

01 - 479 Principles of Human Genetics

479 Principles of Human Genetics

Genes, the Environment, and Disease PART 16 J. Larry Jameson, Peter Kopp

Principles of Human

Genetics IMPACT OF GENETICS AND GENOMICS

ON MEDICAL PRACTICE Over the past four decades, novel insights into human genetics and genomics have fundamentally impacted the practice of medicine, ushering in a new area with a deeper understanding of the genetic basis of numerous health conditions, novel diagnostic technologies, disease prevention and management, personalized medicine, and targeted therapies. Human genetics refers to the study of individual genes, their role and function in disease, and their mode of inheritance. Genomics refers to an organism's entire genetic information, the genome, and the function and interaction of DNA within the genome, as well as with environmental or nongenetic factors, such as a person's lifestyle. With the characterization of the human genome, genomics not only complements traditional genetics in our efforts to elucidate the etiology and pathogenesis of disease, but it also plays a prominent and expanding role in diagnostics, prevention, and therapy (Chap. 480). These transformative developments, originally emerging from the Human Genome Project, have been variably designated genomic medicine, personalized medicine, or precision medicine. Precision medicine aims at customizing medical decisions to an individual patient. For example, a patient's genetic characteristics (genotype) can be used to optimize drug therapy and predict efficacy, adverse events, and drug dosing of selected medications (pharmacogenomics) (Chap. 72). The characterization of the mutational profile of a malignancy allows the identification of driver mutations or overexpressed signaling molecules, which then facilitates the selection of targeted therapies. Genome-wide polygenic risk scores (PRS) for common complex diseases are beginning to emerge and may impact disease prevention in the future. Genetics has traditionally been viewed through the window of relatively rare single-gene diseases. These disorders account for ~10% of pediatric admissions and childhood mortality. Historically, genetics has focused predominantly on chromosomal and metabolic disorders, reflecting the long-standing availability of techniques to diagnose these conditions. For example, conditions such as trisomy 21 (Down's syndrome) or monosomy X (Turner's syndrome) can be diagnosed using cytogenetics. Likewise, many metabolic disorders (e.g., phenylketonuria, familial hypercholesterolemia) are diagnosed using biochemical analyses. The advances in DNA and RNA

diagnostics have extended the field of genetics to include virtually all medical specialties and have led to the elucidation of the pathogenesis of most monogenic disorders. In addition, virtually every medical condition has a genetic component. As is often evident from a patient's family history, many common disorders such as hypertension, heart disease, asthma, diabetes mellitus, and mental illnesses are significantly influenced by the genetic background. These polygenic or multifactorial (complex) disorders involve the contributions of many different genes, as well as environmental factors that can modify disease risk. Genome-wide association studies (GWAS) have elucidated numerous disease-associated loci and are providing novel insights into the allelic architecture of complex traits. These studies have been facilitated by the availability of comprehensive catalogues of human single nucleotide polymorphism (SNP) haplotypes (HapMap, International Genome Sample Resource). Next-generation DNA sequencing (NGS) technologies have evolved rapidly, and the cost of sequencing whole exomes (the exons within the genome; whole exome sequencing [WES]) or genomes (whole genome sequencing [WGS]) has plummeted. Comprehensive unbiased sequence analyses are now routinely used to characterize individuals

with complex undiagnosed conditions or to determine the mutational profile of advanced malignancies in order to select optimal and targeted therapies. The assembly of diploid genomes, i.e., the characterization of the complete genetic information from both sets of chromosomes in an individual's genome, will further enhance the complete resolution of genetic variation and should provide further insights into heritability and disease mechanisms. Cancer has a genetic basis because it results from acquired somatic mutations in genes controlling growth, apoptosis, and cellular differentiation (Chap. 76). In addition, the development of many cancers is associated with a hereditary predisposition. Characterization of the genome (and epigenome) in various malignancies has led to fundamental new insights into cancer biology and reveals that the genomic profile of mutations is in many cases more important in determining the appropriate therapy than the organ in which the tumor originates. The Cancer Genome Atlas (TCGA) initiative of the National Cancer Institute and the National Human Genome Research Institute has already characterized the genomic landscape across >30 malignancies. TCGA consists of comprehensive analyses of genomic and proteomic alterations and provided fundamental new insights into the molecular pathogenesis of cancer. These data, together with comprehensive catalogues of somatic mutations identified in human cancer, have direct clinical ramifications that impact cancer taxonomy, as well as the development and choice of targeted therapies. Genetic and genomic approaches have proven invaluable for the detection of infectious pathogens and are used clinically to identify agents that are difficult to culture such as mycobacteria, viruses, and parasites, or to track infectious agents locally or globally. In many cases, molecular genetics has improved the feasibility and accuracy of diagnostic testing and has opened new avenues for therapy, including gene and cellular therapies (Chap. 483). Molecular genetics has also provided the opportunity to characterize the microbiome, a field that characterizes the population dynamics of bacteria, viruses, and parasites that coexist with humans and other animals (Chap. 483). The microbiome has significant effects on normal physiology as well as various disease states, and the field is now focusing on defining the mechanisms underlying these interactions. Molecular biology has significantly changed the treatment of human disease. Peptide hormones, growth factors, cytokines, and vaccines can be produced in large amounts using recombinant DNA and RNA technology (e.g., mRNA vaccines against SARS-CoV-2; small interfering RNA [siRNA] to treat hypercholesterolemia). Targeted modifications of recombinant peptides provide improved therapeutic tools, as illustrated by genetically modified insulin analogues with more favorable

kinetics or glucagon-like peptide 1 (GLP-1) agonists for treatment of type 2 diabetes and for weight management. The rate at which new genetic and genomic information is being generated presents many challenges for health care providers and systems. Although many functional aspects of the genome remain unknown, there are many clinical situations where genetic and genomic information optimize patient care. Much genetic information resides in databases that provide easy access to the expanding information about the human genome, genetic disease, and genetic testing (Table 479-1). For example, several thousand monogenic disorders are summarized in a large, continuously evolving compendium, the Online Mendelian Inheritance in Man (OMIM) catalogue (Table 479-1). The constant refinement of bioinformatics and big data analytics, together with the widespread adoption of electronic health records (EHRs), are simplifying the access, analysis, and integration of this daunting amount of new information. Importantly, genomic data can be integrated readily into EHRs and thus impact clinical practice. ■ ■

THE HUMAN GENOME

Structure of the Human Genome

The Human Genome Project was initiated in the mid-1980s as an ambitious effort to characterize

TABLE 479-1 Selected Databases Relevant for Genomics and Genetic Disorders

| SITE | URL | COMMENT |
|---|---|---|
| National Center for Biotechnology Information (NCBI) | http://www.ncbi.nlm.nih.gov/ | Broad access to biomedical and genomic information, literature (PubMed), sequence databases, software for analyses of nucleotides and proteins |
| Extensive links to other databases, genome resources, and tutorials | | |
| PART 16 Genes, the Environment, and Disease | | |
| National Human Genome Research Institute | http://www.genome.gov/ | An institute of the National Institutes of Health focused on genomic and genetic research; links providing information about the human genome sequence, genomes of other organisms, genomic research, and legislation |
| Catalog of Published Genome-Wide Association Studies | https://www.ebi.ac.uk/gwas/ | Published high-resolution genome-wide association studies (GWAS) |
| Ensembl Genome browser | http://www.ensembl.org | Maps and sequence information of eukaryotic genomes |
| Online Mendelian Inheritance in Man | http://www.ncbi.nlm.nih.gov/omim | Online compendium of Mendelian disorders and human genes causing genetic disorders |
| American College of Medical Genetics and Genomics | http://www.acmg.net/ | Extensive links to other databases relevant for the diagnosis, treatment, and prevention of genetic disease |
| American Society of Human Genetics | http://www.ashg.org | Information about advances in genetic research, professional and public education, and social and scientific policies |
| The Cancer Genome Atlas | https://cancergenome.nih.gov/ | Comprehensive, multidimensional characterization of the genomic and proteomic landscape of malignancies with high public health impact |
| COSMIC Catalogue of Somatic Mutations in Cancer | https://cancer.sanger.ac.uk/cosmic | Comprehensive catalogue of somatic mutations in human cancer |
| Genetic Testing Registry | https://www.ncbi.nlm.nih.gov/gtr/ | International directory of genetic testing laboratories and prenatal diagnosis clinics; reviews and educational materials |
| Genomes Online Database (GOLD) | http://www.genomesonline.org/ | Information on published and unpublished genomes |
| HUGO Gene Nomenclature | http://www.genenames.org/ | Gene names and symbols |
| GENECODE | https://www.encodegenes.org/ | High-quality reference gene annotation and experimental validation for human and mouse genomes |
| MITOMAP, a human mitochondrial genome database | http://www.mitomap.org/ | A compendium of polymorphisms and mutations of the human mitochondrial DNA |
| The International Genome Sample Resource (IGSR) | http://www.internationalgenome.org | Public catalogue of human variation and genotype data from numerous ethnic groups |
| Human Genome Variation Society | https://www.hgvs.org/ | Collection and documentation of genomic variations including population distribution and phenotypic associations |

ENCODE <http://www.genome.gov/10005107> Encyclopedia of DNA Elements; catalogue of all functional elements in the human genome Dolan DNA Learning Center, Cold Spring Harbor Laboratories <http://www.dnalc.org/> Educational material about selected genetic disorders, DNA, eugenics, and genetic origin The Online Metabolic and Molecular Bases of Inherited Disease (OMMBID) <http://ommbid.mhmedical.com> Online version of the comprehensive text on the metabolic and molecular bases of inherited disease Online Mendelian Inheritance in Animals (OMIA) <https://www.omia.org/home/> Online compendium of Mendelian disorders in animals The Jackson Laboratory <http://www.jax.org/> Information about murine models and the mouse genome Mouse genome informatics <http://www.informatics.jax.org> Mouse genome informatics, potential mouse models of human disease, information on phenotypic similarity between mouse models and human patients Note: Databases are evolving constantly. Pertinent information may be found by using links listed in the few selected databases. the entire human genome and culminated in the completion of the DNA sequence for the last of the human chromosomes in 2006. The scope of a WGS analysis can be illustrated by the following analogy. Human DNA consists of ~3 billion base pairs (bp) of DNA per haploid genome, which is nearly 1000-fold greater than that of the *Escherichia coli* genome. If the human DNA sequence were printed out, it would correspond to about 120 volumes of Harrison's Principles of Internal Medicine. In addition to the human genome, the genomes of thousands of organisms have been sequenced completely or partially (Genomes Online Database [GOLD]; Table 479-1). They include, among others, eukaryotes such as the mouse (*Mus musculus*), *Saccharomyces cerevisiae*, *Caenorhabditis elegans*, and *Drosophila melanogaster*; bacteria (e.g., *E. coli*); and archaea, viruses, organelles (mitochondria, chloroplasts), and plants (e.g., *Arabidopsis thaliana*). Genomic information of infectious agents has significant impact for the characterization of infectious outbreaks and epidemics. Other ramifications arising from the availability of genomic data include, among others, (1) the comparison of entire genomes (comparative genomics); (2) the study of large-scale expression of RNAs (functional genomics), proteins (proteomics), or protein families (e.g., the kinome, the complete set of

protein kinases) to detect differences between various tissues in health and disease; (3) the characterization of the variation among individuals by establishing catalogues of sequence variations and SNPs; and (4) the identification of genes that play critical roles in the development of polygenic and multifactorial disorders. CHROMOSOMES The human genome is divided into 23 different chromosomes, including 22 autosomes (numbered 1-22) and the X and Y sex chromosomes (Fig. 479-1). Adult cells are diploid, meaning they contain two homologous sets of 22 autosomes and a pair of sex chromosomes. Females have two X chromosomes (XX), whereas males have one X and one Y chromosome (XY). As a consequence of meiosis, germ cells (sperm or oocytes) are haploid and contain one set of 22 autosomes and one of the sex chromosomes. At the time of fertilization, the diploid genome is reconstituted by pairing of the homologous chromosomes from the mother and father. With each cell division (mitosis), chromosomes are replicated, paired, segregated, and divided into two daughter cells. STRUCTURE OF DNA DNA is a double-stranded helix composed of four different bases: adenine (A), thymine (T), guanine (G), and cytosine (C). Adenine is paired to thymine, and guanine is paired

Guanine Cytosine H O O H N N H H C C C O O P C C N N N O- C C H H C N N C N H O H H Thymine
 Adenine O N H H3C C N H C O O P C C C C N O- C N H H N C C C N N H O T Double-strand DNA
 without histones G A C A T C G Nucleosome core Histone H2A, H2B, H4 Metaphase chromosome

Nucleosome fiber p, short arm Centromere Solenoid q, long arm Telomere Supercoiled chromatin

FIGURE 479-1 Structure of chromatin and chromosomes. Chromatin is composed of double-strand DNA that is wrapped around histone and nonhistone proteins forming nucleosomes. The nucleosomes are further organized into solenoid structures. Chromosomes assume their characteristic structure, with short (p) and long (q) arms at the metaphase stage of the cell cycle.

to cytosine, by hydrogen bond interactions that span the double helix (Fig. 479-1). DNA has several remarkable features that make it ideal for the transmission of genetic information. It is relatively stable, and the double-stranded nature of DNA and its feature of strict base-pair complementarity permit faithful replication during cell division. Complementarity also allows the transmission of genetic information from DNA → RNA → protein (Fig. 479-2). mRNA is encoded by the so-called sense or coding strand of the DNA double helix and is translated into proteins by ribosomes. The presence of four different bases provides surprising genetic diversity. In the protein-coding regions of genes, the DNA bases are arranged into codons, a triplet of bases that specifies a particular amino acid. It is possible to arrange the four bases into 64 different triplet codons (4³). Each codon specifies 1 of the 20 different amino acids, or a regulatory signal such as initiation and stop of translation. Because there are more codons than amino acids, the genetic code is degenerate; that is, most amino acids can be specified by several different codons. By arranging the codons in different combinations and in various lengths, it is possible to generate the tremendous diversity of primary protein structure. DNA length is normally measured in units of 1000 bp (kilobases, kb) or 1,000,000 bp (megabases, Mb). In the human genome, only ~1% of DNA accounts for protein-coding sequences. The noncoding DNA has multiple functional and structural roles including (1) sequences that form introns; (2) regulatory elements (promoters, enhancers, silencers, insulators); (3) sequences that generate RNAs that do not code for proteins; (4) centromeres and telomeres; (5) regions defining chromatin

structure and histone modifications; (6) various forms of repetitive sequences of variable length; and (7) pseudogenes and regions without currently discernible functional or structural roles (Fig. 479-1).

