immune response in cancer patients. SEREX is not limited to tumor types that can be grown in cell culture nor does it depend on T-cell clones that recognize the autologous tumor. SEREX-defined antigens need to be evaluated following an algorithm of several analytical steps before they become new target antigens for active immunotherapy: expression analysis to evaluate tumor association, serological analysis with sera from tumor patients and normal individuals to prove tumor-associated immunogenicity, identification of potential peptide epitopes for CD8 and CD4 T-cells, and evaluation in T-cell assays to demonstrate their potential use as vaccine targets. We recently identified a new breast cancer differentiation antigen designated as NY-BR-1 in an autologous breast cancer SEREX screening. The different steps of further evaluation are summarized in this chapter.

Key Words: Immunotherapy; serological analysis of recombinant tumor cDNA expression libraries; SEREX; reverse immunology; tumor antigens.

1. Introduction

The immune system interacts with tumor cells during the natural course of a malignant disease. It was shown for different tumor types by several groups that the presence of tumor-infiltrating lymphocytes is associated with a better prognosis in individual patients ***(1–4)***. This observation suggests that tumor-infiltrating lymphocytes recognize tumor-associated antigens expressed by the tumor cells or by the surrounding tumor stroma (*see* Chapter 14, Volume 1).

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols*
Volume I, Emerging Strategies for Targets and Biomarker Discovery
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

Furthermore, this immune response seems to be effective, resulting in a better prognosis and prolonged survival of the individual patient. A number of different cloning strategies have been developed over the past 15 yr to identify tumor antigens recognized by T-cells (*see* Chapter 14, Volume 1): the T-cell epitope cloning technique introduced by T. Boon's group ***(5)*** was successful in identifying tumor antigens based on spontaneous T-cell (cytotoxic T-lymphocyte [CTL]) responses. The first tumor-associated antigen identified with this technique was the melanoma antigen MAGE-1. During the following years, new tumor-associated antigens such as the BAGE and GAGE gene family were identified using the same cloning strategy ***(6,7)***. These antigens share some characteristic features: their expression in normal tissues is restricted to germ cells in the testis and the ovary, whereas these antigens are widely expressed in tumor tissues. Therefore, this group of antigens was recognized as a seperate category, named "cancer-testis" antigens ***(8,9)***. A second category of antigen comprises the differentiation antigens that are expressed in normal cells and tissue, such as melanocytes and the malignant counterpart, the melanoma. Using the same T-cell epitope cloning technique, Melan-A ***(10)***, tyrosinase ***(11)***, gp100 ***(12)***, and gp75 ***(13)*** were identified as CTL-defined tumor antigens. Tumor-specific mutated antigens (CDK4 and p53) also can be targets for CTL ***(14,15)***. Mutated antigens are a category of individual antigens; typically, the mutation is not shared by other individuals, with few exceptions such as the k-*ras* oncogene mutated in pancreatic cancer. Viral and retroviral antigens are frequently expressed in cancers such cervical carcinoma, and they also can be recognized by T-lymphocytes ***(16)***.

A different cloning strategy analyzes peptide sequences eluted from separated major histocompatibility complex (MHC) class I complexes isolated from tumor cells lysates. By comparing the peptide sequence to protein databases, the relevant tumor antigen can often be identified ***(17)***.

The identification of tumor antigens that elicited spontaneous T-cell responses in cancer patients and the discovery of the mechanisms of antigen expression, processing, and MHC class I-restricted presentation of antigen-derived peptides made it possible to design peptide-based immunotherapy protocols. Patients with antigen-positive tumors were vaccinated with the relevant peptides that are naturally processed and presented in the context of the respective MHC class I molecule ***(18)***. It was shown in several phase I and II clinical studies that peptide-specific cellular immune responses (CD4 and CD8) can be induced by immunizing patients with the respective peptide ***(19–25)***. Objective clinical remissions of single metastases and long-term clinical stabilizations were observed in some immunized patients, whereas progression of antigen-negative lesions was frequently documented. This tumor escape phenomenon occurring in the presence of potent antigen-specific T-cell responses reflects the

development of clonal heterogeneity in most solid tumors. To circumvent the selection and outgrowth of antigen-negative tumor cell clones that have a proliferation advantage under the selection pressure of antigen specific immunotherapy, polyvalent immunotherapy approaches targeting several tumor antigens expressed at the same time by an individual tumor need to be developed. To design such polyvalent immunotherapy strategies, it will be necessary to dissect the antigenic profile of an individual tumor.

2. Serological Analysis of Recombinant Tumor cDNA Expression Libraries

The development of the new cloning technique SEREX with autologous serum ***(26)*** to identify tumor antigens based on spontaneous antibody responses in cancer patients made it possible to analyze tumor systems that typically do not grow in cell culture. Another advantage is that SEREX does not rely on tumor-specific T-cell lines ***(27)***. It was shown for several SEREX-defined antigens, e.g., the cancer-testis (CT) antigen NY-ESO-1, that spontaneous CD8 T-cell responses correlate with a spontaneous humoral immune response in patients with antigen-positive tumors. The repertoire of antibody-defined tumor antigens correlates tightly with the CD4 T-cell antigen repertoire of an individual patient. These antigens can potentially be recognized by CD8 T-cells ***(21,28)***. SEREX has been used to identify new tumor-associated antigens in different tumor types, which led to the identification of HOM-MEL-40, a gene identical to the synovial sarcoma/X breakpoint 2 gene (SSX2) involved in the t(x:18) translocation in synovial sarcoma ***(29)***, as well as other CT antigens, including NY-ESO-1 ***(30)***, CT7/MAGE-C1 ***(31)***, SCP-1 ***(32)***, cTAGE-1 ***(33)***, OY-TES-1 ***(34)***, HOM-TES-85 ***(35)***, CAGE ***(36)***, cTAGE ***(33)***, and recently NY-SAR-35 ***(37)***. We were successful using SEREX to identify new differentiation antigens, such as RAB38 ***(38)*** and NY-BR-1 ***(39)***, as well as overexpressed antigens, such as tumor protein D52, NY-BR-62, and NY-BR-85. SEREX represents a very useful technique to define the individual CD4 T-cell tumor antigen repertoire. Today, more than 1200 antigens have been identified by SEREX and are deposited in the SEREX database of the Ludwig Institute for Cancer Research (http:// www. licr.org/SEREX.html).

3. Critical Steps in Antigen Validation

Several steps of analysis are necessary for SEREX-defined antigens to represent potential target antigens for active immunotherapy strategies in cancer patients. In the first step, a careful expression analysis for each individual antigen has to be performed by comparing the cDNA sequence to expressed sequence tag (EST) databases. Antigens with restricted expression by database analysis are being tested by reverse transcription (RT)-polymerase chain reaction by using a panel of normal tissue and tumor tissue cDNAs to confirm the restricted mRNA expression pattern.

Antigens that are tumor- or tissue-restricted expressed undergo a serological analysis to define the frequency of spontaneous humoral immune responses in sera derived from cancer patients as well as healthy individuals. A limited number of sera can be tested using the phage-plaque assay, for larger scale serology, the recombinant protein has to be produced and tested in Western blot or ELISA assays.

Antigens with tumor- or tissue-specific expression and detectable humoral immune responses in tumor patients are being further analyzed for T-cell recognition. We are following an approach called "reverse immunology" where we test the protein sequence of the respective antigen for potential MHC class I binding sequences (http://www2.licr.org/, CancerImmunomeDB, SYFPEITHI database of MHC ligands, and peptide motifs at http://www.bmi-heidelberg.com/ syfpeithi/). We focus on HLA-A2 epitopes because HLA-A2 represents the most frequent MHC class I allele in Western European populations. Peptide-specific CTLs can be generated upon in vitro stimulation, and recognition of target cells expressing the antigenic protein will be documented. CD8 T-cell reactivity to the identified epitope can be analyzed in peripheral blood and tumor lesions of a large series of HLA-A2–positive patients with antigen-positive tumors.

Tumor antigens that fall into the categories of CT antigens or differentiation antigens, which represent target antigens for CD8 T-cells, can be used to immunize HLA-A2–positive patients with antigen-positive tumors. The peptides that are recognized by CD8 T-cells will be used for intradermal injection. The readout includes delayed-type hypersensitivity reaction, induction of peptide-specific CD8 T-cells (ELISPOT and Multimer), and clinical development of the tumor disease.

4. A New Breast Cancer Differentiation Antigen, NY-BR-1

In a recent autologous SEREX screening of a breast cancer cDNA expression library, we identified a previously unknown antigen designated as NY-BR-1. About half of the clones identified in the screening were derived from NY-BR-1, whereas the other clones were derived from distinct known and unknown genes. Comparison with EST databases showed only some cDNA sequences matching the NY-BR-1 cDNA. Interestingly, some sequences were derived from testis; one sequence was derived from normal breast. All other antigens identified in this screening matched to cDNA sequences derived from a wide array of normal tissues, and they were not further analyzed.

Following a combined approach of rapid amplification of cDNA ends-PCR and database analysis for overlapping cDNA sequences to complete the 5′ cDNA sequence, we were able to clone the full-length NY-BR-1 cDNA, which is approx 6 kilobases, coding for a protein of approx 160 kDA. RT-PCR in a normal tissue

panel by using NY-BR-1–specific primers showed NY-BR-1 mRNA to be exclusively expressed in normal breast, normal testis, and variable in normal prostate. In tumor tissues, NY-BR-1 is expressed in 70–80% of breast cancers, and 25% of prostate cancers; all other tumor types are NY-BR-1 negative ***(39)***.

