Basic Concepts of Molecular Pathology

Edited by

Philip T. Cagle, MD
Weill Medical College of Cornell University, New York
The Methodist Hospital, Houston, Texas

and

Timothy Craig Allen, MD, JD
University of Texas Health Science Center at Tyler
Tyler, Texas

Springer
To my wife, Kirsten  

Philip T. Cagle

To my wife, Fran; and to Caitlin and Erin.  

Timothy Craig Allen
Preface

The increasing role of molecular pathology in the daily practice of medicine, including the increasingly routine use of targeted molecular therapy in personalized patient care, has created a need for textbooks that bridge the gap between medical practice and basic molecular science. Each book in the *Molecular Pathology Library Series* by Springer Science+Business Media provides a review of current molecular pathology for a specific organ system. The purpose of *Basic Concepts of Molecular Pathology* is to provide a succinct background of molecular pathology terminology and concepts to serve as a foundation and reference for understanding the medically oriented organ-specific information in the other books in the series. As such, *Basic Concepts of Molecular Pathology* serves as a companion to books in the *Molecular Pathology Library Series* and also stands alone as a quick review of molecular pathology principles.

Philip T. Cagle, MD
Houston, TX
Timothy Craig Allen, MD, JD
Tyler, TX
Series Preface

The past two decades have seen an ever-accelerating growth in knowledge about the molecular pathology of human diseases, which received a large boost with the sequencing of the human genome in 2003. Molecular diagnostics, molecular-targeted therapy, and genetic therapy are now routine in many medical centers. The molecular field now impacts every field in medicine, whether clinical research or routine patient care. There is a great need for basic researchers to understand the potential clinical implications of their research, whereas private practice clinicians of all types (general internal medicine and internal medicine specialists, medical oncologists, radiation oncologists, surgeons, pediatricians, family practitioners), clinical investigators, pathologists, medical laboratory directors, and radiologists require a basic understanding of the fundamentals of molecular pathogenesis, diagnosis, and treatment for their patients.

Traditional textbooks in molecular biology present basic science and are not readily applicable to the medical setting. Most medical textbooks that include a mention of molecular pathology in the clinical setting are limited in scope and assume that the reader already has a working knowledge of the basic science of molecular biology. Other texts emphasize technology and testing procedures without integrating the clinical perspective. There is an urgent need for a text that fills the gap between basic science books and clinical practice.

In the Molecular Pathology Library series, basic science and technology are integrated with the medical perspective and clinical application. Each book in the series is divided according to neoplastic and non-neoplastic diseases for each of the organ systems traditionally associated with medical subspecialties.

Each book in the series is organized to provide (1) a succinct background of the essential terminology, concepts, and technology of molecular biology; (2) an overview of the broad application of molecular biology principles to disease; and (3) specific application of molecular pathology to the pathogenesis, diagnosis, and treatment of neoplastic and nonneoplastic diseases specific to each organ system. These broad section topics are divided into succinct chapters, averaging about 15–20 pages each, to cover a very specific disease entity. The chapters are written by established authorities on the specific topic from academic centers around the world. In one book, diverse subjects are included that the reader would have to pursue from multiple sources to have a clear understanding of the molecular pathogenesis, diagnosis, and treatment of specific diseases. Attempting to hunt for the full information from basic concept to specific applications for a disease from the varied sources is time consuming and frustrating. By providing this quick and user-friendly reference, understanding and application of this rapidly growing field are made more accessible to both expert and generalist alike.

As books that bridge the gap between basic science and clinical understanding and practice, the Molecular Pathology Series serves the basic scientist, the clinical researcher, and the practicing physician or other health care provider who require more understanding of the application of basic research to patient care, that is, from “bench to bedside.” This series is unique and an invaluable resource to those who need to know about molecular pathology from a clinical, disease-oriented perspective. These books will be indispensable to physicians and health care providers in multiple disciplines, as already noted, to residents and fellows in these multiple disciplines as well as their teaching institutions, and to researchers who increasingly must justify the clinical implications of their research.

Philip T. Cagle, MD
Series Editor
Houston, TX
## Contents

<table>
<thead>
<tr>
<th>Contributors</th>
<th>Chapter 1 Genes, Gene Products, and Transcription Factors</th>
<th>Chapter 2 Receptors, Signaling Pathways, Cell Cycle, and DNA Damage Repair</th>
<th>Chapter 3 Cell Adhesion Molecules</th>
<th>Chapter 4 Apoptosis and Cell Death</th>
<th>Chapter 5 The Role of Mutation and Epimutation in the Development of Human Disease</th>
<th>Chapter 6 Bioinformatics and Omics</th>
<th>Chapter 7 General Approach to Molecular Pathology</th>
<th>Chapter 8 Applications in Anatomic Pathology</th>
<th>Chapter 9 Polymerase Chain Reaction and Reverse Transcription-Polymerase Chain Reaction</th>
<th>Chapter 10 Array Comparative Genomic Hybridization in Pathology</th>
<th>Chapter 11 Loss of Heterozygosity</th>
<th>Chapter 12 In Situ Hybridization: Principles and Applications</th>
<th>Chapter 13 Proteomics</th>
<th>Chapter 14 Immunohistochemistry of Biomarkers</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
<td>9</td>
<td>19</td>
<td>29</td>
<td>41</td>
<td>55</td>
<td>61</td>
<td>69</td>
<td>73</td>
<td>87</td>
<td>97</td>
<td>109</td>
<td>119</td>
<td>133</td>
</tr>
</tbody>
</table>
Contributors

Timothy Craig Allen
Department of Pathology, University of Texas Health Science Center at Tyler, Tyler, TX, USA

Roberto Barrios
Department of Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, NY, USA
The Methodist Hospital, Houston, TX, USA

Gregory L. Blakey
Department of Pathology, University of Oklahoma Health Sciences Center, Oklahoma City, OK, USA

Philip T. Cagle
Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, NY, USA
The Methodist Hospital, Houston, TX, USA

Chung-Che (Jeff) Chang
Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, NY, USA
The Methodist Hospital, Houston, TX, USA

William B. Coleman
Department of Pathology and Laboratory Medicine, UNC Lineberger Comprehensive Cancer Center, University of North Carolina School of Medicine, Chapel Hill, NC, USA

Kumarasen Cooper
Department of Pathology, University of Vermont/Fletcher Allen Healthcare, Burlington, VT, USA

Sanja Dacic
Division of Anatomic Pathology, Department of Pathology, University of Pittsburgh Medical Center, Pittsburgh, PA, USA

Daniel H. Farkas
Center for Molecular Medicine, Grand Rapids, MI, USA

Patrick L. Fitzgibbons
Department of Pathology, St. Jude Medical Center, Fullerton, CA, USA

Larry Fowler
Department of Pathology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

Wieslaw Furmaga
Department of Pathology, University of Texas Health Science Center at San Antonio, San Antonio, TX, USA

Thomas J. Goodwin
Department of Biomedical Research and Operations Branch, NASA/Johnson Space Center, Houston, TX, USA
Kevin C. Halling
Molecular Cytology and Imaging Laboratory, Department of Laboratory Medicine and Pathology,
Mayo Clinic, Rochester, MN, USA

Louise Hecker
Division of Pulmonary & Critical Care Medicine, University of Michigan Health System,
Internal Medicine, Ann Arbor, MI, USA

Jennifer L. Hunt
Surgical Pathology, Head and Neck/Endocrine Pathology, AP Molecular Diagnostics Unit,
Department of Anatomic Pathology, Cleveland Clinic, Cleveland, OH, USA

Jeffrey M. Jordan
Department of Pathology, University of Texas Medical Branch, Galveston,
TX, USA

Dwight Oliver
Department of Pathology and Laboratory Medicine, University of Texas-Houston Medical School,
Houston, TX, USA

Sem H. Phan
Department of Pathology, University of Michigan, Ann Arbor, MI, USA

Sharon C. Presnell
Cell and Tissue Technologies, Becton Dickinson, Research Triangle Park, NC, USA

Ashley G. Rivenbark
UNC Lineberger Comprehensive Cancer Center, Department of Biochemistry and Biophysics,
University of North Carolina School of Medicine, Chapel Hill,
NC, USA

Pothana Saikumar
Department of Pathology, University of Texas Health Science Center at San Antonio,
San Antonio, TX, USA

Munir Shahjahan
Pathology and Laboratory Medicine, The Methodist Hospital, Houston, TX, USA

Travis Solley
Department of Surgery, University of Texas Medical Branch, Galveston, TX, USA

Victor J. Thannickal
Division of Pulmonary & Critical Care Medicine, University of Michigan Health System,
Internal Medicine, Ann Arbor, MI, USA

Reinhard Ullmann
Department of Human Molecular Genetics, Max Planck Institute for Molecular Genetics, Berlin,
Germany

Manjeri A. Venkatachalam
Department of Pathology, University of Texas Health Science Center, San Antonio, TX, USA

Roger A. Vertrees
Department of Surgery, University of Texas Medical Branch, Galveston, TX, USA

Belinda J. Wagner
Scientific & Technical Operations, Tengion, Inc., Winston-Salem, NC, USA

Amy J. Wendel
Department of Laboratory Medicine and Pathology, Mayo Clinic, Rochester,
MN, USA

Youli Zu
Pathology and Laboratory Medicine, Weill Medical College of Cornell University, New York, NY, USA
The Methodist Hospital, Houston, TX, USA
Introduction

Molecular pathology employs an ever-expanding array of special techniques to study nucleic acids, genes, gene products, receptors, signaling pathways, the cell cycle, and mutations. This chapter and the others in this section provide a quick review of basic terminology and concepts for the understanding of subsequent chapters.

Nucleic Acids, Genes, and Gene Products

Genes are the bits of information that code for the proteins which are necessary for structure and metabolic reactions in living tissues. Genes and the molecules that construct their protein products using the blueprints or genetic code in the genes are composed of nucleic acids. Nucleotides are the building blocks of the nucleic acids, deoxyribonucleic acid (DNA) and ribonucleic acid (RNA), the essential compounds that make up genes and transcribe the genetic code into proteins, respectively. Nucleotides are basic compounds composed of a sugar-phosphate backbone and a nitrogenous base. Nucleotides consist of two types of bases: purines and pyrimidines. In DNA, there are two purines (adenine, abbreviated as A, and guanine, abbreviated as G) and two pyrimidines (thymine, abbreviated as T, and cytosine, abbreviated as C). In RNA, there are also two purines (A and G) and two pyrimidines (uracil, abbreviated as U, replaces T and C). DNA is typically double stranded, with the nucleotide bases paired together as described below, and RNA is typically single stranded. When nucleotides are assembled together in a nucleic acid, formation of covalent phosphodiester bonds results in a free 5′-phosphate at the origin of the nucleic acid and a free 3′-hydroxyl at the end (terminus) of the nucleic acid. For this reason, the synthesis of the nucleic acid is said to occur in a 5′-to 3′-direction (see below).

DNA is composed of nucleotides arranged sequentially in a deliberate order that encodes as genes for a matching sequential order of amino acids which will form proteins, as further discussed below. The nucleotides in genes are arranged in a double-stranded right-handed helix in the cell nucleus so that the purine A always binds with the pyrimidine T and the purine G always binds with the pyrimidine C in base pairs. As a result, the DNA in a double helix is arranged in complementary strands: the sequence of nucleotides in one strand of DNA is a “mirror image” of the nucleotide sequence in the other DNA strand. Groups of DNA base pairs are wrapped around small proteins called histones, forming arrangements of DNA called nucleosomes, allowing the DNA to fit within the cell nucleus. Genes are located on chromosomes that consist of DNA packaged with histone and nonhistone proteins. There are 23 pairs of chromosomes in the human, for a total of 46. Each gene is located at a specific site or locus on a specific pair of chromosomes. Because the chromosomes are in pairs and each chromosome has a locus for each gene, the genes occur as two copies or alleles, one copy or allele on each of the members of the chromosome pair. Chromosomes are ordinarily indistinct in the nuclear chromatin, but are discrete during mitosis or cell division.

The genome is the entirety of the DNA sequence or chromosomes of an organism, or its “complete genetic complement.” Genomics is the sequencing and study of genomes and cytogenetics is the study of chromosomes, traditionally through visualization of the karyotype or set of chromosomes of an organism. The somatic cells are diploid with a pair of each of the chromosomes and, therefore, two copies or alleles of each gene. During fertilization, the nuclear material of the two gametes combines, restoring the diploid number of chromosomes and alleles to the diploid number in the fertilized egg.

The genotype is the genetic information in an individual’s DNA, and the phenotype is how the genotype is manifested or expressed. Genotype and phenotype can differ. If two alleles are the same in one individual, they are said to be homozygous, and if the two alleles are different, they are...
said to be heterozygous. If a person is heterozygous for a gene, one allele may be expressed preferentially over the other allele, in which case the former allele is considered dominant and the latter allele is considered recessive. In this situation, the dominant allele codes for features that mask the features coded for by the recessive allele, so that the phenotype is different from the genotype. For example, a person is heterozygous for eye color and has an allele for brown eyes on one chromosome and an allele for blue eyes on the other member of that pair of chromosomes. That person will have brown eyes (the phenotype) because the allele for brown eyes is dominant over the allele for blue eyes. On the other hand, if the person is homozygous and has two recessive alleles for blue eyes, that person will have blue eyes.1–9

Single nucleotide polymorphisms, or SNPs (pronounced “snips”), are inherited, naturally occurring variations in one base between the DNA sequences in the same gene in two individuals and account for most of the genetic variation between individuals. Alleles or SNPs that are in close proximity on a chromosome are often inherited together as a haplotype. Polymorphisms are differences in DNA between individuals, and the most simple polymorphism is the SNP.1–9

Before cell division, new DNA must be synthesized from an existing strand of DNA, a process called replication. The synthesis of new DNA is a tightly controlled phase of the cell cycle (the S phase). The cell cycle is described in greater detail in Chap. 2. Before initiation of DNA replication, a prereplicative complex is constructed. This prereplicative complex is composed of the minichromosome maintenance protein complex (MCM), the origin recognition complex (ORC), and Cdc6/Cdc18. The S-phase kinases Cdc7 and Cdk (cyclin-dependent kinase) activate the prereplicative complex to yield an initiation complex at the origin with binding of Cdc45 to MCM. Replication is initiated at specific points in the DNA, referred to as origins of replication, and this creates a Y-shaped replication fork where the parental DNA duplex splits into two daughter DNA duplexes. As further described below, the duplex DNA is unwound with the assistance of special enzymes called topoisomerases, and then replication proteins, including DNA polymerases, bind to the unwound DNA.15–60

The synthesis of RNA, including messenger RNA (mRNA), from a strand of DNA (referred to as the DNA template) is known as transcription and is a fundamental step in the formation of the protein for which the DNA or gene codes. Condensed, inactive DNA at the periphery of the nucleus is called heterochromatin, and less-condensed DNA available for transcription is referred to as euchromatin, which is generally found in the central part of the nucleus.1–9

DNA polymerase is an enzyme that synthesizes DNA using single-stranded DNA as a substrate and requires a small segment of double-stranded DNA to initiate new DNA synthesis. RNA polymerase is an enzyme that synthesizes an RNA transcript from a DNA template during transcription. RNA polymerase first binds to a section of bases on the DNA called the transcription initiation site (TIS) or promoter “upstream” of the gene that is being transcribed. RNA polymerase I transcribes genes encoding for ribosomal RNAs (rRNAs), RNA polymerase II transcribes genes encoding for mRNAs (mRNAs), and RNA polymerase III transcribes genes encoding for transfer RNAs (tRNAs).61–66

The double-stranded helix of DNA must be unraveled and separated into single strands of DNA before it can undergo either transcription or replication. Topoisomerases are enzymes that break or “nick” a DNA strand, releasing the tension of the coiled helix and allowing the DNA to unwind. Transient DNA single-strand breaks are induced by topoisomerase I, and transient DNA double-strand breaks are induced by topoisomerase II. Once the DNA is separated into single strands, the DNA strand that serves as the template for the mRNA during transcription is referred to as antisense, and the complementary DNA strand, which has the identical sequence of bases as the mRNA (except that U replaces T), is referred to as sense.67–78

During transcription, base pairs are matched with the single strand of antisense DNA template to form a strand of mRNA. The resulting mRNA strand is a “mirror image” of the DNA template, except that uracil replaces thymine such that a DNA template with nucleotide sequence AGTC results in a strand of mRNA nucleotide sequence UCAG.1–9

A codon is a series of three base pair nucleotides in a gene that codes for a specific amino acid, and a series of base pair codons codes for a precise sequence of specific amino acids, resulting in the synthesis of a specific protein. The gene product is the final molecule, usually a protein, for which the gene codes that generates the effect of the gene. During translation, the mRNA, derived from the DNA template through transcription, is used as a template for the assembly of the protein product. The assembly of the protein product occurs in association with ribosomes, a component of which are rRNAs, and tRNAs add the amino acids to the protein under assembly. Each tRNA has a specific acceptor arm that attaches a specific amino acid. The tRNA ensures that the amino acid is added to the protein in the correct sequence based on the mRNA template, because the tRNA also has a specific anticodon that binds to the corresponding specific codon in the mRNA. The assembly of the protein product via mRNA is referred to as gene expression. Most gene expression is controlled at the level of transcription.79–81

Genes are made up of DNA segments called exons and introns. Exons are translated into the gene product, and introns are intervening DNA segments that are believed to play a regulatory role or serve as “punctuation” in the gene. As such, introns are spliced out of the sequence at the mRNA level, and the splice junction is the site between an exon and an intron where the splicing occurs.1–9 A short tandem repeat (STR) consists of a sequence of two to five nucleotides that are repeated in tandem, frequently dozens of times,
in introns. These STRs are found in microsatellite DNA, which is important in certain types of cancers such as colon cancer, although not significant in lung cancer.

The end regions of chromosomes are composed of the nucleotide sequence TTAGGG repeated hundreds of times and are called telomeres. Telomere sequences are lost each time that a cell replicates until the cell loses the ability to divide as part of the aging process. Telomerase is a specialized DNA polymerase that replaces the DNA sequences at the telomeres of the chromosomes. Telomerase allows cells to divide indefinitely, a factor that can be important in cancer.

**Posttranslational Modifications of Gene Products**

The specific sequence of amino acids in a protein imparts unique physicochemical properties that cause the polypeptide chain to fold into a tertiary structure which gives the protein its three-dimensional functional form. Domains are compact, spherical units of the three-dimensional tertiary structure.

Dimerization is the binding of two proteins together. Binding of proteins to other proteins can enhance or inhibit their function. Dimers are frequently encountered, but trimers (three proteins), tetramers (four proteins), or other combinations can occur. Homodimers consist of two identical proteins bound together, and heterodimers consist of two different proteins bound together.

Many proteins present in a cell are inert until they are activated by posttranslational modifications such as proteolytic cleavage or phosphorylation and become functional. The activation and inactivation of proteins by posttranslational modifications is essential in control of receptors, signaling pathways, transcription factors, and the cell cycle.

**Phosphorylation and Acetylation**

Phosphorylation is the addition of a phosphate group to a protein that is catalyzed by enzymes called kinases. De-phosphorylation is removal of a phosphate group from a protein that is catalyzed by enzymes called phosphatases. Many of the proteins in signaling pathways, including transcription factors, and the cell cycle are activated or inactivated by kinases and phosphatases. Depending on the domain that is phosphorylated, phosphorylation causes varying effects to a transcription factor. Phosphorylation can cause translocation (movement of a protein from the cytosol into the nucleus) and transactivation of genes, or inhibit binding proteins from binding to DNA.

Acetylation is the addition of an acetyl group to a protein that is catalyzed by acetyltransferases, and deacetylation is the removal of an acetyl group from a protein that is catalyzed by deacetylases. Similar to kinases and phosphatases, acetyltransferases and deacetylases activate and inactivate proteins involved in various molecular events.

**Protein Degradation and Ubiquitinylation**

To limit signaling proteins and remove damaged or abnormal proteins, protein degradation is necessary. Ubiquitinylation or poly ubiquitinylation (the ubiquitin-proteasome pathway) is the rapid degradation of proteins by reversible cross-linkage to a polypeptide called ubiquitin. Ubiquitin-activating enzyme (E1) activates ubiquitin, and the activated ubiquitin is transferred to a ubiquitin-conjugating enzyme (E2). The activated ubiquitin is transferred to the specific target protein by ubiquitin ligase (E3). Multiple ubiquitins are added to the protein, resulting in a polyubiquitinated protein that is degraded by a large protease complex known as the proteasome. During this process, the ubiquitin is released to participate in more cycles of ubiquitinylation. Ubiquitinylation rapidly removes cell-cycle regulators and signaling proteins, including those involved in cell survival and cell death (apoptosis).

**Transcription Factors**

Gene expression is primarily controlled at the level of transcription initiation. The transcriptional unit of DNA starts with the 5′-regulatory sequences and ends with the 3′-terminator signal of the gene. Gene-activating proteins are blocked from DNA by the tight binding of histone proteins to the DNA blocks. Histone acetyltransferases acetylate histones, which allows the gene-activating proteins to bind to the DNA. By blocking this process, histone deacetylases silence gene transcription.

Transcription factors, also called trans-acting factors or transactivators, are proteins that bind to DNA and regulate the activity of RNA polymerase. Transcription factors affect gene expression directly by induction or activation of the gene or by reducing transcription levels, causing silencing or inhibition of the gene.

Transcription factors that stimulate transcription or the synthesis of an RNA molecule from a DNA template are called transcriptional activators. In most cases, transcriptional activators have two domains: the DNA-binding domain recognizes and binds to a specific DNA sequence, and the transactivation domain interacts with the transcriptional machinery to induce transcription. Transcription factors can be categorized into families according to their DNA-binding domains; for example, zinc finger, leucine zipper, copper fist, basic helix-loop-helix, helix-turn-helix, and bZIP.

Trans-acting DNA sequences encode for diffusible transcription factors that bind to distant cis-acting DNA regulatory sequences but may sometimes bind to other proteins which, in turn, bind to DNA or the transcription machinery. There are two categories of diffusible transcription factors.
that bind to DNA. (1) General transcription factors are part of the basic transcription machinery by directly interacting with the RNA polymerase complex. Cis-acting DNA sequences that bind general transcription factors and function in all genes are called promoters. (2) Regulatory transcription factors activate or inactivate specific genes. Cis-acting DNA sequences that bind regulatory transcription factors to induce specific genes are called enhancers.\textsuperscript{134–148}

The initiation of transcription by RNA polymerase requires general transcription factors. A cis-acting DNA regulatory sequence that contains adenine-thymidine-rich nucleotide sequences, referred to as a \textit{TATA box}, is found in the promoters of many genes. The TATA-binding protein (TBP) and TBP-associated factors bind to form the general transcription factor TFIIID. This transcription factor, combined with other general transcription factors (TFIIB, TFIIF, TFIIE, and TFIIH), initiate transcription by binding RNA polymerase II to the promoter. A transcription bubble is formed when the transcriptional preinitiation complex binds to a specific sequence of nucleotides and there is separation or melting of the double-stranded DNA in conjunction with histone acetylation. After separation from the preinitiation complex, the transcribing enzyme moves down the DNA template along the reading frame. During transcription elongation, the transcription bubble moves down the DNA template in a 5' to 3'-direction (as noted earlier). Once transcription is terminated, the resultant mRNA is freed and processed before it is actively transported into the cytoplasm. In the cytoplasm, the mRNA enters the ribosome for translation of the protein product.\textsuperscript{149–161}

Loops in the DNA bring enhancers into proximity of the transcription initiation sites even when they are located a distance away in sequence. This proximity allows the enhancers to interact with general transcription factors or RNA polymerase complexes at the promoter, allowing enhancers to stimulate gene transcription above the basal level.

An example of transcription factors is the Myc/Max/Mad network of transcription factors that regulate cell growth and death. The Myc family includes N-myc, c-myc, and L-myc. The Mad family includes Mad1, Mxi1, Mad3, Mad4, Mnt, and Mga. The Mad family functions in part as antagonists of the Myc family. These proteins form heterodimers that determine their effect. Myc/Max heterodimers activate transcription causing cell growth, proliferation, and death. Mad/Max heterodimers competitively inhibit the Myc/Max-induced transcription, causing differentiation, cell survival, and inhibition of growth and proliferation.\textsuperscript{162–172}

References

1. Genes, Gene Products, and Transcription Factors


Cell-Surface Receptors and Signal Transduction

Ligands are extracellular messenger molecules such as growth factors, inflammatory cytokines, and hormones that bind to specific receptors on the cell surface (i.e., growth factor receptors, cytokine receptors, and hormone receptors). Binding of the ligands to their receptors causes activation of second messengers in the cytosol and, eventually, activation of nuclear transcription factors. (Transcription factors are discussed in Chap. 1.) The transcription factors then direct the transcription of a gene product as a result of the extracellular message (e.g., a growth factor may stimulate a growth factor receptor on the cell surface, causing activation of second messengers that eventually cause a transcription factor to cause transcription of a protein involved in cell growth). This cascade or activation and inactivation of protein messengers from the cell-surface receptors through proteins in the cytosol to the transcription factors in the nucleus is known as signal transduction. The series of steps that occurs during this process is called the signal transduction pathway or signaling pathway. Much of the activation and inactivation of proteins in signaling pathways occurs through reversible phosphorylation of tyrosine, serine, or threonine in the pathway proteins (see Chap. 1). Phosphorylation is accomplished by tyrosine kinases and serine/threonine kinases with phosphates donated from adenosine triphosphate (ATP) or guanosine triphosphate (GTP). Tyrosine kinases are much more common in signaling pathways than are serine/threonine kinases. This discussion focuses on growth factor receptors and cytokine receptors and their associated signaling pathways.

Growth factor receptors are a common type of cell-surface receptor. Polypeptide growth factors such as epidermal growth factor (EGF) serve as ligands that bind to cell-surface receptor protein-tyrosine kinases, which causes activation of the receptor by dimerization, resulting in autophosphorylation. The activated receptor binds other proteins within the cell, leading to their phosphorylation and activation of their enzyme activity as part of the signaling pathway. The type I growth factor receptor tyrosine kinase family consists of epidermal growth factor receptor (EGFR), and ErbB1, ErbB2, ErbB3, and ErbB4 make up the type I growth factor receptor-tyrosine kinase family. In addition to EGF, EGFR has multiple ligands, including transforming growth factor-α (TGF-α). Several signaling pathways are well studied and important to human disease. Signaling pathways transmit the “message” from extracellular ligands such as growth factors, cytokines, and steroid hormones. Signaling pathways are involved in regulation of cell proliferation, cell differentiation, cell death or apoptosis, and cell survival. Some of the more noteworthy and established pathways are briefly reviewed. Most signaling pathways have multiple complex interactions and “crosstalk” with other pathways, so discussion of specific pathways is limited here to an abbreviated overview.

The mitogen-activated protein kinase (MAPK) family is involved in multiple signaling pathways influencing cell growth, differentiation, and apoptosis, including the Ras/Raf-1/MAPK pathway mentioned later. The MAPK family includes the extracellular signal-regulated kinases (ERK1 and ERK2); the c-Jun NH2-terminal kinases (JNK1, JNK2, and JNK3); and p38 (p38 MAP kinases α, β, and δ). The MAP kinase kinase kinases (MKKK) are activated by a wide range of agents, including growth factors, oxidative stress, inflammatory cytokines, and ultraviolet radiation. Activated MKKK activate the MAP kinase kinases (MKK), which subsequently activate the MAP kinases. Examples of MKKK include Raf-1, TGF-β-activated kinase (TAK), apoptosis signal-regulating kinase 1 (ASK1), MAP/ERK kinase kinases (MEKK), germlinal center kinase (GCK), and p21-activated kinase (PAK). The ERKs have about 160 substrates and are antiapoptotic and involved in cell proliferation, cellular differentiation, and cell-cycle progression. On the other hand,
The Ras/Raf-1/MAPK pathway is significant in carcinogenesis. H-Ras, K-Ras, and N-Ras are members of the Ras family, a class of small GTP-binding proteins that are downstream targets of receptor tyrosine kinases. Ras is located at the plasma membrane inner surface. Activation of growth factor receptors converts Ras from its inactive guanosine diphosphate (GDP)-bound state to its active GTP-bound state. Activated Ras recruits Raf protein–serine/threonine kinase from the cytosol to the plasma membrane where kinases activate Raf. ERK, a member of the MAPK family, is activated by Raf through MAP/ERK kinase (MEK). In turn, the activated ERK phosphorylates and activates multiple other proteins, including other protein kinases. Activated ERK also translocates into the nucleus where it phosphorylates and activates transcription factors including Elk-1. Guanosine triphosphate hydrolysis by GTPase-activating proteins (GAPs) inactivates Ras. Ras is involved in many pathways, and epithelial cell proliferation is one of several possible results of Ras activation.

The JAK/STAT pathway is linked with cytokine receptors. When ligands stimulate cytokine receptors, the signal transducers and activators of transcription (STAT) proteins associate with the activated cytokine receptors. The STATs are phosphorylated by the JAK nonreceptor protein tyrosine kinases that are members of the Janus kinase (JAK) family, undergo dimerization, and translocate into the nucleus where they function as transcription factors for their target genes. The JAK/STAT pathway can also be activated in growth factor receptor pathways.

In the TGF-β/Smad pathway, cytokines in the transforming growth factor-β (TGF-β) superfamily inhibit the growth of many types of epithelial cells by formation of a complex of TGF-β type II and type I serine/threonine kinase receptors (TβRI and TβRII). The TGF-β1 ligand binds to TβRII, which subsequently phosphorylates and activates TβRI. Next, TβRI phosphorylates the receptor-regulated Smads (R-Smads) Smad2 and Smad3. Activated Smad2/3 complex with Smad4 (Co-Smad), translocate into the nucleus, and function as transcriptional modulators of TGF-1-regulated genes. On the other hand, the inhibitory Smads, Smad6 and Smad7, inhibit TGF-β1 signaling. The inhibitory Smads bind to Smad4 to prevent it from complexing with Smad2/3 or by binding to TβRI, blocking phosphorylation of Smad2 and Smad3.

The Wnt/B-catenin pathway involves Wnt (“Wingless,” derived from fruit fly studies), which binds to Frizzled cell-surface receptors. Signaling from Frizzled phosphorylates and activates Disheveled, which, in turn, inhibits the protein kinase glycogen synthase kinase-3 (GSK-3). β-Catenin phosphorylated by GSK-3 forms a complex with the adenomatous polyposis coli (APC) protein and the axin protein, restricting the quantity of free β-catenin in the cytosol. When activated Disheveled inhibits GSK-3, dephosphorylated β-catenin is freed from the APC–axin complex. This pathway is the classic Wnt signaling pathway, known as the canonical Wnt signaling pathway, and there are noncanonical Wnt signaling pathways that are not discussed here. β-Catenin associates with the TCF/LEF transcription factors, converting them to gene activators from gene repressors and, after translocation into the nucleus, binds to transcription factor TCF4, which induces Myc. β-Catenin also has roles in cell adhesion, which is discussed in Chap. 3.

The PI3K/Akt/mTOR pathway regulates cell survival. Akt is a protein serine/threonine kinase. Cell membrane phosphatidylinositol 4,5-bisphosphate in the cell is phosphorylated by phosphatidylinositol 3-kinase (PI3K), resulting in inositol 1,4,5-trisphosphate (PIP3). The PIP3 binds to the protein serine/threonine kinase Akt and recruits it to the inner surface of the cell membrane. At the inner surface of the cell membrane, Akt is phosphorylated and activated. Activated Akt phosphorylates proteins directly involved in cell survival as well as transcription factors and other protein kinases.

The nuclear factor-B (NF-B) transcription factor and NF-B signaling pathways regulate many proteins of the immune system, proteins that inhibit apoptosis and proteins which promote cell survival and proliferation. Nuclear factor-B consists of various dimers of the Rel protein family: Rel (c-Rel), RelA (p65), RelB, NF-B1 (p50 and its precursor p105), and NF-B2 (p52 and its precursor p100), of which the p50–p52 dimer is the most common. NF-B complexes bind to promoters to assist transcription in most situations, but homodimer complexes of p50 or p52 may inhibit transcription. Nuclear factor-B proteins are maintained in the cytoplasm in resting cells by associating with members of the inhibitory I.B family (I.B-a, I.B-β, and I.B-e).

Inhibitory-B must be degraded for NF-B to be activated. Inhibitory-B kinases (IKKs) are activated by MAPKKK or by ligands for Toll-like receptors (TLRs), interleukin (IL)-1/IL-18 receptors, the TNF receptor superfamily, and B- and T-cell receptors. Activated IKKs phosphorylate I.B, which is subsequently bound by E3I.B ubiquitin ligase complex-TrCP-SCF, which ubiquinylates I.B. The 26S proteasome degrades the ubiquinylated I.B, releasing NF-B complex to translocate into the nucleus where it binds to specific B sites on DNA. The NF-B complex is a transcription factor that regulates expression of proinflammatory cytokines, chemokines, adhesion molecules, cyclooxygenase-2, inducible nitric oxide synthase, major histocompatibility complex, IL-2, IL-12, and interferon, in addition to antiapoptotic and apoptotic genes.

Apoptosis is discussed in Chap. 4.

In the Hedgehog-Patched-Smoothened signaling pathway, after attachment of a lipid, Sonic Hedgehog (Shh) polypeptide binds to Patched on the cell surface, which prevents inhibition of Smoothened (Smo), a G protein-coupled receptor, by Patched. Smoothened activates the serine/threonine kinase Fused and the zinc finger transcription factor Gli (first detected in gliomas as a mutation), which induces Wnt signaling.
Notch is a receptor for direct cell-to-cell signaling. Delta binds to Notch, resulting in proteolytic cleavage of Notch. The intracellular domain of Notch is released and translocates into the nucleus, where it interacts with a transcription factor,117-122

The Cell Cycle

The cell cycle is the tightly regulated, sequential series of events or phases that govern cell proliferation, including preparation for DNA replication, DNA replication (see Chap. 1), preparation for cell division, cell division, and cell rest. The cell cycle provides orderly control of DNA replication and cell division in response to external and internal stimuli. The cell cycle is divided into several phases: G1 (cell at rest), G1 (preparation for DNA synthesis), S (DNA synthesis or replication), G2, and M (mitosis with nuclear and cellular division). Progression through the series of steps in the cell cycle is tightly regulated by cyclin-dependent kinases (Cdks) after they form complexes with proteins called cyclins. These complexes activate and inactivate proteins by phosphorylation, including proteins that otherwise act as "brakes" on progression through the cell cycle and the proliferation process. There are many interacting pathways and positive and negative feedback loops that control passage through the cell cycle. The cell cycle may be stimulated appropriately or inappropriately in various inflammatory diseases, and loss of cell-cycle regulation is a very important step in uncontrolled cell proliferation during carcinogenesis.

Checkpoints in the cell cycle ordinarily prevent the passage of damaged DNA to daughter cells. Checkpoints temporarily arrest the cell cycle at specific steps in the cell cycle to allow repair of damaged DNA or programmed cell death or apoptosis if the damage is too severe to be repaired (discussed in Chap. 4). The major checkpoint in the cell cycle is the restriction point where "commitment" to the cell cycle occurs in G1, as is further discussed below. In addition to the G1-S checkpoint, there are an S-phase checkpoint and a G2-M checkpoint.123-142

The Rad9–Rad1–Hus1 heterotrimer complex (9-1-1 complex) and the Rad17–RFC complex are damage sensor proteins that detect DNA damage at the checkpoints. The 9-1-1 complex is loaded around DNA by the Rad17–RFC complex. The ATR [ataxia-telangiectasiamutated (ATM) and Rad3-related protein kinase]-mediated and ATM-mediated phosphorylation and activation of Chk1 and Chk2 follow, and ATM and Chk2 phosphorylate and stabilize p53. Cyclin-dependent kinases are inactivated by the regulation of Chk1 and Chk2 phosphates and stabilize p53. Cyclin-dependent kinases are inactivated by the regulation of Cdc25, Wee1, and p53, which causes cell-cycle arrest, DNA damage repair is discussed later.143-146 After DNA damage repair, the DNA damage checkpoint is silenced, and the cell cycle restarts in a process called recovery, involving polo-like kinase (Plk1).147,148

Growth factor signaling initiates the cell cycle and maintains the transition through the G1 phase. When the cell passes through the restriction point of the cell cycle, the cell no longer requires growth factor signaling to complete the cell cycle, and the cell is "committed" to the cell cycle. Passage through the restriction point depends on phosphorylation of the retinoblastoma (Rb) gene product, pRb. The Rb product governs progression past the restriction point of the cell cycle and governs the expression of genes involved in DNA synthesis.149-152 Activation of cyclin D is necessary for progression of the cell cycle.153-159

In response to stimuli for mitosis, such as growth factor signaling, complexes of cyclin D with Cdk4 and Cdk6 phosphorylate pRb during G1 in response to stimuli for mitosis. In addition, cyclin E–Cdk2 complexes phosphorylate pRb just before the S phase. Families of Cdk inhibitors (the INK4 family, including p16INK4A, and the p21WAF1/Cip1/p27Kip1/p57Kip2 family) control these cyclin–Cdk complexes.160-171 Phosphorylation inactivates pRb in G1, releasing E2F transcription factors that activate transcription of numerous genes involved in DNA replication, such as c-Myc,185-187 initiating the S phase. As complexes with Cdks, cyclin A functions in both G1–S phase transition and in mitosis,188 and cyclin B is involved in entry into mitosis.189

One of the primary roles of p53, the product of the TP53 gene, is to "protect" the DNA through the arrest of the cell cycle at checkpoints in response to DNA damage or to help induce apoptosis when damage is beyond repair. Because of these roles, p53 has been referred to as the guardian of the genome. Part of the p53 arrest of the cell cycle is by activation of p21WAF1, which blocks cyclin–Cdk complexes necessary for cell-cycle progression.190-196 A member of the TGF-β superfamily of cytokines, TGF-β1, is involved in inhibition of cell-cycle progression.197,198

It is apparent that Rb and p53 play very crucial roles in the management of the cell cycle. Abnormalities of Rb and p53 are the most common abnormalities associated with the cell-cycle dysregulation of malignancy. However, because there are so many redundancies, interacting pathways, and positive and negative feedback loops, there are many other abnormalities that can produce effects similar to the direct loss of Rb or p53.

The CDKN2A gene encodes for two completely unrelated protein products, p16INK4A, the Cdk inhibitor mentioned earlier,204-209 and p14ARF. Both these proteins are transcribed from different exons of the CDKN2A gene. Abnormalities of either of these genes or their products can produce effects similar to abnormalities of the Rb or p53 genes themselves. By inhibiting Cdk4/6 kinase, p16 blocks phosphorylation and inactivation of pRb. Loss of p16 function results in loss of Rb function because Cdk4/6 kinase inactivation of pRb is not blocked. The other CDKN2A product, p14, destabilizes the MDM2 protein, which binds and degrades p53. Excessive levels of MDM2 resulting from loss of p14 function cause excess degradation of p53 and loss of p53 function. Abnormalities of CDKN2A, p16, p14, and MDM2 and other genes and their products upstream or
downstream of Rb and p53 can produce loss of control of the cell cycle similar to the direct loss of Rb and p53.

DNA Damage Repair

DNA is regularly damaged by endogenous factors (such as oxygen radicals), extracellular factors (such as chemicals, radiation, ultraviolet light), and errors in replication (such as stalled replication forks). As a result of exposure to these harmful agents, DNA undergoes depurination, deamination, hydrolysis, and nonenzymatic methylation (alkylation), which attach chemical groups called adducts to the DNA. DNA damage repair during cell-cycle checkpoints typically involves excision of the damaged DNA and filling of the resultant gap by newly synthesized DNA using the undamaged complementary DNA strand as template. Depending on the type of damage, there are several DNA repair pathways. The DNA damage repair pathways are important in individual susceptibility to cancer and in response to therapy.

The base excision repair (BER) pathway repairs small lesions such as oxidized or reduced single bases and fragmented or nonbulky adducts. In the BER pathway, a single damaged base is excised by base-specific DNA glycosylases (e.g., oxidized 8-oxoguanine is excised by 8-oxoguanine DNA glycosylase or OGG1). Some glycosylases are bifunctional and have an apurinic/apyrimidinic lyase activity to incise the phosphodiester bond of the intact apurinic/apyrimidinic site. An apurinic/apyrimidinic endonuclease (APE1/APEX1) is required for mono-functional glycosylates (e.g., oxidized 8-oxoguanine is excised by 8-oxoguanine DNA glycosylase or OGG1). Some glycosylases are bifunctional and have an apurinic/apyrimidinic lyase activity to incise the phosphodiester bond of the intact apurinic/apyrimidinic site. An apurinic/apyrimidinic endonuclease (APE1/APEX1) is required for mono-functional glycosylates to incise the apurinic/apyrimidinic site. DNA polymerase fills in the single nucleotide gap and a DNA ligase III/X-ray repair cross-complementing group 1 (XRCC1) complex seals the nick.

The nucleotide excision repair (NER) repairs lesions large enough to deform the DNA helical structure, such as pyrimidine dimers, bulky chemical adducts, and cross-links, by excising damaged bases as part of an oligonucleotide. Xeroderma pigmentosum (XP) proteins are an important part of the NER pathway. A protein complex including xeroderma pigmentosum group C protein (XPC) and hHR23B recognizes helical distortion by bulky chemical adducts. TFIHH is composed of nine protein subunits, including p62, p52, p44, p34, Cdk7, cyclin H, MAT1, and the two DNA helicases XPD (xeroderma pigmentosum group D protein), also known as ERCC2 (excision repair cross-complementation group 2), and XBP (xeroderma pigmentosum group B protein), also known as ERCC3 (excision repair cross-complementation group 3). TFIHH, xeroderma pigmentosum group A protein (XPA), and replication protein A (RPA) accumulate at the damage site, and the XPB and XBP helicases of TFIHH unwind the DNA double helix. This mechanism permits excision of the damaged single-stranded DNA fragment (usually about 27–30 bp) by a complex that includes ERCC1 and xeroderma pigmentosum group F protein (XPF).

DNA polymerases synthesize a new strand of DNA using the undamaged complementary DNA strand as template to complete the repair process.

Mismatch repair genes (MMR) participate in an excision repair pathway by scanning newly replicated DNA for mismatched base pairs such as deamination of a nucleotide into a different nucleotide. Heterodimers from MMR (including MLH1, MSH2, PMS1, and PMS2) cause cell-cycle arrest, permitting DNA repair.

The enzyme O6-meG–DNA methyltransferase (MGMT/AGT) repairs O6-meG and other alkylated bases in the direct damage reversal (DR) pathway. O6-Methylguanine (O6-meG) formed by alkylating compounds in tobacco smoke may mismatch with thymine during DNA replication if not repaired.

The DNA damage response or DSB repair (double-strand break repair) pathway (or during the S-phase checkpoint, the DNA replication stress response pathway) occurs in response to DSB DNA damage. The DSB repair pathway includes a cascade of events: sensing of the DNA damage and transduction of the damage signal to multiple pathways (cell-cycle checkpoints, DNA repair, responses to telomere maintenance, and apoptosis) (see Chap. 4). The DSB repair process involves many genes and their products, including the MRE11–Rad50–NBS1 complex (MRN); X-ray repair cross-complementing (XRC); the PI3K-like protein kinases (PIKKs) DNA-PKcs, ATM (mutated in ataxia telangiectasia), and ATR (ATM–Rad3-related); and ATM substrates NBS1 (Nijmegen breakage syndrome protein 1), SMC1 (structural maintenance of chromosomes 1), Chk1, Chk2, MRE11, p53, MDM2, BRCA1 (breast cancer protein 1), BRCA2/FANCD1 (breast cancer protein 2/Fanconi anemia protein D1), and FANCD2 (Fanconi anemia protein D2).

The ATM pathway reacts to DSBs in all phases of the cell cycle. The ATR pathway reacts to DSBs more slowly than ATM and reacts to factors that impede the function of replication forks. The ATM pathway activates many downstream proteins of the ATR pathway. In response to replication stress, ATM and ATR activate members of the Chk kinase family. The ATM phosphorylates Chk2, which, in turn, phosphorylates p53 and Cdc25A, blocking Cdk2.

References

2. Receptors, Signaling Pathways, Cell Cycle, and DNA Damage Repair


42. Pellegrini S, Dusanten-Fourt I. The structure, regulation and function of the Janus kinases (JAKs) and the signal transducers and activators of transcription (STATs). Eur J Biochem. 1997;248:615–633.


2. Receptors, Signaling Pathways, Cell Cycle, and DNA Damage Repair


Cell adhesion molecules, also termed cell adhesion receptors, are one of three classes of macromolecules — along with extracellular matrix molecules and adhesion plaque proteins — that mediate cell adhesion, an activity which is critical for the commencement and maintenance of the three-dimensional structure and normal function of tissues. Cell adhesion molecules are predominantly transmembrane glycoproteins that mediate binding to extracellular matrix molecules or to associated receptors on other cells, in a manner that determines the specificity of cell–cell or cell–extracellular matrix interactions. There are five families of adhesion receptors — integrins, cadherins, immunoglobulin cell adhesion molecules (Ig CAMs), selectins, and CD44. Complexes formed by cell adhesion receptors are not static, but are dynamic units capable of obtaining and incorporating extracellular environmental signals, and are indeed the foundation of two-way signaling between the cell and its external environment. These cell adhesion molecule families are also involved with signaling between the interior and exterior of the cell, and as such are important in cell growth, proliferation, spatial organization, motility, migration, signaling, differentiation, apoptosis, and gene transcription in normal physiological growth and development as well as in pathological conditions such as inflammation and wound healing.

The intricate processes involved in cancer invasion and metastasis require tumor cell detachment from the primary tumor, then tumor cell entry and exit from the lymphatic or vascular systems in turn, culminating in tumor cell growth at distant tissue sites. These steps can only take place with the dysregulation of normal cell–cell adhesion and cell–matrix interactions that are mediated by the cell adhesion receptor families. An increased understanding of cell adhesion molecules and their intercellular and extracellular interactions may provide opportunities for the development of diagnostic and therapeutic modalities for cancer.

### Integrins

The integrin superfamily, the largest known family of cell adhesion molecules, are widely expressed receptors via which cells attach to extracellular matrices or to other cells. They are expressed in various combinations in all cells. Integrins are transmembrane glycoproteins that form heterodimers consisting of one alpha (α) and one beta (β) subunit. Both subunits are required for interaction of integrins with the cytoskeleton and extracellular matrix, and they collaborate to bind extracellular matrix molecules, the specificity of which is determined by the subunit combination and cell type. The specific combination of the subunits determines the ligand-binding specificity, and various specific combinations of α- and β-subunits exhibit binding specificity for collagen receptors, fibronectin receptors, laminin receptors, vitronectin receptors, and other integrin receptors. Most integrins are substrate adhesion molecules that mediate interactions between cells and extracellular matrix components; however, some integrins are cell–cell adhesion molecules (CAMs). Lymphocytes express integrins that mediate heterotypic cell–cell adhesion, binding some Ig CAMs. The integrin receptor family plays a critical role in complex cellular events such as cellular differentiation, proliferation, and migration and is involved in biological processes related to organogenesis, wound healing, and the altered adhesive and invasive properties of tumor cells. Integrins transmit signals bidirectionally across plasma membranes. Alterations in integrin secretion or functional activity may regulate the development and progression of cancers, and changes in integrins in vivo leading to more aggressive tumor behavior have been identified in cancers of the lung, breast, colon, prostate, stomach, pancreas, liver, kidney, ovary, skin, and endometrium. β Integrins have been associated with tumor cell migration in cell lines including fibrosarcoma, bladder cancer, and colon cancer.
been identified immunohistochemically in bronchial epithelium, endothelium, and smooth muscle. In melanomas, increased expression of αvβ3 has been shown to correlate with a more malignant phenotype. Integrins may be important in the development of drug resistance in cancer cells. The inhibition of αvβ3 integrin has been shown to enhance the antitumor effect of radiotherapy. Integrin alterations may be necessary for the development of cisplatin resistance in cancer cells.

β1 Integrins

β1 Integrins make up the largest number of integrins, and there are at least seven receptors in this subfamily, each with different ligand specificity. αβ1, α5β1, α6β1, and αvβ1 are among the most studied of these receptors. α4β1 Integrin is a cell adhesion receptor, primarily found on lymphocytes, monocytes, and eosinophils, that may be a target for treatment of chronic inflammatory diseases. The regulation of β1 integrins by cytokines and other inflammatory mediators is important for eosinophil adhesion to endothelium and transendothelial migration. α4 Integrins are heterodimeric cell surface molecules important for leukocyte–cell and leukocyte–matrix adhesion interactions.

CXCR4 Chemokine Receptor

Cancer cells must pass through blood or lymphatic circulation and through their vessel walls to metastasize. Metastatic tumor cells are thought to co-opt signals normally controlling leukocyte movement, such as chemokine-mediated cell migration. Emigration from vessels is regulated by a sequence of distinct molecular signals, one of which involves chemokines that activate integrins and direct migration of leukocytes. The chemokine stromal cell-derived factor-1 (SDF-1/CXCL12) is a CXC chemokine expressed in bone marrow stromal cells. CXCR4, the receptor for CXCL12, has a critical role for the homing of hematopoietic stem cells in the bone marrow microenvironment. The CXCR4/CXCL12 axis may regulate migration and metastasis of several cancers, and the neutralization of the CXCR4/CXCL12 axis has been shown in vivo to inhibit or attenuate metastases of breast cancer.

Integrin-Linked Kinase

Integrin-linked kinase (ILK) is an ubiquitously expressed serine/threonine kinase that binds to the cytoplasmic domain of the β1 integrin subunit. It is involved in the regulation of cell survival and plays a role in cancer, with its upregulation identified in cancers of the colon, breast, prostate, and lung, and in melanoma. ILK is upregulated, often in a stage- and grade-dependent manner, and is associated with metastases. ILK is a central structural and signaling protein that serves as a link between the extracellular matrix and the cytoskeleton via integrins, paxillin, and parvins, and with adaptor proteins particularly interesting new cysteine-histidine-rich protein (PINCH) and Nck2, and between integrins and receptor tyrosine kinases. It also activates various signaling pathways through its kinase activity.

Focal Adhesion Kinase (FAK)

As integrin receptors are not catalytically active, they must recruit and activate other signaling molecules. FAK, a non-receptor tyrosine kinase enriched in focal adhesions and ubiquitously expressed during development, plays a key role in cell migration, proliferation, and survival. FAK may be a critical mediator of integrin signaling and has been linked to the integrin- and growth factor receptor-signaling pathways that regulate several biological processes concerning neoplastic transformation, invasion, and metastasis, including cell adhesion, migration, and apoptosis. FAK phosphorylation and kinase activity are regulated by integrin-mediated and matrix-dependent cell adhesion in tumor cells of many cancers. FAK overexpression and phosphorylation are related to increased cell tumor cell motility, invasion, and cytoskeleton alteration.

FAK is important for tumor progression. Elevated FAK mRNA levels have been found in human carcinomas and in acute lymphoblastic leukemias. The gene encoding FAK, termed protein-tyrosine kinase-2, located on chromosome 8q24, is increased in number in tumor cells. FAK overexpression in cancer contributes to an invasive phenotype resulting from increased formation of invadopodia. Further examination of the role of FAK role in cancer progression and invasion may allow for the development of targeted cancer therapy.

Integrins and adhesion of cells to extracellular membranes confer higher resistance to ionizing radiation and cytotoxic drugs. The survival-promoting effects of integrins has been identified in tumor response to conventional chemotherapeutic agents such as paclitaxel. Integrins, as well as integrin-associated proteins, modulate the radiation response of cells. ILK has been found in high levels in radiosensitive tumor cells. Because integrin-associated proteins are involved in all major signal transduction pathways regarding cell proliferation and survival, they are strong potential candidates for targeted therapy.

Cadherins

The cadherin family consists of calcium-dependent cell–cell adhesion molecules that are highly conserved transmembrane glycoproteins with similar domains for homophilic binding, calcium binding, and interaction with intracellular proteins. The cytoplasmic domain of cadherins interacts with catenins (α-catenin, β-catenin, γ-catenin), and the resulting complexes associate with cortical actin filaments.
The cadherin–catenin interaction is necessary for cadherin-mediated adhesion and association of the complexes with the cytoskeleton.\(^5,\)\(^7\) The cadherin family contains 16 members, of which the most significant is E-cadherin (epithelial cadherin), which is found in epithelial tissues and is involved in formation and maintenance of cell histocarchitecture.\(^6\) Other important cadherins include N-cadherin (neural cadherin), found in neural and muscle tissues, P-cadherin (placental cadherin), R-cadherin (retinal cadherin), and VE-cadherin (vascular endothelial cadherin).\(^6,\)\(^8\) Loss of function or secretion of the E-cadherin–catenin complex, or any of its components, eliminates a cell’s capacity to adhere, with resulting loss of normal tissue architecture.\(^6\) Altered or absent E-cadherin expression has been identified in various cancers, including stomach, head and neck, bladder, prostate, breast, and colon.\(^5,\)\(^9\)\(^-\)\(^6,\)\(^59\)\(^-\)\(^63\) \(\alpha\)-Catenin absence or alteration has been identified in breast, gastric, and esophageal cancers.\(^6,\)\(^64\)\(^-\)\(^65\) and alteration of \(\beta\)-catenin expression or phosphorylation has been found in cancers of the esophagus, stomach, and colon.\(^6,\)\(^66\)\(^-\)\(^68\) Reduced E-cadherin expression has also been identified in lung cancer.\(^6\) Assembly and disassembly of the cadherin–catenin complex, composed of E-cadherin, \(\beta\)-catenin, and \(\alpha\)-catenin, is highly regulated and is essential for proper cell–cell adhesion and development, the establishment of cellular polarity, and the maintenance of tissue organization.\(^6,\)\(^57\) Loss of the complex is important in cancer development.\(^6,\)\(^75\)

\(\beta\)-Catenin

\(\beta\)-Catenin is a cytoplasmic plaque protein involved in linking cadherin family receptors to the actin cytoskeleton.\(^6,\)\(^8\) It also is a cotranscriptional activator of genes in the nucleus together with lymphoid enhancer factor 1/T-cell factor.\(^6,\)\(^8\) It is important not only in the cadherin–catenin complex but also as a critical component of the Wnt signaling pathway.\(^5,\)\(^7\) In the presence of Wnt, stabilized \(\beta\)-catenin is translocated to the nucleus where it binds T-cell factor/lymphoid enhancer-binding factor-1 transcription factors, inducing target gene expression, including the cancer-related genes MYC, cyclin D1, and matrix metalloproteinase 7.\(^7,\)\(^67\)\(^-\)\(^69\) Both loss of \(\beta\)-catenin expression, and gain-of-function mutations, are common in human cancers, resulting in loss of cell–cell adhesion and increased gene transcription.\(^5,\)\(^70\) Its functions are probably controlled not by the regulation of the transcriptional level of its gene but rather by controlling the level of its cytoplasmic pool.\(^6,\)\(^8\)

\(\alpha\)-Catenin

\(\alpha\)-Catenin was first examined in human cancer cell lines such as the lung cancer cell line PC-9, which has a homozygous deletion of part of the \(\alpha\)-catenin gene.\(^6,\)\(^71\) \(\alpha\)-Catenin is involved in cancer progression via both canonical and noncanonical roles.\(^6,\)\(^8\) Loss of \(\alpha\)-catenin expression causes increased cell proliferation, and it is involved in apoptosis and growth factor signaling.\(^6,\)\(^72\)\(^-\)\(^74\) Dissociation of \(\alpha\)-catenin from the cadherin–catenin complex may lead to actin bundling and suppression of Arp2/3-mediated actin assembly, possibly causing decreased cell migration and stabilization of cell–cell adhesion.\(^6,\)\(^75\) As such, \(\alpha\)-catenin in cancers may cause rampant Arp2/3-mediated actin polymerization and increased lamellipodial membrane activity, with resulting decreased cell–cell adhesion, increased cell migration, and tumor cell invasion.\(^6,\)\(^75\)

E-Cadherin

Cadherins are essential for tight junctions between cells, and E-cadherin is the cadherin most strongly expressed in epithelial cells.\(^5,\)\(^75\)\(^-\)\(^76\) Cadherins form a complex with cytoplasmic proteins known as catenins, and the resulting complex, along with other cytoskeletal components such as actin, constitutes the intercellular adherence junction.\(^5,\)\(^75\)\(^-\)\(^77\) Transmembrane single-chain glycoproteins, cadherins have an extracellular domain, a transmembrane domain, and an intracellular domain, and play a primary role in maintaining physical cell–cell adhesion by mediating calcium-dependent homotypic intercellular adhesion in epithelial tissues.\(^5,\)\(^78\)\(^-\)\(^79\) Cadherin-mediated cell adhesion suppresses invasion of cancer cells in vitro, and dysfunction of the E-cadherin system correlates with cancer invasion in human cancers.\(^75,\)\(^80\)\(^-\)\(^84\) The human lung cancer cell line PC9 expresses an aberrant \(\alpha\)-catenin mRNA, and the cells have very loose cell–cell associations.\(^5,\)\(^85\)\(^-\)\(^86\) \(\alpha\)-Catenin is considered by some authors to be indispensable for cadherin-mediated cell–cell adhesion.\(^5,\)\(^75\)\(^-\)\(^77\) Researchers have identified reduced or heterogeneous E-cadherin expression and/or \(\alpha\)-catenin expression in undifferentiated invasive cancers, and impaired expression of E-cadherin and \(\alpha\)-catenin has been associated with increased lymph node metastases in breast, esophageal, and head and neck cancers.\(^5,\)\(^59\)\(^-\)\(^63\)\(^64\)\(^-\)\(^75\) The prognostic value of reduced E-cadherin expression in cancer patients has not been widely studied.\(^5,\)\(^59\)\(^-\)\(^61\)\(^75,\)\(^87\)\(^-\)\(^89\)

The relationship between E-cadherin and \(\alpha\)-catenin is mediated by \(\beta\)-catenin, and \(\beta\)-catenin in turn mediates the interactions of the cadherin–catenin complex with the c-erbB-2 gene product and epidermal growth factor receptor (EGFR).\(^5,\)\(^75\)\(^-\)\(^83\) APC protein, a tumor suppressor gene product, interacts with \(\beta\)-catenin and plakoglobin and is important in the E-cadherin-mediated cell adhesion system, influencing tumor invasion and metastasis.\(^75\)

Cadherins and Cyclooxygenase 2

Cyclooxygenase 2 (COX-2) and its metabolite prostaglandin E\(_2\) (PGE\(_2\)) are critical for regulating diverse cellular functions under both physiological and pathological conditions.\(^94\)\(^-\)\(^97\) COX-2 is overexpressed in human non-small cell lung cancers, and its inhibition causes tumor reduction in vivo in murine lung cancer models.\(^94\)\(^-\)\(^98\) COX-2 activity is identified throughout the progression of a premalignant
lesion to the metastatic phenotype.\textsuperscript{94,99} Higher COX-2 expression has been identified in lung adenocarcinoma lymph node metastases.\textsuperscript{94,99} COX-2 overexpression has been associated with angiogenesis, decreased host immunity, and enhanced invasion and metastasis, and therefore COX-2 has been considered to have an important role in multiple pathways in lung cancer carcinogenesis, suggesting it has a multifaceted role in conferring malignant and metastatic phenotypes.\textsuperscript{94,100–104} COX-2 may be a central element in orchestrating the multiple genetic alterations required for lung cancer invasion and metastasis.\textsuperscript{94,104} COX-2-dependent invasive capacity in lung cancer is caused by PGE\textsubscript{2}-mediated regulation of CD44 and matrix metalloproteinase-2.\textsuperscript{94} Normal cell–cell adhesion disruption leads to enhanced tumor cell migration and proliferation, with resulting invasion and metastasis.\textsuperscript{94,105,106} Downregulating the cadherin family or catenin family members, or activation of signaling pathways that prevent cell–cell cadherin junction assembly, can cause this disruption.\textsuperscript{94,105} As such, extracellular matrix and cell–cell adhesion are significant barriers to tumor metastasis.\textsuperscript{94,105} The E-cadherin–catenin complex is required for intercellular adhesiveness and normal tissue architecture maintenance.\textsuperscript{94,105} E-cadherin reduction has been linked to tumor invasion, metastasis, and poor prognosis.\textsuperscript{94,107} E-cadherin loss, along with increased COX-2 expression, has been identified in familial adenomatous polyposis.\textsuperscript{94,103}

### Cadherins and EGFR

Epidermal growth factor receptor (EGFR) is overexpressed in many non-small cell lung cancers (NSCLC), and treatment with the EGFR tyrosine kinase inhibitors gefitinib and erlotinib has shown improved survival in some chemotherapy-resistant NSCLC patients.\textsuperscript{108–111} Nonetheless, about half these NSCLC patients have tumor progression within 8 months and show no treatment benefit.\textsuperscript{108} Activating mutations in the EGFR tyrosine kinase domain have been shown to increase EGFR copy number and/or expression of EGFR protein correlating with response and survival after EGFR tyrosine kinase inhibitor therapy.\textsuperscript{108,112,113} EGFR activation and signaling through its downstream targets are modulated by E-cadherin; specifically, E-cadherin inhibits EGFR ligand activation and enhances AKT activation in neighboring cells.\textsuperscript{108,110–113} High phosphorylated AKT levels may predict tumor response to EGFR tyrosine kinase inhibitors.\textsuperscript{110,114} In lung cancer cell lines, E-cadherin expression is regulated by β-catenin signaling and by zinc finger proteins, including the Slug/Snail family, SIP1 and ZEB1.\textsuperscript{108,115} These transcription factors regulate gene expression by interaction with two 5′-CACCTG (E-box) promoter sequences.\textsuperscript{108,116} This regulation is facilitated by interaction with the transcriptional corepressor CtBP, which recruits histone deacetylases (HDAC), causing chromatin condensation and gene silencing.\textsuperscript{108,117}

### H-Cadherin

The H-cadherin gene, a cadherin superfamily member that has been isolated and mapped to 16q24, lacks the cytoplasmic domain, in contrast to E-cadherin, P-cadherin, and N-cadherin.\textsuperscript{118} Genetic abnormalities in the H-cadherin gene have been identified in human cancer cell lines and in lung, stomach, and ovarian cancers.\textsuperscript{118}

### Epithelial–Mesenchymal Transition

Epithelial–mesenchymal transition, a primary cellular process involved in remodeling tissues during development, plays an important role in the progression of cancers toward a more aggressive phenotype that is associated with increased motility and invasiveness.\textsuperscript{119,120} Changes in cadherin expression accompany this process, with loss of E-cadherin and upregulation of N-cadherin leading to increased invasive progression, possibly because of a functional cooperation between N-cadherin and the GFGR, which may potentiate ERK signaling to upregulate matrix metalloproteases and downregulate E-cadherin, yielding localized epithelial–mesenchymal transformation.\textsuperscript{119}

### Selectins

The selectin family of adhesion molecules, consisting of L(eukocyte) selectin, E(ndothelial) selectin, and P(latelet) selectin, are calcium-dependent type I transmembrane glycoproteins with extracellular lectin-like domains that interact, for example, with sialylated carbohydrate-determinant and mucin-like glycoproteins.\textsuperscript{121} They have been studied predominantly by examination of the recruitment of leukocytes from the circulation.\textsuperscript{4,122} Adhesion with selectins is calcium dependent, and the ligands are cell-surface glycans possessing a specific sialyl-Lewis X-type structure also found in blood group antigens.\textsuperscript{14} P selectin binds to P-selectin glycoprotein ligand (PSGL-1), L selectin interacts with GlyCAM-1 and CD34, and E selectin possibly reacts with ESL-1.\textsuperscript{4} Selectins, similar to Ig CAMs, are expressed on cell surfaces in low levels.\textsuperscript{4} P selectin is recruited from storage in Weibel–Palade bodies to the cell surface after inflammatory stimuli activate cells, and E selectin is synthesized and transported to the cell surface upon exposure to inflammatory mediators.\textsuperscript{4,122} L selectin, present on leukocyte surfaces, is shed from the cell surface into circulation.\textsuperscript{4,123} Enhanced cell-surface selectin expression is related to slowing and rolling of leukocytes at the endothelial cell wall.\textsuperscript{4,124} Targeted disruption of P selectin in mice confirms the selectin–ligand interaction is vital for leukocyte recruitment from the circulation, and this step has been examined for possible therapeutic intervention for inflammatory diseases, with antibodies against selectins and against carbohydrate moieties used to block inflammatory cell recruitment in animal models.\textsuperscript{4,125,126}
The selectin family of cell adhesion molecules interacts with their cognate glycoprotein ligands to mediate tethering, rolling, and weak adhesion. The integrin family of cell adhesion molecules interacts with their ligands of the Ig CAM superfamily to mediate firm adhesion and signal transduction, eventually triggering shape changes of the adherent leukocytes and transendothelial migration. P selectin rapidly translocates to the cell surface by exocytosis and mediates leukocyte rolling to the activated endothelial cells and heterotypic aggregation of activated platelets to leukocytes, upon thrombogenic and inflammatory challenges. P selectin interacts with a leukocyte surface sialomucin, PSGL-1, expressed on a variety of human leukocytes. P selectin has been found to bind to several human cancers and cancer cell lines, including NSCLC and squamous cell carcinoma (SCC), colon cancer, breast cancer, malignant melanoma, gastric cancer, and neuroblastoma. NKI-4 cells, a human malignant melanoma cell line, bind P selectin and express glycoprotein ligands for P selectin that are functionally and structurally distinct from leukocyte PSGL-1.