GENES A gene is a functional unit that is regulated by transcription (see below) and encodes an RNA product, which is most commonly, but not always, translated into a protein that exerts activity within or outside the cell (Fig. 479-3). Historically, genes were identified because they conferred specific traits that are transmitted from one generation to the next. Now, they are frequently characterized based on expression in various tissues (transcriptome). The size of genes is quite broad; some genes are only a few hundred base pairs, whereas others are extraordinarily large (2.3 Mb). The number of genes greatly underestimates the complexity of genetic expression, because single genes can generate multiple spliced messenger RNA (mRNA) products (isoforms), which are translated into proteins that are subject to complex posttranslational modification such as phosphorylation. Exons refer to the portion of genes that are eventually spliced together to form mRNA. Introns refer to the spacing regions between the exons that are spliced out of precursor RNAs during RNA processing. The gene locus also includes regions that are necessary to control its expression (Fig. 479-2). Current estimates predict roughly 20,000 protein-coding genes in the human genome with an average of about four different coding transcripts per gene. Remarkably, the exome only constitutes 1.14% of the genome. Of note, the number of transcripts is close to 200,000 and includes thousands of noncoding transcripts (RNAs of various length such as microRNAs [miRNA] and long noncoding RNAs [lncRNA]). These noncoding RNAs are involved in

numerous cellular processes such as transcriptional and posttranscriptional regulation of gene expression, chromatin remodeling, and protein trafficking, among others. Not surprisingly, aberrant expression and/or mutations in these RNAs play a pathogenic role in numerous diseases. CHAPTER 479 Principles of Human Genetics Histone H1 SINGLE-NUCLEOTIDE POLYMORPHISMS On average, a typical genome differs from the reference human genome at 4 to 5 million sites. Some of these variants have no impact on health, whereas others may increase or lower the risk for developing a specific disease. Remarkably, however, the primary DNA sequence of humans has ~99.9% similarity compared to that of any other human. An SNP is a variation of a single base pair in the DNA. Across human populations from distinct ethnic backgrounds, there are more than 1 billion validated SNPs (Fig. 479-3). SNPs are the most common type of sequence variation and account for

90% of all sequence variation. They occur on average every 100–300 bases and are the major source of genetic heterogeneity. SNPs that are in proximity are inherited together (e.g., they are linked) and are referred to as haplotypes (Fig. 479-4). Haplotype maps describe the nature and location of these SNP haplotypes and how they are distributed among individuals within and among populations, information that has been facilitating GWAS designed to elucidate the complex interactions among multiple genes and lifestyle factors in multifactorial disorders (see below). Moreover, haplotype analyses are useful to assess variations in responses to medications (pharmacogenomics) and environmental factors, as well as the prediction of disease predisposition. COPY NUMBER VARIATIONS Copy number variations (CNVs) are relatively large genomic regions (1 kb to several Mb) that have been duplicated or deleted on certain chromosomes and hence alter the diploid status of the DNA (Fig. 479-5). It has been estimated that 5–10% of the genome can display CNVs. When comparing the genomes of two individuals, ~0.4–0.8% of their genomes differ in terms of CNVs scattered throughout the genome. Some CNVs can increase or decrease gene dosage, potentially leading to detrimental effects if essential genes are impacted. Of note, de novo CNVs have been observed between monozygotic twins, who otherwise have identical genomes. Replication of DNA and Mitosis Genetic information in DNA is transmitted to daughter cells under two different circumstances: (1) somatic cells divide by mitosis, allowing the diploid (2n) genome to replicate itself completely in conjunction with cell division; and (2) germ cells (sperm and ova) undergo meiosis, a process that enables the reduction of the diploid (2n) set of chromosomes to the haploid state (1n).

Steroids Ca²⁺ Cytokines Growth factors Hormones Light UV-light Mechanical stress PART 16
Genes, the Environment, and Disease Regulation of Gene Expression Enhancer Silencer Nuclear
receptor Nuclear receptor HAT CoA CBP TAF GTF CREB CREB TBP Transcription factor CRE RE CAAT
TATA

mRNA Processing Posttranslational Processing FIGURE 479-2 Flow of genetic information. Multiple extracellular signals activate intracellular signal cascades that result in altered regulation of gene

expression through the interaction of transcription factors with regulatory regions of genes. RNA polymerase transcribes DNA into RNA that is processed to mRNA by excision of intronic sequences. The mRNA is translated into a polypeptide chain to form the mature protein after undergoing posttranslational processing. CBP, CREB-binding protein; CoA, co-activator; COOH, carboxyterminus; CRE, cyclic AMP responsive element; CREB, cyclic AMP response element-binding protein; GTF, general transcription factors; HAT, histone acetyl transferase; NH₂, aminoterminus; RE, response element; TAF, TBP-associated factors; TATA, TATA box; TBP, TATA-binding protein. Prior to mitosis, cells exit the resting, or G₀ state, and enter the cell cycle. After traversing a critical checkpoint in G₁, cells undergo DNA synthesis (S phase), during which the DNA in each chromosome is replicated, yielding two pairs of sister chromatids (2n → 4n). The process of DNA synthesis requires stringent fidelity in order to avoid transmitting errors to subsequent generations of cells. Genetic abnormalities of DNA mismatch/repair include xeroderma pigmentosum, Bloom's syndrome, ataxia telangiectasia, and hereditary nonpolyposis colon cancer (HNPCC), among others. Many of these disorders strongly predispose to neoplasia because of the rapid acquisition of additional mutations (Chap. 76). After completion of DNA synthesis, cells enter G₂ and progress through a second checkpoint before entering mitosis. At this stage, the chromosomes condense and are aligned along the equatorial plate at metaphase. The two identical sister chromatids, held together at the centromere, divide and migrate to opposite poles of the cell. After formation of a nuclear membrane around the two separated sets of chromatids, the cell divides and two daughter cells are formed, thus restoring the diploid (2n) state. Assortment and Segregation of Genes During Meiosis Meiosis occurs only in germ cells of the gonads. It shares certain features with mitosis but involves two distinct steps of cell division that reduce the chromosome number to the haploid state. In addition, there is active recombination that generates genetic diversity. During the first cell division, two sister chromatids (2n → 4n) are formed for each chromosome pair and there is an exchange of DNA between homologous paternal and maternal chromosomes. This process involves the formation of chiasmata, structures that correspond to the DNA segments that cross over between the maternal and paternal homologues (Fig. 479-6). Usually there is at least one crossover on each chromosomal arm; recombination occurs more frequently in female meiosis than in male meiosis. Subsequently, the chromosomes segregate randomly. Because there are 23 chromosomes, there exist 2²³ (>8 million) possible combinations of chromosomes. Together with the genetic exchanges that

Cytoplasm Nucleus RNA polymerase II DNA Transcription

hRNA

-Poly-A Tail 5' -Cap mRNA Translation Protein NH₂- -COOH occur during recombination, chromosomal segregation generates tremendous diversity, and each gamete is genetically unique. The process of recombination and the independent segregation of chromosomes provide the foundation for performing linkage analyses, whereby one attempts to correlate the inheritance of certain chromosomal regions (or linked genes) with the presence of a disease or genetic trait (see below). After the first meiotic division, which results in two daughter cells (2n), the two chromatids of each chromosome separate during a second meiotic division to yield four gametes with a haploid state (1n). When the egg is fertilized by sperm, the two haploid sets are combined, thereby restoring the diploid state (2n) in the zygote. ■ ■REGULATION OF GENE EXPRESSION Regulation by Transcription Factors The expression of genes is regulated by DNA-binding proteins that activate

or repress transcription. The number of DNA sequences and transcription factors that regulate transcription is much greater than originally anticipated. Most genes contain at least 15–20 discrete regulatory elements within 300 bp of the transcription start site. This densely packed promoter region often contains binding sites for ubiquitous transcription factors. However, factors involved in cell-specific expression may also bind to these sequences. Key regulatory elements may also reside at a large distance from the proximal promoter. The globin and the immunoglobulin genes, for example, contain locus control regions that are several kilobases away from the structural sequences of the gene. Specific groups of transcription factors that bind to these promoter and enhancer sequences provide a combinatorial code for regulating transcription. In this manner, relatively ubiquitous factors interact with more restricted factors to allow each gene to be expressed and regulated in a unique manner that is dependent on developmental state, cell type, and numerous extracellular stimuli. Regulatory factors also bind within the gene itself, particularly in the intronic regions. The

SNPs (612,977) Known Genes (1260) p22.3 p22.1 p21.3 p15.3 p15.1 p14.3 p13 p12.3 p14.1 p21.1
Chromosome 7 116.90 Mb 116.94 Mb 116.98 Mb 117.02 Mb 117.06 Mb CFTR Gene SNPs Intronic
Splice site Coding region, synonymous Coding region, nonsynonymous
FIGURE 479-3 Chromosome 7 is shown with the density of single nucleotide polymorphisms (SNPs) and genes above. A 200-kb region in 7q31.2 containing the CFTR gene is shown below. The CFTR gene contains 27 exons. Close to 2000 mutations in this gene have been found in patients with cystic fibrosis. A 20-kb region encompassing exons 4–9 is shown further amplified to illustrate the SNPs in this region.
FIGURE 479-4 The origin of haplotypes is due to repeated recombination events occurring in multiple generations. Over time, this leads to distinct haplotypes. These haplotype blocks can often be characterized by genotyping selected Tag single nucleotide polymorphisms (SNPs), an approach that facilitates performing genomewide association studies (GWAS).

CHAPTER 479 Principles of Human Genetics q31.2 q31.31 q31.33 q32.1 p21.13 q11.21 q11.22
q11.23 q21.11 q22.1 q22.3 q35 q36.1 q36.3 p12.1 p11.2 q21.3 q31.1 q33 q34 200 Kb 20 Kb
Coding region, frameshift transcription factors that bind to DNA represent only the first level of regulatory control. Other proteins—co-activators and co-repressors—interact with the DNA-binding transcription factors to generate large regulatory complexes. These complexes are subject to control by numerous cell-signaling pathways and enzymes, leading to phosphorylation, acetylation, sumoylation, and ubiquitination. Ultimately, the recruited transcription factors interact with, and stabilize, components of the basal transcription complex that assembles at the site of the TATA box and initiator region. This basal transcription factor complex consists of >30 different proteins. Gene transcription occurs when RNA polymerase begins to synthesize RNA from the DNA template. A large number of identified genetic diseases involve transcription factors (Table 479-2). The field of functional genomics is based on the concept that understanding alterations of gene expression under various physiologic and pathologic conditions provides insight into the underlying functional role of the gene. The ENCODE (Encyclopedia of DNA Elements) project aims at identifying and annotating all functional sequences in the human genome. By revealing specific gene expression profiles, this knowledge can be of diagnostic and therapeutic relevance. The large-scale study of expression profiles is referred to as transcriptomics because the complement of mRNAs transcribed by the cellular genome is called the transcriptome. Most studies of gene expression have focused on the regulatory DNA elements of genes that control transcription. However, it must be

log₂ (ratio)

-1 -2 Chromosome 8 FIGURE 479-5 Copy number variations (CNV) encompass relatively large regions of the genome that have been duplicated or deleted. Chromosome 8 is shown with a CNV detected by genomic hybridization. An increase in the signal strength indicates a duplication, whereas a decrease reflects a deletion of the covered chromosomal regions. emphasized that gene expression requires a series of steps, including mRNA processing, protein translation, and posttranslational modifications, all of which are actively regulated (Fig. 479-2). Epigenetic Regulation of Gene Expression (see Chap. 497)

Epigenetics describes mechanisms and phenotypic changes that are not a result of variation in the primary DNA nucleotide sequence but are caused by secondary modifications of DNA or histones. These modifications include heritable changes such as X-inactivation and imprinting, but they can also result from dynamic posttranslational protein modifications in response to environmental influences such as diet, age, or drugs. The epigenetic modifications result in altered expression of individual genes or chromosomal loci encompassing multiple genes. The term epigenome describes the constellation of covalent modifications of DNA and histones that impact chromatin structure, as well as noncoding transcripts that modulate the transcriptional activity of DNA. Although the primary DNA sequence is usually identical in all cells of an organism, sex- and tissue-specific changes in the epigenome contribute to determining the transcriptional signature of a cell (transcriptome) and hence the protein expression profile (proteome). Mechanistically, DNA and histone modifications can result in the activation or silencing of gene expression (Fig. 479-7). DNA methylation involves the addition of a methyl group to cytosine residues. This is usually restricted to cytosines of CpG dinucleotides, which are abundant throughout the genome. Methylation of these dinucleotides is thought to represent a defense mechanism that minimizes the expression of sequences that have been incorporated into the genome such as retroviral sequences. CpG dinucleotides also exist in so-called CpG islands, stretches of DNA characterized by a high CG content, which are found in the majority of human gene promoters. CpG islands in promoter regions are typically unmethylated, and the lack of methylation facilitates transcription. Histone methylation involves the addition of a methyl group to lysine residues in histone proteins (Fig. 479-7). Depending on the specific lysine residue being methylated, this alters chromatin configuration, making it either more open or tightly packed. Acetylation of histone proteins is another well-characterized mechanism that results in an

open chromatin configuration, which favors active transcription. Acetylation is generally more dynamic than methylation, and many transcriptional activation complexes have histone acetylase activity, whereas repressor complexes often contain deacetylases and remove acetyl groups from histones. Other histone modifications include, among others, phosphorylation and sumoylation. Furthermore, noncoding RNAs and RNA regulatory networks that bind to DNA have a significant impact on transcriptional activity. Physiologically, epigenetic mechanisms play an important role in several instances. For example, X-inactivation refers to the relative silencing of one of the two X chromosome copies present in females. The inactivation process is a form of dosage compensation such that females (XX) do not generally express twice as many X-chromosomal gene products as males (XY). In a given cell, the choice of which chromosome is inactivated occurs randomly in

humans. But once the maternal or paternal X chromosome is inactivated, it will remain inactive, and this information is transmitted with each cell division. The X-inactive specific transcript (Xist) gene encodes a long non-coding RNA (lncRNA) that mediates gene silencing on one of the X chromosomes. The inactive X chromosome is highly methylated and has low levels of histone acetylation. While the majority of X-chromosomal genes are silenced by X-inactivation, ~15% escape inactivation and are expressed. Epigenetic gene inactivation also occurs on selected chromosomal regions of autosomes, a phenomenon referred to as genomic imprinting. Through this mechanism, a small subset of genes is only expressed in a monoallelic fashion. Imprinting is a

a A a a A B b B b b B C c C c c C D d D d d D Chromatids Homologous chromosomes A a a A A a a A A a a A B b b B B b b B B b b B c c C C c c C C C c c C d d D D D d d D D d d D Crossover Double crossover No crossover A a a A A a A a a A B b b B B b B b b B b b B c c C C c c C C C c c C D d D d d d D D D d d D Recombination in gametes Recombination in gametes No recombination in gametes

FIGURE 479-6 Crossing-over and genetic recombination. During chiasma formation, either of the two sister chromatids on one chromosome pairs with one of the chromatids of the homologous chromosome. Genetic recombination occurs through crossing-over and results in recombinant and nonrecombinant chromosome segments in the gametes. Together with the random segregation of the maternal and paternal chromosomes, recombination contributes to genetic diversity and forms the basis of the concept of linkage.

