Basic Genetic Principles

1

Haluk Kavus | Wendy K. Chung

INTRODUCTION

The human genome refers to the complete set of human DNA (with the suffix -ome arising from the Greek for "all" or "complete"). A copy of our genome comprises approximately 3 billion base pairs (bp) and about 20,000 protein-coding genes. The Human Genome Project was a significant contribution toward understanding the organization, structure, and sequence of the human genome.[1,2] With these developments, *genomic medicine* has emerged as a new discipline to analyze the genome and genetic information as a part of clinical care.

Having in-depth knowledge about the genome and the types and consequences of genomic variations is important for all medical professionals, especially neonatologists. Recognizing the most common chromosomal and monogenic disorders and genetic concepts such as inheritance, genomic imprinting, uniparental disomy (UPD), and X chromosome inactivation can help clinicians understand the origins of genetic conditions and risk of recurrence to patients and their families. Knowing the types of available clinical genomic tests, along with their utility and limitations, is critical for appropriate clinical use. In this chapter, we will also review prenatal diagnosis, clinical physical examination of the dysmorphic child, the future of newborn sequencing, and therapeutic approaches for monogenic diseases.

GENOMIC ORGANIZATION

DNA AND RNA

Deoxyribonucleic acid (DNA) and ribonucleic acid (RNA) are long polymers of nucleotides. Each nucleotide has three elements: (1) a nitrogenous base, (2) a sugar molecule, and (3) a phosphate molecule. The nitrogenous bases fall into two types: purines and pyrimidines. The purines include adenine and guanine; the pyrimidines include cytosine, thymine, and uracil. The primary difference between RNA and DNA is related to a base composition, such that RNA contains uracil, whereas DNA contains thymine. The other difference between RNA and DNA is in their sugar-phosphate backbones: RNA contains ribose, and DNA contains 2-deoxyribose. Deoxyribose confers resistance to hydrolysis, which gives DNA chemical stability, supporting the fidelity of the information as cells divide. While DNA is packaged in chromosomes in the nucleus, RNA carries that message from the nucleus to the cytoplasm, where it is made into proteins.

The double-helix structure of DNA was elucidated in 1953.[3] Hydrogen bonds zip up the complementary strands of DNA in which A pairs with T, and C pairs with G. Because of the complementary sequence of the two strands of DNA, there is redundancy in the information content, increasing the fidelity of the code. Replication of the double-stranded structure of DNA molecules requires a separation of the two strands followed by the synthesis of two new complementary strands. In contrast to DNA, RNA molecules are single-stranded and short-lived.

STRUCTURE OF CHROMOSOMES

Humans usually have 46 chromosomes in 23 pairs: 44 are autosomes, and 2 are the sex chromosomes (X and Y) involved in sex determination. The homologous pairs are numbered from 1 to 22 in order of decreasing size, with one member of each pair inherited from one parent.

Every chromosome consists of a long, single, and continuous DNA molecule located in the nucleus. DNA is packaged as chromatin complexed with histone proteins to condense the DNA into the nucleus. The five major types of histone proteins play a critical role in packaging of the chromatin. Two copies of the four core histones H2A, H2B, H3, and H4 constitute an octamer, around which a segment of DNA winds. A fifth histone, H1, binds to DNA at the tip of each nucleosome. Approximately 140 bp of DNA are linked with each histone core, making just under two turns around the octamer. After a short (20- to 60-bp) "spacer" segment of DNA, the pattern repeats, giving chromatin the look of beads on a string. Each complex of DNA with its core histones is called a *nucleosome*, which is the basic structural unit of chromatin.

Between cell divisions, the chromatin is unwound where genes are being expressed. With cell division, the chromosomes condense and become visible as the structures we observe in a karyotype. Noncoding RNA molecules play an essential role in gene regulation. For example, Xist, a noncoding RNA molecule, is a central regulator of X chromosome inactivation. It coats the inactivated X chromosome, which is structurally condensed, with most (but not all) genes being transcriptionally inactive.[4]

MITOCHONDRIAL GENOME

Mitochondria are organelles within cells that transform the energy from food into a form that cells can use. Each cell contains thousands of mitochondria, each containing several copies of a small circular mitochondrial chromosome. The mitochondrial DNA molecule is 16 kb in length and encodes 37 genes, all of which are fundamental for both normal mitochondrial function and also required for the function of ribosomal and transfer RNA molecules in the mitochondria. Mitochondrial genes are solely inherited from the mother.