Sialyl Lewis x (sLe-x) and sialyl Lewis a (sLe-a) are cancer-associated carbohydrate antigens involved in metastasis. Both serve as ligands for P selectin, E selectin, and PSGL-1, expressed on the extracellular domain of the protein, and typically these molecules span the cell membrane and contain only a short cytoplasmic tail. The molecules are important in nervous system development, embryonic development, and immune and inflammatory responses. NCAM is a family member that is critical for the preservation and integrity of the nervous system. Ig CAMS called ICAMS are important for cellular adhesion in the immune system. ICAMs, including VCAM-1 and MadCAM-1, serve as ligands for integrins. PECAM, another Ig CAM, is vital for leukocyte adhesion, and is important in cell–cell contact to promote adhesion between endothelial cells and leukocytes. Modulation in ICAM expression may help regulate the type of leukocyte recruited to an area of inflammation and the temporal pattern of leukocyte recruitment. ICAM-3, found on leukocytes and endothelial cells, interacts with lymphocyte function-associated antigen-1 and is involved in leukocyte intercellular adhesion. CAM-3 may also be important for angiogenesis. ICAM-3 may play a role in tumor progression, and studies have shown a relationship between ICAM-3 and some diseases. Radiation-resistant cervical cancer cases have shown increased ICAM-3 expression in tumor stromal endothelial cells and lymphocytes.

Some adhesion molecules serve not only as adhesion substances but also as regulators of other cell functions by influencing signaling, a process termed “outside-to-in signaling.” Also called β1 integrins, β2 integrins, and CD28 induce costimulatory signals in the binding of T cells to antigen-presenting cells via multiple cellular signaling molecules, including FAKs, causing cell activation and cytokine production. Adhesion molecules are regulated by intracellular signaling induced by various cellular stimuli in the process termed “inside-to-out signaling.” ICAM-1 expression is highly regulated by locally produced inflammatory cytokines such as interleukin-1β, tumor necrosis factor-α, interleukin-6, and interferon-γ. The ICAM-1 leukocyte function-associated antigen-1 pathway regulates important cell–cell interactions such as...
leukocyte adhesion and migration, including tumor cell killing by natural killer cells and cytotoxic T lymphocytes.\textsuperscript{144,147} While several tumor cells highly express ICAM-1, a potent ligand for leukocyte function-associated antigen-1 on cytotoxic T lymphocytes in vitro, many tumor cells remain viable against killing by cytotoxic T lymphocytes in vivo.\textsuperscript{144}

**Cd44**

Adhesion molecule Cd44, a cell-surface transmembrane glycoprotein, is distributed extensively and can be detected on lymphocytes and fibroblasts.\textsuperscript{148–150} Cd44 is a principal cell-surface receptor for hyaluronan, a major extracellular matrix component, and it communicates cell–matrix interactions into cells via "outside-in signaling."\textsuperscript{151} Cd44 assists in cell-to-cell interactions and cell-to-matrix interactions; it has been identified in several isoforms on tumor cells, and serum levels have been shown to be prognostically relevant in stomach cancer patients.\textsuperscript{153,152} It is involved in lymphocyte activation, recirculation, and homing, adhesion of extracellular matrix, angiogenesis, cell proliferation, and cell migration.\textsuperscript{149,151} Tumor expression of Cd44 has been found to be an independent predictor of cancer recurrence with both colorectal cancer and NSCLC.\textsuperscript{135,154} Several isoforms are expressed by tumor cells, and serum levels have been reported to have prognostic significance in gastric cancer.\textsuperscript{135,150} Cd44v expression, related to tumor progression, metastasis, and prognosis, has been detected in cancers of the lung, colon, esophagus, liver, cervix, and kidney, as well as reticulo-sarcoma.\textsuperscript{148}

The secretion of E-selectin, an endothelial ligand for sialyl carbohydrate antigens expressed on the surface of tumor cells, is reportedly stimulated by tumor cells, and serum E-selectin titers have reportedly correlated with prognosis in breast cancer, and, in a few small studies, in NSCLC.\textsuperscript{135} D’Amico et al., examining 196 stage I NSCLC patients, examined serum samples and found that decreasing serum E-selectin levels (P = 0.002), increasing serum C44 levels (P = 0.001), and increasing serum urokinase plasminogen activator receptor (P = 003) predicted NSCLC recurrence before clinical or radiographic determination.\textsuperscript{135} Future research regarding such markers might one day allow for the prediction of NSCLC recurrence in patients who have undergone resection therapy, so that earlier and improved systemic therapy might be provided.

Cd44, expressed in a large number of normal and malignant tissues, interacts with its ligands, hyaluronate and osteopontin, and is involved in lymphocyte homing, as well as T-lymphocyte activation and tumor metastasis.\textsuperscript{155} Several Cd44 isoforms have been identified, arising from mRNA alternative splicing, with Cd44s, also termed Cd44h, being the isoform expressed with hematopoietic cells, fibroblasts, glial cells, and melanoma cells.\textsuperscript{155} The isoform generally identified on epithelial cells and epithelial tumor cells is high molecular weight Cd44, containing one or more alternatively spliced exons.\textsuperscript{155} Cd44 isoform alterations are associated with the transformation of normal cells into cancer cells.\textsuperscript{155} Cd44 isoforms, including high molecular weight Cd44 isoforms such as Cd44v6, have been shown to be overexpressed in non-small cell lung cancers.\textsuperscript{155} Cd44 isoform function in tumors varies with the organ involved.\textsuperscript{155} Cd44S overexpression in melanoma and lymphoma enhances tumorigenicity, whereas upregulation of Cd44S in colon cancer reduces tumorigenicity.\textsuperscript{155} Cd44 expression reduction in non-small cell lung cancer has been shown to be associated with lymph node metastases and shortened disease-free survival.\textsuperscript{155} The interplay between Cd44 and extracellular matrix proteins such as osteopontin may regulate osteoclast motility, responsible for osteolytic metastases.\textsuperscript{156}

**Conclusion**

Further research regarding the various mechanisms of action of each of the cell adhesion molecules, and their interactions with other molecular pathways, in the development and spread of cancer is ongoing. For example, the effect of glycosylation of cell adhesion molecules on their cell–cell and cell–extracellular matrix interactions are being studied.\textsuperscript{157} Ultimately, more effective cancer prognostic markers and therapeutic regimens may arise from these studies.

**References**


Apoptosis and Cell Death

Pothana Saikumar and Manjeri A. Venkatachalam

Introduction

Cell death plays a central role in multicellular organisms during their early development in sculpting the body parts and in their adult life by controlling cell numbers (homeostasis), and cell death further protects the organism overall by removing all cells damaged by disease, aging, infection, genetic mutation, and exposure to toxic agents. Kerr et al. coined the term “apoptosis,” referring to a particular morphology of physiological cell death involving cell shrinkage, nuclear condensation, membrane blebbing, and cellular and nuclear fragmentation into membrane-bound apoptotic bodies (Fig. 4.1). The lipid changes that occur in the membrane eventually lead to phagocytosis of the apoptotic cellular fragments. Often apoptosis is used synonymously with programmed cell death, implying that death results from the regulated activation of a preexisting death program that is encoded in the genome. The first evidence of a genetic program that orchestrates physiological cell death came from the developmental studies of the nematode *Caenorhabditis elegans.* Our knowledge of cell death and the mechanisms of its regulation increased dramatically in the past two decades with the discovery of several death genes in *C. elegans* and their counterparts in mammals. Now it is clear that apoptotic cell death has important biological roles not only in development and homeostasis but also in the pathogenesis of several disease processes. Dysregulation of apoptosis is found in a wide spectrum of human diseases, including cancer, autoimmune diseases, neurodegenerative diseases, ischemic diseases, and viral infections. Interestingly, death programs other than apoptosis with different morphological features were also considered in controlling the cell numbers. These death programs could be directed by the condemned cell itself or by neighboring cells, with or without the help of humoral factors.

Alternative Forms of Cell Death

Previously, cell death was broadly categorized into only two distinct types: apoptosis and necrosis. However, it has become increasingly evident that such a categorization is an oversimplification. There are 12 different types of cell death that have been described in the literature, which can be grouped into only five major types: apoptosis, necrosis, autophagy, paraptosis, and autoschizis. Other forms of death could be classified under one of these headings. For example, anoikis and oncosis are forms of apoptosis (triggered by cell detachment) and necrosis, respectively. Because of overlapping and shared signaling pathways among different death programs, it is difficult to provide exclusive definitions for each one of these cell death programs. However, there is evidence that inhibition of one form of cell death may lead to another. Although the term “programmed cell death” was used specifically to describe apoptosis, other forms of cell death may also fall into this category, if gene activation is required for their execution.

Necrosis

Necrosis results from a variety of accidental and lethal actions by toxins or physical stimuli or in association with pathological conditions, such as ischemia. Necrosis is characterized by cellular edema, dissolution of nuclear chromatin, disruption of the plasma membrane, and release of intracellular contents into the extracellular space, resulting in inflammation (see Fig. 4.1). In contrast, membrane damage occurs very late in the apoptotic process, where dead cells are engulfed by neighboring cells or phagocytes, leading to little or no inflammation (see Fig. 4.1). Although necrosis has mostly been regarded as an accidental form of cell death, more recent
data have suggested that necrosis can also occur as a programmed form of cell death. There is growing evidence that necrotic and apoptotic forms of cell death may share some similarities. Nevertheless, necrosis has been shown to occur in cells having defective apoptotic machinery or upon inhibition of apoptosis, and this form of cell death is emerging as an important therapeutic tool for cancer treatment. Some forms of cell death may have a mixture of morphological features of both necrosis and apoptosis and are referred to as “aponecrosis.”

**Autophagy**

Autophagy, which is referred to as “macroautophagy” or type II programmed cell death, is characterized by sequestration of cytosol and organelles by double or multilamellar structures called autophagic vacuoles, followed by degradation of the contents of these vacuoles by fusing to the cell’s own lysosomes (see Fig. 4.1). Although the precise role of autophagy in cell death is not clear, yet it has long been regarded as a cell survival mechanism whereby starving...
cells produce energy-generating basic molecules from complex polymers such as proteins, lipids, polysaccharides, and nucleic acids through sequestration and degradation of some of the cytoplasm and organelles. In this regard, it is argued that autophagy may help cancer cells to survive under nutrient-limiting and low-oxygen conditions and against ionizing radiation.10,11 However, recent observations that there is decreased autophagy during experimental carcinogenesis and heterologous disruption of Beclin 1 (Atg6),12-14 an autophagy gene, in breast and ovarian cancers suggest that breakdown of autophagic machinery may contribute to development of cancer.15 In addition, Atg6 heterozygous mice develop tumors spontaneously, suggesting tumor suppressor activity of Atg6.12 Similarly, mice deficient in another autophagy gene, Atg4C, a cysteine protease, are susceptible to chemical carcinogenesis.16 However, the mechanism of how autophagy suppresses tumorigenesis is still unclear. Other studies have shed some light on the relationship between autophagy and apoptosis. For example, reduction of Atg7 and beclin 1 inhibited zVAD-induced death in human U937 cells17 and knockdown of Atg5 and beclin 1 protected Bax/Bak double-knockout cells from staurosporine- or etoposide-induced nonapoptotic death.18,19 However, these studies were done in cells whose apoptotic pathways had been compromised. Thus, it remains to be seen whether cells with intact apoptotic machinery can also die by autophagy and whether apoptosis-dependent cells lacking autophagy genes will be resistant to other death stimuli. In the in vivo setting, the complicated relationship between autophagy and apoptosis needs to be clarified.

Because some autophagic cells would undergo caspase-independent gene-activated cell death but not display any of the characteristic ultrastructural features of apoptosis including DNA laddering, autophagy is still considered as programmed cell death. Similar to apoptosis, cells do require de novo gene expression with an increase in the expression of the ubiquitin-like gene.20,21 The ubiquitin-like protein conjugation system and formation of protein complexes that direct membrane docking and fusion of the lysosomes and vacuoles are main components of autophagy. Although the molecular details are still being elucidated, this process appears to be regulated by various kinases, phosphatases, and guanosine triphosphatases (GTPases).22,23

Paraptosis

Paraptosis was recently described as a form of cell death characterized by extensive cytoplasmic vacuolation (see Fig. 4.1) involving swelling of mitochondria and endoplasmic reticulum. This form of cell death was not inhibited by the caspase inhibitors but is inhibited by translation and transcription inhibitors, cycloheximide and actinomycin D, suggesting a requirement for new protein synthesis.24 Paraptosis has been shown to be triggered by the tumor necrosis factor (TNF) receptor family TAJ/TROY and insulin-like growth factor I receptor. TAJ/TROY-induced paraptotic cell death was enhanced by overexpression of programmed cell death 5 (PDCD5).25 This form of cell death was shown to be mediated by mitogen-activated protein (MAP) kinases and inhibited by AIP1/ALIX, a protein interacting with the calcium-binding death-related protein ALG-2.26 Inhibition by Na+/H+ exchanger also led to cell death resembling paraptosis.27 Recently, glucocorticoids were shown to induce retinal toxicity by mechanisms associated with paraptosis.28 Although this form of cell death has been shown to be evoked by a variety of reagents, the exact molecular mechanism is far from clear.

Cytoplasmic vacuolation, as seen in paraptosis, has been shown to occur in a wide range of cell lines either spontaneously or as induced by a variety of stimuli. The extent to which a cell becomes vacuolated depends on the cell type. The process of vacuolation seems to follow a definite pattern, with the vacuole number and size increasing gradually. Cells can recover from vacuolation up to a certain threshold, beyond which they succumb to death.29 The presence of a definite underlying program that decides whether a cell undergoing cytoplasmic vacuolation should be destined to death is currently unknown. However, there are instances in which a decidedly vacuolated cell undergoes cell death, which suggests the existence of some underlying programs. During the salivary gland development in metamorphosis of blowfly larva (Calliphora vomitoria), cells are eliminated by a process of intensive vacuolation in a particular order.30,31 Furthermore, in Dictyostelium discoideum, stalk cells undergo extensive cytoplasmic vacuolation while proceeding toward programmed cell death.32 In addition, cytoplasmic vacuolation was also observed in aspirates of lobular and ductal breast carcinoma.33 These studies further emphasize that cytoplasmic vacuolation-induced cell death has a physiological role and warrants further investigation. The most studied cytoplasmic vacuolation-induced cell death is autophagy.

Autoschizis

A new form of cell death, which has been shown to be triggered by oxidative stress,34 differs from apoptosis and necrosis and is characterized by reduction of cytoplasm to a narrow rim around the nucleus with chromatin marginating the entire nucleus from inside.35 Mitochondria and other organelles aggregate around the nucleus as a consequence of cytoskeletal damage and loss of cytoplasm. Interestingly, the rough endoplasmic reticulum is preserved until the late stages of autoschizis, in which cells fragment and the nucleus becomes condensed and breaks into smaller fragments.35 The nuclear envelope dissipates eventually with the remaining organelles after cell death. In this type of death, cells lose cytoplasm by self-morsellation or self-excision (see Fig. 4.1). Autoschizis usually affects contiguous groups of cells both in vitro and in vivo but can also occasionally affect
scattered individual cells trapped in subcapsular sinuses of lymph nodes.36

Apoptosis

Genetic studies in the nematode worm C. elegans led to the initial characterization of apoptosis. Activation of specific death genes during the development of this worm results in death of exactly 131 cells, leaving 959 cells intact. Further studies revealed that apoptosis can be divided into three successive stages: (1) commitment phase, in which death is initiated by specific extracellular or intracellular signals; (2) execution phase; and (3) cleanup phase, in which dead cells are removed by other cells with eventual degradation of the dead cells in the lysosomes of phagocytic cells.37 The apoptotic machinery is conserved through evolution from worm to human.38 In C. elegans, execution of apoptosis is mediated by CED-3 and CED-4 proteins. Commitment to a death signal results in the activation of CED-3 by CED-4 binding. The CED-9 protein prevents activation of CED-3 by binding to CED-4.39

Mechanisms of Apoptosis

The mechanisms of apoptosis are very complex and involve a cascade of energy-requiring molecular events. Past research indicates that there are two main apoptotic pathways: the extrinsic or death receptor pathway, and the intrinsic or mitochondrial pathway. An additional pathway also operates during T-cell-mediated cytotoxicity involving perforin-granzyme-mediated cell killing. The perforin/granzyme pathway can induce apoptosis via either granzyme B or granzyme A. All three pathways then converge on the same terminal execution pathway, which is initiated by the cleavage and activation of effector caspase-3 and -7, resulting in the degradation of cytoplasmic, cytoskeletal, and nuclear proteins, fragmentation of nuclear DNA, formation of apoptotic bodies (membrane blebbing), expression of cell-surface ligands for phagocytosis, and finally uptake by phagocytic cells.

Caspases

The executors in both intrinsic and extrinsic pathways of cell death are the caspases,40 which are cysteine proteases with specificity to cleave their substrates after aspartic acid residues. The first member of the caspase family, caspase-1, was initially known as interleukin-1β-converting enzyme (ICE), an enzyme required for the maturation of IL-1β. Later, the C. elegans cell death gene ced-3 was found to be similar to ICE and a developmentally regulated protein Nedd2 (now referred to as caspase-2). Subsequently, many caspase genes have been cloned from mammalian and non-mammalian sources. The central role of caspases in apoptosis is underscored by the observation that apoptosis and all classic changes associated with apoptosis can be blocked by inhibition of caspase activity. To date, 12 mammalian caspases (caspase-1 to -10, caspase-14, and mouse caspase-12) have been identified. Caspase-13 was later found to represent a bovine homologue, and caspase-11 appears to be a murine homologue of human caspase-4 and -5, respectively. These caspases have been broadly categorized into initiators (caspase-2, -8, -9, -10), effectors or executioners (caspase-3, -6, -7), and inflammatory caspases (caspase-1, -4, -5). While murine caspase-11 is reported to regulate apoptosis and cytokine maturation during septic shock, caspase-12 appears to mediate endoplasmic reticulum-specific apoptosis and cytotoxicity by amyloid-β. Caspase-14 is highly expressed in embryonic tissues but not in adult tissues.41

Caspases are normally produced as inactive zymogens containing an N-terminal prodomain followed by a large and a small subunit that constitute the catalytic core of the protease. They have been categorized into two distinct classes: initiator and effector caspases. The upstream initiator caspases contain long N-terminal prodomains and one of the two characteristic protein–protein interaction motifs: the death effector domain (DED; caspase-8 and -10) and the caspase activation and recruitment domain (CARD; caspase-1, -2, -4, -5, -9, and -12). The downstream effector caspases (caspase-3, -6, and -7) are characterized by the presence of a short prodomain. Apart from the structural differences, a prominent difference between initiator and effector caspases is their basal state. Both the zymogen and the activated forms of effector caspases exist as constitutive homodimers, whereas initiator caspase-9 exists predominantly as a monomer both before and after proteolytic processing.42 Initiator caspase-8 has been reported to exist in equilibrium between monomers and homodimers.43 Although the initiator caspases are capable of autocatalytic activation, the activation of effector caspases requires formation of oligomeric complexes with their adapter proteins and often intrachain cleavage within the initiator caspase.

Caspases have also been divided into three categories based on substrate specificity. Group I members (caspase-1, -4, and -5) have a substrate specificity for the WEHD sequence with high promiscuity; group II members (caspase-2, -3, -7 and CED-3) prefer the DEXD sequence and have an absolute requirement for aspartate (D) at P4; and members of group III (caspase-6, -8, and -9 and the "aspase" granzyme B) have a preference for the (I/L/V)EXD sequences. Several reports have suggested a role for group I members in inflammation and that of group II and III members in apoptotic signaling events.

The roles of various caspases in apoptotic pathways and their relative importance for animal development have been examined in genetic studies involving knockout of different caspase genes (Table 4). A caspase-1 (interleukin-1β-converting enzyme, ICE) knockout study suggested that ICE plays an important role in inflammation
by activating cytokines such as IL-1β and IL-18. However, caspase-1 was not required to mediate apoptosis under normal circumstances and did not have a major role during development. Interestingly, ischemic brain injury was significantly reduced in caspase-1 knockout mice compared with wild-type mice, suggesting that inflammation may contribute to ischemic injury. Caspase-3 deficiency leads to impaired brain development and premature death. Also, functional caspase-3 is required for some typical hallmarks of apoptosis such as formation of apoptotic bodies, chromatin condensation, and DNA fragmentation in many cell types. Lack of caspase-8 results in the death of embryos at day 11 with abnormal formation of the heart, suggesting that caspase-8 is required for cell death during mammalian development. In support of this finding, knockout of FADD, which is required for caspase-8 activation, resulted in fetal death with signs of abdominal hemorrhage and cardiac failure. Moreover, caspase-8-deficient cells did not die in response to signals from members of the TNF receptor family. However, cells lacking either FADD or caspase-8, which are resistant to TNF-α-mediated or CD95-mediated death, are susceptible to chemotherapeutic drugs, serum deprivation, ceramide, γ-irradiation, and dexamethasone-induced killing. In contrast, caspase-9 has a key role in apoptosis induced by intracellular activators, particularly those that cause DNA damage. Deletion of caspase-9 resulted in perinatal lethality, apoptotic failure in developing neurons, enlarged brains, and craniofacial abnormalities. In caspase-9-deficient cells, caspase-3 was not activated, suggesting that caspase-9 is upstream of caspase-3 in the apoptotic cascade. As a consequence, caspase-9-deficient cells are resistant to dexamethasone or irradiation, whereas they retain their sensitivity to TNF-α-induced or CD95-induced death because of the presence of caspase-8, the initiator caspase involved in death receptor signaling that can also activate caspase-3. Overall, these observations support the idea that different death signaling pathways converge on downstream effector caspases (Fig. 4.2). Indeed, caspase-3 is regarded as one of the key executioner molecules activated by apoptotic stimuli originating either at receptors for exogenous molecules or within cells through the action of drugs, toxins, or radiation.

A cascade of caspases plays the central executioner role by cleaving various mammalian cytosolic and nuclear proteins that play roles in cell division, maintenance of cytoskeletal structure, DNA replication and repair, RNA splicing, and other cellular processes. This proteolytic carnage produces the characteristic morphological changes of apoptosis. Once the caspase cascade is initiated, the process of cell death has crossed the point of no return.

### Extrinsic Death Pathway

The extrinsic pathway involves binding of death ligands such as tumor necrosis factor-α (TNF-α), CD95 ligand (Fas ligand), and TNF-related apoptosis-inducing ligand (TRAIL) to their cognate cell-surface receptors TNFR1, CD95/Fas, TRAIL-R1, TRAIL-R2, and the DR series of receptors, resulting in the activation of initiator caspase-8 (also known as FADD-homologous ICE/CED-3-like protease, or FLICE) and subsequent activation of effector caspase-3 (see Fig. 4.2). The cytoplasmic domains of death receptors contain the “death domain,” which plays a crucial role in transmitting the signal from the cell’s surface to intracellular signaling molecules. Binding of the ligands to their cognate receptors results in receptor trimerization and recruitment of adapter proteins to the cell membrane, which involves homophilic interactions between death domains of the receptors and the adapter proteins. The adapter protein for the receptors TNFR1 and DR3 is TNFR-associated death domain protein...
Fig. 4.2. Schematic representation of apoptotic signaling pathways. **Extracellular Pathway:** Following the binding of peptides such as tumor necrosis factor (TNF)-α or Fas ligand (FASL), the receptors oligomerize and recruit adapter proteins [Fas-associated death domain (FADD), tumor necrosis factor receptor (TNFR)-associated death domain (TRADD)] to form death-inducing signaling complexes, causing the activation of the initiator caspase-8, which sequentially activates effector caspsases (e.g., caspase-3). Other adapter proteins (FLASH), inhibitory proteins (FLIP), or proteins involved in survival pathways as well as death mechanisms (receptor interaction protein, RIP) may participate in complex mechanisms that determine life or death. The TNF-α-TNFR1 complex can also elicit an antiapoptotic response by recruiting TRAF2, which results in NF-κB-mediated upregulation of antiapoptotic genes. In cytotoxic T-lymphocyte-induced death, granzyme B, which enters the cell through membrane channels formed by the protein perforin, activates caspsases by cleaving them directly or indirectly. **Intracellular Pathway:** Lack of survival stimuli (withdrawal of growth factor, hypoxia, genotoxic substances, etc.) is thought to generate apoptotic signals through poorly defined mechanisms, which lead to translocation of proapoptotic proteins such as Bax to the outer mitochondrial membrane. In some cases, transcription mediated by p53 may be required to induce proteins such as Bax. Translocated Bax undergoes conformational changes in the outer membrane to form oligomeric structures (pores) that leak cytochrome c from mitochondria into the cytosol. Formation of a ternary complex of cytochrome c, the adapter protein Apaf-1, and the initiator caspase-9 called the “apoptosome” results in the activation of caspase-9, followed by sequential activation of effector caspase(s) such as caspase-3 and others. The actions of caspsases, endonucleases, and possibly other enzymes lead to cellular disintegration. For example, the endonuclease CAD (caspase-activated DNAse) becomes activated when it is released from its inhibitor ICAD upon cleavage of ICAD by an effector caspase. Antiapoptotic proteins such as Bcl-2 and Bcl-xL inhibit the membrane-permeabilizing effects of Bax and other proapoptotic proteins. Cross-talk between extra- and intracellular pathways occurs through caspase-8-mediated Bid cleavage, which yields a 15-kDa protein that migrates to mitochondria and releases cytochrome c, thereby setting in motion events that lead to apoptosis via caspase-9.
Upon recognizing the infected cells, the lymphocytes bind and secrete granules onto the surface of infected cells. Perforin then assembles into transmembrane channels to allow the entry of granzyme B into the target cell. Upon entry, granzyme B, which cleaves after aspartate residues in proteins (“asparate”), activates one or more of the apoptotic proteases (caspase-2, -3, -7, -8, and -10) to trigger the proteolytic death cascade (see Fig. 4.2). Fas ligand/Fas and perforin/granzyme B systems are the main apoptotic machinery that regulates homeostasis in immune cell populations.

**Intrinsic Death Pathway**

Cells can respond to various stressful stimuli and metabolic disturbances by triggering apoptosis. Drugs, toxins, heat, radiation, hypoxia, and viral infections are some of the stimuli known to activate death pathways. Cell death, however, is not necessarily inevitable after exposure to these agents, and the mechanisms determining the outcome of the injury are a topic of active interest. The current consensus appears to be that it is the intensity and the duration of the stimulus that determine the outcome. The stimulus must go beyond a threshold to commit cells to apoptosis. Although the exact mechanism used by each stimulus may be unique and different, a few broad patterns can be identified. For example, agents that damage DNA, such as ionizing radiation and certain xenobiotics, lead to activation of p53-mediated mechanisms that commit cells to apoptosis, at least in part through transcriptional upregulation of proapoptotic proteins. Other stresses induce increased activity of stress-activated protein kinases, which results ultimately in apoptotic commitment. These different mechanisms converge in the activation of caspases.

All stimuli described above change the outer mitochondrial membrane permeability and release two main groups of normally sequestered proapoptotic proteins from the intermembrane space into the cytosol. The first group consists of cytochrome c, Smac/DIABLO, and the serine protease HtrA2/Omi. Cytochrome c binds the adapter protein, Apaf-1, and procaspase-9 to form an “apoptosome” complex and activates caspases. The clustering of procaspase-9 in the apoptosome complex leads to caspase-9 activation. IAPs (inhibitors of apoptosis proteins) in the cell block activate caspase-9 and cause the activation of downstream effector caspases. Smac/DIABLO and HtrA2/Omi are reported to promote apoptosis by inhibiting the activity of IAPs. The second group of proapoptotic proteins, AIF, endonuclease G, and CAD (caspase-activated DNase), are released from the mitochondria during apoptosis, but this is a late event that occurs after the cell dies. The control and regulation of these mitochondrial events during apoptosis occurs through members of the Bcl-2 family of proteins, as discussed next.

In *C. elegans*, biochemical and genetic studies have indicated a role for CED-4 upstream of CED-3. Upon receiving death commitment signals, CED-4 binds to pro-CED-3 and releases active CED-3. However, when overexpressed,
CED-9 can inhibit the activation of pro-CED-3 by binding to CED-4 and sequestering it away from pro-CED-3. Therefore, CED-3 and CED-4 are involved in activation of apoptosis, and CED-9 inhibits apoptosis. After the discovery of caspases as CED-3 homologues, a search for activators and inhibitors analogous to CED-4 and CED-9 led to the discovery of diverse mammalian regulators of apoptosis. The plethora of these molecules and their functional diversity allowed them to be classified into four broad categories: (1) adapter proteins, (2) the Bcl-2 family of regulators, (3) inhibitors of apoptosis (IAPs), and (4) other regulators.

Adapter Proteins

As stated earlier, two major pathways of apoptosis, involving either the initiator caspase-8 or the initiator caspase-9 (see Fig. 4.2), have been recognized. Signaling by death receptors (CD95, TNFRI) occurs through a well-defined process of recruitment of caspase-8 to the death receptor by adapter proteins such as FADD. Recruitment occurs through interactions between the death domains that are present on both receptor and adapter proteins. Receptor-bound FADD then recruits caspase-8 through interactions between DEDs common to both caspase-8 and FADD, forming a DISC. In the DISC, caspase-8 activation occurs through oligomerization and autocatalysis. Activated caspase-8 then activates downstream caspase-3, culminating in apoptosis. The inhibitory protein FLIP was shown to block Fas-induced and TNF-α-induced DISC formation and subsequent activation of caspase 8.61 Of particular interest is cellular FLIP, which stimulates caspase-8 activation at physiologically relevant levels and inhibited apoptosis upon high ectopic expression.73 Cellular FLIP contains two DEDs that can compete with caspase-8 for recruitment to the DISC; this limits the degree of association of caspase-8 with FADD and thus limits activation of the caspase cascade. It also forms a heterodimer with caspase-8 and caspase-10 through interactions between both the DEDs and the caspase-like domains of the proteins, thus activating both caspase-8 and caspase-10.74 Apoptotic protease activating factor-1 (Apaf-1), a CED-4 homologue in mammalian cells, affects the activation of initiator caspase-9.75 This factor binds to procaspase-9 in the presence of cytochrome c and 2′-deoxyadenosine 5′-triphosphate (dATP) or adenosine triphosphate (ATP) and activates this protease, which in turn activates a downstream cascade of proteases (see Fig. 4.2).69 By and large, Apaf-1 deficiency is embryonically lethal, and the embryos exhibit brain abnormalities similar to those seen in caspase-9 knockout mice.76 These genetic findings support the idea that Apaf-1 is coupled to caspase-9 in the death pathway. Unlike CED-4 in nematodes, Apaf-1 requires the binding of ATP and cytochrome c to activate procaspase-9. The multiple WD40 repeats in the C-terminal end of Apaf-1 have a regulatory role in the activation of caspase-9.77

The Bcl-2 Family of Proteins

The CED-9 homologue in mammals is the Bcl-2 protein. Bcl-2 was first discovered in B-cell lymphoma as a proto-oncogene. Overexpression of Bcl-2 was shown to offer protection against a variety of death stimuli.79 The Bcl-2 protein family includes both proapoptotic (Bcl-2, Bcl-xL, Bcl-w, Mcl-1, Nr13, and A1/Bfl1) and antiapoptotic proteins (Bax, Bak, Bok, Diva, Bcl-Xs, Bik, Bim, Hrk, Nip3, Nix, Bad, and Bid).71,79 These proteins are characterized by the presence of Bcl-2 homology (BH) domains: BH1, BH2, BH3, and BH4 (Fig. 4.3). The proapoptotic members have two subfamilies: a multidomain and a BH3-only group (see Fig. 4.3). The relative ratio of pro- and antiapoptotic proteins determines the sensitivity of cells to various apoptotic stimuli. The best studied proapoptotic members are Bax and Bid. Exposure to various apoptotic stimuli leads to translocation of cytosolic Bax from the cytosol to the mitochondrial membrane.80 Bax oligomerizes on the mitochondrial membrane along with another proapoptotic protein, Bak, leading to the release of cytochrome c from the mitochondrial membrane into the cytosol.81

Other proapoptotic proteins, mainly the BH3-only proteins, are thought to aid in Bax–Bak oligomerization on the mitochondrial membrane. The antiapoptotic Bcl-2 family members are known to block Bax–Bak oligomerization on the mitochondrial membrane and subsequent release of cytochrome c into the cytosol.81,82 After release from the mitochondria, cytochrome c is known to interact with the WD40 repeats of the adapter protein Apaf-1, resulting in the formation of the apoptosome complex. Seven molecules of Apaf-1, interacting through their N-terminal caspase activation and recruitment domain, form the central hub region of the symmetrical wheel-like structure, the apoptosome. Binding of ATP/dATP to Apaf-1 triggers the formation of the apoptosome, which subsequently recruits procaspase-9 into the apoptosome complex, resulting in its activation.83 Activated caspase-9 then activates executioner caspases, such as caspase-3 and caspase-7, eventually leading to programmed cell death. However, for the caspase-9 to be active, it must reside on the apoptosome complex.84

Inhibitors of Apoptosis Proteins

The IAPs, first discovered in baculoviruses and then in insects and Drosophila, inhibit activated caspases by directly binding to the active enzymes.85 These proteins contain one or more baculovirus inhibitors of apoptosis repeat domains, which are responsible for the caspase inhibitory activity.86 To date, eight mammalian IAPs have been identified: they include X-linked IAP (XIAP), c-IAP1, c-IAP2, Melanoma IAP (ML-IAP)/Livin, IAP-like protein-2 (ILP-2), neuronal apoptosis-inhibitory protein (NAIP), Bruce/Apollon, and Survivin. In mammals, caspase-3, -7, and -9 are inhibited by IAPs.83 There are reports suggesting aberrant expression of IAPs in many cancer tissues. For example, cIAP1
Apoptosis and Cell Death

is overexpressed in esophageal squamous cell sarcoma; the cIAP2 locus is translocated in mucosa-associate lymphoid lymphoma; and Survivin has been shown to be upregulated in many cancer cells.

Other Regulators

The caspase inhibitory activity of IAPs is inhibited by proteins containing an IAP-binding tetrapeptide motif. The founding member of this family is Smac/DIABLO, which is released from the mitochondrial intermembrane space into the cytosol during apoptosis. In the cytosol, it interacts with several IAPs and inhibits their function. The other mitochondrial protein, Omi/HtrA2, is also known to antagonize XIAP-mediated inhibition of caspase-9 at high concentrations. A serine protease, Omi/HtrA2, can proteolytically cleave and inactivate IAP proteins and thus is considered to be a more potent suppressor of IAPs than Smac.

It has been reported that the heat shock proteins Hsp90, Hsp70, and Hsp27 can inhibit caspase activation by cytochrome c by interacting with either Apaf-1 or other players in the pathway. A high-throughput screen identified a compound called PETCM [α-(trichloromethyl)-4-pyridineethanol] as a caspase-3 activator. Further work with PETCM revealed its involvement in apoptosis regulation. This pathway also includes oncoprotein prothymosin-α and tumor suppressor putative HLA-DR-associated proteins. These proteins were shown to promote caspase-9 activation after apoptosisome formation, whereas prothymosin-α inhibited caspase-9 activation by inhibiting apoptosisome formation.

Protein Targets of Caspases

In an apoptotic cell, the regulatory, structural, and housekeeping proteins are the main targets of the caspases. The regulatory proteins mitogen-activated protein/extracellular signal-regulated kinase kinase-1, p21-activated kinase-2, and Mst-1 are activated upon cleavage by caspases. Caspase-mediated protein hydrolysis inactivates other proteins, including focal adhesion kinase, phosphatidylinositol-3 kinase, Akt, Raf-1, IAPs, and inhibitors of caspase-activated DNAse (ICAD). Caspases also convert the antiapoptotic protein Bcl-2 into a proapoptotic protein such as Bax upon cleavage. There are many structural protein targets of caspases, which include nuclear lamins, actin, and regulatory proteins such as spectrin, gelsolin, and fodrin.

Degradation of nuclear DNA into internucleosomal chromatin fragments is one of the hallmarks of apoptotic cell death that occurs in response to various apoptotic stimuli in a wide variety of cells. A specific caspase-activated DNase (CAD), that cleaves chromosomal DNA in a caspase-dependent manner, is synthesized with the help of ICAD. In proliferating cells, CAD is always found to be associated with ICAD in the cytosol. When cells are undergoing apoptosis, caspases (particularly caspase-3) cleave ICAD to release CAD and allow its translocation to the nucleus to cleave...
chromosomal DNA. Thus, cells that are ICAD deficient or which express caspase-resistant ICAD mutant do not exhibit DNA fragmentation during apoptosis.

**Conclusion**

Cell death has become an area of intense interest and investigation in science and medicine because of the recognition that cell death, in general, and apoptosis, in particular, are important features of many biological processes. Apoptosis is a carefully regulated energy-dependent process in which caspase activation plays a central role. Although many of the key apoptotic proteins have been identified, the molecular mechanisms of their action are not yet fully understood and thus continue to be a research focus. Because apoptosis is a vital component of both health and disease that is initiated by various physiological and pathological stimuli, as well as having widespread involvement in the pathophysiology of disease, it lends itself to therapeutic intervention.

**References**

33. Nijhawan R, Rajwanshi A, Gautam U, Gupta SK. Cytoplasmic vacuolation, intracytoplasmic lumina, and DPAS stain-
4. Apoptosis and Cell Death


The Role of Mutation and Epimutation in the Development of Human Disease

Ashley G. Rivenbark and William B. Coleman

Introduction

The paradigm of disease causation holds that disease represents the manifestation or several manifestations of an underlying process that has one or more root causes. Thus, human diseases reflect a spectrum of pathologies and mechanisms of disease pathogenesis. The general categories of disease affecting humans include (a) hereditary diseases, (b) infectious diseases, (c) inflammatory diseases, and (d) neoplastic diseases. Pathological conditions representing each of these general categories have been described for every tissue in the body. Despite the grouping of diseases by the common features of the general disease type, the pathogenesis of each of the various diseases is unique, and in some cases multiple mechanisms can give rise to a similar pathology (disease manifestation). Disease causation may be related to intrinsic factors or extrinsic factors, but many or most diseases are multifactorial, involving a combination of intrinsic and extrinsic factors. It is now well recognized that most major diseases are ultimately the result of aberrant gene expression and that susceptibility to disease is significantly influenced by patterns of gene expression in target cells or tissues for a particular type of pathology. It follows that gene mutations and other genetic alterations are important in the pathogenesis of many human diseases. Similarly, nongenetic alterations affecting the expression of key genes, called epimutations, may also contribute to the genesis of disease at many tissue sites. In this chapter, general concepts related to the molecular basis for the major disease types are reviewed. This review is not intended to be comprehensive. Rather, the current state of understanding related to the genes and molecular mechanisms (genetic and epigenetic) that contribute to illustrative diseases is described.

Mutation and Epimutations

Mutation refers to changes in the genome that are characterized by alteration in the nucleotide sequence of a specific gene and/or other alterations at the level of the primary structure of DNA. Point mutations, insertions, deletions, and chromosomal abnormalities are all classified as mutations. In contrast, epimutation refers to alterations in the genome that do not involve changes in the primary sequence of the DNA. Aberrant DNA hypermethylation or hypomethylation and/or abnormal histone modifications resulting in alterations in chromatin structure are considered epimutations. Despite the differences between mutation and epimutation, the consequences of these molecular processes on the normal expression or function of critical genes and proteins may be the same: alteration of normal gene expression and/or normal protein function. These alterations may reflect (a) loss or reduction of normal levels of gene expression with consequent loss of protein function, (b) loss of function caused by loss of protein or synthesis of defective protein, (c) increased levels of gene expression with consequent overexpression of protein, or (d) gain-of-function mutation with consequent altered protein function. Although many human diseases can be attributed to genetic alteration or epimutation affecting a single gene (or a few genes), the actual molecular consequences of these changes can be very dramatic, resulting in major alterations in gene expression patterns secondary to the primary genetic or epigenetic gene defect.

Genetic Alterations

Disease-related genetic alterations can be categorized into two major groups: nucleotide sequence abnormalities and chromosomal abnormalities. Examples of both these forms
of molecular lesion have been characterized in familial and acquired diseases affecting various human tissues.

Nucleotide sequence alterations include changes in individual genes involving single nucleotide changes (missense and nonsense) and small insertions or deletions (some of which result in frameshift mutations). Single nucleotide alterations that involve a change in the normal coding sequence of the gene (point mutations) can give rise to an alteration in the amino acid sequence of the encoded protein. Missense mutations alter the translation of the affected codon, whereas nonsense mutations alter codons that encode amino acids to produce stop codons. This action results in premature termination of translation and the synthesis of a truncated protein product. Small deletions and insertions are typically classified as frameshift mutations because deletion or insertion of a single nucleotide (for instance) will alter the reading frame of the gene on the 3′-side of the affected site. This alteration can result in the synthesis of a protein that bears very little resemblance to the normal gene product or production of an abnormal/truncated protein because of the presence of a stop codon in the altered reading frame. In addition, deletion or insertion of one or more groups of three nucleotides will not alter the reading frame of the gene, but will alter the resulting polypeptide product, which will exhibit either loss of specific amino acids or the presence of additional amino acids within its primary structure.

Chromosomal alterations include the gain or loss of one or more chromosomes (aneuploidy), chromosomal rearrangements resulting from DNA strand breakage (translocations, inversions, and other rearrangements), and gain or loss of portions of chromosomes (amplification, large-scale deletion). The direct result of chromosomal translocation is the movement of a segment of DNA from its natural location into a new location within the genome, which can result in altered expression of the genes that are contained within the translocated region. If the chromosomal breakpoints utilized in a translocation are located within structural genes, then hybrid (chimeric) genes can be generated. The major consequence of a chromosomal deletion (involving a whole chromosome or a large chromosomal region) is the loss of specific genes that are localized to the deleted chromosomal segment, resulting in changes in the copy number of the affected genes. Similarly, gain of chromosome number or amplification of chromosomal regions results in an increase in the copy numbers of genes found in these chromosomal locations.

Epigenetic Alterations

In contemporary terms, epigenetics refers to modifications of the genome that are heritable during cell division but do not involve a change in the DNA sequence. Therefore, epigenetics describes heritable changes in gene expression that are not simply attributable to nucleotide sequence variation. It is now recognized that epigenetic regulation of gene expression reflects contributions from both DNA methylation as well as complex modifications of histone proteins and chromatin structure. Nonetheless, DNA methylation plays a central role in nongenomic inheritance and in the preservation of epigenetic states and remains the most accessible epigenomic feature due to its inherent stability. Thus, DNA methylation represents a target of fundamental importance in the characterization of the epigenome and for defining the role of epigenetics in disease pathogenesis.

Diseases Associated with Chromosomal Alterations

Chromosomal Alterations in Cancer

Chromosomal alterations have been widely documented in the various forms of cancer, and some of these chromosomal aberrations are recognized to be causally related to cancer induction and/or progression. In some cases, these alterations represent loss of individual chromosomes, chromosome arms, or specific chromosomal segments, consistent with deletion of a tumor suppressor locus. In other cases, these alterations represent gain of individual chromosomes or specific chromosomal segments, consistent with activation of a positive mediator of cell proliferation. In addition, numerous complex chromosomal rearrangements (such as translocation) have been characterized in certain forms of cancer. In each of these cases, the result is alteration of gene dose (either as gene loss or gene amplification) and function (altered product or altered expression). Cyto genetic methods to characterize chromosomal alterations have been applied for many years, but recent developments have automated this process and opened the analysis to difficult samples (solid tumors). For instance, array-based comparative genomic hybridization (CGH) is a technique that analyzes global genomic changes by documenting gains and losses of chromosomal regions in diseases such as cancer. Recent studies have shown the potential of array CGH in detecting copy number alterations in gastric cancer, chronic lymphocytic leukemia, fallopian tube carcinomas, oral squamous carcinoma, bladder cancer, and pancreatic cancer. A few well-characterized examples from major human cancers are given in the sections that follow.

Gene Amplifications in Cancer

Several chromosomal regions have been characterized to be overrepresented in lung cancer, suggestive of amplification of chromosomal regions. One such frequently overrepresented region is chromosome 8q, which harbors the c-myc proto-oncogene. The c-myc proto-oncogene is a member of the basic helix-loop-helix superfamily of nuclear transcription factors. The myc protein heterodimerizes with max, and the resulting myc–max protein complex binds to and transcriptionally activates genes that contain a CAGCTG consensus binding sequence.
has been reported for both small cell lung carcinoma (SCLC) and non-small cell lung carcinoma (NSCLC).\textsuperscript{23–27} Similarly, amplification of the c-myc locus has been observed in both SCLC and NSCLC,\textsuperscript{28} but it may be more prevalent in SCLC.\textsuperscript{29} Thus, overexpression of c-myc in lung cancer is the frequent consequence of gene amplification at \(8q24.\textsuperscript{30–32}\) Gene amplification and overexpression of c-myc occur more frequently in advanced neoplasms and metastatic lesions, suggesting a role for this event in tumor progression, and partially explaining the significant correlation between c-myc amplification and poor prognosis.\textsuperscript{31} In contrast to other cancers (such as lymphoma), point mutation of c-myc and c-myc gene translocation related to specific chromosomal alterations have not been reported in lung cancer.

Several studies have examined chromosomal alterations in breast cancer using comparative genomic hybridization approaches.\textsuperscript{33–35} These studies (as well as those using cytogenetic techniques) have identified a number of gene amplifications in breast cancer. For instance, the \(TM4SF1\) gene is located at \(3q25.1\) and is amplified frequently in BRCA1-related breast cancers.\textsuperscript{36} Similarly, the \(CYP24\) gene that is located at \(20q13.2\) is subject to recurrent amplification in breast cancer.\textsuperscript{37} Amplification of the \(Her2/neu\) gene on \(17q\) was first detected using Southern blotting\textsuperscript{38} and is thought to occur in approximately 20–25% of breast cancer patients.\textsuperscript{39,40} This gene amplification event results in multiple copies of the \(Her2/neu\) gene (>5 copies per cell, but perhaps as many as 50–100 copies per cell), and the increase in gene dose correlates with increased mRNA expression and overexpression of the \(p185/erbB2\) protein. \(Her2/neu\) is a member of the epidermal growth factor receptor family that also includes \(Her1, Her3,\) and \(Her4\). Breast cancers that are positive for \(Her2/neu\) amplification and overexpression of \(p185/erbB2\) protein have a poor prognosis and are inherently resistant to common chemotherapeutic regimens.\textsuperscript{41} However, the \(p185/erbB2\) protein product of the \(Her2/neu\) gene is the target for the monoclonal antibody trastuzumab (Herceptin), and breast cancer patients who have \(Her2/neu\) amplification significantly benefit from combination therapies that include trastuzumab, resulting in improved survival.\textsuperscript{42,43}

**Common Chromosomal Deletions in Cancer**

Allelotype studies of lung cancer have identified several recurring chromosomal deletions. In SCLC, frequent loss of heterozygosity (LOH) occurs at \(3p\) (91%), \(5q\) (71%), \(13q\) (96%), \(17p\) (88%), and \(22q\) (73%).\textsuperscript{44} NSCLCs display frequent LOH at \(2q\) (68%), \(3p\) (82%), \(5q\) (60%), \(9p\) (79%), \(12q\) (63%), \(13q\) (67%), \(17p\) (89%), \(18q\) (86%), and \(22q\) (75%).\textsuperscript{35} However, distinct differences in the patterns of chromosomal deletion have been noted for the histological subtypes of NSCLC. Among squamous cell carcinomas, frequent LOH was noted for \(3p\) (82%), \(9q\) (67%), \(13q\) (60%), and \(17p\) (88%).\textsuperscript{46} In contrast, fewer chromosomal losses were noted among adenocarcinomas, with 51% LOH at \(17p\) representing the most frequent alteration.\textsuperscript{46} In other studies, similar findings have been reported.\textsuperscript{47–49} Deletions affecting a specific chromosomal region (as measured by LOH) may be indicative of the presence of a tumor suppressor gene (or other negative mediator of cell proliferation) at that chromosomal location. Among lung cancers, frequent LOH affecting \(3p, 5q, 13q, 17p,\) and \(22q\) occur in both SCLC and NSCLC. Several regions of chromosome \(3p\) have been implicated in lung cancer, including \(3p12–p14, 3p21,\) and \(3p25.\textsuperscript{50–52}\) These observations suggest that there may be three or more tumor suppressor genes on human chromosome \(3p.\)\textsuperscript{52} Candidate tumor suppressor genes from chromosome \(3p\) include \(FHIT\) at \(3p12–p14\)\textsuperscript{53} and \(RASSF1\) at \(3p21.\)\textsuperscript{54} LOH at chromosome \(5q\) typically corresponds to loss at \(5q13–q21.\)\textsuperscript{55} A number of genes map to this chromosomal region, including \(MCC\) (for mutated in colorectal cancer) and \(APC\) (for adenomatous polyposis coli).\textsuperscript{56–59} Although neither of these genes has been shown to be mutated in lung cancer,\textsuperscript{62} frequent LOH at this chromosomal region suggests the involvement of one of these or other candidate genes localized to this region in the molecular pathogenesis of lung cancer. The tumor suppressor gene \(Rb\) localizes to chromosome \(13q14.1.\)\textsuperscript{61,62} The expression of \(Rb\) is altered in a significant percentage of primary lung cancer.\textsuperscript{63,64} The \(p53\) tumor suppressor gene is located at \(17p13.1.\)\textsuperscript{65} and is lost as a result of chromosomal deletion of this region in all lung cancer types.\textsuperscript{66} The precise nature of the putative lung cancer tumor suppressor locus at chromosome \(22q\) is not yet defined. However, a candidate gene, termed \(SEZ6L,\) has been localized to \(22q12.1\) and shown to be mutated in a SCLC cell line.\textsuperscript{67}

A number of chromosomal regions demonstrate LOH in breast cancer, most commonly affecting \(1p, 3p, 6q, 7q, 9p, 11p, 13q, 16q, 17p, 17q, 18q,\) and \(22q.\)\textsuperscript{68–75} Many of these chromosomal locations harbor known or candidate tumor suppressor genes. BRCA1 localizes to \(17q21.\)\textsuperscript{76} and BRCA2 localizes to \(13q12–q13.\)\textsuperscript{77} Other regions of \(17p\) may also contain tumor suppressor loci that are important in breast cancer.\textsuperscript{78} For instance, the \(p53\) tumor suppressor gene is located at \(17p13.1.\)\textsuperscript{15} and has been implicated in the genesis of some breast cancers. Similarly, genes located in other regions of \(13q\) (including the \(Rb\) locus at \(13q14.1\))\textsuperscript{51,62} may be important in breast carcinogenesis. Chromosome \(16q\) harbors the \(CDH1\) gene, which encodes E-cadherin.\textsuperscript{79} Loss of \(E-cadherin\) has been implicated in breast cancer metastasis.\textsuperscript{80} Other known tumor suppressor genes that localize to chromosomal regions which demonstrate LOH in breast cancer include the \(DCC\) gene at \(18q.\)\textsuperscript{81} Multiple regions of chromosome \(1\) and \(11\) are affected by LOH in breast cancers, suggesting the possibility of multiple distinct tumor suppressor loci on each chromosome.\textsuperscript{82–84} Several studies have examined chromosomal alterations in ductal carcinoma in situ (DCIS) in an attempt to identify genetic changes occurring in the preneoplastic breast that may contribute to malignant conversion. These studies found frequent LOH affecting various chromosomal regions including \(6p, 13q, 16q, 17p,\) and \(17q\) in DCIS.\textsuperscript{85,86} These results are consistent with the idea that
known tumor suppressor genes (such as BRCA1, BRCA2, and p53) may be involved in the early stages of breast carcinogenesis.

**Complex Chromosomal Rearrangements in Cancer**

Chromosomal translocations have been studied from many years in hematopoietic malignancy. In fact, the first consistent chromosomal aberration observed in human cancer was the so-called Philadelphia chromosome, which was described in chronic myelogenous leukemia.\(^{87,88}\) This chromosomal aberration was later found to represent a balanced translocation between chromosomes 9 and 22 [t(9;22)(q34.1;q11.2)].\(^{89}\) Since then, numerous chromosomal translocations have been described in leukemia and lymphoma.

A major chromosomal translocation in acute myelogenous leukemia (AML) involves chromosomes 15 and 17. The resulting chromosomal abnormality [t(15;17) (q21;q21)], which occurs exclusively in acute promyelocytic leukemia (APL), is a balanced and reciprocal translocation in which the PML (for promyelocytic leukemia) gene on chromosome 15 and the RAR\(\alpha\) gene on chromosome 17 are disrupted and fused to form a hybrid gene.\(^{90,91}\) The PML–RAR\(\alpha\) fusion gene (located on chromosome 15) encodes a chimeric mRNA and a novel protein. On the derivative chromosome 15, both the PML and RAR\(\alpha\) genes are oriented in a head-to-tail orientation. There are three major forms of the PML–RAR\(\alpha\) fusion gene, corresponding to different breakpoints in the PML gene.\(^{92-94}\) The breakpoint in the RAR\(\alpha\) gene occurs in the same general location in all cases, involving the sequences within intron 2. Approximately 40–50% of cases have a PML breakpoint in exon 6 (the long form, termed bcr1), 40–50% of cases have the PML breakpoint in exon 3 (the short form, termed bcr3), and 5–10% of cases have a breakpoint in PML exon 6 that is variable (the variable form, termed bcr2). In each form of the translocation, the PML–RAR\(\alpha\) fusion protein retains the 5’-DNA-binding and dimerization domains of PML and the 3’-DNA-binding, heterodimerization, and ligand (retinoic acid)-binding domains of RAR\(\alpha\). Recent studies indicate that the different forms of PML–RAR\(\alpha\) fusion mRNA correlate with clinical presentation or prognosis. In particular, the bcr3 type of PML–RAR\(\alpha\) correlates with higher leukocyte counts at time of presentation.\(^{93,94}\) Both higher leukocyte counts and variant morphology are adverse prognostic findings, and the bcr3 type of PML–RAR\(\alpha\) does not independently predict poorer disease-free survival.\(^{94}\) The presence of the t(15;17) translocation consistently predicts responsiveness to a specific treatment utilizing all-trans-retinoic acid (ATRA). Retinoic acid is a ligand for the retinoic acid receptor (RAR). The gene encoding the RAR is involved in the t(15;17) chromosomal abnormality. ATRA has been suggested to function by overcoming the blockade of myeloid cell maturation, allowing the neoplastic cells to mature (differentiate).\(^{95,96}\)

Progress toward characterization of complex chromosomal rearrangements in solid tumors was hindered by technical limitations until recently, when spectral karyotyping became available. A number of studies using spectral karyotyping of various human cancers, including lung cancer, have now appeared.\(^{48,97,98}\) These studies identified a number of unbalanced chromosomal translocations, in many cases involving some of the same chromosomal regions that are frequently deleted in lung cancer. These complex chromosomal rearrangements may alter the structure or expression of genes localized to the affected chromosomal regions. However, additional investigation will be required to characterize the molecular consequences associated with specific chromosomal rearrangements cancers of solid tissues, such as breast, prostate, colon, and lung.

**Chromosomal Alterations in Nonneoplastic Diseases**

Chromosomal alterations are not limited to neoplastic diseases. Various forms of chromosomal abnormality have been documented in a wide variety of human genetic diseases. A few examples of diseases that are associated with gene inversions, gene deletions, and gene duplications are provided here.

Structural variants affecting specific chromosomes have been identified in the general population to be the cause of genetic disease in the offspring of parents who exhibit certain DNA inversions. Williams–Beuren syndrome has an incidence of 1 in 20,000–50,000 births and is associated with a 1.5-Mb inversion at 7q11.23.\(^{99}\) This chromosomal abnormality occurs with a 5% frequency in the general population.\(^{100}\) Angelman syndrome has an incidence of 1 in 10,000 to 20,000 births and is associated with a 4-Mb inversion at 15q11.2.\(^{101}\) This chromosomal abnormality occurs in 9% of the general population.\(^{102}\) The genetic inversions associated with some disease states are not detected in the general population.\(^{103}\) Patients with hemophilia A carry a 400 Kb inversion in intron 22 of the factor VIII gene, and two copies are found in an inverted orientation.\(^{104}\) This genetic inversion is the product of a nonallelic homologous recombination event that results in the inactivation of the factor VIII gene.\(^{105}\) A smaller genetic inversion in the emerin gene has been identified associated with Emery–Dreifuss muscular dystrophy.\(^{106}\) Hunter syndrome is an X-linked dominant disorder in which nonallelic homologous recombination between the iduronate 2-sulfatase gene (IDS) and an IDS pseudogene generates a genomic inversion, resulting in a disruption of the functional IDS gene. This genetic inversion occurs in approximately 13% of Hunter syndrome patients.\(^{107,108}\) Soto syndrome is a microdeletion syndrome. Fathers of Soto syndrome patients carry a 1.9-Mb inversion at 5q35 that predisposes their offspring to this disease.\(^{109}\) Polymorphic inversion at olfactory receptor gene clusters found at 4p16 and 8p23 occur at frequencies of 12.5% and 26%, respectively.\(^{110}\) These inversion
events predispose affected individuals to a form of genetic instability characterized by constitutional chromosomal translocations. Heterozygous carriers of these translocations exhibit no phenotypic characteristics, whereas their offspring who inherit these translocations show phenotypes from mild dysmorphic features to Wolf–Hirschhorn syndrome, characterized by growth defects and severe mental retardation.

α-Thalassemia affects 5–40% of the population in Africa and 40–80% in South Asia. This disorder results from a homologous deletion of an approximately 4-kb segment of DNA that is flanked by two α-globin genes on chromosome 16q13.3. A nonallelic homologous recombination event between these two copies of the α-globin genes results in the deletion of one functional copy, producing the disease phenotype. Red and green pigment genes are located on Xq23, and individuals that have normal color vision have one copy of the red pigment gene and one or more copies of the green pigment gene. In red-green color blindness, which affects 4–5% of males, genetic deletions or fusions caused by nonallelic homologous recombination occur. In patients with incontinentia pigmenti, an 11-kb deletion occurs secondary to nonallelic homologous recombination between two low-copy repeats with one in the diseased gene (NEMO) and another DNA segment that is 4 kb downstream of the gene. Hereditary neuropathy with liability to pressure palsy (HNPP) is a common autosomal dominant neurological disorder that is caused by a 1.4-Mb deletion on 17p12. The gene NF1 that encodes for neurofibromatosis type 1 is located on 17q11.2, and a 1.5-Mb deletion encompassing this gene accounts for 5–22% of patients with this disease.