TABLE 479-2 Selected Examples of Diseases Caused by Mutations and Rearrangements in Transcription Factors

| TRANSCRIPTION FACTOR CLASS | EXAMPLE | ASSOCIATED DISORDER |
|---|---|---|
| Nuclear receptors | Androgen receptor | Complete or partial androgen insensitivity (recessive missense mutations) |
| | Spinobulbar muscular atrophy (CAG repeat expansion) | Zinc finger proteins |
| | WT1 | WAGR syndrome: Wilms' tumor, aniridia, genitourinary malformations, mental retardation |
| Basic helix-loop-helix | MITF | Waardenburg's syndrome type 2A |
| Homeobox | IPF1 | Maturity onset of diabetes mellitus type 4 (monoallelic mutation/ haploinsufficiency) |
| | Pancreatic agenesis (biallelic mutations) | Leucine zipper |
| | Retina leucine zipper (NRL) | Autosomal dominant retinitis pigmentosa |
| High mobility group (HMG) proteins | SRY | Sex reversal |
| Forkhead | HNF4 α , HNF1 α , HNF1 β | Maturity onset of diabetes mellitus types 1, 3, 5 |
| Paired box | PAX3 | Waardenburg's syndrome types 1 and 3 |
| T-box | TBX5 | Holt-Oram syndrome (thumb anomalies, atrial or ventricular septum defects, phocomelia) |
| Cell cycle control proteins | P53 | Li-Fraumeni syndrome, other cancers |
| Co-activators | CREB binding protein (CREBBP) | Rubinstein-Taybi syndrome |
| General transcription factors | TATA-binding protein (TBP) | Spinocerebellar ataxia 17 (CAG expansion) |
| Transcription elongation factor | VHL | von Hippel-Lindau syndrome (renal cell carcinoma, pheochromocytoma, pancreatic tumors, hemangioblastomas) |
| Autosomal dominant inheritance, somatic inactivation of second allele (Knudson two-hit model) | Runt | RUNX1 |
| Familial thrombocytopenia with propensity to acute myelogenous leukemia | Chimeric proteins due to translocations | PML-RAR |
| Acute promyelocytic leukemia | t(15;17)(q22;q11.2-q12) | translocation |

Abbreviations: CREB, cAMP responsive element-binding protein; HNF, hepatocyte nuclear factor; PML, promyelocytic leukemia; RAR, retinoic acid receptor; SRY, sexdetermining region Y; VHL, von Hippel-Lindau. heritable and leads to the preferential expression of one of the parental alleles, which deviates from the usual biallelic expression seen for the majority of genes. Remarkably, imprinting can be limited to a subset of tissues. Imprinting is mediated through DNA methylation of one of the alleles. The epigenetic marks on imprinted genes are maintained throughout life, but during zygote formation, they are activated or inactivated in a sex-specific manner (imprint reset) (Fig. 479-8), which allows a differential expression pattern in the fertilized egg and the subsequent mitotic divisions. Appropriate expression of imprinted genes is important

for normal development and cellular functions. Imprinting defects and uniparental disomy, which is the inheritance of two chromosomes or chromosomal regions from the same parent, are the cause of several developmental disorders such as Beckwith-Wiedemann syndrome, Silver-Russell syndrome, Angelman's syndrome, and Prader-Willi syndrome (see below). Monoallelic loss-of-function mutations in the *GNAS1* gene lead to Albright's hereditary osteodystrophy (AHO). Paternal transmission of *GNAS1* mutations leads to an isolated AHO

phenotype (pseudopseudohypoparathyroidism), whereas maternal transmission leads to AHO in combination with hormone resistance to parathyroid hormone, thyrotropin, and gonadotropins (pseudohypoparathyroidism type IA). These phenotypic differences are explained by tissue-specific imprinting of the *GNAS1* gene, which is expressed primarily from the maternal allele in the thyroid, gonadotropes, and the proximal renal tubule. In most other tissues, the *GNAS1* gene is expressed biallelically. In patients with isolated renal resistance to parathyroid hormone (pseudohypoparathyroidism type IB), defective imprinting of the *GNAS1* gene results in decreased $G\alpha$ expression in the proximal renal tubules. Rett syndrome is an X-linked dominant disorder resulting in developmental regression and stereotypic hand movements in affected girls. It is caused by mutations in the *MECP2* gene, which encodes a methyl-binding protein. The ensuing aberrant methylation results in abnormal gene expression in neurons, which are otherwise normally developed.

CHAPTER 479 Principles of Human Genetics Remarkably, epigenetic differences also occur among monozygotic twins. Although twins are epigenetically indistinguishable during the early years of life, older monozygotic twins exhibit differences in the overall content and genomic distribution of DNA methylation and histone acetylation, which would be expected to alter gene expression in various tissues. In cancer, the epigenome is characterized by simultaneous losses and gains of DNA methylation in different genomic regions, as well as repressive histone modifications. Hyper- and hypomethylation are associated with mutations in genes that control DNA methylation. Hypomethylation is thought to remove normal control mechanisms that prevent expression of repressed DNA regions. It is also associated with genomic instability. Hypermethylation, in contrast, results in the silencing of CpG islands in promoter regions of genes, including tumor-suppressor genes. Epigenetic alterations are more easily reversible compared to genetic changes; modification of the epigenome with demethylating agents and histone deacetylases is being used in the treatment of various malignancies. ■ ■ TRANSMISSION OF GENETIC DISEASE

Origins and Types of Mutations The term mutation or variant is used to designate the process of generating genetic variations as well as the effect of these alterations. A mutation can be defined as any change in the primary nucleotide sequence of DNA regardless of its functional consequences, although it often has a negative connotation. There has been a shift towards using the more neutral term variant to describe sequence changes, and it is now recommended by several professional organizations and guidelines instead of mutation. Some variants may be lethal, others are less deleterious, and some may confer an evolutionary advantage. Variations can occur in the germline (sperm or oocytes); these can be transmitted to progeny. Alternatively, variants can occur during embryogenesis or in somatic tissues. Variations that occur during development lead to mosaicism, a situation in which tissues are composed of cells with different genetic constitutions. If the germline is mosaic, a mutation can be transmitted to some progeny but not others, which sometimes leads to confusion in assessing the pattern of inheritance. Somatic mutations that do not affect cell survival can sometimes be detected because of variable phenotypic effects in tissues

(e.g., pigmented lesions in McCune-Albright syndrome). Other somatic mutations are associated with neoplasia because they confer a growth advantage to cells. Epigenetic events may also influence gene expression or facilitate genetic damage. With the exception of triplet nucleotide repeats, which can expand (see below), variations are usually stable. Sequence variants are structurally diverse—they can involve the entire genome, as in triploidy (one extra set of chromosomes), or gross numerical or structural alterations in chromosomes or individual genes. Large deletions may affect a portion of a gene or an entire gene, or, if several genes are involved, they may lead to a contiguous gene syndrome. Unequal crossing-over between homologous genes can result in fusion gene mutations, as illustrated by color blindness. Variations involving single nucleotides are referred to as point mutations.

Methylated DNA Cytosine Methylation NH₂ NH₂ N CH₃ N PART 16 Genes, the Environment, and Disease O N O N Histone Acetylation Unmethylated DNA Histone Modifications Acetylation Phosphorylation Methylation Transcription NH₂ FIGURE 479-7 Epigenetic modifications of DNA and histones. Methylation of cytosine residues is associated with gene silencing. Methylation of certain genomic regions is inherited (imprinting), and it is involved in the silencing of one of the two X chromosomes in females (X-inactivation). Alterations in methylation can also be acquired, e.g., in cancer cells. Covalent posttranslational modifications of histones play an important role in altering DNA accessibility and chromatin structure and hence in regulating transcription. Histones can be reversibly modified in their aminoterminal tails, which protrude from the nucleosome core particle, by acetylation of lysine, phosphorylation of serine, methylation of lysine and arginine residues, and sumoylation. Acetylation of histones by histone acetylases (HATs), e.g., leads to unwinding of chromatin and accessibility to transcription factors. Conversely, deacetylation by histone deacetylases (HDACs) results in a compact chromatin structure and silencing of transcription. Substitutions are called transitions if a purine is replaced by another purine base (A ↔ G) or if a pyrimidine is replaced by another pyrimidine (C ↔ T). Changes from a purine to a pyrimidine, or vice versa, are referred to as transversions. If the DNA sequence change occurs in a coding region and alters an amino acid, it is called a missense mutation. Depending on the functional consequences of such a missense mutation, amino acid substitutions in different regions of the protein can lead to distinct phenotypes. Variants can occur in all domains of a gene (Fig. 479-9). A point mutation occurring within the coding region leads to an amino acid substitution if the codon is altered (Fig. 479-10). Point mutations that introduce a premature stop codon result in a truncated or missing protein. Large deletions may affect a portion of a gene or an entire gene, whereas small deletions and insertions alter the reading frame if they do not represent a multiple of three bases. These “frameshift” mutations, also designated as amphigoric amino acid changes, lead to an entirely altered carboxy terminus. Mutations in intronic sequences or in exon junctions may destroy or create splice donor or splice acceptor sites. Variants may also be found in the regulatory sequences of genes, resulting in reduced or enhanced gene transcription. Certain DNA sequences are particularly susceptible to mutagenesis. Successive pyrimidine residues (e.g., T-T or C-C) are subject to the formation of ultraviolet light-induced photoadducts. If these pyrimidine dimers are not repaired by the nucleotide excision repair pathway, mutations will be introduced after DNA synthesis. The dinucleotide C-G, or CpG, is also a hot spot for a specific type of alteration. In this case, methylation of the cytosine is associated with an enhanced rate of deamination to uracil, which is then replaced with thymine. This C → T transition (or G → A on the opposite strand) accounts for at least one-third of point mutations associated with polymorphisms and mutations. In addition to the fact that certain types of mutations (C → T or G → A) are relatively

common, the nature of the genetic code also results in overrepresentation of certain amino acid substitutions. Polymorphisms are sequence variations that have a frequency of at least 1%. Usually, they do not result in a perceptible phenotype; the term variant is now preferred for the description of these sequence changes because allele frequency and functional consequences are often not known. Often, they consist of single base-pair substitutions that do not alter the protein coding sequence because of the degenerate

nature of the genetic code (synonymous polymorphism), although it is possible that some might alter mRNA stability, translation, or the amino acid sequence (non synonymous polymorphism) (Fig. 479-10). The detection of sequence variants poses a practical problem because it is often unclear whether it creates a change with functional consequences or a benign variation. In this situation, the sequence alteration is also described as variant of unknown significance (VUS).

METHYLATION Methylation represents an important cause of genetic diversity as well as disease. Mutation rates are difficult to determine in humans because many mutations are silent and because testing is often not adequate to detect the phenotypic consequences. Mutation rates vary in different genes but are estimated to occur at a rate of $\sim 10^{-10}$ – 10^{-6} /bp per cell division.

MUTATION RATES Germline mutation rates (as opposed to somatic mutations) are relevant in the transmission of genetic disease. Because the population of oocytes is established very early in development, only ~ 20 cell divisions are required for completed oogenesis, whereas spermatogenesis involves ~ 30 divisions by the time of puberty and 20 cell divisions each year thereafter. Consequently, the probability of acquiring new point mutations is much greater in the male germline than the female germline, in which rates of aneuploidy are increased. Thus, the incidence of new point mutations in spermatogonia increases with paternal age (e.g., achondrodysplasia, Marfan's syndrome, neurofibromatosis). It is estimated that about 1 in 10 sperm carries a new deleterious mutation. The rates for new mutations are calculated most readily for autosomal dominant and X-linked disorders and are $\sim 10^{-5}$ – 10^{-6} /locus per generation. Because most monogenic diseases are relatively rare, new mutations account for a significant fraction of cases. This is important in the context of genetic counseling because a new mutation can be transmitted to the affected individual, but this does not necessarily imply that the parents are at risk to transmit the disease to other children. An exception to this is when the new mutation occurs early in germline development, leading to gonadal mosaicism.

UNEQUAL CROSSING-OVER Normally, DNA recombination in germ cells occurs with remarkable fidelity to maintain the precise junction sites for the exchanged DNA sequences (Fig. 479-6). However, mispairing of homologous sequences leads to unequal crossover, with gene duplication on one of the chromosomes and gene deletion on the other chromosome. A significant fraction of growth hormone (GH) gene deletions, for example, involve unequal crossing-over (Chap. 391). The GH gene is a member of a large gene cluster that includes a GH variant gene as well as several structurally related chorionic somatomammotropin genes and pseudogenes (highly homologous but functionally inactive relatives of a normal gene). Because such gene clusters contain multiple homologous DNA sequences arranged in tandem, they are particularly prone to undergo recombination and, consequently, gene duplication or deletion. Duplication of the PMP22 gene because of unequal crossing-over results in increased gene dosage and type IA Charcot-Marie-Tooth disease. In contrast, unequal crossing-over resulting in deletion of PMP22 causes a distinct neuropathy called hereditary neuropathy with liability to pressure palsies (HNPP) (Chap. 457). Glucocorticoid-remediable aldosteronism (GRA) is caused by a gene fusion or rearrangement involving the genes that encode aldosterone synthase (CYP11B2) and steroid 11 β -hydroxylase (CYP11B1), normally

arranged in tandem on chromosome 8q. These two genes are 95% identical, predisposing to gene duplication and deletion by

Maternal somatic cell mat pat Active unmethylated Inactive methylated Inactive methylated
Germline development: Imprint reset Maternal germline Paternal germline mat pat Active unmethylated Active unmethylated Inactive methylated Zygote pat mat Inactive methylated Active unmethylated

FIGURE 479-8 A few genomic regions are imprinted in a parent-specific fashion. The unmethylated chromosomal regions are actively expressed, whereas the methylated regions are silenced. In the germline, the imprint is reset in a parent-specific fashion: both chromosomes are unmethylated in the maternal (mat) germline and methylated in the paternal (pat) germline. In the zygote, the resulting imprinting pattern is identical with the pattern in the somatic cells of the parents. unequal crossing-over. The rearranged gene product contains the regulatory regions of 11 β -hydroxylase fused to the coding sequence of aldosterone synthetase. Consequently, the latter enzyme is expressed in the adrenocorticotrophic hormone (ACTH)-dependent zona fasciculata of the adrenal gland, resulting in overproduction of mineralocorticoids and hypertension (Chap. 398). Gene conversion refers to a nonreciprocal exchange of homologous genetic information. It has been used to explain how an internal portion of a gene is replaced by a homologous segment copied from another allele or locus; these genetic alterations may range from a few nucleotides to a few thousand nucleotides. As a result of gene conversion, it is possible for short DNA segments of two chromosomes to be identical, even though these sequences are distinct in the parents. A practical consequence of this phenomenon is that nucleotide substitutions can occur during gene conversion between related genes, often altering the function of the gene. In disease states, gene conversion often involves intergenic exchange of DNA between a gene and a related pseudogene. For example, the 21-hydroxylase gene (CYP21A2) is adjacent to a nonfunctional pseudogene (CYP21A1P). Many of the

nucleotide substitutions that are found in the CYP21A2 gene in patients with congenital adrenal hyperplasia correspond to sequences that are present in the CYP21A1P pseudogene, suggesting gene conversion as one cause of mutagenesis. In addition, mitotic gene conversion has been suggested as a mechanism to explain revertant mosaicism in which an inherited mutation is "corrected" in certain cells. For example, patients with autosomal recessive generalized atrophic benign epidermolysis bullosa have acquired reverse mutations in one of the two mutated COL17A1 alleles, leading to clinically unaffected patches of skin.