Based on the expression analysis, NY-BR-1 represents a new breast differentiation antigen. We produced a short carboxy-terminal recombinant protein for larger scale serological analysis and found humoral immune responses in approx 10% of patients with NY-BR-1–positive breast cancers. Importantly, all normal sera were antibody negative. To analyze T-cell responses against NY-BR-1, the entire NY-BR-1 protein sequence was screened for potential HLA-A2 binding motifs. Numerous potential binding peptides were synthesized, and CD8 T-cells derived from seropositive HLA-A2–positive breast cancer patients were stimulated once with each of the peptides and then tested in ELISPOT assay for peptide-specific recognition. We identified several peptides that were specifically recognized by CD8 T-cells. We could further demonstrate that those peptides represent naturally processed and presented NY-BR-1 epitopes. In summary, we (1) discovered a new tumor antigen by applying SEREX, (2) demonstrated a tumor-specific expression pattern, and (3) identified NY-BR-1–derived peptide epitopes that are naturally processed and presented to CD8 T-cells. A phase I clinical trial for HLA-A2–positive patients with advanced NY-BR-1–expressing breast cancer is planned. Vaccination with the two NY-BR-1–derived peptides will show whether potent antigen-specific CD8 T-cell responses can be induced. Clinical effects and potential side effects in normal breast tissue will be carefully analyzed. Vaccine strategies aiming at the induction of an integrated immune response (antigen-specific antibody, CD4, and CD8 cells) use longer peptide sequences or the entire protein. Such vaccine approaches been have shown to be efficient in early clinical trials in melanoma patients, and they are currently being tested in phase III trials in high-risk melanoma patients after resection of lymph node metastases.

5. Conclusions

SEREX is one of several techniques that identify tumor-associated antigens that have potential use as target antigens for immunotherapy approaches. In the future, we aim at immunizing patients with a whole array of different tumor-associated antigens expressed by the individual tumor to induce integrated immune responses targeting a maximal percentage of cancer cells.

References

1. Naito, Y., Saito, K., Shiiba, K., et al. (1998) CD8+ T cells infiltrated within cancer cell nests as a prognostic factor in human colorectal cancer. *Cancer Res.* **58,** 3491–3494.

2. Nakano, O., Sato, M., Naito, Y., et al. (2001) Proliferative activity of intratumoral CD8(+) T-lymphocytes as a prognostic factor in human renal cell carcinoma: clinicopathologic demonstration of antitumor immunity. *Cancer Res*. **61,** 5132–5136.
3. Schumacher, K., Haensch, W., Roefzaad, C., and Schlag, P. M. (2001) Prognostic significance of activated CD8(+) T cell infiltrations within esophageal carcinomas. *Cancer Res.* **61,** 3932–3936.
4. Eerola, A. K., Soini, Y., and Paakko, P. (2000) A high number of tumor-infiltrating lymphocytes are associated with a small tumor size, low tumor stage, and a favorable prognosis in operated small cell lung carcinoma. *Clin. Cancer Res.* **6,** 1875–1881.
5. van der Bruggen, P., Traversari, C., and Chomez, P. (1991) A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* **254,** 1643–1647.
6. Boel, P., Wildmann, C., Sensi, M. L., et al. (1995) BAGE: a new gene encoding an antigen recognized on human melanomas by cytolytic T lymphocytes. *Immunity* **2,** 167–175.
7. Van den Eynde, B., Peeters, O., De Backer, O., Gaugler, B., Lucas, S., and Boon, T. (1995) A new family of genes coding for an antigen recognized by autologous cytolytic T lymphocytes on a human melanoma. *J. Exp. Med*. **182,** 689–698.
8. Boon, T. and van der Bruggen, P. (1996) Human tumor antigens recognized by T lymphocytes. *J. Exp. Med.* **183,** 725–729.
9. Scanlan, M. J., Gure, A. O., Jungbluth, A. A., Old, L. J., and Chen, Y. T. (2002) Cancer/testis antigens: an expanding family of targets for cancer immunotherapy. *Immunol. Rev.* **188,** 22–32.
10. Coulie, P. G., Brichard, V., Van Pel, A., et al. (1994) A new gene coding for a differentiation antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas. *J. Exp. Med*. **180,** 35–42.
11. Brichard, V., Van Pel, A., Wolfel, T., et al. (1993) The tyrosinase gene codes for an antigen recognized by autologous cytolytic T lymphocytes on HLA-A2 melanomas. *J. Exp. Med*. **178,** 489–495.
12. Kawakami, Y., Eliyahu, S., Delgado, C. H., et al. (1994) Identification of a human melanoma antigen recognized by tumor- infiltrating lymphocytes associated with in vivo tumor rejection. *Proc. Natl. Acad. Sci. USA* **91,** 6458–6462.
13. Wang, R. F., Robbins, P. F., Kawakami, Y., Kang, X. Q., and Rosenberg, S. A. (1995) Identification of a gene encoding a melanoma tumor antigen recognized by HLA-A31-restricted tumor-infiltrating lymphocytes. [published erratum appears in *J. Exp. Med*. (1995) Mar 1;181(3),1261]. *J. Exp. Med.* **181,** 799–804.
14. Wolfel, T., Hauer, M., Schneider, J., et al. (1995) A p16INK4a-insensitive CDK4 mutant targeted by cytolytic T lymphocytes in a human melanoma. *Science* **269,** 1281–1284.
15. Gnjatic, S., Cai, Z., Viguier, M., Chouaib, S., Guillet, J. G., and Choppin, J. (1998) Accumulation of the p53 protein allows recognition by human CTL of a wild-type p53 epitope presented by breast carcinomas and melanomas. *J. Immunol*. **160,** 328–333.

16. Jochmus, I., Osen, W., Altmann, A., et al. (1997) Specificity of human cytotoxic T lymphocytes induced by a human papillomavirus type 16 E7-derived peptide. *J. Gen. Virol*. **78(7),** 1689–1695.
17. Rammensee, H. G., Weinschenk, T., Gouttefangeas, C., and Stevanovic, S. (2002) Towards patient-specific tumor antigen selection for vaccination. *Immunol. Rev.* **188,** 164–176.
18. Jager, E., Jager, D., and Knuth, A. (2002) Clinical cancer vaccine trials. *Curr. Opin. Immunol.* **14,** 178–182.
19. Jager, E., Nagata, Y., Gnjatic, S., et al. (2000) Monitoring CD8 T cell responses to NY-ESO-1: correlation of humoral and cellular immune responses. *Proc. Natl. Acad. Sci. USA* **97,** 4760–4765.
20. Jaeger, E., Bernhard, H., Romero, P., et al. (1996) Generation of cytotoxic T-cell responses with synthetic melanoma-associated peptides in vivo: implications for tumor vaccines with melanoma-associated antigens. *Int. J. Cancer* **66,** 162–169.
21. Jager, E., Chen, Y. T., Drijfhout, J. W., et al. (1998) A simultaneous humoral and cellular immune response against cancer-testis antigen NY-ESO-1: definition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes. *J. Exp. Med*. **187,** 265–270.
22. Jager, E., Gnjatic, S., Nagata, Y., et al. (2000) Induction of primary NY-ESO-1 immunity: CD8+ T lymphocyte and antibody responses in peptide-vaccinated patients with NY-ESO-1+ cancers. *Proc. Natl. Acad. Sci. USA* **97,** 12,198–12,203.
23. Jager, E., Ringhoffer, M., Dienes, H. P., et al. (1996) Granulocyte-macrophage-colony-stimulating factor enhances immune responses to melanoma-associated peptides in vivo. *Int. J. Cancer* **67,** 54–62.
24. Rosenberg, S. A., Yang, J. C., Schwartzentruber, D. J., et al. (1998) Immunologic and therapeutic evaluation of a synthetic peptide vaccine for the treatment of patients with metastatic melanoma. *Nat. Med.* **4,** 321–327.
25. Marchand, M., van Baren, N., Weynants, P., et al. (1999) Tumor regressions observed in patients with metastatic melanoma treated with an antigenic peptide encoded by gene MAGE-3 and presented by HLA-A1. *Int. J. Cancer* **80,** 219–230.
26. Sahin, U., Tureci, O., Schmitt, H., et al. (1995) Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc. Natl. Acad. Sci. USA* **92,** 11,810–11,813.
27. Scanlan, M. J., Gout, I., Gordon, C.M., et al. (2001) Humoral immunity to human breast cancer: antigen definition and quantitative analysis of mRNA expression. *Cancer Immunity* **1,** 4.
28. Ayyoub, M., Stevanovic, S., Sahin, U., et al. (2002) Proteasome-assisted identification of a SSX-2-derived epitope recognized by tumor-reactive CTL infiltrating metastatic melanoma. *J. Immunol*. **168,** 1717–1722.
29. Tureci, O., Chen, Y. T., Sahin, U., et al. (1998) Expression of SSX genes in human tumors. *Int. J. Cancer* **77,** 19–23.
30. Chen, Y. T., Scanlan, M. J., Sahin, U., et al. (1997) A testicular antigen aberrantly expressed in human cancers detected by autologous antibody screening. *Proc. Natl. Acad. Sci. USA* **94,** 1914–1918.