STRUCTURE OF GENES

Genes in humans are composed of protein-coding sequences called *exons* and the intervening (noncoding) DNA sequences called *introns* (Fig. 1.1). Introns are initially transcribed into RNA in the nucleus and are spliced out to make the mature mRNA. Therefore, the information from the intronic sequences is not typically represented in the final protein product. Exonic sequences determine the amino acid sequence of the protein. Most genes contain at least one and usually numerous introns. The total length of the introns makes up a far greater proportion of a gene's total length for most genes. Genes are also flanked by additional sequences that are transcribed but untranslated, known as the 5′ and 3′ untranslated regions, which play a role in RNA stability and gene expression.

The human genome contains noncoding DNA sequences that act as *regulatory elements*. These sequences include promoters, enhancers, silencers, and locus control regions. They coordinate the regulation of genes in space and time. They include sequences for proteins called *transcription factors* to bind to and either increase or repress transcription. Promoter regions are responsible for the initiation of transcription and are typically found 5′ or upstream of a gene. Enhancers and silencers can be located either 5′ or 3′ of a gene, within the introns or sometimes farther away in neighboring genes. Promoters are binding sites for proteins that increase or repress transcription. Besides genes that are transcribed and made into proteins, there are genes known as non-coding RNAs (ncRNAs), whose functional product is RNA. Some ncRNAs can be quite long (long ncRNAs or lncRNAs) and play roles in gene regulation.[5] There are also small non-coding RNAs, known as *short interfering RNAs (siRNAs)* and *microRNAs (miRNAs)*, that control gene expression. MicroRNAs are short, non-coding RNAs, approximately 22 nucleotides in length, which post-transcriptionally regulate mRNA expression,[6] usually by decreasing expression. siRNAs are homologous to specific mRNAs and degrade the mRNA to decrease the expression of the target gene.

More than half of the human genome consists of various types of repeat sequences that are either clustered together or evenly distributed throughout the genome. These sequences can be short and consist of only a few nucleotides or can be as long as 5000 to 6000 nucleotides. The two best-studied dispersed repetitive elements are the Alu family and the long interspersed nuclear element (LINE) family. The Alu family makes up at least 10% of human DNA, and the LINE family accounts for nearly 20% of the genome. Segmental duplications are another repeat and are highly conserved and make up 5% of the human genome. When these duplicated sequences include genes, structural rearrangements can cause genetic diseases.

CELL DIVISION

Transmission of the genetic information from one generation to the next relies on accurate replication of DNA during reproduction. *Mitosis* is used during somatic cell division to support the development and cellular differentiation. In mitotic division, the usual complement of 46 chromosomes is maintained through a process of DNA replication and subsequent separation of the chromosomes. In contrast, *meiosis* occurs only in cells that become gametes, each of which has only 23 chromosomes (haploid genome). Thus errors in cell division in either somatic or germline cells can cause abnormalities of chromosome number or structure that can be clinically significant.

MOSAICISM

Mosaicism is the presence of at least two cell populations derived from the same zygote. Mitotic nondisjunction, trisomy rescue, or occurrence of a somatic new mutation can lead to the development of genetically different cell lines within the body. Mosaicism can affect any cells or tissue within a developing embryo at any point after conception to adulthood. If the mosaic cells are found only in the placenta and absent in the embryo, this known as *confined placental mosaicism* (*CPM*) (Fig. 1.2A).[7] CPM may be detectable on a chorionic villus sample and may be associated with intrauterine growth restriction but not with congenital anomalies or neurodevelopmental disorders if the genetic anomaly is not present in the fetus. Somatic mosaicism is the presence of two or more cell lineages in tissues that may have a clinically observable phenotype in the part of the body with the genetic aberration (see Fig. 1.2B). In gonadal mosaicism, the mosaic cells are restricted to the gametes and do not have a clinically observable phenotype but can be passed onto the next generation.