Patients with DiGeorge syndrome/velocardiofacial syndrome (DGS/VCFS) can exhibit a 3-Mb deletion within a region-specific repeat unit, LCR22, that is flanked by LCR22A and LCR22B, or a 1.5-Mb deletion that is flanked by LCR22A and LCR22B, located on chromosome 22q11.2. Patients with this congenital disease experience recurrent infection, have heart defects, and exhibit distinct facial features. Smith–Magenis syndrome (SMS) affects 1 in 25,000 individuals and is characterized by atrophy of the muscles in the legs, progressing over time to the hands, forearms, and feet. There are two clinical classes of CMT, type I and type II. In CMT type I (CMT1A), 75% of individuals have a duplication in one of the peripheral myelin protein 22 (PMP22) genes. Duplication events affecting PMP22 on both copies of chromosome 17p12 produce a more severe form of CMT1A, reflecting the presence of four copies of the PMP22 gene. Pelizaeus–Merzbacher disease is a central nervous system disorder affecting the myelin sheath covering the nerve fibers in the brain. The majority of patients with this disease have a duplication of the proteolipid protein gene (PLP1), which is found on Xq21–22.

Diseases Associated with Gene Mutations

Gene Mutations in Cancer

Mutations affecting a variety of genes have been characterized in cancer. Some of these mutations represent activating mutations of proto-oncogenes (or other positive mediators of cell proliferation) and others represent inactivating mutations of tumor suppressor genes (or other negative mediators of cell proliferation). Mutations in these genes synergize with other genetic (chromosomal) and epigenetic abnormalities to drive neoplastic transformation and tumorigenesis in affected cells and tissues. This review focuses on the mutation of ras and p53 in cancer.

Mutation of the ras Gene Family in Cancer

Three genes constitute the ras gene family: H-ras, K-ras, and N-ras. The ras protein functions in cell signaling as a hete-rotrimeric G protein, toggling between inactive and active forms. In response to a specific cell stimulus, inactive ras protein releases bound guanosine diphosphate (GDP) and binds a guanosine triphosphate (GTP) molecule. In this GTP-bound active configuration, cell signaling occurs until the intrinsic GTPase activity of the ras protein itself cleaves the GTP to GDP, resulting in reacquisition of the inactive configuration. Mutant K-ras proteins lack intrinsic GTPase activity and remain in a continuously active form. The constitutive activation of K-ras through this mutational mechanism leads to the induction of multiple signaling pathways involved with cell proliferation and cell survival.

Each of these ras family of oncogenes have been found to be mutated with varying frequencies in major forms of human cancer. K-ras is mutated in cancers of the lung, colon, and pancreas, as well as in myelodysplastic syndrome and some other cancers (such as seminoma). The frequency of K-ras mutation in pancreatic cancer is estimated to be 90%, suggesting that this gene mutation is very important in the malignant conversion of this tissue. H-ras is mutated in cancers of the bladder, kidney, and thyroid, and N-ras is mutated in seminoma, thyroid, and acute myelogenous leukemia. In lung cancer, the majority of ras mutations occur in the K-ras gene. K-ras is mutated in 15–20% of all NSCLC and in 30–50% of lung adenocarcinomas, but it is infrequently mutated in other lung cancer types.

In lung adenocarcinomas, 85% of K-ras mutations affect codon 12. Certain carcinogens found in cigarette smoke, such as benzo[a]pyrene, have been shown to preferentially adduct codon 12 of K-ras, and this adduct is not effectively
repaired.\textsuperscript{138,139} No K-ras mutations are found in adenocarcinomas of nonsmokers, supporting a specific role of tobacco carcinogens in the mutation of K-ras.\textsuperscript{140} The majority of K-ras codon 12 mutations are G to T transversions,\textsuperscript{141} resulting in either glycine to cysteine (GGT to TGT) or glycine to valine (GGT to GTT) amino acid substitutions in the mutant protein.

\textit{Mutation of the p53 Tumor Suppressor Gene in Cancer}

The p53 tumor suppressor gene is one of the most frequently mutated genes in cancer.\textsuperscript{142,143} The p53 protein functions as a transcriptional regulator and mediator of cellular responses to DNA damage and stress.\textsuperscript{144} Point mutation of p53 leads to synthesis of mutant forms of the protein that do not fold properly,\textsuperscript{145} resulting in a nonfunctional protein that will not bind DNA.\textsuperscript{146,147} When this mutant protein oligomerizes with other p53 molecules (normal or mutant), the resultant tetramers (or higher-order oligomers) are nonfunctional.\textsuperscript{148} Thus, cells with mutant p53 (or p53 deficiency due to chromosomal deletion) become susceptible to progressive genomic instability and accumulation of additional genetic damage.\textsuperscript{149,150} Mutational inactivation of p53 has been intensely studied since the discovery of the gene two decades ago. The IARC TP53 Mutation Database (http://www-p53.iarc.fr)\textsuperscript{151,152} catalogues reported mutations in the p53 gene. The most recent compilation (Release 12, November 2007) contained 24,819 mutations, including germline and somatic mutations. Of these mutations, 23,544 were characterized as somatic mutations, the majority (17,358, or 74\%) of which were missense mutations. Although missense mutations represent the prevalent mutation type affecting p53, numerous other mutation types have been documented, including frameshift, nonsense, and splice site mutations among others. p53 mutations have been documented in subsets of most major forms of human cancer, including cancers of the colorectal, lung, breast, head and neck, esophagus, brain, liver, skin, pancreas, and bladder, as well as in hematological malignancies.

Inactivation of the p53 protein can result from mutations of the p53 gene affecting codons encompassing the majority of the gene, with most of the mutations falling within the highly conserved region of the gene that includes exons 5–9,\textsuperscript{143,153,155} Nevertheless, a few hotspots for mutation have been identified. These mutational hotspots correspond to the codons that encode amino acids which are important for the functional interaction of the p53 protein with the DNA molecule.\textsuperscript{156} The mutational spectrum of the p53 gene is also known to reflect the carcinogenic insults encountered by various tissues. For example, consider p53 mutations in liver cancers and lung cancers. In hepatocellular carcinoma, p53 is frequently the target of chromosomal deletion involving 17p13.1, perhaps occurring in as many as 41\% of liver cancers examined.\textsuperscript{157} However, p53 mutation occurs in conjunction with chromosomal deletion of the p53 locus in hepatocellular carcinoma. A summary of numerous studies of p53 mutation in hepatocellular carcinoma suggests that mutations occur in approximately 27\% (547/2029) of these tumors.\textsuperscript{157} Point mutation of codon 249 of the p53 gene represents a hotspot in hepatocellular carcinoma, accounting for approximately 30\% of p53 mutations in hepatocellular carcinoma.\textsuperscript{157} This mutation, which results in a G to T transversion (AGG to AGT, arginine to serine), was first recognized in patients from Qidong, China,\textsuperscript{158–160} but it has also been recognized in hepatocellular carcinomas from Africa and North America.\textsuperscript{161–163} The p53 codon 249 mutation has been attributed to exposure to aflatoxin B$_1$.\textsuperscript{164} Hence, the mutational spectrum of the p53 gene in hepatocellular carcinoma differs significantly between geographic areas with high aflatoxin exposure and those with low exposures.\textsuperscript{162,165,166} Similar to liver cancer, the p53 gene may be involved in the molecular pathogenesis of lung cancer through chromosomal deletion of 17p13.1, as well as through gene mutation. Approximately 50% of NSCLCs and 90% of SCLCs harbor mutations in the p53 gene.\textsuperscript{167–169} The p53 mutational spectrum in lung cancer indicates that G to T transversions dominate and that specific hotspot codons are frequently mutated (including codons 157, 158, 175, 245, 248, 249, and 273).\textsuperscript{170} The types of mutations detected may reflect the interaction of the DNA with specific carcinogens found in cigarette smoke.\textsuperscript{171,172}

\textbf{Gene Mutations in Nonneoplastic Diseases}

The phenylalanine hydroxylase gene (PAH) maps to chromosome 12q23.2\textsuperscript{173,174} and is expressed in the liver and kidney.\textsuperscript{175} Mutations in the PAH gene, which encodes the enzyme l-phenylalanine hydroxylase, cause phenylketonuria (PKU).\textsuperscript{176} PKU is an autosomal recessive inherited disease, causing mental retardation, a musty odor, light pigmentation, and peculiarities of sitting, standing, and walking, as well as eczema and epilepsy.\textsuperscript{176} The average incidence of PKU in the United States is 1 in 8000 individuals.\textsuperscript{177} The metabolic defect in PKU reflects the inability of these patients to hydroxylize phenylalanine to tyrosine, resulting in hyperphenylalaninemia. Untreated hyperphenylalaninemia is toxic to the brain and leads to symptoms of mental retardation.\textsuperscript{176} The \textit{PAH} locus encompasses 1.5 Mb of DNA with single nucleotide polymorphisms (SNPs), repeat sequences, polymorphisms, and \textit{cis} control elements embedded in the sequence, and harbors five other genes. Thus, a wide range of disease-causing mutations can affect the genes contained in this region, resulting in variable disease manifestations.\textsuperscript{176,178} PKU is one of the first genetic disorders to benefit from an effective rational therapy.\textsuperscript{176} PKU can be identified with a biochemical test in newborns and can be treated by a phenylalanine-free, tyrosine-supplemented diet, which permits normal or near-normal cognitive development.\textsuperscript{176}

Cystic fibrosis (CF), an autosomal recessive disease that occurs in approximately 1 in 3,500 newborns, is the
most common lethal inherited genetic disease among the Caucasian population, affecting almost 30,000 Americans. Treatment advances for CF patients have increased life expectancy to more than 36 years today.

The molecular basis for the development of CF is mutation of the cystic fibrosis transmembrane conductance regulator (CFTR). The diagnosis of CF is based on several clinical characteristics, a family history of CF or a positive CF newborn screening test, and mutation in the CFTR gene and/or protein. The symptomatology of CF is highly variable from patient to patient, possibly owing to differences in CFTR function caused by specific mutations.

Given that epithelia in CF patients behave as though they are impermeable to chloride ions, it was expected that the CFTR gene would encode a chloride channel. However, it was found that the CFTR gene encodes an ABC transporter homologue that actually functions as a chloride channel which is directly activated by phosphorylation.

The most common mutation in the CFTR gene, designated ΔF508, results in deletion of phenylalanine at position 508 within nucleotide binding domain 1 of the CFTR protein. The ΔF508 mutation accounts for more than 65% of CFTR mutations, although more than 1,000 other CFTR mutations have been documented. The ΔF508-CFTR protein is not properly folded during posttranslational processing and is rapidly degraded, resulting in deficiency for the CFTR protein.

The best characterized genetic risk factor for emphysema (and COPD) is α1-antitrypsin deficiency. However, α1-antitrypsin deficiency is a rare condition and only accounts for 1–2% of patients with emphysema, typically those with panacinar pathology. α1-Antitrypsin is a member of the serine protease inhibitor superfamily of proteins, which function to inactivate neutrophil elastase and other proteases to maintain the protease-antiprotease balance.

Normally, α1-antitrypsin is synthesized in the liver and secreted into the blood, producing serum concentrations of 20–53 μmol/l. In deficient individuals, α1-antitrypsin levels in the serum fall below 20 μmol/l. A number of different mutant alleles for α1-antitrypsin have been identified and characterized. Most clinically recognized cases of α1-antitrypsin deficiency are caused by the presence of the Z allele, which is characterized by the substitution of lysine for glutamic acid at position 342 (Glu342Lys). The Glu342Lys mutation leads to polymerization of the Z form of α1-antitrypsin within hepatocytes, resulting in impaired secretion and reduction of the plasma levels of α1-antitrypsin to 15% of normal.

To compound the α1-antitrypsin deficiency in these patients, the Z form of α1-antitrypsin does not function properly. Thus, deficiency in the antielastase activity of α1-antitrypsin caused by diminished serum levels and decreased activity of the defective protein (Z form) sets up a chronic condition in the patient characterized by excess elastase activity and destruction of the lung.

Diseases Associated with Epimutation

Epimutations in Cancer

Neoplastic transformation is associated with alterations in DNA methylation, including both global hypomethylation and gene-specific hypermethylation. Gains of DNA methylation in cancer cells typically reflect hypermethylation of CpG islands in gene promoter regions, which contributes to gene silencing. Methylion-dependent gene silencing is a normal mechanism for regulation of gene expression. However, in cancer cells methylation-dependent epigenetic gene silencing represents a mutation-independent mechanism for inactivation of tumor suppressor genes or other negative mediators of neoplastic transformation. A significant number of cancer-related genes have been identified across all chromosome locations that are subject to methylation-dependent silencing, and many of these genes contribute to the hallmarks of cancer. It is likely that more genes are silenced by hypermethylation than through genetic alterations. Additionally, some genes that are silenced though epigenetic mechanisms (including DNA methylation) in many tumor types are not subject to genetic mutation.

Similarly, genetic and epigenetic alterations can function together to alter gene expression in cancer cells. There are several examples of genes (MLH1 and p16INK4A) in which one allele is subject to mutation, and subsequently the remaining normal allele becomes silenced through promoter hypermethylation. Thus, epigenetic modifications complement and collaborate with genetic changes to transform normal cells into neoplastic cells. These observations combine to strongly suggest that epigenetic events, and specifically those involving DNA methylation, represent fundamental aspects of cancer and play key roles in neoplastic transformation and progression.

Epimutations in Lung Cancer

Several genes are known to be hypermethylated in lung cancer, resulting in gene silencing. Among these genes, p16INK4A is found to be methylated in more than 40% of lung cancers examined. Although it is recognized that hypermethylation of p16INK4A occurs frequently in lung cancer, there is debate regarding the importance of this mechanism of inactivation among lung cancer patients who are smokers versus nonsmokers (or never smokers). In one study, deletion or mutation of the p16INK4A gene was only observed in the lung cancers of smokers, while hypermethylation was found only among nonsmokers. In contrast, other studies have found hypermethylation of p16INK4A as the prevalent mechanism of inactivation among lung cancers from smokers. Other epigenetically regulated genes in lung cancer include CDH1, CDKN1A, DAPK1, ESR1, GJB2, GSTP1, HS3ST2, PRDM2, PRKCDBP, RASSF1, and SFN. Similar to p16INK4A, RASSF1 may be subject to inactivation through multiple molecular mechanisms:
Epimutations in Colorectal Cancer

Methylation of multiple gene targets occurs during colorectal carcinogenesis and some of these methylation events occur early in the multistep process. HLTF, hMLH1, MGMT, APC, and ATM are frequently methylated in colorectal adenomas (42–78%), at least one of these genes was methylated in 100% of the adenomas tested, and 61% contained at least three methylated genes. Similarly, another study found that p16INK4A, MGMT, and hMLH1 are methylated in 34%, 49%, and 7% of colorectal adenomas, respectively. These genes were not methylated in normal colonic mucosa and were infrequently methylated in hyperplastic polyps (5%, 10%, and 7.5%, respectively). These observations suggest that methylation events that occur early in colorectal tumorigenesis may significantly contribute to progression of pre-neoplastic lesions toward malignant conversion. In fact, the methylation patterns observed in colorectal adenomas are preserved with progression to colorectal carcinoma. HLTF, hMLH1, MGMT, APC, and ATM are methylated in colorectal carcinomas with similar or greater frequency (47–77%) compared to that observed in adenomas, and 68% contain at least three methylated genes.

Epimutations in Breast and Prostate Cancer

A number of studies have shown the presence of aberrant DNA methylation of genes in breast carcinogenesis, affecting genes including p16INK4A, p14ARF, Cyclin D2, and Slit2. The majority (86%) of breast tumors showed hypermethylation of at least one gene, and there was a significant correlation between methylation in tumor and serum DNA of all genes. In other studies, nipple aspiration was applied as a noninvasive technique to obtain DNA samples for examination of tumor-associated gene methylation. Hypermethylation of one or more genes, including GSTP1, RARβ2, p16INK4A, p14ARF, RASSF1, and DAP-kinase, was detected in DNA from nipple aspirate, and identical gene hypermethylation was found in 82% of matched breast tumor tissue samples. However, no aberrant levels of gene methylation were seen in the normal and benign breast tissue or in nipple aspirates from healthy women. Several studies have examined methylation of GSTP1 in prostate cancers. Hypermethylation of GSTP1 has been analyzed in more than 1,000 prostate tumor samples and has been shown to be methylated in the majority (81%) of these tumors. In addition, a recent study of GSTP1, ARF, p16INK4A, and MGMT methylation levels in urine sediments reported that these four genes were able to identify the majority (87%) of prostate cancer patients with extremely high specificity (100%) compared to control samples. These findings suggest that DNA from urine samples could be utilized as a screening tool to reduce the number of biopsies performed on men with a high risk of developing prostate cancer.

Epimutations in Other Cancers

In addition to the major cancer sites described here, aberrant epigenetic regulation of gene expression has been documented in various other human cancers, including cancers of the oral cavity, cervix, ovary, skin, pancreas, and brain, and leukemia and lymphoma.

Epimutations in Nonneoplastic Disease

Evidence in the literature strongly suggests that loss of CFTR function causes cystic fibrosis (CF), and a large number of mutations in the CFTR gene have been characterized. However, there is some evidence that the CFTR gene may be subject to methylation-dependent epigenetic regulation. The promoter of the CFTR gene is remarkably GC rich. Early studies found hypermethylation of the CFTR promoter in various cell lines analyzed, which correlated with low levels of expression. However, more detailed bisulfite sequencing analysis of the CFTR promoter has not been reported. Nevertheless, these results suggest that the CFTR gene may be subject to epigenetic regulation through promoter hypermethylation, possibly representing a nonmutational mechanism for inactivation of CFTR, leading to cystic fibrosis.

References

47. Tsuchiya E, Nakamura Y, Weng SY, et al. Allelotype of non-small cell lung carcinoma: comparison between loss of


83. Bieche I, Khodja A, Lidereau R. Deletion mapping in breast tumor cell lines points to two distinct tumor-suppressor genes in the 1p32-pter region, one of deleted regions (1p36.2) being located within the consensus region of LOH in neuroblastoma. Oncol Rep. 1998;5:267-272.
5. The Role of Mutation and Epimutation in the Development of Human Disease


104. Upadhyaya M, Ruggieri M, Maynard J, et al. Gross deletions of the neurofibromatosis type 1 (NF1) gene are predominantly...


Bioinformatics and Omics

The term “genomics” originated in 1920 to describe the complete set of chromosomes and their associated genes; however, it has been in the last decade that the use of omics – genomics, transcriptomics and proteomics – and bioinformatics has led to dramatic advances in the understanding of the molecular/genetic basis of disease.\(^1\)\(^-\)\(^9\)

Bioinformatics

Bioinformatics has become an essential part of omics research and requires unique practical and analytical skills for appropriate results interpretation. Bioinformatics uses computers and statistics to perform extensive omics-related research by searching biological databases and comparing gene sequences and protein on a vast scale to identify sequences or proteins that differ between diseased and healthy tissues, or between different phenotypes of the same disease.\(^30\)\(^-\)\(^37\) The techniques used in omics are called high throughput because they involve analysis of very large numbers of genes, gene expression, or proteins in one procedure or combination of procedures. The vast amounts of data generated by these high-throughput studies typically require computers for analysis and comparison of differences between diseased and physiological cells and tissues, a key feature of bioinformatics. Omics and bioinformatics are used not only for the study of the genes and signaling pathways involved in human diseases, but also for identifying potential targets of therapy and the design of therapeutic drugs.

Omics

Omics – a suffix signifying the measurement of the entire complement of a given level of biological molecules and information – today encompasses a variety of new technologies that can help explain normal and abnormal cell pathways, networks, and processes via the simultaneous monitoring of thousands of molecular components.\(^6\)\(^,\)\(^7\)

Genomics

Genomics provides platforms for the study of genomes and their genes, including haplotyping and single nucleotide polymorphism detection, by investigating single nucleotide polymorphisms (SNPs) and mutations using high-throughput genome sequencing techniques such as high-density DNA microarrays/DNA (oligonucleotide) chips.\(^7\)\(^,\)\(^38\)\(^-\)\(^44\) The base sequence of the genes of the human mitochondrial genome was completed in 1981, and in 2003 the base sequence of the genes of the entire human genome was completed.\(^45\)\(^-\)\(^47\) The human genomic sequence data from the International Human Genome Sequencing Consortium can be mined using tools that are now publicly available.\(^48\) Public databases can also be mined for SNPs, human mitochondrial genome, and other human DNA polymorphic markers.\(^49\)\(^-\)\(^54\)

Transcriptomics

Also termed “functional genomics,” transcriptomics provides information about the expression of individual genes at the messenger RNA (mRNA) level and correlates patterns of expression with biological function.\(^7\)\(^,\)\(^55\)\(^-\)\(^64\) A variety of techniques have been developed to investigate gene expression. These techniques include serial analysis of gene expression (SAGE), suppression subtractive hybridization (SSH), differential display (DD) analysis, RNA arbitrarily primer-polymerase chain reaction (RAP-PCR), amplified restriction fragment-length polymorphism (AFLP), total gene expression analysis (TOGA), and use of internal standard competitive template primers in a quantitative multiplex reverse transcription (RT)-PCR method [StaRT-(PCR)], restriction endonucleolytic analysis of differentially expressed sequences (READS), differential screening (DS), high-density cDNA filter hybridization analysis (HDFCA), and gene expression microarrays.\(^65\)\(^-\)\(^67\)
Proteomics

Proteomics investigates individual protein concentrations present in a biological system and studies the structural, functional and regulatory roles of proteins in the cell and in pathways, including how and where they are expressed. Because gene function is ultimately performed by the proteins transcribed from the genes and mRNA, proteomics is essential to comprehend the actual definitive functioning of a gene or pathway. Fortunately, proteomics can be performed on surgical specimens including needle biopsies, cytology specimens, serum, and other fluids. Frequently, two-dimensional gel electrophoresis (two-dimensional polyacrylamide gel electrophoresis) and mass spectrometry are employed in proteomics to initially fractionate the groups of proteins in a specimen. Proteins in a given sample fraction may later be identified using fingerprinting or sequence tag techniques.

Metabolomics

Metabolomics (also termed metabonomics) comprises study of metabolic profiles by investigating the compounds in a process and involves the characterization and quantification of small organic molecules in either circulatory or cell–tissue systems. Nuclear magnetic resonance and mass spectrometry are techniques frequently used in metabolomics.

Glycomics

A challengingly diverse variety of cell-surface glycoconjugates control various biological events, including, among others, cell adhesion, cell differentiation, immunological recognition, and virus–cell recognition. Glycomic analysis, termed “glycomics,” is the systemic study of the variations in cellular glycoconjugates, using mass spectrometry and microarray-based and computational technologies. In mammals, the glycome includes all oligosaccharides from glycoconjugates such as glycolipids, proteoglycans, and glycoproteins. As it is difficult to profile the entire glycome because of the large diversity of oligosaccharides in the structure and function of an organism, glycomics targets specific groups of oligosaccharides, glycoproteins, and glycoconjugates, as well as the characterization of glycans and their corresponding attachments to proteins or lipids, to determine the role of glycan heterogeneity in diseases. Significant challenges in mammalian glycomics include the complexity of the glycans, the complex biosynthesis of the epitopes, the multivalent nature of biological glycan recognition, and the subtle phenotypes of glycan manipulation, often requiring a multicellular environment to become manifest. The integration of glycomics with genomics and proteomics is important, as the relationship between the glycome and genome is uncertain, and as there is constant reevaluation of the relative impacts of genomics, proteomics, and posttranslational modulation of protein levels and functions. High-throughput glycomic technologies, such as lectin microarrays, will assist in the integration of these data sets and ultimately help determine the role of carbohydrates in various biological pathways.

Cytomics

Cytomes are cellular systems, subsystems, and functional components of an organism. Cytomics, the study of the heterogeneity of cytomes, is the multimolecular, single-cell analysis of cell–system heterogeneity using image or flow cytometry. Cytomics may assist in the identification of rare cells lost in the background with genomic and proteomic profiling, such as contaminant cancer cells in autologous transplant tissue that may cause relapse. As technology improves, the use of cytomics, integrated with proteomics, may allow for the use of flow cytometry to sort separate heterogeneous populations of cells using specific phenotypic characteristics, ultimately identifying and validating tumor markers that advance the potential to create personalized molecular medicine.

Physiomics

The physiome is defined as the total integration of genome, proteome, and metabolome, from cells to organisms. Physiomics is a branch of omics using large-scale and experimental databases, and computer algorithms, to examine the physiological phenotypes of, and the relationships between, genes and proteins. Useful in the development of drugs and biosensors and biochips, physiomics will be enhanced by the emergence of new technologies including microtechnologies to probe the functional realm of living cells, ultimately extending electroanalytical chemistry and biosensor technologies from the micro-scale to the nano-scale, allowing for the study of subcellular and organellar physiology.

Techniques

DNA microarrays are employed to simultaneously screen for the presence or expression of large numbers of genes. Suppression subtractive hybridization (SSH) selectively amplifies target cDNA fragments (differentially expressed genes) and suppresses nontarget DNA. Serial analysis of gene expression (SAGE) is used for global analysis of gene expression and provides a comprehensive qualitative and the quantitative expression profile of virtually every gene in a cell population or tissue, and SAGE libraries from different cells and tissues have been created. 2-DE, a computer software developed in the mid-1970s to separate complex mixtures of proteins, has become a powerful tool for allowing
simultaneous resolution of thousands of proteins contained in an organism’s proteome.115,116 Today, a variety of softwares are available for 2-DE experiments, but one must be careful with their use.1 2-DE experiments generate great quantities of data that might require weeks to analyze and interpret for appropriate protein identification.1 Poor experimental design, poor statistical treatment, and high rates of false-positive results hinder appropriate protein identification with these softwares.1 The massive amounts of data produced by these studies have become problematic for journal article reviewers, who may find themselves faced with many pages of raw bioinformatics data when reviewing an article.117

Microarray technologies have become a powerful tool for simultaneously monitoring expression patterns of thousands of genes; however, a important task is to now propose analytical methods to identify groups of genes manifesting similar expression patterns that are activated by similar conditions.118 Other concerns include the longevity of data sets and the robustness of conclusions built upon them, the desire to integrate data from diverse scientific disciplines, and the facility to productively mine accrued data.119 Ultimately, omics technologies in biomarker discovery and cancer detection may allow for the production of diagnostic tools that are simple, fast, robust, portable, and cost-effective, yielding clinical diagnosis systems for use outside the laboratory, including the clinic and hospital bedside.120

Conclusion

The need for better and more accurate diagnosis, prognosis, and treatment response indicators has yielded increased utilization of omics technologies, particularly in the area of gene expression profiling.121 U.S. Food and Drug Administration-approved proprietary testing platforms based on microarray technologies already exist and will continue to expand.122 Despite the rapid evolution of, and widespread adoption of, DNA microarray technology technologies, there remain considerable uncertainty and skepticism regarding data obtained using these technologies.123 Although rapid medical advances in the laboratory have not necessarily translated into rapid treatment advances, continuing omics research to understand the conceptual framework of disease – including disease progression and treatment response – along with improved and more efficient bioinformatics tools to analyze the great amounts of data originating from omics investigations may permit future diagnostic, prognostic, and therapeutic benefit to arise from these technologies.6 The rapid progress of omics and bioinformatics has evolved into a promising new era for personalized patient care.124

References


6. Bioinformatics and Omics

76. Scaros O, Fisler R. Biomarker technology roundup: from discovery to clinical applications, a broad set of tools is required to translate from the lab to the clinic. *Biotechniques*. 2005;Suppl.30:32.
General Approach to Molecular Pathology

Gregory L. Blakey and Daniel H. Farkas

Introduction

Once a highly specialized subdiscipline of laboratory medicine, molecular diagnostics now infiltrates all of anatomic and clinical pathology. The shift from dependence on a few, relatively cumbersome methods to a wider range of technologies has facilitated this expansion. In addition, the completion of the Human Genome Project and the growing amount of sequence data related to infection, cancer, and other disease states have yielded additional applications of molecular biology for the clinical laboratory (Fig. 7.1). As the various phases of testing can be automated in many instances, molecular biological experience is no longer a prerequisite. In fact, performance of nucleic acid extraction and amplification in a tabletop unit is possible. Increasingly, miniaturization will further move molecular testing to the point of care.

Although varying from protocol to protocol, the workflow for many molecular pathological tests follows a similar scheme (Fig. 7.2). As with other laboratory tests, preanalytical variables can dramatically affect molecular assay results. Therefore, appropriate care should be paid to specimen collection, transport, processing, and storage. In particular, fresh samples submitted for analysis of RNA must be processed rapidly to prevent degradation.

Nucleic Acid Extraction

Although the traditional organic means of extracting nucleic acids, with phenol and chloroform, are useful in the clinical laboratory, less cumbersome methods now dominate. For instance, silica-based spin column technologies couple ease of use with adaptability. With these methods, a digested sample is loaded onto a column to which nucleic acids bind in the presence of chaotropic salts. Microcentrifugation of the column followed by washing removes proteins and other macromolecules. Finally, nucleic acids are eluted with buffer or water. As with more traditional techniques, conventional spin column-based nucleic acid isolation suffers from the need for manual labor, not an insignificant factor in the busy clinical laboratory.

Automation of nucleic acid extraction techniques has increased laboratory efficiency and decreased turnaround time. Of particular importance for quantitative testing, reduced inter- and intraoperator variability also results. Commonly used automated nucleic acid extraction techniques often rely on magnetized glass beads. After cell lysis, nucleic acids bind the beads, which can be robotically manipulated with a magnet. The beads are washed, and then nucleic acids are eluted with water or an appropriate buffer. Throughput of automated extraction robots ranges from approximately 6 to 96 or more samples per run. Depending on the model, these machines are capable of processing various sample types, including plasma, whole blood, fresh tissue, and formalin-fixed tissue. Typical isolation choices include total nucleic acid, DNA, and RNA.

Although nucleic acids are most commonly purified for use in molecular protocols, crude cell lysates can be used in selected cases. Such an approach may be especially useful when only minute amounts of samples are available, as with macro- or microdissected cells. In these instances, the decrease in nucleic acid yield because of the extraction process might be unacceptable. Crude lysates have been successfully used as starting material for polymerase chain reaction (PCR)-based clinical diagnostics.

Amplification Technologies

Target Amplification

Polymerase Chain Reaction

In combination with a variety of other methods (Fig. 7.3), PCR remains the core molecular pathological technique. It results in a manyfold increase in target sequences by virtue of repeated cycles in which product becomes additional template. Despite the superior sensitivity and convenience of
PCR versus Southern blotting and other venerable methods, the potential for contamination of the laboratory space and reagents with reaction products - amplicons - slowed entry of molecular diagnostics into traditional laboratory settings. The relatively recent availability of thermal cyclers that can continually monitor product formation has mitigated this

Fig. 7.1. Molecular diagnostics encompasses the use of nucleic acids to diagnose infection, malignancy (hematological and solid tumors), and genetic diseases. Other major foci include the investigation of human remains and crime scenes (forensics) and evaluation of transplant donors and recipients [human leukocyte antigen (HLA) testing].

Fig. 7.2. The general steps in many molecular diagnostic protocols are similar. For example, a blood sample received for genetic testing first may be processed to isolate white blood cells. Then, DNA is extracted from these cells and used to set up polymerase chain reaction. A multitude of detection methods are available (see text for further discussion).

Fig. 7.3. Numerous techniques allow detection and analysis of polymerase chain reaction (PCR) products, amplicons. Amplicons can be sized by gel electrophoresis, detected in real time with fluorescent probes, and sequenced with a variety of methods. Mass spectrometry approaches allow precise fragment sizing or base composition determination, and DNA “chips” permit parallel hybridization of amplicon and thousands of probes. (Sequencing diagram modified from Blakey and Farkas,7 by permission of the Colorado Association for Continuing Medical Laboratory Education).
problem. These instruments allow real-time PCR in which subsequent detection steps requiring opening of the reaction vessel are usually not needed. Besides lowering the risk of contamination, real-time PCR allows rapid turnaround time; the high-speed thermal cyclers used rapidly change the reaction temperature of metal blocks or air in which reaction vessels are suspended. Applications of real-time PCR include genotyping, pathogen identification, and target quantification.

Besides the standard reagents [primers flanking the target site, heat-resistant polymerase, deoxyribonucleotide triphosphates (dNTPs), and various buffer components], real-time PCR requires fluorescent molecules for ultimate product detection. Fluorescent dyes such as SYBR Green that nonspecifically bind double-stranded DNA can be used, but false-positive results may occur. More commonly, polymorphism and mutation detection are accomplished with fluorescently labeled oligonucleotide probes. Some of the most frequently used of the many such probe systems are hybridization probes, hydrolysis probes, and molecular beacons. Paired hybridization probes anneal to adjacent sites on the target sequence. One hybridization probe has an acceptor fluorophore, the other a donor fluorophore. When in close proximity, the energy from the donor excites the acceptor via fluorescence resonance energy transfer (FRET; Fig. 7.4a). The real-time thermal cycler measures this signal. A hydrolysis probe consists of a single oligonucleotide labeled...
with reporter and quencher fluorophores; fluorescence is inhibited by FRET between the two. Upon dissociation of the two fluorophores by the exonuclease activity of the DNA polymerase during the extension phase of PCR, the reporter emits light (Fig. 7.4b). Similarly, molecular beacons contain a stem-loop structure that brings together reporter and quencher fluorophores attached to the respective ends of the probe. When the probe binds to the target sequence, the reporter and quencher are physically separated (Fig. 7.4c).

Melting curve analysis with hybridization probes can be used to distinguish different target sequences. Hybridized strands separate at the melting temperature ($T_m$). To determine $T_m$, after PCR is completed, the thermal cycler slowly increases the temperature while continuously monitoring fluorescence. As the temperature rises, probes that are less tightly bound to target, either by design or because of mismatches, dissociate first. Stronger, more stable, target-probe hybrids melt at higher temperatures. By plotting the change in fluorescence produced versus temperature, different target amplicon species (e.g., normal versus mutant or different pathogens) can be distinguished (Fig. 7.5).

Real-time PCR permits relatively easy quantification of analytes such as microbial genomes or fusion transcripts. As the initial concentration of an analyte of interest increases, the reaction cycle number at which fluorescence exceeds background decreases. Therefore, external calibrators can be used to construct a standard curve, thus permitting quantification.

Polymerase chain reaction systems with 96- and 384-well plates increase throughput. Multiplexing (more than one reaction per tube or well) enhances this. However, untoward interactions between the various components of a multiplexed reaction limit the number of analytes that can be interrogated simultaneously. High-throughput gene expression profiling may be accomplished by advances such as highly parallel picoliter-scale PCR. With potentially superior dynamic range and reproducibility versus microarrays (see later), real-time PCR may allow gene expression profiling to enter more widespread clinical use.

Fig. 7.5. Amplicons can be distinguished by melting curve analysis. Two hybridization probe pairs differentiate between a mixture of analytes 1 and 2 in this illustration (left). As the thermal cycler ramps up the temperature, one of the probes for analyte 2 dissociates first (middle), resulting in drop in total fluorescence. The temperature at which this loss of fluorescence occurs is the melting temperature for analyte 2. A second drop indicates the dissociation of at least one of the analyte 1 probes (right).
Other Target Amplification Techniques

Besides PCR, target amplification techniques include the proprietary nucleic acid sequence-based amplification and transcription-mediated amplification, both of which rely on the isothermal production of RNA intermediates.

Signal Amplification

Instead of increasing the number of targets to boost sensitivity, signal amplification techniques rely on multiplying the detection signal. Signal amplification techniques include Hybrid Capture, Invader, and branched chain DNA (bDNA), all proprietary.

With Hybrid Capture, in wide use for the identification of human papillomavirus, hybrids of viral DNA and probe RNA are bound by antibodies attached to a reaction well. Labeled antibodies complete the sandwich of the hybrids, allowing chemiluminescent detection.

Invader assays depend on recognition of the combination of the target, a probe, and a so-called invader oligonucleotide by a proprietary enzyme (Cleavase); probe cleavage results. A probe fragment in turn interacts with a looped DNA structure labeled with reporter and quencher fluorophores (FRET cassette). Cleavage again results, releasing the reporter, which fluoresces. Invader chemistry is isothermal. Thus, a thermal cycler is not required. In addition, as amplicons are not produced, the risk of contamination is minimized. Genotyping is a common application.

bDNA technology uses a combination of capture molecules and labels to amplify the amount of signal. Applications include quantification of human immunodeficiency, hepatitis B, and hepatitis C viruses.

Restriction Fragment Analysis and Southern Blotting

Restriction Fragment Length Polymorphism

Restriction fragment length polymorphism (RFLP) analysis, previously a mainstay of genotyping, involves cutting target molecules (usually after amplification with PCR) with a restriction endonuclease, followed by gel electrophoresis. The resulting fragment patterns can be compared to those of controls to deduce the genotype. Many RFLP assays have been supplanted by those based on real-time PCR.

Southern Blotting

A particularly laborious and time-consuming (turnaround times of many days) technique, the Southern blot is still used in the clinical laboratory, particularly when target sequences are too long to safely amplify with PCR. In Southern blotting, relatively large quantities (microgram) of DNA are digested with a restriction endonuclease, fractionated via gel electrophoresis, and then transferred to a solid support, such as a nylon membrane. Labeled probes then hybridize to complementary target sequences, permitting detection.

Sequencing

Sequencing is used widely in genetics laboratories, particularly for the identification of heterogeneous mutations within a given gene or exon.

Sanger Sequencing

In Sanger sequencing, fluorescently labeled deoxyribonucleotide triphosphates are incorporated into growing nucleotide strands, causing chain termination. The resulting fragments are resolved with slab gel or, more commonly, capillary electrophoresis, to yield the sequence. Sequence read lengths are in the hundreds of base pairs. Common sequencing primers and protocols, high-throughput 384-capillary sequencers, and sophisticated analysis software form the large-scale sequencing pipeline required for the timely detection of polymorphisms and mutations in the high-volume clinical molecular genetics laboratory.

Real-Time Sequencing

Pyrosequencing, or real-time sequencing, depends on a proprietary chemical cascade that converts the pyrophosphate by-product of nucleotide incorporation to light (Fig. 7.6). With
inkjet printer technology, the pyrosequencing instrument adds individual dNTPs in a specified order to reaction vessels. Only when a dNTP incorporates at a given position is light generated and detected by the instrument; therefore, the sequence of the target can be inferred.

Although Sanger sequencing remains the predominant sequencing technique, real-time sequencing offers several advantages. Sequence reads can begin immediately after the sequencing primers. In addition, detection occurs during the sequencing reaction, reducing turnaround time. Compared with real-time PCR, pyrosequencing has the advantage of quicker assay development in some instances, given that design and optimization of fluorescently labeled probes is not required. Limitations of pyrosequencing include difficulty reading homopolymeric repeats (several consecutive identical bases) and the relatively short sequence read lengths (less than approximately 100 base pairs).

Advanced Sequencing
Rapid progress in high-throughput sequencing technologies hastens the goal of quickly and inexpensively sequencing a human genome for clinical purposes. One recently introduced machine, based on pyrosequencing chemistry and the so-called shotgun sequencing approach, can sequence a bacterial genome in as little as a few hours. An effort to more fully characterize genomic changes of malignancies, the National Institutes of Health’s Cancer Genome Atlas project (http://cancergenome.nih.gov), could eventually generate new clinical assays based on new sequencing technologies.

Methylation Detection Methods
Epigenetic changes, modifications to the genome that do not alter the sequence itself, are involved in the regulation of gene expression. Means to identify one such class of epigenetic change, methylation of cytosines, include methylation-sensitive restriction enzymes that generate differential RFLP patterns and chemical conversion (bisulfite treatment) of methylated bases followed by sequencing.

Fluorescence In Situ Hybridization and Microarrays

Fluorescence In Situ Hybridization
Fluorescently labeled probes hybridized to histological and cytological preparations allow the detection of translocations, amplifications, and deletions. Fluorescence in situ hybridization (FISH) is used widely in the study of hematological malignancies, solid tumors, and genetic disorders. Platforms with automated hybridization and analysis will increase the practicality of FISH in the clinical laboratory. Although generally less sensitive than PCR, FISH has the advantage of preserving histological context.

Chromosomal Microarrays
Composed of hundreds or thousands of probes or clones representing every chromosome, chromosomal microarrays achieve a higher resolution than the staining techniques of classic cytogenetics. Thus, this technique, also known as array comparative genomic hybridization, may detect small deletions and duplications missed by traditional karyotyping. Before hybridization, samples may undergo whole genome amplification to increase sensitivity. Preliminary studies have shown that formalin-fixed cells and tissues may be successfully used with chromosomal microarrays, raising the possibility of its routine use in anatomic pathology.

Oligonucleotide Microarrays
Oligonucleotide microarrays are constructed by spotting hundreds to thousands of oligonucleotide probes onto a glass slide or other substrate. Alternatively, probes can be incorporated in situ during construction of the array. By hybridizing cDNA, derived from cellular mRNA via reverse transcription, to these arrays, the expression levels of the various genes in a cell can be measured. Gene expression profiling has been repeatedly demonstrated to generate "signatures" that correlate with prognosis of different cancers, although variabilities in the means of assay performance and statistical analysis have sometimes rendered interlaboratory comparison difficult. Platform standardization and other efforts to improve reproducibility will likely pave the way for gene expression profiling to more fully enter routine clinical use. Selected example applications of microarray technologies are listed in Table 7.1.

Table 7.1. Applications of microarray technologies.

<table>
<thead>
<tr>
<th>Application</th>
<th>Format</th>
<th>Comment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gene expression (mRNA) profiling</td>
<td>Solid phase, liquid bead</td>
<td>Can identify disease-specific signatures</td>
</tr>
<tr>
<td>Global chromosomal abnormality detection</td>
<td>Solid phase</td>
<td>Resolutions depends on number of clones or probes; may not detect balanced translocations</td>
</tr>
<tr>
<td>MicroRNA expression profiling</td>
<td>Solid phase, liquid bead</td>
<td>Robust tumor-specific signatures are obtainable from fixed tissues</td>
</tr>
<tr>
<td>Pathogen detection</td>
<td>Solid phase, liquid bead</td>
<td>Has potential to replace many conventional virological and bacteriological methods</td>
</tr>
<tr>
<td>Polymorphism and mutation detection</td>
<td>Solid phase, liquid bead</td>
<td>A pharmacogenetic test is the first U.S. Food and Drug Administration- approved microarray-based assay</td>
</tr>
<tr>
<td>Sequencing</td>
<td>Solid phase</td>
<td>Alternative to traditional sequencing techniques</td>
</tr>
</tbody>
</table>
7. General Approach to Molecular Pathology

Liquid Bead Microarrays

Liquid bead arrays consisting of sets of labeled polystyrene microbeads can be used to simultaneously interrogate numerous targets for genotyping, pathogen identification, and other purposes. As with traditional flow cytometry, beads are passed single file through a laser-based detector. One laser identifies the bead based on its unique color while another quantifies attached analyte. Besides nucleic acids, proteins, cytokines, and other biomolecules can be attached to the beads, enhancing the multiplexing power of this technology.

Mass Spectrometry

Although not “molecular pathology” - the diagnosis of disease using nucleic acids - in the strict sense, proteomic profiling with mass spectrometry is touted by some as a powerful tool for the screening and diagnosis of malignancies and the identification of pathogens. Numerous variants of mass spectrometry separate proteins by mass and charge (Fig. 7.7a). The resultant protein mass fingerprints reveal the relative quantities of the protein and peptide species present (Fig. 7.7b). Such patterns derived from blood may constitute proteomic disease signatures. Confirmation of these sometimes controversial, nascent diagnostic approaches is needed before implementation in the clinical laboratory.

Systems Biology Approaches

The emerging field of systems biology seeks to more fully understand life in terms of its myriad molecular networks. This outlook contrasts with the reductionist bent of molecular biology and the common clinical practice of employing one or a few biomarkers to diagnose a given disease. Much more predictive, less reactive medical care is promised by use of multiparameter, high-throughput molecular tools. Although molecular diagnostic assays now serve primarily as adjuncts to other methods in surgical pathology, the use of global, systems-based approaches may challenge the dominance of histopathology.

Implementing Molecular Pathology

Laboratory Design Considerations

To prevent contamination, laboratories manipulating amplicons generated by PCR or other target amplification technologies must strictly observe a unidirectional workflow, from sample processing to detection (Fig. 7.8). Reagent preparation should ideally be done in an area of the laboratory separate from sample processing and nucleic acid extraction. Both these activities benefit from positive air pressure or “still boxes” and ultraviolet light to minimize airborne contaminants. Most importantly, the laboratory’s postamplification room, where reaction vessels are opened after PCR, must be kept separate from the rest of the laboratory. Dedicated gloves and gowns should be kept in this area. In addition, negative air flow can prevent escape of amplicons.

Fig. 7.7. (a) Mass spectrometry can be used to separate serum proteins. Following ionization or other means of dissociation, proteins move through the mass analyzer based on the ratio of mass to charge. A detector measures this time of flight (TOF). (b) Mass spectrometry TOF yields quantitative “proteomic profiles,” as shown in this simplified diagram. The species with the lowest mass/charge ratio (triangle) is more abundant than the two species with higher mass/charge ratios (rectangle and diamond). Serum proteomic signatures may be composed of the shed proteins and peptides of cancer and other diseases.

Fig. 7.8. Unidirectional workflow in the molecular pathology laboratory prevents contamination of new samples and reagents with amplicons produced in previously performed polymerase chain reactions (PCRs).
Laboratory Staffing

Early molecular pathology laboratories relied on staff trained in molecular biology to perform and troubleshoot assays. Now, those with more traditional medical technologist training increasingly perform these functions in part because of better educational opportunities. The American Society for Clinical Pathology, the American Board of Bioanalysis, and the National Credentialing Agency for Laboratory Personnel offer certification in molecular diagnostics. In addition, there are several degree-granting programs. The Training and Education section of the Association for Molecular Pathology website provides an updated list of molecular training available for technologists (http://www.amp.org/T&E/training&edu.htm).

References

Introduction

Molecular testing in anatomic pathology is becoming increasingly important, and molecular assays are used both diagnostically and prognostically, particularly in the workup of neoplasia and for identification or subclassification for certain infectious processes. Fresh and frozen tissues are the optimal source of DNA and RNA to serve as the template for targeted molecular analysis; however, archival formalin-fixed, paraffin-embedded tissue is a frequently used alternative source of DNA for clinical testing. Paraffin-embedded tissue is especially critical as a source of nucleic acid when pathological evaluation renders an unexpected diagnosis, and it also provides the advantage of allowing for archival analysis with correlation to outcome.

Molecular analysis can be performed at different levels of resolution, from the whole chromosome down to the specific nucleotide sequence. At the chromosomal level, classical cytogenetics discerns chromosome structure and number and can detect many major translocations and deletions. This technique requires fresh cells that are capable of entering into cell division. The cells are arrested in metaphase so that the chromosomes can be visualized individually for analysis. A newer technique that can be combined with classical cytogenetics is spectral karyotyping (SKY). For this specialized analysis, the metaphase chromosomes are subjected to fluorescent in situ hybridization with specific probes that are labeled with combinations of five different fluorescent tags. The result involves the “painting” of each chromosome with a unique fluorescent signal that can be differentiated with the assistance of computer detection system. This technique allows for resolution of complex karyotypes.

In situ hybridization [chromogenic (CISH) or fluorescent (FISH)] enables detection of known sequences of DNA and can be performed on whole cells or sections cut from paraffin or frozen tissue. The probes used in ISH can highlight amplifications, specific translocations mutations, or genomic material from infectious organisms. In paraffin-embedded tissues, FISH is a particularly powerful tool for molecular analysis, given the known difficulties in DNA extraction from fixed material.

Comparative genomic hybridization (CGH) is another relatively new technique that can detect and map changes in copy number for specific DNA sequences. The DNA from a test genome (e.g., tumor) and a reference genome (e.g., normal tissue) are differentially labeled and hybridized to normal metaphase chromosomes. In the past few years, microarray-based formats for CGH (array CGH) have been developed and are beginning to be widely used in preference to chromosome-based CGH. Overrepresentation (amplifications) or underrepresentation (deletions) of the test sample signal can be resolved with computer software. This technique is excellent for discovery of unique genetic alterations (deletions and amplifications), but it is not commonly used as a diagnostic tool.

In anatomic pathology laboratories, genetic material is often analyzed at the nucleic acid level, especially with the ubiquitous technique of polymerase chain reaction (PCR). Currently, many of the PCR-based diagnostic molecular tests yield positive or negative results to identify disease-specific genetic changes.

As molecular testing becomes more focused on quantitation of molecular targets, the need for relatively pure cell populations will increase. Tissue microdissection is an excellent method to obtain relatively pure cellular samples of morphologically confirmed cell types. This cellular purity results in more accurate test results and will limit the contamination from normal DNA that usually reduces assay sensitivity. Microdissection can be performed in a variety of ways, all of which have different advantages and disadvantages that have been reviewed elsewhere. Basically, these methods range from simple and inexpensive manual methods to laser-capture microdissection (LCM) methods that require expensive and complex equipment. Microdissection of target tissue is followed by DNA or RNA extraction. The common sequence of tissue processing in anatomic clinical or developmental molecular pathology laboratories is illustrated in Fig. 8.1.
Anatomic Pathology Testing to Detect or Characterize Neoplasia

Rapid advances in molecular technologies and expanding knowledge about lung tumor carcinogenesis have resulted in many studies of potential diagnostic and prognostic biomarkers. Based on current understanding, most tumors have multiple genetic and epigenetic alterations, including inactivation of tumor suppressor genes and activation of oncogenes. These molecular changes can be correlated with the sequential morphological changes of multistep carcinogenesis. The status of the many genes involved in tumorigenesis has been explored with multiple different molecular techniques, including loss of heterozygosity, array CGH, direct gene sequencing, gene expression profiling, fluorescent in situ hybridization, and proteomics. Each of these assays is discussed in the relevant chapters for specific assays throughout this book. In the following discussion, a more general overview of molecular testing in carcinomas is presented.

Oncogenes

“Proto-oncogenes” are wild-type, universally present genes that can stimulate carcinogenesis when mutated. The mutant forms are designated as “oncogenes.” Oncogenes act as dominant genes in which a mutation in only one copy of the gene leads to activation, often with overexpression of a protein product. A variety of mutational events can transform a proto-oncogene to an oncogene, including point mutations, translocations, and amplifications.

Detection techniques for oncogene mutations in tumor samples depend entirely on what type of mutation is occurring. For example, point mutations can be detected using techniques for gene sequencing. Translocations and amplifications are generally best detected in paraffin-embedded tissues using FISH-based techniques. Because many of these mutations are activating mutations and cause overexpression, the aberrant protein product may be detectable by immunohistochemistry (IHC). However, overexpression of an oncogene protein product by IHC alone does not necessarily signify a mutation, because other nonmolecular mechanisms can also be responsible for protein overexpression.

In surgical pathology, some of the most common oncogene assays detect translocations, particularly for sarcomas and hematological malignancies. These tumorigenic translocations often reposition an oncogene partner next to a constitutively active gene. The oncogene is then aberrantly and constitutively activated in the cells harboring the translocation. Depending on the clustering of breakpoints, translocation assays can use reverse transcription-polymerase chain reaction (RT-PCR), PCR alone, or in situ hybridization. Very few epithelial tumors are known to harbor consistent translocations.

Most oncogenes that have been associated with epithelial tumors involve specific point mutations in notorious genes. One of the most prevalent oncogene mutations occurs in the KRAS gene. KRAS gene mutations occur in many different tumor types, including lung carcinomas, pancreatic carcinomas, thyroid tumors, and colon carcinomas. KRAS mutations usually involve one of three important codons (12, 13, or 61). In some tumor types, such as lung carcinoma, KRAS mutations are known to associate with the genetic damage from tobacco. In other organ systems, such as colon carcinoma, there is an emerging role for KRAS testing for predicting nonresponse to some targeted drugs.

The clustering of mutations in KRAS mutations predominantly in exon 2 of the gene makes it feasible to use a simple and rapid PCR-based sequencing assay for detection of these mutations. Even in small or low cellularity samples, such as cytology samples, sequencing assays can be performed. One caveat to all traditional sequencing reactions for detecting oncogene mutations, however, is that the lower limits of sensitivity are around 20–25% in most studies. This limitation means that samples that are heavily contaminated with normal adjacent tissue may yield false-negative results using a sequencing assay for point mutations.

Unfortunately, KRAS mutations alone are not specific for carcinoma in most organ systems. In lung carcinoma, for example, they have also been detected in the sputum of patients whose tumors are negative for KRAS mutations and even in individuals with no clinically detectable carcinoma. This finding highlights the fact that mutations in KRAS are only one component of the carcinogenesis pathway.

Tumor Suppressor Genes

Tumor suppressor genes (TSG) in the wild-type state have two active copies (alleles). These genes are thought to functionally suppress carcinogenesis in their routine cellular activity. Because both copies of the TSG must be mutated
for tumorigenesis, TSGs are designated as recessive genes, in contrast to dominant proto-oncogenes, in which only one mutation is needed. We rarely test for both mutations when assessing the status of TSGs.

The most common assays for identifying TSGs include loss of heterozygosity (LOH), CGH, array CGH (aCGH), and FISH. LOH studies have been used to identify and define the location of the many TSGs involved in carcinogenesis. Alterations in TSGs are generally thought to be cumulative and progressive throughout tumorigenesis in many different organ systems. Advanced tumors frequently show widespread complete or partial loss of each chromosomal arm, whereas precursor lesions tend to show more focal and smaller chromosomal losses.

Molecular Anatomic Testing for Targeted Therapies

The advancement in development of novel targeted therapies, such as the inhibitors of the epidermal growth factor receptor (EGFR) and other tyrosine kinases, has opened a new chapter for the molecular testing of tumors. It is clear from the experience with targeted therapy for breast cancer that standardized assays for assessing and predicting the effects of therapeutic agents should be ideally developed in parallel with targeted therapies. The first assays used in the search for biomarkers for drug therapy with EGFR inhibitors were tests for EGFR protein expression by IHC and gene copy number by FISH. There is a good correlation between IHC and FISH results, meaning that amplification often does correlate with increased protein expression in some tumor systems, such as lung carcinoma and glioblastoma multiforme. Unfortunately, these assays have not proven to be useful in predicting patient response to EGFR inhibitors. EGFR activating point mutations have also been described, but these appear to be exclusive to lung cancer; only extremely rare mutations have been identified in other types of cancer, particularly in colorectal carcinoma. The same deletion in exon 19 as is seen in lung cancer has also been rarely detected in squamous cell carcinoma of the head and neck.

Anatomic Pathology Testing for Infectious Agents

As microbiology has become more and more reliant on molecular technology, sophisticated assays can also be readily applied to tissues in anatomic pathology. Molecular testing for infectious agents is increasingly useful as a diagnostic tool. IHC and ISH remain the popular ancillary techniques for surgical pathologists, because they can be applied to routine diagnostic material coupled with histological evaluation. However, paraffin-embedded tissues can also serve as excellent source material for PCR-based organism identification assays; this is particularly useful when tissue was not sent directly for cultures or when suspicion is high but IHC and ISH are negative or inconclusive. Nucleic acids extracted from formalin-fixed paraffin-embedded tissue are usually more fragmented than the high molecular weight DNA that can be obtained from fresh or frozen tissue. Several laboratories are now equipped to do testing for Mycobacteria species and viruses, such as cytomegalovirus (CMV), from paraffin-embedded tissues, through specially designed and optimized assays for fragmented DNA.

Despite the increase in isolates of Mycobacterium tuberculosis in the United States since 1985, there are now more isolates of nontuberculous mycobacteria (NTB) such as Mycobacterium avium complex. The diagnosis of tuberculosis for many years has been based on special stains of smears for acid-fast bacilli (AFB) and mycobacterial cultures. The special stains for AFB are not sensitive and do not allow the identification of the different species of mycobacteria. Mycobacterial cultures are specific, but results are usually not available for 2–3 weeks or longer. PCR using oligonucleotide primers specific for gene fragments in M. tuberculosis has been used to identify tuberculosis organisms from archival formalin-fixed paraffin-embedded tissue. This method may detect organism with high specificity and sensitivity and is much faster than conventional cultures. The sensitivity of PCR is better than that of the special stains, in which a few mycobacteria might be missed on routine histological sections. However, PCR results should always be interpreted in a clinical context. In diagnostic molecular pathology, sampling error is an important source of false-positive and false-negative results. The percentage of false-negative results after PCR amplification of M. tuberculosis from formalin-fixed, paraffin-embedded tissue varies among studies from 2 to 19%, depending on the PCR technique employed. The DNA extraction used for this method is performed from several whole thick sections, as the rate of PCR positivity seems to be at least partially related to the quantity of DNA used for amplification. Therefore, PCR techniques may have some disadvantages such as possibly increased effect of tissue inhibitors or increased possibility of false-positive results caused by contamination. Because of the relatively low amount of template DNA needed for amplification, PCR-based analysis is now increasingly performed on microdissected areas from tissue sections. For surgical pathologists, a molecular identification of mycobacteria can be the method of choice when the tissue obtained by biopsy is not sent for culture.

Several types of amplification assays are available for quantitation of CMV and other viruses, most of which use a form of real-time or quantitative PCR. Currently, most assays are designed for peripheral blood, and therefore surgical pathologists are rarely involved in ordering these tests.

References


Polymerase Chain Reaction

Polymerase chain reaction (PCR) enables one to determine if a specific needle is present in a haystack, and it can be used as a step toward the characterization of the needle. It is a quick, powerful, inexpensive DNA amplification technique that has become a fundamental tool in molecular pathology.

Theory

The PCR is one of the most significant technical innovations in molecular biology. The PCR was devised by Kary Mullis and colleagues at Cetus Corporation in California and was first described in a 1985 paper demonstrating its application in the prenatal diagnosis of sickle cell anemia and then further described in an ensuing paper. These works detailed how a DNA sequence could be enzymatically amplified in vitro using specific oligonucleotide primers and bacterial (Klenow) DNA polymerase. With refinement of PCR over the next 3 years, it was found that a robust PCR using a thermostable polymerase could amplify a DNA sequence by a factor of more than $10^7$, even when the target DNA made up only 1 of 100,000 DNA strands in a reaction. Since then, additional improvements and variations to the original reaction have been made, affording even more efficiency, sensitivity, and utility to this tool. Its application specifically to diseases of the lungs has ranged from detection of infectious diseases to study of inflammatory mechanisms to use in mutation analysis, to detection of tumors and metastases.

Principles

The principle of PCR is illustrated in Fig. 9.1. The target DNA to be amplified in vitro can be human genomic, bacterial, viral, plasmid, or previously PCR-amplified DNA and is represented in the figure by the target’s nucleotide base letters A, C, G, and T and the sugar phosphate backbone. Other components of PCR include a thermostable DNA polymerase such as Taq polymerase, two oligonucleotide primers, four deoxynucleotide triphosphates, magnesium, buffer, and a thermocycler. PCR achieves amplification of the DNA by repeating a three-step cycle over and over. These three steps are denaturation, annealing, and extension.

In the denaturation step, target double-stranded DNA (dsDNA) is heated to a high temperature (94–95°C) to break the hydrogen bonds between nucleotide bases on opposing strands. The dsDNA denatures, splitting into two intact single strands (ssDNA) that are complementary to each other. In the second (annealing) step, the reaction mix is cooled, typically to 50–65°C, allowing ssDNA oligonucleotide primers to bind (anneal) to the portion of the target ssDNA to which they have specifically been designed. The investigator must know the sequence of the target DNA (at least in the region of the primers) to design the primers. Two primers are required: one that is complementary to the 3'→5' oriented target strand and the reverse primer. Because these primers are in overwhelming abundance compared with the target DNA, they will anneal to the target DNA much more frequently than the full-length target DNA will anneal to its full-length complementary strand as the reaction is cooled. In the third (extension) step, as the temperature is raised to its working optimum (72°C), Taq polymerase recognizes these partially double-stranded DNAs and uses the forward and reverse primers as initiation points to begin extending the primers via polymerization in a 5'→3' direction along the two new (nascent) strands. DNA polymerase does this by selecting high-energy deoxynucleotide triphosphates (dNTPs) from solution and placing them in the nascent strands directly across from their complementary base in the template (target) strand. During the first round (cycle) of PCR, the DNA polymerases extend the nascent strands for a relatively long distance before falling off. At the end of the first cycle, two double-stranded copies of a portion of the target DNA have been generated from one copy.
Fig. 9.1. Polymerase chain reaction steps. A single double-stranded DNA fragment is denatured and cycled through three steps – denaturation, annealing, and extension – to create an exact copy of the original. Taq polymerase extends primers by selecting deoxyribonucleotide triphosphates from the reaction solution based on the nucleotide sequence of the target DNA.
During the second cycle (Fig. 9.2, where DNA strands are symbolized by lines) the same process happens, but now the polymerase can only proceed as far down the DNA as the point where the opposite primer started. At the end of the second cycle the PCR products (amplicons) include two relatively short fragments of ssDNA whose two ends now correspond exactly to the locations of the forward and reverse primers. Both the genomic DNA and new amplicons can serve as templates in successive PCR cycles. In a perfect PCR the amount of dsDNA doubles with each cycle so that after 30 cycles there are more than 1 billion copies of the original dsDNA (2^{30} = 1.1 billion) and more than 1 trillion copies after 40 cycles (2^{40} = 1.1 trillion). Because the primers recognize only the target DNA they are designed for, only a specific segment of DNA is amplified, even if it makes up only a fraction of all the different DNA sequences in a reaction. This preferential amplification greatly facilitates post-PCR analysis of the target sequence.

PCR cycling is done in a thermocycler, a small automated tabletop instrument programmed by the investigator. A typical program (Table 9.1) starts with a 5–10-min denaturation at 94–95°C to ensure that the majority of the DNA (especially large chromosomal strands) is denatured. This step is followed by 30–50 cycles of brief denaturation/annealing/extension. The PCR concludes with a 5–10-min final polymerization step at 72°C to ensure nearly all amplicons are extended to their full length.