31. Chen, Y. T., Gure, A. O., Tsang, S., et al. (1998) Identification of multiple cancer/testis antigens by allogeneic antibody screening of a melanoma cell line library. *Proc. Natl. Acad. Sci. USA* **95,** 6919–6923.
32. Tureci, O., Sahin, U., Zwick, C., Koslowski, M., Seitz, G., and Pfreundschuh, M. (1998) Identification of a meiosis-specific protein as a member of the class of cancer/testis antigens. *Proc. Natl. Acad. Sci. USA* **95,** 5211–5216.
33. Eichmuller, S., Usener, D., Dummer, R., Stein, A., Thiel, D., and Schadendorf, D. (2001) Serological detection of cutaneous T-cell lymphoma-associated antigens. *Proc. Natl. Acad. Sci. USA* **98,** 629–634.
34. Ono, T., Kurashige, T., Harada, N., et al. (2001) Identification of proacrosin binding protein sp32 precursor as a human cancer/testis antigen. *Proc. Natl. Acad. Sci. USA* **98,** 3282–3287.
35. Tureci, O., Sahin, U., Koslowski, M., et al. (2002) A novel tumour associated leucine zipper protein targeting to sites of gene transcription and splicing. *Oncogene* **21,** 3879–3888.
36. Cho, B., Lim, Y., Lee, D. Y., et al. (2002) Identification and characterization of a novel cancer/testis antigen gene CAGE. *Biochem. Biophys. Res. Commun.* **292,** 715–726.
37. Lee, S. Y., Obata, Y., Yoshida, M., et al. (2003) Immunomic analysis of human sarcoma. *Proc. Natl. Acad. Sci. USA* **100,** 2651–2656.
38. Jager, D., Stockert, E., Jager, E., et al. (2000) Serological cloning of a melanocyte rab guanosine 5′-triphosphate-binding protein and a chromosome condensation protein from a melanoma complementary DNA library. *Cancer Res*. **60,** 3584–3591.
39. Jager, D., Stockert, E., Gure, A. O., et al. (2001) Identification of a tissue-specific putative transcription factor in breast tissue by serological screening of a breast cancer library. *Cancer Res*. **61,** 2055–2061.

16

Identification of Tumor Antigens by Using Proteomics

François Le Naour

Summary

The recent progress of proteomics has opened new avenues for tumor-associated antigen discovery. Here, I describe a two-dimensional (2D), gel-based Western blot approach for screening and identification of proteins eliciting a humoral response in cancer. Sera from patients are used in 2D Western blot experiments for screening of autoantibodies, and the immunoreactive target proteins are subsequently identified by mass spectrometry. Applied to several types of cancer, this proteomic-based approach has revealed a high frequency of autoantibodies in sera from cancer patients and has led to the identification of novel tumor antigens. Relevant examples are described.

Key Words: Autoantibodies; mass spectrometry; proteomics; two-dimensional (2D) Western blot; tumor antigens.

1. Introduction

Cancer immunology has a long history since it has been worded more than a century ago that tumor cells could be considered as "nonself" ***(1)***. This basic concept, suggesting that the immune system could be able to recognize tumor cells and then destroy them, has been strengthened by the observation of spontaneous tumor regression in some patients with different types of cancer. To develop immunotherapeutic intervention, also called immunotherapy or cancer vaccine, in cancer patients, the field of cancer immunology has been dominated by the search of tumor-specific and tumor-associated antigens (TAAs) as well as by the characterization of the immune cells involved in tumor recognition by the autologous host. It has now been demonstrated that cancer may elicit autologous immune response involving CD8+ cytotoxic T-lymphocytes (CTLs), CD4+ T-helper cells as well as antibodies ***(2–4)***. Tumor rejection is critically

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarket Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

dependent on CD8+ cytotoxic T-lymphocytes. Thus, attention has been focused on tumor-specific CD8+ T-cells. Since the cloning of MAGE-1, the first reported antigen to encode a human tumor antigen recognized by autologous CTLs ***(5)***, an array of autoimmunogenic tumor antigens have been identified ***(6)***. Although the characterization of tumor antigen recognized by cytotoxic T-cells is crucial, a limitation stems from this approach requiring the establishment of autologous target cells in culture and the isolation of stable lines of autologous CTLs that recognize antigens. In addition, some epithelial cell types are poorly susceptible to CTLs in vitro. In addition, the B-cell response (antibodies) also has been demonstrated in cancer patients. Recently, a method called serological analysis of recombinant tumor cDNA expression libraries (SEREX) has been used for the identification of tumor antigens ***(7)***. SEREX is based on screening of autoantibodies in sera from patients with cancer against an expression library made with mRNAs from the autologous tumor (*see* Chapter 15, Volume 1). By applying this strategy to human tumors of different origin, an unexpected frequency of tumor antigens that elicit specific immune responses in the autologous host has been demonstrated ***(7–10)***. However, SEREX is limited by the necessity to construct expression libraries, and the analysis is usually restricted to one or a few patients ***(7–11)***.

The recent progress of proteomics, allowing large-scale analysis of proteins within a single experiment, has opened new avenues for TAA discovery. This chapter focuses the utility of proteomics for identification of tumor-associated antigens with emphasis on proteins eliciting humoral response in cancer patients.

2. Proteomics: The State of the Art

Proteomics designates the study of the proteome. It involves a variety of approaches combining high-resolution separation techniques with identification methods such as mass spectrometry. For many years, two-dimensional (2D) polyacrylamide gel electrophoresis (PAGE) has been the primary technique for proteomic-based biomarker discovery. Thus, 2D electrophoresis (2DE) allows the separation of the proteins according to their charge in a first dimension where they stop their migration after reaching their isoelectric point (p*I*) and then in a second dimension according to their molecular mass. The proteins can be visualized by different staining, such as Coomassie blue or silver, and digitalized. 2DE is a highly resolutive technique leading to the visualization of at least 1000 protein spots in one experiment. Therefore, 2DE allows comparative analysis of protein expression levels and the visualization of various protein isoforms. The identification of the proteins can be performed after enzymatic digestion of the proteins in-gel followed by analysis of the resulting peptides by mass spectrometry. This process can be performed by mass fingerprinting, usually using matrix-assisted laser desorption ionization/time of flight mass spectrometry or

by "shot gun" analysis by using tandem mass spectrometry (MS/MS). In both methods, spectra are used for database searches, leading to protein identification. It should be noted that MS/MS mass spectrometry allows the identification of a protein from a single peptide. Indeed, after fragmenting that peptide, the second spectrum (MS/MS or MS^2) is used for determining its sequence or a sequence tag. Recent advances in proteomics have been realized by coupling liquid chromatography with mass spectrometry (LC-MS/MS). LC-MS/MS is a powerful tool for analyzing the composition of complex mixtures ***(12)***.

3. 2D Gel-Based Western Blot Approach

3.1. Strategy

A proteomic-based approach, using 2DE followed by Western blot experiments with sera from cancer patients, has enabled the identification of tumor-associated antigens that elicit a humoral response. The methodology to identify proteins eliciting humoral response in cancer patients involves two essential steps: first, the screening of autoantigens after serological analysis; and second, the identification of the target molecule. The use of 2D-PAGE allows separating simultaneously several thousand individual cellular proteins from tumor tissue or tumor cell lines. Separated proteins are transferred onto membranes. Sera from cancer patients obtained at time of diagnosis are used as primary antibody for Western blot analysis and are screened individually for antibodies that react against separated proteins. The comparison of series of Western blot obtained with sera from cancer patients with those performed with sera from healthy individuals led to determination of the occurrence of relevant autoantigens. An important advantage of this methodology is that isoforms can be separately analyzed. Notably, some modifications, such as glycosylation, can be immunogenic.

For identification, the recurrent protein spots that specifically react with sera from cancer patients are located on silver-stained 2D gels after superimposition with the blot. The proteins of interest are extracted from the gel and identified by mass spectrometric analysis.

To confirm the identification of TAAs, a validation step is usually necessary. This step can be performed using classical methods such as Western blot experiments when antibodies are available. The relevance to cancer in terms of expression and location may be investigated using ELISA, immunohistochemistry, or by other techniques **(Fig. 1)**.

3.2. Practical and Technical Aspects

2D gels are performed with 100–200 µg of proteins from whole cell extracts of cell lines or biopsies (*see* **Subheading 6.**). The first dimension can be realized using different pH gradients on different sizes of strips. A cocktail of 7 *M* urea, 2 *M* thiourea, and 2% CHAPS can be used as a regular buffer for