MEIOSIS

Meiosis is the process by which gametes are formed. In contrast with mitosis, in which a single cell division and an exact duplication of the genetic material occurs, meiosis involves two cell divisions, starting with a diploid parental cell and random reassortment and reduction of genetic material so that each of the four daughter cells has the haploid DNA content (i.e., 23 chromosomes). In this way, meiosis yields four haploid gametes (sperm or eggs). In female meiosis, the second meiotic division is completed only after fertilization, and advancing maternal age is associated with nondisjunction of the chromosomes. A polar body, containing a complete set of chromosomes, is extruded, leaving the egg with a single remaining haploid set of chromosomes. The second polar body is useful for preimplantation genetic diagnosis.

RECOMBINATION

During the prophase of the first meiotic division, homologous pairs of chromosomes are held together by the synaptonemal complex, which extends along the entire length of the paired chromosomes. Recombination between chromatids of the homologous chromosomes occurs at this stage, resulting in the exchange of DNA between the original parental chromosomes (Fig. 1.3). In males, the X and Y chromosomes are physically associated only at the tips of their short arms during meiotic prophase. This short region is called the *pseudoautosomal* region because recombination between the X and Y chromosomes occurs there (and thus it behaves as an autosome in terms of Mendelian inheritance). Recombination

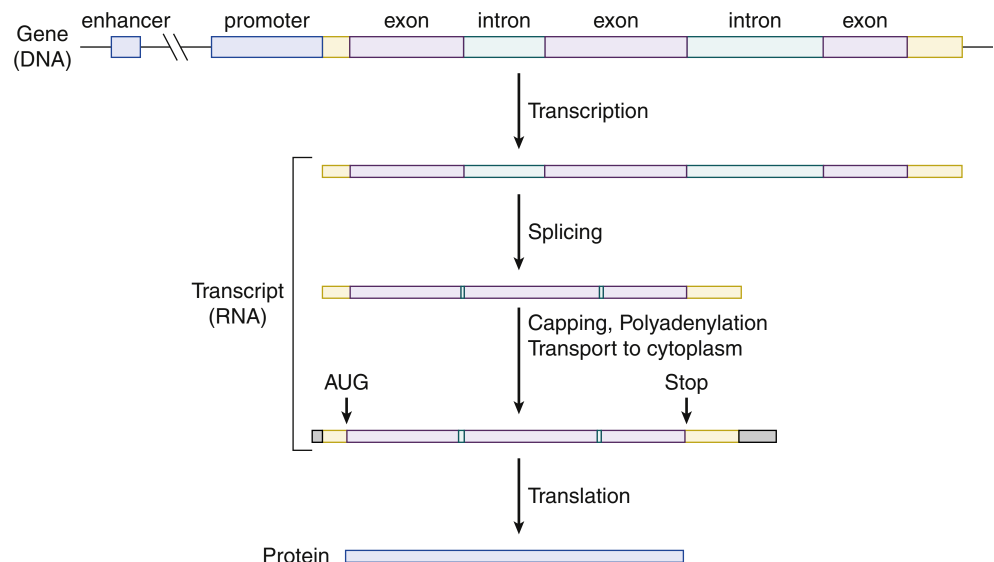


Fig. 1.1 Gene structure *(top)* and the flow of genetic information from DNA to protein. Tan boxes indicate the regions of exons that do not encode amino acid sequences; gray boxes indicate posttranscriptional modifications. AUG is a codon that specifies the amino acid methionine and is also used to specify the first amino acid of a protein.

involves the exchange of genetic material between the two homologs during meiosis I, and it is also critical for proper chromosome segregation during meiosis. Failure to recombine correctly can lead to nondisjunction of the chromosomes in meiosis I and is a frequent cause of aneuploidy (incorrect chromosome number) leading to pregnancy loss and congenital anomalies.