Paternal somatic cell pat mat CHAPTER 479 Active unmethylated Principles of Human Genetics
INSERTIONS AND DELETIONS Although many instances of insertions and deletions occur as a consequence of unequal crossing-over, there is also evidence for internal duplication, inversion, or deletion of DNA sequences. The fact that certain deletions or insertions appear to occur repeatedly as independent events indicates that specific regions within the DNA sequence predispose to these errors. For example, certain regions of the DMD gene, which encodes dystrophin, appear to be hot spots for deletions and result in muscular dystrophy (Chap. 460). Some regions within the human genome are rearrangement hot spots and lead to CNVs. pat mat Inactive methylated

ERRORS IN DNA REPAIR Because mutations caused by defects in DNA repair accumulate as somatic cells divide, these types of mutations are particularly important in the context of neoplastic disorders. Several genetic disorders involving DNA repair enzymes underscore their importance. Patients with xeroderma pigmentosum have defects in DNA damage recognition or in the nucleotide excision

and repair pathway (Chap. 81). Exposed skin is dry and pigmented and is extraordinarily sensitive to the mutagenic effects of ultraviolet irradiation. Variants in more than 10 different genes have been shown to cause the different forms of xeroderma pigmentosum. Ataxia-telangiectasia is a multisystem disorder that includes progressive neurodegenerative cerebellar ataxia, immunologic defects, telangiectatic lesions, lymphomas and leukemias, and hyper sensitivity to ionizing radiation (Chap. 450). The discovery of the ataxia-telangiectasia mutated (ATM) gene revealed that it is homologous to genes involved in DNA repair and control of cell cycle checkpoints. Mutations in the ATM gene give rise to defects in meiosis as well as increasing susceptibility to damage from ionizing radiation. Fanconi's anemia is also associated with an increased risk of multiple acquired genetic abnormalities. It is characterized by diverse congenital anomalies and a strong predisposition to develop aplastic anemia and acute myelogenous leukemia (Chap. 109). Cells from these patients are susceptible to chromosomal breaks caused by a defect in genetic recombination. It can be caused by mutations in the multiple genes forming the Fanconi's anemia pathway, which is involved in DNA repair and replication. HNPCC (Lynch syndrome) is characterized by autosomal dominant transmission of colon cancer, young age (<50 years) of presentation, predisposition to lesions in the proximal large bowel, and associated malignancies such as uterine cancer and ovarian cancer. HNPCC is predominantly caused by mutations in one of several different mismatch repair (MMR) genes including MutS

A *

PART 16 Genes, the Environment, and Disease

intron 2 intron 1 Poly A Promoter 5'UTR ε Gγ Aγ ψβ
 β δ -10 kb 0 kb 10 kb 20 kb 30 kb 40 kb 50 kb 60 kb β-Globin Gene Cluster

FIGURE 479-9 Point mutations causing a thalassemia as example of allelic heterogeneity. The β-globin gene is located in the globin gene cluster. Point mutations occur in the promoter, the CAP site, the 5'-untranslated region, the initiation codon, each of the three exons, the introns, or the polyadenylation signal. Many mutations introduce missense or nonsense mutations, whereas others cause defective RNA splicing. Not shown here are deletion mutations of the β-globin gene or larger deletions of the globin locus that can also result in thalassemia. , promoter mutations; *, CAP site; , 5'UTR; 1 , initiation codon; , defective RNA processing; , missense and nonsense mutations; A, Poly A signal.

homologue 2 (MSH2), MutL homologue 1 and 6 (MLH1, MLH6), MSH6, PMS1, and PMS2 (Chap. 86). These proteins are involved in the detection of nucleotide mismatches and in the recognition of slipped-strand trinucleotide repeats. Germline mutations in these genes lead to microsatellite instability and a high mutation rate in colon cancer. Genetic screening tests for this disorder are now being used for families considered to be at risk. Recognition of HNPCC allows early screening with colonoscopy and the implementation of prevention strategies using nonsteroidal anti-inflammatory drugs.

UNSTABLE DNA SEQUENCES Trinucleotide repeats may be unstable and expand beyond a critical number. Mechanistically, the expansion is thought to be caused by unequal recombination and slipped mispairing. A premutation represents a small increase in trinucleotide copy number. In subsequent generations, the expanded repeat may increase further in length and result in an increasingly severe phenotype, a process called dynamic mutation (see below for discussion of anticipation). Trinucleotide expansion was first recognized as a cause of the fragile X syndrome, one of the most common causes of intellectual disability. Other disorders arising from a similar mechanism include

Wild-type AA DNA A GCA CTC L CTA S TCG H CAC A GCT R
 CGG E GAG G GGC E L Silent mutation AA DNA A CGT GCA L CTC L CTA S TCG H CAC A GCT R
 GAG G GGC E E Missense mutation AA DNA A CCG E GCA L CTC L CTA S TCG H CAC A GCT GAG G

GGC E P Nonsense mutation AA DNA A GCA L CTC L CTA S TCG H CAC A GCT R CGG E GAG G GGC 1 bp Deletion with frameshift AA DNA A GCA L CTC L CTA CGC ACG CTC GGG AGG GCG R T L G R A A B FIGURE 479-10 A. Examples of mutations (now commonly referred to as variations). The coding strand is shown with the encoded amino acid sequence. B. Chromatograms of sequence analyses after amplification of genomic DNA by polymerase chain reaction.

Huntington's disease, X-linked spino bulbar muscular atrophy, and myotonic dystrophy. Malignant cells are also characterized by genetic instability, indicating a breakdown in mechanisms that regulate DNA repair and the cell cycle. Functional Consequences of Mutations Functionally, mutations can be broadly classified as gain-of-function and loss-of-function mutations. Gain-of-function mutations are typically dominant (e.g., they result in phenotypic alterations when a single allele is affected). Inactivating mutations are usually recessive, and an affected individual is homozygous or compound heterozygous (e.g., carrying two different mutant alleles of the same gene) for the disease-causing mutations. Alternatively, mutation in a single allele can result in haploinsufficiency, a situation in which one normal allele is not sufficient to maintain a normal phenotype. Haploinsufficiency is a commonly observed mechanism in diseases associated with mutations in transcription factors (Table 479-2). Remarkably, the clinical features among patients with an identical mutation often vary significantly. One mechanism underlying this variability consists in the influence of modifying genes. Haploinsufficiency can also affect the expression of rate-limiting enzymes. For example, haploinsufficiency in enzymes involved in heme synthesis can cause porphyrias (Chap. 428). An increase in dosage of a gene product may also result in disease, as illustrated by the duplication of the DAX1 (NR0B1) gene in dosage-sensitive sex reversal (Chap. 402). Mutation in a single allele can also result in loss of function due to a dominant-negative effect. In this case, the mutated allele interferes with the function of the normal (wild type) gene product by one of several different mechanisms: (1) a mutant protein may interfere with the function of a multimeric protein complex, as illustrated by mutations in type 1 collagen (COL1A1, COL1A2) genes in osteogenesis imperfecta (Chap. 425); (2) a mutant protein may occupy binding sites on proteins or promoter response elements, as illustrated by thyroid hormone resistance β , a disorder in which Wild-type GAA N AAT E GAG S AGC F T T C T A C C D G A C F T T C I A T A C T G C GAA N AAT E GAG S AGC Heterozygous point mutation F T T C I A T A C T G C F T T C T A C C D G A C GAA N AAT E GAG S AGC T A C Y TAA AAT GAG AGC Homozygous point mutation X F T T C T A C C Y T A C F T T C I A T A C T G C AAA ATG AGA GC K M R

inactivated thyroid hormone receptor β binds to target genes and functions as an antagonist of normal receptors (Chap. 394); or (3) a mutant protein can be cytotoxic as in α 1 antitrypsin deficiency (Chap. 303) or autosomal dominant neurohypophyseal diabetes insipidus (Chap. 393), in which the abnormally folded proteins are trapped within the endoplasmic reticulum and ultimately cause cellular damage. Genotype and Phenotype • ALLELES, GENOTYPES, AND HAPLO TYPES An observed trait is referred to as a phenotype; the genetic information defining the phenotype is called the genotype. Alternative forms of a gene or a genetic marker are referred to as alleles. Alleles may be polymorphic variants of nucleic acids that have no apparent effect on gene expression or function. In other instances, these variants may have subtle effects on gene expression, thereby conferring adaptive advantages associated with genetic diversity. On the other hand, allelic variants may reflect mutations that clearly alter the function of a gene product. The common Glu6Val (E6V) sickle cell mutation in the β -globin gene and the Δ F508 deletion of phenylalanine (F) in the CFTR gene are examples of allelic variants of these genes that result in

disease. Because each individual has two copies of each chromosome (one inherited from the mother and one inherited from the father), an individual can have only two alleles at a given locus. However, there can be many different alleles in the population. The normal or common allele is usually referred to as wild type. When alleles at a given locus are identical, the individual is homozygous. Inheriting identical copies of a mutant allele occurs in many autosomal recessive disorders, particularly in circumstances of consanguinity or isolated populations. If the alleles are different on the maternal and the paternal copy of the gene, the individual is heterozygous at this locus (Fig. 479-10). If two different mutant alleles are inherited at a given locus, the individual is said to be a compound heterozygote. Hemizygous is used to describe males with a mutation in an X chromosomal gene or a female with a loss of one X chromosomal locus. Genotypes describe the specific alleles at a particular locus. For example, there are three common alleles (E2, E3, E4) of the apolipoprotein E (APOE) gene. The genotype of an individual can therefore be described as APOE3/4 or APOE4/4 or any other variant. These designations indicate which alleles are present on the two chromosomes in the APOE gene at locus 19q13.2. In other cases, the genotype might be assigned arbitrary numbers (e.g., 1/2) or letters (e.g., B/b) to distinguish different alleles. A haplotype refers to a group of alleles that are closely linked together at a genomic locus (Fig. 479-4). Haplotypes are useful for tracking the transmission of genomic segments within families and for detecting evidence of genetic recombination if the crossover event occurs between the alleles (Fig. 479-6). As an example, various alleles of the histocompatibility locus antigens (HLA) at the major histocompatibility complex (MHC) on chromosome 6 are used to establish haplotypes associated with certain disease states. For example, 21-hydroxylase deficiency, complement deficiency, and hemochromatosis are each associated with specific HLA haplotypes. It is now recognized that these genes lie in close proximity to the HLA locus, which explains why HLA associations were identified even before the disease genes were cloned and localized. In other cases, specific HLA associations with diseases such as ankylosing spondylitis (HLA-B27) or type 1 diabetes mellitus (HLA-DR4) reflect the role of specific HLA allelic variants in susceptibility to these autoimmune diseases. The characterization of common SNP haplotypes in numerous populations from different parts of the world has provided the necessary tools for association studies designed to detect genes involved in the pathogenesis of complex disorders (Table 479-1). The presence or absence of certain haplotypes can also be relevant for the customized choice of medical therapies (pharmacogenomics) or may have value for preventive strategies. Genotype-phenotype correlation describes the association of a specific mutation and the resulting phenotype. The phenotype may differ depending on the location or type of the mutation in some genes. For example, in von Hippel-Lindau disease, an autosomal dominant multisystem disease that can include renal cell carcinoma, hemangioblastomas, and pheochromocytomas, among others, the phenotype varies greatly, and the identification of the specific mutation can be clinically useful in order to predict the spectrum of disease manifestations.

ALLELIC HETEROGENEITY Allelic heterogeneity refers to the fact that different mutations in the same genetic locus can cause an identical or similar phenotype. For example, many different mutations of the β -globin locus can cause β thalassemia (Table 479-3) (Fig. 479-9). In essence, allelic heterogeneity reflects the fact that many different mutations can alter protein structure and function. For this reason, maps of inactivating mutations in genes usually show a near-random distribution. Exceptions include (1) a founder effect, in which a particular mutation that does not affect reproductive capacity can be traced to a single individual; (2) "hot spots" for mutations, in

which the nature of the DNA sequence predisposes to a recurring mutation; and (3) localization of mutations to certain domains that are particularly critical for protein function. Allelic heterogeneity creates a practical problem for genetic testing because one must often examine the entire genetic locus for mutations, because these can differ in each patient. For example, ~2000 variants have been identified in the CFTR gene to date, although some of them are very rare and some may not be disease-causing (Fig. 479-3). Mutational analysis may initially focus on a panel of mutations that are particularly frequent (often taking the ethnic background of the patient into account), but a negative result does not exclude the presence of a mutation elsewhere in the gene. Until recently, mutational analyses tended to focus on the coding region of a gene without considering regulatory and intronic regions. However, disease-causing mutations may be located outside the coding regions, so negative results need to be interpreted with caution. The increasingly wide spread access to comprehensive sequencing technologies, WES and WGS, greatly facilitates unbiased mutational analyses. However, comprehensive sequencing can result in significant diagnostic challenges because the detection of a sequence alteration is not always sufficient to establish that it has a causal role (VUS).

CHAPTER 479 Principles of Human Genetics PHENOTYPIC HETEROGENEITY Phenotypic heterogeneity occurs when more than one phenotype is caused by allelic mutations (e.g., by different mutations in the same gene) (Table 479-3). For example, laminopathies are monogenic multisystem disorders that result from mutations in the LMNA gene, which encodes the nuclear lamins A and C. Multiple autosomal dominant and recessive disorders are caused by mutations in the LMNA gene. They include several forms of lipodystrophies, Emery-Dreifuss muscular dystrophy, progeria syndromes, a form of neuronal Charcot-Marie-Tooth disease (type 2B1), and a group of overlapping syndromes. Remarkably, hierarchical cluster analysis has revealed that the phenotypes vary depending on the position of the mutation (genotype-phenotype correlation). Similarly, identical mutations in the FGFR2 gene can result in very distinct phenotypes: Crouzon's syndrome (craniofacial synostosis) or Pfeiffer's syndrome (acrocephalopolysyndactyly).