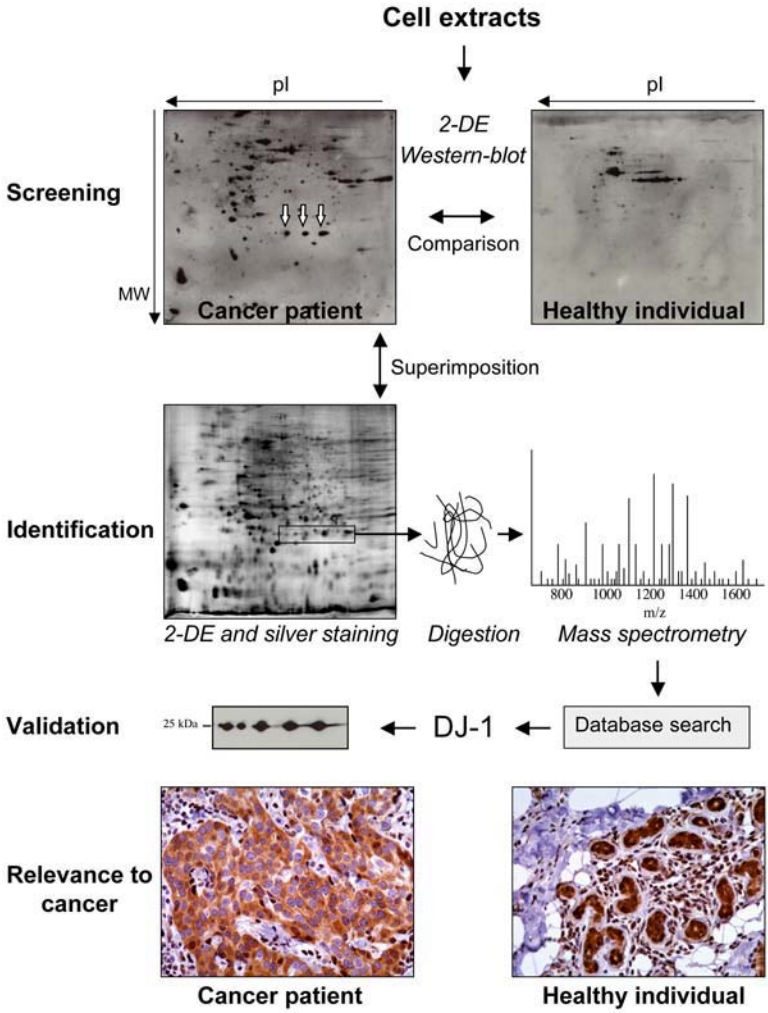


Fig. 1. To illustrate the proteome-based approach for targeting autoantigens in cancer, an example is given from a study on breast cancer ***(15)***. The proteins from a cancer cell line (SUM-44) were solubilized and separated by two-dimensional electrophoresis (2DE). Proteins were transferred onto Immobilon-P polyvinylidene difluoride (PVDF) membranes. The screening of autoantigens was performed by Western blot using sera from cancer patients or healthy individuals as the primary antibody. The comparison of the patterns allowed visualization of specific protein spots in cancer. In our example, three protein spots were specifically reactive (white arrows). For identification, the protein spots were located on the gel after staining of the proteins and superimposition with the blot. The proteins of interest were cut out and in-gel digested. The resulting peptides were analyzed by mass spectrometry. The spectra were

isoelectric focalization in reducing conditions. The second dimension corresponds to a regular sodium dodecyl sulfate-PAGE in reducing conditions. However, a gel gradient is indicated for optimal separations. For better reproducibility of 2D patterns, it is recommended to migrate several 2D gels at a time. After electrophoresis, proteins can be stained using Coomassie blue or silver staining compatible with mass spectrometry analysis ***(13)***, or transferred to polyvinylidene fluoride (PVDF) membrane. To monitor the transfer efficiency and the amount of proteins before hybridization, a rapid staining can be performed using red ponceau. This dye is water soluble and compatible with following Western blotting. The hybridization step is realized with serum from patient at a dilution 1/100–1/300. If background is generated, a lower dilution can be used with subsequent extensive washes. For protein identification, trypsin digestion is performed as described previously ***(14,15)***. After mass spectrometry, spectra are used for database search by using programs MS-Fit, Profound, or Mascott (http://prospector.ucsf.edu; www.hgmp.mrc.ac.uk/GenomeWeb/prot-fragment.html, or www.matrixscience.com).

3.3. Relevant Examples

A study using the 2D gel-based Western blot approach has been used for tumor antigen identification in breast cancer ***(15)***. This approach is described in **Fig. 1** as an example. In this study, sera from patients with breast cancer and healthy individuals were used. The breast cancer cell line SUM-44 was used as a source of tumor cell proteins for 2D-PAGE and for Western blot analysis. A restricted reactivity against a set of three proteins with an estimated molecular mass of 25 kDa was observed in 13% of the patients. These protein spots were analyzed using mass spectrometry and were identified as isoforms of the protein DJ-1.

Other interesting examples can be cited. Indeed, sera from patients with lung adenocarcinomas exhibited a high frequency (60%) and specificity of autoantibodies directed against annexins I and II ***(16)***. A high frequency of autoantibodies also has been observed in sera from 70% patients with hepatocellular carcinoma (HCC) ***(17)***. However, by contrast with lung adenocarcinoma, reactivity against 13 protein spots has been observed. This observation may reflect the

---

Fig. 1. *(Continued)* then used for database search and protein identification. The three protein spots were identified as the protein DJ-1. A validation of the identification has been performed using Western blotting with specific monoclonal antibodies. Finally, a relevance to cancer has definitely been established after immunochemistry experiments that led to the observation of differential expression level and location of the protein between breast cancer and the healthy tissue.

heterogeneity of HCC, which is a multistage process with the involvement of a multifactorial etiology and many gene–environment interactions.

Finally, the proteomic-based approach, also called serological proteome analysis (SERPA) ***(18)*** or proteome-based target evaluation combined with immunoreactive target structure identification explored by sera (Proteomex) ***(19)*** has been applied to different types of cancer, such as renal cellular carcinoma; neuroblastoma, lung, pancreatic, and breast cancers; and HCC ***(15–25)***. Several autoantigens have been identified that may have clinical applications in diagnosis as well as in determining prognosis.

3.4. Alternative Proteomic-Based Approaches and Future Directions

The screening and identification of proteins that elicit humoral response in cancer have been performed recently by alternative methods (*see* Chapters 14 and 15, Volume 1). Indeed, a proteomic-based methodology called antibody-mediated identification of antigens (AMIDA) has been set up ***(26)***. The approach is based on immunoprecipitation experiments using antibodies from serum of healthy individuals or patients with cancer. The method has been applied for TAA identification in head and neck cancer, and 12 proteins were identified ***(27)***. Three of the proteins exhibited overexpression in carcinomas by using immunochemistry experiments. In addition, elevated circulating antibodies were observed in serum from patients with cancer ***(27)***. Proteomics is now moving on the analysis of whole proteins in complex mixtures by using gel-free separation techniques. For example, protein biochips (or microarrays) are emerging in different formats as an alternative to 2DE for differential expression analysis at the protein level ***(28)***. Microarrays of tumor cell-derived proteins have uncovered immunogenic proteins and autoantibodies in serum from patients with prostate, colon, and lung cancer (*see* also Chapter 17, Volume 1) ***(28–30)***.

3.5. Conclusions

Proteomics has emerged recently and is making great progress very rapidly. Proteomic-based approaches have already permitted screening and identification of TAAs that were not obtained using classical approaches. These proteins have a potential clinical utility and may serve as novel cancer markers in screening, diagnosis, or prognosis. It is likely that proteomics will play an important role in the discovery of new TAAs and perhaps in the development of personalized medicines.

4. Notes

1. 2D-PAGE is a very high-resolution separation technique.
2. 2D-PAGE is not adapted for proteins having extreme molecular weight or for proteins that are highly hydrophobic (e.g., membrane proteins).

3. 2D-PAGE is usually performed in reducing conditions, which may be an issue for detection of certain antigens for which epitope(s) involving disulfide bonds would be lost.

References

1. Happ, S. S. (1997) *A Commotion in the Blood: Life, Death, and the Immune System.* The Spoon Technology Series. Henry Holt and Company Inc., NY, NY.
2. Jager, E., Chen, Y. T., Drijfhout, J. W., et al. (1998) Simultaneous humoral and cellular immune response against cancer-testis antigen NY-ESO-1: definition of human histocompatibility leukocyte antigen (HLA)-A2-binding peptide epitopes. *J. Exp. Med.* **187,** 265–270.
3. Jager, D., Jager, E., and Knuth, A. (2001) Vaccination for malignant melanoma: recent developments. *Oncology* **60,** 1–7.
4. Nakatsura, T., Senju, S., Ito, M., Nishimura, Y., and Itoh, K. (2002) Cellular and humoral immune responses to a human pancreatic cancer antigen, coactosin-like protein, originally defined by the SEREX method. *Eur. J. Immunol.* **32,** 826–836.
5. van der Bruggen, P., Traversari, C., Chomez, P., et al. (1991) A gene encoding an antigen recognized by cytolytic T lymphocytes on a human melanoma. *Science* **254,** 1643–1647.
6. Renkvist, N., Castelli, C., Robbins, P. F., and Parmiani, G. (2001) A listing of human tumor antigens recognized by T cells. *Cancer Immunol. Immunother.* **50,** 3–15.
7. Sahin, U., Türeci, O., Schmitt, H., et al. (1995) Human neoplasms elicit multiple specific immune responses in the autologous host. *Proc. Natl. Acad. Sci. USA* **92,** 11,810–11,813.
8. Old, L. J. and Chen, Y. T. (1998) New paths in human cancer serology. *J. Exp. Med.* **187,** 1163–1167.
9. Türeci, O., Sahin, U., Zwick, C., Neumann, F., and Pfreundschuh M. (1999) Exploitation of the antibody repertoire of cancer patients for the identification of human tumor antigens. *Hybridoma* **18,** 23–28.
10. Chen, Y. T. (2000) Cancer vaccine: identification of human tumor antigens by SEREX. *Cancer J. Sci. Am.* **6,** s208–s217.
11. Nishikawa, H., Tanida, K., Ikeda, H., et al. (2001) Role of SEREX-defined immunogenic wild-type cellular molecules in the development of tumor-specific immunity. *Proc. Natl. Acad. Sci. USA* **98,** 14,571–14,576.
12. Patterson, S. D. and Aebersold, R. H. (2003) Proteomics: the first decade and beyond. *Nat. Genet.* **33,** *Suppl*, 311–323.
13. Gharahdaghi, F., Weinberg, C. R., Meagher, D. A., Imai, B. S., and Mische, S. M. (1999) Mass spectrometric identification of proteins from silver-stained polyacrylamide gel: a method for the removal of silver ions to enhance sensitivity. *Electrophoresis* **20,** 601–605.
14. Shevchenko, A., Wilm, M., Vorm, O., and Mann, M. (1996) Mass spectrometric sequencing of proteins silver-stained polyacrylamide gels. *Anal. Chem.* **68,** 850–858.