**Table 9.1** Typical polymerase chain reaction (PCR) thermocycler program.

<table>
<thead>
<tr>
<th>Step</th>
<th>Temperature (°C)</th>
<th>Time</th>
</tr>
</thead>
<tbody>
<tr>
<td>Initial denaturation</td>
<td>95</td>
<td>10 min</td>
</tr>
<tr>
<td>Forty cycles of Denaturation</td>
<td>95</td>
<td>30 s</td>
</tr>
<tr>
<td>Annealing step</td>
<td>57</td>
<td>30 s</td>
</tr>
<tr>
<td>Extension step</td>
<td>72</td>
<td>45 s</td>
</tr>
<tr>
<td>Final polymerization</td>
<td>72</td>
<td>10 min</td>
</tr>
<tr>
<td>Post-PCR hold</td>
<td>4</td>
<td>Indefinite</td>
</tr>
</tbody>
</table>
Practical Polymerase Chain Reaction

Thermocyclers must be properly programmed to create an efficient PCR. Most programs are roughly similar (see Table 9.1), except for the annealing temperature and time intervals of each step. The annealing temperature is dependent on the melting temperature of the primers, and the time spent at the different steps is dictated by the size of the target DNA. The denaturation temperature is generally the same in all PCRs (94–95°C), just as the extension temperature is usually 72°C, because most polymerases used in PCR work best at this temperature (but see later discussion of polymerases and real-time PCR).

Denaturation Step Programming

The initial denaturation, which occurs before any cycling, is typically 94–95°C for 5–10 min; this gives human chromosomal DNA time to unravel and split into single strands. Smaller (e.g., viral) DNA targets may require only 3–5 min. Because ssDNA tends to reanneal while cooling, the dsDNA must be redenatured at 94°C for 10–60 s at the beginning of each cycle. GC-rich targets may require hotter/longer denaturations, while formamide or dimethylsulfoxide (DMSO) can be added to promote denaturation. Temperatures above 95°C or excessive cycles should be avoided because Taq stability decreases under these conditions.

Annealing Step Programming and Primer Design

The annealing temperature varies with every PCR and is a critical factor in the PCR program, as it helps dictate lower limit of detection, sensitivity, and specificity of the PCR. It must be calculated for each unique PCR and is dependent on the melting temperature of the two oligonucleotide primers. The melting temperature \( T_m \) of a DNA fragment is the temperature at which half of it has denatured into the single-stranded form (e.g., the primer is not annealed to its complementary target) and half is still double-stranded (primer is annealed to its target), assuming the number of copies of the two complementary strands is equal. Melting temperature calculation of an oligonucleotide can be very complex, with formulas that employ thermodynamics and depend on nearest neighbor nucleotides and the salt concentrations in the reaction. Numerous websites, including those of companies that manufacture custom oligonucleotides, have free \( T_m \) calculators; after a primer sequence is typed in, the \( T_m \) is instantly calculated. Software is also available that will generate a list of potential primer pairs, with their \( T_m \)s, once the entire target sequence is entered.

Manual calculation of the approximate \( T_m \)s of short oligonucleotides can also be done using the abbreviated Wallace rule:

\[
T_m = 2(A + T) + 4(C + G)
\]

Here \( T_m \) is in degrees centigrade, and A, T, C, and G stand for the number of each of these bases in the single-stranded oligonucleotide sequence. Thus, the approximate \( T_m \) of a 22-base pair (bp) (22-mer) poly (A) oligonucleotide would be only 44°C, whereas the 21-mer CACGCTGACGCTGACGCTGACGCTG would have an approximate \( T_m \) of 76°C. Because Cs and Gs base pair with three hydrogen bonds, more heat is required to melt them apart in a dsDNA → ssDNA conversion compared with As and Ts, which base pair by sharing only two hydrogen bonds.

To design primers, one must first isolate the region of the gene of interest that will be amplified. The human genome and the genomes of many viruses and bacteria are available on the National Center of Biotechnology Information (NCBI) web site (www.ncbi.nlm.nih.gov). Avoid placing primers in regions with known polymorphisms or splice variants or, in the case of viruses and bacteria, where any subtype/strain variations have been reported in the region of the proposed primers.

To ensure genomic DNA (and not reverse-transcribed mRNA) is being amplified, primers should be located at exon–intron junctions or within introns. To ensure reverse-transcribed RNA is being amplified (from the cDNA; see later), primers should span introns and be within exons. To create a robust PCR, the size of the DNA region amplified (distance of the two primers from each other) should be less than 500 bp and preferably around 200–300 bp. Amplicons larger than 1 kb (kilobase, or 1,000 bp) often require special polymerases (see later) with enhanced processivity. These forward and reverse primers should be only 18–25 bp long, have a 40–60% G + C content, and have \( T_m \)s in the 55–65°C range. Melting temperatures outside this range may work, and in fact are sometimes necessary, but they have a greater likelihood of resulting in a less efficient or more nonspecific PCR. In addition, the primers should have \( T_m \)s that are within 2°C of each other: add or remove bases to meet this goal. Ideally, the last five bases in the primer should include three Cs or Gs, and the 3’-end of the primers should be a C or G to promote tight base pairing at the point of Taq recognition and initiation of polymerization.

Just as a primer will anneal to its complementary region in the target DNA, so a pair of primers might anneal to complementary regions within themselves or within each other. Thus, TGGCCCATTACACTTGGCCATT is a poor primer choice because there will be some tendency for the boldface sequences to anneal to each other, forming a hairpin stem-loop structure within itself or to form primer dimers between two similar primers. Likewise, a 5’-TAGG-3’ sequence in the reverse primer could transiently anneal to a 5’-CCTA-3’ sequence in the forward primer. Many of the web sites that calculate \( T_m \)s also have a tool to check for these types of structures, as these aberrant forms can significantly reduce the yield of PCR product due to effective reduction of the available primer supply. Finally, avoid repetitive bases at the 3’-end of a primer, as this promotes slippage (“out of register”) errors by Taq polymerase. Once all (or as many as feasible) of the foregoing rules are met and two primer
sequences have been found, the primers should be checked to be certain they are not complementary to DNA sequences unrelated to the target sequence by using the NCBI BLAST (basic local alignment search tool) database at www.ncbi.nlm.nih.gov/BLAST/. This website will list all published DNA sequences that exactly or closely match the oligonucleotide sequences submitted. If one or both primers are close matches to nontarget DNA that may be present in the samples tested, new primer(s) may have to be designed.

The rules are empirical and do not guarantee a large PCR yield for reasons that are not always obvious or easily tested. It is often more expeditious to simply design more than one set of primers using the above rules and test all on a specimen to determine which pair gives the most robust PCR.

What annealing temperature should now be used in the PCR? This choice also requires trial and error. Initially the annealing temperature programmed into the thermocycler should be 5–10°C below the lowest mole fraction of the forward and reverse primers. At such a low temperature there will be mispriming onto nontarget sequences, resulting in amplification of nonspecific products. To enhance the specificity and perhaps even the sensitivity of the PCR, the temperature should then be systematically raised until amplicon yield drops and, hopefully, nonspecific amplicons disappear. Alternatively, the annealing temperature can be lowered, potentially increasing sensitivity, at the risk of generating nonspecific amplicons in addition to the desired PCR product.

Polymerase Chain Reaction Components

Thermostable Polymerase

The first PCRs used an Escherichia coli DNA polymerase that was thermolabile and had to be replaced at each cycle; extension at 37°C would allow the nonspecific priming of numerous genomic sites and thus formation of nonspecific amplicons. Because the cycling of PCR repeatedly raises the reaction to 95°C, a thermostable polymerase is required. Taq polymerase, the most frequently used, was first purified from the bacterium Thermus aquaticus in 1976, years before PCR. The enzyme has optimal 5' → 3' polymerase activity at 80°C (but will inefficiently extend primers at much lower temperatures), requires a divalent cation (Mg2+), extends at a rate of 60 nucleotides/s, and has a polymerization per binding event (processivity) of 50–80 bases. Taq has double-strand-specific 5' → 3' exonuclease activity (see later discussion of real-time PCR) but does not have 3' → 5' exonuclease (proofreading) ability; its estimated error rate is 2.1 × 10^-4 errors per base per duplication. Taq is inhibited in samples containing heparin, hemoglobin, phenol, ethanol, urea, Triton X-100 are often used, as well as gelatin. Fortunately, the above components (depending on manufacturer) are included in optimized 10× PCR buffer solutions provided by most Taq polymerase suppliers and do not have to be added individually to the reaction.

Magnesium

The magnesium concentration is important because it affects PCR specificity and efficiency through its interaction with Taq, whose function is dependent on the divalent cation. A free [Mg2+] of 1.2–1.3 mM is optimal for Taq, whereas much higher Mg2+ levels result in increased error rates caused by base substitutions and frameshift errors. Magnesium is often included in 10× PCR buffer solutions at a [Mg2+] of 1.5 mM, which decreases to 1.3 mM in the presence of 0.2 mM dNTPs because of equimolar binding of dNTPs and Mg2+ by Taq polymerase.

Polymerase Chain Reaction Setup

Before starting experiments using PCR, accommodations must be made in the laboratory. Because of the amplification power of PCR, contamination of even minute amounts
of DNA from one sample to the next must be avoided. Master mixes, containing all components of the PCR except the DNA, should be set up in a separate, dedicated room or area (hood) away from specimens or post-PCR solutions. DNA or RNA should be isolated from samples in a second room or area. DNA of the samples can then be added to the master mix in a third area, ideally. The thermocycler should also be in a separate room or area of the lab. Keep a unidirectional flow of material from pre-PCR to post-PCR: do not allow PCR-amplified material into the master mix preparation or DNA isolation areas. Label dedicated pipetters and use only aerosol-resistant tips to prevent contamination of pipettor barrels. Wear protective disposable gloves at all times, and change them and laboratory coats when going from one room or area to the next. Use ultraviolet irradiation inside hoods or on benchtops to destroy possible contaminating DNA. Use only autoclaved molecular biology-grade water in master mixes. Make certain that all plastic tubes and tips are DNAse- and RNase free, as these enzymes will digest the target nucleic acids in specimens.

There is no universal recipe for the PCR mix, but a typical mix, including the DNA, is shown in Table 9.2. However, this is just a start. All PCRs need some adjusting – annealing temperatures, cycle step times, reagents (concentration of each primer, Mg²⁺ concentration, etc.), primer sequences – to determine the optimal conditions to give the greatest yield, fastest time, or highest specificity.

### Postprocedure Analysis

PCR alone does not provide answers to an investigator’s questions, but, because of the huge increase in a specific product, it makes the product’s analysis much easier. To prove that the target sequence was present in a specimen, amplicons are very useful, but, because of the huge increase in a specific product, PCR alone does not provide answers to an investigator’s questions. Variations and improvements in PCR have been introduced over the past 20 years to meet the needs of researchers and clinical molecular diagnostics labs.

**Hot start PCR** is a technique preventing Taq from extending primers until a temperature of 60–80°C is reached, usually done by withholding Taq from the reaction until these temperatures have been reached. This step prevents Taq extension of primer dimers or primers that have annealed to nonspecific regions of the specimen DNA at low temperatures, such as during the preparation of the reagent mix. The result is improved specificity and yield of the PCR and is especially helpful when the target DNA is a small percentage of the total DNA. There are two common hot start methods. The first uses a wax plug to separate key reagents (e.g., dNTPs from Taq) in the PCR tube; the wax melts at a temperature well above the primer Tm and allows mixing of all reagents. The second method uses a modified Taq that is activated only after the initial 95°C 10-min denaturation step.

**Nested PCR** uses two pairs of primers to improve amplicon yield and specificity. The first pair (“outer primers”) is designed to amplify a larger fragment of the target DNA. These amplicons are then used, usually in a second PCR, as the template DNA for the second set (“inner primers”), which necessarily makes a smaller PCR product. Even if two sequential PCRs are run there may still be four products in the final PCR, the smallest being defined by only the inner primers. If all four primers are added at once and only one PCR is run, the Tm of the inner primers should be lower than that of the outer primers.

**Methylation-specific PCR** is used on genomic DNA to determine if the CpG islands within the promoter of a gene are methylated on the cytosine residue (blocking the gene expression in the cell) or unmethylated (potentially allowing gene expression). For example, imprinted genes and silenced genes on X chromosomes are methylated, as are the promoters of tumor suppressor genes in many cancers. Methylation-specific PCR is based on the realization that a cytosine converts to a uracil after bisulfite treatment, whereas a methylated cytosine (5-methylcytosine) is refractory and remains as cytosine. An unmethylated CpG will be converted to UpG after bisulfite, but a methylated one will remain as CpG. The change in DNA sequence of an unmethylated promoter compared with a methylated promoter after bisulfite treatment allows one to design primers that can discriminate between the two. Whichever of the two primers yields a PCR product indicates the methylation status of the promoter.

**Multiplex PCR** amplifies multiple different regions of DNA at one time by using multiple primer pairs in one reaction. Several targets can be analyzed in one specimen, including housekeeping genes and variably expressed genes, multiple microorganisms, or multiple mutations in a genetic disorder or malignancy. Although ostensibly a timesaver, designing primers and optimizing reaction conditions so as

### Table 9.2. Typical polymerase chain reaction (PCR) reagent mix.

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>H₂O (nuclease-free) to final total volume of</td>
<td>25.0 µl</td>
</tr>
<tr>
<td>dNTPs at 5–10 mM each (20–40 mM total)</td>
<td>0.5 µl</td>
</tr>
<tr>
<td>Forward primer at 25 µM</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>Reverse primer at 25 µM</td>
<td>0.3 µl</td>
</tr>
<tr>
<td>10x PCR buffer with Mg²⁺</td>
<td>2.5 µl</td>
</tr>
<tr>
<td>Taq polymerase at 1 unit/µl</td>
<td>1.0 µl</td>
</tr>
<tr>
<td>Target DNA 100–500 ng</td>
<td>1.0 µl</td>
</tr>
</tbody>
</table>

*One unit of Taq is defined as the amount of enzyme that will incorporate 10 nmole of deoxynucleotide triphosphates into acid-insoluble material in 30 min at 75°C*
to ensure equally robust amplification of all targets in multiplex PCR is a challenge. First, design individual primer pairs (but avoid primer dimers with every other primer), and program the thermocycler to allow optimal amplification of all targets on an individual basis. Then, combine (equimolar) primers and run the same program to see which targets are weakly amplified. Adjust primer concentrations, annealing temperatures, \([\text{Mg}^{2+}]\), and so forth to equalize yields of targets. Primer software is also available, but in all cases trial and error are necessary to arrive at the final setup.

Other variations of PCR, each with its unique benefits, include the amplification refractory mutation system,\(^{55}\) allele-specific oligonucleotide probes,\(^{56}\) rapid amplification of cDNA ends,\(^{57}\) and in situ PCR.\(^{58}\) One variant that deserves special attention is real-time PCR.

**Real-Time Polymerase Chain Reaction**

Real-time PCR\(^{59-62}\) is a recent innovation that has quickly become very popular in molecular biology research and molecular diagnostics. It circumvents the need for time-consuming post-PCR analysis, and it can detect DNA targets, quantify the original (before amplification) copy number of the target DNA present in a specimen, or detect specific mutations. Although very similar to conventional PCR, real-time PCR is based on two additional principles. First, Taq polymerase has a bonus 5'-exonuclease activity on partially double-stranded DNA\(^{63}\); it will digest it to a single-stranded target just before polymerizing it into a double-stranded amplicon. Second, energy can be transferred between fluorescent molecules (fluorophores) attached to oligonucleotides when these small DNA fragments are used as probes against specific DNA targets. This second principle, first proposed in 1948 by Förster\(^{64}\) and later supported by Stryer and Haugland,\(^{65}\) indicates that electronic excitation energy can be transferred over short distances (≤50 Å) via dipole–dipole resonance between an energy donor and acceptor chromophore.\(^{66}\) Transfer efficiency is proportional to the inverse sixth power of the distance separating the donor and acceptor.\(^{65}\)

Fluorescence resonance energy transfer (FRET) in real-time PCR is demonstrated in the five frames in Fig. 9.3. A single oligonucleotide (TaqMan\(^{8}\) type) probe is designed so that it specifically anneals to its complementary region of a target DNA, somewhere between the forward and reverse primers (top frame of Fig. 9.3). A donor/emitter fluorophore covalently attached to one end of this oligonucleotide is stimulated by monochromatic light from a laser, and the energy is transferred to a quencher fluorophore at the other end. In this kind of probe, light is either not reemitted by the quencher or is reemitted at a wavelength different from that of the donor fluorophore. During the primer extension step of PCR (see Fig. 9.3, second frame) the probe remains intact, and the light remains quenched, until degraded into single nucleotides by the 5'-exonuclease activity of Taq as it passes through during polymerization of the target DNA (see Fig. 9.3, third frame). Once cleaved, the emitter nucleotide drifts too far away for its emissions to be quenched, and its fluorescence is recorded in each PCR cycle by a sensitive photodetector and entered into the system's computer. With each PCR cycle, the number of amplicons and the quantity of light double as more probes are cleaved.

Dual-probe real-time PCR (LightCycler\(^{6}\) type or hybridization probes; see Fig. 9.3, last two frames) uses donor/emitter and acceptor probes, designed to anneal side by side on the same target DNA. Fluorescent energy is transferred from the fluorophore of the 5'-donor probe (on the left) to the 3'-acceptor probe (on the right) during the annealing step of PCR. The acceptor immediately reemits at a wavelength unique to its fluorophore, and its signal strength doubles as the number of amplicons doubles with each PCR cycle. The probes separate, and FRET is terminated as the temperature is increased for the denaturation step.

Whether in single- or double-probe real-time PCR, the unique emission spectra of all fluorophores are captured and analyzed throughout each cycle by the real-time instrument software to generate curves such as those in Fig. 9.4. Cycle number is plotted versus fluorescence, and each of the colored curves represents a different specimen. Flat lines indicate no target DNA was in the specimen, and curves that begin rising at a low cycle number represent specimens that had more initial target DNA than those curves that rise later. Note that an exponential increase in fluorescence (and thus amplicon number) occurs only after numerous cycles, when PCR becomes its most efficient. To compare specimen target DNA copy number, a single horizontal threshold line can be drawn through all curves at this exponential phase and then a vertical line dropped from each intersection point to the X (cycle number)-axis. This point on the X-axis is the cycle threshold number of the specimen, and it increases as the original DNA copy number in a specimen decreases. Real-time PCR can be quantitative: a series of controls with a range of known DNA copy numbers are run to generate a standard curve of cycle threshold versus copy number. The cycle threshold values of specimens run simultaneously with the controls can then be converted into original DNA copy number by extrapolating from the standard curve.

Real-time instruments are able to detect fluorescence over a broad range of amplification: dynamic ranges of 7–8 logs can be obtained\(^{61}\) (e.g., specimen DNA copy numbers from 10\(^4\) to 10\(^8\) or from 10\(^2\) to 10\(^9\) can be detected on the same run). In addition, probes are able to detect even a single base pair change in target DNA, such as in allelic discrimination.\(^{67,68}\)

In the design of real-time PCR, the target should be short, preferably less than 150 bp, to maximize amplification efficiency. In some real-time instrument programs the annealing and extension steps are combined into one and run at a temperature of ~60°C. Probes should be designed before the primers and can bind the sense or antisense strand. They
Fig. 9.3. Real-time polymerase chain reaction (PCR). The top three frames demonstrate single-probe real-time PCR. Quenching ceases once the nucleotide bound to the emitter (yellow) is cleaved by Taq. In the third frame, the forward portion of Taq cleaves the probe (releasing nucleotides), while the back part of Taq extends the forward primer (consuming other nucleotides). The bottom two frames show a dual-probe assay in which probes attach to the target DNA and transfer energy from the yellow to the red fluorophore during the annealing step. They float away during the extension or denaturation step, terminating fluorescence energy transfer, and will not be cleaved by Taq.
Polymerase Chain Reaction and Reverse Transcription-Polymerase Chain Reaction

should be ≤35 bp long, have a GC content of 30–80%, should not have runs of four or more of the same nucleotide, should not partially anneal to each other or to either primer, and should have the 3′-end blocked by phosphorylation to prevent extension by Taq. TaqMan® probes should have a \( T_m \) of 68–70°C (about 8–10°C higher than their associated primers, which should have a \( T_m \) of 58–60°C), and the 5′-fluorophore should not be bound to a G. LightCycler® probes should optimally be 1–3 bp apart, have a \( T_m \) between the extension and annealing step temperatures, and be 5–10°C higher than the primer \( T_m \); the \( T_m \) of the two probes should be within 2°C of each other unless mutation detection is desired. Mutation detection is facilitated by looking for differential melting (denaturation and consequently loss of FRET) of probes from wild-type versus mutant target DNA sequences (melting curve).

Reverse Transcription

All normal cells in a human’s body, with few exceptions, have the same chromosomal DNA sequence, that is, the same genetic code. Thus, genomic information obtained from the DNA of easily obtained normal white blood cells would be applicable to the genetic makeup of normal lung, brain, or colon cells. This idea does not apply to malignant cells, which can have a genetic composition that is profoundly different from that of normal cells.

Function and structure of various cell types differ because of the mRNA that they transcribe and ultimately the proteins that are translated. In other words, it is the protein expression profile of cells that differentiates them. Presently, the most practical way to study the specific genes expressed in a particular cell type is to analyze the mRNA the cells make. Because RNA is unstable and therefore difficult to work with in the laboratory, it can be converted into the complementary DNA (cDNA) by a process known as reverse transcription. The resultant cDNA is much more stable than the mRNA. Reverse transcription is so named because RNA is used as the template to direct the production of DNA: the reverse of normal cellular transcription, where DNA is used by RNA polymerase to direct the production of mRNA.

A reverse transcriptase enzyme is an RNA-directed DNA polymerase made and used by some RNA viruses to complete their life cycle within a host. Viral reverse transcriptases have been characterized and/or cloned, and the enzymes are commercially available for use in research and clinical molecular laboratories.

Reverse transcriptase, similar to DNA polymerase, requires a DNA primer (Fig. 9.5) to initiate its function. Because mRNA has a poly(A) tail at its 3′-end, an ideal primer for reverse transcription of mRNA species would be a poly(T) oligonucleotide (oligo dT). A replete collection of short DNA primers with random sequences can also be used; these primers are recommended if reverse transcription of ribosomal RNA (rRNA) is also desired along with mRNA. The enzyme starts transcription at the 3′-end of template RNA [the 5′-end of the nascent (new) cDNA strand] and proceeds in a 5′ → 3′-direction on the nascent strand ("first strand synthesis"). In this fashion, all the mRNA (or total RNA) present in a cell can be transcribed into complementary DNA. Those mRNA sequences that are present at a high copy number in the cell will be reverse transcribed to a high cDNA copy number compared with those mRNA sequences which are rare in a cell.

A typical reverse transcription protocol is given in Table 9.3. The two most commonly used reverse transcriptases are from bird and mouse viruses: avian myeloblastosis virus...
RNAse inhibitors and diethylpyrocarbonate-treated water are needed to preserve the unstable RNA. The initial 70°C heating is to remove secondary structures from the RNA; the 42°C (AMV and some MMLV products) or 37°C (some MMLV products) incubations are the working temperature of the enzymes. The 90–95°C step is needed to inactivate the enzymes.

The cDNA made by reverse transcription of mRNA (and/or rRNA) can then be used as a template for PCR if the appropriate primers for the target DNA are present. During the first cycle of the PCR, only one (the forward) primer is needed because only one strand is polymerized, but this new strand will serve as the template for the opposite primer during the second PCR cycle, and polymerization of both strands will continue with each cycle. [Note that some bacteria such as *Thermus thermophilus* have an enzyme (*Tth*) that can both reverse transcribe RNA and polymerize DNA, (AMV) and Moloney murine leukemia virus (MMLV).]

Their recommended buffers should not be interchanged. RNAse inhibitors and diethylpyrocarbonate-treated water are needed to preserve the unstable RNA. The initial 70°C heating is to remove secondary structures from the RNA; the 42°C (AMV and some MMLV products) or 37°C (some MMLV products) incubations are the working temperature of the enzymes. The 90–95°C step is needed to inactivate the enzymes.

**Table 9.3. Typical reverse transcription reaction protocol.**

<table>
<thead>
<tr>
<th>Reagent</th>
<th>Volume</th>
</tr>
</thead>
<tbody>
<tr>
<td>RNA 1–2 μg</td>
<td>1.0 μl</td>
</tr>
<tr>
<td>DEPC-treated H₂O</td>
<td>8.5 μl</td>
</tr>
<tr>
<td>Oligo dT or random primers at 40 μM</td>
<td>2.0 μl</td>
</tr>
<tr>
<td>70°C ↓ 5 min</td>
<td></td>
</tr>
<tr>
<td>4°C ↓ 5 min</td>
<td></td>
</tr>
<tr>
<td>Quickly add 20.0 μl of prepared RT master mix</td>
<td></td>
</tr>
<tr>
<td>DEPC-treated H₂O; add to total final volume</td>
<td>20.0 μl</td>
</tr>
<tr>
<td>MMLV (10×) or AMV (5×) buffer</td>
<td>2.0–4.0 μl</td>
</tr>
<tr>
<td>dNTP at 10 mmole</td>
<td>3.0 μl</td>
</tr>
<tr>
<td>RNAse inhibitor 10–40 units</td>
<td>1.0 μl</td>
</tr>
<tr>
<td>MMLV reverse transcriptase 200 units*</td>
<td>1.0 μl</td>
</tr>
<tr>
<td>AMV reverse transcriptase 30 units*</td>
<td></td>
</tr>
<tr>
<td>37 or 42°C ↓ 60 min</td>
<td></td>
</tr>
<tr>
<td>90–95°C ↓ 5 min</td>
<td></td>
</tr>
<tr>
<td>Use ~1–3 μl in the PCR reaction</td>
<td></td>
</tr>
</tbody>
</table>

*One unit of reverse transcriptase is defined as the amount of enzyme that will incorporate 1 nmole of deoxynucleoside triphosphate into acid-insoluble material in 10 min at 37°C using poly(rA), oligo(dT) as template primer.

**AMV** avian myeloblastosis virus; **DEPC** diethylpyrocarbonate; **dNTP** deoxynucleotide triphosphate; **MMLV** Moloney murine leukemia virus; **PCR** polymerase chain reaction; **RT** reverse transcription.

**Fig. 9.5. Reverse transcription.** Reverse transcriptase uses oligo dT as the primer on the target mRNA and polymerizes in the 5'–3' direction on the new DNA strand. The original mRNA strand is then cleaved by an RNAse domain within the reverse transcriptase (not shown), thus allowing polymerization of the single-stranded DNA into double-stranded DNA during polymerase chain reaction.
allowing reverse transcription and PCR to proceed simulta-
neously in a single tube.]

Reverse transcription-PCR (RT-PCR) is thus an impor-
tant tool that allows the investigator to study the genes
expressed or not expressed in specific cells after isolation of
the mRNA.\textsuperscript{80–83} Additional (post-PCR) techniques such as gel
electrophoresis, single-stranded conformation polymorphis-
microarrays, and so forth, can be applied to also determine if the genes expressed have mutations.

Under- or overexpression of a particular gene in neoplastic
or reactive cells can be investigated by comparing their expres-
sion levels in normal cells, which could be done by comparing
band strengths on Northern (RNA) blots. However, in these
methods one must control for the number of tumor/reactive
cells being the same as the number of normal cells. Analysis is
much easier if done by real-time PCR, in which the ratio of the
expression level of the gene of interest is compared with the
expression level of a constitutively expressed housekeeping
gene such as β-actin, 18S rRNA, cyclophilin, glyceraldehyde-
3-phosphate dehydrogenase, and β,-microglobulin. This ratio is
calculated in both the normal and neoplastic/reactive cells,
and then the ratios are compared to see if there is relative up-
or downregulation of the gene of interest.

References

351–355.

beta-globin genomic sequences and restriction site analysis for

3. Mullis K, Faloona F, Scharf S, et al. Specific enzymatic amplifi-
cation of DNA in vitro: the polymerase chain reaction. \textit{Cold

amplification of DNA with a thermostable DNA poly-

5. Saboor SA, Johnson NM, McFadden J. Detection of mycobac-
terial DNA in sarcoidosis and tuberculosis with polymerase

6. Myerson D, Lingenfelter PA, Gleaves CA, et al. Diagnosis of
cytomegalovirus pneumonia by the polymerase chain reaction
with archived frozen lung tissue and bronchoalveolar lavage

monary aspergillosis using polymerase chain reaction-based

scripts for tumor necrosis factor-alpha, interleukin-2, and inter-
feron-gamma in human pulmonary allografts. \textit{J Heart Lung
Transplant}. 1995;14:512–518.

Th2 cytokines and allergen on normal and asthmatic bronchial

surfactant protein C gene associated with familial interstitial

11. Pan Q, Pao W, Ladanyi M. Rapid polymerase chain reac-
tion-based detection of epidermal growth factor receptor

12. Westra WH, Baas IO, Hruban RH, et al. K-ras oncogene activa-
tion in atypical alveolar hyperplasias of the human lung. \textit{Cancer

13. Pulte D, Li E, Crawford BK, et al. Sentinel lymph node map-
ing and molecular staging in nonsmall cell lung carcinoma.


15. Bremnes RM, Sirera R, Camps C. Circulating tumour-derived
DNA and RNA markers in blood: a tool for early detection,

16. Eisenstein BI. The polymerase chain reaction. A new method of

17. Wenham PR. DNA-based techniques in clinical biochemistry:
a beginner’s guide to theory and practice. \textit{Ann Clin Biochem}.

18. Remick DG, Kunkel SL, Holbrook EA, Hanson CA. Theory and
applications of the polymerase chain reaction. \textit{Am J Clin

19. Chakrabarti R, Schutt CE. The enhancement of PCR
amplification by low molecular-weight sulfones. \textit{Gene}.
2001;274:293–298.

20. Laksalamlai P, Pavlov AR, Slesarev AL, Robb FT. Stabiliza-
tion of \textit{Tag} DNA polymerase at high temperature by protein
folding pathways from a hyperthermophilic archaeon, \textit{Pyrococ-

duplex stability from the base sequence. \textit{Proc Natl Acad Sci U

22. SantaLucia J Jr, Allawi HT, Seneviratne PA. Improved nearest-
neighbor parameters for predicting DNA duplex stability. \textit{Bio-

23. Wallace RB, Shaffer J, Murphy RF, et al. Hybridization of
synthetic oligodeoxyribonucleotides to phi chi 174 DNA:
the effect of single base pair mismatch. \textit{Nucleic Acids Res}.
1979;6:3543–3557.

24. Chien A, Edgar DB, Trela JM. Deoxyribonucleic acid poly-
merase from the extreme thermophile \textit{Thermus aquaticus}. \textit{J

DNA polymerase from \textit{Pyrococcus sp} strain KOD1 and its appli-

polymerase thioredoxin binding domain enhances the proces-
sivity and fidelity of \textit{Tag} DNA polymerase. \textit{Nucleic Acids Res}.
2003;31:4702–4709.

27. Koehavong P, Thilly WG. Fidelity of DNA polymerases in DNA

28. Wilson IG. Inhibition and facilitation of nucleic acid amplifica-

29. Khan G, Kangro HO, Coates PJ, Heath RB. Inhibitory effects
of urine on the polymerase chain reaction for cytomegalovirus


10
Array Comparative Genomic Hybridization in Pathology

Reinhard Ullmann

Principle of Comparative Genomic Hybridization

Comparative genomic hybridization (CGH) is a molecular cytogenetic method for the detection and mapping of chromosomal gains and losses. It is based on the cohybridization of differentially labeled test and reference DNA onto metaphase spreads, which usually have been prepared from peripheral blood lymphocytes of a healthy donor. The signal intensity ratios of the two labels along the chromosomes then reflect DNA copy number changes in the test genome relative to the reference genome. Although CGH has tremendously contributed to our knowledge of chromosomal aberrations, its resolution, unfortunately, is limited to about 3–10 Mb. Resolution of CGH has significantly improved when samples were no longer hybridized to metaphase spreads but to DNA targets that have been arrayed on a glass substrate. This modification to the original technique has been named array CGH or matrix CGH, respectively. In theory, resolution of array CGH is only limited by the number and quality of DNA targets arrayed on the slide. The principle of array CGH is illustrated in Fig. 10.1.

Array CGH Platforms

Arrays Based on Clone Inserts

cDNA Arrays

The first genome-wide application of array CGH was based on cDNA arrays, with each spot representing one reversely transcribed mRNA. Apart from the fact that these arrays were readily available, the main advantage of cDNA arrays was that they facilitated a direct comparison of DNA copy number changes with gene expression data derived from the same tumor. However, cDNA arrays are extremely gene focused and thus only a small percentage of the genome is actually represented. DNA copy number changes concerning introns or intergenic sequences are missed. Another problem arises from paralogous genes or shared sequence motifs. Together with the low signal-to-noise ratio usually obtained with this kind of arrays, these disadvantages have limited the use of cDNAs for the detection of DNA copy number changes.

Large-Insert Clone Arrays

Array CGH platforms based on large-insert clones are typically made up of bacterial artificial chromosome (BAC) and P1 artificial chromosome (PAC) clones and, to a lesser extent, also cosmids (Fig. 10.2). The first large-insert clone sets for whole genome analysis provided a resolution of about 1 Mb. Subsequently, a more comprehensive clone set has been assembled, and in 2004, the first sub-megabase-resolution whole genome tiling path array CGH study has been published, which was based on an array comprising 32,433 large-insert clones covering the whole genome in an overlapping fashion (http://bacpac.chori.org/). Because of the overlap of clones, tiling path arrays can provide a theoretical resolution that is below the average insert size of a BAC of 150 kb. Figure 10.3 shows the array CGH analysis of a squamous cell lung carcinoma using a sub-megabase-resolution whole genome tiling path with more than 36,000 clones.

Despite their robustness and widespread application, BAC arrays have some limitations. The low copy number of BACs and PACs within the propagating bacteria requires special isolation methods that preserve the integrity of the clone insert while at the same time eliminate as much bacterial genomic DNA as possible. Most laboratories are not spotting this isolated DNA directly but instead amplify the material to generate a renewable stock of amplicons that can be printed several times (see below for discussion on amplification...
Fig. 10.1. Principle of array comparative genomic hybridization (CGH). Differentially labeled test and reference DNA (green and red spheres, respectively) are cohybridized onto an array of DNA spots printed on a glass slide. In case of a deletion in the test DNA, fewer test DNA will bind to the corresponding spots and the red label of the reference DNA will prevail; gains in the test genome can be identified by a dominance of the green label of the test DNA. Spots representing sequences with the same copy number in the test genome relative to the reference genome appear yellow. For BAC arrays, an excess of repetitive Cot DNA (blue spheres) has to be added to suppress otherwise unspecifically binding repetitive sequences.

Fig. 10.2. Bacterial artificial chromosome (BAC) array. Each spot on a BAC array represents the very specific part of the genome that is contained in the BAC. A whole genome tiling path BAC array comprises as many BAC clones as necessary to cover the whole genome in an overlapping manner (~32,400 for the human genome). The upper part of the image is based on a screenshot of the UCSC Human Genome Browser. The red lines illustrate the selection of overlapping clones from a comprehensive BAC library.
methods). A protocol for the high-throughput isolation and purification of BAC/PAC clone inserts can be downloaded from our website (http://www.molgen.mpg.de/~abt_rop/molecular_cytogenetics/Protocols.html). Whatever protocol is followed, the setup of a comprehensive BAC array platform remains time consuming and costly. Other shortcomings of BAC arrays are directly related to specific features of the respective genomic sequence. Otherwise unspecific binding repetitive sequences must be blocked using a considerable excess of (expensive) Cot-DNA. Low copy repeats, that is, stretches of DNA that are longer than 1 kb and have a sequence similarity of more than 90% to other locations in the genome, can lead to ambiguous results. This problem especially applies to tiling path arrays because low-resolution arrays usually avoid low copy repeats. Finally, with the coming of tiling path arrays, BAC arrays have met the limits of resolution, which are simply given by clone insert size.

**Repeat-Free and Nonredundant Sequence Arrays**

In the light of the problems connected to the presence of repetitive sequences in the genome, researches have set out to generate genomic arrays that are depleted for repetitive and redundant sequences. This depletion, for example, has been accomplished by means of selective amplification using sequence-specific primers. Mantipragada et al have used this approach to create arrays focusing on the DiGeorge region (22q11 deletion syndrome). Another array, based on sequence-specific polymerase chain reaction (PCR) products of 162 exons of five genes, has been generated to test a spectrum of inherited human disorders. However, workload and high costs associated with this approach, as well as the upcoming of commercial oligonucleotide arrays, have hampered the widespread use of this approach.

**Oligonucleotide Arrays Using Presynthesized Oligonucleotides**

In contrast to the oligonucleotide platforms described below, these arrays are either generated by printing prefabricated, commercially available sets of oligonucleotides on glass slides or by coupling oligonucleotides to beads that are assembled on the slide afterward. One example for the use of printed arrays is reported by Carvalho et al, who have used a set of 18,861 oligos to identify DNA copy number changes in several tumor cell lines. Presynthesized oligos coupled to beads are used for a platform developed by Illumina (http://www.illumina.com). The company provides several designs that are dedicated to either gene expression, linkage,
or DNA copy number analysis. Some of their array designs enable the simultaneous detection of DNA copy number changes and loss of heterozygosity (LOH). 17

Although prefabrication of oligonucleotides enables highly efficient synthesis, at the same time it also reduces flexibility in terms of sequences on the array. Custom design becomes considerably expensive and requires a minimal batch size to pay off.

Oligonucleotide Arrays Based on In Situ Synthesis

Meanwhile, there are numerous ways to synthesize an oligonucleotide directly on the slide. Despite this diversity, the common principle is shared and is already known from PCR primer synthesis: the growing oligonucleotide is alternatively exposed to As, Gs, Cs, and Ts, but only when the last oligonucleotide of the growing chain is activated by splitting off a protective group can a new nucleotide be attached. The main difference between the platforms comes from how this protective group is inactivated. Some companies use light, selectively distributed through fixed photolithographic masks (www.affymetrix.com) or micro-mirrors (http://www.nimblegen.com/http://www.febit.de). Others de-protect by means of a current-induced change of pH value (http://www.combimatrix.com/) or control synthesis by specifically addressing each spot separately with high-resolution printers (http://www.home.agilent.com). Usually, oligonucleotides on such arrays are designed to be both isothermal (i.e., they share the same melting temperature, Tm) and single copy sequences; this sometimes results in an uneven distribution of oligonucleotides, leading to considerable variability in terms of resolution across the genome. Nevertheless, given the current developments, it can be expected that oligonucleotide arrays will replace all other platforms in the near future. Shortcomings with respect to hybridization kinetics (see below) and coverage will be compensated by the incredible increase of features on the array. Oligonucleotide arrays with more than 500,000 features are readily available. Those platforms that are not dependent on fixed photolithographic masks especially can offer extreme flexibility, which is limited only by the setup fees charged by some companies. Single nucleotide polymorphism (SNP) arrays, consisting of short oligonucleotides in the range of 16–20 mers and originally dedicated to linkage analysis, have been successfully used for the simultaneous detection of LOH (loss of heterozygosity) and DNA copy number changes. 18

General Platform Considerations

Before setting up an array CGH facility, several decisions have to be made. The first one refers to the expected number of array CGH experiments. In many cases, it will be much cheaper to cooperate with other groups that already have established the technique or send the samples to a company which is offering a hybridization service. The next decision concerns the type of array that should be used. Certainly, this decision depends on the scientific problem that should be addressed with the analysis, but often the consequences of this decision are far ranging. Frequently, the choice implies the purchase of expensive machines only useful for arrays sold by the same company; this is especially true for the most expensive devices necessary for array CGH analysis, namely, the hybridization machine and the scanner. A hybridization machine is designed to provide controlled temperature and even circulation of the hybridization mix to promote hybridization efficiency (see below). Some of these machines also accomplish the posthybridization washing of slides. Important criteria when selecting a particular machine could be flexibility in terms of slide formats, handling, maintenance/follow-up costs, and, most important, performance in one’s own laboratory. High-quality scanners are essential for the errorless readout of hybridization results. Reliability, flexibility, and resolution, the latter especially in the light of the continuing minimization of feature sizes on the array, are important issues. Other arguments can be the availability of autoloading to support high-throughput analysis or the need for more than two color channels. Note that there are devices that are scanning from the back of the slide and need transparent substrates.

DNA Preparation and Hybridization

DNA Isolation and Quality

DNA quality has a great influence on the outcome of an array CGH experiment. In general, oligonucleotide platforms, especially when used for one-color experiments, are more sensitive to compromised DNA quality than are BAC arrays. One source of trouble can come from the DNA isolation procedure, for example, detergents that could hamper the subsequent labeling process and introduce noise in the data. Although such problems can be avoided quite easily by improving or changing the isolation protocol, the situation is much more difficult when it comes to retrospective studies using formalin-fixed, paraffin-embedded (FFPE) material. This kind of fixation very frequently results in damage of DNA. 19 To cope with such damage, several adaptations of the DNA isolation and labeling protocols are required. Modifications can include the prolongation of tissue digestion time, with adding fresh proteinase each day, or the switch from enzymatic to chemical labeling systems. For other applications, some labs have even tried to repair fixation-induced DNA strand breaks by protocols following the principle of nick translation assay, or employed sodium thiocyanate to revert DNA-protein cross-linking. Unfortunately, it is not always possible to reliably predict the performance of FFPE DNA in an array CGH experiment, but verifying the average fragment size of the single-stranded DNA by gel electrophoresis and
testing DNA performance in PCR reactions with differently sized amplicons can give a good estimate.

In the light of the increasing impact of molecular (cytogenetic) techniques on research and routine diagnostics, optimization of fixation protocols gains increasing importance. Short-term measurements include the correct buffering of formalin and regular control of the pH value, but also reducing fixation time to the minimum necessary to ensure proper histological evaluation, and storage needs, should be considered.

In the long run, it may be beneficial to overcome old traditions and switch to alternative fixation protocols. In this context, it is also worth noting that many institutes have started to establish tissue banks where, in addition to the usual formalin fixation, tissues are especially preserved for various applications.

**Microdissection and DNA Amplification**

In many instances, for example, when trying to avoid normal cell contamination in compact growing tumors, it is sufficient to scratch out regions of interest from the section manually using a needle. Unfortunately, this low-budget solution may not be applicable to other problems in pathology. For single cells or small cell clusters interspersed in the tissue, such as suspected tumor stem cells, micrometastasis, or preneoplasias, laser microdissection must be employed. Figure 10.4 demonstrates the microdissection of a bronchiolar columnar cell dysplasia.

The amount of DNA isolated from laser-microdissected samples usually is not sufficient for array CGH analysis. Thus, uniform amplification of the whole genome is required to make such samples amenable to array CGH. Fortunately, in the meantime, numerous protocols for whole genome amplification (WGA) exist and many companies are providing advanced WGA kits. WGA methods can be roughly divided into PCR-based and non-PCR-based approaches. Probably the best known PCR-based technique is degenerate oligonucleotide primer (DOP) PCR. The name of this technique refers to the specific primer, which contains a central cassette of six degenerated bases (at each of these six positions, every one of four bases can be present). In the first cycles of PCR with very low annealing temperature, DOP promotes the priming of DNA synthesis from multiple evenly dispersed sites within the genome, resulting in amplicons that are flanked by the primer sequence, which then can serve as the annealing site in the subsequent amplification step using high stringency conditions (Fig. 10.5a). The initial priming is supposed to occur at approximately $10^6$ sites in

---

**Fig. 10.4.** Laser microdissection. The selective isolation of bronchiolar columnar cell dysplasia is shown (a) before and (b) after laser microdissection.

**Fig. 10.5.** Whole genome amplification. Two examples of PCR-based whole genome amplification techniques are shown: (a) degenerate oligonucleotides primer (DOP)-PCR; (b) linker-mediated PCR. Primer and oligo-complexes, respectively, are depicted in red; template DNA is drawn in blue. See text for explanation.
the human genome, resulting in a genomic representation sufficient for chromosomal CGH. However, with the coming of high-resolution array-based CGH techniques, many laboratories have shifted to alternative methods, which promise better coverage of the genomic complexity. Among the PCR-based methods, this capability is mainly ascribed to ligation-mediated PCR techniques, which have already been successfully used for the analysis of single cells by CGH. The principle of ligation-mediated PCR is based on either enzymatic, hydrodynamic, or chemical fragmentation (http://www.rubicongenomics.com) of DNA followed by the ligation of an adapter complex, which serves as a universal priming site for uniform amplification (Fig. 10.5b). Certainly, PCR-based approaches imply the risk of amplification bias and experimental errors, and therefore some researchers prefer non-PCR-based approaches such as strand displacement amplification using such enzymes as phi29 DNA polymerase. A more comprehensive review and discussion about the pros and cons of diverse WGA is given by Hughes et al.

Unfortunately, WGA cannot distinguish DNA that should be amplified from incidentally introduced DNA. Therefore, precautions have to be taken to avoid any contamination with DNAs from other sources. This caution starts with the cleaning of the carving board before sectioning the resected tumor, regular cleaning or replacement of microtome blades, exchange of water used to stretch paraffin sections, and so forth. Especially when dealing with FFPE, one has to be aware that nondegraded, contaminating DNA is preferentially amplified. Amplicons generated in previous PCR reactions, distributed through aerosols, are an extremely good template. Therefore, strict compliance to the general rules of PCR setup is mandatory to produce reliable results.

Reference DNA

In a typical array CGH experiment, test DNA is compared to DNA of a healthy donor or a DNA pool of healthy individuals. Many laboratories hybridize in a sex-matched manner, that is, test and reference DNA have the same sex, while others prefer sex-mismatched hybridizations, where the ratio changes at the sex chromosomes can serve as an internal control of hybridization quality, but eventually render the interpretation of DNA copy number changes at the sex chromosomes complicated. Using DNA from the same individual as the test DNA, but isolated from normal-looking tissue far away from the tumor, may be advantageous in terms of hybridization quality, but it contains the risk of missing DNA copy number changes present in both DNAs; for example, constitutional changes that may predispose to tumor formation or chromosomal imbalances also present in the tumor microenvironment.

For chromosomal CGH, it has already been observed that matching the quality of test and reference DNA can significantly improve the results of an experiment. Thus, it has been realized that combining amplified test DNA with nonamplified reference DNA can introduce hybridization artifacts, which can be avoided when a PCR-amplified reference DNA is used. A related observation has been made in our laboratory when working with FFPE DNA, where the use of reference DNA also isolated from FFPE tissue has significantly improved results. In one-color array CGH experiments using short oligonucleotides arrays, the reference DNA is not tested within the same experiment, but data are compared to a reference dataset, either generated in the same laboratory by analyzing a cohort of several normal individuals or provided by the array-selling company. This in silico comparison can hardly compensate for intraexperimental variation and thus is more prone to noise.

DNA Labeling

Problems in labeling DNA can result in artifacts and failure of the array CGH experiment. Despite a great variety of direct and indirect labeling protocols, at the moment the prevailing method for DNA labeling for noncommercial slides is direct incorporation of fluorochromes into the DNA by means of a random priming assay, which, as a nice side effect, also results in a net gain of DNA. In some instances, however, for example when dealing with highly degraded DNA isolated from FFPE tissue, chemical labeling may be superior as this method does not depend on long DNA fragments and DNA synthesis. As it is true for DNA quality, in our experience it is essential that test and reference DNA do also match in the way they have been labeled, so that a possible bias introduced by problems in labeling of specific sequences can be compensated. Commercial arrays frequently require different labeling protocols and the use of specialized kits, usually sold by the same company.

Hybridization

Disadvantageous hybridization kinetics is the greatest problem of array CGH based on oligonucleotide arrays. Even long oligonucleotides represent an extreme low complex target: a diversity of 3 billion bases in the hybridization mix versus ~60 bases represented by each specific spot (Fig. 10.6). Circulation of the hybridization mix can only partially compensate for this discrepancy. Representational oligonucleotide microarray analysis (ROMA) is one technical approach to address the problem of complexity. ROMA is based on a linker-mediated PCR (see Fig. 10.5b) using the restriction enzyme BglII. The following PCR is optimized to result in the preferential amplification of fragments smaller than 1.2 kb and thus to lead in a reduction of complexity down to about 2.5%. Given the known recognition sequence of the restriction enzyme and the maximum spacing between two sites, the authors expected the amplification of about 200,000 sites across the genome and designed a corresponding array specific for the expected sites. It is clear...
that this elegant approach is technically and computationally demanding and requires optimal settings and standardization to avoid the introduction of artifacts.

Data Analysis

A typical array CGH experiment yields several thousands of data points, which have to be displayed in a comprehensive and illustrative way. Hardly ever are raw data presented in a manuscript. Instead, data have already passed several steps of manipulation. For many readers, the computational analysis appears like a black box, and the brief descriptions of the procedures in many manuscripts are not always able to shed light into this “box.” However, availability and understanding of information about data analysis are not only essential for appreciating the quality of a microarray experiment and the functional (real) resolution of an experiment but also to judge whether a comparison of two different studies is feasible. It is beyond the scope of this chapter to provide in-depth insights into data analysis. Instead, the following is meant as a rough overview that should enable a basic understanding of the data analysis workflow and highlight key points of data interpretation of an array CGH experiment. For a more comprehensive review on computational aspects of array CGH, see Chari et al. 33

Image Processing

The wet lab part of a typical two-color array CGH experiment ends with putting the hybridized array into the scanner, where the signal intensities of the two different fluorochromes are recorded as two gray-scale images, either simultaneously or one by one. Usually the scanner output is two 16-bit TIFF (tagged image file format) images, which provide $2^{16}$ (65,536) different gray scales. These images are then imported into specialized image processing software, frequently sold as a package together with the laser scanner. There, DNA spots are defined by superimposing a grid that reflects the architecture of the array in terms of rows and columns and links each spot with information on clone/sequence identity and chromosomal location.

Background Subtraction

Global background subtraction is based on the averaged signal intensities of all pixels outside those areas that have been identified as DNA spots. This method does not take into account the possible uneven distribution of background signal intensities across the array. Therefore, most people favor local subtraction methods based on background intensity values determined in the vicinity of each spot to cope with spatial bias. In our laboratory, we do not subtract background at all.

At this stage, data already can be exported to array CGH-specific analysis tools.

Normalization

The absolute signal intensities measured in each channel can be subject to systematic, spatial, or intensity-dependent bias, which can influence array CGH results 34 and has to be removed before calculating the signal intensity ratios. Systematic bias of signal intensities can arise from differences in input of test and reference DNA, fluorochromes and labeling efficiency, laser settings (laser power as well as amplification through photomultiplier tubes), etc. The simplest way to eliminate systematic bias is to normalize by equalizing the median intensities of the two channels.

However, calculating the global median intensity does not take into account spatial effects, that is, areas that appear to show trends of higher and lower intensity, respectively. For this reason, many researchers apply this normalization method for small subgrids separately. Meanwhile, more sophisticated programs are also available that can detect and remove spatial bias. 35 Sometimes evaluation of intensity scatter-plots reveals an intensity-dependent trend; for example, the ratios of low-intensity spots always show the same tendency. Normalization algorithms such as LOWESS or subgrid LOWESS, which are based on regression models, can be employed to tackle this problem.

Whatever algorithm is applied to normalize the data, it relies on the assumption that for many, if not most, of the data points of DNA copy number are the same in test and reference DNA. This assumption may be a problem in tumors with complex aberrations or customized arrays focusing on very small regions of interest. Therefore, for some projects it
may be advisable to skip normalization completely to avoid loss of dynamic range.

Identifying an Aberration

Despite the availability of several sophisticated tools dedicated to the analysis and visualization of array CGH data, defining the presence or absence of an aberration remains challenging. The simplest way to identify an aberration would be to determine fixed thresholds, and every clone/oligo whose signal ratio exceeds these thresholds would be considered as an indicator of aberrant copy number of the respective genomic region. Alternatively to the use of fixed thresholds, which does not take into account variable data quality, thresholds can be defined based on standard deviation. However, given the considerable number of outliers that can occur in one experiment, false positives would pose a problem. Smoothing data by means of sliding windows (moving average) has been frequently applied to be less sensitive to outliers (Fig. 10.7). Although this method dramatically reduces the false-positive rate, it goes along with loss of resolution, which depends on the size of the window, i.e., the number of clones/oligos that are averaged. To avoid this disadvantage, advanced algorithms such as hidden Markov models (HMM), BioHMM, circular binary segmentation (CBS), and wavelet-based approaches have been designed to determine copy number changes in an objective and automatic fashion.

Especially in tumor cytogenetics, biology introduces another layer of complexity that renders data interpretation difficult. Just imagine the following scenario: a tetraploid cell loses one chromosome. In a tetraploid cell with four copies of each chromosome, this means a relative loss of 25%. There may also be some polyclonality in the tumor specimen, with some cells having this aberration and others not. Some cells may be even euploid, especially those normal cells that are contaminating the tumor sample. To make a long story short, DNA copy number changes are not necessarily changing as integers, and it is up to the scientist to determine those thresholds that reflect the best compromise between sensitivity and specificity. This decision can dramatically influence the outcome of an array CGH study and can hamper the comparability of different array CGH studies. In this context, it is of great interest that authors of array CGH papers are increasingly following the so-called MIAME (minimal information about a microarray experiment) criteria, which also include the deposition of primary data to a public database such as Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo; GEO) or Array Express (http://www.ebi.ac.uk/arrayexpress/). Every scientist interested in a specific published dataset can download the respective files and perform an in silico reanalysis of the data sets with different analysis parameters or different focus. Meanwhile, a number of array CGH analysis and visualization tools are available for free (for review) that can assist in the exploitation of this valuable data source.

---

**Fig. 10.7.** Data smoothening by sliding window. In this schematic illustration, a sliding window of three clones/oligonucleotides is shown. Instead of the original ratio (row a; upper part of the triangles), the averaged ratio of three adjacent clones/oligonucleotides is plotted (row b; tip of each triangle). The differentially colored triangles symbolize the moving of the windows (blue triangle is the latest window). Note that the size of the window determines the quality of data smoothening and also the functional resolution of the array.
However, not only technical issues complicate the identification of causative aberrations. Because of the high resolution of array CGH, a number of DNA copy number variants have been found to occur in the normal population (a comprehensive database can be found under the following link: http://projects.tcag.ca/variation/). Although these variants can be several hundred kilobases in size, many of them are supposed to be phenotypically neutral, but they also may pre-dispose to disease.

**Conclusion**

The accumulation of chromosomal imbalances is a typical feature of tumor development. The specific composition of chromosomal changes creates a unique genomic setting that influences the biological behavior of the tumor. Currently, array CGH using high-resolution platforms is the technique most often applied to detect DNA copy number changes. Because of the introduction of commercial platforms, the technique is no longer restricted to highly specialized laboratories, and therefore molecular cytogenetics profiling by means of array CGH will gain increasing importance in routine pathology also. Array CGH data can be employed to identify chromosomal signatures useful for differential diagnosis and therapeutic decisions. Some aberrations may serve as discriminators in differential diagnosis, especially for tumors that are hardly distinguishable by classical pathological approaches, while the finding of other chromosomal changes creates a unique genomic setting that influences the biological behavior of the tumor. Some aberrations may serve as discriminators in differential diagnosis, especially for tumors that are hardly distinguishable by classical pathological approaches, while the finding of other chromosomal changes may be useful to anticipate the risk of metastasis or the likelihood of recurrence. The detection of genes frequently amplified in tumors will help to identify new therapy targets, and customized therapy arrays will prove useful to predict response to such therapies.

**References**


Introduction

The most common molecular alteration observed in human cancers, loss of heterozygosity (LOH), is a significant mechanism by which critical genes involved in growth regulation and homeostasis become inactivated, or silenced, during disease evolution. This chapter provides a review of LOH and its implications in various cancers as well as a review of LOH in nonmalignant diseases. Only 0.08% of those base pairs within the entire human genome (3 billion base pairs) vary between any two humans, and only 0.02% of those variations actually result in an expressed protein with a different amino acid as a result of the change. Even more remarkable, 90% of those variations are changes that are common in the population and lead to normal variation in traits among individuals; eye color, for example.

In a normal individual’s DNA, every genetic locus is composed of two alleles, one inherited from each parent. Genetic loci can be homozygous, meaning that both copies of the gene are exactly the same, or heterozygous, meaning that the two copies of the chromosome are different. The most common differences between two chromosomes are single nucleotide polymorphisms (SNPs), where the DNA sequence differs by one nucleotide. SNPs can be found on average every 1,200 base pairs. Most (90%) of this variance is normal within the population, but approximately 10% of the heterozygous sites in any individual’s DNA are the result of deleterious genetic variations that have the potential to cause disease.

Loss of one allele can occur at any genetic locus, homozygous, or heterozygous (Fig. 11.1). Allelic loss at a heterozygous site, called loss of heterozygosity (LOH), leaves the cell with a single version of a gene. If the remaining copy is aberrant, LOH has the potential to disrupt cellular homeostasis. LOH can occur when a whole or part of a chromosome is lost (Fig. 11.1b) or if part of a chromosome is replaced by translocation (Fig. 11.1c). During mitosis, recombination between chromosomes can also cause localized LOH (Fig. 11.1d).

The importance of LOH in cancer was recognized by Knudsen in what has become a classic model of tumor suppressor gene inactivation. The two-hit model recognized that one copy of an important regulatory or homeostatic gene could become defective by mutation or gene silencing (e.g., via promoter hypermethylation or small intragenic deletion) and the remaining copy could be lost through an LOH event. In hereditary cancers, individuals inherit one defective copy and thus are predisposed to disease because LOH can occur subsequently by several mechanisms (Fig. 11.1). The two-hit model is now a well-documented mechanism by which tumor suppressor genes are lost during the formation and progression of many types of cancer, including lung, breast, bladder, esophagus, and liver.

Detecting LOH

Methods to detect LOH rely on the ability to assess copy number and determine that one copy is missing or reduced compared to the other copy. Technical methods such as comparative genomic hybridization (CGH) allow detection of regions of chromosomal gain or loss but generally do not have the resolution to detect LOH at specific gene loci. Classical methods for detection of LOH involved using heterozygous restriction fragment length polymorphisms (RFLP) and Southern blotting. Two drawbacks to this classical approach are low throughput and the relatively sparse coverage of the genome available for study with RFLP. Much like geopositioning satellite (GPS) technology, which has brought about more accurate pinpointing of our physical position on the globe, higher-density markers that leverage simple sequence length polymorphisms (SSLPs, or microsatellites) and single nucleotide polymorphisms (SNPs) have generated more comprehensive polymerase chain reaction (PCR)-based “allelotyping” studies of LOH. LOH can now be surveyed in larger sample sets, and studies can identify clusters or linkages of multiple loci altered by LOH in specific disease...
The availability of high-density polymorphic probes, gene arrays, and analysis algorithms is expanding the scope and increasing the throughput of genome-wide LOH assessment. Minimally deleted regions (MDR) have been identified for many different cancers, pointing researchers to the most likely locations for tumor suppressor genes whose inactivation contributes to the progression of primary tumor development or metastases.

Although the classic LOH detection methods established the importance of LOH in a wide variety of cancers, SNP-based screening technologies are becoming the dominant technology. SNPs are high-density markers. When multiple adjacent SNPs show a consistent genetic alteration, researchers’ confidence in concluding that an LOH event has occurred is higher. Affymetrix, in particular, has invested in a series of SNP gene array chips at various densities, providing the research community with a commercially available resource that allows high-throughput testing of target populations and a common database on which the results of different groups can be compared.

Regardless of the methods used, studies of LOH typically provide several key measurements helpful in understanding the potential role of LOH in various cancers. In addition to defining hotspots for allelic loss, or minimally deleted regions, likely to contain key genes whose alteration plays a role in the development and/or progression of the cancer, LOH analysis defines the extent of allelic loss in a given sample. Fractional allelic loss (FAL) is defined as the number of LOH events in a sample divided by the total number of informative heterozygous markers in the corresponding normal DNA. Breakpoints can also be determined, and are defined as the junction between a specific marker displaying LOH and an adjacent marker that retains heterozygosity (LOH:HET).

Conducting LOH analysis requires gaining access to tumor cells. In the case of surgical resection or postmortem analysis of specimens, cells can be obtained directly from tissue via macro- or microdissection. Cells have also been harvested successfully from biopsies, bronchial brushings, and urine. The purity of the harvested cell population must be assessed because contaminating stroma, blood vessels, lymphocytes, and other normal cells are inevitably present in clinical specimens. LOH sensitivity to nontumor DNA depends on the method and type of probe.

Classical LOH detection methods that rely on copy number analyses (e.g., CGH, RFLP) lose sensitivity if the proportion of DNA contributed by tumor cells is not ≥80% of the total sample DNA. Flow cytometry has been used to increase the percentage of tumor cells in the sample DNA. Alternatively, LOH has been detected in tissues with fluorescent in situ hybridization (FISH) in clinical situations where pure samples cannot be easily obtained and the loci of LOH are well known. However, SNP arrays are significantly less sensitive to nontumor DNA contamination because they do not rely on copy number analysis. Studies using genome-wide SNP probes have revealed new LOH loci that would have gone undetected by copy number analysis techniques. Another caveat of using copy number analysis techniques to detect LOH is the phenomenon by which one allele becomes genetically amplified. When one allele is present in significantly more copies than the other, copy number analysis data can sometimes indicate LOH when in reality the nonamplified allele remains intact.

Yet another source of false negatives using classical LOH detection methods with <100% pure samples is when both copies of a locus are missing in tumor cells and the heterozygous signal arises solely from contaminating nontumor DNA.
SNP-based technologies have expanded the scope and utility of LOH detection. SNPs have helped pinpoint minimally deleted regions and focus researchers’ efforts to identify new tumor suppressor genes. SNPs have helped dissect the temporal order of mutations that affect prognosis and the potential for metastatic spread of tumors. The high throughput available with SNP-based LOH detection will likely continue to spur new applications and new understanding of the genetic basis of tumor initiation, disruption of cellular homeostasis, and metastases.

Environmental Factors Influencing LOH

Mechanisms by which LOH occurs are varied (see Fig. 11.1), and all have been associated with at least one type of cancer. Environmental factors can contribute to LOH and have been linked to certain cancers. Carcinogenic compounds in the air we breathe, the food we eat, and even some prescription drugs can induce mutation either directly or through metabolic by-products. Pathogens have also been shown to play a contributing role in cancer progression.

LOH is a downstream event that happens some time after exposure to a carcinogen. The initial effect is usually a small change at the single nucleotide level. Carcinogens cause mutations through highly reactive intermediate compounds produced by the metabolic or elimination processes of the body – either in the digestive tract or by immune system surveillance. These compounds can chemically react with DNA, modifying its structure. These so-called DNA adducts occur in multiple forms, depending on the chemical and molecular structure of the carcinogen. Some DNA adducts cause DNA polymerases to insert an incorrect nucleotide during DNA replication. Other DNA adducts link DNA strands to proteins or other DNA molecules, causing a physical barrier to replication. Other carcinogens cause DNA strand breaks.