LOCUS OR NONALLELIC HETEROGENEITY AND PHENOCOPIES Nonallelic or locus heterogeneity refers to the situation in which a similar disease phenotype results from mutations at different genetic loci (Table 479-3). This often occurs when more than one gene product produces different subunits of an interacting complex or when different genes are involved in the same genetic cascade or physiologic pathway. For example, osteogenesis imperfecta can arise from mutations in two different procollagen genes (COL1A1 or COL1A2) that are located on different chromosomes and can involve multiple other genes (Chap. 425). The effects of inactivating mutations in these two genes are similar because the protein products comprise different subunits of the helical collagen fiber. Similarly, muscular dystrophy syndromes can be caused by mutations in various genes, consistent with the fact that it can be transmitted in an X-linked (Duchenne or Becker), autosomal dominant (limb-girdle muscular dystrophy type 1), or autosomal recessive (limb-girdle muscular dystrophy type 2) manner (Chap. 460). Mutations in the X-linked DMD gene, which encodes dystrophin, are the most common cause of muscular dystrophy. This feature reflects the large size of the gene (2.3 MB, 79 exons), as well as the fact that the phenotype is expressed in hemizygous males because they have only

TABLE 479-3 Selected Examples of Phenotypic Heterogeneity and Locus Heterogeneity

| GENE | PROTEIN | PHENOTYPE | INHERITANCE | OMIM |
|--------------------------------|-------------|--|-------------|-------------------------------------|
| LMNA | Lamin A/C | Emery-Dreifuss muscular dystrophy (AD) | PART 16 | Genes, the Environment, and Disease |
| Familial partial lipodystrophy | Dunnigan AD | | | |

Hutchinson-Gilford progeria AD

Atypical Werner's syndrome AD

Dilated cardiomyopathy 1A AD

Familial atrial fibrillation 3 AD

Charcot-Marie-Tooth type 2B1 AR

KRAS Noonan's syndrome AD

Cardio-facio-cutaneous syndrome 1 AD

Locus Heterogeneity PHENOTYPE GENE CHROMOSOMAL LOCATION PROTEIN Familial hypertrophic cardiomyopathy MYH7 14q11.2 Myosin heavy chain beta Genes encoding sarcomeric proteins TNNT2 1q32.1 Troponin-T2 TPM1 15q22.2 Tropomyosin alpha MYBPC3 11p11q Myosin-binding protein C TNNC1 19q13.4 Troponin 1 MYL2 12q24.11 Myosin light chain 2 MYL3 3p21.31 Myosin light chain 3 TTN 2q31.2 Cardiac titin ACTC 15q14 Cardiac alpha actin MYH6 14q11.2 Myosin heavy chain alpha MYLK2 20q11.21 Myosin light-peptide kinase CAV3 3p25 Caveolin 3 Genes encoding nonsarcomeric proteins MT-T1 Mitochondrial tRNA isoleucine MT-TG Mitochondrial tRNA glycine PRKAG2 7q36.1 AMP-activated protein kinase γ 2 subunit DMPK 19q13.32 Myotonic protein kinase (myotonic dystrophy) FRDA 9q21.11 Frataxin (Friedreich's ataxia) Polycystic kidney disease PKD1 16p13.3 Polycystin 1 (AD) PKD2 4q22.1 Polycystin 2 (AD) PKHD1 6p21.1-p12.2

Fibrocystin/polyductin (AR) Noonan's syndrome PTPN11 12q24.13 Protein-tyrosine phosphatase 2c KRAS 12p12.1 KRAS Abbreviations: AD, autosomal dominant; AR, autosomal recessive; OMIM, Online Mendelian Inheritance in Man. a single copy of the X chromosome. Dystrophin is associated with a large protein complex linked to the membrane-associated cytoskeleton in muscle. Mutations in several different components of this protein complex can also cause muscular dystrophy syndromes. Although the phenotypic features of some of these disorders are distinct, the phenotypic spectrum caused by mutations in different genes overlaps, thereby leading to nonallelic heterogeneity. It should be noted that mutations in dystrophin are also associated with allelic heterogeneity. For example, mutations in the DMD gene can cause either Duchenne's or the less severe Becker's muscular dystrophy, depending on the severity of the protein defect. Recognition of nonallelic heterogeneity is important for several reasons: (1) the ability to identify disease loci in linkage studies is reduced by including patients with similar phenotypes but different genetic disorders; (2) genetic testing is more complex because several different genes need to be considered along with the possibility of different mutations in each of the candidate genes; and (3) novel information is gained about how genes or proteins interact, providing unique insights into molecular physiology. Phenocopies refer to circumstances in which nongenetic conditions mimic a genetic disorder. For example, features of toxin- or drug-induced neurologic syndromes can resemble those seen in Huntington's

AD

disease, and vascular causes of dementia share phenotypic features with familial forms of Alzheimer's dementia (Chap. 442). As in nonallelic heterogeneity, the presence of phenocopies has

the potential to conduct linkage studies and genetic testing. Patient history and subtle differences in phenotype can often provide clues that distinguish these disorders from related genetic conditions.

VARIABLE EXPRESSIVITY AND INCOMPLETE PENETRANCE The same genetic mutation may be associated with a phenotypic spectrum in different affected individuals, thereby illustrating the phenomenon of variable expressivity. This may include different manifestations of a disorder variably involving different organs (e.g., multiple endocrine neoplasia [MEN]), the severity of the disorder (e.g., cystic fibrosis), or the age of disease onset (e.g., Alzheimer's dementia). MEN 1 illustrates several of these features. In this autosomal dominant tumor syndrome, affected individuals carry an inactivating germline mutation that is inherited in an autosomal dominant fashion. After somatic inactivation of the alternate allele (loss of heterozygosity; Knudson two-hit model), patients can develop tumors of the parathyroid gland, endocrine pancreas, the pituitary gland, and dermatologic lesions (Chap. 400). However, the pattern of tumors in the different glands, the age at which tumors develop, and the types of hormones produced vary among

affected individuals, even within a given family. In this example, the phenotypic variability arises, in part, because of the requirement for a second somatic mutation in the normal copy of the MEN1 gene, as well as the large array of different cell types that are susceptible to the effects of MEN1 gene mutations. In part, variable expression reflects the influence of modifier genes, or genetic background, on the effects of a particular mutation. Even in identical twins, in whom the genetic constitution is essentially the same, one can occasionally see variable expression of a genetic disease. Interactions with the environment can also influence the course of a disease. For example, the manifestations and severity of hemochromatosis can be influenced by iron intake (Chap. 426), and the course of phenylketonuria is affected by exposure to phenylalanine in the diet (Chap. 431). Other metabolic disorders, such as hyperlipidemias and porphyria, also fall into this category. Many mechanisms, including genetic effects and environmental influences, can therefore lead to variable expressivity. In genetic counseling, it is particularly important to recognize this variability, because one cannot always predict the course of disease, even when the mutation is known.

Penetrance refers to the proportion of individuals with a mutant genotype that express the phenotype. If all carriers of a mutant express the phenotype, penetrance is complete, whereas it is said to be incomplete or reduced if some individuals do not exhibit features of the phenotype. Dominant conditions with incomplete penetrance are characterized by skipping of generations with unaffected carriers transmitting the mutant gene. For example, hypertrophic obstructive cardiomyopathy (HCM) caused by mutations in the myosin-binding protein C gene is a dominant disorder with clinical features in only a subset of patients who carry the mutation (Chap. 267). Patients who have the mutation, but no evidence of the disease, can still transmit the disorder to subsequent generations. In many conditions with postnatal onset, the proportion of gene carriers who are affected varies with age. Thus, when describing penetrance, one must specify age. For example, for disorders such as Huntington's disease or familial amyotrophic lateral sclerosis, which present later in life, the rate of penetrance is influenced by the age at which the clinical assessment is performed.

Imprinting can also modify the penetrance of a disease. For example, in patients with AHO, mutations in the Gs α subunit (GNAS1 gene) are expressed clinically only in individuals who inherit the mutation from their mother (Chap. 422).

SEX-INFLUENCED PHENOTYPES Certain mutations affect males and females quite differently. In some instances, this is because the gene resides on the X or Y sex chromosomes (X-linked disorders and Y-linked disorders). As a result, the phenotype of mutated X-linked genes will be expressed fully in males but variably in heterozygous females, depending on the degree of X-inactivation and the function of the gene. For

example, most heterozygous female carriers of factor VIII deficiency (hemophilia A) are asymptomatic because sufficient factor VIII is produced to prevent a defect in coagulation (Chap. 121). On the other hand, some females heterozygous for the X-linked lipid storage defect caused by α -galactosidase A deficiency (Fabry's disease) experience mild manifestations of painful neuropathy, as well as other features of the disease (Chap. 429). Because only males have a Y chromosome, mutations in genes such as SRY, which causes male-to-female sex reversal, or DAZ (deleted in azoospermia), which causes abnormalities of spermatogenesis, are unique to males (Chap. 402). Other diseases are expressed in a sex-limited manner because of the differential function of the gene product in males and females. Activating mutations in the luteinizing hormone receptor cause dominant male-limited precocious puberty in boys (Chap. 403). The phenotype is unique to males because activation of the receptor induces testosterone production in the testis, whereas it is functionally silent in the immature ovary. Biallelic inactivating mutations of the follicle-stimulating hormone (FSH) receptor cause primary ovarian failure in females because the follicles do not develop in the absence of FSH action. In contrast, affected males have a more subtle phenotype, because testosterone production is preserved (allowing sexual maturation) and spermatogenesis is only partially impaired

(Chap. 403). In congenital adrenal hyperplasia, most commonly caused by 21-hydroxylase deficiency, cortisol production is impaired and ACTH stimulation of the adrenal gland leads to increased production of androgenic precursors (Chap. 398). In females, the increased androgen level causes ambiguous genitalia, which can be recognized at the time of birth. In males, the diagnosis may be made on the basis of adrenal insufficiency at birth, because the increased adrenal androgen level does not alter sexual differentiation, or later in childhood, because of the development of precocious puberty. Hemochromatosis is more common in males than in females, presumably because of differences in dietary iron intake and losses associated with menstruation and pregnancy in females (Chap. 426).

CHAPTER 479 Principles of Human Genetics Chromosomal Disorders Chromosomal disorders and the techniques used for their characterization have been discussed in detail in previous editions of this textbook. Chromosomal or cytogenetic disorders are caused by numerical (aneuploidy) or structural aberrations (deletions, duplications, translocations, inversions, dicentric and ring chromosomes, Robertsonian translocations) in chromosomes. They occur in ~1% of the general population, in 8% of stillbirths, and in close to 50% of spontaneously aborted fetuses. Indications for cytogenetic and cytogenomic chromosome analyses are summarized in Table 479-4. Contiguous gene syndromes (e.g., large deletions affecting several genes) have been useful for identifying the location of new disease-causing genes. Because of the variable size of gene deletions in different patients, a systematic comparison of phenotypes and locations of deletion breakpoints allows positions of particular genes to be mapped within the critical genomic region. Monogenic Mendelian Disorders Monogenic human diseases are frequently referred to as Mendelian disorders because they obey the principles of genetic transmission originally set forth in Gregor Mendel's classic work. The continuously updated OMIM catalogue lists several thousand of these disorders and provides information about the clinical phenotype, molecular basis, allelic variants, and pertinent animal models (Table 479-1). The mode of inheritance for a given phenotypic trait or disease is determined by pedigree analysis. All affected and unaffected individuals in the family are recorded in a pedigree using standard symbols (Fig. 479-11). The principles of allelic segregation, and the transmission of alleles from parents to children, are illustrated in Fig.

479-12. One dominant (A) allele and one recessive (a) allele can display three Mendelian modes of inheritance: autosomal dominant, autosomal recessive, and X-linked. About 65% of human monogenic disorders are autosomal dominant, 25% are autosomal recessive, and 5% are X-linked. Genetic testing is now readily available for the characterization of monogenic disorders and plays an important role in clinical medicine (Chap. 480). TABLE 479-4 Indications for Cytogenetic and Cytogenomic Analysis across the Life Span

| Timing | Indications for Testing |
|--|---|
| Prenatal | Advanced maternal age, Abnormalities on ultrasound, Increased risk for genetic disorder |
| Maternal serum screen | Increased risk for genetic disorder |
| Neonatal and childhood | Multiple congenital anomalies, Intellectual disability, Autism spectrum disorders, Developmental delay, Failure to thrive, Short stature, Disorders of sexual development |
| History of familial chromosomal alteration | Cancer, Adult Infertility, Recurrent miscarriage, Familial cancer |

Female Male Unknown sex

PART 16 Genes, the Environment, and Disease

| Parental Genotype | Offspring Genotype | Sex | Phenotype |
|-------------------|----------------------|---------------------------|---------------------|
| Multiple siblings | Spontaneous abortion | Deceased | Affected male |
| Proband | Affected female | Heterozygous male | Heterozygous female |
| | Female carrier | Carrier of X-linked trait | |

Mating Consanguineous union II

Monozygotic twins Dizygotic twins FIGURE 479-11 Standard pedigree symbols. AUTOSOMAL DOMINANT DISORDERS In autosomal dominant disorders, mutations in a single allele are sufficient to cause the disease. In contrast to recessive disorders, in which disease pathogenesis is relatively straightforward because there is a biallelic loss of gene function, dominant disorders can be caused by various disease mechanisms, many of which are unique to the function of the genetic pathway involved. Mechanistically, the mutation may confer constitutive activation (gain of function), exert a dominant negative effect, or result in loss of function and haploinsufficiency. In autosomal dominant disorders, individuals are affected in successive generations; the disease does not occur in the offspring of unaffected individuals. Males and females are affected with equal frequency because the defective gene resides on one of the 22 autosomes (Fig. 479-13A). Autosomal dominant mutations alter one of the two alleles at a given locus. Because the alleles segregate randomly at meiosis, the probability that an offspring will be affected is 50%. Unless there is a new germline mutation, an affected individual has an affected parent. Children with a normal genotype do not transmit the disorder. Due to differences in penetrance or expressivity (see above), the clinical manifestations of autosomal dominant disorders may be variable. Because of these variations, it is sometimes challenging to determine the pattern of inheritance. It should be recognized, however, that some individuals acquire a mutated gene from an unaffected parent due to de novo germline mutations. They occur more frequently during later cell divisions in gametogenesis, which explains why siblings are rarely affected. As noted, new germline mutations occur more frequently in fathers of advanced age. For example, the average age of fathers with new germline mutations that

| Parental Genotype | Offspring Genotype | Phenotype |
|-------------------|--------------------|------------|
| Aa | aa | Unaffected |
| AA | Aa | Affected |
| Aa | Aa | Affected |
| Aa | aa | Unaffected |
| Aa | Aa | Affected |
| Aa | AA | Affected |
| Aa | Aa | Affected |
| aa | aa | Unaffected |

50:50

25:50:25 FIGURE 479-12 Segregation of alleles. Segregation of genotypes in the offspring of parents with one dominant (A) and one recessive (a) allele. The distribution of the parental alleles to their offspring depends on the combination present in the parents. Filled symbols = affected individuals.

Autosomal dominant A B Autosomal recessive Autosomal recessive with pseudodominance X-linked C Mitochondrial D

FIGURE 479-13 A. Dominant, B. recessive, C. X-linked, and D. mitochondrial (matrilinear) inheritance. cause Marfan's syndrome is ~37 years, whereas fathers who transmit the disease by inheritance have an average age of ~30 years.

AUTOSOMAL RECESSIVE DISORDERS In recessive disorders, the mutated alleles result in a complete or partial loss of function. They frequently involve enzymes in metabolic pathways, receptors, or proteins in signaling cascades. In an autosomal recessive disease, the affected individual, who can be of either sex, is a homozygote or compound heterozygote for a single-gene defect. With a few important exceptions, autosomal recessive diseases are rare and occur more often in the context of parental consanguinity. The relatively high frequency of certain recessive disorders such as sickle cell anemia, cystic fibrosis, and thalassemia, is partially explained by a selective biologic advantage for the heterozygous state (see below). Although heterozygous carriers of a defective allele are usually clinically normal, they may display subtle differences in phenotype that only become apparent with more precise testing or in the context of certain environmental influences. In sickle cell anemia, for example, heterozygotes are normally asymptomatic. However, in situations of dehydration or diminished oxygen pressure, sickle cell crises can also occur in heterozygotes (Chap. 103).

aa In most instances, an affected individual is the offspring of heterozygous parents. In this situation, there is a 25% chance that the offspring will have a normal genotype, a 50% probability of a heterozygous state, and a 25% risk of homozygosity for the recessive alleles (Figs. 479-12 and 479-13B). In the case of one unaffected heterozygous and one affected homozygous parent, the probability of disease increases to 50% for

each child. In this instance, the pedigree analysis mimics an autosomal dominant mode of inheritance (pseudodominance). In contrast to autosomal dominant disorders, new mutations in recessive alleles are rarely manifest because they usually result in an asymptomatic carrier state.