15. Le Naour, F., Misek, D. E., Krause, M. C., et al. (2001) Proteomics-based identification of RS/DJ-1 as a novel circulating tumor antigen in breast cancer. *Clin. Cancer Res.* **11,** 3328–3335.
16. Brichory, F. M., Misek, D. E., Yim, A. M., et al. (2001) An immune response manifested by the common occurrence of annexins I and II autoantibodies and high circulating levels of IL-6 in lung cancer. *Proc. Natl. Acad. Sci. USA* **98,** 9824–9829.
17. Le Naour, F., Brichory, F., Misek, D. E., Bréchot, C., Hanash, S. M., and Beretta L. (2001) A distinct repertoire of autoantibodies in hepatocellular carcinoma identified by proteomic analysis. *Mol. Cell Proteomics* **1,** 197–203.
18. Klade, C. S., Voss, T., Krystek, E., et al. (2001) Identification of tumor antigens in renal cell carcinoma by serological proteome analysis. *Proteomics* **1,** 890–898.
19. Lichtenfels, R., Kellner, R., Bukur, J., et al. (2002) Heat shock protein expression and anti-heat shock protein reactivity in renal cell carcinoma. *Proteomics* **2,** 561–570.
20. Prasannan, L., Misek, D. E., Hinderer, R., Michon, J., Geiger, J. D., and Hanash, S. M. (2000) Identification of β-tubulin isoforms as tumor antigens in neuroblastoma. *Clin. Cancer Res.* **6,** 3949–3956.
21. Brichory, F., Beer D., Le Naour, F., Giordano, T., and Hanash, S. M. (2001) Proteomics-based identification of PGP 9.5 as a tumor antigen that induces a humoral immune response in lung cancer. *Cancer Res.* **61,** 7908–7912.
22. Le Naour, F. (2001) Contribution of proteomics to tumor immunology. *Proteomics* **1,** 1295–1302.
23. Hanash, S., Brichory, F., and Beer, D. (2001) A proteomic approach to the identification of lung cancer markers. *Dis. Markers* **17,** 295–300.
24. Shalhoub, P., Kern, S., Girard, S., and Beretta, L. (2001) Proteomic-based approach for the identification of tumor markers associated with hepatocellular carcinoma. *Dis. Markers* **17,** 217–223.
25. Hong, S. H., Misek, D. E., Wang, H., et al. (2004) An autoantibody-mediated immune response to calreticulin isoforms in pancreatic cancer. *Cancer Res.* **64,** 5504–5510.
26. Rauch, J., Ahlemann, M., Schaffrik, M., et al. (2004) Allogenic antibody-mediated identification of head and neck cancer antigens. *Biochem. Biophys. Res. Commun.* **323,** 156–162.
27. Gires, O., Munz, M., Schaffrik, M., et al. (2004) Profile identification of disease-associated humoral antigens using AMIDA, a novel proteomics-based technology. *Cell Mol. Life Sci.* **61,** 1198–1207.
28. Henash, S. M. (2003) The emerging field of protein microarrays. *Proteomics* **3** (special issue), 2075.
29. Imafuku, Y., Omenn, G. S., and Hanash, S. (2004) Proteomics approaches to identify tumor antigen directed autoantibodies as cancer biomarkers. *Dis. Markers* **20,** 149–153.
30. Qiu, J., Madoz-Gurpide, J., Misek, D. E., et al. (2004) Development of natural protein microarrays for diagnosing cancer based on an antibody response to tumor antigens. *J. Proteome Res.* **3,** 261–267.

17

Protein Arrays

A Versatile Toolbox for Target Identification and Monitoring of Patient Immune Responses

Lina Cekaite, Eivind Hovig, and Mouldy Sioud

Summary

Functional proteomics is a promising technique for the rational identification of novel therapeutic targets and biological markers. The studies of protein–protein interactions have been gained from the development of high-throughput technologies such as the yeast two-hybrid system, protein arrays, phage display, and systematic analysis of interaction maps for the prediction of protein functions. Because antibodies are used extensively as diagnostic and clinical tools, the characterization of their antigen specificity is of prime importance. Indeed, screening protein arrays with sera from patients with either cancer or autoimmune diseases would facilitate the identification of autoantibody signatures that can be used for diagnosis and/or prognosis of patients. The usefulness of multiplexed measurements lies not only in the ability to screen many individual marker candidates but also in evaluating the use of multiple markers in combination. Here, we review the advantage of protein and serum screening of peptides and cDNA repertoires displayed on phages as well as the fabrication of protein microarrays for probing immune responses in patients.

Key Words: Autoimmunity; cancer; phage display; protein microarray.

1. Introduction

1.1. Yeast and Phage Technologies as High-Throughput Proteomics Toolbox

Protein–protein interactions play pivotal roles in various aspects of the structural and functional organization of the cell. Today, the identification of proteins that interact with a protein of interest is a focus of the rapidly growing field of proteomics.

From: *Methods in Molecular Biology, vol. 360, Target Discovery and Validation Reviews and Protocols Volume I, Emerging Strategies for Targets and Biomarker Discovery*
Edited by: M. Sioud © Humana Press Inc., Totowa, NJ

Yeast two-hybrid assays and phage display are classical approaches for systematic protein interaction screening capable of probing numerous interactions. The principle of the yeast two-hybrid technique ***(1)*** is based on chimeric proteins containing parts of GAL4: the GAL4 DNA-binding domain fused to a protein "X" and a GAL4-activating region fused to a protein "Y". If X and Y can form protein–protein complex and reconstitute proximity of the GAL4 domains, transcription of a regulated marker gene occurs. The first genome-wide two-hybrid screen was performed by Bartel and co-workers for the study of protein interactions in bacteriophage T7 ***(2)***. The first genome-wide studies of a eukaryotic organism were performed by Uetz and co-workers and Ito and colleagues in *Saccharomyces cerevisiae* ***(3,4)***. Also, a three-hybrid approach for scanning the proteome for targets of small-molecule kinase inhibitors have been developed ***(5)***.

Another strategy was introduced by G. P. Smith ***(6)***. It uses filamentous M13 bacteriophages. A single binding molecule may be displayed as a protein or peptide fused to the surface of a bacteriophage and identified by sequencing the encoding genome. Originally, phage display technology has been used for selection of ligands for well characterized pure target proteins (e.g., monoclonal antibodies and cell surface receptors) immobilized on a solid surface ***(7–9)***. In this respect, a large number of peptides and single-chain Fv antibodies have been selected for their ability bind various target proteins ***(10,11)***. Furthermore, phage display techniques have been applied in profiling immune responses in patients with autoimmune disease or cancers, and peptides or proteins binding serum antibodies were identified ***(9,12–20)***. Moreover, cancer cell binding peptides were selected from random peptide phage libraries ***(21)***.

An advantage of two-hybrid assays and phage display techniques is that no previous knowledge about the interacting proteins/receptors is necessary for a screen to be performed. Moreover, two-hybrid screening can be done in a colony array format, in which each colony expresses a defined pair of proteins ***(22)***. Recently, we have demonstrated that a high-throughput analysis of the immune response in cancer patients can be performed by phage-based microarray technique ***(18)***.

1.2. Examples of Protein Microarrays

Arrays of proteins have been used for both analysis of protein–protein interactions and biochemical analysis of protein functions. The low volume requirement and the multiplexed detection capability of microarrays enable optimal use of precious clinical samples. In addition, the assays are rapid and amenable to automation, thus large sample sets, as required for biomarker studies, can be processed and analyzed. Also, through the use of optimized detection methods and rigorous quality control, the assays can be very sensitive,

Table 1
Classification of Protein Arrays by Capture/Target Molecules

Capture molecular (immobilized)	Target molecular (hybridized with)	References
Monoclonal antibodies, single-chain Fv antibodies	Serum, cellular lysates, antibodies	***(27,28,54–56)***
*A*tigens (proteins, peptides, enzyme complexes, ribonucleoprotein complexes, DNA)	Serum	***(29–32,57)***
Protein fraction from chromatographic separation (that may be further identified with mass spectrometry) (reverse phase arrays)	Serum	***(33,34)***
Yeast two-hybrid system (colony arrays)	Binding constructs	***(3–5,58)***
Recombinant proteins	Serum, antibodies, kinases	***(37–39,41,59,60)***
Random peptides/cDNA libraries displayed on phage	Serum, antibodies	***(11,13–16,18–20, 43,45,61–66)***
Ribosome display	Serum, antibodies	***(48,49)***

reproducible, and quantitative ***(23–27)***. Together, these experimental features provide the technology with great potential to have a significant impact on cancer research.

A protein microarray may consist of antibodies, proteins, protein fragments, peptides, aptamers, or carbohydrate elements that are coated or immobilized in a grid-like pattern on small surfaces. The arrayed molecules are then used to screen and assess patterns of interaction with samples containing distinct proteins or classes of proteins. There is great variation in use of the protein microarray technique, when considering selection of the capture molecules. Aspects of protein coupling to surfaces, sensitivity, dynamic range of detection techniques, and standardization of data remain to be solved.

One of the critical factors for construction of protein microarrays is the choice of captures molecules that could optimally find proteins/antibodies in a complex sample. A diverse set of capture molecules have been used for screening of serum/antibodies **(Table 1)**. Notably, screening protein arrays with sera from patients with either cancer or autoimmune diseases would facilitate the identification of potentially new autoantigens that could be valuable for diagnosis and classification of patients. Here, a summary is provided of

different approaches priming humoral immune responses by a protein microarray technique.