HOW GENES FUNCTION

FLOW OF GENETIC INFORMATION

TRANSCRIPTION

The first step in gene expression is the production of an RNA molecule from the DNA template. The RNA acts as a molecular messenger, carrying the genetic information out of the nucleus to the cytoplasm. The synthesis of mRNA is called *transcription* because the genetic information in DNA is transcribed. During transcription, the two DNA strands separate, and one functions as a template for the synthesis of single-stranded RNA molecules by RNA polymerases. The initial RNA transcripts are quite long because they include both introns and exons from the gene. The intronic sequences are cut out, and the remaining exons are spliced together. To form the mature mRNAs that leave the nucleus, a methylated guanine nucleotide called a *cap* is added to the 5′ end, and a string of 200 to 250 adenine (polyA tail) bases is added to the 3′ end. The cap is necessary for ribosomal binding to initiate protein synthesis, and the polyadenosine stretch at the 3′ end increases the stability of the mRNA.

Transcriptional control is central for the development and proper functioning of every organism. Transcriptional regulation is accomplished by modifying the DNA (e.g., cytosine methylation) or by protein binding to specific DNA sequences to activate or repress transcription of a gene. There are many sequence-specific transcription factors that are differentially active by cellular and tissue type and time in development. Several regulatory sequences have been identified in promoters that are important for transcriptional initiation by RNA polymerase II, including the TATA box, so-called because it consists of a run of T and A base pairs. The TATA box is located approximately 30 bases before the transcription start site and functions as the binding site for a large, multisubunit complex of transcription factors (including RNA polymerase). A second conserved region, the so-called *CAT box*, is a few dozen base pairs farther upstream. Specific sequence elements that form promoters and enhancers are required for binding the ~1400 sequence-specific proteins that bind to DNA and regulate transcription. Mutations in these regulatory sequence elements can lead to significant alterations in transcription and also can lead to genetic disorders.

The boundary between the introns and exons consists of a 5′ donor GT dinucleotide and a 3′ acceptor AG dinucleotide. Besides the canonical splice sequences, there are also splicing regulatory elements such as exonic splicing enhancers (ESEs) and exonic splicing silencers (ESSs). ESEs and ESSs correspond to six to eight nucleotides that serve as docking sites for splicing activator or splicing repressor proteins, thereby influencing the recruitment and activity of the splicing machinery.[8] Most human genes undergo alternative splicing and hence encode more than one protein for each gene. Alternative polyadenylation creates further diversity. Some genes have more than one promoter, and these alternative promoters may result in tissue-specific isoforms. Alternative splicing of exons is also seen with individual exons present in only some isoforms.

TRANSLATION

The production of protein from a mRNA template is called *translation* because the genetic information that is encoded in DNA is translated into a sequence of amino acids in the protein. The genetic information is stored in the genetic code. Each of the three adjacent nucleotides is a unit of information called a *codon* and specifies an amino acid or the start or stop of translation. The linear codons in the DNA sequence specify the sequence of amino acids in a protein. Because each of the three sites in a codon can be one of four possible nucleotides, a total of 4^3, or 64, different codons are possible. Three of these 64 possible codons, UAA, UAG, and UGA, are called *termination codons*. The remaining 61 codons specify one of the 20 amino acids, leading to some degeneracy in coding certain amino acids. A consequence of degeneracy is that some DNA variants do not result in a change in the amino acid sequence (synonymous variants).