The body has mechanisms that monitor and repair DNA or initiate the death of cells that harbor deleterious mutations. However, chronic or high-level exposures to a carcinogen can overwhelm the body’s intrinsic repair mechanisms. Some individuals inherit DNA repair-associated alleles that make them more or less susceptible to disease from many sources of mutation, including environmental carcinogens. In most cases, disease emerges months or years after an exposure event. Therefore, establishing linkages between environmental factors and cancers can take many years of research with large study populations.

Tobacco

Exposures to tobacco smoke from cigarette use and tobacco in smokeless products have both been linked to increased risk of cancer. Lung cancer is the most well known association, but tobacco smoking is a major cause of bladder cancer as well. The evidence linking exposure to tobacco smoke and lung cancer is overwhelming, with smoking named as the greatest etiological factor contributing to an individual’s risk of developing lung cancer. The mechanisms by which tobacco carcinogens contribute to the development and progression of lung cancer are multivariate and are only partially understood. Tobacco smoke produces reactive chemical species, including benzo[a]pyrenes and polycyclic aromatic hydrocarbons (PAH). Smokeless tobacco is a major source of nitrosamines and is linked to increased risk of oral, esophageal, and pancreatic cancers.

When adolescents smoke, it is believed that normal lung epithelium is “preconditioned” by exposure to tobacco carcinogens at a time of critical lung development, resulting in somatic mutations in an entire “field” of epithelium rather than in a single cell or cluster of cells. These genetically altered cells are replicated during normal adolescent lung growth, leading to the presence of a large number of cells that are “primed” for subsequent events (such as allelic loss), thus rendering the individual highly susceptible to the development of lung cancer. Smoking during adolescence may increase the risk that if lung cancer develops it will develop as many clonal cancers that progress concomitantly and aggressively, a concept first described as “field cancerization.”

Asbestos

Asbestos is a well-known pulmonary carcinogen, linked in particular to the formation of mesotheliomas, lung tumors that usually develop in individuals with a history of asbestos exposure, typically with a long period of latency between exposure and clinical presentation. The specific mechanisms by which asbestos contributes to LOH are not known. Mutagenicity studies carried out using a lymphocyte model demonstrated that the LOH rate upon exposure to asbestos fibers was greater than the spontaneous LOH rate of the assay, suggesting a direct connection between the fibers and LOH. In patients exposed to asbestos, allelic deletion of the FHIT gene on chromosome 3p14.1 and LOH on regions of chromosome 6p have been reported, showing a correlation between asbestos exposure and LOH in human disease.

Carcinogens in Food

Food is another source of carcinogens that lead to LOH and disease. Aflatoxin, a metabolite of certain fungi, can be introduced into food sources by grains such as corn and livestock feed. Aflatoxins often occur in crops in the field before harvest. Postharvest contamination can occur if crop drying is delayed and during storage of the crop if water is allowed to exceed critical values for mold growth. Insect or rodent infestations facilitate mold invasion of some stored commodities.

Whether sodium nitrate and sodium nitrate, used to preserve meats, are carcinogens has been a matter of debate.
Chemically, the possibility exists for these compounds to be converted into nitrosoamines, which have been demonstrated as participating in the carcinogenesis resulting from tobacco smoke exposure. However, regulations on maximal levels imposed by the FDA and USDA and changes in industry processes have lowered the risk of cancer from dietary intake of preserved meats. Despite the lower risk of exposure from preserved meats, nitrosamines can be generated from the digestion of meats that have been charred during cooking and are ingested with smokeless tobacco products, as noted above.

Alcohol

Alcoholic beverage consumption was first added as a known human carcinogen to the U.S. Department of Health and Human Services biannual Report on Carcinogens in 2000 because it was demonstrated to increase the risk of upper gastrointestinal tract cancers (e.g., mouth, pharynx, larynx, and esophagus) and because evidence of a linkage to pancreatic cancer is accumulating. The formation of acetaldehyde seems to be the most important mechanism by which alcoholic beverage consumption causes human cancer. Although the N\textsuperscript{2}-ethyl-2\textsuperscript{¢}-deoxyguanosine DNA adduct is increased in white blood cells obtained from human alcohol abusers, a more likely cause of the mutagenic effects of alcohol is a different DNA adduct, 1,N\textsuperscript{2}-propano-2\textsuperscript{¢}-deoxyguanosine (PdG), that can form when acetaldehyde is in the presence of histones and other basic molecules. PdG has been shown to be responsible for the genotoxic and mutagenic effects of crotonaldehyde through the formation of secondary lesions, including DNA-protein cross-links and DNA interstrand cross-links. Individuals who smoke tobacco and are heavy consumers of alcoholic beverages are at even greater risk of upper gastrointestinal cancers than individuals who are exposed to either substance alone.

Pathogens

Inflammation associated with chronic infection can be a contributing risk factor for carcinogenesis. Examples include infection of the gastric wall with Helicobacter pylori as a predisposing factor for gastric cancer; chronic infections of hepatitis virus leading to hepatocellular carcinoma (HCC); and the linkage of lung cancer with chronic Chlamydia pneumoniae and Mycobacterium tuberculosis (M-TB) infections. Studies indicate that the mechanism of DNA damage in relation to infections is indirect and comes from the host’s activated inflammatory cells sent to fight the infection. Cumulative evidence suggests that free radicals, such as NO, produced by activated inflammatory cells can contribute to cancer, as they are known to be able to cause direct damage to DNA.\textsuperscript{47}

Patterns of LOH in Cancers

LOH has been detected on all chromosomes. Table 11.1 summarizes some associations between LOH and cancers, and Table 11.2 lists individual genes that are targets for allele loss in cancers. The following sections outline in more detail selected research on LOH in disease, with particular focus on known or putative tumor suppressor genes and their role in disease initiation or progression.

LOH at 1p

LOH at 1p has been associated several cancers, including lung, colorectal, and gastrointestinal stromal tumors and hepatocellular carcinoma. LOH at 1p36 was found to
be a prognostic indicator for decreased overall survival in hepatocellular carcinoma patients.\textsuperscript{64} 1p36 encompasses several candidate tumor suppressor genes, including p73 and tumor necrosis factor receptor-2 (TNFR2).\textsuperscript{48} P73 is a member of the p53 family and is capable of mimicking some of the effector functions of p53, including induction of permanent growth arrest and promotion of apoptosis.\textsuperscript{55} The TNFR2 binds to and mediates signals from lymphotoxin-\(\alpha\) (LT\(\alpha\)), lymphotoxin-\(\beta\) (LT\(\beta\)), and TNF, three cytokines associated with receptor-mediated induction of cell death.\textsuperscript{56}

In oligodendroglial tumors, LOH at 1p is associated with better prognosis.\textsuperscript{57} Further characterization of these tumors revealed that in cells with LOH at 1p, the expression of GLUT-1, a component of the glucose transport machinery, was reduced compared to cells that retained heterozygosity at 1p. Stockhammer et al\textsuperscript{57} speculated that the disappearance of GLUT-1 in cells might affect the tumor’s sensitivity to chemotherapy drugs. Further study is needed to elucidate the details, but this research highlights how LOH detection can inform treatment planning.

LOH at 3p

The best described and best documented gene targeted for LOH in nearly all lung cancers is the fragile histidine triad (FHIT), on 3p14. LOH at 3p14 is strongly correlated with early age.\textsuperscript{37} In two studies, LOH at a locus near the hMLH1 gene on chromosome 3p was correlated strongly with early age of smoking initiation and with the level of DNA adducts.\textsuperscript{35,64}

The FHIT gene encodes a small protein with diadenosine triphosphate hydrolase activity, but its tumor suppressor activity is presumed to be independent from this enzymatic activity.\textsuperscript{62} The microregion of the FHIT gene is frequently targeted for LOH in tumors; however, this association alone does not confirm tumor suppressor activity. Evidence for FHIT as a tumor suppressor gene includes the observations that FHIT gene-deficient mice are more susceptible to carcinogen-induced tumor formation,\textsuperscript{55} and that expression of FHIT suppresses the growth of cancer cells through promotion of apoptosis and growth inhibition.\textsuperscript{86}

FHIT genetic and protein losses have been shown in tumors from other cancers in addition to lung: esophageal,\textsuperscript{67-69} breast,\textsuperscript{70} and kidney.\textsuperscript{71,72} In addition, other tumor suppressor gene candidates targeted by LOH are located in the same region (3p12–3p22) as the FHIT gene. These other genes include transforming growth factor receptor beta-2 (TGF\(\beta\)R2), MLH/HNPPCC2, deleted in lung cancer-1 (DLC1), RASSF1A, retinoic acid receptor-beta (RAR\(\beta\)), and BRCA1-associated protein-1 (BAP1).\textsuperscript{20,48,58,59,73,74} LOH at 3p25, which includes the von Hippel-Lindau (VHL) locus, has also been reported as a frequent event in several cancers.\textsuperscript{20,75}

LOH at 4q

LOH at 4q has been associated with breast,\textsuperscript{76} cervical,\textsuperscript{77-79} mouth,\textsuperscript{80} throat,\textsuperscript{81-83} bladder,\textsuperscript{84-86} colon,\textsuperscript{87,88} lung,\textsuperscript{48,89,90} and liver cancers\textsuperscript{91-94} as well as acute lymphoblastic leukemia.\textsuperscript{95} Jiang et al\textsuperscript{95} further analyzed 4q LOH patterns in colorectal carcinoma tumors from 83 patients to evaluate relationships between LOH patterns and clinicopathological features. LOH at 4q13.2 was significantly higher in tumors larger than 5 cm in diameter. LOH in 4q21.23 was significantly associated with metastasis. Suspected tumor suppressor genes are located in each region, but cause-and-effect relationships between specific genes and tumor diameter or metastasis have not yet been established.

LOH at 5q

SCLC is characterized by a high frequency of LOH at 5q32-ter, and although no candidate tumor suppressor genes have been identified, the SPARC (secreted protein acidic and rich in cysteine) gene maps to 5q32 and has been associated with LOH in idiopathic pulmonary fibrosis (IPF).\textsuperscript{96} SPARC, a protein involved in the regulation of cell adhesion and growth, is methylated aberrantly in lung cancers.\textsuperscript{97} Although nonsmall cell lung cancer (NSCLC) has not been associated with LOH at 5q32, there is often LOH at 5q21.3–31, a region that contains several key tumor suppressor genes including MCC (mutated in colorectal cancer), APC (adenomatous polyposis coli), and IRF1.\textsuperscript{48} LOH at 5q21 has been documented in preneoplastic cells, suggesting it may be an early genetic change, and it is more prevalent in squamous cell carcinomas compared to adenocarcinomas.\textsuperscript{98,99} However, as in other LOH events, studies have failed to provide strong correlation between these events and patient prognosis. In lung cancer, it appears that LOH at the APC locus is relatively frequent but that the mechanism by which the remaining allele is silenced is usually promoter methylation, rather than mutation.\textsuperscript{100}

LOH at 8p

LOH on chromosome 8p21–23 is a frequent event in many cancers and is often associated with more aggressive disease.\textsuperscript{101,102} Intense research has focused on 8p21 to identify tumor suppressor genes and look for connections between LOH and clinical characteristics. Shi et al\textsuperscript{103} investigated DBC2 (deleted in breast cancer 2) expression in bladder cancer. In these tumors, LOH was not frequent, but they discovered that promoter methylation caused downregulation of DBC2 expression, mimicking the effect of LOH at DBC2 observed in other cancers.
In head and neck squamous cell cancer (SCC), 8p21–22 was the most frequent site of LOH. Ye et al. investigated expression of another candidate tumor suppressor gene identified in the region, MTUS1 (mitochondrial tumor suppressor gene 1), and found downregulation of transcription and mutations in exon-coding sequences. Coon et al. found that LOH in a region of 8p that included 8p21 was strongly associated with shorter survival in head and neck SCC patients. Ye et al. discovered that LOH was observed mainly in invasive and metastatic SCC.

LOH at 8p21 is a frequent and early occurrence in NSCLC, believed to occur after LOH events at 3p and 9p, and hepatocellular carcinoma. Despite consistent observations of LOH in this region in lung cancer, specific tumor suppressor genes have not been mapped to 8p21–23 and implicated in the development and progression of lung cancer. LOH at 8p21–23 is also a frequent event in hepatocellular carcinoma (HCC), and a gene encoding a growth inhibitory protein, HCRP1 (hepatocellular carcinoma related protein-1) has been mapped to that area and implicated in HCC.

LOH at 9p

Chromosome 9p is frequently altered in NSCLC, not only by LOH but also by homozygous deletion and gene silencing via promoter hypermethylation. The redundancy in mechanisms aimed at silencing genes in this region and the prevalence of these alterations in smoke-exposed preneoplastic epithelium point to the importance of 9p in the development of lung cancer. The gene most frequently identified in LOH studies of 9p (9p21) in NSCLC is P16 INK4A, a gene that encodes a cell-cycle protein, HCRP1 (hepatocellular carcinoma related protein-1) has been mapped to that area and implicated in HCC.

LOH at 10q

LOH at 10q22–23 is a frequent observation in SCLC. Although no tumor suppressor genes in this region have been definitively associated with lung cancer, multiple studies have identified LOH targeted to the PTEN/MMAC locus at 10q23. The encoded PTEN protein is a lipid phosphatase that negatively regulates the phosphatidylinositol 3-kinase/AKT (PI3K/AKT) pathway. Loss of PTEN function results in reduced apoptosis and stimulation of cellular proliferation and migration.

LOH at 13q

Chromosome 13q12–14 is a prevalent hotspot for LOH in many cancers. The retinoblastoma (RB1) gene is a well-characterized tumor suppressor gene, at 13q12, the product of which is a key regulator of entry into S phase of the cell cycle. The critical pathway regulated by RB, p16INK4A, cyclin-D1, and cyclin-dependent kinases (CDKs) is disturbed, usually via multiple components, in nearly every case of lung cancer, thus rendering cells insensitive to the signaling involved in regulating mitosis. LOH on 13q12.1–13.1 can be identified in cells obtained from bronchial washing specimens from SCLC and NSCLC patients and is proposed as one of a set of markers that could be used for early detection of lung cancer. Another gene targeted to 13q12–14 is BRCA2, a gene associated with cancer susceptibility and with possible linkages to p53 in the context of DNA damage repair.

LOH at 17p

The TP53 gene at 17p13 is the most frequently altered tumor suppressor gene in human cancers. TP53 is targeted by mutation, methylation, and homozygous deletion, in addition to LOH. The observations of redundant mechanisms of TP53 inactivation and frequency of LOH at TP53 gene point to its critical role in the development of malignant disease. TP53 modulates a broad network of cellular responses, including cell-cycle arrest, apoptosis, DNA repair, cellular senescence, and inflammation, and therefore plays a central role in homeostasis. TP53 mutations can be found in precancerous lesions, but to date studies of LOH at 17p indicate that allelic loss of TP53 occurs during disease progression, after LOH events on chromosome 3p.

LOH at 19p

LOH at 19p13.3 is a very frequent event in NSCLC but is not targeted in SCLC. One tumor suppressor gene candidate residing at this locus is STK11/LKB1, a gene implicated in Peutz-Jeghers syndrome. LOH at this locus has been documented in breast cancer and in brain metastases from a variety of human cancers.

Summary

Abundant research has demonstrated that certain patterns of LOH are seen in multiple cancers. Other LOH loci are seen in only certain cancers or occur at a particular stage of disease progression. The availability of high-throughput assays and high-resolution LOH probes is increasing the breadth and depth of studies to elucidate the contribution of LOH to disease, prognosis, or effectiveness of treatment.

In head and neck squamous cell cancer (SCC), 8p21–22 was the most frequent site of LOH. Ye et al. investigated expression of another candidate tumor suppressor gene identified in the region, MTUS1 (mitochondrial tumor suppressor gene 1), and found downregulation of transcription and mutations in exon-coding sequences. Coon et al. found that LOH in a region of 8p that included 8p21 was strongly associated with shorter survival in head and neck SCC patients. Ye et al. discovered that LOH was observed mainly in invasive and metastatic SCC.

LOH at 8p21 is a frequent and early occurrence in NSCLC, believed to occur after LOH events at 3p and 9p, and hepatocellular carcinoma. Despite consistent observations of LOH in this region in lung cancer, specific tumor suppressor genes have not been mapped to 8p21–23 and implicated in the development and progression of lung cancer. LOH at 8p21–23 is also a frequent event in hepatocellular carcinoma (HCC), and a gene encoding a growth inhibitory protein, HCRP1 (hepatocellular carcinoma related protein-1) has been mapped to that area and implicated in HCC.

LOH at 9p

Chromosome 9p is frequently altered in NSCLC, not only by LOH but also by homozygous deletion and gene silencing via promoter hypermethylation. The redundancy in mechanisms aimed at silencing genes in this region and the prevalence of these alterations in smoke-exposed preneoplastic epithelium point to the importance of 9p in the development of lung cancer. The gene most frequently identified in LOH studies of 9p (9p21) in NSCLC is P16 INK4A, a gene that encodes a cell-cycle protein which inhibits CDK4, CDK6, and cyclin-dependent phosphorylation of the Rb gene product. Loss of the P16INK4A gene effectively removes key negative regulation of the cell cycle at the G1 → S-phase transition. Interestingly, experimental evidence suggests that the silencing of the remaining P16INK4A allele after allelic loss is predominantly caused by epigenetic methylation rather than mutation.

LOH at 10q

LOH at 10q22–23 is a frequent observation in SCLC. Although no tumor suppressor genes in this region have been definitively associated with lung cancer, multiple studies have identified LOH targeted to the PTEN/MMAC locus at 10q23. The encoded PTEN protein is a lipid phosphatase that negatively regulates the phosphatidylinositol 3-kinase/AKT (PI3K/AKT) pathway. Loss of PTEN function results in reduced apoptosis and stimulation of cellular proliferation and migration.

LOH at 13q

Chromosome 13q12–14 is a prevalent hotspot for LOH in many cancers. The retinoblastoma (RB1) gene is a well-characterized tumor suppressor gene, at 13q12, the product of which is a key regulator of entry into S phase of the cell cycle. The critical pathway regulated by RB, p16NK4A, cyclin-D1, and cyclin-dependent kinases (CDKs) is disturbed, usually via multiple components, in nearly every case of lung cancer, thus rendering cells insensitive to the signaling involved in regulating mitosis. LOH on 13q12.1–13.1 can be identified in cells obtained from bronchial washing specimens from SCLC and NSCLC patients and is proposed as one of a set of markers that could be used for early detection of lung cancer. Another gene targeted to 13q12–14 is BRCA2, a gene associated with cancer susceptibility and with possible linkages to p53 in the context of DNA damage repair.

LOH at 17p

The TP53 gene at 17p13 is the most frequently altered tumor suppressor gene in human cancers. TP53 is targeted by mutation, methylation, and homozygous deletion, in addition to LOH. The observations of redundant mechanisms of TP53 inactivation and frequency of LOH at TP53 gene point to its critical role in the development of malignant disease. TP53 modulates a broad network of cellular responses, including cell-cycle arrest, apoptosis, DNA repair, cellular senescence, and inflammation, and therefore plays a central role in homeostasis. TP53 mutations can be found in precancerous lesions, but to date studies of LOH at 17p indicate that allelic loss of TP53 occurs during disease progression, after LOH events on chromosome 3p.

LOH at 19p

LOH at 19p13.3 is a very frequent event in NSCLC but is not targeted in SCLC. One tumor suppressor gene candidate residing at this locus is STK11/LKB1, a gene implicated in Peutz-Jeghers syndrome. LOH at this locus has been documented in breast cancer and in brain metastases from a variety of human cancers.

Summary

Abundant research has demonstrated that certain patterns of LOH are seen in multiple cancers. Other LOH loci are seen in only certain cancers or occur at a particular stage of disease progression. The availability of high-throughput assays and high-resolution LOH probes is increasing the breadth and depth of studies to elucidate the contribution of LOH to disease, prognosis, or effectiveness of treatment.
regimens. Currently, there are few loci for which a cause-and-effect relationship to cancer has been established. However, researchers are now armed with an arsenal of tools with which to discover the temporal sequence of LOH events in tumor initiation and progression and contribution of specific tumor suppression genes to cellular and tissue growth regulation.

LOH in Benign Diseases

There is evidence to suggest that benign conditions, such as chronic obstructive pulmonary disease (COPD) and asthma, have a genetic predisposition and probably arise via the interaction between multiple gene products. Studies of microsatellite instability and LOH have been employed to study these diseases with the goals of identifying causative genes and/or develop genetic screening tools for use in epidemiology. LOH in several prevalent noncancer diseases is discussed next.

LOH in Idiopathic Pulmonary Fibrosis

Idiopathic pulmonary fibrosis (IPF) is a serious disease believed to be the result of immune response to tissue damage in the lung. Although IPF is a benign disease, it progresses to bronchogenic carcinoma in about 10% of IPF patients. The observation that the incidence of lung cancer in IPF patients is much higher than the incidence in the general population lends support to the hypothesis that IPF contains precancerous lesions that may progress to peripheral-type lung tumors through the inactivation of critical tumor suppressor genes. Indeed, LOH at several common loci (3p21, 5q32, 9p21, and 17p13) has been documented in IPF.

LOH in Sarcoidosis

Sarcoidosis is a multisystem disease characterized by the formation of noncaseating granulomatous lesions in affected organs, especially the lungs. These lesions can progress to cause fibrosis and, similar to IPF, lead to a higher incidence of lung cancer. Studies of LOH have identified targeted loci on 9p, 9q, and 17q. The genes targeted by LOH were within or proximal to DNA mismatch repair (MMR) genes and genes associated with lymphocyte activation. This pattern is somewhat distinct from lung cancer or IPF and may reflect the absence of linkage with exposure to tobacco smoke.

LOH in Chronic Obstructive Pulmonary Disease and Asthma

LOH has been detected in chronic obstructive pulmonary disease (COPD) and asthma, although the number of studies is limited and these have focused on analysis of chromosomal regions that contain genes suspected in the diseases. There are genes implicated in the development and progression of asthma throughout the genome, although the “hotspot” for alterations via LOH is chromosome 14q, which contains several target genes that have been implicated in asthma, including prostaglandin E receptor 2 (PTGER2), arginase II (ARG2), and alpha-1-antichymotrypsin precursor (AACT). Within asthmatic patients, those with the greater number of genetic alterations have higher mean immunoglobulin E and blood eosinophils, both of which are indicators of inflammation and bronchial hyperresponsiveness. Analyses of COPD samples demonstrate that LOH occurs most frequently at the thyroid hormone receptor alpha-1 (THRA1) locus on chromosome 17q21. It is believed that exposure to tobacco smoke is a risk factor for the genetic changes associated with COPD, and individuals with COPD (similar to those with IPF) carry a greater risk for the development of lung cancer.

Summary

Clearly, LOH is a well-documented occurrence in benign, premalignant, and malignant disease. Although some genes targeted by LOH represent key early events in pathogenesis, many occur at later stages of disease progression, thus making their specific contributions difficult to discern. Comprehensive genome-wide studies of LOH in diseases using SNPs and bioinformatics analysis to analyze patterns of multiple genetic alterations have given clinicians more precise tools for diagnosis and treatment of cancers. Of particular interest is the importance of LOH research for discerning the incremental effects of accumulating mutations and how specific LOH events or clusters of LOH events can influence a patient’s response to treatment or inform prognosis. The availability of scalable, high-throughput LOH screening techniques promises to give researchers and clinicians the information they need to design more effective and customized treatment plans for patients with cancer or at higher risk for cancer.

References


11. Loss of Heterozygosity


Loss of Heterozygosity

11. Loss of Heterozygosity


Introduction

In situ hybridization (ISH) has become an extremely useful tool for the clinical pathology laboratory to aid oncologists, geneticists, and infectious disease specialists in the diagnosis and treatment of their patients. ISH utilizes nucleic acid (DNA or RNA) probes to assess intact cells for various types of genetic alterations. Common applications of in situ hybridization include its use to detect cancer cells in cytological specimens, chromosomal alterations in resected tumor specimens that predict prognosis, and response to therapy of certain cancer types and microorganisms in various specimen types. ISH is typically performed either with fluorescently labeled probes or with probes that are subsequently visualized with a chromogen such as diaminobenzidine. If performed with fluorescently labeled probes, the technique is referred to as Fluorescence In Situ Hybridization, or FISH. Alternatively, if the technique is performed with a probe that requires subsequent visualization with a chemical reaction that produces a colored chemical at the site of the probe (a chromogen), the technique is referred to as Chromogenic In Situ Hybridization, or CISH.

In Situ Hybridization: General Principles

To understand how in situ hybridization works, one should be familiar with (1) DNA and RNA composition and structure, (2) principles of base pairing, (3) denaturation and hybridization, and (4) factors that influence denaturation and hybridization. It is not possible to cover these topics extensively; however, the following paragraphs briefly touch on the most important aspects of these topics.

DNA Composition and Structure

DNA and RNA are each composed of four nucleotides that form long chains. The nucleotides that comprise DNA contain one of four bases: adenine, cytosine, guanine, or thymine (Fig. 12.1). RNA is composed of nucleotides that also contain the same nucleotides except that uracil substitutes for thymine (Fig. 12.1). RNA is composed of nucleotides that also contain the same nucleotides except that uracil substitutes for thymine. Another important difference between DNA and RNA is that DNA is double stranded and RNA is single stranded.

Principles of Base Pairing

To comprehend hybridization, it is essential to understand how base pairing occurs in double-stranded DNA or DNA-RNA hybrids. In double-stranded DNA, nucleotides with an adenine (A) or thymine (T) pair with one another (referred to as “base pairing”) through two hydrogen bonds whereas guanine (G) and cytosine (C) pair with one another through three hydrogen bonds (see Fig. 12.1). The only difference with DNA-RNA hybrids is that uracil (U) rather than thymine being assessed for may be present in only a small percentage of the cells. Other molecular techniques in which the tissue is homogenized before analysis often cannot detect these alterations because of dilution by cells not carrying the alteration (e.g., by benign stromal or inflammatory cells that are invariably present in tumors).
Nucleotides are the building blocks of both DNA and RNA. Nucleotides are composed of the sugar deoxyribose (DNA) or ribose (RNA), a phosphate group, and a base. The four bases found in DNA are the purines guanine and adenine, and the pyrimidines cytosine and thymine. Uracil substitutes for thymine in RNA. Single-stranded DNA is a polymer of nucleotides. The 5'-end of a single-stranded DNA molecule is the end that has a free phosphate group and the 3'-end is the end that has a free 3'-OH group. Double-stranded DNA is composed of two single-stranded DNA molecules that run in opposite directions. In double-stranded DNA, guanine pairs with cytosine through three hydrogen bonds and adenine pairs with thymine through two hydrogen bonds.
base pairs with adenine. It is the pairing of A’s to T’s (or U’s in the case of RNA) and of C’s to G’s that keeps the two strands of DNA or the two strands of a DNA-RNA hybrid paired to one another. Because GC base pairs have three hydrogen bonds, they are stronger and require more energy to break than AT base pairs. If two strands of DNA have bases that pair perfectly, that is, every A, C, G, or T on one strand is matched with a T, G, C, or A, respectively, on the opposite strand, then the two strands are said to be 100% complementary.

**Denaturation, Renaturation, and Influencing Factors**

Denaturation refers to the process of making double-stranded DNA (or DNA-RNA hybrids) single stranded. Denaturation is also sometimes referred to as “melting” the DNA. Denaturation is brought about by breaking the hydrogen bonds that hold the two strands of DNA together; this is most commonly achieved by applying heat (i.e., raising the temperature of the sample). In addition, certain chemicals such as formamide can be used to promote denaturation. Because G/C base pairs are stronger than A/T or A/U base pairs, double-stranded DNA fragments (or DNA-RNA hybrids) with a higher percentage of G’s and C’s require more heat (i.e., higher temperatures) to denature than will double-stranded DNA fragments (or DNA-RNA hybrids) with a higher percentage of A’s and T’s. Renaturation refers to the process of bringing two complementary strands of DNA (or complementary DNA and RNA) back together.

Hybridization occurs when two strands of complementary single-stranded DNA or RNA molecules “stick to” one another. The greater the complementarity of the two single-stranded molecules, the more likely that two strands will hybridize to one another. In addition, due to stronger base pairing, DNA-DNA or DNA-RNA hybrids that are 100% complementary will be more stable and require more energy to break apart than those which are less complementary. Two strands that are less complementary (e.g., only 95% complementary) can hybridize to one another under certain conditions, which are referred to as “less stringent” conditions. The tendency for a DNA or RNA probe to stick to sequences that are not complementary is referred to as nonspecific hybridization. Nonspecific hybridization can be prevented by maintaining conditions, referred to as “stringent conditions,” that favor hybridization of the most complementary sequences but prevent the hybridization of less-complementary sequences. Stringent conditions are generally achieved by keeping the temperature at the highest possible temperature that allows the most complementary single-stranded DNA or RNA probe to anneal to its completely complementary target without allowing it to anneal to targets that are less complementary. Depending on how similar the noncomplementary sequence is to the complementary sequence, the difference between stringent and nonstringent conditions may be just a degree or two. For example, 37°C may favor the annealing of a probe that is 100% complementary to its target sequence while preventing a probe that is only 95% complementary from binding to its target sequence. However, dropping the hybridization temperature just a couple of degrees, to 35°C, may favor the binding of both probes to the target sequence.

The melting temperature of double-stranded DNA in solution is approximated by the following formula:

$$T_m = 81.5 + 16.6 \log (\text{Na}^+) + 0.41(\% \text{ GC}) - 0.63 \times (\% \text{ formamide}) - (300 + 2000(\text{Na}^+)) / N$$

where $T_m$ is the melting temperature in degrees centigrade, $\text{Na}^+$ is the molar concentration of sodium ions, $\% \text{ GC}$ is the percentage of GC base pairs in the hybridized molecules, $\% \text{ formamide}$ is the percentage of formamide (volume/volume), and $N$ is the length in bases of the hybrid. This formula may appear somewhat daunting for those of us who are less mathematically inclined, but conceptually, the important points to note are that increasing the salt concentration inhibits melting and increasing formamide promotes melting; this is because high salt concentrations favor hydrogen bonding of the two strands of a DNA-DNA or DNA-RNA hybrid while formamide helps break those hydrogen bonds. Similarly, it is important to note that under identical conditions, shorter fragments of DNA melt at lower temperatures than longer fragments of DNA. A familiarity with this formula can be helpful when trying to troubleshoot problems with nonspecific or incomplete hybridization.

**Probes**

**Types of Probes**

All in situ hybridization techniques utilize DNA or RNA probes. These probes are designed to hybridize to specific target sequences of interest such as genes that have been implicated in inherited diseases or cancer as well as to microorganisms of various types. DNA probes that are commonly used for genetic and/or oncologic applications have been categorized as chromosome enumeration probes (CEP), locus-specific indicator (LSI) probes, telomeric probes, and chromosome paints.

**CEP Probes**

CEP probes hybridize to repetitive DNA sequences found near the centromeres of chromosomes, which are referred to as alpha-satellite DNA. These regions are composed of ~171-bp sequences that are tandemly repeated hundreds to thousands of times and span approximately...
250,000–5,000,000 bases. The repeat regions of the different chromosomal centromeres exhibit substantial sequence divergence of approximately 20–40%. Because of this, probes specific for each of the centromeres are available for most, but not all, chromosomes. A few chromosomes, for example, chromosomes 13 and 21, 14 and 22, and 5 and 19, have alpha-satellite repeat sequences that are too similar to allow one to distinguish these chromosomes from one another as a consequence of cross-hybridization.

CEP probes are used to enumerate the number of copies of a given chromosome in a cell. CEP probes are able to enumerate chromosome copy number because if the centromere of a chromosome is lost, the whole chromosome will generally be lost. One advantage of CEP probes is that because they hybridize to sequences that have high copy number, they provide strong (i.e., bright) signals. In addition, because these regions are tightly compacted in the heterochromatic regions of the chromosome, the signals provided with CEP probes are generally tight (“crisp”) rather than diffuse.

**Locus-Specific Probes**

Locus-specific probes hybridize to unique sequences (i.e., nonrepetitive DNA sequences) and are generally used to determine if specific genes are amplified (e.g., HER2), deleted (e.g., p53 or p16), or translocated (e.g., BCR/ABL translocation). These probes typically hybridize to a region that ranges from 40 to 500 kilobases (kb). Probes that hybridize to regions smaller than 40 kb often produce weak signals that can be difficult to see. Probes larger than this can produce diffuse signals that are difficult to distinguish as single spots.

**Telomeric Probes**

Telomeric probes are not actually probes to telomeric sequences but to unique DNA sequences found very near the telomeres (so called subtelomeric probes). Unique telomeric probes are available for 41 of the 46 chromosomal telomeres. The only telomeres for which there are not probes are for the p arms of the acrocentric chromosomes 13, 14, 15, 21, and 22. However, these regions are composed primarily of ribosomal DNA, and deletions or duplications of these regions are not generally thought to have clinical significance.

**Chromosomal Paints**

Chromosomal paints are mixtures of probes that hybridize to the entire length of one or more chromosomes. The probes that comprise chromosomal paints are generally prepared by isolating individual chromosomes by flow cytometry and then performing PCR amplification with degenerate oligonucleotide primers. The fluorescently labeled probes generated from this can then be used as a “paint” that highlights the entire chromosome homogeneously along its length. Chromosomal paints can be used to identify specific chromosomes in a metaphase spread. This can be particularly helpful when standard karyotyping is unable to identify a chromosome (e.g., marker chromosomes).

**Probe Preparation**

Probes are generally prepared from fragments of DNA that have been cloned into bacterial (BAC), P1 (PAC), or yeast (YAC) artificial chromosomes. Probes that are used for FISH can either be directly or indirectly labeled with a fluorophore. A fluorophore is a molecule that fluoresces when excited by light of specific wavelength. There are a wide variety of fluorophores that can be used for FISH. Two of the more commonly used fluorophores are fluorescein isothiocyanate (FITC) and Texas Red, which fluoresce green and red, respectively.

**Direct Labeled Probes**

Direct labeled probes are prepared by incorporating a fluorophore-labeled nucleotide into the probe, usually by nick translation or random priming. Because the fluorescently labeled probe is bound to its cellular target in a single hybridization step, no further processing is required to visualize the probe. Directly labeled CEP and LSI probes are commercially available singly or as probe mixtures of up to four different probes and are generally labeled with red, green, aqua, or yellow fluorophores.

**Indirect Labeled Probes**

Indirectly labeled probes contain nucleotides that have covalently attached reporter molecules such as biotin or digoxigenin. These probes require an additional step after hybridization in which the probe is generally visualized by applying fluorophore-labeled avidin or fluorophore-labeled antidigoxigenin in a second step. A significant disadvantage of the use of indirectly labeled probes is that it adds extra steps to the procedure. However, potential advantages of indirectly labeled probes include that it is often possible to obtain stronger signals with indirectly labeled probes because of the ability to achieve greater signal amplification and that indirect labeling allows for greater versatility in probe use as one can use this technique to visualize any probe that has been labeled with biotinylated 2'-deoxyuridine 5'-triphosphate (dUTP) or digoxigenin.

**“In-House” Developed Probes**

Most diagnostic laboratories utilize commercially available probes. However, a lab may want to prepare their own “in-house” probe if a probe is not commercially available. If a lab prepares its own probes for diagnostic use, it is important to extensively validate the probe before clinical implementation to make sure that it hybridizes to the intended target and provides the expected results on normal and abnormal specimens. Wiktor et al. have recently provided recommendations for the validation of probes for clinical use.

**Indirect Labeled Probes**
Fluorescence In Situ Hybridization (FISH)

Overview of Steps Involved in FISH

The steps involved in performing FISH are as follows: (1) obtain specimen for FISH, (2) prepare specimen for hybridization ("prehybridization"), (3) hybridize probes to target DNA in specimen, (4) remove nonspecifically bound probe by washing, and (5) assess signals in specimen using fluorescence microscopy (Fig. 12.2). Each of these steps is described in more detail below.

Specimens

FISH can be performed on a wide range of specimens including peripheral blood, paraffin-embedded tissue, or cytology specimens (e.g., urine, sputum, or endoscopic brushings). Frozen tissue will not work well because freezing results in intracellular ice crystals that disrupt the morphology of the cells. The fixative that gives the best FISH results for paraffin-embedded tissue is neutral buffered formalin. However, tissue that has been left in formalin too long (>48 h) before paraffin embedding will not provide
good FISH results.7 FISH on specimens that have been fixed in B5 fixative, which is often used for hematological specimens, may require isolation of nuclei before FISH to obtain good results.8 Some fixatives such as Prefer work poorly if at all for FISH.9

Prehybridization

The goal of this step is to prepare the cells in the specimen so that the probe can efficiently hybridize to its cellular DNA target but do so without significantly disrupting the morphology of the cells. Detailed protocols for prehybridization and hybridization can be found elsewhere.1011 The pretreatment required for different specimen types (e.g., fresh cells vs. paraffin-embedded tissue) is similar but frequently has some important differences. For example, certain specimens such as paraffin-embedded tissue generally need to be treated with a protease such as pepsin to increase the accessibility of the probe DNA to its nuclear DNA target. It is important to not over- or underdigest the tissue with pepsin. Overdigestion can lead to a decrease in signal intensity and destroy nuclear morphology. Underdigestion can lead to auto-fluorescence, and an underestimation of signal copy number may result. Other methods that have been used to facilitate entry of the probe into the cell include treatment with dilute acid or non-ionic detergents.12

Denaturation and Hybridization

This step involves denaturing the probe and cellular DNA and then allowing the probe to hybridize to its cellular DNA target. The probe hybridization solution contains not only the probe DNA but a type of DNA that is referred to as COT-DNA. COT-DNA is DNA that hybridizes to highly repetitive DNA sequences which are present throughout the genome. Without COT-DNA, the probe DNA would nonspecifically bind to these repetitive DNA sequences, resulting in many nonspecific signals rather than the specific signals that are desired. In one of the more widely utilized FISH hybridization protocols, the probe and target DNA are codenatured at about 73°C for about 3 min in 50% formamide. Formamide is added to lower the temperature at which the probe and cellular DNA melt. This step is important because high temperatures can destroy the cellular morphology that one is attempting to maintain. The temperature is then lowered to about 37°C to allow the probe DNA to hybridize to its specific target. The hybridization temperature is less fastidious for FISH assays than for PCR assays because FISH probes are much longer than “PCR probes.” Hybridization is generally allowed to occur over about 4–12 h. Factors that can influence the efficiency and specificity of the hybridization include the probe sequence and hybridization temperature. If the probe sequence is not unique, additional signals may present that represent cross-hybridizing sequences. If the hybridization temperature is too high, the signals may be weak, and if the hybridization temperature is too low, there may be an abundance of nonspecific background signal.

Removing Nonspecifically Bound Probe

Following hybridization, the slide is washed with a specific washing solution. The goal of this step is to remove any probe that is not specifically bound to the desired target without removing the probe which has bound to the desired target. This step is generally done by washing the slide in a solution that is heated to 73°C and contains 0.4× SSC/0.3% NP40 (nonformalin-fixed samples) or 2.0× SSC/0.3% NP40 (formalin-fixed samples). A higher salt concentration (less-stringent wash) is used for paraffin-embedded specimens because the probe-target hybrids are less strong for paraffin-embedded specimens than for non-formalin-fixed samples; this may result from the shorter length of target DNA sequences in paraffin-embedded tissue or the persistence of protein in the hybridized region that causes hybrid destabilization.10 If the posthybridization wash is too stringent, all or most of the probe may be washed off, and there may be either no signals or weak signals when viewed under the microscope. On the other hand, if the wash is not stringent enough, there may be an excessive amount of nonspecific hybridization background when viewed under the microscope. If indirectly labeled probes are used, an additional step is required at this point to attach the fluorophore to the probe. The procedure used will depend on the type of reporter molecule (biotin, digoxigenin, other) that has been previously chosen.

Finally, approximately 10 μl of a solution that contains compounds known as DAPI and “Antifade” is placed on the slide, and the slide is cover-slipped. DAPI (4′,6-diamidino-2-phenylindole) is a nuclear counterstain that fluoresces a blue-gray hue when viewed with a fluorescent microscope. DAPI allows one to see the nuclei. Without DAPI, the FISH signals would appear to be free floating in a dark background. Antifade is used to inhibit photobleaching, which is the tendency of fluorophores to fade with exposure to light and heat.13 Because fluorophores are susceptible to photobleaching, it is best to store hybridized slides and probe in a refrigerator and in the dark. In addition, when viewing the slide on the microscope, it is important to remove the slide from the path of light by closing the condenser when not viewing. If a slide has lost its fluorescence, it can frequently be rehybridized with good results because the target DNA is still intact. We have successfully rehybridized slides that are several years old.

Fluorescence Microscopy

Basic Principles

Following hybridization, FISH signals are assessed with a fluorescence microscope (usually an epi-fluorescence microscope) equipped with the appropriate filters necessary to see the probe signals. A detailed discussion of fluorescence microscopy is beyond the scope of this chapter, but there are several excellent reviews on this subject.14
Nonetheless, it is important to understand at least a few of the basic aspects of how fluorescence microscopy works. As already noted, a fluorophore is a molecule that absorbs light of a certain wavelength (actually a certain range of wavelengths) and then reemits that energy as light of longer wavelengths. For example, as shown in Fig. 12.3, the fluorophore that is referred to as Spectrum Green absorbs light that ranges from about 470 to 510 nm and then reemits it (i.e., fluoresces) as light of higher wavelengths that range from about 500 to 550 nm. The goal of FISH is to cause the fluorophore-labeled DNA or RNA probe to fluoresce and highlight the target of interest in the specimen. To do this, the fluorescence microscope must have a light source that is capable of exciting the fluorophore. Mercury or xenon arc lamps are typically relied upon as light sources for fluorescence microscopes. These lamps emit light of wavelengths that can be absorbed by the fluorophores. A combination of filters (excitation, beam splitting interference, and barrier) select an appropriate wavelength of light to excite the fluorophore and ensure that the main light that reaches the observer’s eye is the desired fluorescence wavelength. These filters are generally incorporated into a single housing device that is referred to as a filter cube. The appropriate combination of filters depends on the fluorophore(s) that is (are) being utilized. Cubes are available for most fluorophores. Most cubes contain filters that allow one to look at only one probe at a time. However, dual-pass and triple-pass filters that allow one to see red and green signals or the red signal, green signals, and DAPI counterstain simultaneously are available and very useful.

It is frequently necessary to take a picture of the cells and their fluorescent signals. This requires the use of imaging systems and software that enable one to capture the different colored probe signals in a single composite image. When a multicolor probe set has been used, it may be particularly difficult to take a picture that shows all the signals when using a normal camera because the signals are often present in different planes. This limitation can be overcome with a device known as a Z-stack camera that allows one to easily capture all the signals in a cell as a single image.

**Chromogenic In Situ Hybridization (CISH)**

CISH is another type of in situ hybridization that is very similar to FISH except that the probe is visualized by a chromogenic reaction after it has hybridized to its target rather than being visualized as a fluorescent signal. There are several ways in which one can perform CISH. One method incorporates biotinylated-dUTP into the probe. After hybridization of the probe to its target, horseradish peroxidase-labeled streptavidin is added to the slide. The streptavidin attaches to the biotin. Diaminobenzidine (DAB) is then added, and the horseradish peroxidase converts the DAB to a brown chromogen at the site of the probe. Another CISH method utilizes digoxigenin-labeled probes instead of biotin-labeled probes but also utilizes peroxidase and DAB to visualize the probe.

CISH has several advantages over FISH. A major advantage that CISH has over FISH is that CISH slides can be viewed with a light microscope, and the tissue architecture is much easier to discern with CISH; this makes it easier to correlate molecular alterations with histological or cytological features with CISH than with FISH. In addition, CISH can be performed in conjunction with immunohistochemistry to assess for correlations between chromosomal alterations and protein expression. A final important advantage of CISH is that the permanent slides obtained after CISH analysis do not fade and thus are suitable for long-term archiving.

Disadvantages of CISH relative to FISH include that it can only easily identify a single target at a time, and the signals produced with CISH are less discrete than those produced by FISH, which makes signal enumeration less precise than that which is possible by FISH. The inability of CISH to identify more than one target may or may not be important depending on the application. For example, CISH is frequently used to identify viruses such as Epstein-Barr virus or cytomegalovirus in lung tissue sections (Fig. 12.4). For this particular application, the inability of CISH to identify more than a single target is not a limitation.

---

**Fig. 12.3.** Excitation (absorption) and emission spectra for more commonly used fluorophores.

**Fig. 12.4.** Identification of cytomegalovirus (CMV) in lung tissue section utilizing chromogenic in situ hybridization (CISH). The purple-stained nuclei contain CMV.
However, the inability to detect more than one target at a time could be a limitation for some applications. For example, one commercially available FISH assay for HER2 amplification utilizes two probes, a green probe to the chromosome 17 centromere and a red probe to the HER2 gene. With this FISH assay, a HER2/CEP17 signal ratio >2.0 is considered evidence of HER2 amplification and justification for Herceptin therapy. The rationale for including the CEP17 probe in this probe set was that some tumors may exhibit gains (i.e., more than two copies) of the HER2 gene because of polysomy (i.e., extra copies) of chromosome 17 in the tumor but not because of true amplification of the HER2 gene. It is not fully known at this time whether the absolute HER2 copy number by itself is predictive of response to Herceptin or whether knowing the ratio of HER2/CEP17 is important. However, if knowing the ratio is important, that information would not be available with the CISH assay. Other probe sets, for example, a FISH probe set for bladder cancer detection, utilize as many as four different probes in a single cocktail to achieve high sensitivities for bladder tumor detection. Other Techniques that Utilize In Situ Hybridization

Comparative Genomic Hybridization (CGH)

CGH is a powerful technique that can, similar to conventional karyotyping, be used to analyze the entire genome of cells for chromosomal abnormalities. CGH is particularly useful for assessing solid tumors for chromosomal abnormalities. As first described, CGH is preformed by isolating DNA from tumor tissue and labeling it with a green fluorophore, then isolating DNA from normal tissue and labeling it with a red fluorophore. Equal amounts of the green-labeled tumor DNA and red-labeled normal DNA are then cohybridized to a normal metaphase spread (i.e., a karyotype without any abnormalities). If there are no chromosomal abnormalities in the tumor tissue, then equal amounts of green and red probe will hybridize to the chromosomes in the normal metaphase spread, and the color of all the chromosomes in the metaphase spread will be yellow (because equal amounts of red and green make yellow). However, if there are, for example, three copies of chromosome 1 in the tumor, then there will be more green-labeled chromosome 1 DNA than red-labeled chromosome 1 DNA, and chromosome 1 in the metaphase spread will appear green. Likewise, areas in the metaphase spread that appear green represent chromosomal areas that have been gained in the tumor while areas that appear red represent chromosomal areas which have been deleted in the tumor. In this fashion, CGH can be used to identify many of the chromosomal abnormalities present in a tumor. CGH is generally not performed with metaphase spreads anymore but instead with microarrays of DNA clones that are largely representative of all the loci along the length of the 24 chromosomes. This technique, which is referred to as array CGH, has the potential for higher resolution and greater reproducibility than can be obtained with metaphase spreads. A representative example of a CGH array is shown in Fig. 12.5.

CGH has several advantages over conventional karyotyping, including (1) the fact that the tumor be grown in culture before analysis, (2) the cells do not have to be in metaphase to determine the types of chromosomal abnormalities present, and (3) it has higher resolution, typically down to about 1 megabase, whereas karyotyping has a resolution down to about 5 megabases. The only disadvantage of CGH relative to conventional karyotyping is that it cannot detect balanced chromosomal alterations (e.g., recip-

---

**Fig. 12.5.** Array comparative genomic hybridization. A genomic array spotted with DNA from 287 targets in the genome is shown in (a). The targets include subtelomeric regions, oncogenes, tumor suppressor genes, and microdeletion loci. Each target is spotted in triplicate. *Green-labeled* tumor DNA from a single tumor (a bile duct tumor) and *red-labeled* normal DNA were cohybridized to the DNA spots on the array. Gene amplifications within the tumor appear as *green spots* (in triplicate) and deletions within the tumor appear as *red spots* (in triplicate). The chromosomal location of the gene amplifications and deletions for this particular tumor are illustrated in (b). The tumor shows deletions of 5q, 6q, 12q, 14q, 18q, 19, and 21, gain of 15, and amplification of the MDM2 gene at 12q14.3-q15.
rocal translocations). Although CGH is generally not used clinically, it has been a powerful research tool for identifying the most common chromosomal abnormalities that are present in various tumor types, for example, lung cancer.\textsuperscript{20–37}

## Multicolor Whole Chromosome Painting (M-FISH and Sky)

Spectral karyotyping (SKY) and multicolor FISH (M-FISH) are highly sophisticated techniques that allow one to generate a color-coded karyotype.\textsuperscript{38,39} In other words, these techniques produce a karyotype in which each chromosome has its own unique color. The generation of chromosomal paints for each chromosome is accomplished with combinatorial labeling of the chromosome specific probes with just five fluorophores. SKY and M-FISH can be helpful in identifying chromosomes that are difficult to identify, such as marker chromosomes or chromosomes which are involved in complex rearrangements. This aspect can be particularly helpful for solid tumor genetics because solid tumors often have very complex karyotypes that are difficult to elucidate (Figs. 12.6 and 12.7).

## Summary

ISH is a technique that utilizes nucleic acid (DNA or RNA) probes to assess intact cells for various types of alterations. To understand how in situ hybridization works, one should be familiar with DNA and RNA composition and structure, principles of base pairing, and factors that influence denaturation and hybridization. Techniques that utilize ISH include FISH, CISH, comparative genomic hybridization, spectral karyotyping, and multicolor FISH. These techniques have found many important applications in oncology, genetics, and infectious disease.

## References


“Omics” is a term designating a complete analysis of biological systems in which entire metabolic pathways are studied. “Omics” changes the methodological approach from small-scale research of one gene, one protein, or one metabolic reaction to the large comprehensive level of synthetic study in which the entire genome, protein composition, or metabolic pathway is studied simultaneously. The main breakthrough to the “omics era” was a confirmation of the complete human genome sequence. Entering this new research territory was possible because of an improvement in scientific instrumentation that has been able to perform high-sensitivity testing at high-throughput mode. The software-directed, “walk-away” DNA and protein microarray or highly sensitive, operator-friendly mass spectrometry have become a driving force for omics research.

These instruments become data factories that produce enormous amounts of information, transcending the capacity of the human brain. The need to store and analyze such a large amount of data has stimulated the advancement of new bioinformatics tools that contribute significantly to the development of “omics” disciplines. Despite its short history, this comprehensive methodology has proved its effectiveness by advancing our understanding of physiological and pathological processes, which brings hope for more accurate diagnosis and treatment of diseases.¹

**Genomics and Transcriptomics**

In light of omics as a “method of thinking large,” genomics can be defined as a comprehensive study of DNA structure and function. Genomics includes intensive efforts to determine the entire DNA sequence of organisms and fine-scale genetic mapping. The field also includes studies of intragenomic phenomena such as heterosis, epistasis, pleiotropy, and other interactions between loci and alleles within the genome.²

The analysis of gene expression has evolved into transcriptomics, an independent field with specific methodology and instrumentation. Transcriptomics examines gene expression on the mRNA level and describes the genes that are being actively expressed at any given time.³

The main shortcoming of both genomics and transcriptomics is their inability to predict phenotypic and functional consequences of mutations or single nucleotide polymorphisms. Only analysis of the proteins, on an effector level, is able to fill this gap and bring another “omics” into play – proteomics.

**Proteomics**

The aim of proteomics is to perform a systematic study of proteins and to present a complete explanation of their structural, functional, and regulatory roles in a particular biological system. Similarly to the other “omics” disciplines, proteomics is defined as a comprehensive study of the entire protein composition that is characteristic for a particular physiological or pathological state of the cell or whole organism. Technological and methodological progress has made it possible to go beyond a simple biochemical analysis of a single protein and to progress into an investigation of complex protein mixtures. In conjunction with the constantly growing field of bioinformatics, proteomics is able to explain physiological and pathological processes in both normal and diseased cells, and therefore it has the potential to become an essential component in laboratory diagnostics.

The first tool that assisted the evolution of protein biochemistry into proteomics was two-dimensional gel electrophoresis. Despite its imperfections, two-dimensional gel electrophoresis is still used for protein profiling and/or for quantitation of protein expression. Low sensitivity is a major limitation that makes this method unfeasible for analysis of posttranslational modifications and protein-protein interactions; however, rapid progression in mass spectrometry and protein microarray technology has offered solutions to the two-dimensional electrophoresis limitations. The new generation of user-friendly mass spectrometers is able to analyze a protein mixture with high sensitivity and specificity as well as to perform testing in a high-throughput mode on different
types of specimens (whole blood, serum, body fluids, lavages, and tissue and cellular extracts). The improvements in protein microarray technology and in affinity chromatography permit efficient isolation of protein complexes and study of protein-protein interactions. All these bring proteomics closer to being a useful diagnostic tool for any human disorders, and for hematological disease particularly.

Proteomics: A New Diagnostic Tool for Hematology

Genomics and transcriptomics measure gene expression and transcriptional activities of the cells. It has been widely recognized that the changes in protein expression cannot be predicted at the level of gene amplification or at the level of mRNA production. For this reason, as well as because proteins are the true effectors that change the functional state of cells, it is the proteins which should be examined as the most pertinent markers for cellular function. Performed on cytological or surgical specimens, proteomics shows a complete picture of cellular activity and, along with other omics, provides new information about physiology, pathology, and cellular reaction to treatment. If cytological or histological diagnostic material is not accessible, or if not a cellular activity but discovery of a new marker of disease is of particular interest, protein profiling on biological fluids such as blood, urine, bronchoalveolar lavage (BAL), or serosal transudate or exudate as well as synovial fluid may be performed. The primary objective for protein profiling is the discovery of a protein or its posttranslational modification that is different in the pathological than in the physiological state. This profiling is accomplished by a comparison of an entire collection of proteins present in the specimen derived from healthy as well as diseased organisms, and it is named comparative proteomics.

The Ideal Proteomics

The ideal proteomics can be described as a method, or a series of methods, that can analyze an unmodified specimen in its entirety in both the control and testing groups and is able to detect 100% of the proteins and their modifications present in these sets. The results of proteomics experiments are submitted for an evaluation, which analyzes the entire body of data and recognizes the proteins that are different in testing compared to the control group (proteins of difference). These selected proteins are subsequently used as an organism and become a basis for the explanation of biochemical pathways or for diagnosis of diseases.

This simple model of proteomics becomes complicated by what is referred to as abundance. There are approximately 35,000 human genes encoding proteins. Most of the proteins undergo posttranslational modifications, so that the actual number of functionally active proteins exceeds 1 million. Therefore, if a particular metabolic process involves 0.1% of the total protein composition (proteome) to fulfill the proteomics requirement that the entire spectrum of proteins must be detected, the protein collection characteristic for this state exceeds 1,000 individual particles.

The problem of abundance affects not only the laboratory (separation and identification) but also the analytical (data analysis) portion of the proteomics process. To efficiently separate and identify proteins, especially those at low concentration, that are present in the unprocessed sample, the highest possible sensitivity is desired which overcomes the “masking effect” created by highly concentrated particles. However, testing with high sensitivity instrumentation detects “artificial peptides” that are not naturally produced. Also, for the data analysis process, this abundance makes it more difficult to obtain sufficient statistical power to prove that the detected differences in protein composition between testing and control groups are systematic, not incidental.

Shotgun and Protein Profiling

For diagnostic purposes, proteomics methods are not restricted to surgically removed bone marrow or lymphatic tissues, but they can also be used to analyze fine-needle aspirates, effusion fluids, urine, and serum. For all these specimens, proteomics provides a unique opportunity to evaluate almost the entire protein content. Because the amount of analyzed proteins is colossal, to overcome “abundance” effects the original specimen is fractionated and divided into portions containing proteins with similar physicochemical properties.

The process of fractionation is performed by gel electrophoresis or by mass spectrometry, and the information received during this process is used in two different ways. The first one is the “shotgun” protocol, in which the protein identification is performed on a chosen fraction that separates the control from the testing group. This fraction of difference is represented by a spot on the electrophoretic gel or by a peak on spectrometric spectra that is unique for the testing or control samples. The proteins present in this fraction and identified by fingerprinting and/or a sequence tag methodology are subsequently evaluated as a potential diagnostic marker.

The second approach is “protein profiling,” in which a pattern of all electrophoretic spots or a pattern of all spectrometric peaks obtained from the control and test groups is compared. In this methodology, the pattern that exclusively represents one of these groups (discriminatory pattern) or which is highly specific for a particular cellular state (or state of the organism) becomes a biomarker. Subsequently, this discriminatory pattern is evaluated as a potential marker, such that the identification of its single components is not necessary.

The electrophoretic pattern is defined as the presence or absence of protein spots at particular locations, as well as the intensity of their staining. The mass spectrometry pattern is described as either the presence or absence of spectral peaks
Protein Identification

There are two general methods of protein identification - database dependent and database independent.14 In the database-dependent method, a simple protein mixture or a fraction chosen from two-dimensional electrophoresis is enzymatically digested, or chemically cleaved, and submitted for mass spectrometry analysis. As a result, molecular masses of the peptides from the mixture are obtained that are subsequently used for identification of the proteins from which they derive. This identification is made using protein mass fingerprinting15 and/or peptide sequence tagging16,17 methodology.

The fingerprinting protein identification approach utilizes the molecular masses of the peptides from the mixture and compares them against a database containing a collection of proteins and peptides derived from them. Based on the similarity of masses, a cluster of protein candidates that may be related to the particular peptide is selected. For each of these protein candidates, a matching score is calculated that reflects the probability that this candidate is a factual protein present in the specimen. The higher the score, the more certain it is that a given protein is truly present in the specimen.

In the peptide sequence tagging identification method, the information is collected from a tandem MS/MS. The first MS measures masses of intact peptides present in the mixture. The role of the second MS is to assess masses of a series of ions generated from a parent peptide in the collision chamber located between mass spectrometers. The derived ions are arranged according to increased molecular masses. The difference of masses between subsequent ions serves as a criterion for establishing which amino acid is added and thus determines the sequence. These sequenced fragments (short sequenced tags) are used for peptide and protein identification, which requires comparison with a sequence database. The protocol for database searching and score calculation varies depending on the goal,18,19 but the principles of this process are similar for both the fingerprinting and tagging methods. Although each method, fingerprinting and tagging, can be used separately, the combination of the two brings the most efficient results.

A relatively new protein identification method is based on comparing the tandem MS/MS spectra obtained during experimentation with the spectra and corresponding sequences stored in the spectral library. Specially created software performs the search, and as a result a sequence of the potential peptide is attained.20

The database-independent method of protein identification is usually performed for those organisms for which little or no sequence information is available. One of them is the de novo sequencing protocol21 that is used not only in cases of incomplete database information about particular protein but also for validation of the database-dependent identification or for identification of new peptides.

Data Validation

As already mentioned, the amount of data produced by proteomics instrumentation is so sizeable that it requires a special method for its storage and analysis. The need for that method has stimulated a dynamic development of bioinformatics. There are two general levels of proteomics investigation where bioinformatics plays a key role. The first is on the instrument level, where “raw” data are produced. What are referred to as raw data are, in fact, data that are already processed by the instrument, which is referred to as “spectra deconvolution.”14 Spectra deconvolution is the process whereby continuous data obtained by the instrument are transformed into a list of spectral peaks. This list of peaks is subsequently used as input data for the second level of
proteomics investigation, which leads to protein identification (database search, score calculation, etc.) and peak validation as a potential marker. The process of “deconvolution” consist of subtraction of background, noise removal, charge state deconvolution, and deisotoping.22 As a result, the peaks that are extracted from the continuing line of the mass spectrometry recording are obtained and submitted for the next level of elaboration involving peak clustering and identification of the peaks of difference. Peak clustering consists of grouping of peaks that are characteristic for a particular testing and control group (peaks of difference). The peaks of difference are identified by a variety of statistical methods, such as hierarchical clustering and principal component.

All commercial mass spectrometer systems support the first level of data processing, but the process is not standardized, and often the final product differs across different brands of instrumentation. There is thus a real need for standardization in this area, as differences in peak identification affect not only subsequent database identification but also the de novo sequencing process.

The second level of data analysis is also not standardized among different systems. The principles for peak clustering as well as the criteria for identification of the peaks of difference vary significantly from system to system, such that the evaluation and comparison of results from different laboratories is impossible. A new initiative for standardization of these two processes and a requirement for multicenter validation of obtained results slightly improves this situation, but much must yet be done to decrease confusion and inconsistency in this area.

Proteomics Instrumentation

Two-Dimensional Gel Electrophoresis

The first protein separation method that met omics criteria was two-dimensional polyacrylamide gel electrophoresis (2D PAGE). This technique is able to accept minimally pre-prepared specimens and to separate all proteins present in the mixture. Because of its contribution to proteomics and its popularity among researchers, the name two-dimensional (2D) electrophoresis has become almost synonymous with the term proteomics. Two-dimensional PAGE includes two procedural steps. The first is isoelectric focusing (IEF), which separates proteins based on differences in isoelectric (pI) points. During IEF separation, proteins migrate along a pH gradient until they reach a point where the pH is equal to the pI. At that point, the net charge of the protein becomes zero, and migration stops. At the end of this step, the original mixture is divided into fractions that contain proteins with the same pI. The second step of 2D PAGE separates proteins based on their molecular mass and is performed on a sodium dodecyl sulfate (SDS)-polyacrylamide gel. The final result of these two separation steps is a collection of the fractions that contain proteins with the same pI and the same molecular mass. To visualize separated proteins located in the spots, staining with silver or Coomassie Blue G-250 is performed. In an ideal situation, in which the sensitivity of separation is 100%, each spot would contain one protein. Unfortunately, the sensitivity of 2D PAGE is not perfect, which contributes to several serious methodological problems.