1.2.1. Protein/Peptide–Antibody Interaction

Miller and colleagues ***(28)***. Constructed an antibody microarray containing 184 different antibodies. Forty of the antibodies targeted 32 unique proteins that are typically found in the serum of healthy adults, another 13 antibodies targeted nine proteins that have been detected in the serum of cancer patients, and the rest of the antibodies targeted normal intracellular proteins. The serum samples from patients with prostate cancer and healthy control serum samples were screened using this autoantigen array, and specific reactivity profile was detected in patients. Indeed, five proteins (von Willebrand factor, IgM, α1-antichymotrypsin, villin, and IgG) exhibited different reactivity levels between the prostate cancer samples and the controls.

Robinson and colleagues fabricated antigen microarrays containing 196 distinct biomolecules, including proteins, peptides, and enzyme complexes; ribonucleoprotein complexes; and DNA and posttranslationally modified antigens. Using these arrays, they demonstrated sensitive and specific detection of autoantibodies in serum derived from patients with various autoimmune diseases. No autoantigen reactivity was demonstrated in a healthy individual ***(29)***. Quintana and co-workers used protein microarrays consisting of 266 different antigens to confirm that the future response of mice to induced diabetes could be predicted by autoantibody repertoires ***(30)***. An investigation of antiviral antibody responses to vaccine trials with a simian–human immunodeficiency virus (SHIV), a model for HIV, was performed by Neuman de Vegvar and colleagues ***(31)***. They produced antigen microarrays with 430 different proteins and overlapping peptides spanning the whole SHIV proteome and identified eight immunodominant epitopes. Joos and colleagues created a microarray-based immunoassay containing 18 known autoantigens, which are used as serological markers in predominant autoimmune diseases ***(32)***.

Together, these studies demonstrate the feasibility of using antibody–antigen–protein microarrays for large-scale, high-throughput screening studies. Antibodies–antigens are one of the most prominent capture agents with high affinity and specificity to target molecules. However, these types of capture molecules should be linked to the investigate diseases before use in high throughput screening; therefore, uncharacterized proteins will be absent, limiting this approach as a discovery tool.

1.2.2. Reverse Phase Microarray

Other microarray formats have been designed to detect and quantify antibodies in human serum. Madoz-Gurpide and co-workers created microarrays of protein

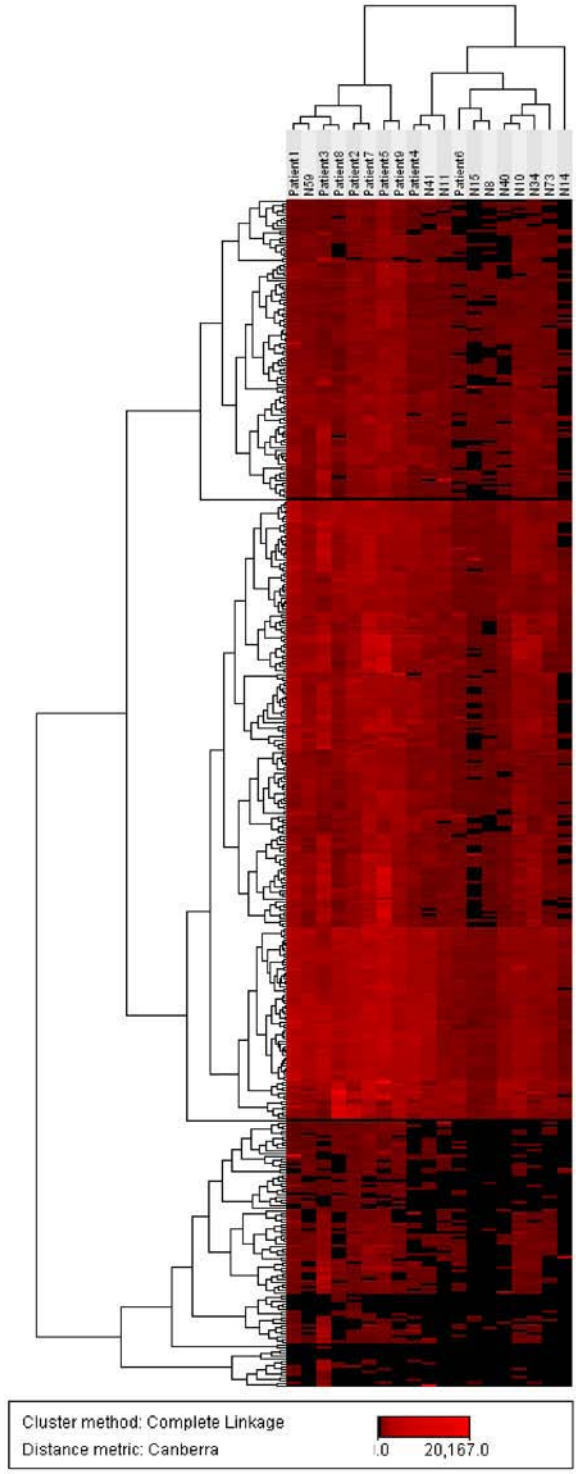
fractions from chromatographic separations of complex protein mixtures ***(33)***. The authors proposed arraying fractions from high-resolution multidimensional chromatography separations, which serially combine different separation phases such as reverse phase and ion exchange chromatography, producing as many as 2000 fractions from a single protein pool. Twenty protein fractions from the A549 lung cancer cell line were separated by isoelectric focusing and arrayed. Solubilized proteins from the LoVo colon adenocarcinoma cell line were separated into fractions, arrayed onto nitrocellulose-coated slides, and hybridized with individual sera from patients with colon cancer, with lung cancer, and healthy subjects ***(34)***. A distinct pattern of reactivity was observed with sera from colon cancer relative to lung cancer. One fraction that exhibited reactivity with 9 of 15 colon cancer sera was subjected to mass spectrometry, leading to the identification of ubiquitin C-terminal hydrolase isoenzyme 3 as a potential autoantigen. Notably, reverse phase microarray has the improvement that novel cancer antigens may be identified.

1.2.3. Peptide-on-Plasmid Display

The advantage of combination of large cDNA expression libraries with microarray technology is the direct connection of the DNA sequence information from a particular clone to its recombinant, expressed protein. For the generation of protein biochips, high-throughput subcloning of open reading frames from the genome of humans, *Saccharomyces cerevisiae* ***(35,36)***, *Arabidopsis thaliana* ***(37,38)***, or *Caenorhabditis elegans* ***(39)*** was performed. This cloning approach is strongly dependent on the progress in genome sequencing projects and the annotation of the human sequences ***(40)***. Thus, previously uncharacterized proteins will be absent. Additionally, a clear determination of the expressed sequence remains difficult because of differential splicing or posttranslational modifications. For these reasons, this approach has proved most valuable in the production of arrays containing proteins from well characterized organisms. However, the use of cDNA expression libraries eliminates the need to construct individual expression clones for every protein of interest and allows identification of unknown sequences/proteins. In this respect, Büssow and colleagues described a human fetal brain expression library ***(41)***. Recently, a mouse TH1 expression cDNA library was constructed ***(42)***.

1.2.4. Phage-Displayed Peptide/Recombinant Proteins

Peptide and cDNA phage display libraries also have been used to determine the specificities of antibodies present in the whole sera of patients where information about the parental antigens is unknown. This phage display-based approach (termed “fingerprinting”) enables isolation of mimic peptides of antigens eliciting a humeral response ***(19,20,43)***. Antibody fingerprinting is a


Patient1
N59
Patient3
Patient8
Patient2
Patient7
Patient5
Patient9
Patient4
N41
N11
Patient6
N15
N8
N40
N10
N34
N73
N14
Cluster method: Complete Linkage
Distance metric: Canberra
.0
20,167.0

combinatorial screening in which phage display random peptide libraries are screened on patient antibodies; when successful, such approach can lead to epitope mapping as well as the identification of the corresponding native tumor antigens. Because cancer cells are expected to express several different antigens, fingerprinting might yield markers of disease aggressiveness or targets for therapy.

Hansen and colleagues identified several consensus peptide motifs that mimic some potential tumor antigens, such as nuclear autoantigen sp100 and an uncharacterized 66-kDa protein, which has recently been identified as dihydrolipoamide *S*-acetyltransferase. Interestingly, serum reactivity to the identified peptides correlated with patient prognosis. Recently, Vidal and colleagues identified a consensus peptide motif that is selectively recognized by antibodies derived from ascites from ovarian cancer patients and identified the corresponding native antigen mimicked by the motif: heat-shock protein of 90 kDa (HSP90). Furthermore, they showed that although HSP90 is widely expressed in ovarian cancer, the humoral response against HSP90 is tumor associated and stage specific ***(44)***.

In addition to peptides, the cDNA repertoires from cancer cells can be displayed on phage and screened with patient sera ***(15,19,20,45)***. Compared with random peptide phage libraries, phage-displaying cDNA repertoires from cells of interest may have an advantage for the identification of the parental antigens, because they represent naturally expressed antigens. In this respect, cDNA repertoires from breast cancer cell lines T47D and MCF-7 were cloned in the three open reading frames as fusion with the M13 pVI protein ***(20)***. When the libraries were screened with serum antibodies from patients with breast cancer, phage-encoded cDNA products were selected. Sequence analysis of the selected clones identified important antigens, including p53, centromere-F, int-2, pentraxin I, integrin β5, cathepsin L2, and S3 ribosomal protein. The selected phage-displayed cDNA products were recognized by a significant number of breast cancer sera compared with sera from normal individuals.