X-LINKED DISORDERS Males have only one X chromosome; consequently, a daughter always inherits her father's X chromosome in addition to one of her mother's two X chromosomes. A son inherits the Y chromosome from his father and one maternal X chromosome. Thus, the characteristic features of X-linked inheritance are (1) the absence of father-to-son transmission and (2) the fact that all daughters of an affected male are obligate carriers of the mutant allele (Fig. 479-13C). The risk of developing disease due to a mutant X-chromosomal gene differs in the two sexes. Because males have only one X chromosome, they are hemizygous for the mutant allele; thus, they are more likely to develop the mutant phenotype, regardless of whether the mutation is dominant or recessive. A female may be either heterozygous or homozygous for the mutant allele, which may be dominant or recessive. The terms X-linked dominant and X-linked recessive are therefore only applicable to expression of the mutant phenotype in women. In addition, the expression of X-chromosomal genes is influenced by X chromosome inactivation.

Y-LINKED DISORDERS The Y chromosome has a relatively small number of genes. One such gene, the sex-region determining Y factor (SRY), which encodes the testis-determining factor (TDF), is crucial for normal male development. Normally, there is infrequent exchange of sequences on the Y chromosome with the X chromosome. The SRY region is adjacent to the pseudoautosomal region, a chromosomal segment on the X and Y chromosomes with a high degree of homology. A crossing-over event occasionally involves the SRY region with the distal tip of the X chromosome during meiosis in the male. Translocations can result in XY females with the Y chromosome lacking the SRY gene or XX males harboring the SRY gene on one of the X chromosomes (Chap. 402). Point mutations in the SRY gene may also result in individuals with an XY genotype and an incomplete

female phenotype. Most of these mutations occur de novo. Men with oligo spermia/azoospermia frequently have microdeletions on the long arm of the Y chromosome that involve one or more of the azoospermia factor (AZF) genes. Exceptions to Simple Mendelian Inheritance Patterns •

MITOCHONDRIAL DISORDERS Mendelian inheritance refers to the transmission of genes encoded by DNA contained in the nuclear chromosomes. In addition, each mitochondrion contains several copies of a small circular chromosome (Chap. 481). The mitochondrial DNA (mtDNA) is ~16.5 kb and encodes transfer and ribosomal RNAs and 13 core proteins that are components of the respiratory chain involved in oxidative phosphorylation and ATP generation. The mitochondrial genome does not recombine and is inherited through the maternal line because sperm does not contribute significant cytoplasmic components to the zygote. A noncoding region of the mitochondrial chromosome, referred to as D-loop, is highly polymorphic. This property, together with the absence of mtDNA recombination, makes it a valuable tool for studies tracing human migration and evolution, and it is also used for specific forensic applications. Inherited mitochondrial disorders are transmitted in a matrilineal fashion; all children from an affected mother will inherit the disease, but it will not be transmitted from an affected father to his children (Fig. 479-13D). Alterations in the mtDNA that involves enzymes required for oxidative phosphorylation lead to reduction of ATP supply, generation of free radicals, and induction of apoptosis. Several syndromic disorders arising from mutations in the mitochondrial genome are known in humans, and they affect both protein-coding and tRNA genes. The broad clinical spectrum often involves (cardio) myopathies and encephalopathies because of the high dependence of these tissues on oxidative phosphorylation. The age of onset and the clinical course are highly variable because of the unusual mechanisms of mtDNA transmission, which replicates independently from nuclear DNA. During cell replication, the proportion of wild-type and mutant

mitochondria can drift among different cells and tissues. The resulting heterogeneity in the proportion of mitochondria with and without a mutation is referred to as heteroplasmy and underlies the phenotypic variability that is characteristic of mitochondrial diseases.

Acquired somatic mutations in mitochondria are thought to be involved in several age-dependent degenerative disorders affecting predominantly muscle and the peripheral and central nervous system (e.g., Alzheimer's and Parkinson's diseases). Establishing that an mtDNA alteration is causal for a clinical phenotype is challenging because of the high degree of polymorphism in mtDNA and the phenotypic variability characteristic of these disorders. Certain pharmacologic treatments may have an impact on mitochondria and/or their function. For example, treatment with the antiretroviral compound zidovudine (AZT) causes an acquired mitochondrial myopathy through depletion of muscular mtDNA. **CHAPTER 479 Principles of Human Genetics MOSAICISM** Mosaicism refers to the presence of two or more genetically distinct cell lines in the tissues of an individual. It results from a mutation that occurs during embryonic, fetal, or extrauterine development. The developmental stage at which the mutation arises will determine whether germ cells and/or somatic cells are involved. Chromosomal mosaicism results from nondisjunction at an early embryonic mitotic division, leading to the persistence of more than one cell line, as exemplified by some patients with Turner's syndrome (Chap. 402).

Somatic mosaicism is characterized by a patchy distribution of genetically altered somatic cells. The McCune-Albright syndrome, for example, is caused by activating mutations in the stimulatory

G protein α ($G\alpha$) that occur postzygotically in early development (Chap. 422). The clinical phenotype varies depending on the tissue distribution of the mutation; manifestations include ovarian cysts that secrete sex steroids and cause precocious puberty, polyostotic fibrous dysplasia, café-au-lait skin pigmentation, GH-secreting pituitary adenomas, and hypersecreting autonomous thyroid nodules.

X-INACTIVATION, IMPRINTING, AND UNIPARENTAL DISOMY According to traditional Mendelian principles, the parental origin of a mutant gene is irrelevant for the expression of the phenotype. There are, however, important exceptions to this rule. X-inactivation prevents the expression of most genes on one of the two X chromosomes in every cell of a female. Gene inactivation through genomic imprinting occurs on selected chromosomal regions of autosomes and leads to inheritable preferential expression of one of the parental alleles. It is of pathophysiologic importance in disorders where the transmission of disease is dependent on the sex of the transmitting parent and, thus, plays an important role in the expression of certain genetic disorders. Two classic examples are the Prader-Willi syndrome and Angelman's syndrome. Prader-Willi syndrome is characterized by diminished fetal activity, obesity, hypotonia, intellectual disability, short stature, and hypogonadotropic hypogonadism. Deletions of the paternal copy of the Prader-Willi locus located on the short arm of chromosome 15 result in a contiguous gene syndrome involving missing paternal copies of the *necdin* and *SNRPN* genes, among others. In contrast, patients with Angelman's syndrome, characterized by intellectual disability, seizures, ataxia, and hypotonia, have deletions involving the maternal copy of this region on chromosome 15. These two syndromes may also result from uniparental disomy. In this case, the syndromes are not caused by deletions on chromosome 15 but by the inheritance of either two maternal chromosomes (Prader-Willi syndrome) or two paternal chromosomes (Angelman's syndrome). Lastly, the two distinct phenotypes can also be caused by an imprinting defect that impairs the resetting of the imprint during zygote development (defect in the father leads to Prader-Willi syndrome; defect in the mother leads to Angelman's syndrome). Imprinting and the related phenomenon of allelic exclusion may be more common than currently documented because it is difficult to examine levels of mRNA expression from the maternal and paternal alleles in specific tissues or in individual cells. Genomic imprinting, or uniparental disomy, is involved in the pathogenesis of several other disorders and malignancies. For example, hydatidiform moles contain a normal number of diploid chromosomes, but they are exclusively of

paternal origin. The opposite situation occurs in ovarian teratomata, with 46 chromosomes of maternal origin. Expression of the imprinted gene for insulin-like growth factor 2 (*IGF-2*) is involved in the pathogenesis of the cancer-predisposing Beckwith-Wiedemann syndrome (BWS). These children show somatic overgrowth with organomegalies and hemihypertrophy, and they have an increased risk of embryonal malignancies such as Wilms' tumor. Normally, only the paternally derived copy of the *IGF2* gene is active, and the maternal copy is inactive. BWS can be caused by several genetic defects that result in overactivity of *IGF-2*, or a missing active copy of *CDKN1C*, that result in inhibition of cell proliferation. They include paternal uniparental disomy (UPD) of chromosome 11, aberrant methylation of this region, maternal chromosomal rearrangements, or deletions within the locus.

PART 16 Genes, the Environment, and Disease Alterations of the epigenome through gain and loss of DNA methylation and altered histone modifications play an important role in the pathogenesis of malignancies.

SOMATIC MUTATIONS Cancer can be considered a genetic disease at the cellular level (Chap. 76). Cancers are monoclonal in origin, indicating that they have arisen from a single

precursor cell with one or several mutations in genes controlling growth (proliferation or apoptosis) and/or differentiation. These acquired somatic mutations are restricted to the tumor and its metastases and are not found in the surrounding normal tissue. The molecular alterations include dominant gain-of-function mutations in oncogenes, recessive loss-of-function mutations in tumor-suppressor genes and DNA repair genes, gene amplification, and chromosome rearrangements. Chromothripsis refers to a mutational process including multiple clustered chromosomal rearrangements in close vicinity, for example, after injury by ionizing radiation. Rarely, a single mutation in certain genes may be sufficient to transform a normal cell into a malignant cell. In most cancers, however, the development of a malignant phenotype requires several genetic alterations for the gradual progression from a normal cell to a cancerous cell, a phenomenon termed multistep carcinogenesis. Genome-wide analyses of cancers using deep sequencing often reveal somatic rearrangements resulting in fusion genes and mutations in multiple genes (Table 479-1 and Fig. 479-14). Comprehensive sequence analyses, now also possible through single-cell sequencing (SCS), provide insight into the evolution and genetic heterogeneity within malignancies; these include intratumoral heterogeneity among the cells of the primary tumor, intermetastatic and intrametastatic heterogeneity, and interpatient

Mutations per Mb

A

Histology HPV clade HPV integration APOBEC mutagenesis UCEC-like EMT score Purity iCluster
 PIK3CA (26%) EP300 (11%) FBXW7 (11%) PTEN (8%) HLA-A (8%) ARID1A (7%) NFE2L2 (7%) HLA-B
 (6%) KRAS (6%) ERBB3 (6%) MAPK1 (5%) B CASP8 (4%) TGFBR2 (3%) SHKBP1 (2%) C Synonymous
 In-frame indel Other non-synonymous Missense Splice site Frameshift Nonsense 3q (66%) CD274
 (8%) PTEN (8%) YAP1 (16%) BCAR4 (16%) D $0.1 \leq \log_2[\text{CN}] < 0.4$ $\log_2[\text{CN}] \leq -0.4$ $-0.4 < \log_2[\text{CN}] \leq -0.1$ $\log_2[\text{CN}] \geq 0.4$ Gene-level SCNAs

FIGURE 479-14 Somatic alterations in cervical cancer. A. Cervical carcinoma samples ordered by histology and mutation frequency; B. clinical and molecular platform features; C. significantly mutated genes (SMGs); and D. select somatic copy number alterations. SMGs are ordered by the overall mutation frequency and color-coded by mutation type. Adeno, adenocarcinomas; Adenosq, adenosquamous cancers; CN, copy number; SCNAs, somatic copy number alterations; Squamous, squamous cell carcinomas. (Reproduced from The Cancer Genome Atlas Research Network. Integrated genomic and molecular characterization of cervical cancer. *Nature* 543:378–384, 2017.)

differences. These analyses further support the notion of cancer as an ongoing process of clonal evolution, in which successive rounds of clonal selection within the primary tumor and metastatic lesions result in diverse genetic and epigenetic alterations that require targeted (personalized) therapies (precision medicine). The heterogeneity of mutations within a tumor can also lead to resistance to targeted therapies because cells with mutations that are resistant to the therapy, even if they are a minor part of the tumor population, will be selected as the more sensitive cells are eliminated. Telomeres, repeats of conserved sequences, protect the ends the chromosomes from DNA damage or fusion with neighboring chromosomes. Telomere length shortens with age. Most human tumors express telomerase, an enzyme formed of a protein and an RNA component, which adds telomere repeats at the ends of chromosomes during replication. This mechanism

impedes shortening of the telomeres and is associated with enhanced replicative capacity in cancer cells. Telomerase inhibitors provide a strategy for treating advanced human cancers. In many cancer syndromes, there is frequently an inherited predisposition to tumor formation. In these instances, a germline mutation is inherited in an autosomal dominant fashion inactivating one allele of an autosomal tumor-suppressor gene. If the second allele is inactivated by a somatic mutation or by epigenetic silencing in a given cell, this will lead to neoplastic growth (Knudson two-hit model). Thus, the defective allele in the germline is transmitted in a dominant mode, although tumorigenesis results from a biallelic loss of the tumor-suppressor gene in an affected tissue. The classic example to illustrate this phenomenon is retinoblastoma, which can occur as a sporadic or hereditary tumor. In sporadic retinoblastoma, both copies of the retinoblastoma (RB) gene are inactivated through two somatic events. In hereditary retinoblastoma, one mutated or deleted RB allele is inherited in an autosomal dominant manner and the second allele is inactivated by a subsequent somatic mutation. This two-hit model applies to other inherited cancer syndromes such as MEN 1 (Chap. 400) and neurofibromatosis types 1 and 2 (Chap. 95). In contrast, in the autosomal dominant MEN 2 syndrome, the predisposition for tumor formation in various organs is caused by a gain-of-function mutation in a single allele of the RET gene (Chap. 400).