Methodological Obstacles of 2D PAGE

For both shotgun and protein profiling approaches, precise and reproducible placement of spots on the gel is critical. Even minute changes in protocol and/or environment may lead to the misplacement of identical spots, which could be mistakenly evaluated as spots of difference. During IEF, even minimal changes in the pH gradient from one experiment to another deposit the same proteins in different locations on the strip and change their placement on the final gel. This problem was significantly reduced when an immobilized pH gradient was introduced. Despite improvements in the procedure, large hydrophobic proteins failed to be separated by IEF because of their tendency to aggregate and precipitate around their pI. The problem of precipitation also plagues separation of strongly acidic and basic proteins, which occurs on the extreme points of the IEF strip. The second step of the 2D PAGE separation requires a precise placement of IFE strips on the SDS polyacrylamide gel slab because any inconsistency changes the final location of the same protein from run to run. To secure precise and reproducible placement of the IFE strip on the gel slab, new hardware has been introduced that is capable of a consistent transfer of immobilized pH gradient strips onto the SDS-PAGE. If performed on polyacrylamide gel, the second dimension of separation is limited to the proteins with sizes larger than 10 kDa. The smaller proteins, as well as low-abundance proteins with fewer than 1,000 copies per cell, do not separate. Hence, this system is unable to separate membrane proteins because of their large hydrophobic domain, as well as posttranslationally modified proteins because of their low concentration, which is far below the sensitivity of this technique.

Protein Staining

The next step in the 2D PAGE separation is protein visualization. The staining not only reveals individual proteins but also, by its intensity, which is measured by densitometry, provides information about protein concentration. Silver (without glutaraldehyde) and Coomassie Blue G-250 stains are the most commonly used, with a preference for the former because of its significantly higher sensitivity. The drawback of the silver staining is its low sensitivity resulting from the staining of the background. To minimize this effect and to increase the sensitivity for protein detection, glutaraldehyde is used for destaining of the background. There is no endpoint for the silver staining reaction, which means that the
intensity of the color is directly proportional to time of staining. The silver stain procedure is also not compatible with mass spectrometry because of protein oxidation and covalent modification during contact with silver and formaldehyde, as well as because peptide-peptide binding occurs during background decoloration with glutaraldehyde. The Coomassie Brilliant Blue stains proteins by chemical interaction with arginine residues. The staining process is not time dependent but relates to amino acid composition and stops when all arginine amino acids are saturated. Because the relationship between staining intensity and protein concentration is not linear, and changes from run to run are significant, this stain cannot be used for protein semiquantitation. Compared to silver, the Coomassie Brilliant Blue stains possess lower sensitivity but higher compatibility with mass spectrometry. To obtain high sensitivity and compatibility with MS in a single experiment, both staining techniques should be used in tandem.

The sensitivity of protein visualization has been significantly improved through the use of fluorescent labeling agents such as ruthenium II-Tris, SYPRO dyes, and cyanine dyes, to the extent that posttranslationally modified proteins can be revealed as distinct spots. Another advantage of using fluorescent dyes is related to the intensity of the staining, which is not time dependent and produces minimal background staining. Although special protocols and equipment are required, a fluorescent stain can be used for the detection of low-abundance proteins. This method cannot, however, be used for protein quantitation, because the intensity of staining is proportional to the amino acid composition rather than to the protein concentration. To quantify proteins on a gel, a special staining with stable isotopes has been developed. In this procedure, both light and heavy isotopes of hydrogen, nitrogen, and carbon are used for in vitro or in vivo protein labeling. In vitro labeling is used mostly for relative quantification of the differences in protein concentrations between two states of a cellular culture or an organism. The concentration of proteins that are present at different cellular states can be calculated from the difference of signal intensities detected by mass spectrometry between the heavy and the light forms of proteins. Thus far, in vitro labeling has generally been tested on samples of low complexity, such as commercially available proteins or simple systems such as viruses and prokaryotic cells. Despite its promise, this method requires significant improvement before utilization with more complex systems such as eukaryotic organisms.

Enzymatic and Nonenzymatic Digestion

The spots present on 2D PAGE contain groups of similarly sized and charged proteins. To perform protein identification, the spots are excised and digested using enzymatic or nonenzymatic methods. During this process the proteins are cleaved into smaller peptides suitable for mass spectrometry. The enzyme most commonly used for this process is trypsin. Its popularity is related to its low cost and to generation of peptides of a size that is ideal for mass spectrometry analysis. Other proteases such as Glu-C endoprotease are used sparingly because they produce large amounts of autoproteolytic peptides. Cyanogen bromide is the most commonly used nonenzymatic cleaving agent. It is not very popular because it produces peptides that are too large for mass spectrometry and thus require additional digestion with trypsin.

All the enzymes used in preparatory digestion cleave proteins at well-defined sites. If the sequence of the proteins is known, it is possible to deduce the number, sequence, and molecular weight of all peptides produced by a particular enzyme. These data have been stored in databases; they are readily accessible on the Internet and they are used for protein identification. Similar to any other PAGE-based procedure, enzymatic preparatory digestion carries several methodological problems that need to be resolved. For example, trypticization of proteins with small lysine and arginine residues produces large fragments not suitable for mass spectrometric analysis. In addition, tryptic fragmentation of proteins is a very imprecise process that not only skips certain cleavage sites but also targets amino acids other than lysine and arginine, which complicates the fingerprinting method of protein identification.

Mass Spectrometry

Mass spectrometry became a potential diagnostic technology at the time when electrospray ionization (ESI) and matrix-assisted laser desorption/ionization (MALDI) instruments were developed. These tools combine both high sensitivity and easy utility, which makes them ideal for commercialization. The new mass spectrometry technology coupled with automated and semiautomated specimen delivery systems provide the ability to perform high-throughput operations, an important aspect of diagnostic tools. Analyzing proteins in their original complex forms and performing this operation in high-throughput mode fulfills the omics criteria and makes modern mass spectrometry useful for proteomics procedures.

Matrix-Assisted Laser Desorption/Ionization

Originally described in 1987, MALDI is a mass spectrometer in which the specimen is embedded into a matrix in a 1:1000 ratio. The matrix is composed of small molecules that absorb ultraviolet waves, inducing solvent vaporization and matrix crystallization. During crystallization, the specimen is incorporated into matrix crystals and subsequently vaporized. The vaporized clouds of ionized crystals containing the protein particles are directed into a detector, most commonly the time-of-flight (TOF) mass analyzer. The MALDI-TOF is capable of analyzing thousands of different components (not exclusively proteins), but its role in proteomics seems to be the most significant. It is the first instrument that competently
detects entire proteins and measures their molecular masses with a precision that allows for fingerprinting identification, even when the proteins are present in the mixture at very low concentrations. 28

Two modifications of the MALDI mass spectrometer allow for protein identification based on the amino acid sequence: MALDI-Oq-TOF (Oq indicates tandem quadrupole/time-of-flight mass spectrometry) and MALDI-TOF/TOF. 29 These instruments operate with a collision chamber located between the first and second mass analyzers. The first analyzer measures molecular masses of peptides present in the primary mixture, and the second analyzer measures the molecular masses of the fragments derived from the peptide that was chosen for sequencing. This chosen peptide is directed from the first MS to the collision chamber for fragmentation. The molecular masses of these fragments are measured in the second MS, and they are the base for the sequencing process.

**Imaging Mass Spectrometry**

A new feature of the mass spectrometry development has been attained with an innovative MALDI technology called imaging mass spectrometry (IMS). In this technology, the mass spectrometer produces a two-dimensional ionic image of the sample in which the localization of the targeted molecules is tagged. 30 Two techniques are used for obtaining an image. The first one is a blotting technique that requires fresh tissue to be placed on a methanol-wetted organic membrane for protein binding. After the tissue is removed, the blotted area is submitted for MALDI-TOF mass spectrometry analysis. The second technique calls for frozen tissue to be placed on a flat metal MALDI target plate and covered by the matrix. The image is constructed by mapping the m/z value over the target area. A unique feature of IMS technology is that it can reveal a distribution of hundreds of unknown compounds in a single measurement. Conventional methods for visualizing molecules such as chemical staining, immunohistochemistry, and radiolabeling require prior knowledge about targeting molecules and preparation of specific molecular tags. This requirement significantly limits the utility of these methods. MS/MS imaging offers an analysis of the tryptic fragments that are obtained after on-slide enzymatic digestion. The potential benefits promised by this method may be not only an identification of the proteins directly from the slide but also a localization of high molecular weight, membrane-bound or membrane-associated proteins that are difficult to detect with direct MALDI-MS analysis. 31

In addition to obtaining information about proteins distribution directly from the tissue, IMS allows for measuring their relative concentration and local composition. This technique has been successfully used for protein visualization in glioblastoma of the brain and in non-small cell carcinoma of the lung. In clinical practice, it can be used to identify different populations of cells based on their molecular differentiation, which can be used to clear surgical margins at the molecular level and for discovering of a new marker for ovarian cancer. 30, 32

Although this technology seems extremely attractive, several serious limitations must be overcome before its diagnostic implementation. The size of the analyzing specks is limited by the laser spot, which is 30 μm in diameter. It means that MALDI-IMS produces only about 1,000 discrete images from the cells that contain tens of thousands of proteins, 31 which translates into instrumental sensitivity significantly too low to produce any meaningful results. In addition to low sensitivity, the overshadowing effect limits the usage of this technique in anatomic pathology. In the overshadowing effect, a low signal is suppressed by a higher one, which most often is related to hemoglobin that is present in high concentration on the cut surface. An additional difficulty is formed by an abundance of the produced data. The doubling of the resolution increases the amount of data fourfold with concomitant increases in acquisition time and data file size. To overcome these difficulties, significant improvement in the process of data analysis and storage is required. 31 In addition to the outlined specific limitations, every limitation of the MALDI is also pertinent in this system. 33

**Surface-Enhanced Laser Desorption/Ionization Mass Spectrometry**

SELDI mass spectrometry is another spectrometric system that has gained interest for its potential diagnostic usefulness. This system uses microchip technology and is promoted by the Bio-Rad Diagnostic Company. A minimally prepared or unprepared specimen containing a mixture of proteins is applied to an aluminum or steel microchip support platform measuring 1–2 mm in diameter. The surface of the miniplate is coated with chemical matrices (“bites”) similar to chromatographic columns and composed of hydrophobic, cationic, or anionic phases. With these types of bites, proteins are separated based on their physicochemical properties such as the isoelectric point (pI). The mixture of proteins and peptides present in the specimen is applied on the microchip, and groups of proteins with similar physicochemical properties bind to the surface. The unbound proteins are removed during the washing process. The coating matrices of the chips can be more specific, such as antibodies, enzymes, ligands, proteins, receptors, or DNA oligonucleotides. These specifically reacting chips are designed to interact with a single target protein and are used to study specific molecular interactions such as protein-protein, protein-antibody, protein-receptor, or protein-DNA. 34 The complexes of “bite” and proteins of interest are subsequently analyzed by MALDI-TOF or MALDI-TOF/TOF mass spectrometry. Microchips are suitable for analyzing all possible specimens, including serum, whole blood, urine, cellular lysate, body fluids, and lavages. To enhance the efficiency of the result analysis, the instrument is equipped with a biomarker discovery software packet. This software packet supports recognition of the spectrometric peaks and combines them into clusters that
are characteristic for a particular group. It also extracts the peaks that are characteristic for healthy and diseased cells or for different stages of disease or different states of normal organisms (peaks of difference).

The entire process of the spectrometric spectra evaluation provides a basis for discovering single biomarkers and for a diagnostic strategy called proteomics pattern diagnostics. The SELDI mass spectrometer along with provided software was tested during development and validation of the spectral patterns as a diagnostic marker for ovarian, prostate, and breast cancer. As a result, a large amount of very complex data was obtained that could not be evaluated by classical statistics. To validate whether spectrometric peak patterns can be utilized in cancer diagnosis, an artificial neuronal network algorithm was used, integrated with the SELDI software packet. An artificial neural network is a statistical tool that imitates the learning ability of the human brain. PROPeak software provided by Bio-Rad is based on an artificial neural network and is able to “learn” to recognize universal elements based on repeated analysis of the complex data from at least two experimental groups. Based on the “experience” gained during the training set, the instrument develops criteria and rules that are validated during subsequent test runs. Each run is used to validate and modify the rules, similar to the typical learning process.

The usage of the pattern of mass spectrometric peaks as a marker for disease is a completely new approach in laboratory diagnostics and carries several methodological and technical issues that need to be addressed. The most serious ones are related to the low sensitivity of SELDI instrumentation, its limited reproducibility, and difficulties in protein identification. In addition, several questions have been raised that are not specifically related to the SELDI instrumentation, such as the validation of results among different laboratories, the influence of different sample handling on results, and different experimental designs.

Other Types of Spectrometry

The other type of mass spectrometry used in proteomics is ESI mass spectrometry. This MS operates on soft ionization, which allows the generation of ions from large (>100 kDa) nonvolatile analytes, such as proteins, with no fragmentation. The main difference between MALDI and ESI is that the latter produces multicharged ions and consequently more complex spectra. ESI collects more information; however, its results seem to be more difficult to interpret than the less-complex spectra from MALDI. At the same time, MALDI possesses lower resolution and limited sequence information, which makes it less useful for identification of protein isoforms, posttranslational modifications, and protein complexes. In the recent years, ESI has become a more user-friendly instrument and is more widely used than MALDI in proteomics investigations.

Quadrupole mass spectrometers, which are also named quadrupole mass filters, are another commonly used MS. The wide popularity is accounted for by their flexibility, easy operation, adequate performance for most applications, relatively low cost, and well-developed software. The main functional element consists of four parallel electrically conductive rods arranged into a channel through which the ions from the sample pass. While passing through this channel the mixture of ions is sorted so that only the ion with a selected, very narrow m/z range is able to reach the detector. Because only chosen ions are detected by this system, the quadrupole mass spectrometer is often referred to as the “mass filter.” The instrument operates in two basic modes: selective and scanning. The selective mode is used to detect only one ion with precisely defined m/z. In the scanning mode, a wider range of m/z is chosen in which ions across a full range of m/z are detected.

Quadrupole mass spectrometry is predominantly used in target-driven proteomics where a single ion is targeted for identification, as well as for multiple reactions monitoring MS-based biomarker validation.

A “Dream Team” Mass Spectrometry System

The technological challenge for today’s proteomics is to develop a system that will allow rapid fractionation of the sample and efficient identification of proteins in the mixture in high-throughput mode. In 2004, Schrader and Klein published a paper describing the ideal proteomics system. This scheme consists of a high performance liquid chromatography-based separation system coupled with Fourier transform ion cyclotron resonance (LC FTICR) mass spectrometry. The Schrader system helps to overcome the limited sensitivity of gel-based separation and limited accuracy of mass spectrometry. High performance liquid chromatography (HPLC) is able to fractionate an entire specimen in a liquid environment into fractions containing proteins with precisely defined m/z. The FTICR detects ions in the mixture by measuring their frequency of resonance. Because frequency of resonance changes with mass, a precise measurement of frequency allows for mass measurement with high resolution and accuracy. It seems to be the solution for problems related to the currently used, less-accurate mass spectrometers. The FTICR is also able to analyze unfractinated samples. It can provide fast fractionation of a complex protein mixture, and it is able to perform mass measurement with precision, allowing for library-based protein identification using very narrow mass tolerances.

At the same time, an efficient fragmentation method for FTICR has been developed: infrared multiphoton dissociation (IRMPD). IRMPD produces peptide fragments similar to those obtained in the collision chamber in a tandem mass spectrometry system. The analysis of fragments is similar to that obtained by a tandem mass spectrometry system, but with significantly higher mass accuracy and efficiency.

The unmatched advantages of this system are high resolution, large dynamic range, and supreme sensitivity, which provide the highest quality data for biological analyses, especially in proteomics. It can be used for both the
“bottom-up” method that identifies proteins based on accurate mass measurements of peptides derived from the digested mixture as well as the “top-down” approach which is able to identify intact proteins. Since this technology has been used, especially with the improvement of dissociation techniques, significant progress in providing sequencing and structural information about proteins is noted. Recently, a few papers describing an attempt to couple a desorption electrospray ionization (DESI) technique with a Fourier transform ion cyclotron resonance mass spectrometer have been published. This system combines the unique ability of DESI to ionize a natural and unprepared sample with the high sensitivity of FTICR. In this system, unfracti onated samples are directed by DESI to the FTICR for top-down analysis and sequencing of the peptides present in the samples. If successful, the presence of such a system may open new opportunities for sample processing and protein identification, without prepreparation and in high-throughput output mode.

Proteomics Protocols

Protein Identification

Protein Mass Fingerprinting Identification

The commonly used detecting systems that provide information allowing for protein identification are based on single (MS) or tandem (MS/MS) mass spectrometry. The specimens for these systems consist of prefractionated mixtures of proteins that are enzymatically digested or chemically cleaved into derived peptides. To recognize proteins present in the sample, the molecular masses of the derived peptides are measured with the highest possible precision, and they become a basis for protein identification by the mass fingerprinting procedure.

Protein mass fingerprinting identification is the first and the most often utilized identification method by proteomics investigators. This method is based on the fact that proteolytic enzymes cleave proteins at precisely defined points such that for any known protein the number of derived peptides, their amino acids sequences, and molecular masses can be predicted based on amino acid composition and the characteristics of enzymatic action.

The information about proteins and theoretical sets of peptides derived from them after enzymatic digestion is stored in a database. To identify proteins present in the specimen, molecular masses of experimental peptides are compared against molecular masses present in the database. If they match, the peptide and its “mother” protein are identified. In real life, this perfect picture of protein identification is complicated by several factors. The most important of these are the mass spectrometer’s instrumental sensitivity (a degree of precision for molecular mass measurement) and the completeness of the database for these particular proteins. A higher sensitivity and a more detailed database give a better chance for the protein to be identified. To increase the chance for correct protein recognition, it is crucial to measure the mass of the ions with the highest possible accuracy. The number of peptides that differ from each other in the mass range of 0.5 Da is four times higher than the number of different peptides in the mass range higher than 0.5 Da. For the mass spectrometer with a mass resolution higher than 0.5 Da, the probability that dissimilar peptides are differentiated is higher than if the resolution is lower than 0.5 Da. The instrumental sensitivity is even more important for the detection of posttranslational modification or for the uncovering of a single element from a complex eukaryotic proteome.

Another difficulty confront by fingerprinting identification is related to the sizes of the peptide fragments. The general rule for peptide matching is that the shorter fragments are less specific but have a higher chance to be present in the experimental mixture and to be found in the database. On the other hand, the longer peptides are more specific for the “mother” protein, but they are present in the mixture in lower concentrations. To obtain a result of the database search with higher specificity, longer peptides are more desirable. Unfortunately, large peptides not only are present in low concentration, but they also do not ionize as well as small ones do during the spectrometric procedure. Thus, their chances to be detected are diminished. For this reason, the optimal results for protein identification are obtained when a larger number of small peptides containing 8–10 amino acids is detected and submitted for evaluation. The peptides with these characteristics are produced by trypsin, the proteolytic enzyme that is widely used during preparatory digestion. Another problem related to size is that small proteins have a lower chance of being identified by the fingerprinting method than do large ones; this occurs because the number of digested fragments produced from small proteins is low, and consequently the number of peptides available for comparison against the database is small, which decreases the probability that the mother protein will be identified.

Some of the proteins present in the primary mixture may be lost during preparation and/or fractionation. As it was mentioned before, two-dimensional PAGE is still widely used for sample prefractionation. During this process, large proteins with a molecular mass higher than 100 kDa are missed due to precipitation during the first IEF separation, and subsequently they are not identified.

It is important to remember that each mass spectrometer has its own characteristics with variabilities. For that reason, the final product of protein identification can be distorted by the type and internal properties of the particular instrument. A case in point is MALDI-based protein identification. As a general rule for successful protein identification, only the most intense peaks should be analyzed. However, peak intensity is not always related to peptide abundance. An example is MALDI-TOF, which creates peaks whose intensity depends more on arginine content than on
peptide concentration. This result indicates that arginine-rich proteins have a greater chance of being identified by MALDI-TOF spectrometry than proteins with low arginine content.

A highly advanced tool for fingerprint identification is “iterative analysis,” which analyzes in depth the 2D PAGE protein distribution and protein composition within a single two-dimensional spot. The first step for interactive analysis is the removal of masses of contaminants such as keratins, products of trypsin autolysis, and dyes. The second step of this analysis consists of removing the peaks caused by neighboring spots. The final results of protein identification are reevaluated for the major components and contaminants.\(^{35}\)

**Sequence-Based Identification**

It is possible that, in a digested protein mixture, especially one derived from eukaryotic organisms, there are a number of peptides with the same molecular mass but with different amino acid compositions. It is also possible that the same mixture contains peptides with slightly different masses that are not discriminated by the mass spectrometer. For these cases, molecular mass alone cannot be the basis for protein identification, and to increase scoring and certainty for protein recognition, amino acid sequencing is added to the protocol. Peptide sequencing is performed via tandem mass spectrometry, in which the mass spectrometers are separated by a collision chamber. When the mixture of ions reaches the first mass spectrometer, the molecular masses are measured and one of them is chosen for future sequencing. This selected ion is subsequently directed to the collision chamber, where it is fragmented into a unique spectrum of derived ions whose masses are measured in the second mass spectrometer. As a result, a series of increasing m/z is obtained, progressing from a single C amino acid to the entire peptide, where each subsequent ion differs from the previous one by the mass of one amino acid.\(^{46}\) The key elements for this sequencing process are precise measurements of the molecular masses produced in the collision chamber, and their arrangement from lowest to highest one allows for construction of amino acid sequencing. For protein recognition, the spectra and sequenced fragments are compared with a comprehensive protein sequence database by using one of the identification algorithms.

The simplest algorithm starts with a match of the experimental ion mass with the masses of peptides present in the database. Usually, a wide window of mass tolerance is chosen, and a broad range of peptides is selected. In the next step, a theoretical sequence for each peptide candidate is created, which is then compared with the sequence of the experimental one. The identification process by collision-induced spectra is very efficient, because sequence analysis brings supportive information to the identification process performed by the fingerprinting method.

**Database Searching**

Several available databases allow for comprehensive searches for molecular masses and for sequence matching. No matter which database is used, the peaks produced by internal standards and the peaks of autolytic peptides must be excluded from the search. The search process and data interpretation require precise information about search parameters, such as specific instrumentation and fragmentation or cleavage methods. There are a few search engines that support multiple fragmentations. For comprehensive fingerprinting and sequence tag protein identification, the RefSeq, NCBI Entrez Protein, UniProt, TrEMBL, and Swiss-Prot databases are most widely used. The Swiss-Prot is not so complete as the others, but it contains several tools for sequence analysis.

The NCBI is an example of a sequence repository database that contains sequence data translated from the nucleotide sequences of the DDBJ/EMBL/GenBank. The NCBI is constantly updated by sequence information from other protein database such as Swiss-Prot, Protein Information Resource (PIR), and Protein Data Bank (PDB).\(^{47}\) To be efficient, the search process requires some specific information to be entered, such as keratin contamination, the proteolytic enzyme used, the number of allowed missed cleavages, the usage of iodoacetamide for cysteine modification, and the fixative modification in cysteine residues. This information narrows the spectrum of the search and reduces searching time.\(^{48}\)

The RefSeq collection’s main feature includes explicitly linked nucleotide and protein sequences, which are automatically updated to reflect current knowledge of sequence data, data validation, and review status in each record.\(^{47}\)

The UniProt database was developed with the aim to be the most comprehensive protein database. It merges other collections such as Swiss-Prot, TrEMBL, and PIR-PSD. All suitable PIR-PSD sequences that are missing from Swiss-Prot or TrEMBL were incorporated into UniProt. To allow fast tracking of any PIR-PSD entries, cross references between Swiss-Prot, TrEMBL and PIRPSD have been created. This comprehensive UniProt database is an initiative that provides a foundation for uniform information flow among researchers.\(^{47}\)

For database searching, there are several commercial and publicly available engines, but not all provide the same quality of search, and only a few cover the fragmentation method that is used by a particular instrument. Commercial search engines such as MASCOT and Phenyx support most instrumentation with its fragmentation methods. MASCOT provides a complex search report expressed by numerical scores. A score of 100 identifies a protein, but any score between 70 and 100 indicates that a chosen protein is a good candidate for identification. MASCOT also provides a list of unassigned masses. The most common reasons for unmatched peaks are tryptophan oxidation, methylation...
of aspartic and glutamic acids, and (not as often, but most excitingly) a new protein.

Publicly available search engines such as VEMS, OMSSA, and S!Tandem return results of a similar quality as those attained by MASCOT or Phenyx. 47

There are several reasons for obtaining unassigned peaks. If investigation proves that the presence of unassigned peaks is the result of the presence of an unknown or posttranslationally modified protein, it must be sequenced by tandem mass spectrometry or retested with one of the modification targeted methods such as multiple reaction monitoring (MRM).

Validation of Protein Match

To validate matching accuracy for each candidate protein, various scoring methods are available. In the simplest one, the peak intensities of the experimental peptides are compared with those of the theoretical group. The most serious bias for this method is that the intensity of peaks depends on the collision energy. Any inconsistency in collision energy significantly changes the matching results. The most popular scoring systems using signal intensity produced by tandem mass spectrometry is MASCOT. More advanced scoring methods calculate the matching score based on intensity of the collision energy and internal properties of the peptides. 49 The popularity of these analytical methods is very low, and the simpler MASCOT and SEQUEST systems still dominate the market. 50

The sequence-based identification process may be obscured if the reference database for a particular organism is not available; if the identified protein underwent posttranslational modification not listed in the database; or if spectrometric peak intensity is doubtful because of a low signal-to-noise ratio. These situations can be overcome by using a de novo sequencing method. There are three basic approaches to de novo sequencing method: the exhaustive listing method, the subsequencing method, and the “graph theoretical” approach. Detailed descriptions of these methods are available elsewhere. 51 A simpler version of de novo sequencing identification is the partial sequence method. It is based on the mass sequence-mass protocol, which matches a molecular mass and establishes a sequence for an experimental peptide with a theoretical partial sequence and the masses of peptide candidates. 50 The other group of methods is based on the calculating of theoretical mass from the peptide sequences in the database and comparing the results with observed peaks.

The method that has recently gained popularity is a calculation of the false discovery rate (FDR) for reverse sequences obtained from a real database. The popularity of the FDR scheme is related to its simplicity and to the fact that it can be used with any search engine and scoring system. It is a method that offers a practical way to establish cutoff values for searching scores, and is especially valuable for validation of assignments for complex posttranslational modifications. 52

Ascore is another computational scoring system that works in most instances except in cases of ambiguous phosphorylation sites, when the potential score error is too high, so that manual validation is still preferred by most researchers. 53

In the instances when phosphorylated peptides are validated, the VEMS program may be used. It offers the option to generate all possible sites of phosphorylation for a particular peptide and rank them using the VEMS score.

Although a large number of methods is used for validation of sequence-based matching, it is difficult to assess which one is the best because there is no consistent evaluation protocol that can be used to compare these methods. 31 As of today, there is no method that would satisfy all possible situations, and it seems obvious that the comprehensive identification and validation approach is the best choice. It would include fingerprinting and sequencing identification with computing validation followed by manual verification. This comprehensive tactic increases the chance for correct identification not only of common cases but also of proteins with extremely low and extremely high molecular masses, posttranslationally modified as well as unknown ones.

Protein Quantification

For higher organisms, the old concept of “one gene, one protein” no longer holds for the obvious reason that a single gene is the source of a large number of protein isoforms that subsequently undergo extensive posttranslational modifications. The different isoforms of the same protein play important regulatory roles and may have the opposite effect on cellular function. The idea that one protein participates in cellular metabolism as a single factor has been revised and changed to the theory that a multiple protein network governs cellular activities. The function of these complexes is regulated not only by the actions of different isoforms but also by different concentrations of the participating proteins. Although the importance of different protein concentrations for cellular metabolism is well known, to this day there is no good method that measures absolute concentrations of proteins or their isoforms. The only well-established method uses a stable isotope and measures a relative concentration of a protein in prokaryotic cells. This strategy is based on the ability of mass spectrometry to separate chemically identical proteins that differ by a stable isotope. In recent years, some hope has been brought to this field by the development of a new technology that is based on the multiphoton detection methodology. The sensitivity of this method allows not only for detection of as few as 600 molecules per liter but also for separation of different protein isoforms. 54

The general methods for protein quantitation can be classified into densitometry that is used for 2D electrophoresis, relative quantitation by chemical and metabolic isotopic labeling, and spectral intensity profiling.
On 2D PAGE the protein concentration on stained spots can be quantified by measuring the absorbance of each gel fraction by a densitometer. The currently available densitometers are equipped with a microprocessor that calculates the area under each peak of absorbance, which represents a protein spot, and reports the peak as a percent of the total or as an absolute concentration. Reliable densitometric quantitation requires an appropriate light wavelength, a linear relation between absorbance and concentration, and a transparent background of the protein-free gel.

Quantitation techniques by stable isotope labeling are based on comparing the measured intensity of the mass of the unlabeled “light peptide” and the intensity of the mass of the labeled “heavy peptides.” A more advanced form of the same method is called “multiple encoding,” and it uses multiple stable isotopes that produce multiple “heavy peptides.” This method allows for the comparison of more than two samples in the same MS experiment. Although different quantitative methods are based on the same principles, the calculating algorithm and interpretation of the results are different and specialized. ASAPratio, RelEx, MSquant, and VEMS are examples of the tools used for these calculations.14

In the spectral intensity profiling method, the intensities of spectral peaks received from different LC-MS are compared. Because of the low reproducibility of LC, the bias must be corrected by duplicate measurement and calculation of the experimental variances55 or by adding an internal standard to the mixture.

There is no single method that can be used for simultaneous protein identification and quantification. The most widely used tactic is to identify and quantify individual components separately and subsequently integrate these results using analytical tools. The 2D PAGE is an example for this approach, where concentration of proteins is assessed by measuring stain intensity of the separated spots. For protein identification, the selected spot is excised, digested, and submitted for mass spectrometry testing. The gel-based approach carries several limitations described earlier, of which the most important are inability to detect low-abundance proteins, to recognize posttranslational modification, and to separate protein complexes.51

The non-gel approach for protein identification and quantification is a mass spectrometry-based protocol in which an entire protein mixture is submitted for analysis. Although more effective than the gel-based approach, mass spectrometry protocols are still imperfect. For example, the analytical sensitivity of single-dimension mass spectrometry is too low to detect low-abundance proteins. In addition, the relationship between peak intensity and the amount of the tested substances in the mixture is too poor to be used for assessing their concentration. To overcome these problems, an internal standard must be used for a more precise assessment of the protein concentrations, and a “dream team” mass spectrometry system ought to be employed, such as HPLC coupled with LC-FTICR for surmounting low instrumental sensitivity.

Detection of Posttranslational Modification

Posttranslational modification is a reversible event serving as a regulatory process that modifies protein activity. During this process, proteins become mature and ready for their physiological function. It is possible that the only cellular reaction to diseases or environmental modification is an alteration in the concentration of modified proteins without simultaneous changes in protein expression.

There is no single and efficient method to determine posttranslational modification of proteins in a biological sample or in any protein mixture. The most popular method of posttranslational modification detection is the use of different enzymes interacting with posttranslationally modified proteins/peptides and subsequent analysis of the differences in molecular masses. Proteins undergo more than 300 different posttranslational modifications, of which the most commonly observed are phosphorylation, splicing, and disulfide bridges.

Phosphorylation is an important regulatory mechanism and occurs at the serine, threonine, and tyrosine residues. During this modification, 80 Da is added to the molecular mass of the peptide, which can be detected by comparative mass measurement before and after treatment with alkaline phosphatase. However, this method is complicated by a low consistency in peptide ionization and by unpredictable protein fragmentation.56 The nonenzymatic approach for detection of phosphorylation and other posttranslational modifications is based on a specific database search aimed at the evaluation of modified and unmodified peptides.57 The main complication of this analysis is technical difficulty of detecting phosphoproteins by mass spectrometry in the presence of nonphosphorylated species.

Splicing is a regulatory mechanism for gene expression present in higher organisms that leads to the production of multiple proteins from a single transcript. Splicing can be detected by the amino acid sequencing as a complementary identification method to the fingerprinting protocol. Disulfide bridging is formed between cysteine residues and can be produced in vivo or during sample preparation. It can occur within the same protein, producing a change in the tertiary structure, or between molecules with the formation of multimers. This modification changes the molecular mass and subsequently the behavior during electrophoretic separation. Formation of multimers is suspected if the same amino acid sequence is detected in several bands or spots on the electrophoretic gel and can be confirmed by the presence of a molecular mass two or more times higher than that of a basic peptide.

Several methods are alternatively used for the detection of posttranslational modification. The immunoprecipitation method is most commonly used to detect phosphoproteins, and it uses antibodies against phosphorylated amino acids. The precipitated proteins are subsequently separated via
gol electrophoresis and analyzed by MS. Another method is chemical derivatization with alkali solution, which eliminates \( p \)-serine and \( p \)-threonine, and subsequent tagging with biotin. The phosphoproteins after being tagged with biotin are separated and identified by affinity chromatography.\(^{58,59}\) Other methods are immobilized metal affinity chromatography and affinity purification, both of which use the high affinity of phosphoproteins to positively charged metal ions such as iron, aluminum, or zirconium. Because of this high affinity, the method is biased toward multiply phosphorylated proteins, and also it does not give quantitative information.\(^{58}\)

Modification of proteins not only occurs in vivo as a regulatory process of protein activity but can also be artificially produced during laboratory procedures. For example, during electrophoresis, proteins react with nonpolymerized acrylamide monomers and produce unusually modified amino acids such as alkalysated cysteine or propionamide. Similarly, during the staining procedure in the presence of acetic acid and methanol, asparic and glutamic acids are methylated.\(^{60}\) For practical reasons it is important to be aware of these artificial modifications and not misinterpret them as naturally occurring.

The methodology recently implemented into proteomics, multiple reaction monitoring (MRM), gives a significant boost for mass spectrometry in its role to detect and monitor the post-translational status of proteins. The MRM experiment requires mass spectrometry that runs at high sensitivity to be able to detect targeted components. This method has been developed and used by the pharmaceutical industry to detect and quantify drugs and their metabolites. The potential value for the proteomics investigation was shown by Cox et al.\(^ {61}\) who identified a low level of phosphorylated MEF2A by using a highly sensitive MRM experiment and a hybrid quadrupole linear ion-trap instrument for identification of targeted ions. The phosphorylated targets were deduced based on specific information for MEF2A, which included the primary sequence, the type of phosphorylation (serine/threonine vs. tyrosine), and the acceptor sites for phosphorus (consensus peptide that is targeted by a kinase). The principle of MRM can be applied to studying other types of posttranslational modifications that occur on proteins with known molecular masses and deduced derived peptides. This information about the protein molecular masses and derived peptides can be used to design an MRM experiment specifically for the identification of potential sites of posttranslational modifications that have become a target for mass spectrometry.

**Protein Complexes and Protein Interactions**

In a physiological environment, proteins do not exist as single entities but rather as a complex system responsible for cellular function and regulation. The methods of studying protein complexes must be able to evaluate not one but multiple elements, as one-on-one interaction between proteins is more an exception than a rule. Proteins interact not only with other proteins but also with nucleic acids, small molecules, and other components, such as drugs. The classic low-throughput system for studying interactions between proteins or between a protein and other molecules is a design in which one component is used as bait for the isolation of its binding partner. The most serious obstacle to this system is its low reproducibility, even if the same bait is used for protein binding.\(^ {62}\)

The protein–protein interaction depends on variations in protein affinity, which is regulated by changes in the cytoplasmic environment. Because the environment of mass spectrometry is completely different from that in the cytoplasm, mass spectrometry is not a good method for studying protein-protein interactions. In recent years, new and more advanced methods for protein complex study have been developed. One is the “complex walking” strategy, in which the complex is studied sequentially. The other method uses stable isotopes for labeling the different components of the protein complexes. This method allows for the identification of dynamic changes in protein composition relative to different cellular states.

There are also several MS-based methods that use modified tagged proteins as a bait, forming physiological complexes with other proteins. These complexes are created in vivo and are subsequently pulled out. Extracted proteins with the tagged bait are identified using the MALDI MS. This method has been used on a large scale for the characterization of protein complexes in yeasts and has been named TAP (tandem affinity purification). The high-throughput modification of this method is called the high-throughput mass spectrometric protein complex identification (HMS-PCI).\(^ {63}\)

Another method of mapping protein-protein interaction is with microchip technology. This method uses a large number of purified proteins that are attached to a microchip surface at known locations and form baits for protein complexes. These bait proteins are preserved in their folded conformation so that they can interact with other proteins. The specimen that contains the mixture of proteins to be tested is incubated with a chip. If specific interaction between bait proteins and protein(s) from the solution occurs, it is detected and characterized by the position on the chip. A substantial advantage of microchip technology over the conventional methods is its high sensitivity and small amount of sample needed for a run. In addition, short turnaround time for a single sample and a large amount of samples that can be tested simultaneously, making this technology attractive, especially in the perspective of diagnostic usage.\(^ {64}\)

All these high-throughput protein-protein interaction experiments produce large amounts of data, which become the foundation for new biological discoveries. However, simultaneously a great challenge to bioinformatics is created to manage, analyze, and model these data.


14
Immuno-Histochemistry of Biomarkers
Patrick L. Fitzgibbons and Kumarasen Cooper

Introduction

Immunohistochemistry (IHC) has an essential role in the diagnostic evaluation of cancers, most commonly to help identify and subclassify tumors, but its utility in assessing biomarkers that are predictive of benefit or lack of benefit from specific chemotherapies is becoming increasingly important. When IHC is used for predictive marker testing, the assessment of the extent of protein expression may be even more important than the presence or absence of expression, but many variables affect the quantitative measurement of protein expression in routinely processed tissues, and these can have profound influences on the test results.

Immunohistochemistry (IHC) has been a critical tool for pathologists for more than 20 years and is still used most frequently to determine cell lineage and tumor type. Because of its far greater specificity, IHC has essentially replaced most traditional histochemical stains once used for this purpose. IHC is also used to identify specific cellular constituents (e.g., basal and myoepithelial cells), which may help to determine the presence of malignancy and/or invasion and to diagnose infectious diseases, but its role in predictive marker testing to select or exclude patients for specific therapies is growing rapidly in frequency and importance.

Until recently, only a few predictive markers were routinely assessed in all patients, but an increasing number are becoming a regular part of clinical management. Such tissue-based biomarkers, many assessed solely by IHC, may predict responsiveness (or lack of response) to specific drugs or to entire classes of chemotherapeutic agents; however, along with the discovery of new biomarkers and their clinical significance, we are increasingly recognizing problems related to variability in laboratory assessment.

Role of Immunohistochemistry in Biomarker Testing

For some markers, such as steroid hormone receptors in breast cancer, IHC is the standard method of identifying patients likely to respond to specific drugs. Patients with tumors that are estrogen receptor positive [ER(+) or ER(+)] (about 70% of breast cancers) are more likely to respond to selective estrogen receptor modulators (e.g., tamoxifen) or aromatase inhibitors than those with ER(−) tumors. Indeed, ER expression as assessed by IHC has been shown to be the single most important biomarker in breast cancer: tamoxifen treatment of ER(+) patients reduces annual breast cancer death rates by nearly one-third.1 Similarly, responsiveness to the drug trastuzumab is much more likely in patients with tumors that over-express HER2 protein (or have amplification of the HER2 gene). Other examples of IHC-based identification of potential therapeutic responsiveness include CD20 expression in B-cell non-Hodgkin lymphoma (rituximab), CD117 staining of gastrointestinal stromal tumors (imatinib mesylate), and topoisomerase II alpha in breast cancer (anthracyclines).

More recently, detection of the excision repair cross-complementation group 1 (ERCC1) protein by IHC has been shown to be useful as both a prognostic indicator and a predictor of response to chemotherapy in patients with non-small cell lung cancer. Patients with ERCC1(+) tumors survive longer than those whose tumors do not express ERCC1. More importantly, patients with ERCC1(−) tumors may benefit from adjuvant cisplatin whereas those with ERCC1(+) tumors appear resistant to platinum-based chemotherapy.2

For some tumors, there is little correlation between over-expression of a specific protein and responsiveness to agents...
targeting that same molecule. Cetuximab is a monoclonal antibody that has a high affinity for the epidermal growth factor receptor (EGFR). The drug was initially shown to be effective in some patients with EGFR(+) colorectal carcinoma, but similar response rates were subsequently described in patients whose tumors were negative for EGFR by IHC. Similarly, response to the EGFR tyrosine kinase inhibitors gefitinib and erlotinib in non-small cell lung cancer is related to the presence of activating mutations in the EGFR receptor, but not to the presence of EGFR positivity by IHC. In such cases, assessment of the EGFR status of the tumor by IHC seems to have little or no value in terms of predicting response to therapy.

In addition to its utility in identifying potential therapeutic response, IHC can sometimes be used as a surrogate test to identify underlying molecular abnormalities. As previously noted, IHC can reveal overexpression of cell-surface or cytoplasmic proteins because of gene amplification, for instance, HER2 and TOP2A in breast cancer, but can also provide evidence of specific translocations and mutations, which may be of great value for diagnosis and patient management. For example, follicular lymphoma is classically associated with a t(14;18)(q32;q21) translocation that brings the BCL2 gene adjacent to the immunoglobulin heavy chain gene. This change results in aberrant overexpression of functionally active bcl-2 protein, which can be detected by IHC and used diagnostically.

Soft tissue tumors are also commonly associated with specific translocations that result in proteins detectable by IHC. Desmoplastic small round cell tumor, for example, has a specific chromosomal translocation, t(11;22)(p13;q12), that fuses EWS on chromosome 22 to WT1 on chromosome 11; this results in a EWS/WT1 fusion protein that can be detected by immunostaining for WT-1 (usually with antibodies to the C-terminal of WT-1). Similarly, about 90% of primitive neuroectodermal tumor/Ewing’s sarcomas have a t(11;22)(q24;q12) translocation that results in a EWS/FLI-1 fusion gene and overexpression of the FLI-1 protein.

Immunohistochemistry can also be used to identify abnormal protein expression resulting from gene mutations. Mutations in the tumor suppressor gene p53 are one of the most common genetic changes found in human cancers. Although the protein produced by wild-type p53 has a very short half-life and is undetectable by IHC, p53 gene mutation results in an inactive but stable protein that is easily detected by IHC stains. Similarly, mutations in the adenomatous polyposis coli (APC) gene result in abnormal intranuclear accumulation of β-catenin protein, and IHC stains for β-catenin are very useful in diagnosing tumors such as mesenteric fibromatosis.

The opposite staining pattern is seen with other markers. Hereditary nonpolyposis colon carcinoma (HNPCC) most often arises because of a germline mutation in one or more genes for DNA mismatch repair, principally MLH1 and MSH2. Although the proteins produced by these genes are detectable by IHC in normal colonic epithelial cells, mutation in tumor cells results in loss of functionality of the corresponding proteins and, in most cases, absence of staining by IHC. As such, IHC can be used as a screening tool to identify patients likely to have mutations and for whom genetic sequencing may be useful.

Identification of abnormal mismatch repair status in colorectal carcinoma provides useful prognostic information and may be of predictive value in identifying patients who are resistant to 5-fluorouracil-based chemotherapy. Although IHC is a sensitive method for identifying defects in mismatch repair proteins, the results are not specific for HNPCC. About 15% of sporadic colorectal cancers also have microsatellite instability, usually because of hypermethylation of the MLH1 promoter and silencing of the gene, and such tumors exhibit loss of expression of the corresponding protein similar to that seen in HNPCC.

Importance of Method Standardization in Biomarker Testing

IHC stains that are used to determine cell lineage are almost always done as part of an overall diagnostic evaluation and interpretation, and the data obtained are generally meaningful only in conjunction with the clinical and histological features. The significance of any one lineage stain depends on the context in which it is assessed, and this must be determined by the pathologist evaluating the entire case. Because such immunostains are usually done in panels, minor staining differences caused by variations in antigenicity, fixation, or laboratory methods may have little consequence. In this setting, the presence or absence of staining is usually much more important than the extent or intensity of staining.

Predictive biomarker testing, by contrast, has several important differences. Although the significance of a lineage stain cannot be separated from the overall diagnostic evaluation, a predictive marker stain can be considered a discrete laboratory test that provides independent information leading directly to a treatment decision. In this setting, the extent or intensity of staining may be more important than its presence or absence and can be of paramount importance in determining whether a patient is eligible for drug therapy. Thus, minor differences in staining may have a major impact on patients.

Few data were available on the overall accuracy of IHC testing for various biomarkers until recently, but it was generally assumed that experienced laboratories using modern equipment and methods were providing reliable results. This perception changed with the publication of studies showing significant discordances in the results of HER2 testing among different laboratories. In these studies, central testing failed to confirm the results of testing done in local hospital laboratories in up to a quarter of cases. This finding
was alarming because of the impact on patients with inaccurate test results. A false-negative HER2 test would result in denial of a potentially life-saving therapy, whereas a false-positive test might lead to very expensive but unnecessary treatment with a drug that has significant cardiotoxicity in some patients.

These papers drew attention to the underappreciated reality that relatively common variations in preanalytical and analytical phases of routine testing can lead to a significant number of incorrect test results. Not surprisingly, this situation was deemed unacceptable and led directly to the creation of new HER2 testing guidelines by the College of American Pathologists and the American Society of Clinical Oncology. Lessons learned from HER2 have had an impact on predictive marker testing in general and underscore the need for greater stringency when testing for biomarkers. Some of the most important issues that must be addressed include fixation, assay validation, adherence to standardized methods, and ongoing laboratory performance assessment.

### Preanalytical Factors

The single most important preanalytical factor influencing consistency of biomarker testing is variability in fixation (time of fixation and fixative type). Incomplete fixation produces heterogeneous staining across tissue sections whereas prolonged fixation can result in partial or complete loss of immunoreactivity. Goldstein et al, in a study of the effect of time of formalin fixation on the results of estrogen receptor staining, showed that tissues must be fixed at least 6–8 h to achieve reliable results. This 6-h minimum is irrespective of specimen size; it is not the rate of tissue penetration by formalin that is important, but rather the time needed to complete the chemical reactions that comprise formalin “fixation.” Although penetration of aqueous formalin into tissues is relatively fast (about 1 mm per hour), the biochemical cross-linking needed for complete fixation is much slower (up to 24 h). Thus, all specimens, including needle biopsies, need to be fixed for at least 6 h before processing. Placing specimens in the tissue processor before they have been completely formalin fixed leads to completion of fixation in alcohol (a coagulation fixative). This event is an added source of variability that can result in false-positive and false-negative IHC test results.

The adverse effects of underfixation on IHC have been shown to be greater than those of overfixation. While problems with overfixation can be overcome to a significant extent by heat-induced antigen retrieval, there are no such remedies for underfixed tissues.

When using an in vitro diagnostic assay that specifies specimen handling conditions, one must follow the manufacturer’s instructions on fixation and processing. Most IHC assays have been validated on tissues fixed in 10% neutral buffered formalin, and use of an alternative fixative may invalidate the predictive aspects of the assay. Laboratories that choose to use another fixative must validate that fixative’s performance by testing the same tissues that have also been fixed in buffered formalin and showing equivalent results.

### Analytical Factors

Formalin fixation alters the three-dimensional configuration of protein as a result of protein–protein cross-links, thereby potentially altering antibody-antigen interaction. The introduction of antigen retrieval techniques in the early 1990s revolutionized IHC by drawing attention to methodologies to unmask these protein cross-links. Microwave pretreatment of sections immersed in a buffer enhanced immunostaining reactions several fold. Subsequently, other techniques employing heat (e.g., pressure cooking, steam heating, and high-temperature water bath) and proteolytic enzyme digestion (e.g., proteinase K, trypsin, pepsin, etc.) and combinations thereof have been employed to great success; however, caution with respect to overdigestion must be exercised, because excessive tissue digestion may result in destruction of antigenicity with subsequent false-negative immunoreactions. Conversely, extreme tissue digestion may also unmask “unwanted” antigens, resulting in false positivity. Both pitfalls can be avoided with cautious optimization and monitoring of antigen retrieval techniques plus use of appropriate positive and negative controls.

Antibodies used to assess biomarkers can vary significantly in sensitivity and specificity. Until recently, most antibodies used in clinical practice were rabbit polyclonal or mouse monoclonal reagents. Newer rabbit monoclonal antibodies generally have greater affinity and specificity than mouse monoclonal antibodies. As a result, these antibodies are often quickly adopted for routine practice; however, because the intensity or extent of staining is of such critical importance in predictive biomarker testing, conclusive evidence of superiority over existing reagents in predicting outcome should be required before changing methods.

The choice of an immunodetection system is of critical importance in accurate and reproducible biomarker testing. The most widely employed detection system is the avidin–biotin IHC technique, but this is being gradually replaced by more sensitive polymer-based detection systems. The latter involve a two-step procedure, comprising a dextran polymer “backbone” to which are bound peroxidase molecules and secondary antibodies (to bind to the unlabeled primary antibody). This highly recommended system enjoys the advantage of beingavidin-free, thus avoiding potential nonspecific binding with endogenous biotin in some tissues. Newer techniques employing tyramide-based amplification technology (or catalyzed reporter deposition) provide even more sensitive methods of antigen detection. Although these latter techniques are useful for detecting low
levels of antigenicity, they do carry the potential problem of spurious false-positive immunoreactions, and, hence, cautious interpretation with appropriate positive and negative controls is always mandatory.

When assessing biomarkers by IHC, one must always consider the possibility of intratumoral heterogeneity. A significant portion of cases with equivocal immunohistochemical staining for HER2 also show heterogeneity of gene amplification. Lewis et al., in a study of cases that were equivocal (Score 2) by IHC, found that when staining multiple additional blocks from the same case, more than half of cases had a score that was different from the initial score. This study confirmed that equivocal staining is often the result of intratumoral genetic heterogeneity and/or low-level gene amplification.

All laboratory assays must be validated before results can be reported on patient specimens, but validation is particularly important for biomarker assays. For initial IHC test validation, labs should show that their assay agrees with a reference method in a high percentage of cases. The ASCO/CAP HER2 Guidelines require at least 95% concordance between IHC and fluorescence in situ hybridization (FISH) (or with another lab’s validated IHC assay) when testing the same blocks. The validation study must be done on at least 25–100 cases, which include negative and positive samples. Assays must be revalidated whenever there is a significant change to test methods (new antibody type, new detection system, new automated platform, etc.).

Postanalytical Factors

Besides strict adherence to specimen handling guidelines and assay performance specifications, ongoing assessment of laboratory performance is essential. This step includes training and regular competency assessment of all laboratory personnel, including the pathologists who interpret the stain slides.

Importantly, some form of continuing quality assessment is needed for each biomarker assay offered by the laboratory. This assessment is usually done by participating in an external proficiency testing program, but if such a program is not available, for example, for new or rare analytes, an alternative form of proficiency testing may be done. This effort may take the form of split sample analysis with other labs or comparing the results of staining with an alternative test method (e.g., FISH). Such steps must be done on a regular, ongoing basis (e.g., several challenges two to four times per year).

Conclusions

Molecular tests that directly analyze nucleic acids can provide highly specific diagnostic and predictive information, but IHC continues to play an essential role in biomarker testing and clinical decision making. Because the quantitative aspect of predictive marker testing by IHC is subject to significant analytical variability, steps must be taken to ensure that standardized methods are used throughout the testing process.

References


Introduction

The concept of flow cytometry has been in existence for more than five decades. With the emergence of antibodies and fluorochrome chemistry and the advances in laser and computer technology, flow cytometry has become a powerful and indispensable tool for the diagnosis of hematopoietic and lymphoid disorders. Although limited, it can be also used for determining cell types of solid tissue neoplasms. Current flow cytometers have the capability of simultaneously measuring multiple parameters of individual cells in a cell suspension. Thus, a large number of cell specimens can be processed with a quick turnaround time. In addition, flow cytometry is also highly sensitive and can detect immunophenotype of cells in a specimen with thousands of cells. The parameters analyzed by flow cytometry include physical properties of cells; the size, cytoplasmic granularity, and amount of DNA contents; and cell antigens/markers (surface, cytoplasmic, and nuclear) that can be recognized by specific antibodies. By using appropriate antibody panels, flow cytometry can reveal the cell type (hematopoietic, lymphoid, or nonhematopoietic), cell lineage (B- and T cells, natural killer cells, myeloid/monocytic cells, neuro/neuroendocrine cells, and epithelial cells), cell maturation stage (precursors vs. matured cells), and B-cell clonality of expressing immunoglobulin light chains. Moreover, flow cytometry allows us to sort cells based on their distinct parameters for clinical and research purposes. In this chapter, we outline the basic principles of flow cytometry for understanding when flow cytometry will be helpful and how to interpret flow cytometric data for disease diagnosis. Importantly, flow cytometry analysis results must be correlated with clinical presentations and pathological findings for accurate diagnosis.

Basic Principles of Flow Cytometry Analysis

Flow cytometry has a variety of clinical applications include phenotyping analysis (diagnosis) of leukemia and lymphoma, following-up of residual disease, assessment of cell size and membrane integrity (viability), detection of antigens used as therapeutic targets, monitoring CD4/CD8 T-cell ratio in human immunodeficiency virus (HIV)-infected individuals, detection of DNA/RNA content (cell-cycle analysis), assessment of reticulocyte count, detection of platelet antibody, diagnosis of paroxysmal nocturnal hemoglobinuria, detection of nonhematopoietic neoplastic cells, transplant applications (HLA typing), and estrogen receptor analysis. Flow cytometry analysis mainly relies on identification of cell antigens/markers by using specific antibodies (immunophenotyping study). However, the quantity and quality of the diagnostic markers may differ in different cells. Thus, the diagnostic value of flow cytometry varies among diseases. Generally speaking, flow cytometry is

- Most powerful for diagnosis and classification of B-cell lymphoma/leukemia
- Very powerful for diagnosis of plasma cell neoplasm
- Very useful for diagnosis of precursor B- and T-cell neoplasms
- Very useful for diagnosis and classification of acute myeloid leukemia
- Useful for diagnosis of T-cell lymphoma/leukemia
- Helpful in diagnosis of myelodysplastic syndromes and chronic myeloproliferative disorders
- Not useful for diagnosis of nonhematopoietic/lymphoid disorders
In comparison to immunohistochemical techniques, flow cytometry analysis has its incomparable advantages including simultaneous multiparametric analysis of individual cells, shorter turnaround time (minutes to hours), and the need for many fewer specimens (thousands of cells) for diagnosis (Table 15.1). Flow cytometry also provides quantitative results with less subjective interpretation. However, sampling bias without morphological correlation may diminish its accuracy. All these factors need to be taken into consideration when ordering a flow cytometry analysis.

Technical Aspects

Antibodies

One of the most significant advancements in the immunophenotyping analysis of cell-surface antigens/markers is the development of highly specific monoclonal antibodies. Individual monoclonal antibodies are produced by cloned antibody-secreting cells and specifically recognize the antigen that was used in immunolization for the antibody production. The cluster of differentiation (abbreviated as CD) was initially a protocol used for the identification and distinction of cell-surface molecules (antigens) present on leukocytes. In 1982, the CD nomenclature was formally established in the first International Workshop and Conference on Human Leukocyte Differentiation Antigens (HLDA). This system was used for the classification of the monoclonal antibodies based on their recognized CD molecules. Each surface molecule that is recognized by antibodies is assigned a CD number individually. Using the CD system, we can identify cells by the presence or absence of particular surface molecules. Cells are usually defined using a “+” or a “−” symbol to indicate whether they express or lack the CD molecules. For example, a “CD3+ and CD20−” cell is one that expresses CD3 but not CD20. Currently, the CD system has been expanded and is widely used in all types of cells.15–17

Fluorochromes and Fluorescent Dye

The characterization of cells by flow cytometry is mainly based on the presence of distinct cell antigens/markers and cell contents, which can be detected through the fluorescent signals directly or indirectly. Fluorescent dyes can directly bind to certain cellular content, such as DNA and RNA, and allow us to perform quantitative analysis on individual cells. However, in most cases fluorochromes are conjugated with monoclonal antibodies, which specifically target cellular antigens/markers. Fluorochromes are substances that can be excited by certain light source (such as laser) and emit a fluorescent signal at a single wavelength. Table 15.2 provides a list of commonly used fluorochromes and dyes. Interestingly, although some of them can be excited by the same light source, the different fluorochromes may emit fluorescent signals with different wavelengths/colors. Thus, multiple fluorochromes can be simultaneously excited by a light source and detected by their emission fluorescent signals with different wavelengths, respectively. Under the current consensus guidelines, a minimum of three fluorochromes should be used in a multicolor flow cytometry for the analysis of leukemia and lymphoma.13 As illustrated in Fig. 15.1, by utilizing different fluorochrome-conjugated antibodies, multiple antigens/markers in the same cells can be simultaneously detected.

Flow Cytometers

For cell marker detection and analysis, the basic components of a flow cytometer include (1) the flow system or fluidics, (2) an optical system to stimulate and detect fluorescent signals, (3) an electronic system, and (4) a computer system (Fig. 15.2). The flow system has a coaxial streaming nature that transports and presents single cells in suspension to the laser beam(s) and the optical path for the detectors through hydrodynamic focusing. The coaxial stream consists of the inner sample fluid surrounded by an outer sheath of isotonic fluid that creates a laminar flow, allowing the cells in the

---

**Table 15.1. Comparison of immunophenotypic techniques.**

<table>
<thead>
<tr>
<th>Flow cytometry</th>
<th>Immunohistochemistry</th>
</tr>
</thead>
<tbody>
<tr>
<td>Shorter turnaround time (minutes to hours)</td>
<td>Longer turnaround time (hours to days)</td>
</tr>
<tr>
<td>Less subjective result interpretation</td>
<td>Subjective result interpretation</td>
</tr>
<tr>
<td>Quantitative results</td>
<td>Semiquantitative results</td>
</tr>
<tr>
<td>Good to detect immunoglobulin light chain restriction</td>
<td>High background staining for immunoglobulin light chains</td>
</tr>
<tr>
<td>Multiple antibodies/fluorochromes per test</td>
<td>Usually limited to a single antibody per slide</td>
</tr>
<tr>
<td>Greater antibody selection</td>
<td>Fewer antibodies available</td>
</tr>
<tr>
<td>Data/results can be electronically transferred</td>
<td>Slides can be shipped by mail or courier service</td>
</tr>
<tr>
<td>Need fresh cells or tissue</td>
<td>Can use fixed/archived tissue</td>
</tr>
<tr>
<td>Limited morphologic correlation</td>
<td>Architectural and cytologic correlation</td>
</tr>
<tr>
<td>Cannot assess nonviable cells</td>
<td>Can assess nonviable “ghost” cells</td>
</tr>
</tbody>
</table>

**Fluorochromes and Fluorescent Dye**

Table 15.2. Characteristics of fluorochromes commonly used in flow cytometry.

<table>
<thead>
<tr>
<th>Fluorochromes conjugated to antibodies</th>
<th>Excitation wavelength (nm)</th>
<th>Emission wavelength (nm)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fluorescein isothiocyanate (FITC)</td>
<td>488*</td>
<td>530</td>
</tr>
<tr>
<td>Phycoerythrin (PE)</td>
<td>488</td>
<td>580</td>
</tr>
<tr>
<td>PE-Texas Red</td>
<td>488</td>
<td>615</td>
</tr>
<tr>
<td>PE-Cy5</td>
<td>488</td>
<td>670</td>
</tr>
<tr>
<td>Peridinin chlorophyl protein (PerCP)</td>
<td>488</td>
<td>670</td>
</tr>
<tr>
<td>Allophycocyanin (APC)</td>
<td>633*</td>
<td>670</td>
</tr>
<tr>
<td>APC-Cy7</td>
<td>633</td>
<td>767</td>
</tr>
<tr>
<td>Propidium iodide (dye to stain DNA)</td>
<td>488</td>
<td>620</td>
</tr>
</tbody>
</table>

*Argon laser is commonly used for 488 nm and HeNe laser for 633 nm excitation wavelength.
center to pass individually through the interrogation point of the laser beams. The optical system carries laser beam(s) to excite the cells and fluorochromes and uses a series of optical devices including filters, mirrors, and detectors to collect the incident light, reflected or scattered light, and emitted fluorescent signals (see Fig. 15.2). The most commonly used lasers, argon ion lasers, produce a laser beam of 488-nm wavelength that is capable of exciting several fluorochromes and dyes with different excitation wavelengths/colors, as listed in Table 15.2. The multicolored fluorescent signals with different wavelengths are then separated by the combined usage of filters, lenses, and dichroic mirrors. In the electronic system, after traveling through the optical system, the light and fluorescent signals are collected by the corresponding detectors and then converted into electronic impulses proportional to the number of photons. Although these impulses/signals are analog, they can be further amplified and digitized for computer analysis. An external computer system can be directly connected to the flow cytometer and interfaced with it in two ways. As a navigator, it controls many operating functions of the flow cytometer such as establishing the instrument setting, setting the compensation, monitoring operation conditions, gating cell populations, and directing the cell-sorting controls. The computer system also plays an important role in acquisition, storage, analysis, and presentation of cytometric data. Most often data are graphically displayed as dot plots or two-parameter histograms, although a variety of other data presentation formats, such as contour plot and isometric plot, are available. Contemporary cytometers are generally capable of analyzing between 5 and 13 parameters (forward scatter, side scatter, and as many as 11 different fluorescent wavelengths/colors).

Specimens Suitable for Flow Cytometry

Theoretically, any specimens from which a single cell suspension can be generated are suitable for flow cytometry analysis. However, a lack of distinct antigens or markers in the cells of interest or tissues limits the diagnostic value of flow cytometry. Common specimens suitable for flow cytometry analysis include peripheral blood, bone marrow, body fluids, cerebrospinal fluid, urine, lymph node (cells or fresh tissues), any fine-needle aspirates, and fresh tissues suspicious for hematopoietic and lymphoid disorders. For blood and bone marrow specimens, anticoagulants such as ethylenediaminetetraacetic acid (EDTA), heparin, or acid citrate...
Dextrose are needed. Fresh tissue specimens are best transported and stored in sterile tissue culture medium such as RPMI1640. Although specimens may be stored at room temperature, refrigeration is preferred, particularly when there is a delay for flow cytometric analysis. For flow cytometry analysis, single-cell suspensions of the fresh tissues can be achieved by mechanical dissociation.