In addition to fingerprinting immune responses in patients, phage display libraries are a valuable tool for the selection of cancer cell binding peptides. In this respect, we have explored the possibility of selecting small peptides that bind specifically to breast cancer cell lines. Phage displaying an LTVSPWY peptide sequence exhibited a specific binding to breast cancer cells. None of the selected peptides bound to human primary cells from different tissue origin (e.g., epithelial, endothelial, and hematopoetic) ***(45)***. In vivo screening of a peptide library in a patient also was performed by Arap and coworkers ***(46,47)***. They selected 47,160 motifs that localized to different organs. This large-scale screening indicates that the tissue distribution of circulating peptides is

Fig. 1. *(Opposite page)* Hierarchical clustering of patient immunoreactivity against phage-displayed peptides ***(23)***.

nonrandom. High-throughput analysis of the motifs revealed similarities to ligands for differentially expressed cell surface proteins, and a candidate ligand–receptor pair was identified and validated. These data represent a step toward the construction of a molecular map of human vasculature, and they may have broad implications for the development of targeted therapies.

Phage display is a well established technology that allows detection and characterization of protein–protein interactions. However, screening of large number of clones is labor-intensive. Previously, we have demonstrated a proof of principle of implementation of phage-displaying libraries with microarrays. The recombinant phages were used for fabricating phage arrays that were further screened against either breast cancer or healthy donor serum antibodies. A significant tumor effect was found with most of the selected phage-displayed peptides, suggesting that recombinant phage microarrays can serve as a tool in monitoring humoral responses toward phage-displayed peptides ***(23)***. As shown in **Fig. 1**, the immunoreactivity clusters of patients and healthy donors are distinguishable, indicating that recombinant phage microarrays can monitor humoral immune response in patients.

1.3. Ribosome Display

Ribosome display is an in vitro protein display system. The concept is to link individual proteins to their genes ("phenotype linked to genotype") and to create libraries of such linkages. Similar to phage display, selection of a protein from the library simultaneously captures its encoding gene, which can then be expressed or manipulated. Key features of ribosome display are that it is a wholly cell-free technology that can generate exceptionally large protein libraries in vitro. The link between protein and gene (mRNA) is made through the ribosome as a protein–ribosome–mRNA complex ***(48,49)***. DNA elements essential for transcription and translation such as T7 promoter sequence, translation initiation site, and immobilization tag sequence are added through PCR assembly of DNA fragments. A coupled in vitro transcription and translation system (such as rabbit reticulocyte lysate) is used to produce individual tagged proteins on a tag-binding surface, which then are screened for binding to other proteins such as enzymes and antibodies.

1.4. Conclusions and Future Challenges

Priming of the humoral arm of the immune system by tumor cells, other cancer-causing agents, or both is printed in patient sera as a specific antibody response. Therefore, probing of such humoral immune responses could provide an effective strategy for cancer screening and therapy. Cancer antigens that may be accessible to immune system at some stage of the disease may be used for identification, classification, choice of therapy, or a combination.

The ideal proteomics-based analytical tool would consist of a microarray containing a large number of high-affinity, high-specificity protein ligands. For humans, this outcome would mean isolating on the order of 100,000 good monoclonal antibodies or the equivalent capture proteins. In reality, the number would be much higher, because the detection of different posttranslationally modified forms of a protein is one of the principal advantages of moving from nucleic acids to protein-based analytical techniques. However, the high cost of monoclonal antibody manufacturing is a bottleneck for large-scale production of antibody microarrays. Thus, it is most likely the field will proceed in a stepwise manner, and a reasonable intermediate goal would be to construct arrays of small amount of protein-binding ligands directed against proteins of interest in a particular disease state. Reverse phase arrays and the aforementioned display techniques allow the rapid and efficient isolation of high-affinity and -specificity protein ligands. Furthermore, the uses of display techniques allow identification of unknown biomarkers. It is increasingly evident that posttranslational modifications of biomolecules could play an essential role in coding for both new and native properties. However, the selection of antigens with modifications is not accessible by peptide-on-plasmid or phage display techniques. Regardless, the establishment of a robust system that may combine the advantages of display strategies continues to be a key challenge. To achieve certain desired display advantages, several different viral systems have been used to display peptides, including lysogenic filamentous phages ***(6)*** and lytic lambda phage ***(50,51)***, T7 bacteriophage ***(52)***, and T4 bacteriophage ***(53)***. Lysogenic filamentous phages remain the most commonly used phage display system. For cases in which displayed proteins may be toxic to filamentous phage assembly or incompatible with the bacterial secretion pathway, lytic phages can be used that allow displayed sequences to minimize negative selection. However, selection of recombinant proteins displayed on T7 phage by arrays seems to be difficult because negative phage immobilized on the surface showed reactivity to patient antibodies as well (Cekaite et al., unpublished data). One way of addressing this problem is to reduce binding valency by including a binding competitor, thereby increasing the binding affinity threshold.

The examples in this chapter show that protein microarray technology is still in its infancy, given the diversity of proposed approaches as well as raised problems. However, this technology is promising tool for priming of the humoral arm of the immune system.

References

1. Fields, S. and Song, O. -K. (1989) A novel genetic system to detect protein Â–protein interactions. **340,** 245–246.
2. Bartel, P. L., Roecklein, J. A., SenGupta, D., and Fields, S. (1996) A protein linkage map of Escherichia coli bacteriophage T7. *Nat Genet*. **12,** 72–77.

3. Uetz, P., Giot, L., Cagney, G., et al. (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae*. *Nature* **403,** 623–627.
4. Ito, T., Chiba, T., Ozawa, R., Yoshida, M., Hattori, M., and Sakaki, Y. (2001) A comprehensive two-hybrid analysis to explore the yeast protein interactome. *Proc. Natl. Acad. Sci. USA* **98,** 4569–4574.
5. Becker, F., Murthi, K., Smith, C. W., et al. (2004) A three-hybrid approach to scanning the proteome for targets of small molecule kinase inhibitors. *Chem. Biol*. **11,** 211–223.
6. Smith, G. P. (1985) Filamentous fusion phage: novel expression vectors that display cloned antigens on the virion surface. *Science* **228,** 1315–1317.
7. Winter, G., Griffiths, A. D., Hawkins, R. E., and Hoogenboom, H. R. (1994) Making antibodies by phage display technology. *Annu. Rev. Immunol*. **12,** 433–455.
8. Sheets, M. D., Amersdorfer, P., Finnern, R., et al. (1998) Efficient construction of a large nonimmune phage antibody library: the production of high-affinity human single-chain antibodies to protein antigens. *Proc. Natl. Acad. Sci. USA* **95,** 6157–6162.
9. Dybwad, A., Lambin, P., Sioud, M., and Zouali, M. (2003) Probing the specificity of human myeloma proteins with a random peptide phage library. *Scand. J. Immunol*. **57,** 583–590.
10. Souriau, C. and Hudson, P. J. (2003) Recombinant antibodies for cancer diagnosis and therapy. *Expert Opin. Biol. Ther.* **3,** 305–318.
11. Moghaddam, A., Borgen, T., Stacy, J., et al. (2003) Identification of scFv antibody fragments that specifically recognise the heroin metabolite 6-monoacetylmorphine but not morphine. *J. Immunol. Methods* **280,** 139–155.
12. Dybwad, A., Forre, O., Kjeldsen-Kragh, J., Natvig, J. B., and Sioud, M. (1993) Identification of new B cell epitopes in the sera of rheumatoid arthritis patients using a random nanopeptide phage library. *Eur. J. Immunol*. **23,** 3189–3193.
13. Hansen, M. H., Dybwad, A., Forre, O., and Sioud, M. (2000) Probing antinuclear antibody specificities by peptide phage display libraries. *Clin. Exp. Rheumatol*. **18,** 465–472.
14. Hansen, M. H., Nielsen, H., and Ditzel, H. J. (2001) The tumor-infiltrating B cell response in medullary breast cancer is oligoclonal and directed against the autoantigen actin exposed on the surface of apoptotic cancer cells. *Proc. Natl. Acad. Sci*. *USA* **98,** 12,659–12,664.
15. Hansen, M. H., Ostenstad, B., and Sioud, M. (2001) Identification of immunogenic antigens using a phage-displayed cDNA library from an invasive ductal breast carcinoma tumour. *Int. J. Oncol.* **19,** 1303–1309.
16. Hansen, M. H., Ostenstad, B., and Sioud, M. (2001) Antigen-specific IgG antibodies in stage IV long-time survival breast cancer patients. *Mol. Med.* **7,** 230–239.
17. Hansen, M. H., Sode, L. L., Hyldig-Nielsen, J. J., and Engberg, J. (1997) Detection of PNA/DNA hybrid molecules by antibody Fab fragments isolated from a phage display library. *J. Immunol. Methods* **203,** 199–207.