EPIGENETICS

In addition to the classic transcription factors that bind to specific sequence elements in genes, gene expression is controlled by enzymes that modify DNA-bound proteins and DNA itself. The

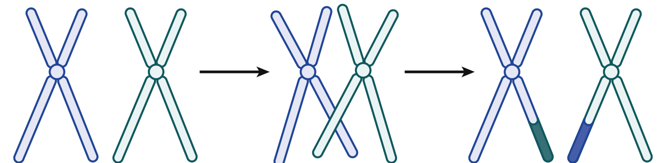
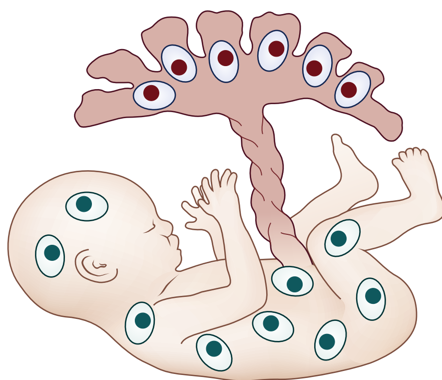


Fig. 1.3 Recombination. In this simplified view of recombination, the two members of a homologous pair of chromosomes line up during the first meiotic prophase. Segments of the two chromosomes "cross over," and breakage and rejoining of the DNA strands occur.



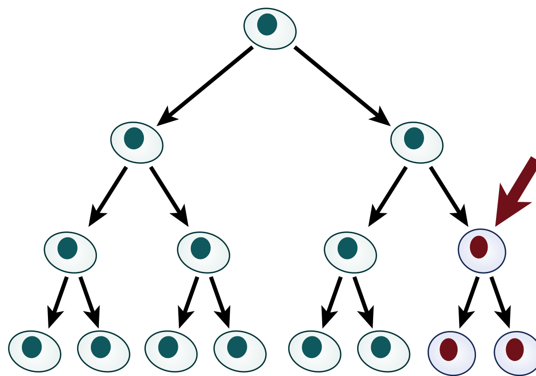


Fig. 1.2 Mosaicism. (A) Confined placental mosaicism. Presence of mutant cells only in the placental tissue, not in the fetus. (B) Somatic mosaicism. Presence of two or more mutant cell lineages in tissues and may have a clinically observable phenotype in the part of the body with the genetic aberration.

principal mechanism by which DNA is modified is by methylation of cytosine residues adjacent to guanosine. Methylation of these CpG dinucleotides by DNA methylases leads to transcriptional inactivation, while demethylation by demethylases alters the conformation of chromatin, leading to transcriptional activation. Histone proteins are extensively modified by many enzymes, including acetylases, kinases, and methylases. The pattern of histone modification, particularly on lysine residues, controls whether a particular region of chromatin will be transcriptionally active or inactive.

Modifications of chromatin proteins and DNA can be inherited through multiple cell divisions. Such alterations that do not change the DNA sequence itself and are called *epigenetic*. Genetic diseases that affect this process exemplify the importance of epigenetics. For example, mutations in MeCP2, a protein that binds to methylated DNA to repress the expression of associated genes, cause Rett syndrome, an X-linked neurodegenerative disease. Rubinstein-Taybi syndrome is caused by mutations in the CBP gene, encoding CREB-binding protein, which acts to acetylate the histone proteins that are major components of chromatin.

GENETIC VARIATION

A *locus* is a particular position on a chromosome for a specific gene and related DNA elements. *Alleles* refer to an alternative version of the DNA sequence at a locus. Generally, one of the alternative alleles is found in more than half of the population and is called is the *major allele*. The other versions of that gene refer to variants or minor alleles. Allele frequencies vary significantly in different populations. If an allele frequency is greater than 1%, it is said to be a *polymorphism* (multiple forms).

Mutation is generally meant to signify a DNA sequence that is deleterious and associated with a human disease. Mutations can be germline and inherited from one or both parents, or somatic and acquired over the life of an individual. Mutations can vary by the size of the altered DNA sequence. The size of mutations can range from a single nucleotide to the rearrangements of an entire chromosome. By convention, we have a reference genome that is used to compare genetic variants. This reference genome is updated as we understand the human genome better.