**Data Analysis and Interpretation**

At the time of flow cytometry analysis, it is important to collect data on sufficient events (cell numbers) and all available parameters to ensure detection of the cells of interest. The immunophenotyping characterization of hematopoietic and lymphoid neoplasms relies solidly on the principle that neoplastic cells frequently exhibit abnormal or aberrant antigenic expression that differs from normal cells.\[^{20-22}\] Abnormal/aberrant antigenic expression can be grouped into four basic categories:

- Abnormally increased or decreased levels of antigenic expression (aberrant expression)
- Gain of antigens not normally expressed in the cell type or lineage
- Expression of antigens not synchronized with normal development and maturation stage of the cell type or lineage
- Homogeneous expression of antigen(s) by a cell population that normally show more heterogeneous expression

**Basic Parameters and Windows of Cell Populations**

*Forward light scatter (FSC) and side light scatter (SSC).* FSC collects light at 180° from the point at which the laser beam intersects the cells, usually on a linear scale. It is correlated with cell size, and thus can distinguish normal lymphocytes (small), monocytes (intermediate), and neoplastic cells (generally they are large in size). SSC collects right-angle light at 90° and is correlated with cytoplasmic granularity and nuclear configuration. The combination of both FSC and SSC can distinguish normal lymphocytes, granulocytes, and monocytes (Fig. 15.3a-c). The detection of lymphocytes and monocytes provides a reliable internal control to evaluate the size of the cells of interest.

*CD45 and SSC.* As the first step, it is most important to determine whether the cells of interest are hematopoietic/lymphoid. Generally speaking, all hematopoietic/lymphoid cells express CD45 antigens (CD45+). Thus, a histogram of CD45 on a logarithmic scale vs. SSC on a linear scale is indispensable as a starting point of flow cytometry analysis (Fig. 15.3d-f). Based on antigen expression, cells are divided into CD45+ and CD45− groups. Among the CD45+ group, the cells can further separated into subgroups (windows in the histogram) based on expression levels of CD45 and intensity of cytoplasmic granularity.\[^{23-25}\]

- **CD45-negative cells**
  - Nonhematopoietic/lymphoid cells
  - Exception: erythroid and megakaryocytic precursors, giant platelets, and abnormal plasma cells
- **CD45-positive cells (hematopoietic/lymphoid)**
  - Lymphocyte window
  - Myeloblast window
  - Lymphoblast window
  - Granulocyte window
  - Monocyte window

**Gating Cell Populations and Immunophenotyping Analysis**

Gating is the most important first step in immunophenotyping analysis. It is critical particularly in a specimen that contains mixed cell populations, such as bone marrow aspirate. To accurately target and gate the cell population(s) of interest, one must correlate the patient’s history, clinical presentation, and laboratory tests, and the clinician’s concerns and possible pathology findings. Generally, there are three ways to gate cells:

1. **By cell distribution in the CD45 vs. SSC histogram:** This is most useful in a specimen containing mixed cell populations (see Fig. 15.3d). The grouped cells in individual windows represent different cell lineages (Fig. 15.4).
2. **By cell size:** In FSC vs. SSC histograms, neoplastic cells (usually large in size) can be gated by using lymphocytes (small) and monocytes (intermediate) as an internal size control (see Fig. 15.3a). Once the cells of interest are gated, further analysis of cell lineage can be performed.
3. **By cell lineage-specific antigens (immunophenotype):** If cells are CD45+ but do not fit into particular windows in the CD45 vs. SSC histogram (see Fig. 15.3e), identification of lineage-specific antigen expression is needed (Table 15.3).

**Immunophenotyping Analysis and Interpretation**

When the cell populations of interest have been accurately gated, detailed immunophenotyping analysis can be carried out by the presence and absence of specific cell antigens/markers. Figures 15.4–15.8 outline the basic principles of data interpretation for individual cell windows/cell lineage step by step. Generally, analysis can start with the CD45 vs. SSC histogram, and the cells of interest are gated for further characterization based on clinical indication. In specimens
Fig. 15.3. FSC vs. SSC and CD45 vs. SSC histograms. (a, d) Normal bone marrow aspirate. In the FSC vs. SSC histogram, identification of monocytes (intermediate size) and lymphocytes (small size) provides a useful internal size control. In the CD45 vs. SSC histogram, cells are initially divided into CD45(−) nonhematopoietic/lymphoid cells and CD45(+) hematopoietic/lymphoid cells. The CD45(+) cells can be further classified by their distribution in individual windows: (1) lymphocyte window; (2) lymphoblast/hematogone window; (3) myeloblast window; (4) monocyte window; and (5) granulocyte window (see Fig. 15.4 for details). (b, e) B-cell lymphoma biopsy. In the FSC vs. SSC histogram, cell size can be determined by evaluating cell FSC as indicated. In addition, the CD45 vs. SSC histogram reveals that cells are CD45(+) hematopoietic/lymphoid cells and present within the lymphocyte window. (c, f) Small cell carcinoma (SCC) tissue biopsy. The FSC vs. SSC histogram demonstrates that SCC cells are intermediate in size (intermediate FSC) and have a complicated nuclear configuration (intermediate to high SSC). In the CD45 vs. SSC histogram, SCC cells are CD45(−) and a small group of background lymphocytes are CD45(+).

Fig. 15.4. Identification of cell windows/lineages in CD45 vs. SSC histogram. As the first step, cells are divided into CD45(+) and CD45(−) groups. The majority of CD45(−) cells are nonhematopoietic/lymphoid, with rare exceptions. CD45(+) hematopoietic/lymphoid cells can be further divided into different windows/lineages as illustrated in Fig. 15.3. Detailed immunophenotyping analysis of the cells in each window is shown in Figs. 15.5–15.8.
Table 15.3. Cell lineage associated antigens.

<table>
<thead>
<tr>
<th>Cell lineage</th>
<th>Antigens</th>
</tr>
</thead>
<tbody>
<tr>
<td>All leukocytes</td>
<td>CD45&lt;sup&gt;a&lt;/sup&gt;</td>
</tr>
<tr>
<td>B cells</td>
<td>Cyto CD22&lt;sup&gt;b&lt;/sup&gt;, CD19, CD20, CD22, CD79a, CD79b</td>
</tr>
<tr>
<td>T cells</td>
<td>Cyto CD3&lt;sup&gt;a&lt;/sup&gt;, CD2, CD3, CD4, CD5, CD7, CD8</td>
</tr>
<tr>
<td>Plasma cells</td>
<td>CD38 (bright), CD138 (bright)</td>
</tr>
<tr>
<td>NK cells</td>
<td>CD2, CD16, CD56, CD57</td>
</tr>
<tr>
<td>Myeloid cells</td>
<td>CD13, CD15, CD33</td>
</tr>
<tr>
<td>Monocytes</td>
<td>CD14, CD36, CD68</td>
</tr>
<tr>
<td>Platelet/megakaryocyte</td>
<td>CD41, CD42, CD61</td>
</tr>
<tr>
<td>Erythroid cells</td>
<td>CD71, glycophorin A</td>
</tr>
<tr>
<td>Myeloblasts</td>
<td>CD34, CD117</td>
</tr>
<tr>
<td>Epithelial cells</td>
<td>EpCam</td>
</tr>
<tr>
<td>Neuroendocrine cells</td>
<td>CD56, EpCam, CD138</td>
</tr>
</tbody>
</table>

<sup>a</sup>CD45, cytoplasmic CD22, and cytoplasmic CD3 are cell lineage specific antigens

Fig. 15.5. Lymphocyte window. For B cells, initially the κ/λ light chain ratio should be checked in total cells (normal range, 0.5–4), as well as in CD5<sup>+</sup> or CD10<sup>+</sup> cell populations. For T cells, the expression level of T-cell antigens should be examined (normal T cells show same or similar expression as CD2, CD3, CD5, and CD7, which are also equal to the sum of CD4<sup>+</sup> and CD8<sup>+</sup> T cells). r/o, rule out; NK cells, natural killer cells; ATLL, adult T-cell leukemia/lymphoma; AITL, anaplastic large cell lymphoma; ALCL, angio-immunoblastic T-cell lymphoma; MF, mycosis fungoides/Sézary syndrome; LGL, large granular lymphocytes; CLL/SLL, chronic lymphocytic leukemia/small lymphocytic lymphoma; MCL, mantle cell lymphoma; FCL, follicular lymphoma; BL, Burkitt lymphoma; LBCL, large B-cell lymphoma; MZBCL, marginal zone B-cell lymphomas; HL, hairy cell leukemia.

Containing relatively homogeneous cells, analysis may start with cell lineage identification by using the specific antibodies listed in Table 15.3. In cases in which the cell type of interest is unknown, gating large cells in the FSC vs. SSC histogram might be helpful for cell lineage determination (neoplastic cells are usually large). This step is particularly useful in a specimen with limited and mixed cells, by which distinct cell populations or windows cannot be established in the CD45 vs. SSC histogram. Again, to make an accurate interpretation of flow cytometry data, clinical correlation and discussion with clinicians are indispensable. Consultation with experts is also valuable.
Fig. 15.6. Myeloblast window. r/o, rule out; PML, acute promyelocytic leukemia; AMLs, acute myeloid leukemias excluding acute erythroid and megakaryoblastic leukemias.

Fig. 15.7. Lymphoblast window. r/o, rule out; LBL, lymphoblastic leukemia/lymphoma.

Fig. 15.8. Granulocyte and monocyte windows. MDS, myelodysplastic syndromes; MPD, myeloproliferative diseases.
References

Introduction

Stem cells are defined by their potential to self-renew and differentiate into more specialized cell types within a given tissue. Stem cells can be found in the embryo and in adult tissue and are termed embryonic stem cells and adult stem cells, respectively. Embryonic stem cells are responsible for the development of an entire organism, whereas adult stem cells provide a mechanism for maintenance of tissue homeostasis by replacing damaged cells throughout the life of the organism. Stem cell biology has provided a platform for addressing important scientific questions in the areas of development, repair/regeneration, and disease. Here, we discuss some basic concepts regarding stem cells, including definitions of commonly used terminology and the current status of the stem cell field. A glossary of terms defined in this chapter is given in Table 16.1 (Table 16.1).

Embryonic Stem Cells

Embryonic stem cells are derived from the inner cell mass of a 4- to 5-day-old embryo (blastocyst), before implantation of the embryo in the uterus (Fig. 16.1). These embryonic stem cells are pluripotent, defined as having the potential to differentiate into any of the tissues derived from the three major germ layers—endoderm, ectoderm, and mesoderm—and ultimately into all the cells that make up the fully developed organism. They express telomerase to maintain telomere length, which would account for their extended capacity to replicate and self-renew. In addition to certain cell-surface markers, embryonic stem cells characteristically express Oct-4 and Nanog, as part of a group of transcription factors that are considered to be important in suppressing gene expression and maintenance of their pluripotent and undifferentiated state. Another distinctive characteristic is the absence of, or highly abbreviated, G1 checkpoint in the cell cycle.

Adult Stem Cells

Adult stem cells are undifferentiated cells that can be found throughout the body of a mature organism. Adult stem cells can self-renew as do embryonic stem cells; however, they are multipotent rather than pluripotent. Adult stem cells have been identified in tissues that have developed from all three germ layers, including the brain, bone marrow, lungs, peripheral blood, blood vessels, skeletal muscle, skin, digestive tract, cornea, retina, dental pulp, liver, and pancreas. Therefore, adult stem cells have been identified in tissues that develop from all three germ layers (see Fig. 16.1). However, none of these adult stem cell populations is capable of differentiating into cells from all three germ layers. Instead, the major function of adult stem cells is to serve as a source of cells to regenerate the tissues in which they reside, both for homeostasis and for reparative responses to injury.

Although the origin of embryonic stem cells is well defined, much less is known with respect to adult stem cells, partly because they are present in very small numbers within defined tissue niches and are difficult to identify because of the absence of specific markers. This lack of distinct cell-surface markers also contributes to the technical challenges for their isolation and purification. A combination of markers, both negative and positive, is often required to define a particular stem cell population for purification; in the absence of specific stem cell markers, negative selection based on the absence of markers that define more differentiated cells is necessary. Hematopoietic stem cells (HSCs) represent an example of adult stem cells that have been well characterized, as they were the first to be discovered and isolated. By using fluorescence-activated cell sorting (FACS) and a panel of monoclonal antibodies that recognize cell-surface antigens specific to HSCs, a nearly homogeneous population of HSCs has been isolated in mice and in humans. Further research aimed to better characterize cell-surface markers and/or a gene expression profile that identify specific adult stem cell populations will allow for improved isolation techniques and better characterization of tissue-resident adult stem cells and their origins.
Table 16.1. Glossary of terms.

**Determination:** The irreversible commitment of a cell to follow a specific developmental pathway.

**Differentiation:** The overall process by which stem/progenitor cells are activated to become more specialized cell types.

**Fate:** Term that describes what cells will become as a result of differentiation.

**Multipotent stem cells:** Cells that self-renew and differentiate into several different specialized cell types, often within a tissue (e.g., hematopoietic stem cells).

**Plasticity:** The ability of an adult stem cell from one tissue to develop into differentiated cell types from a different tissue; this remains largely unproven as an in vivo physiological process.

**Potential:** Term that describes what cells are able to become.

**Progenitor cell:** An undifferentiated precursor cell with the capacity to undergo differentiation into specialized cell types; unlike putative stem cells, they do not retain the capacity for self-renewal.

**Pluripotent stem cells:** Cells capable of self-renewing and differentiating into any of the three germ layers (endoderm, ectoderm, and mesoderm); unlike totipotent cells, they do not give rise to embryonic components of the trophoblast and placenta.

**Totipotent stem cells:** Cells derived from the first few divisions of the fertilized egg; these cells have the potential to give rise to all the differentiated cells of the fully developed organism.

**Self-renewal:** A defining property of stem cells that allows them to undergo repeated mitotic cell divisions to create at least one daughter cell equivalent to the mother cell that retains latent capacity for differentiation.

---

**Fig. 16.1.** Schematic depicting the hierarchy and differentiation potential of embryonic and adult stem cells. Pluripotent embryonic stem cells are derived from the inner cell mass (ICM) of a 4- to 5-day-old embryo (blastocyst) and can differentiate into the tissues that comprise the endoderm, ectoderm, and mesoderm. Multipotent stem cells that may also derive from all three germ layers have been identified in adult tissues; these adult stem cells self-renew within tissue niches, maintain tissue homeostasis, and participate in repair/regeneration following injury. Both embryonic and adult stem cells undergo asymmetrical mitotic cell division to form progenitor cells, which eventually differentiate into specialized cell types.
Hierarchy of Stem Cell Differentiation

Both embryonic and adult stem cells can undergo mitotic cell division to form progenitor cells (sometimes referred to as a transit-amplifying population) with the loss of some degree of multipotency before reaching their specialized fate. The convergence of intrinsic cellular signals and extrinsic microenvironmental cues from the surrounding niche regulates differentiation, but the specific signals involved are poorly understood. For example, the mechanisms by which the daughter cells of a stem cell either remain a stem cell or become a progenitor cell are not entirely clear, although it can involve the inheritance of cell fate determinants from the mother cell and/or niche factors. In contrast to stem cells, progenitor cells have not been shown to possess unlimited self-renewal capacity. There are, however, some poorly defined elements with respect to this and other aspects (e.g., degree of plasticity) of the definition of a progenitor cell versus a stem cell that have been used interchangeably to reflect this uncertainty. Progenitor cells will eventually differentiate into mature specialized cell types. During differentiation, activation of a program of gene expression orchestrated by a set of transcription factors and epigenetic changes results in the phenotypic maturation of a cell with specialized functions. For example, bone marrow stromal cells or mesenchymal stem cells (MSCs) have the capacity to differentiate into adipocytes, bone, and cartilage in culture and upon transplantation into donors.3–5 In general, there exists a hierarchy of stem cell differentiation, where stem cell division yields progenitor cells that become progressively more restricted in their developmental potential, assuming the phenotype and function of specialized cells (see Fig. 16.1).

Stem Cell Plasticity

Stem cells remain quiescent within their niche, but possess extraordinary proliferative capacity compared to other cell types. For example, a single hematopoietic stem cell (HSC) can supply all the blood cells throughout the life of a mouse via self-renewal and multilineage differentiation capabilities.6 Some adult stem cells appear to exhibit extended plasticity, in that they can differentiate into tissues other than the tissue from which they originated.7,8 However, there is also controversy with respect to the degree of plasticity of many of these adult stem cells, including HSCs, which have been claimed to give rise to parenchymal cells of many organs, including the liver, lung, and heart.6–10 An added complication is that fusion of these stem cells with the target differentiated cell at a distal site may give the appearance of transdifferentiation, a phenomenon that is known to occur.9,10 Further research is necessary to conclusively establish the extent of plasticity of these adult stem cells in vivo. A noteworthy circulating cell of limited plasticity has been described with the capacity to differentiate to fibroblasts, consequently referred to as fibrocytes.11,12 Presumably derived from HSCs, these cells are thought to contribute to the fibroblast populations at sites of injury and repair. With a more restricted plasticity are the resident stem cells found in many tissues that may give rise to the differentiated parenchymal elements within the tissue or organ itself.13 Further studies are required to determine the origin(s), niche, cell-specific markers, and plasticity of these tissue-resident stem/progenitor cells.

Mechanisms of Homeostasis and Response to Injury

A primary function of adult tissue-resident stem/progenitor cells is to maintain cellular homeostasis within the tissue/organ and replace damaged cells following injurious insults. The mechanisms for repair/regeneration in humans are organ specific, with distinct resident stem cell populations giving rise to the differentiated elements of the tissue of origin. This mechanism may involve the potential for differentiated cells to act as precursors to other differentiated cell types in response to microenvironmental cues, particularly in organs that turn over slowly, such as the liver, lung, and pancreas. In skeletal muscle tissues, a population of stem cells that reside between the plasma membrane of multinucleated skeletal muscle cells and the basal lamina that surrounds each muscle cell are called “satellite stem cells” and function as classic adult tissue-resident stem cells.14,15 Satellite cells are normally nondividing, quiescent cells unless they are triggered to proliferate as a result of injury. Transplant-amplifying satellite cells can produce myoblasts that fuse with preexisting multinucleated skeletal muscle help repair damaged tissue. Similarly, progenitors of lung and intestinal epithelium have been described that reside within or subjacent to the epithelium itself and serve as the precursors for the differentiated epithelium to replace cells lost by normal sloughing or toxic injury.

Stem Cells and Disease

Recent studies have begun to implicate stem cells in diverse human diseases. In carcinogenesis, for example, work by several groups has led to the notion that cancer stem cells may be responsible for the growth and maintenance of tumors. The cancer stem cell hypothesis suggests that malignant tumors are initiated and maintained by a population of cancer stem cells that share biological properties similar to those of normal adult stem cells.16 A cancer stem cell has been defined as a self-renewing cell within a tumor that has the capacity to regenerate the phenotypic diversity of the original tumor.17 The first evidence for this concept was provided by the finding that a subpopulation of CD34+/CD38− cells from acute myeloid leukemia patients initiated leukemia when transplanted into NOD/SCID mice.18 Cancer stem cells have been
implicated in a growing list of cancers including those in the breast, brain, pancreas, colon, and ovary. Current efforts are focused on dissecting the molecular pathways by which cancer stem cells are initiated and propagated. Specific genes and pathways, such as Shh and Bmi-1, are just beginning to be identified in these processes. The ability to selectively target cancer stem cells in human patients would revolutionize the field of cancer treatment.

The role of adult stem cells in other diseases is also receiving increasing interest. For example, discovery of the fibrocytes suggests that bone marrow-derived cells may contribute to the pathogenesis of organ fibrosis; this is in addition to the well-known role of inflammation in disease and the HSC origin of the inflammatory cell participants. In addition to their role as the source of cells that directly participate in disease, deficiencies in stem cell function or numbers may play a role in pathogenesis. Moreover, a number of age-related degenerative disorders, and even aging itself, may be related to a deficiency in stem cell responses and tissue-renewal capacity. Deficient stem cell function or replicative capacity has significant consequences on the ability to produce essential cells for normal function of the affected organ. For example, dyskeratosis congenita is characterized by bone marrow stem cell failure caused by impaired telomerase function and consequent associated aplastic anemia.

**Significance and Future Research**

Embryonic stem cell research will likely uncover fundamental events during development and lead to a better understanding of the causes of birth defects and novel strategies to allow us to prevent or correct them. Adult stem cells have recently received a great deal of research interest because of their potential for future clinical applications. Two obvious advantages for clinical applications include these: (a) adult stem cells may be isolated from the patient’s own tissues and thus provide an autologous cell source that obviates issues of immune incompatibility; and (b) ethical concerns surrounding the use of embryonic stem cells can be avoided. More recently, another alternative is provided by the discovery that fibroblasts can be reprogrammed or induced to revert to a phenotype comparable in many respects to that of a pluripotent stem cell. These so-called induced pluripotent stem cells (iPSCs) can be artificially derived from a nonpluripotent somatic cell by inducing expression of three or four specific genes that are normally expressed in pluripotent stem cells. Thus, human iPSCs are obtained by inducing expression of Oct3/4, Sox2, Klf4, and c-Myc with a retroviral system, while another study achieved a comparable outcome by inducing expression of Oct4, Sox2, and Nanog using a lentiviral system. The hope is that these iPSCs and other stem cells will provide novel cell-based therapies that facilitate or directly contribute to the regeneration of damaged or lost tissues following injury or as a result of aging. In this regard, embryonic and adult stem cells, and potentially iPSCs, have been shown to contribute to the repair or replacement of damaged tissues in animal models with varying degrees of success. Recent observations of the protective effects of MSCs in acute tissue injury and inflammation represent another example of the potential utility of stem cells in therapy. To realize the full potential of cell-based therapies for regenerative medicine, we must take into account alterations to tissue niches in disease states and in aging that may influence the capacity for successful differentiation of transplanted cells, much of which remains to be discovered and elucidated.

**References**

Animal models afford the opportunity for investigators to study the mechanisms of human disorders by experimentally manipulating a number of controlled variables. Experimental models can be arbitrarily divided into (a) spontaneous disease models, which are mutant animals that carry a disease similar to a human condition; (b) models obtained through genetic manipulation, in other words, gene-modified models including: transgenic and knock out animals and (c) chemically or physically induced changes. Researchers have to choose from a number of possibilities such as species and strain of animal, environment, and the genome to investigate the molecular interactions involved in the pathogenesis of many diseases.  

Experimental models also provide a unique opportunity to test potential therapeutic intervention. Well-developed animal models should share features with specific human disorders. Although the basic molecular biology terminology is covered in other chapters of this book, it is convenient to briefly review the basic technology for the development of transgenic mice.

Transgenic Mouse Models

Transgenic mice technology has provided a powerful tool for both neoplastic and non-neoplastic disease investigation. Injection of exogenous DNA into fertilized eggs (zygotes) is known as transgenesis and is the most common method to modify the germline of mice. The offspring of transgenic founders are usually examined for a specific phenotype and are used to characterize genes and gene regulatory elements. The procedure involves a construct made of a segment of DNA that contains the sequence of interest and that has been cut with restriction enzymes. This segment may be ligated to a promoter that will induce expression of the transgene. Injection of the construct into mouse embryos results in a random integration and variable copy numbers into the mouse genome. The result will be expression of the transgene contained in the construct. Random integration and the copy number of the transgene are factors that cannot be controlled in the classical transgenic mouse technology. The procedure necessitates the generation and characterization of several founders to ensure that the phenotype is dependent on the transgene and not a result of alterations in the function of an endogenous gene. The standard protocol involves injection of the DNA construct into the pronucleus of fertilized eggs (0.5-day-postcoital embryos). These injected eggs are then implanted into the oviduct of a pseudopregnant foster mother. The injected DNA integrates randomly into the genome in a fraction of the implanted embryos. Founders are mated to demonstrate germline transmission and to establish a transgenic line.

We describe here some examples of experimental models to illustrate the powerful tool that they represent.

A good example of the role of lymphocytes in host defense against tumor development was provided in the paper by McCabe et al. They observed an interesting role for B lymphocytes in the surveillance of two experimental tumors in the rat. Following intravenous (i.v.) injection of MADB106 tumor cells into syngeneic WAG rats, most animals develop lung tumors. Previously, these tumors had been shown to be under control by natural killer (NK) cells. It was noted that after i.v. injection of these tumors, a rapid influx of B cells into the lung occurred. The study demonstrated that homing of B lymphocytes to the potential site of tumor development plays an important role in defense against these tumors.

Another area of a common human disease that gained knowledge from animal models is chronic obstructive pulmonary disease (COPD). COPD is one of the most common pulmonary diseases seen in clinical practice. Smoking is by far the most important risk factor for COPD. Because COPD is produced by a combination of environmental and genetic factors, animal models are ideal to study the pathogenesis of this condition. A number of species to study COPD have been proposed in the literature. In recent years, mice have been shown to be excellent for these studies because mouse and human genomes are very similar, and because...
of the ability to modify the genetic constitution of the mouse by either inducing a protein, knocking out genes, or creating new mutations. In addition, the physiology and general aspects of mouse biology are well known and a large number of inbred strains are commercially available. There are several experimental models of COPD and emphysema. Some depend on endobronchial perfusion of proteolytic enzymes in guinea pigs, hamsters, and mice. Others depend in tobacco smoke. There are also mouse strains that develop emphysema spontaneously.

An imbalance between collagen synthesis and degradation and between proteases and their inhibitors has been proposed for many years as a possible pathogenetic mechanism for the development of emphysema. It is not surprising that this hypothesis has been studied in experimental models. In theory, this imbalance may be caused by an excessive production of proteases by inflammatory cells or by reduced synthesis or increased breakdown of antiproteases. This hypothesis generated a number of experimental models that included intratracheal instillation of papain, trypsin, porcine elastase, and human neutrophil elastase as well as a large number of other proteases.

More recently, transgenic mice have allowed expression or overexpression of proteases, such as human interstitial collagenase (MMP-1). These mice develop emphysema that is very similar to that seen in human alpha-1 antitrypsin deficiency. The emphysematous changes are induced by degradation of type III collagen. Overexpression of interleukin 13 (IL-13) or interferon-gamma increases the expression of matrix metalloproteinases (MMPs) that are most likely responsible for the emphysematous changes. Knockout of TIMP-3 and, in other models, knockout of surfactant protein D, induces "spontaneous," age-related emphysema.

A number of the previously mentioned models provided support to the hypothesis of imbalance between protein formation and degradation, a hypothesis that has also been proposed for the development of interstitial fibrosis.

Asthma is another common disease that has been studied through a large number of animal models. These models have aided in the study of mechanisms and pathways leading to the development of airway hyperreactivity and inflammation. The majority of described models depend on previous sensitization and challenge with the offending antigen. Most of these models are reminiscent of the changes seen in individuals with airway hyperreactivity; however, the morphological changes seen in these animals do not resemble the changes that the pathologist sees in patients with asthma. Sensitization and challenge with interleukins such as IL-13 have provided new knowledge related to the role of leukotrienes and the glutathione pathway in this disease.

With the discovery of the cystic fibrosis (CF) gene in 1989, a number of models have been attempted. Three years later, the first CF mouse model was described. Approximately 11 CF mouse models have been described. CF, the most common genetic disease among Caucasians, is caused by mutations in the gene encoding CFTR (cystic fibrosis transmembrane conductance regulator). However, CFTR null mice did not develop pathology resembling human CF lesions. A more promising model was that of Mall et al., in which they enhanced sodium absorption in mouse airways by manipulating the epithelial sodium channel (ENaC). This model has provided new insight in the pathogenesis of CF and the study of ion transport defects.

An important clinical problem is the development of fibrosis of unknown causes in several organs, such as pulmonary interstitial fibrosis, some examples of “cryptogenic” cirrhosis, and even focal segmental glomerulosclerosis in the kidney. All these are common diseases that are associated with high morbidity and mortality, and their pathogenesis is not clear. To understand the pathogenesis and possible therapeutic interventions in interstitial diseases that progress to fibrosis, animal models provide valuable help. Pathologists have known for decades that the pathogenesis of fibrosis was associated with a number of conditions, such as infections, collagen vascular disease, allergic alveolitis, and trauma. However, the origin and mechanisms of development of idiopathic interstitial pulmonary fibrosis remain uncertain, probably in part from a lack of a good experimental model analogous to human interstitial fibrosis. Despite the number of experimental models, we still lack a model that recreates all the characteristics of human disease. Similar to the approach for emphysema, researchers have attempted to instill fibrogenic agents intratracheally in various experimental models. While these approaches attempt to elucidate some mechanisms of fibrogenesis, it is uncertain whether they really can be extrapolated to human disease. Nevertheless, our understanding of the basic pathology of the development of fibrosis has increased dramatically thanks to many of these models. A number of pioneer studies demonstrated that mice and other species were susceptible to the development of fibrosis after bleomycin (a profibrotic drug) instillation.

These models have been well studied; however, they do not resemble human idiopathic pulmonary fibrosis. Paraquat-induced fibrosis has also been well described in experimental models, but the lesions tend to be heterogeneous. With the development of transgenic technology, induced abnormal expression of extracellular matrix proteins, cytokines, and proteases has provided additional information. Although some of these models mimic pulmonary fibrosis, interpretation of the results requires a critical analysis. At the present time, it is impossible to reproduce the natural history of human pulmonary fibrosis in an experimental model. Recent advances such as microarray analysis, however, are providing better scrutiny of the initial mechanisms of disease. For example, it has been shown that bleomycin alters the gene transcription pattern in the mouse lung by increasing genes associated with inflammation, which reached maximum levels at 5 days after bleomycin administration, whereas genes involved in the development of fibrosis increased gradually up to 14 days after bleomycin treatment. These changes in
gene expression signature were well correlated with observed histopathological changes. This approach will provide a sensitive method to assess gene expression and may help identify genes involved in clinical pulmonary fibrosis.

An interesting example of transgenic mice that help elucidate mechanisms of development of glomerular lesions was published by Brantley et al. They found that increased expression of Cux-1, a murine homeobox gene, results in cell proliferation and mesangial expansion. In addition, these changes are potentially related to disruption of podocyte architecture leading to loss of filtration. These results suggest that expression of Cux-1 is sufficient to induce the early events of mesangio proliferative glomerulonephritis.

A large number of animal models of neoplastic diseases have been developed. Urethane-induced lung tumors in some strains of mice is a well-known model that has been studied for many years, and the mechanisms for the increased susceptibility of some mouse strains have been studied with and without transgenic manipulation. There are, however, many complications in establishing animal models of lung cancer, among them a good correlation between histological patterns of human and animal malignancies, natural strain susceptibility, and time frames.

The first oncogene targeted specifically to the lung was the Simian virus large T antigen (Tag). A model of pulmonary adenocarcinomas has been produced in transgenic mice harboring a chimeric gene comprising the SV40 large T antigen under control of a transcriptional region derived from the human surfactant protein C (SP-C) gene. In these studies, transgenic mice succumbed with pulmonary tumors within 4–5 months of age. By histology, the tumors were adenocarcinomas with lepidic, papillary, and solid growth patterns that were indistinguishable from adenocarcinomas occurring in humans. This model has been useful in our understanding of regulatory pathways disrupted during tumor progression.

SP-C promoter has also been used to express c-myc, epidermal growth factor, the receptor d’origine nantais (RON), receptor tyrosine kinase (a member of the MET proto-oncogene family), and Raf-1. These models have been very useful for specifically targeting certain cell types such that other lung cell types are not directly affected. Animal models for squamous cell carcinomas and small cell carcinomas are needed. Failure to develop specific tumor types is probably the result of variability of the transgene expression early in lung development. Another limiting factor in the study of these models is that once transcription of the transgene is initiated, it is irreversible. Apparently, conditional transgenic models have overcome these limitations. The ligand-inducible binary transgenic systems provide effective regulatory models; they consist of at least two transgene constructs, a regulatory transgene and a target transgene to provide regulated expression of a specific gene. The regulator transgene encodes a transcription factor whose activity is determined by the administration of an exogenous compound.

The regulator is placed under the control of a tissue-specific promoter to express the transcription factor in the tissue of interest. This regulator does not activate transcription of the target transgene until the animal receives an exogenous compound. When this compound is administered, the regulator activates only the target transgene. The other construct that contains the target gene contains the sequence of a protein of interest under the transcriptional control of cis-acting DNA elements that are responsive to the DNA-binding domain of the regulator transgene. The currently used ligand-inducible binary transgenic systems are the tetracycline transactivator inducible system, the mifepristone gene switch, and the adysese regulatory system.

### Knockout Mouse Models

Knockout mouse production implies mutation or ablation of an endogenous gene by homologous recombination in embryonic stem (ES) cells. Basically, the ES cells with the appropriate target are injected into the blastocyst of a mouse embryo; mice born contain cells from both the host embryo and the targeted ES cells. If these ES cells incorporate themselves into the germline, the mutation can be transmitted to future generations. There are limitations for the use of these models in lung cancer because the mutated or knockout gene either provides a silent phenotype or will not allow the study of new neoplastic transformation in the adult mouse.

Ligand-dependent Cre recombinase provides a clever mechanism to the production of controlled lesions, which is possible by using Cre recombinases. The technique depends on the Cre protein, which is encoded by the coliphage P1. It is a 38-kDa protein that efficiently promotes both intra- and intermolecular synopsis and recombination of DNA both in *Escherichia coli* and in vitro. Recombination occurs at a specific site, called lox, and does not require any other protein factors. The Cre protein causes synopsis of DNA and site-specific recombination in a mammalian cell line. Cre protein activity is directly regulated by a ligand, which binds to the Cre recombinase and causes changes in conformation that allow the recombinase to edit floxed genes. Cre, the product of the *Cre* (cyclization recombination) gene of bacteriophage P1, catalyzes the reciprocal recombination of genomic segments that are flanked by loxP sequences. The recombinase Cre acts on the DNA site *loxP*. If there are two *loxP* sites in the same orientation near each other, Cre can act to loop out the sequence between the two sites, leaving a single *loxP* site in the original DNA and a second *loxP* in a circular piece of DNA containing the intervening sequence. Therefore, a properly designed targeting construct containing *loxP* sites can be used for introducing subtle mutations or for a temporally or spatially controlled knockout (see also Chap. 1 in this volume). Although this technology appears promising, it has not been applied successfully to the study of lung malignancies.
Malignant mesothelioma has been linked to asbestos exposure and generally has a poor prognosis because it is often diagnosed in advanced stages and is refractory to conventional therapy. Mouse models of pleural and peritoneal mesotheliomas have been produced by exposure to asbestos fibers, radionuclides, particulate nickel compounds, and chemicals such as 3-methylcholanthrene. The role of SV40 virus as a cofactor with asbestos fibers in the development of diffuse malignant mesotheliomas in humans has also been explored in animal models. Some models have shown that SV40 virus alone induces mesotheliomas in hamsters. Because human malignant mesotheliomas frequently show hypermethylation or deletions at the Cdkn2a/Arf and Cdkn2b gene loci and deletions or mutations at the NF2 gene locus, heterozygous Nf2 (+/−) experiments with mice exposed to crocidolite asbestos fibers have been published. These mice exhibited accelerated development of malignant mesotheliomas compared to wild-type littermates.

An interesting study by Altomare et al in a mouse model of mesothelioma demonstrated that mesotheliomas from asbestos-treated Nf2 (+/−) mice show somatic genetic changes, including homozygous deletion of the tumor suppressor genes p16(INK4A), p14(ARF)/p19(Arf), and/or p15(INK4B) that are very similar to events found in human malignant mesotheliomas. Moreover, they show that a similar reciprocal pattern of ARF loss versus p53 alteration is present in both mouse and human malignant mesotheliomas. Taken together, these data implicate a common set of cellular perturbations in both human and mouse malignant mesotheliomas. Thus, from these studies it seems that alterations of the p53/ARF and p16(INK4A) cell-cycle regulatory pathways and the AKT and p21-activated kinase-merlin signal transduction pathways are critical events that cooperate to drive malignant mesothelioma tumorigenesis in both human and murine malignant mesotheliomas. These findings are consistent with the view that cancer is a multistep process involving the accumulation of somatic genetic changes that enable tumor cells to override fail-safe mechanisms regulating normal cell proliferation.62,63

Conclusions

Animal models of lung disease have provided extraordinary information to help us understand human disease. They are powerful tools that enable the study of the mechanisms and natural history of human diseases. Several species have provided good models for certain diseases; however, murine models have several intrinsic advantages compared to other animal models including lower cost, easier maintenance, and rapid reproduction rate. Transgenic or knockout mice can be generated in the laboratory in a relative short time compared to other species. Nevertheless, anatomic and immunological differences between mice and humans mean that murine models have limitations that must be considered when interpreting the results obtained from experimental models and applying these to the pathogenesis of human diseases. The methodology is limited by a number of factors including species differences, lack of models that truly resemble human disease, and strain variations. Although transgenic and knockout mice have been used in research for many years, the sequencing of the mouse and human genomes and high-density DNA expression analysis has recently added powerful tools to researchers using animal models. Furthermore, new knowledge of certain pathways altered in some diseases, obtained from animal models, may provide an opportunity for pharmacological intervention.

References

17. Animal Models of Lung Disease


Introduction

The use of tissue cultures as a research tool to investigate the pathophysiological bases of diseases has become essential in the current age of molecular biomedical research. Although it will always be necessary to translate and validate the observations seen in vitro to the patient or animal, the ability to investigate the role(s) of individual variables free from confounders is paramount toward increasing our understanding of the physiology and their role in disease. Additionally, it is not feasible to conduct certain research in humans because of ethical constraints, yet investigators may still be interested in the physiological response in human tissues; in vitro characterization of human tissue is an acceptable choice.

Tissue culture techniques have been utilized extensively to investigate questions pertaining to physiology and disease. The isolation and propagation of human epithelial cells has allowed investigators to begin to characterize the interactions and reactions that occur in response to various stimuli. Moreover, the culture of other human tissue has allowed researchers to investigate pathological cascades as well as other physiological responses. Combining cultured endothelial cells and leukocytes together in vitro under laminar flow conditions has helped elucidate the critical interactions that occur in rolling and emigration of leukocytes during the inflammatory response. Cultured embryonic stem cells that had been subjected to various growth conditions have advanced our understanding of cellular differentiation and growth.

Numerous cancer cell lines have been established to investigate their responses to chemotherapy and determine their biological properties. Overall, the use of cultured human tissue has provided a windfall of information on the pathogenesis of diseases that affect the body. In light of this, we briefly discuss the history and principles behind the utilization of tissue culture. We then discuss the current use of tissue culture to examine many of the unanswered questions involved in physiology and pathology.

History of Tissue Culture

The technique of tissue culture is generally accepted to have arisen following the experiment of Ross Harrison, around the turn of the twentieth century. In 1907, Harrison began by adapting a previously established bacteriology technique, the “hanging-drop” method, to culture a frog neuron. In 1912, Alexis Carrel built upon this work by successfully culturing small tissue samples from an 18-day-old chick embryo heart, thereby becoming the first scientist to propagate mammalian cells in vitro. Carrel’s demonstration that cells could be passed 18 times, remain viable over 3 months, and continue to maintain cardiac rhythm was the first to show that cardiac tissues in vitro could retain normal characteristics for a prolonged period of time. These elegant studies, conceived by Carrel, initiated the modern day art of histoculture as it is now known.

Although only a small “sect” of researchers embraced early tissue culture as a methodology to investigate the pathogenesis of disease, it is appropriate to describe Carrel as the father of mammalian tissue culture. In fact, Sven Gard, in his presentation speech for the Nobel Prize in Physiology or Medicine in 1954, referred to tissue culture as a “tissue cult. . . with Carrel as their high priest.” In that year, Drs. Weller, Enders, and Robbins shared the prize for their work in propagating poliovirus in tissue culture. This work was the first Nobel Prize awarded in medicine and physiology for work accomplished primarily utilizing tissue culture as a methodology.

An additional important milestone in the use of tissue culture in biomedical research was the establishment of the first human cell line. In 1951, cervical cancer cells from Henrietta Lacks were cultivated into the first immortal cell line – “HeLa.” These HeLa cells are still one of the most widely used cell lines today. Since the 1950s, tissue culture has become firmly established as a mechanism to answer many questions in biomedical research. Today, tissue culture is widely used to investigate diseases that affect the body, and through this work we have been able to increase our understanding of the pathological cascades that occur in diseases, as well as normal physiologies.
Types of Tissue Culture

Tissue culture is a commonly used generic term for the in vitro cultivation of cells, attributed to the early cultures that generally consisted of heterogeneous cultures of crudely disaggregated tissues. Currently, many terms are used that can be encompassed by the term: organ culture, cell culture, primary explants, and ex vivo propagation all concern the in vitro cultivation of cells or tissues. Cell culture in general can be applied either to primary cells (e.g., those with a finite lifespan) or to cell lines (e.g., HeLa cells). Additionally, these cultures can be either a homogeneous or a heterogeneous group of cells.

Primary cell culture involves the isolation of cells from a tissue by disaggregation. Single-cell suspensions from tissues can be completed through either enzymatic digestion of extracellular matrix surrounding the cells – such as with ethylenediaminetetraacetic acid, trypsin, or collagenase – or mechanical disaggregation. These disaggregation procedures have the disadvantage of possibly injuring cells. If the cells of interest are adherent viable cells, they will be separated from nonviable cells when the medium is changed. Alternatively, viable cells can be separated from nonviable cells before culture by subjecting the single-cell suspension to density-gradient centrifugation (e.g., Hypaque). Primary cells have an advantage of possessing many of the biological properties that they possessed in vivo because they are not transformed. Primary cells, unlike cell lines, are not immortal and have only a finite survival time in culture before becoming senescent. Variant cells, however, as well as those obtained from neoplastic tissue, may proliferate infinitely, thus becoming immortal in vitro. This development will eventually allow the immortal cell to take over the culture and can be thought of as a cell line. In general, primary human cultures will survive for 30–80 passages in vitro, although this number is dependent on cell type, conditions, and possibly other unknown factors. Primary cells are widely used to examine the effects of toxins, infectious agents, or other cellular interactions that would not be feasible in vivo. Primary cells have a disadvantage of being a heterogeneous mixture of cells upon primary isolation, with the type of cell obtained generally a component of the disaggregation method used. The most common contaminant seen following isolation of primary cells are cells of mesenchymal origin (e.g., fibroblasts). However, advances have been made that allow the culture of homogeneous populations of cells. For instance, cell-surface molecules specific for the cells of interest may be tagged with monoclonal antibodies. Techniques such as fluorescence-activated cell sorting or the use of magnetic beads can be utilized to enrich the single-cell suspension for the cell type of interest. Additionally, some investigators have recently exploited unique characteristics of certain cells, such as the presence of P-glycoprotein or multidrug resistance-associated proteins expressed on endothelial cells, to poison other contaminating cells in culture.

Another type of primary cell culture are “primary explants.” This type of culture is not subjected to a disaggregation procedure as is the primary cell technique described earlier. Therefore, single-cell suspensions do not occur. Briefly, tissue samples are dissected and finely minced. These tissue pieces are then placed onto the surface of a tissue culture plate. Following plating of tissue pieces, cells have been shown to migrate out of the tissue and onto the tissue culture surface. This technique is useful when cells of interest may become damaged or lost in the disaggregation technique described earlier and is often used to culture human bronchial epithelial cells.

Cell lines are another useful source of cells to investigate questions in biomedical research. These cells have the advantage of being immortal as opposed to the finite lifespans that primary cells possess. Additionally, they are generally well studied and characterized, leaving few experimental variables to worry about. These cells however, are prone to dedifferentiation, a process by which they lose the phenotypic characteristics of the cell from which they began. Many of the early cell lines were established from tumor tissue and as such possess abnormal growth characteristics. Newer cell lines have been established by molecular techniques such as inserting a telomerase gene into a cell to allow it to replicate infinitely. Because of the phenotypic changes that allow cell lines to replicate infinitely in culture, they are often a first choice for experiments; however, they are also highly criticized in light of their nonnatural phenotype.

Organ culture, as the name implies, involves ex vivo culture of the whole or significant portion of the organ. The main advantage to this type of culture is the retention and preservation of the original cell-cell interaction and extracellular architecture. This type of culture may be particularly important when experimental design necessitates the use of an ex vivo system, but researchers still need to retain the original organ architecture to answer questions posed. These types of cultures do not grow rapidly, however, and are therefore not suitable for experiments needing large numbers of a particular cell type.

Advantages and Limitations of Tissue Culture

Tissue culture has become the penultimate tool of the reductionist biologist. The utilization of tissue culture as a research methodology has allowed investigators to study isolated interactions in its near-normal environment. These experiments by their very nature introduce artifacts; however, they do minimize the number of confounding variables that may affect a particular experiment. For instance, tissue culture allows investigators to determine the effects of one particular treatment on a particular cell type, which would not be feasible in vivo. Additionally, tissue culture models
of disease allow investigators to obtain samples and make observations more readily than those done in vivo. However, it is the relative simplicity of experiments done in vitro that allows models of disease or physiology to come under frequent and warranted criticism. These models do not take into consideration the complexity of biological systems. Diminishing possible confounding variables by culturing cells in vitro brings up the constant criticism of how applicable results are because of alterations of the normal cellular environment in vivo. For example, cell–cell interactions in vitro are reduced and unnatural. Moreover, the culture does not contain the normal heterogeneity and three-dimensional architecture that are seen in vivo. This said, however, tissue culture biology has proved to be successful in many ways.

Cell Culture and the Study of Disease Processes

We have discussed the advantages that experimental systems using tissue culture afford researchers studying physiology and pathogenesis. Because of its ability to isolate individual variables and determine their role(s) in physiology, cell culture has become an integral tool in deciphering the pathological cascades that occur in human disease.

Many diseases that affect humans are multifactorial. This begs the question how can cell culture, because of its reductionist nature only dealing with a minimal number of variables, help to solve the unknown questions and decipher the components involved in disease? Often, clinical observations, and the questions arising therein, have been the launching pad for investigation.

For instance, observations of massive inflammation in the bronchoalveolar lavage samples of patients with acute respiratory disease syndrome (ARDS), consistent with damage seen in histological samples, prompted investigators to determine the role(s) of inflammation in the etiology of ARDS. Through the use of cell culture, investigators were able to determine individual interactions that occurred in the disease process. Investigators have utilized culture models employing microcapillary endothelial cells under flow conditions to understand the role of proinflammatory cytokines in the cytokinesis and emigration of neutrophils in disease. Using a model of pulmonary endothelium under flow conditions allowed investigators to demonstrate the importance of certain proinflammatory cytokines in ARDS.12

The role of inhaled toxicants in lung injury, and the mechanism(s) by which they cause disease, is another area of investigation that has utilized cell culture. Scientists have developed diverse and unique tissue culture systems that contain air–liquid barriers of lung epithelium and subjected these cells to various gaseous toxicants to determine what occurs following inhalation of various chemicals. Utilizing these types of systems, investigators are able to control the exposure time and other variables that may be difficult when determining inhaled toxicant effects in vivo. Moreover, the use of tissue culture, as opposed to an animal model, allows investigators to observe effects kinetically, without undue changes (e.g., sacrifice) and expense in the experimental model.12

A tissue culture model also permits an investigator to observe multiple changes in real time, such as cellular integrity, cell signaling and intracellular trafficking, protein expression changes, and oxidant-induced cellular damage. Deciphering each of these changes in an animal model would be extremely difficult; by employing a tissue culture model, researchers are able to tightly control the experimental system while isolating the events of interest. Further examples of how tissue culture models are currently being used to elucidate questions in physiology and disease are discussed later.

Biology of the Cultured Cell

Culture Environment

Maintaining cells in vitro was initially a very difficult task. Many characteristics need to be fulfilled before a successful cell culture occurs. Some of these characteristics are dependent on the type of tissue being studied; others may depend on specific requirements of the individual cells. Various chemically defined media are now available commercially to support the growth and differentiation of numerous cell types. The creation of defined media has allowed investigators to culture a multitude of cell types while controlling the local environment to answer pertinent questions. For example, glucose can be removed from a culture medium to study its effects on cellular metabolism, relative position in the cell cycle, and many other effects. Each chemical component is known in these media. Additionally, investigators can add growth factors to nourish their cell cultures.

The medium chosen when culturing cells in tissue culture must fit two main requirements: (1) it must allow cells to continue to proliferate in vitro, and (2) it must allow the preservation of the certain specialized functions of interest.8 The most common medium formulations used currently are Dulbecco’s modified Eagle’s medium, minimum essential medium, RPMI 1640, and Ham’s F-12. Occasionally, investigators develop new medium types to attain a formulation that optimizes their own experimental conditions. Fetal bovine serum is a common additive to most tissue culture media, although some investigators choose to forgo this additive for more defined supplementation. Additionally, others may choose sera from other sources such as human serum when culturing cells of human origin. Inactivation of complement by heat treating serum for 1 h at 56°C was initially very popular in tissue culture. However, it has become clear that this treatment may in fact damage some of the proteinaceous growth factors present in the medium, rendering it less effective. Currently, many experts
recommend heat inactivation only if the cell type of interest is particularly sensitive to complement.15

When deciphering if the current culture conditions are sufficient for the experimental design, the investigator must determine which cellular characteristics are important. Not only are the general characteristics, such as adhesion, multiplication, and immortalization of cell types important, but so are tissue-specific characteristics.

Cell Adhesion

Nearly all normal or neoplastic human epithelial cells will attach with relative ease to tissue culture surfaces. However, for culture cells that may loosely adhere, or may not adhere at all, scientists coat tissue culture surfaces with extracellular matrix proteins. Incubating tissue culture surfaces with serum, as well as laminin, fibronectin, or collagen, before culture has been shown to improve attachment of finicky cells.9 These treatments also help in replicating the normal attachment of cells to extracellular matrix proteins in vivo.

Development of Continuous Cell Lines

The development of continuous cell lines may be serendipitous, as was the development of early cell lines. In brief, many investigators would continue splitting primary cell cultures until one or more cell clones became immortal. Unfortunately, the changes that generally occurred in culture led to cells with abnormal phenotypes that had undergone dedifferentiation. Today, many investigators choose to use molecular biology techniques, exploiting our current knowledge of oncogenic viruses and enzymatic processes of cellular aging to transform primary cells in vitro to an immortal phenotype. It is known that the large T antigen present in the SV (Simian virus) 40 virus is capable of transforming cells to an abnormal phenotype.12,14,15 Moreover, transfection of primary cells with a transposase enzyme has also been shown to induce an immortal phenotypic change while preserving most normal cellular functions and phenotypes.12

Dedifferentiation

A commonly encountered problem in tissue culture is dedifferentiation. This loss of phenotype may be insignificant to the research at hand or it may be critical, and it must be dealt with on a case-by-case basis. When a cell culture undergoes dedifferentiation, it is often unclear whether undifferentiated cells took over the culture of terminally differentiated cells or whether a primary cell of interest became immortal under the culture conditions.

Functional Environment

The functional environment in which cells are cultured is critical when correlating experimental results to those seen in vivo. We previously alluded to the importance of the environment in which cells are cultured when discussing the advantages and limitations of tissue culture. Investigators have frequently striven to replicate integral in vivo environments in vitro to increase the significance of their experimental results.

The development of cell culture insert wells (e.g., Transwells, Corning) has allowed investigators to culture bronchial or alveolar epithelial cells at an air-liquid interface. This ability allows investigators to begin to replicate a significant aspect of these cells’ functional environment in vitro, thereby increasing their understanding of the effects of gaseous particles on pulmonary epithelial cells. Alternatively, scientists have also cultured epithelial cells on a roller bottle apparatus. This method allows investigators to determine the amount of time the apical epithelial cell surface is in contact with the air.

Capillary cell cultures have also come under frequent criticism when cultured in a monolayer in a tissue culture plate. Investigators have been able to utilize gel matrices in which capillary cells form tubule-like structures, more closely replicating the architecture these cells maintain in vivo. Additionally, endothelial cells are constantly under flow conditions in vivo. Addressing this condition in vitro has allowed investigators to look at the role of endothelial cells during inflammation, helping to increase the understanding of the role endothelium plays in various diseases.

At times, researchers may also choose to determine the effects of soluble factors (e.g., cytokines, hormones, neurotransmitters) from acute patients or animal models in a cell culture model. The milieu of soluble factors present in the serum that may play a role in a disease state is considerable. Moreover, these factors may have actions alone that are different when combined with other soluble factors. Reconstituting every factor presents a difficulty in vitro and leaves the possibility that an unknown factor may be missing. To address this, investigators have harvested sera from patients or animal models and used these samples as additives in their media formulations. For instance, through the use of serum samples from an animal model of smoke- or burn injury-induced acute lung injury, investigators have demonstrated that use of arteriovenous CO2 removal in acute lung injury significantly reduces apoptotic cell death in epithelial cells.16

Epithelial Cell Culture Models

Epithelial cells provide many important and equally diverse physiological roles in vivo. Toward the end of replicating these roles in vitro, investigators have collected numerous epithelial cells, both primary and immortalized, to help advance our understanding of the physiological significance.
**Gastrointestinal Epithelial Cell Cultures**

Results from a three-dimensional organotypical tissue culture experiment from adult murine colon performed by Bareiss et al.\(^7\) showed that this model could be useful in studying epithelial cell–cell interactions, cellular signaling, and microbiological infections. Primary cell cultures of bovine colon epithelium can further be used in toxicological studies in vitro as a model for the colon epithelium.\(^18\)

**Hepatocyte Tissue Cultures**

Hepatocytes play an integral role in the metabolism of numerous drugs, nutrients, and other constituents in vivo. In vitro culture of hepatocytes has facilitated extending our understanding of the physiological steps important in mediating these physiological functions. Additionally, in vitro culture has allowed investigators to understand the importance of viral infection, drug interactions, and other physiological signals that alter metabolism and normal cellular function.

**Hepatocytes in Drug Metabolism**

Studies performed by Nussler et al.\(^19\) concluded that the use of a collagen sandwich or three-dimensional membrane bioreactor can be used to study drug metabolism of human hepatocytes. Additionally, to study drug metabolism in human hepatocyte cells, Zeilinger et al.\(^20\) developed a three-dimensional bioreactor culture model of liver cells under continuous medium perfusion with decentralized mass exchange of integral oxygenation.

**Hepatocellular Carcinoma**

Zen et al.\(^21\) using tissue culture methods, discovered support for the maintenance of the cancer cell hierarchy in human hepatocellular carcinoma. Utilizing a radial flow bioreactor system, Kosuge et al.\(^22\) prompted differentiation of human hepatocellular carcinoma cells in three-dimensional culture. The resulting differentiated cells could be useful for improvements in genetic or pharmacological reinforcement and for the monitoring of bioartificial livers.\(^22\)

**Lymphatic Tissue Culture**

The complexity of the lymphatic system has made the engineering of functional tissues difficult, but progress has been made in the engineering of lymphatic capillaries and the development of a bioreactor designed to culture lymph nodes.\(^26\)

**Pulmonary Endothelial Cells**

Pulmonary endothelial cells represent a unique type of endothelium because of their paradoxical responses to hypoxia. This uniqueness highlights the need to utilize cell culture models of pulmonary endothelium as opposed to other endothelia when interested in investigating their role(s) in pulmonary physiology. Several investigators have described the isolation and culture of pulmonary endothelial cells. Persistent pulmonary hypertension of the newborn, also known as neonatal pulmonary hypertension, is caused by a disorder of the pulmonary vasculature from fetal to neonatal circulation, culminating in hypoxic respiratory failure and death. The inciting events that culminate in neonatal pulmonary hypertension are multifactorial. Despite this, decreased production of vasodilator molecules such as nitric oxide and prostaglandin I\(_2\) in the pulmonary endothelium has been shown to be a critical component of disease progression.\(^27\)

**Infectious Diseases**

Infectious diseases play a unique role in lung pathology in light of their roles as either important contributors or consequences of many lung diseases. For instance, certain lung diseases may predispose patients to infection: patients afflicted with obstructive lung diseases, as well as cystic fibrosis patients, commonly suffer from severe and recurrent bacterial infections. Additionally, patients may become superinfected following a viral respiratory infection. Systemic infections, such as gram-negative bacterial sepsis, may lead to lung diseases such as ARDS.
Human Type II Alveolar Pneumocytes and Acute Lung Injury/Acute Respiratory Distress Syndrome

Pulmonary alveolar type II cells are a unique cell subset that carries out highly specialized functions which include synthesis and secretion of surfactant, a unique composition of lipoproteins that act to reduce surface tension at the alveolar air–liquid interface. Defining the molecular mechanisms leading to production of surfactant by type II pneumocytes is important in many disease processes.

The pathogenic sequence that results in ARDS, the most severe manifestation of alveolar lung injury, is generally thought to be initiated by a systemic inflammatory response. Despite this knowledge, there still exist many questions about the initial triggers and pathological steps that occur in ARDS. Greater understanding of these steps may help to develop new treatment regimes. Currently, treatment of ARDS consists of mechanical ventilation, which helps to stabilize blood gases. However, mechanical ventilation itself may provoke further inflammation in the alveoli, thereby decreasing compliance and gas exchange in the alveoli.

The cell type of particular interest in ARDS and diffuse alveolar damage is the type II pneumocyte. Until recently, studies trying to decipher the pathological sequence in acute lung injury have had to rely on standard lung epithelial cell lines. Recently, however, human type II alveolar epithelial cells (pneumocytes) have been successfully isolated from fetal human lung tissue by collagenase digestion. Briefly, fetal lung tissues were minced and incubated in a serum-free medium containing dibutylryl cyclic adenosine monophosphate for 5 days. The tissue explants were then treated with collagenase and incubated with DEAE-dextran to eliminate contaminating fibroblasts. Cells were then plated onto tissue culture dishes treated with extracellular matrix derived from MDCK cells and cultured overnight in Weymouth’s medium containing 10% serum. These steps resulted in relatively pure populations of human type II pneumocytes that were then cultured at an air–liquid interface. Using these methods, Alcorn et al were able to maintain a primary culture that retained the morphological and biochemical characteristics of type II pneumocytes for up to 2 weeks.

Three-Dimensional Biology

Conventional Bioreactors and Three-Dimensionality: The Origins of Three-Dimensional Culture

Carrel postulated that tissue development was linked to access to nutrient supply, noting that peripheral cells grew readily, and internal cells became necrotic, presumably based on their distance from the nutrient source. To circumvent this issue, Carrel implemented cultures on silk veils, preventing the plasma clots of the growth media from deforming or becoming spherical, thus facilitating the internal cell’s ability to obtain nutrient replenishment. Many attempts were made in standard culture systems (bioreactors) and other culture apparatuses to escape the constraints of two-dimensional cell culture, with the intent of yielding high-fidelity human and mammalian tissues, and thus emphasizing the need for development of three-dimensional biology.

Another famous researcher, Leighton, improved on Carrel’s techniques in the 1950s and 1960s. Leighton’s major contribution to three-dimensional culture technology was the introduction of the idea of a sponge matrix as a substrate on which to culture tissues. Leighton first experimented on cellulose sponges surrounded by plasma clots resident within glass tubes. He devised a system to grow tissue explants 1–5 mm² in area on sponges, using small amounts of chick plasma and embryo extract. After the mixture solidified on the sponge, Leighton added the nutrient media and inserted the “histoculture” in a roller apparatus to facilitate nutrient mass transfer. He experimented with many sponge combinations, discovering that collagen-impregnated cellulose sponges were optimal for sustaining the growth of native tissue architecture.

Leighton was successful in growing many different tissue types on the sponge-matrix cultures. Leighton also found that C3HBA mouse mammary adenocarcinoma cells, when grown on sponge-matrix histoculture, aggregated “much like the original tumor, forming distinct structures within the tumors such as lumina and stromal elements, and glandular structures.” An extremely important difference of this three-dimensional histoculture from the standard two-dimensional culture is the apparent quiescence of the stromal component and the balanced growth of these cells with regard to the overall culture. Leighton further advanced the concept of three-dimensional histoculture to histophysiological gradient cultures. These cultures are conducted in chambers that allow metabolic exchange between “the pool of medium and the culture chamber by diffusion across a membrane.” Histophysiological gradient cultures mimic, to some degree, diffusion in tissues.

From the pioneering work of Carrel and Leighton, other methods of emulating three-dimensional cultures have been developed, such as embedding cells and tissues in collagenous gels of rat tail as per the techniques of Nandi and colleagues. Clusters of mammary cells were initiated into cultures, and Nandi and colleagues observed three-dimensional formation of ductile structures and sustained growth over a period of several weeks concordant with proliferation at the edge of the cell mass with minimal fibroblast contamination. Thus, mammary tissue cultured in the collagen gels resembled, histologically, in vivo tissue, differing significantly from two-dimensional monolayer cultures. Nandi et al suggested an important criterion for successful growth of mammalian tissue-like structures in collagen gels, versus the
monolayer culture, was maintenance of mammmary cell shape in three dimensions as opposed to conventional monolayer culture. Many of the advantages of three-dimensional cultures seen by Leighton, Nandi, and others may be attributed to permitting the cells to retain their normal shape and special associations. This global concept will be important as we begin to understand and recall the physical and environmental characteristics of the rotating-wall vessel systems.

Other methods of three-dimensional culture encompass a technique known as organ culture or culture on a filter, a strategy developed by Strangeways and Fell and Robison. Tissue explants were grown on lens paper in a watch glass containing liquid culture medium. Browning and Trier found “that for some tissues, it is critical to keep the cultures at the air-liquid interface,” thus allowing the tissues to experience conditions similar to the in vivo environment.

Another strategy is the use of three-dimensional cultures known as proto-tissues, or aggregates of cells, used to form spheroids. This technique was popularized by Sutherland and colleagues more than 20 years ago when they manipulated aggregates of cells into a spherical configuration by spinning agitation of the cells in spinner flasks. This technique produced pseudo-tissue-like organoids useful for research evaluations. Each of these methodologies will be of benefit as we continue to examine strategies for achieving three-dimensional lung tissue constructs.