NUCLEOTIDE REPEAT EXPANSION DISORDERS Several diseases are associated with an increase in the number of nucleotide repeats above a certain threshold (Table 479-5). The repeats are sometimes located

| Synonymous | Non-synonymous | Other | Squamous | Adenosq. | Adeno. | Negative | A9 | A7 | Yes | Low | High | No |
|------------|----------------|-------|----------|----------|--------|----------|--------|--------------|-------------|--------|------------|----|
| No | Yes | No | 0.96 | -3.76 | 0.22 | 1.15 | Adeno. | Keratin-high | Keratin-low | APOBEC | Non-APOBEC | |

Mutations Gain Loss

TABLE 479-5 Selected Trinucleotide Repeat Disorders

| DISEASE | LOCUS | REPEAT | X-chromosomal |
|-------------------------------------|---------------------------------------|-------------------|------------------|
| spinobulbar muscular atrophy (SBMA) | Xq12 | CAG 11–34/40–62 | XR |
| Androgen receptor | Fragile X | | |
| syndrome (FRAXA) | Xq27.3 | CGG 6–50/200–300 | XR |
| FMR-1 protein | Fragile X | | |
| syndrome (FRAXE) | Xq28 | GCC 6–25/>200 | XR |
| FMR-2 protein | Dystrophia | | |
| myotonia (DM) | 19q13.32 | CTG 5–30/200–1000 | AD, |
| variable penetrance | Myotonin protein kinase | | |
| Huntington’s disease (HD) | 4p16.3 | CAG 6–34/37–180 | AD |
| Huntingtin | Spinocerebellar ataxia type 1 (SCA1) | 6p22.3 | CAG 6–39/40–88 |
| AD | Spinocerebellar ataxia type 2 (SCA2) | 12q24.12 | CAG 15–31/34–400 |
| AD | Spinocerebellar ataxia type 3 (SCA3); | | |

Machado-Joseph disease (MD) 14q32.12 CAG 13–36/55–86 AD Ataxin 3 Spinocerebellar ataxia type 6 (SCA6, CACNA1A) 19p13 CAG 4–16/20–33 AD Alpha 1A voltage-dependent L-type calcium channel Spinocerebellar ataxia type 7 (SCA7) 3p14.1 CAG 4–19/37 to >300 AD Ataxin 7 Spinocerebellar ataxia type 12 (SCA12) 5q32 CAG 6–26/66–78 AD Protein phosphatase 2A Dentatorubral pallidoluysian atrophy (DRPLA) 12p13.31 CAG 7–23/49–75 AD Atrophin 1 Friedreich’s ataxia (FRDA1) 9q21.11 GAA 7–22/200–900 AR Frataxin Abbreviations: AD, autosomal dominant; AR, autosomal recessive; XR, X-linked recessive. within the coding region of the genes, as in Huntington’s disease or the X-linked form of spinal and bulbar muscular atrophy (SBMA; Kennedy’s syndrome). In other instances, the repeats probably alter gene regulatory sequences. If an expansion is present, the DNA fragment is unstable and tends to expand further during cell division. The length of the nucleotide repeat often correlates with the severity of the disease. When repeat length increases from one generation to the next, disease manifestations may worsen or be observed at an earlier age; this phenomenon is referred to as anticipation. In Huntington’s disease, for example, there is a correlation between age of onset and length of the triplet codon expansion

(Chap. 435). Anticipation has also been documented in other diseases caused by dynamic mutations in tri nucleotide repeats (Table 479-5). The repeat number may also vary in a tissue-specific manner. In myotonic dystrophy, the CTG repeat may be 10-fold greater in muscle tissue than in lymphocytes (Chap. 460). Complex Genetic Disorders The expression of many common diseases such as cardiovascular disease, hypertension, diabetes, asthma, psychiatric disorders, and certain cancers is determined by a combination of genetic background, environmental factors, and lifestyle. A trait is called polygenic if multiple genes contribute to the phenotype or multifactorial if multiple genes are assumed to interact with environmental factors. Genetic models for these complex traits need to account for genetic heterogeneity and interactions with other genes and the environment. Complex genetic traits may be influenced by modifier genes that are not linked to the main gene involved in the pathogenesis of the trait. This type of gene-gene interaction, or epistasis, where the expression of a gene is altered by the expression of one or several independently inherited genes, plays an important role in polygenic traits. In aggregate, variants in multiple genes need to be present simultaneously to result in a pathologic phenotype. Type 2 diabetes mellitus provides a paradigmatic example of a multifactorial disorder, because genetic, nutritional, and lifestyle factors are intimately interrelated in disease pathogenesis (Table 479-6)

(Chap. 415). The identification of genetic variations and environmental factors that either predispose to or protect against disease is essential for predicting disease risk, designing preventive strategies, and developing novel therapeutic approaches. The study of rare monogenic diseases may provide insight into some of the genetic and molecular mechanisms important in the pathogenesis of complex diseases. For example, the identification of the genes causing monogenic forms of permanent neonatal diabetes mellitus or maturity-onset diabetes defined them as candidate genes in the pathogenesis of diabetes mellitus type 2 (Tables 479-2 and 479-6) (Fig. 479-15). Genome scans have identified numerous genes and loci that may be associated with

TRIPLET LENGTH (NORMAL/DISEASE) INHERITANCE GENE PRODUCT CHAPTER 479 Principles of Human Genetics susceptibility to development of diabetes mellitus in certain populations (Fig. 479-16). Efforts to identify susceptibility genes require very large sample sizes, and positive results may depend on ethnicity, ascertainment criteria, and statistical analysis. Association studies analyzing the potential influence of (biologically functional) SNPs and SNP haplotypes on a particular phenotype have revealed new insights into the genes involved in the pathogenesis of these common disorders. Large variants ([micro]deletions, duplications, and inversions) present in the human population also contribute to the pathogenesis of complex disorders, but their contributions remain poorly understood. Linkage and Association Studies There are two primary strategies for mapping genes that cause or increase susceptibility to human disease: (1) classic linkage can be performed based on a known genetic model or, when the model is unknown, by studying pairs of affected relatives; or (2) disease genes can be mapped using allelic association studies (Table 479-7). GENETIC LINKAGE Genetic linkage refers to the fact that genes are physically connected, or linked, to one another along the chromosomes. Two fundamental principles are essential for understanding the concept of linkage: (1) when two genes are close together on a chromosome, they are usually transmitted together, unless a recombination event separates them (Figs. 479-6); and (2) the odds of a crossover, or recombination event, between two linked genes is proportional to the distance that separates them. Thus, genes that are farther apart are more likely to undergo a recombination event than genes that are very close together. The detection of

chromosomal loci that segregate with a disease by linkage can be used to identify the gene responsible for the disease and to predict the odds of disease gene transmission in genetic counseling. Polymorphic variants are essential for linkage studies because they provide a means to distinguish the maternal and paternal chromosomes in an individual. On average, 1 out of every 1000 bp varies from one person to the next. Although this degree of variation seems low (99.9% identical), it means that >3 million sequence differences exist between any two unrelated individuals and the probability that the sequence at such loci will differ on the two homologous chromosomes is high (often >70-90%). These sequence variations include variable number of tandem repeats (VNTRs), short tandem repeats (STRs), and SNPs. Most STRs, also called polymorphic microsatellite markers, consist of di-, tri-, or tetranucleotide repeats that can be characterized readily using the polymerase chain reaction (PCR). Characterization of SNPs, using DNA chips or beads, permits comprehensive analyses of genetic variation, linkage, and association studies. Although these sequence variations often have no apparent functional consequences, they provide much of the basis for variation in genetic traits.

TABLE 479-6 Examples of Genes and Loci Involved in Mono- and Polygenic Forms of Diabetes DISORDER GENES OR SUSCEPTIBILITY LOCUS Monogenic permanent neonatal diabetes mellitus KCNJ11 (inwardly rectifying potassium channel Kir6.2) 11p15.1 AD GCK (glucokinase) 7p13 AR PART 16 Genes, the Environment, and Disease INS (insulin) 11p15.5 AR, hyperproinsulinemia ABCC8 (ATP-binding cassette, subfamily c, member 8; sulfonylurea receptor) 11p15.1 AD or AR GLIS3 (GLIS family zinc finger protein 3) 9p24.2 AR, diabetes, congenital hypothyroidism Maturity-onset diabetes of the young (MODY): monogenic forms of diabetes mellitus MODY 1 HNF4 α (hepatocyte nuclear factor 4 α) 20q13.12 AD inheritance MODY 2 GCK (glucokinase) 7p13 MODY 3 HNF1 α (hepatocyte nuclear factor 1 α) 12q24.31 MODY 4 IPF1 (insulin receptor substrate) 13q12.2 MODY 5 (renal cysts, diabetes) HNF1 β (hepatocyte nuclear factor 1 β) 17q12 MODY 6 NeuroD1 (neurogenic differentiation factor 1) 2q31.3 MODY 7 KLF1 (Kruppel-like factor 1) 19p13.13 MODY 8 CEL (carboxyl ester lipase) 9q34.13 MODY 9 PAX4 (paired box transcription factor 4) 7q32.1 MODY 10 INS (insulin) 11p15.5 MODY 11 BLK (B-lymphocyte-specific tyrosine kinase) 8p23.1 MODY 12 ABCC8 (ATP-binding cassette, subfamily c, member 8; sulfonylurea receptor) 11p15.1 MODY 13 KCNJ11 (inwardly rectifying potassium channel Kir6.2) 11p15.1 Diabetes mellitus type 2; loci and genes linked and/or associated with susceptibility for diabetes mellitus type 2 Genes and loci identified by linkage/association studies Heavily influenced by diet, energy expenditure, obesity PPARG, KCNJ11/ABCC8, TCF7L2, HNF1B, WFS1, SLC30A8, FTO, HHEX, IGF2BP2, CDKN2A/B, CDKAL1, TSPAN8, ADAMT5, CDC123/CAMK1D, JAZF1, NOTCH2, THADA, KCNQ1, DUSP8, MTNR1B, IRS1, SPRY2, SRR, ZFAND6, GCK, KLF14, TP53INP1, PROX1, PRC1, BCL11A, ZBED3, RBMS1, HNF1A, DGKB/ TMEM195, CCND2, C2CD4A/C2CD4B, PTPRD, ARAP1/CENTD2, HMGA2, TLE4/ CHCHD9, ADCY5, UBE2E2, DUSP9, GCKR, COBLL1/GRB14, HMG20A, VPS26A, ST6GAL1, AP3S2, HNF4A, BCL2, LAMA1, GIPR, MC4R, TLE1, KCNK16, ANK1, KLHDC5, ZMIZ1, PSMD6, FITM2/R3HDML/HNF4A, CILP2, ANKRD55, GLIS3, PEPD, GCC1/PAX4, ZFAND3, MAEA, BCAR1, RBM43/RND3, MACF1, RASGRP1, GRK5, TMEM163, SGCG, LPP, FAF1, TMEM154, MPHOSPH9, ARL15, POU5F1/ TCF19, SSR1/RREB1, HLA-B, INS-IGF2, GPSM1, LEP, SLC16A13, PAM/PPP5K2, SLC16A11, CCDC63, C12orf51, CCND2, HNF1A, TBC1D4, CCDC85A, INAFM2, ASB3, FAM60A, ATP8B2, MIR4686, MTMR3, DMRTA1, SLC35D3, GLP2R, GIP, MAP3K11, PLEKHA1, HSD17B12, NRXN3, CMIP, ZZEF1, MNX1, ABO, ACSL1, HLA-DQA1 Abbreviations: AD, autosomal dominant; AR, autosomal recessive; MODY, maturity-onset diabetes of the young. In order to identify a chromosomal locus that segregates with a disease, it is necessary to characterize polymorphic DNA markers from affected and unaffected individuals of

one or several pedigrees. One can then assess whether certain marker alleles cosegregate with the disease. Markers that are closest to the disease gene are less likely to undergo recombination events and therefore receive a higher linkage score. Linkage is expressed as a lod (logarithm of odds) score—the ratio of the probability that the disease and marker loci are linked rather than unlinked. Lod scores of +3 (1000:1) are generally accepted as supporting linkage, whereas a score of -2 is consistent with the absence of linkage. ALLELIC ASSOCIATION, LINKAGE DISEQUILIBRIUM, AND HAPLO TYPES Allelic association refers to a situation in which the frequency of an allele is significantly increased or decreased in individuals affected by a particular disease in comparison to controls. Linkage and association differ in several aspects. Genetic linkage is demonstrable in families or sibships. Association studies, on the other hand, compare a population of affected individuals with a control population. Association studies can be performed as case-control studies that include unrelated affected individuals and matched controls or as family-based

CHROMOSOMAL LOCATION OTHER FACTORS studies that compare the frequencies of alleles transmitted or not transmitted to affected children. Allelic association studies are particularly useful for identifying susceptibility genes in complex diseases. When alleles at two loci occur more frequently in combination than would be predicted (based on known allele frequencies and recombination fractions), they are said to be in linkage disequilibrium. Evidence for linkage disequilibrium can be helpful in mapping disease genes because it suggests that the two loci are tightly linked. Detecting the genetic factors contributing to the pathogenesis of common complex disorders is challenging. In many instances, these are low-penetrance alleles (e.g., variations that individually have a subtle effect on disease development, and they can only be identified by unbiased GWAS) (Catalog of Published Genome-Wide Association Studies; Table 479-1) (Fig. 479-16). Most variants occur in noncoding or regulatory sequences but do not alter protein structure. The analysis of complex disorders is further complicated by ethnic differences in disease prevalence, differences in allele frequencies in known susceptibility genes among different populations, locus and allelic heterogeneity, gene-gene and gene-environment interactions, and the possibility of phenocopies. Catalogues of human variation and genotype data

Rare alleles Mendelian disease High Effect size 3.0 Intermediate 1.5 Modest Rare variants with small effect: difficult to identify 1.1 Low 0.001 0.005 0.05 Common Very rare FIGURE 479-15 Relationship between allele frequency and effect size in monogenic and polygenic disorders. In classic Mendelian disorders, the allele frequency is typically low but has a high impact (single-gene disorder). This contrasts with polygenic disorders that require the combination of multiple low-impact alleles that are frequently quite common in the general population. (HapMap, International Genome Sample Resource) have greatly facilitated GWAS for the characterization of complex disorders. Adjacent SNPs are inherited together as blocks, and these blocks can be identified by genotyping selected marker SNPs, so-called Tag SNPs, thereby reducing cost and workload (Fig. 479-4). The availability of this information permits the characterization of a limited number of SNPs to identify the set of haplotypes present in an individual (e.g., in cases and controls). This, in turn, permits performing GWAS by searching for associations of certain haplotypes with a disease phenotype of interest, an essential step for unraveling the genetic factors contributing to complex disorders. POPULATION GENETICS In population genetics, the focus changes from alterations in an individual's genome to the distribution pattern of different genotypes in the population. In a case where there are only two alleles, A and a, the frequency of the genotypes will be $p^2 + 2pq + q^2 = 1$, with p^2 corresponding to the frequency of AA, $2pq$ to the frequency of Aa, and q^2 to aa. When

the frequency of an allele is known, the frequency of the genotype can be calculated. Alternatively, one can determine an allele frequency if the genotype frequency has been determined. Allele frequencies vary among ethnic groups and geographic regions. For example, heterozygous mutations in the CFTR gene are relatively common in populations of European origin but are rare in the African population. Allele frequencies may vary because certain allelic variants confer a selective advantage. For example, heterozygotes for the sickle cell mutation, which is particularly common in West Africa, are more resistant to malaria infection because the erythrocytes of heterozygotes provide a less favorable environment for Plasmodium parasites. Although homozygosity for the sickle cell mutation is associated with severe anemia and sickle 'crises, heterozygotes have a higher probability of survival because of the reduced morbidity and mortality from malaria; this phenomenon has led to an increased frequency of the mutant allele. Recessive conditions are more prevalent in geographically isolated populations because of the more restricted gene pool.