SINGLE NUCLEOTIDE VARIATIONS

When a "point mutation" of one or a small number of nucleotides occurs in a part of the gene that codes for a protein and alters the protein by changing the codon of which it is a part, it is called a *nonsynonymous variant* (Fig. 1.4A). Because the genetic code is degenerate, it is possible to have a point mutation that does not change the amino acid that is encoded. This is called a *synonymous variant*. Insertion or deletion of a nucleotide in the protein-coding portion of a gene is called a *frameshift mutation* because it changes the entire reading frame of the gene at every codon distal to the site of the mutation. *Nonsense mutations* are those point mutations that result in one of the three codons (UAA, UAG, UGA) that do not code for amino acids but rather truncate proteins, often producing proteins with little or no activity if they occur early enough in the protein. Point mutations occurring at the boundaries between introns and exons can cause improper splicing of RNA precursors, resulting in aberrant splicing or RNA instability.

Regulation of gene expression can be affected by mutations occurring in control elements, such as promoters and enhancers; the effect of such mutations is to quantitatively affect the amount of protein produced, such as occurs in some forms of thalassemia.

A different mutational mechanism involves the expansion of triplet repeat sequences *(dynamic mutations)*, caused by an increase in the number of copies of triplet repeats in the coding or noncoding region of a gene. These disorders include myotonic dystrophy, fragile X syndrome, and Huntington disease. The repeat size can increase with successive generations, and as the repeat number increases, the age of onset of the disease decreases, giving rise to the phenomenon of *anticipation*.

HOW MANY MENDELIAN CONDITIONS ARE THERE?

There are 8319 Mendelian phenotypes described. A total of 5489 of them (66%) have an associated known underlying gene. A total of 20% (3912/19,580) of human genes are known to underlie a Mendelian phenotype.[9] This information is curated in Online Mendelian Inheritance in Man (OMIM),[10] which is a continuation of Mendelian Inheritance in Man (MIM), published between 1966 and 1998.

CHROMOSOMAL MUTATIONS

Mutations in chromosome structure include deletions, duplications, inversions, and translocations. Because chromosomal alterations usually result in the disruption of multiple genes, they often have profound clinical consequences that include more than one organ system (see Fig. 1.4B).

If the total amount of genetic material is normal (just simply rearranged), then the karyotype is *balanced*. A *deletion* is the loss of a part of a chromosome and results in monosomy for that segment of the chromosome. An *insertion* is the addition of a segment of DNA into a chromosome. An insertion is often associated with an *unbalanced* chromosome complement. An *inversion* is a two-break rearrangement involving a single

A. Single-Gene Mutations

Sequence	Type
ATG · CTA · CGC · TGG · ACA · AGC Met · Leu · Arg · Try · Thr · Ser	Normal
ATG · CCA · CGC · TGG · ACA · AGC Met · Pro · Arg · Try · Thr · Ser	Missense
ATG · CTT · CGC · TGG · ACA · AGC Met · Leu · Arg · Try · Thr · Ser	Silent
ATG · CTA · CGC · TGA · ACA · AGC Met · Leu · Arg · (Stop)	Nonsense
ATG · CGT · ACG · CTG · GAC · AAG · C Met · Arg · Thr · Leu · Asp · Lys	Frameshift (insertion)

B. Chromosomal Mutations

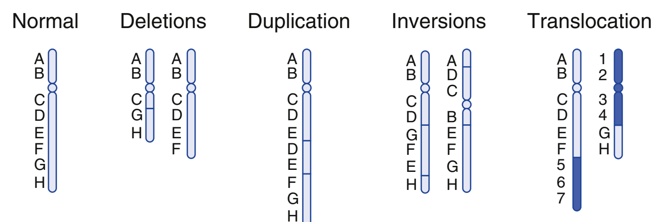


Fig. 1.4 Mutation. (A) Single-gene mutations. A prototypical normal gene sequence is shown on the first line, with the corresponding amino acid sequence. Examples of four types of common mutations also are shown. The substituted or inserted nucleotides are indicated by arrows, and the affected amino acids are underlined. (B) Chromosomal mutations. A prototype normal chromosome is shown, with genes A through H. Examples of gross chromosomal mutations are shown to the right, and their effects on gene content and arrangement are indicated. In the translocation example, the two chromosomes are not members of a homologous pair.