18. Cekaite, H. L., Myklebost, O., Aldrin, M., et al. (2004) Analysis of the humoral immune response to immunoselected phage-displayed peptides by a microarray-based method. *Proteomics* **4,** 2572–2582.
19. Sioud, M., Hansen, M., and Dybwad, A. (2000) Profiling the immune responses in patient sera with peptide and cDNA display libraries. *Int. J. Mol. Med.* **6,** 123–128.
20. Sioud, M. and Hansen, M. H. (2001) Profiling the immune response in patients with breast cancer by phage-displayed cDNA libraries. *Eur. J. Immunol.* **31,** 716–725.
21. Barry, M. A., Dower, W. J., and Johnston, S. A. (1996) Toward cell-targeting gene therapy vectors: selection of cell-binding peptides from random peptide-presenting phage libraries. *Nat. Med.* **2,** 299–305.
22. Cagney, G., Uetz, P., and Fields, S. (2000) High-throughput screening for protein-protein interactions using two-hybrid assay. *Methods Enzymol.* **328,** 3–14.
23. Cekaite, H. L., Myklebost, O., Aldrin, M., et al. (2004) Analysis of the humoral immune response to immunoselected phage-displayed peptides by a microarray-based method. *Proteomics* **4,** 2572–2582.
24. Borrebaeck, C. A., Ekstrom, S., Hager, A. C., Nilsson, J., Laurell, T., and Marko-Varga, G. (2001) Protein chips based on recombinant antibody fragments: a highly sensitive approach as detected by mass spectrometry. *BioTechniques* **30,** 1126–1130, 1132.
25. Haab, B. B., Dunham, M. J., and Brown, P. O. (2001) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol*. **2,** RESEARCH0004.
26. Kim, T. E., Park, S. W., Cho, N. Y., et al. (2002) Quantitative measurement of serum allergen-specific IgE on protein chip. *Exp. Mol. Med.* **34,** 152–158.
27. Wingren, C., Steinhauer, C., Ingvarsson, J., Persson, E., Larsson, K., and Borrebaeck, C. A. (2005) Microarrays based on affinity-tagged single-chain Fv antibodies: sensitive detection of analyte in complex proteomes. *Proteomics* **5,** 1281–1291.
28. Miller, J. C., Zhou, H., Kwekel, J., et al. (2003) Antibody microarray profiling of human prostate cancer sera: antibody screening and identification of potential biomarkers. *Proteomics* **3,** 56–63.
29. Robinson, W. H., DiGennaro, C., Hueber, W., et al. (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat. Med.* **8,** 295–301.
30. Quintana, F. J., Hagedorn, P. H., Elizur, G., Merbl, Y., Domany, E., and Cohen, I. R. (2004) Functional immunomics: microarray analysis of IgG autoantibody repertoires predicts the future response of mice to induced diabetes. *Proc. Natl. Acad. Sci. USA* **101 (Suppl. 2),** 14,615–14,621.
31. Neuman de Vegvar, H. E., Amara, R. R., Steinman, L., Utz, P. J., Robinson, H. L., and Robinson, W. H. (2003) Microarray profiling of antibody responses against simian-human immunodeficiency virus: postchallenge convergence of reactivities independent of host histocompatibility type and vaccine regimen. *J. Virol.* **77,** 11,125–11,138.

32. Thomas, O., Joos, M. S., Höpfl, P., et al. (2000) A microarray enzyme-linked immunosorbent assay for autoimmune diagnostics. *Electrophoresis* **21,** 2641–2650.
33. Madoz-Gurpide, J., Wang, H., Misek, D. E., Brichory, F., and Hanash, S. M. (2001) Protein based microarrays: a tool for probing the proteome of cancer cells and tissues. *Proteomics* **1,** 1279–1287.
34. Nam, M. J., Madoz-Gurpide, J., Wang, H., et al. (2003) Molecular profiling of the immune response in colon cancer using protein microarrays: occurrence of autoantibodies to ubiquitin C-terminal hydrolase L3. *Proteomics* **3,** 2108–2115.
35. Zhu, H., Klemic, J. F., Chang, S., et al. (2000) Analysis of yeast protein kinases using protein chips. *Nat. Genet.* **26,** 283–289.
36. Zhu, H., Bilgin, M., Bangham, R., et al. (2001) Global analysis of protein activities using proteome chips. *Science* **293,** 2101–2105.
37. Kersten, B., Feilner, T., Kramer, A., et al. (2003) Generation of *Arabidopsis* protein chips for antibody and serum screening. *Plant Mol. Biol.* **52,** 999–1010.
38. Feilner, T., Hultschig, C., Lee, J., et al. (2005) High-throughput identification of potential *Arabidopsis* MAP kinases substrates. *Mol. Cell Proteomics* **4,** 1558–1568.
39. Walhout, A. J., Sordella, R., Lu, X., et al. (2000) Protein interaction mapping in *C. elegans* using proteins involved in vulval development. *Science* **287,** 116–122.
40. Heyman, J. A., Cornthwaite, J., Foncerrada, L., et al. (1999) Genome-scale cloning and expression of individual open reading frames using topoisomerase I-mediated ligation. *Genome Res.* **9,** 383–392.
41. Bussow, K., Cahill, D., Nietfeld, W., et al. (1998) A method for global protein expression and antibody screening on high-density filters of an arrayed cDNA library. *Nucleic Acids Res.* **26,** 5007–5008.
42. Gutjahr, C., Murphy, D., Lueking, A., et al. (2005) Mouse protein arrays from a TH1 cell cDNA library for antibody screening and serum profiling. *Genomics* **85,** 285–296.
43. Mintz, P. J., Kim, J., Do, K. A., et al. (2003) Fingerprinting the circulating repertoire of antibodies from cancer patients. *Nat. Biotechnol.* **21,** 57–63.
44. Vidal, C. I., Mintz, P. J., Lu, K., et al. (2004) An HSP90-mimic peptide revealed by fingerprinting the pool of antibodies from ovarian cancer patients. *Oncogene* **23,** 8859–8867.
45. Ansuini, H., Cicchini, C., Nicosia, A., Tripodi, M., Cortese, R., and Luzzago, A. (2002) Biotin-tagged cDNA expression libraries displayed on lambda phage: a new tool for the selection of natural protein ligands. *Nucleic Acids Res.* **30,** e78.
46. Shadidi, M. and Sioud, M. (2002) Identification of novel carrier peptides for the specific delivery of therapeutics into cancer cells. *FASEB J.* **17,** 256–258.
47. Arap, W., Kolonin, M. G., Trepel, M., et al. (2002) Steps toward mapping the human vasculature by phage display. *Nat. Med.* **8,** 121–127.
48. He, M. and Taussig, M. J. (2003) DiscernArray technology: a cell-free method for the generation of protein arrays from PCR DNA. *J. Immunol. Methods* **274,** 265–270.

49. Yuko Kawahashi, N. D., Takashima, H., Tsuda, C., et al. (2003) In vitro protein microarrays for detecting protein-protein interactions: application of a new method for fluorescence labeling of proteins. *Proteomics* **3,** 1236–1243.
50. Santini, C., Brennan, D., Mennuni, C., et al. (1998) Efficient display of an HCV cDNA expression library as C-terminal fusion to the capsid protein D of bacteriophage lambda. *J. Mol. Biol.* **282,** 125–135.
51. Sternberg, N. and Hoess, R. H. (1995) Display of peptides and proteins on the surface of bacteriophage lambda. *Proc. Natl. Acad. Sci. USA* **92,** 1609–1613.
52. Danner, S. and Belasco, J. G. (2001) T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proc. Natl. Acad. Sci. USA* **98,** 12,954–12,959.
53. Ren, Z. J., Lewis, G. K., Wingfield, P. T., Locke, E. G., Steven, A. C., and Black, L. W. (1996) Phage display of intact domains at high copy number: a system based on SOC, the small outer capsid protein of bacteriophage T4. *Protein Sci*. **5,** 1833–1843.
54. Steinhauer, C., Wingren, C., Hager, A. C., and Borrebaeck, C. A. (2002) Single framework recombinant antibody fragments designed for protein chip applications. *BioTechniques Suppl*, 38–45.
55. Holt, L. J., Enever, C., de Wildt, R. M., and Tomlinson, I. M. (2000) The use of recombinant antibodies in proteomics. *Curr. Opin. Biotechnol.* **11,** 445–449.
56. de Wildt, R. M., Mundy, C. R., Gorick, B. D., and Tomlinson, I. M. (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat. Biotechnol*. **18,** 989–994.
57. Poetz, O., Ostendorp, R., Brocks, B., et al. (2005) Protein microarrays for antibody profiling: specificity and affinity determination on a chip. *Proteomics* **5,** 2402–2411.
58. Uetz, P. (2002) Two-hybrid arrays. *Curr. Opin. Chem. Biol*. **6,** 57–62.
59. Lueking, A., Horn, M., Eickhoff, H., Bussow, K., Lehrach, H., and Walter, G. (1999) Protein microarrays for gene expression and antibody screening. *Anal. Biochem*. **270,** 103–111.
60. Lueking, A., Possling, A., Huber, O., et al. (2003) A nonredundant human protein chip for antibody screening and serum profiling. *Mol. Cell Proteomics* **2,** 1342–1349.
61. Sioud, M., Dybwad, A., Jespersen, L., Suleyman, S., Natvig, J. B., and Forre, O. (1994) Characterization of naturally occurring autoantibodies against tumour necrosis factor-alpha (TNF-alpha): in vitro function and precise epitope mapping by phage epitope library. *Clin. Exp. Immunol*. **98,** 520–525.
62. Yip, Y. L. and Ward, R. L. (2002) Application of phage display technology to cancer research. *Curr. Pharm. Biotechnol*. **3,** 29–43.
63. Portefaix, J. M., Fanutti, C., Granier, C., et al. (2002) Detection of anti-p53 antibodies by ELISA using p53 synthetic or phage-displayed peptides. *J. Immunol. Methods* **259,** 65–75.
64. Rodi, D. J., Soares, A. S., and Makowski, L. (2002) Quantitative assessment of peptide sequence diversity in M13 combinatorial peptide phage display libraries. *J. Mol. Biol*. **322,** 1039–1052.

65. Lee, K. J., Mao, S., Sun, C., et al. (2002) Phage-display selection of a human single-chain fv antibody highly specific for melanoma and breast cancer cells using a chemoenzymatically synthesized G(M3)-carbohydrate antigen. *J. Am. Chem. Soc.* **124,** 12,439–12,446.
66. Huls, G., Heijnen, I. A., Cuomo, E., et al. (1999) Antitumor immune effector mechanisms recruited by phage display-derived fully human IgG1 and IgA1 monoclonal antibodies. *Cancer Res.* **59,** 5778–5784.

Index