APPROACH TO THE PATIENT Inherited Disorders For the practicing clinician, the family history remains an essential step in recognizing the possibility of a hereditary predisposition to disease. When taking the history, it is useful to draw a detailed

Rare: Common variants with high effect on complex disease CHAPTER 479 Low-frequency variants with intermediate effect Principles of Human Genetics Typical: Common variants with low effect on complex disease Rare Low frequency Allele frequency pedigree of the first-degree relatives (e.g., parents, siblings, and children), because they share 50% of genes with the patient. Standard symbols for pedigrees are depicted in Fig. 479-11. The family history should include information about ethnic background, age, health status, and deaths, including infants. Next, the physician should explore whether there is a family history of the same or related illnesses to the current problem. An inquiry focused on commonly occurring disorders such as cancers, heart disease, and diabetes mellitus should follow. Because of the possibility of age-dependent expressivity and penetrance, the family history will need intermittent updating. If the findings suggest a genetic disorder, the clinician should assess whether some of the patient's relatives may be at risk of carrying or transmitting the disease. In this circumstance, it is useful to confirm and extend the pedigree based on input from several family members. Emerging artificial intelligence tools analyzing facial features can aid the clinician in diagnosing patients with genetic conditions. In aggregate, this information may form the basis for genetic counseling, carrier detection, early intervention, and disease prevention in relatives of the index patient (Chap. 480). In instances where a diagnosis at the molecular level may be relevant, it is important to identify an appropriate laboratory that can perform the appropriate test. Genetic testing is available for a large number of monogenic disorders through commercial laboratories. For uncommon disorders, the test may only be performed in a specialized research laboratory. Approved laboratories offering testing for inherited disorders can be identified in continuously updated online resources (e.g., Genetic Testing Registry; Table 479-1). If genetic testing is considered, the patient and the family should be counseled about the potential implications of positive results, including psychological distress and the possibility of discrimination. The patient or caretakers should be informed about the meaning of a negative result, technical limitations, and the possibility of false-negative and inconclusive results. For these reasons, genetic testing should only be performed after obtaining informed consent. Published ethical guidelines address the specific aspects that should be considered when testing children and adolescents.

IDENTIFYING THE DISEASE-CAUSING GENE Precision medicine aims to enhance the quality of medical care through the use of genotypic analysis (DNA testing) to identify genetic predisposition to disease, to select more specific

Significant Loci:

PART 16 Genes, the Environment, and Disease African and African-American East Asian European Hispanic/Native American South Asian

Initial sample size Replication sample size

Linkage or candidate gene GWAS or Metabochip Exome array Genome or exome sequencing

Total sample size (1000s)

PPARG KCNJ11 TCF7L2 SLC30A8 MC4R SLC16A11

PubMed ID

2003 2006

Year

FIGURE 479-16 Genome-wide association studies (GWAS) across ancestries and discovery of loci over time. The pie charts represent type 2 diabetes GWAS, as well as candidate gene or sequencing studies. The x axis shows the year of publication, and the y axis shows discovery sample size. The inner circles are scaled in proportion to discovery sample size, and the outer circles are scaled in proportion to total (discovery + replication) sample size. Significant loci are defined as a p value of 5×10^{-8} . At the end of 2022, 534 type 2 diabetes distinct intervals (520 autosomal, 14 X chromosomal) were defined. (Reproduced with permission from BF Voight.)

pharmacotherapy, and to design individualized medical care based on genotype. Genotype can be deduced by analysis of protein (e.g., hemoglobin, apoprotein E), mRNA, or DNA. Many (pathogenic) variants can be readily identified by DNA analyses; technical advances in RNA sequencing now add increasing depth to genetic and genomic investigations (e.g., for the detection of gene fusions or aberrant gene expression patterns). DNA testing is performed by mutational analysis or linkage studies in individuals at risk for a genetic disorder known to be present in a family. Mass screening programs require tests of high sensitivity and specificity to be cost-effective. The benefits and risks of screening newborns with genomic sequencing, and the potential impact on surveillance, preventative health care, and personalized treatment options are topics of current research (BabySeq Project). Prerequisites for the success of genetic screening programs include the following: that the disorder is potentially serious; that it can be influenced at a presymptomatic stage by changes in behavior, diet, and/or pharmaceutical manipulations; and that the screening does not result in any harm or discrimination. Screening in Jewish populations for the autosomal recessive neurodegenerative storage disease Tay-Sachs has reduced the number of affected individuals. In contrast, screening for sickle cell trait/disease in African Americans has led to unanticipated problems of discrimination by health insurers and employers. Mass screening programs harbor additional potential problems. For example, screening for the most common genetic alteration in cystic fibrosis, the $\Delta F508$ mutation

with a frequency of ~70% in northern Europe, is feasible and seems to be effective. One has to keep in mind, however, that there is pronounced allelic heterogeneity and that the CFTR gene can be affected by >2000 other mutations. While the search for less common mutations has been challenging in the past, next-generation genome sequencing now permits comprehensive and cost-effective mutational analyses. However, the bioinformatic analysis and the classification of the detected variants as pathogenic or benign alterations is still challenging. Occupational screening programs aim to detect individuals with increased risk for certain professional activities (e.g., α 1 antitrypsin deficiency and smoke or dust exposure). Integrating genomic data into electronic medical records is evolving and can provide significant decision support at the point of care, for example, by providing the clinician with genomic data and decision algorithms for the prescription of drugs that are subject to pharmacogenetic influences. Mutational Analyses DNA sequence analysis is widely used as a diagnostic tool and has significantly enhanced diagnostic accuracy. It is used for determining carrier status and for prenatal testing in monogenic disorders. Numerous techniques, discussed in previous versions of this chapter, are available for the detection of mutations. Analyses of large alterations in the genome are possible using classic methods such as karyotype analysis, cytogenetics, fluorescent in situ hybridization (FISH), and array- or bead-based techniques that search for multiple single exon deletions or duplications.

TABLE 479-7 Genetic Approaches for Identifying Disease Genes INDICATIONS AND ADVANTAGES LIMITATIONS METHOD Linkage Studies Classical linkage analysis (parametric methods) Analysis of monogenic traits Difficult to collect large informative pedigrees Suitable for genome scan Difficult to obtain sufficient statistical power for complex traits Control population not required Useful for multifactorial disorders in isolated populations Allele-sharing methods (nonparametric methods) Suitable for identification of susceptibility genes in polygenic and multifactorial disorders Difficult to collect sufficient number of subjects Affected sib and relative pair analyses Suitable for genome scan Difficult to obtain sufficient statistical power for complex traits Sib pair analysis Control population not required if allele frequencies are known Reduced power compared to classical linkage, but not sensitive to specification of genetic mode Statistical power can be increased by including parents and relatives Association Studies Case-control studies Suitable for identification of susceptibility genes in polygenic and multifactorial disorders Requires large sample size and matched control population Linkage disequilibrium Suitable for testing specific allelic variants of known candidate loci False-positive results in the absence of suitable control population Transmission disequilibrium test (TDT) Facilitated by comprehensive catalogs of genotypes and variants Candidate gene approach does not permit detection of novel genes and pathways Whole-genome association studies Does not necessarily need relatives Susceptibility genes can vary among different populations Next-Generations Sequencing Technologies Whole exome or genome sequencing Unbiased approach, analysis can be performed without reference sequences from parents or siblings Requires appropriate bioinformatics, may have low sensitivity if CNV analysis is not included, detects numerous VUS, can lead to the detection of unrelated deleterious alleles Targeted sequencing of gene panels Captures multiple candidate genes and loci with hybridization techniques followed by deep sequencing Permits analyses of multiple candidate genes in parallel; facilitates molecular characterization of disorders with locus heterogeneity Abbreviations: CNV, copy number variation; VUS, variants of unknown significance. The analysis of more discrete sequence alterations often rely on the use of PCR, which allows rapid gene amplification and

analysis. Moreover, PCR makes it possible to perform genetic testing and mutational analysis with small amounts of DNA extracted from leukocytes or even from single cells, buccal cells, or hair roots. DNA sequencing can be performed directly on PCR products. The advent of comprehensive sequencing technologies analyzing the whole

exome or genome, of selected chromosomes, or of numerous candidate genes in a single run with NGS platforms is now fundamentally transforming the characterization of patients with rare disorders and advanced malignancies. These techniques have the advantage of an unbiased comprehensive approach, and they are increasingly cost-effective. Analysis of cell-free DNA (cfDNA; also referred to as "liquid biopsy") present in body fluids is playing a growing role for minimally invasive diagnostics and disease monitoring. Genomic tests are also widely used for the detection of pathogens and for the identification of viral or bacterial sequence variations. CHAPTER 479 Principles of Human Genetics The integration of genomic tests into clinical medicine is associated with a number of ongoing challenges related to variable sensitivities of the tests, bioinformatics analyses, storage and sharing of data, and the difficulty of interpreting all genetic variants identified with comprehensive testing. The discovery of incidental (or secondary) findings that are unrelated to the indication for the sequencing analysis, but indicators of other disorders of potential relevance for patient care can pose a difficult ethical dilemma. It can lead to the detection of undiagnosed medically actionable genetic conditions but can also reveal deleterious mutations that cannot be influenced, as numerous sequence variants are of unknown significance. A general algorithm for the approach to mutational analysis in patients with a suspected genetic disorder and (advanced) malignancies is outlined in Fig. 479-17. The importance of a detailed characterization of the clinical phenotype cannot be overemphasized. This is the step where one should also consider the possibility of genetic heterogeneity and phenocopies. If obvious candidate genes are suggested by the phenotype, they can be analyzed directly. After identification of a mutation, it is essential to demonstrate that it segregates with the phenotype. The functional characterization of novel mutations remains labor intensive and may require analyses in vitro or in transgenic models in order to document the relevance of the genetic alteration. Prenatal diagnosis of numerous genetic diseases in instances with a high risk for certain disorders is possible by direct DNA analysis. Amniocentesis involves the removal of a small amount of amniotic fluid, usually at 16 weeks of gestation. Cells can be collected and submitted for karyotype analyses, FISH, and mutational analysis of selected genes (Table 479-4). The main indications for amniocentesis include advanced maternal age (>35 years), presence of an abnormality of the fetus on ultrasound examination, an abnormal serum "quad" test (α fetoprotein, β human chorionic gonadotropin, inhibin-A, and unconjugated estriol), a family history of chromosomal abnormalities, or a Mendelian disorder amenable to genetic testing. Prenatal diagnosis can also be performed by chorionic villus sampling (CVS), in which a small amount of the chorion is removed by a transcervical or transabdominal biopsy. Chromosomes and DNA obtained from these cells can be submitted for cytogenetic and mutational analyses. CVS can be performed earlier in gestation (weeks 9–12) than amniocentesis, an aspect that may be of relevance when termination of pregnancy is a consideration. Later in pregnancy, beginning at ~18 weeks of gestation, percutaneous umbilical blood sampling (PUBS; cordocentesis) permits collection of fetal blood for analysis. Prenatal cfDNA allows DNA analyses from the mother and fetus from a maternal blood sample to screen for certain chromosomal abnormalities and fetal sex. These approaches enable screening for clinically relevant and deleterious alleles inherited from the parents, as well as for de novo germline mutations, and they have the potential to identify genetic disorders in the prenatal

setting. In combination with in vitro fertilization (IVF) techniques, it is possible to perform genetic diagnoses in a single cell removed from the four- to eight-cell embryo or to analyze the first polar body from an oocyte. Preconceptual diagnosis thereby avoids therapeutic abortions but is costly and labor intensive. It should be emphasized that excluding a specific disorder by any of these approaches is never equivalent to the assurance of having a normal child.

Characterization of phenotype Familial or sporadic genetic disorder PART 16 Genes, the Environment, and Disease Pedigree analysis Gene unknown Gene known or candidate genes Targeted sequencing Deep sequencing of DNA Deep sequencing of RNA (RNAseq) Deep sequencing (Linkage analysis and sequencing of linked region) Mutational analysis Determine functional properties of identified mutations in vitro and in vivo Genetic counseling Testing of other family members Therapy integrating genetic and genomic information Treatment based on pathophysiology FIGURE 479-17 Approach to genetic disease. Postnatal indications for cytogenetic analyses in infants or children include multiple congenital anomalies, suspicion of a known cytogenetic syndrome, developmental delay, dysmorphic features, autism, short stature, and disorders of sexual development, among others (Table 479-4). Mutations in certain cancer susceptibility genes such as BRCA1 and BRCA2 may identify individuals with an increased risk for the development of malignancies and result in risk-reducing interventions. The detection of cytogenetic alterations and mutations is an important diagnostic and prognostic tool in leukemias, and it has also transformed the management of solid tumors. In addition to providing diagnostic information, mutational analysis can inform the choice of targeted therapies (“actionable mutations”), characterize the mutational load, identify gene signatures associated with effective immunotherapies, and be used for surveillance. The demonstration of the presence or absence of mutations and polymorphisms is also relevant for the field of pharmacogenomics, including the identification of differences in drug treatment response or metabolism as a function of genetic background Gene therapy through the introduction of a normal gene or the ability to make site-specific modifications to the human genome has, so far, limited clinical application. However, several gene transfer methods have now been approved for clinical use, for example, for the treatment of Leber congenital amaurosis, B-cell acute lymphoblastic leukemia, spinal muscular atrophy, and hereditary transthyretin-mediated amyloidosis. Genome editing (or gene editing) with CRISPR-Cas9 is a promising novel approach for the treatment of various diseases, for example cystic fibrosis, certain cancers, hemophilia, and sickle cell disease. The first therapies using this technology for the treatment of sickle cell disease were approved by the U.S. Food and Drug Administration in 2023 (Chap. 483). ETHICAL ISSUES Determination of the association of genetic defects with disease, comprehensive data of an individual’s genome, and studies of genetic variation raise many ethical and legal issues. Genetic information is generally regarded as sensitive information that should not be

Patient with (advanced) cancer Tumor biopsy: Somatic analysis Peripheral cells: Germline analysis DNA and RNA extraction Bioinformatics Tumor board readily accessible without explicit consent (genetic privacy). The disclosure of genetic information may risk possible discrimination by insurers or employers. The scientific components of the Human Genome Project have been paralleled by efforts to examine ethical, social, and legal implications. An important milestone emerging from these endeavors is the Genetic Information Nondiscrimination Act (GINA), signed into law in 2008, which aims to protect asymptomatic individuals against the misuse of genetic information for health insurance and employment. It does not, however, protect the symptomatic individual. Provisions of the U.S. Patient Protection and Affordable Care Act, effective in 2014, have, in part,

filled this gap and prohibit exclusion from, or termination of, health insurance based on personal health status. Potential threats to the maintenance of genetic privacy include the increasing integration of genomic data into electronic medical records, compelled disclosures of health records, and direct-to-consumer genetic testing. It is widely accepted that identifying disease-causing genes can lead to improvements in diagnosis, treatment, and prevention. However, the information gleaned from genotypic results can have quite different impacts, depending on the availability of strategies to modify the course of disease. For example, the identification of mutations that cause MEN 2 or hemochromatosis allows specific interventions for affected family members. On the other hand, at present, the identification of an Alzheimer's or Huntington's disease gene does not currently alter therapy and outcomes. Most genetic disorders are likely to fall into an intermediate category where the opportunity for prevention or treatment is significant but limited. However, the progress in this area is unpredictable, as underscored by the finding that angiotensin II receptor blockers appear to slow disease progression in Marfan's syndrome. Genetic test results can generate anxiety in affected individuals and family members. Comprehensive sequence analyses are particularly challenging because most individuals can be expected to harbor several serious recessive gene mutations. Moreover, the sensitivity of comprehensive sequence analyses is not always greater, for example, if CNV analysis is not integrated. Genetic manipulation and patient selection for gene therapy approaches have raised ethical controversy and safety concerns that remain unresolved.

Revision #1

Created 2026-01-06 16:36:07 UTC by Omar Ayman

Updated 2026-01-06 16:36:07 UTC by Omar Ayman