Bioreactors and three-dimensional systems have been and are used by medical researchers and by pharmaceutical production plants to create mimics of human, mammalian, microbial, and plant physiology. As reviewed, many approaches have been employed to create these synthetic physiologies; however, none of these has been entirely satisfactory for broad-based mammalian cell culture use. The impact of molecular genetic engineering in the early 1980s was for the most part confined to pharmaceutical proteins such as interferon, growth hormones, and insulin molecules. During processing of these valuable, but low-yield, products, emphasis was placed on downstream recovery rather than on bioconversion or bioprocesses, let alone three-dimensional tissue development. Advances in modern biotechnological product recovery are expanding to medium- and high-volume enzyme products and limited forms of bulk products. In the world of high-volume, high-value products, cell culture bioreactor productivity contributes an important role in determining the economic influence of bioprocessing. Ideal bioprocessing bioreactors will in future have high reactor productivity with an in situ separation capability of products.

Membrane bioreactors fit this description and already contribute remarkably to the bioprocessing industry. Membrane bioreactors are capable of retaining enzymes, organelles, and microbial, animal, and plant cells behind a membrane barrier, trapped in a matrix or adherent to the membrane surface. In 1963, Gallup and Gerhardt initially used the membrane bioreactor for dialysis culture of Serratia marcescens. Immobilized enzyme microencapsulation was pioneered by Chang but Butterworth et al first developed the enzyme membrane reactor to successfully accomplish starch hydrolysis with α-amylase. Similarly, for animal cell culturing, Knazek et al cultured human choriocarcinoma cells on compacted bundles of Amicon fibers. Many reviews on the particular applications of hollow-fiber and immobilized bioreactor bioreactors for enzyme catalysts, microbial cells, and animal cell culture are available.

Prolonged human residence in microgravity results in the atrophy of healthy skeletal muscle. Continued space exploration requires a better understanding of this phenomenon. Muscle wasting is likely caused by changes at the system level marked by increased circulating glucocorticoids and decreased circulating growth hormone and at local levels as seen by decreased myofiber resting tensions. Differentiated skeletal muscle cultures have served as model systems for gaining a better understanding of the roles of exogenous and endogenous cytokines and of muscle fiber tension in regulating muscle cell growth. This effort has led to the beginning of tissue engineering studies by the space science community. Engineering these tissues into three-dimensional bioartificial muscle (BAM) constructs has extended their utility for space flight investigations as a prelude to developing countermeasures for the microgravity environment. Serendipitously, these bioreactor models in part mimic diseases of the musculature such as dystrophy.

BAMs have been sustained in a viable condition for as much as a month in perfused hollow-fiber bioreactor cartridges. Further, investigators have documented that growth hormone and/or insulin-like growth factors are potential protein therapeutics, which may in part attenuate skeletal muscle degeneration in space as well as show promise for the treatment of muscle-wasting diseases of the terrestrial community.

An associated application of space bioscience to an Earth-based disease is the investigation of bone degeneration. The microgravity effect on bone decalcification has been documented beginning with early space flights and subsequently found to mimic conditions analogous to osteoporosis frequently seen in aging females. Over the past three decades, the relationship between cellular morphology and metabolism and the role of physical stress in bone loss has been studied. Physiomechanical stresses induce shape changes in osteoblasts, possibly mediated by reorganization of focal contacts. Thus, gravity variations (Gz) have been studied to understand their influence on osteoblast adhesion of ROS 17/2.8 rat osteosarcoma cells during 15–30 parabolic flights. Non trivial flight-induced cell shape changes consisting of decreased area concordant with contact plaque reordering have been observed; however identical periods of continuous mechanical stress induced by centrifugation (2 Gz) or clinostatic rotation (Gz randomization) had no discernible effect on cell adhesion. Synchronization of ROS 17/2.8 G2/M accomplished via treatment in nocodazole inhibited
the flight-induced decreases in adhesion parameters. Investigators thus concluded that ROS 17/2.8 cells possess the ability to sense Gz triggers and that adaptation is related to cytoskeletal function.59

As presented previously, tissue engineering applications of three-dimensional function and structure are well known in medical science research.60 In microgravity, three-dimensional aggregates form, facilitating the expression of differentiated organotypic assemblies. Investigations to determine the effect of composite matrices, spiked with esterified hyaluronic acid and gelatin, to augment osteochondral differentiation of cultured, bone marrow-derived mesenchymal progenitor cells, and the effects of the matrix on cellular differentiation have been examined in vitro and in vivo.60

Briefly, empty and populated matrices cultured for 28 days with and without transforming growth factor (TGF)-β1 demonstrated the following results. Cells implanted in the matrix produced a robust type II collagen extracellular matrix in vitro. Matrices placed in immunodeficient mice yielded no differentiation in empty constructs, osteochondral differentiation in loaded implants, and an enhanced level of differentiation in preimplantation in vitro-cultured matrices containing TGF-β1. These results demonstrate the utility of a three-dimensional matrix for presentation of bone mesenchymal progenitor cells in vivo for repair of cartilage and bone defects as well as indicate the efficacy for in vitro tissue engineering regimes.60 These techniques lend themselves to microgravity and ground-based research tissue cultures alike.

Many earth-based laboratories are researching and developing hematopoietic bone marrow cultures of stem cell origin, and three-dimensional configurations are providing promising results, as illustrated by Schoeters and coworkers.61 They report that murine bone marrow cells, cultured under long-term hematopoietic conditions, produce mineralized tissue and bone matrix proteins in vitro but only when precipitated by the presence of adherent bone stroma cells in three-dimensional collagen matrices. At a concentration of 8 × 10⁶ stromal cells, mineralization occurs in 6 days. In contrast, two-dimensionally oriented marrow fragments at 1 × 10⁷ cells require requires more than 10 days before mineralization can similarly be detected.61

Two-dimensional long-term marrow culture facilitates and enhances expansion of the stromal component and rudimentary differentiation of osteogenic-like cells in the adherent stromal layer, as verified by type I collagen or cells positive for alkaline phosphatase. Production of osteonectin and osteocalcin, a bone-specific protein, combined with calcification is observed only in three-dimensional cultures. These studies demonstrate the need for and benefit of three-dimensionality and the application to the microgravity environment.61 As we can see, this further reinforces the quest for three-dimensionality and the potential of modeling the microgravity environment.

Three-Dimensional Models for Physiological Study

Investigations clearly show the need for the application of three-dimensional study techniques in pathophysiological studies. Interestingly, three-dimensional biology has facilitated full-scale investigations into most areas of tissue engineering, cell biology and physiology, immunology, and cancer research.

Several important factors are necessary when considering the successful recapitulation of a three-dimensional mammalian tissue. These factors include specific elements of differentiation which, when replicated accurately, result in the functionality of the original tissue being mimicked. Figure 18.1 graphically represents the stages of “assembly” and differentiation needed to result in a functionally accurate tissue. These stages are three-dimensional assembly, three-dimensional growth, cellular matrix formation [extracellular matrix (ECM) and basement membranes], differentiation (associated with cellular specialization), and vascular formation (or pseudo-vasculature).

To accomplish the successful recapitulation of an ex vivo tissue, a series of defined steps are necessary. The creation of a three-dimensional structure is accomplished by systematic tissue engineering of the desired tissue architectures. Figure 18.2 demonstrates the assembly process and illustrates some of the variables that may be employed in the engineering process.

The assembly of complex functional mammalian tissues in 1× gravity is problematic because of the effects of shear stress, turbulence, and inadequate oxygenation in conventional cell culture systems. This study describes the culture and three-dimensional assembly of baby hamster kidney (BHK) mammalian cells on microcarriers under controlled oxygenation, low shear stress, and turbulence in the NASA-designed integrated rotating-wall vessel (RWV).62–64 Anchoragedependent cells are widely cultured on microcarriers.65 Studies show that for the purposes of improved surface-to-volume ratio and scale-up, the microcarrier suspension culture provides excellent potential for high-density cell growth.66 In addition; microcarriers serve well as structural supports for three-dimensional assembly, the composite of which is the basis for three-dimensional tissue growth.67

Conventional culture systems for microcarrier cultures (i.e., bioreactors) use mechanical agitation to suspend microcarriers and thus induce impeller strikes as well as fluid shear and turbulence at the boundary layer between the wall and the fluid. Investigators have attempted to make a complete study of the most efficient bioreactor designs and agitation regimens.68 They concluded that virtually all stirred-tank bioreactors operate in the turbulent regimen. It has been demonstrated that bead-to-bead bridging of cells is enhanced significantly at lower agitation rates in a stirred reactor.69 Aggregates of as many as 12–15 microcarriers
Fig. 18.1. Stages of cellular differentiation requisite for tissue functionality.

Fig. 18.2. Development and tissue assembly process for engineering of a three-dimensional tissue.
bridged with bovine embryonic kidney cells (BEK) have been reported. Excessive agitation from either stirring or gas bubble sparging has been documented as a cause of cell damage in microcarrier cell cultures. To overcome the problems induced by these mechanisms, investigators developed alternative culture techniques such as porous microcarriers to entrap cells, increased viscosity of culture medium, bubble-free oxygenation, and improved methods for quiescent inoculation. These steps decreased the damage attributed to turbulence and shear forces but failed to significantly rectify the problems. Reactor systems of substantially increased volume exhibit less agitation-related cell damage, presumably because of the decreased frequency of cell-microcarrier contact with the agitation devices in the systems. Research-scale investigations do not afford the luxury of experimenting with large-scale production systems. Therefore, if a large-volume system is indeed more quiescent, an improved bioreactor system should emulate the fluid dynamics present in the upper regions of large-scale reactors in which cells and microcarriers reside with minimal agitation. Microcarriers and cells in large-volume reactors are semibuoyant, rendering them practically weightless, a condition that we hypothesize may be analogous to the environment of microgravity. It is stated that “Unless a cell culture is growing in an environment free of gravitational forces, moderate levels of agitation are required to suspend microcarriers that are not neutrally buoyant.” The ability to grow cells in a rudimentary horizontally rotated chamber has been previously demonstrated and is intended to simulate in the laboratory the effects of weightlessness or microgravity on cells. It was shown that human embryonic kidney (HEK) cells attach to microcarriers in microgravity (Space Shuttle experiment) and that attachment and spreading may be enhanced in a weightless environment. Review articles of cell biology performed in space flight experiments delineate the growth of many cell types in an environment devoid of gravitational influence. The problem, then, is to suspend microcarriers and cells without inducing turbulence or shear while providing adequate oxygenation and nutritional replenishment. One environment that possesses these attributes is the microgravity of space flight. Another is the system referenced here, which randomizes the forces of gravity by classical methods, thus simulating some aspects of microgravity.

The term rotating-wall vessel (RWV) comprises a family of vessels, batch fed and perfused, that embody the same fluid dynamic operating principles. These principles are (1) solid body rotation about a horizontal axis that is characterized by (a) colocation of particles of different sedimentation rates, (b) extremely low fluid shear stress and turbulence, and (c) three-dimensional spatial freedom; and (2) oxygenation by active or passive diffusion to the exclusion of all but dissolved gasses from the reactor chamber, yielding a vessel devoid of gas bubbles and gas-fluid interface (zero head space) (Fig. 18.3).

Fig. 18.3. NASA Rotating-Wall Bioreactor. United States Patent: Schwarz et al. (Courtesy of NASA).
The baby hamster kidney (BHK) model (Fig. 18.4) described and illustrated here has been used extensively as a means of transfecting viral and bacterial DNA into cells to accomplish gene splicing protocols. These three-dimensional models are proving extremely valuable in viral, microbial, and genetic research.

Normal Tissue Models

As reviewed in several of the previous sections, investigators have cultured bone, cartilage (chondrocytes), muscle tissue, cells of the immune system, and others. Baker and Goodwin in 1997 employed the RWV to culture chondrocytes (Fig. 18.5) for 36 days to observe the influence of low-shear and quiescent culture conditions on the development of three-dimensional differentiation and extracellular matrix formation in bovine chondrocytes.

Chondrocytes from bovine cartilage were inoculated into the RWV with 5 mg/ml Cytodex-3 microcarriers in non-adherent petri dishes with identical conditions and were initiated with microcarriers as standard tissue culture controls. Differentiated chondrocytes were observed in all sections of RWV material through 36 days, while few were observed in the sections of petri dish material. These results indicate that the unique conditions provided by the RWV afford access to cellular processes that signify the initiation of differentiation as well as production of normal matrix material. These experiments demonstrated the ability to create a synthetic model of cartilage or bone without leaving Earth’s orbit.
Figure 18.6 illustrates a 34-day time course of three-dimensional culture of normal human chondrocytes (HCH). Increased Actin (alexa 555 phalloidin; red) and nuclear staining (syto16; green) is observed over the time course, resulting in yellowish-red staining by day 34. Goodwin et al.\(^{63}\) describe the results of experiments where normal human small intestine epithelium and mesenchyma were cultured in rotating-wall vessel cultures. They found that the rotating-wall vessels are a major advance toward constructing a functionally accurate, large-scale, in vitro, three-dimensional (3-D) tissue model of the small intestine. Because of the presence of large tissue-like masses that express differentiated epithelial, mesenchymal, and developing endothelial cells, a multitude of possibilities are afforded for cell biological investigations, and further, functional epithelial cell brush borders with extracellular matrix and basal lamina components that represent ordering of tissue and cellular polarity were nurtured by the molecular conditions and physical orientations of the culture system. An important additional finding was that this 3-D model demonstrates a significantly diminished requirement of complex culture media, which suggests specific cell-cell interactions and the production of paracrine and autocrine factors essential to the growth and development of these fragile tissues.\(^{63}\)

The role of basement membranes and extracellular matrix and their relationship to epithelial mesenchymal development and differentiation are the subjects of considerable research. Current studies indicate that the stromal component exerts a strong and driving influence over developing intestinal mucosa.\(^{82–84}\) Stallmach et al.\(^{85}\) have shown that only organ-specific mesenchyme will produce differentiation in epithelium from a given organ site and that embryonic mesenchyme of the same age but from different organs was ineffective. Studies of rat small intestine development have detailed a shift in the membrane molecular components of crypt cells as measured by monoclonal antibody binding in fetal versus newborn rat epithelium. Quaroni\(^{86}\) postulates that the presence of specific markers in the crypts indicates undifferentiated crypt cells, which may be able to perform specific intestinal functions similar to that of the villus cells. Additional information shows that a single fetal epithelial cell type in the last 2–3 days of gestation may express the function of more than one cell, thereby giving rise to DNA synthesis and proliferation.\(^{87}\)

Cocultures of small intestine produced in the RWV were initiated with adult epithelium composed of dividing and terminally differentiated cells and predominantly adult mesenchymal cells with one exception, a 2-month-old female
Levels of epithelial cell (keratin-positive) and mesenchymal cell-specific (vimentin-positive) immunostaining decreased slightly but consistently over the course of the RWV experiments, while total numbers of cells increased. Levels of endothelial cell (Factor VIII-positive) and epithelial cell (villin-positive) staining remained reasonably constant, while staining for angioblasts (endothelial precursor cells) developed toward the end of each experiment. Functional brush border markers such as sucrase were also present in this coculture system. Other brush border-specific monoclonal antibodies, kindly supplied by Dr. Andrea Quaroni (Cornell University, Ithaca, NY), have similarly proved positive in more recent studies (unpublished results). Collectively, these results indicate that the model, although not histotypically correct, exhibits many facets of functional normal small intestine. Growth conditions, however, may still be skewed slightly in favor of the mesenchymal versus the epithelial cell types. Work will continue on further definition of the correct nutritional and growth-matrix requirements directed toward attaining the best representation possible. Interestingly, all the RWV cocultures exhibited laminin, fibronectin, and type IV collagen production, as well as large amounts of proteoglycan. Hahn states that only undifferentiated, highly proliferative intestinal epithelial cells (usually fetal) synthesize the aforementioned proteins. Of these, only laminin appeared to promote differentiation of intestinal epithelial cells. Type IV collagen and fibronectin had no effect.

The literature reviewed and the data obtained here would suggest that this model of human small intestine embodies many aspects of differentiation observed in other in vitro and in vivo cell and organ models. Primary distinctions would be (1) the overall scale of the model, (2) the ability to culture epithelium for long periods (in excess of 40 days) without loss of functional cell markers, and (3) the ability of the system to respond to extensive analyses and manipulations without the termination of a given experiment.

The significance of a complex three-dimensional in vitro culture system for the growth of normal small intestine should not be underestimated. The molecular basis and clinical treatment of diseases such as inflammatory bowel disease (Crohn’s, ulcerative colitis), malabsorptive syndromes (short-gut syndrome), numerous infectious diseases, and tumors of the small bowel may be investigated with the advent of this new technology. For example, the recent demonstration that the human immunodeficiency virus can replicate in human small intestine and columnar epithelium may be impacted through studies of virus-cell interactions in the RWV culture system. This hypothesis is currently being tested. Additionally, general application of this culture model may lead to advances in understanding growth and differentiation in developing organisms and the potential treatment of a myriad of clinical conditions as well as tissue renewal (Figs. 18.7 and 18.8).

The RWV systems have been demonstrated as useful for the development of sophisticated models, which emulate facets of...
microgravity and at the same moment present characteristics of in vivo tissues. We have examined the production of normal tissue models for research and experimentation; as we will see in the next section, the RWV has the ability to serve as a tremendous tool to construct synthetic tumor tissues, which allows an in-depth understanding of tumor biology and oncology.

Three-Dimensional Models of Lung Disease

Current cell culture models have shortcomings resulting in unreliable tumor growth, uncharacteristic tumor development, nonhuman tumors, and inadequate methods of detection. Cells propagated under traditional culture conditions differ widely in their expression of differentiated markers, adhesion receptors, and growth factor receptors compared with cells in situ or those grown as tissue-like structures. This factor is of concern because the phenotypic changes leading to malignant transformation often stem from alterations in the balanced and multifaceted roles of growth factors, receptors, and cytokines (reviewed by Herlyn et al. With increasing evidence of the importance of adhesive contacts, paracrine cross-talk between different cell types, and signaling cascades that link the cell with a complex substratum, there is now recognition that models must be developed that better simulate these complexities. There is still much to learn about the dynamic relationships among the different phenotypes found in the normal lung and in lung cancers. Until a cell culture system is developed that allows differentiation to occur, it is difficult to make any conclusive statement about relating effects in cell culture to clinical practice. Tissue engineering is very embryonic in development and currently nearly universally focused on building replacement tissues. A new technology developed at the NASA Johnson Space Center used to study colon cancer has been adapted to three-dimensional in vitro lung tissue culture models but has not been reported on to date.

Fig. 18.7. Early-stage scanning electron micrographs of human small intestine RWV cocultures. (From Goodwin et al. Reprinted by permission of the Society for Experimental Biology and Medicine.)
Rotating-wall vessels are horizontally rotating cylindrical tissue culture vessels that provide controlled supplies of oxygen and nutrients with minimal turbulence and extremely low shear. These vessels suspend cells and microcarriers homogeneously in a nutrient-rich environment, which allows the three-dimensional assembly of cells to tissue. Before seeding rotating-wall vessels (Synthecon, Houston, TX, USA), cells were cultured in standard T flasks (Corning, Corning, NY, USA) in GTSF-2 medium (1993 PSEBM) in a humidified 37°C, 5% CO₂ incubator. The rotating-wall vessels were seeded with 1–2 mg/ml Cultispher-GL microcarriers (Hyclone Laboratories, Logan, UT, USA) followed by BEAS2-B or BZR-T33 cells (ATCC, Baltimore, MD, USA) at a density of 2 × 10⁵ cells/ml. Cultures were grown in the rotating-wall vessels for 14–21 days for formation of 3- to 5-mm-diameter tumor masses. Rotating-wall vessel rotation was initiated at 25 rpm and increased as aggregate size became larger. Stationary control cultures were initiated under the same conditions using FEP Teflon bags (American Fluoroseal, Columbia, MD, USA). At 24-h intervals pH, dissolved CO₂, and dissolved O₂ were determined using a Corning 238 model clinical blood gas analyzer. Glucose concentration was determined using a clinical glucose analyzer. Cell samples were harvested every 48 h and fixed for immunohistochemistry or for scanning electron microscopy.

Cancer models already developed by NASA investigators include growth and differentiation of an ovarian tumor cell line, growth of colon carcinoma lines, and three-dimensional aggregate and microvillus formation in a human bladder carcinoma cell line. In support as an appropriate model for cancer, even the most rudimentary three-dimensional cellular structures exhibit different phenotypes than cell lines cultured under two-dimensional conditions. Properties such as responses to TGF-β, drug resistance to cisplatin or cyclophosphamide, and resistance to apoptosis are all altered in various types of cell aggregates.

Many investigations sustain consistent evidence that cells growing in three-dimensional arrays appear more resistant
to cytotoxic chemoagents than cells in monolayer culture. Li et al. found that spheroids were more resistant to cytosine arabinoside by 11-fold and methotrexate by 125-fold when compared with single-cell suspensions. Further monolayer cultures of colon carcinoma cells were sensitive to piericidin C in contrast to responses within in vivo colon tumors or three-dimensional slices of tumors grown in vitro. Numerous other investigations have revealed increased levels of drug resistance of spheroids compared with single-cell monolayers.

Questions of poor diffusion and insufficient drug absorption within spheroids and a relatively frequent high proportion of resting cells have clouded differences in drug resistance, which could be the result of nutrient deprivation and hypoxia. Heppner and colleagues executed precise experiments that confirmed three-dimensional structure and function as the causative agent which was responsible for drug resistance rather than simple inaccessibility to nutrients or the drug concentration. Heppner embedded tumor specimens or cell aggregates in collagen gels, exposed the culture to various cytotoxic drugs, and compared the drug responses of the same cells in monolayers. These experiments revealed an increased resistance in the three-dimensional tumor arrays of a remarkable 1,000-fold greater than in monolayer cultures, and a similar result was seen in three-dimensional histocultures in collagen. The tumor cells grew in the presence of drug concentrations that rendered monolayers to a viability less than 0.1% of control cultures. Amazingly, Heppner observed that the cells became sensitive again when replated as monolayers, and finally showed that even when exposed to melphalan and 5-fluorouracil in monolayer culture, cells transferred to collagen gels were again resistant based on three-dimensional architecture. Thus, the cells were exposed to the drugs as monolayers, facilitating access to the drugs, and, once the cells were transferred after drug exposure to a three-dimensional structure, high resistance to the drugs was sustained.

Based on the caliber of data referenced above, Teicher et al. serially passaged through multiple (ten) transfers EMT-6 tumors in mice that were treated with thiotaenia, cisplatin, and cyclophosphamide over a prolonged 6-month period, thus producing extremely drug-resistant tumors in vivo. When these tumors were grown as monolayer cultures, they were as drug sensitive as the parental cells. Kobayashi and colleagues grew the same in vivo drug-resistant tumor cell lines as spheroids in three-dimensional arrays, and resistance was almost 5,000 times that of the parent line with selected drugs, an example being the active form of cyclophosphamide used in vitro. Similarly extreme resistance was also observed to cisplatin and thiotaenia. This resistance was not seen in monolayer cultures, even when the monolayers were cultured on traditional extracellular matrix substrates. These experiments reconfirmed that cells in a three-dimensional array are more drug resistant than monolayer cells in vitro and demonstrated that three-dimensional cellular configurations can and do become resistant to super-pharmacological doses of drugs by forming compact structures.

Rotating-Wall Vessel Tumor Models

Several important human tumor models have been created in rotating-wall vessel cultures, specifically, lung, prostate, colon, breast carcinoma, and ovarian. Many of these models involve cancers that are leading killers in our society. We present two such examples in this section, colon and prostate carcinoma. As previously reviewed, the literature indicates the remarkable difference between chemotherapeutic cytotoxicity in two-dimensional and three-dimensional cellular constructs, which may be predicated on a number of criteria. Therefore, a three-dimensional tumor model that emulates differentiated in vivo-like characteristics would provide unique insights into tumor biology.

Colon Cancer

Goodwin et al. detailed the first construction of a complex three-dimensional ex vivo tumor in rotating-wall vessel (RWV) culture composed of a normal mesenchymal basal layer (as would be seen in vivo), and either of two established human colon adenocarcinoma cell lines, HT-29, an undifferentiated line, and HT-29KM a stable, moderately differentiated subline of HT-29, were grown in RWV culture. RWVs were used in conjunction with multicellular cocultivation to develop a unique in vitro tissue modeling system. Cells were cultivated on Cytodex-3 microcarrier beads, with and without mixed normal human colonic fibroblasts, which served as the mesenchymal layer. Culture of the tumor lines in the absence of fibroblasts produced spheroid-like growth and minimal differentiation. In contrast, when tumor lines were cocultivated with normal colonic fibroblasts, initial growth was confined to the fibroblast population until the microcarriers were covered. The tumor cells then commenced proliferation at an accelerated rate, organizing themselves into three-dimensional tissue masses that achieved 1.0- to 1.5-cm diameters. Each of these engineered tumor tissues produced tissue-like aggregates (TLAs) with glandular structures, apical and internal glandular microvilli, tight intercellular junctions, desmosomes, cellular polarity, sinusoid development, internalized mucin, and structural organization akin to normal colon crypt development. Necrosis was minimal throughout the tissue masses up to 60 days of culture while achieving >1.0 cm in diameter. Other notable results included enhanced growth of neoplastic colonic epithelium in the presence of mixed normal human colonic mesenchyme. These results mimic the cellular differentiation seen in vivo and are similar to results obtained with other tumor types.

Prostate carcinoma has also been modeled in the RWV system by several investigators. One of the most comprehensive descriptions of these engineered tissues is detailed by Wang et al. In that review, the authors describe the ability of the RWV system to recapitulate human prostate carcinoma (LNCaP) and bone stroma (MG63) to illuminate the evolution of prostate tumorigenesis to the metastatic condition. In particular, the LNCaP and ARCaP models represented in the review are known to be lethal in the human, being androgen independent and metastatic. Rotating-wall
vessel TLA engineering also allowed in-depth study of epithelial and stromal interactions, which are the facilitating elements of the continuance of LNCaP prostate-specific antigen production in vitro. When LNCaP was cultured in three dimensions without stroma, production of prostate-specific antigen ceased, and metastatic markers were not observed. The authors outline the process of malignant transformation, demonstrating that these metastatic models are only possible in three-dimensional TLAs and are achieved by specific geometric relationships in three-dimensional configuration. Furthermore, they show through direct comparison with other culture systems the advantages of the RWV system to allow synergistic relationships to study this disease state.107

The final two carcinoma models to be reviewed here are the subject of original ongoing research, breast carcinoma in Dr. Goodwin’s lab and lung carcinoma in Dr. Vertrees’ lab. As we have seen with the previous data, the modeling of three-dimensional tissues is not only achievable, but also desirable to create an improvement in the fidelity of the tumor model for research and diagnostic purposes. Construction of three-dimensional tumor models using RWVs has revealed striking similarities to tissues grown in vivo.63, 67

As a demonstration of the capability of these advanced tumor models to produce unique results, we have developed three-dimensional breast carcinoma models and subjected them to assault by lymphokine-activated killer (LAK) cells. Tumor-infiltrating lymphocyte (TIL) and lymphokine-activated killer (LAK) cell research has primarily dealt with tumor cells in single suspension cultured in the presence of LAK cells in close proximity using 96-well plates. Although a substantial amount of literature exists in this field, no literature could be found investigating three-dimensional tumor models and the kill response elicited in the presence of LAK cells. We hypothesized that not only would the three-dimensional models more closely resemble in vivo tissue histologically, but they also could demonstrate a useful application of using three-dimensional models. Three breast carcinoma cell lines were chosen for this investigation: BT-20, an undifferentiated estrogen receptor (ER)– cell line; MCF-7, a moderately differentiated estrogen receptor (ER)-positive and one ER-negative cell lines indicated that, in contrast to generally accepted data obtained from standard two-dimensional LAK assay, three-dimensional tumor aggregates present significantly different kill rates for all the breast carcinoma lines than do their two-dimensional analogues. Additionally, these altered kill ratios are associated with dramatic response changes in both the receptor and soluble molecular expression for the cytokines interleukin (IL)-8, tumor necrosis factor (TNF)-α, transforming growth factor (TGF)-β, and epidermal growth factor (EGF).

In Fig. 18.9, a well-developed histology section of MCF-7 is seen with organizational structure before LAK invasion. The scanning electron microscopy (SEM) panel in Fig. 18.10 demonstrates the adherence and invasion of LAK cells at 24 and 48 h post inoculation onto a mature, three-dimensional culture of breast carcinoma. Concomitantly, at 48 h LAK cells are seen digesting the surrounding tumor cells (Fig. 18.10).

Vertrees et al108 reported on a study that compared various characteristics between an immortalized (not malignant) and its malignant transfected counterpart grown both as 2-D and 3-D cultures (Figs. 18.11 and 18.12). Electron microscopy identified significant mitochondrial and granular endoplasmic reticular pathology in the 2-D cells not seen in the 3-D cells.108 The degree of differentiation determined by immunohistochemistry shows that ultrastructure and antibody expressions were more representative of control tissue when cells (both immortalized and transfected) were grown in 3-D culture than when grown as a 2-D culture. Electron microscopy identified the presence of lipid inclusion and lamellar (surfactant) bodies indicative of type II pneumocytes, a differentiation not seen in monolayer cultures. Fluorescence microscopy allowed for determination of individual migrating cells (unpublished data). The coculture experiment (simultaneous cultures of normal and transformed lung cells) in the three-dimensional environment demonstrates the ability of the transformed cell to migrate through the preexisting normal cells.

In contrast to two-dimensional models, these rotating-wall vessel tumor tissues were devoid of metabolic and nutrient deficiencies and demonstrated in vivo-like architecture. These data suggest that the rotating-wall vessel affords a new model for investigation and isolation of growth, regulatory, and structural processes within neoplastic and normal tissues.

Rotating-Wall Vessel Normal Human Tissue Models as Disease Targets

In this section, we explore the utility of rotating-wall vessel tissue-like aggregates (TLAs) as targets for microbial infection and disease. Several studies have been conducted...
Fig. 18.9. Hematoxylin and eosin (H&E)-stained sections (×400) demonstrating the superior organization of the RWV-grown cells at days 14 and 21 (a and c) versus the static control cultures at days 14 and 21 (b and d). (From Goodwin, 2001: original work; supported by NASA grant #962-23-0132).

Fig. 18.10. Scanning electron microscopy at 48 h post-LAK invasion. Notice the increased invasion potential of the LAK cells exposed to the RWV-grown cells (a) as compared to the static control cultures (b). (From Goodwin et al., 2001: original work; supported by NASA grant #962-23-0132).
recently that indicate that three-dimensional tissues respond to infective agents with greater fidelity and with a more in vivo-like response than traditional two-dimensional cultures. Nickerson et al.\textsuperscript{109} describe the development of a three-dimensional TLA engineered from INT-407 cells of the human small intestine, which were used as targets for the study of \textit{Salmonella typhimurium}. In this study, three-dimensional TLAs were used to study the attachment, invasion, and infectivity of \textit{Salmonella} into human intestinal epithelium. Immunocytochemical characterization and scanning and transmission electron microscopic analyses of the three-dimensional TLAs revealed that the TLAs modeled human in vivo differentiated tissues more accurately than did two-dimensional cultures. The level of differentiation in the INT-407 TLAs was analogous to that found in previously discussed small intestine TLAs\textsuperscript{62} and from other organ tissues reconstructed in rotating-wall vessels. Analysis of the infectivity studies revealed \textit{Salmonella} attached and infected in a manner significantly different from that in control two-dimensional cultures. During an identical exposure period of infection with \textit{Salmonella}, the three-dimensional TLAs displayed a minor loss of structural integrity when compared with the two-dimensional INT-407 cultures. Furthermore, \textit{Salmonella} demonstrated a greatly reduced ability to adhere, invade, and induce the apoptotic event in these INT-407 three-dimensional TLAs than in two-dimensional cultures. This result is not unlike the in vivo human response. Two-dimensional cultures were significantly damaged within several hours of contact with the bacteria; conversely, although “pot marks” could be seen on the surfaces of the three-dimensional TLAs, they remained structurally sound.

Cytokine analysis and expression postinfection of three-dimensional TLAs and two-dimensional cultures with \textit{Salmonella} exhibited remarkable differences in expressed levels of interleukin (IL)-1\textalpha, IL-1\beta, IL-6, IL-Ra, and TNF-\textalpha mRNAs. Additionally, noninfected three-dimensional TLAs constitutively demonstrated elevated levels of TGF-\textbeta1 mRNA and prostaglandin E\textsubscript{2} compared with noninfected two-dimensional cultures of INT-407.\textsuperscript{109}

As previously stated, traditional two-dimensional cell monolayers lack adequate fidelity to emulate the infection dynamics of in vivo microbial adhesion and invasion. The respiratory epithelium is of critical importance in protecting humans from disease. Exposed to the environment, the respiratory epithelium acts as a barrier to invading microbes present in the air, defending the host through a multilayered complex system.\textsuperscript{110} The three major layers of the human respiratory epithelium are pseudo-stratified epithelial cells, a basement membrane, and underlying mesenchymal cells. Ciliated, secretory, and basal epithelial cells are connected by intercellular junctions and anchored to the basement membrane through desmosomal interactions. Together with tight junctions and the mucociliary layer, the basement membrane maintains the polarity of the epithelium and provides a physical barrier between the mesenchymal layer and the airway.\textsuperscript{111,112} Infiltrating inflammatory and immune cells move freely between the epithelial and subepithelial compartments.

Airway epithelial cells play a vital role in host defense\textsuperscript{110} by blocking paracellular permeability and modulating airway function through cellular interactions. Ciliated epithelial cells block invasion of countless inhaled microorganisms by transporting them away from the airways.\textsuperscript{113} As regulators of the innate immune response, epithelial cells induce potent immunomodulatory and inflammatory mediators such as cytokines and chemokines that recruit phagocytic and inflammatory cells that remove microbes and enhance protection.\textsuperscript{110,112,114,115}

Ideally, cell-based models should reproduce the structural organization, multicellular complexity, differentiation state, and function of the human respiratory epithelium.
Immortalized human epithelial cell lines, such as BEAS-2B, primary normal human bronchial epithelial cells, and air-liquid interface cultures, are used to study respiratory virus infections in vitro. Traditional monolayer cultures (two-dimensional) of immortalized human bronchoepithelial cells represent homogeneous lineages. Although growing cells in monolayers is convenient and proliferation rates are high, such models lack the morphology and cell-cell and cell-matrix interactions characteristic of human respiratory epithelia. Thus, their state of differentiation and intracellular signaling pathways most likely differ from those of epithelial cells in vivo. Primary cell lines of human bronchoepithelial cells provide a differentiated model similar to the structure and function of epithelial cells in vivo; however, this state is short lived in vitro. Air-liquid interface cultures of primary human bronchoepithelial cells (or submerged cultures of human adenoid epithelial cells) are grown on collagen-coated filters in wells on top of a permeable filter. These cells receive nutrients basolaterally, and their apical side is exposed to humidified air. The result is a culture of well-differentiated heterogeneous (ciliated, secretory, basal) epithelial cells essentially identical to airway epithelium in situ. Although this model shows fidelity to the human respiratory epithelium in structure and function, maintenance of consistent cultures is not only difficult and time consuming but also limited to small-scale production and thus limits industrial research capability.

True cellular differentiation involves sustained complex cellular interactions in which cell membrane junctions, extracellular matrices (e.g., basement membrane and ground substances), and soluble signals (endocrine, autocrine, and paracrine) play important roles. This process is also influenced by the spatial relationships of cells to each other. Each epithelial cell has three membrane surfaces: a free apical surface, a lateral surface that connects neighboring cells, and a basal surface which interacts with mesenchymal cells.

Viral studies by Goodwin et al. and Suderman et al. were conducted with RWV-engineered TLA models of normal human lung. This model is composed of a coculture of in vitro three-dimensional human bronchoepithelial TLAs engineered using a rotating-wall vessel to mimic the characteristics of in vivo tissue and to provide a tool to study human respiratory viruses and host–pathogen cell interactions. The TLAs were bioengineered onto collagen-coated cyclodextran beads using primary human mesenchymal bronchial tracheal cells as the foundation matrix and an adult human bronchial epithelial immortalized cell line (BEAS-2B) as the overlying component. The resulting TLAs share significant characteristics with in vivo human respiratory epithelium, including polarization, tight junctions, desmosomes, and microvilli. The presence of tissue-like differentiation markers, including villin, keratins, and specific lung epithelium markers, as well as the production of tissue mucin, further confirm these TLAs differentiated into tissues functionally similar to in vivo tissues. Increasing virus titers for human respiratory syncytial virus (wtRSV A2) and parainfluenza virus type 3 (wtPIV3 JS) and the detection of membrane-bound glycoproteins (F and G) over time confirm productive infections with both viruses. Viral growth kinetics up to day 21 pi with wtRSV A2 and wtPIV3 JS were as follows: wtPIV3 JS replicated more efficiently than wtRSV A2 in TLAs. Peak replication was on day 7 for wtPIV3 JS [approximately 7 log_{10} particle-forming units (pfu) per milliliter] and on day 10 for wtRSV A2 (approximately 6 log_{10} pfu/ml). Viral proliferation remained high through day 21 when the experiments were terminated. Viral titers for severe acute respiratory syndrome coronavirus were approximately 2 log_{10} pfu/ml at 2 days post inoculation.

Conclusion

Although great strides have been made in recent years in the understanding of the biology of cancer, the role that cellular differentiation and three-dimensional structural organization play in metastasis and malignancy is still unclear. A model that more closely links the expression of specific antibodies in normal lung cells, transformed lung cells, and cancerous lung cells to cellular biology may provide new information about cancer. The development of three-dimensional cell cultures in bioreactors may ultimately provide a model that facilitates discovery and interpretation of more relevant information for the expression of an antibody and its role in cellular pathobiology of the lung.

Human lung tissue-like aggregations mimic aspects of the human respiratory epithelium wall and provide a unique opportunity to study the host-pathogen interaction of respiratory viruses and their primary human target tissue independent of the host immune system, as there can be no secondary response without the necessary immune cells. These rotating-wall vessel-engineered tissues represent a valuable tool in the quest to develop models that allow analysis and investigation of cancers and infectious disease in models engineered with human cells alone.

We have explored the creation of three-dimensional tissue-like aggregates for normal and neoplastic studies and finally as targets for microbial infections. Perhaps Carrel and Leighton would be fascinated to know that from their early experiments in three-dimensional modeling and the contributions they made has sprung the inventive spirit to discover a truly space age method for cellular recapitulation.

References


42. Strangeways T. Tissue Culture in Relation to Growth and Differentiation. Cambridge, England: Heffer and Sons; 1924.


64. Schwarz RP, Wolf DA, Trinh TT. Horizontally rotated cell culture system with a coaxial tubular oxygenator. US patent no 5,026,650; June 25, 1991.


76. Feder J, Tolbert WR. The large-scale cultivation of mammalian cells. Sci Am. 1983;248:36–43.


18. Tissue Culture Models


Index

A
A1-antitrypsin deficiency, 47
Aberrations, identification of, 94–95
Abundance concept, 120
A-Catenin, 21
Acetaldehyde, 100
Acetylation, 3
Acute respiratory distress syndrome (ARDS), 161, 164
Adapter proteins, 36
Adenine (A)
  defined, 1
  in DNA structure, 109–111
Adult stem cells
  differentiation potential of, 148f, 149
  vs. embryonic stem cells, 148
  glossary of terms for, 148t
  injury response of, 149
  overview of, 147
Advanced sequencing, 66
Aflatoxins, 99
Age/aging, stem cells role in, 150
Alcohol/alcoholism, as carcinogen, 100
Alveolar lung tissue injury, 164
Amplicons
  detection of, 63–65
  lab design for, 67–68
  manufacture of, 75
  production techniques for, 62f
Amplification technologies
  with lasers, 91–92
  in PCR, 81f
  types of, 61–65
Anatomic pathology
  for infectious agents, 71
  for neoplasia, 70–71
  techniques in, 69
Animal models, 153–156
Annealing
  overview of, 73–75
  step programming for, 76
Antibodies
  for assessing biomarkers, 135
  monoclonal antibodies, 140
Antigens
  cell lineage-specific, 144t
  role in flow cytometry, 142–145
Apoptosis
  defined, 29
  inhibitors of, 36–37
  mechanisms of, 32–38
  types of, 30–32
Array technology
  bacterial artificial chromosome (BAC) arrays, 87–89
  in CGH, 116f
  large-insert clone arrays, 87–89
  nonredundant sequence arrays, 89
  oligonucleotide microarrays, 89–90
  repeat-free arrays, 89
Asbestos exposure
  loss of heterozygosity in, 99
  mesothelioma from, 156
Asthma
  animal model research in, 154
  loss of heterozygosity in, 103
Autophagy, 30–31
Autoschisis, 30f, 31–32

B
Baby hamster kidney (BHK) tissues, 169f
Background subtraction, 93
Bacteria
  *Helicobacter pylori*, 100
  *Mycobacterium tuberculosis*, 100
  *Salmonella typhimurium*, 177
Bacterial artificial chromosome (BAC) arrays, 87–89
Base excision repair (BER), 12
Base pairing, 109–111
B-Catenin, 21
B-cell lymphoma histogram, 143f
Bcl-2 family of proteins, 36, 37f
Benign diseases, loss of heterozygosity in, 102
Bioartificial muscle (BAM), 165
Bioinformatics, 55
Biomarker testing
  immunohistochemistry in, 133–134
  method standardization in, 134–135
Bioreactors, 164–165, 166, 168
Bladder cancer, loss of heterozygosity in, 99
Bleomycin installation, 154–155
Bone loss, 165–166
Bone marrow, 166
Bone marrow histogram, 143f
Bovine embryonic kidney (BEK) cells, 168
Branched chain DNA (bDNA), 65
BRCA genes, 100t
Breast cancer
  epimutations in, 48
  gene amplifications in, 43–44
  three-dimensional models of, 175
C
Cadherins
cyclooxygenase 2 and, 21–22
types and functions of, 20–22
Cancers/malignancies
animal model research in, 153–156
cell culture research in, 163
chromosomal alterations in, 42
common chromosomal deletions in, 43–44
complex chromosomal rearrangements in, 44
complex in cancer, 44
epimutations in, 47–48
gene amplifications in, 42–43
gene mutations in, 45–46
loss of heterozygosity in, 97, 100–103
stem cell role in, 149–150
See also specific cancers.
Carcinogenesis
from environmental factors, 99–100
oncogenes role in, 70
stem cells role in, 149–150
Carcinogens
cigarette smoke. See Tobacco use.
in foods, 99–100
mutation mechanisms of, 99, 149–150
Carrel, Alexis, 159, 164–165
Caspases
functions of, 33, 34f
protein targets of, 37–38
role in cell death, 32
CD44 adhesion molecule, 24
CD45 antigen
in cell windows, 144f, 145
overview of, 142
vs.SSC, 143f
windows/lineages in, 143f, 144
CDH13 gene, 100t
Cell adhesion molecules
CD44 adhesion molecule, 24
in cell culture, 162
defined, 12
role in cancer invasion, 19–24
Cell culture, primary
biology of, 161–162
cell adhesion in, 162
defined, 160
functional environments for, 162
models for epithelial cultures, 162–163
in study of disease processes, 161
three-dimensional, 164–166
Cell cycle, 11–12
Cell death, alternative forms of, 29–32
Cell lines
characteristics of, 160
development of, 162
Cell populations
basic parameters and windows of, 142–145
for culture models, 162–163
gating of, 142
immunophenotyping analysis in, 142–144
muscle cells, 165
pulmonary alveolar type II cells, 164
windows/lineages in, 143f, 144–145
CEP probes, 111–112
Charcot-Marie-Tooth disease (CMT), 45
Checkpoints (cell cycle), 11
Chlamydia pneumoniae, 100
Chondrocyte cultures, 169–171
Chromogenic in situ hybridization (CISH)
defined, 69, 109
vs. FISH, 115–116
Chromosomal alterations, diseases associated with, 42–45
Chromosomal deletions common in cancer, 43–44
Chromosomal microarrays, 66
Chromosomal paints, 112
Chromosomes
cancer-specific LOH on, 100–103
defined, 1
Chronic obstructive pulmonary disease (COPD)
animal model research in, 153–155
loss of heterozygosity in, 103
Cigarette smoke. See Tobacco use.
Cluster of differentiation (CD), 140
Codon, 2
Colon cancer, 174–175
Color blindness, 45
Colorectal cancer, epimutations in, 48
Comparative genomic hybridization (CGH), 69
array platforms in, 87–90
data analysis in, 93–95
data smoothening in, 94
facility setup for, 90
image processing in, 93
in loss of heterozygosity, 97
normalization in, 93–94
principle of, 87, 88f
in situ type of, 116–117
Complementary DNA (cDNA) arrays, 87
Coomassie Brilliant Blue stains, 123
Cre recombinase, 155
Crotonaldehyde, 100
Culture media, 161–162
Cux-1 gene, 155
CXCR4 chemokine receptor, 20
Cyclooxygenase 2 (COX-2), 21–22
Cystic fibrosis (CF)
animal model research in, 154
overview of, 46–47
Cystic fibrosis transmembrane conductance regulator (CFTR), 47, 48, 154
Cytomegalovirus (CMV), 115f
Cytomics, 56
Cytosine (C)
defined, 1
in DNA structure, 109–111
Cytotoxic T cells, 35

D
Data analysis in CGH, 93–95
Data smoothening in CGH, 94
Data validation in proteomics, 121–122
Database searching for protein identification, 127–128
Dedifferentiation, 162
Denaturation
defined, 111
in FISH, 114
overview of, 73–75
step programming for, 76
Deoxynucleotides, 77
Deoxyribonucleic acid (DNA)
amplification of, 91–92
composition and structure of, 63f, 74f, 80f, 109, 110f
damage repair for, 12
Index

defined, 1
injected into mouse models, 153
isolation and quality of, 90–91
labeling of, 92
melting temperature of, 111
reference DNA, 92
Desorption electrospray ionization (DESI) technique, 126
Diaminobenzidine (DAB), 115
Digestion, preparatory, enzymatic and nonenzymatic, 123
Dimerization, 3
Dimers, 3
Direct labeled probes, 112
Disease
animal model research in, 153–156
categories of, 41
cell culture’s role in, 161
stem cell role in, 149–150
dna adducts, 99
DNA chips, 62f
Drug metabolism, 163
Dual-probe real time PCR, 79–81

E
E-Cadherin, 21
Electrospray ionization (ESI), 125
Embryonic stem cells
vs. adult stem cells, 148
differentiation potential of, 148f, 149
glossary of terms for, 148t
in knockout mouse models, 155–156
overview of, 147
Emphysema, animal model research in, 154
Enhancers, 4
Environmental factors, carcinogenesis-related, 99–100
Enzymatic preparatory digestion, 123
Enzymes
caspases. See Caspases.
kinas. See Kinases.
polymerases. See Polymerase chain reaction (PCR).
for preparatory digestion, 123, 126
replication of, 2
Taq polymerase, 73–75, 77, 79–81
See also specific enzymes.
Epidermal growth factor receptor (EGFR), 22, 71, 133
Epimutations
defined, 41
diseases associated with, 47–48
Epithelial cells, 162–163, 177–178
Epithelial tissue, 177
Epithelial-mesenchymal transition, 22
Estrogen receptor positive [ER(+)] tumors, 133
Euchromatin, 2
Ex vivo propagation, 160
Excision repair cross-complementation group 1 (ERCC1), 133
Extracellular pathway, 34f
Extrinsic death pathway, functions of, 33, 35

F
False discovery rate (FDR), 128
False positives/false negatives in PCR, 71
Fhit gene, 100t, 101
Fibrosis, idiopathic, 154
Field “cancerization”, 99
Fingerprinting identification. See Protein mass fingerprinting identification.
Fixation, incomplete, 135
Flow cytometers, 140–141
Flow cytometry
basic principles of, 139–140
compared to immunohistochemistry (IHC), 140t
data analysis and interpretation in, 142–145
overview of, 139
specimens suitable for, 141–142
technical aspects of, 140–142
Fluorescence, 63–64
Fluorescence in situ hybridization (FISH)
defined, 66, 109
multicolored, 117
specimens for, 113–114
steps involved in, 113–115
vs.CISH, 115–116
Fluorescence microscopy for FISH, 114–115
Fluorescence resonance energy transfer (FRET), 79–81
Fluorescence-activated cell sorting (FACS), 147
Fluorescent dyes, 140
Fluorescent probes, 62f
Fluorochromes, 140
Focal adhesion kinase, 20
Foods, carcinogens in, 99–100
Formalin-fixed paraffin-embedded (FFPE) tissue, 69–71, 90, 135
Forward light scatter (FSC), 142–145
Fourier transform ion cyclotron resonance (FTICR), 125–126
Fractional allelic loss (FAL), 98
Fractionation, 120
Free radicals, cancer-related, 100

G
Gametes, 1
Gastrointestinal epithelial cells, 163
Gel electrophoresis
defined, 62f, 119
two dimensional. See Two-dimensional polyacrylamide gel electrophoresis (2D PAGE).
Gene amplifications in cancers/malignancies, 42–43
Gene expression
alterations of normal, 41–42
control of, 3
defined, 2
Gene products
defined, 2
posttranslational modifications of, 3
Gene sequencing, 62f, 65–66, 89, 127
Genes
defined, 1
disease-specific genes, 155–156
splicing of, 129
tumor-specific genes, 100–101
See also Proteins; Proteomics; specific genes and procedures.
Genome, 1
Genomics, 1, 55, 119
Genotypes, 1–2
GLUT-1 gene, 101
Glycomics, 56
Gox1 gene, 100t
Granulocyte window, 145f
Growth factor
receptors of, 9
signaling of, 11
Guanine (G)
defined, 1
in DNA structure, 109–111
Helicobacter pylori, 100
Hematopoietic stem cells, 147, 149
Hepatocellular carcinoma (HCC), 100, 163
Hepatocyte tissue cultures, 163
HER2 gene, 116, 133, 136
Hereditary nonpolyposis colon carcinoma (HNPCC), 134
Heterochromatin, 2
Heterodimers, 3
Heterozygous status, 2
Histograms
of CD45 vs. SSC, 143f
of FSC vs. SSC, 143f
Histones, 1
Homeostasis mechanisms, 149
Homodimers, 3
Homozygosity, 2, 97
Hot start PCR, 78
HRPT2 gene, 100t
Human embryonic kidney (HEK) cells, 168
Human leukocyte differentiation antigens (HLDA), 140
Hybrid Capture, 65
Hybridization probes/kinetics
chromogenic in situ type of, 69, 109, 115–116
defined, 111
disadvantages in CGH, 92–93
FISH vs. CISH, 115–116
fluorescence in situ type of, 66, 69, 109, 114
functions of, 62f, 63–64, 79, 80f
general principles for in situ, 109–111
overview of, 109
Hydrolysis probes, 63–64
Idiopathic pulmonary fibrosis (IPF), 103
Ig CAMS, 23–24
Image processing in CGH, 93
Imaging mass spectrometry (IMS), 124–127
Immunophenotype cells, 140–141
Immunodetection systems, 135–136
Immunohistochemistry (IHC)
in biomarker testing, 70, 133–136
compared to flow cytometry, 140t
Indirect labeled probes, 112
Induced pluripotent stem cells (iPSCs), 150
Infections
cell culture research in, 163
inflammation associated with, 100
Infectious agents, anatomic pathology testing for, 71
Inflammation associated with infections, 100
Infrared multiphoton dissociation (IRMPD), 125–126
Inhaled toxicants, 161
In-house developed probes, 112
Injury response of stem cells, 149
Inner cell mass (ICM), 148f, 149
Integrin-linked kinase (ILK), 20
Integrins, 19–20
Interstitial diseases, 154–155
Interstitial pulmonary fibrosis (IPF), 154–155
Intracellular pathway, 34f
Intrinsic death pathway, 35–36
Invader assays, 65
Isoelectric focusing (IEF), 122
Karyotypes
vs. CGH, 116–117
defined, 1
Kinases
defined, 3
integrin-linked, 20
Knockout mouse models, 155–156
KRAS gene, 70
Karyotypes
vs. CGH, 116–117
defined, 1
Kinases
defined, 3
integrin-linked, 20
Knockout mouse models, 155–156
KRAS gene, 70
Laboratory performance assessment, 136
Laboratory staffing, 68
Large-insert clone arrays, 87–89
Ligands, 7
LightCycler® PCR, 79, 81
Liquid bead microarrays, 67
Locus-specific probes, 112
Loss of heterozygosity (LOH)
chromosome- and tumor-specific, 100–103
detection of, 97–99
environmental influence on, 99–100
gene- and tumor-specific, 100–103
mechanisms of, 98f
overview of, 97
Lung cancer
array CGH analysis of, 89f
epimutations in, 47–48
FISH detection of, 117f
loss of heterozygosity in, 99
ras mutations in, 45–46
three-dimensional models of, 175
Lung disease
animal models of, 153–156	hree-dimensional models of, 172–174
See also specific diseases.
Lymphatic tissue culture, 163
Lymphoblast window, 145f
Lymphocyte window, 143–144
Lymphokine activated killer (LAK) cells, 175–176
Magnesium in PCR, 77
MASCOT database, 127–128
Mass spectrometry
defined, 62f, 67
“dream team” system for, 125–126
overview of, 123
types of, 123–126
Matrix-assisted laser desorption/ionization (MALDI), 123–124, 126–127
Membrane bioreactors, 165
Mesenchymal stem cells (MCSs), 149–150
Mesothelioma
animal model research in, 156
loss of heterozygosity in, 99
Messenger RNA (mRNA)
defined, 2
in reverse transcription, 81–83
Metabolomics, 56
Methylation
defined, 47
detection methods for, 66
Methylation-specific PCR, 78
Microarray technologies
defined, 56–57
types of, 66–67

Microdissection (tissue)
defined, 69–70
with lasers, 91–92

Mismatch repair genes (MMR), 12

Missense mutations, 46

Mitogen-activated protein kinase (MAPK), 9–10

Mitosis, 11

Molecular beacons, 63–64

Molecular pathology/diagnostics
in anatomic pathology, 69–71
defined, 61, 62f
general approaches to, 61–68
lab design for, 67–68

Molony murine leukemia virus (MMLV), 82

Monocyte window, 145f

Multicolor FISH (M-FISH), 117

Multiple reaction monitoring (MRM), 130

Molecular pathology/diagnostics
in anatomic pathology, 69–71
defined, 61, 62f
general approaches to, 61–68
lab design for, 67–68

Molony murine leukemia virus (MMLV), 82

Monocyte window, 145f

Multicolor FISH (M-FISH), 117

Multiple reaction monitoring (MRM), 130

Molecular pathology/diagnostics
in anatomic pathology, 69–71
defined, 61, 62f
general approaches to, 61–68
lab design for, 67–68

Molony murine leukemia virus (MMLV), 82

Monocyte window, 145f

Multicolor FISH (M-FISH), 117

Multiple reaction monitoring (MRM), 130

N
NASA Johnson Space Center, 172–174
National Center for Biotechnology Information
(NCBI), 127

Necrosis, 29–30

Neoplasia, testing for, 70–71. See also Cancers/malignancies.

Nervous system tissue cultures, 163

Nested PCR, 78

Nitrosamines, 100

Nonenzymatic preparatory digestion, 123

Nonneoplastic diseases
chromosomal alterations in, 44–45
epimutations in, 48
gene mutations in, 46–47

Nonredundant sequence arrays, 89

Nonsense mutations, 42, 46

Normal tissue models, 169–172, 175–178

Normalization
algorithms for, 93
in CGH, 94

Nucleic acid extraction, 61

Nucleic acids, 1

Nucleosomes, 1

Nucleotide excision repair (NER), 12

Nucleotide sequence alterations, 42

Nucleotides, 1

O
Oligodendroglial tumors, 100t, 101

Oligonucleotide microarrays, 66, 89–90

Omics
defined, 55
types of, 119–120

Oncogene targeting, disease-specific, 155–156

Oncogenes, 70

Organ culture, 160

Osteoporosis, 165–166

Osteosarcoma, multicolor FISH of, 117f

P

P53 tumor suppressor gene, 46

Pancreatic cancer, 100

Paraptosis, 30f, 31

Pathogens, cancer-related, 100

Peptides
heavy vs. light peptides, 129
sequencing of, 127

Phenotypes, 1–2

Phenyx database, 127–128

Phosphatases, 3

Phosphorylation
defined, 3
process of, 9, 129

Physiomics, 56

Plasticity of stem cells, 149

Polymerase chain reaction (PCR)
buffers in, 77
components of, 77
false positives/false negatives in, 71
procedure analysis of, 78–79
process of, 61–64
in real time, 63–64, 79–81
recent variations in, 78–79
setup for, 77–78
theory and principles of, 73–75
typical reagent mix for, 78t

Polymers, 2

Posttranslational modifications, 129–130

Prehybridization, 114

Preparatory digestion, 123, 126

Primary explants, 160, 164

Primer design, 76–77

Probes
preparation of, 112
types of, 111–112
See also Hybridization probes/kinetics.

Progenitor cells, 148f, 149

Promoters, 4

Promyelocytic leukemia (PML), 44

Prostate carcinoma, 48, 175

Proteases, 154

Proteasomes, 3

Protein Information Resource (PIR), 127

Protein mass fingerprinting identification, 126–128

Protein profiling, 120–121

Proteins
complexes and interactions in, 130
degradation of, 3
identification of, 121, 126–128
quantification of, 128–129
staining of, 122–123

Proteomics
defined, 56
as diagnostic hematology tool, 120
four methods of protein identification, 121
historical development of, 119–120
ideal proteomics, 120
Index

Proteomics (cont.)
  instrumentation for, 122–126
  protein staining in, 122–123
  proteomics pattern diagnostics, 125
  protocols for, 126–128
Proto-tissues, 165
Pulmonary alveolar type II cells, 164
Pulmonary disease. See Lung disease.
Pulmonary endothelial cells, 163
Purines
  defined, 1
  in DNA structure, 109–111
Pyrimidines
  defined, 1
  in DNA structure, 109–111
Q
  Quadrupole mass spectrometry, 125
  Quantification techniques, 129
  Quenching process, 80f
R
  Ras gene family, 45–46
  Real time sequencing, 65–66
  Receptors, cell surface, 9
  Reference DNA, 92
  RefSeq collection database, 127
  Renaturation, 111
  Repeat-free arrays, 89
  Replication, 2
  Representational oligonucleotide microarray analysis (ROMA), 92
  Restriction fragment length polymorphism (RFLP)
    defined, 65
    in loss of heterozygosity, 97
  Retinoblastoma (RB1) gene, 102
  Reverse transcription
    defined, 81–83
    reaction protocol for, 82t
  Ribonucleic acid (RNA)
    defined, 1
    in reverse transcription, 81–83
  Ribosomal RNA (rRNA), 2
  Rotating wall vessel (RWV)
    applications for, 166–172
    overview of, 173
    superior cells from, 176f
    tumor-specific models from, 174–175
S
  Salmonella typhimurium, 177
  Sanger sequencing, 65
  Sarcoidosis, 103
  Satellite stem cells, 149
  Selectins, types and functions of, 22–24
  Sequencing. See Gene sequencing.
  Short tandem repeat (STR), 2–3
  Shotgun fractionation, 120
  Side light scatter (SSC), 142–145
  Signal amplification, 65
  Signaling pathways, 9
  Simian virus large T antigen (Tag), 155
  Single nucleotide polymorphisms (SNPs)
    defined, 2
  in loss of heterozygosity, 97–99
  Small cell lung cancer (SCLC), 101, 143f
  Smokeless tobacco
    as carcinogen, 99
  See also Tobacco use.
  Sodium dodecyl sulfate (SDS) polyacrylamide gel, 122
  Sodium nitrate, 99–100
  Sodium nitrite, 99–100
  Southern blotting, 65
  Space bioscience, 165–166
  Spectra deconvolution, 121–122
  Spectral karyotyping (SKY), 69, 117
  Splicing, gene, 129
  Stem cells
    adult vs. embryonic, 147–148
    glossary of terms for, 148f
    overview of, 147
    plasticity of, 149
    role in pathogenesis, 149–150
  Surface-enhanced laser desorption/ionization mass spectrometry (SELDI), 124–125
  Surfactant protein C (SP-C), 155
  SV40 virus, 156
  Swiss-Prot database, 127
  Systems biology, 67
T
  Taq polymerase, 73–75, 77, 79–81
  Target amplification, 61–65
  TATA-binding protein (TBP), 4
  Telomeres, 3
  Telomeric probes, 112
  Tetramers (four proteins), 3
  Thermocyclers, 75–76
  Thermostable polymerase, 77
  Three-dimensional biology
    examples of, 164–166
    study models for, 166–169
  Thymine (T)
    defined, 1
    in DNA structure, 109–111
  Time of flight (TOF) mass analyzer, 123–124, 126–127
  Tissue culture models
    2D vs. 3D, 177f
    advantages and limitations of, 160–161
    history of, 159
    types of, 160
  Tissue engineering stages, 167f, 172–174
  Tissue-like aggregates (TLAs), 175–178
  Tobacco use
    carcinogens associated with, 45–46
    COPD-related, 153
    in loss of heterozygosity, 99
  Topoisomerases, 2
  TP53 gene, 102
  Transcription
    defined, 2
    initiation of, 4
  Transcription factors, 3–4
  Transcription initiation site (TIS), 2
  Transcriptomics, 55, 119
  Transfer RNA (tRNA), 2
  Transgenic mouse models, 153–155
  Transgenic systems, 153–155
Translation
  defined, 2
  posttranslational modifications, 129–130
TrEMBL database, 127
Trimers (three proteins), 3
Trypsin, 123, 126
Tumor necrosis factor receptor-2 (TNFR2), 101
Tumor necrosis factor (TNF), 35
Tumor suppressor genes (TSGs), 70–71
Tumorigenesis. See Carcinogenesis.
Tumors
  estrogen receptor positive [ER(+)], 133
  stem cell role in, 149–150
See also Benign diseases; Cancers/malignancies.
Two-dimensional polyacrylamide gel electrophoresis (2D PAGE), 122, 126–127
Two-hit model, 97
Type II pneumocytes, 164

U
  Ubiquitin, 3
  Ubiquitinylation, 3
  Unidirectional workflow, 68
  UniProt database, 127
  Uracil, 1

V
  Vacuolation, 30f, 31
  Validation process (protein match), 128
  Viral studies, 178
  Von Hippel-Lindau (VHL) gene, 100t

W
  Whole genome amplification (WGA), 91–92