chromosome in which a segment is flipped and reversed in position. If the inversion segment includes the centromere, it is termed a *pericentric inversion*. If it involves only one arm of the chromosome and does not involve the centromere, it is known as a *paracentric inversion*. A *translocation* refers to the transfer of genetic material from one chromosome to another. If two different chromosome breaks and the chromosome segments are exchanged, it is called a *reciprocal translocation*. If the translocation occurs between two acrocentric chromosomes or chromosomes with redundant genetic material on the p arm, it is called a *Robertsonian translocation*. Similar to inversions, translocations can be balanced or unbalanced. Individuals who carry translocations but who have a normal amount of genetic material are called *balanced translocation carriers*. Translocations can have severe effects on offspring if the progeny has an unbalanced complement. When counseling a carrier of a balanced reciprocal translocation or Robertsonian translocation, it is necessary to consider the particular rearrangement to determine what the probability is that it could result in the birth of an abnormal baby. Translocation carriers are at increased risk of miscarriage of unbalanced products of conception.

A *ring chromosome* is formed when a break occurs on each arm of a chromosome leaving two "sticky" ends on the central portion of the chromosome that reunite as a ring. The two distal chromosomal fragments are lost, so if the involved chromosome is an autosome, the effects are usually severe.

Structural variant (SV) refers to genomic rearrangements and includes deletions, insertions, inversions, duplications, and translocations. *Copy number variation* (CNV) describes a group of DNA sequence variants (including deletions and duplications) that result in an abnormal number of copies of segments of DNA.[11]

GENETIC DISORDERS

CHROMOSOMAL DISORDERS

Chromosome disorders compose an important category of genetic disease, occurring in approximately 1 out of every 150 live births.[12] They are a common cause of intellectual disability and pregnancy loss. Chromosomal disorders can be divided into two groups: numerical and structural abnormalities. Numerical abnormalities result from one or more chromosome gains or losses, referred to as an *aneuploidy* (e.g., trisomy, monosomy, tetrasomy) or the addition of one or more complete haploid genomes, referred to as *polyploidy* (e.g., triploidy, tetraploidy).

In addition to the loss or gain of whole chromosomes, parts of chromosomes can be lost or duplicated as gametes are formed, and the arrangement of portions of chromosomes can be altered. Structural chromosome abnormalities may be unbalanced (the rearrangement causes a gain or loss of chromosomal material) or balanced (the rearrangement does not produce a loss or gain of chromosome material). Molecular methods including chromosome microarrays are often helpful to sensitively detect gains, losses, and rearrangements that may be missed by standard karyotype alone. Unlike aneuploidy and polyploidy, balanced structural abnormalities less frequently produce serious health consequences, and many are compatible with normal health and behavior. However, unbalanced abnormalities of chromosome structure and even some that are balanced but that disrupt key genes can create severe disease in individuals or their offspring. The phenotype associated with the chromosome disorder tends to run true in the family when inherited, so testing the parents to determine if a chromosome disorder is inherited or de novo is often recommended by a geneticist. Some structural alterations can be caused by translocations (reciprocal or Robertsonian), insertion, deletion, or rearrangement of DNA sequences, so examination of the parents of a child with a complex rearrangement may be helpful in predicting recurrence risk for the parents. Some deletions and insertions involve only a few nucleotides and are generally most easily detected by sequencing the relevant part of the genome. In other cases, a large segment of a gene, an entire gene, or several adjacent genes are deleted, duplicated, inverted, or translocated and create a novel arrangement. Depending on the exact nature of the deletion, insertion, or rearrangement, a variety of different laboratory approaches can be used to detect the genomic alteration, including karyotype, fluorescent in situ hybridization (FISH), chromosome microarray, or sequencing methods.

ANEUPLOIDY

Aneuploidy refers to missing or additional individual chromosomes in the cell (a number other than 46 chromosomes). Aneuploidies of chromosomes 13, 18, and 21 are among the most clinically important of the chromosome abnormalities. They consist of monosomy (the presence of only one copy of a chromosome in an otherwise diploid cell) and trisomy (three copies of a chromosome). Nondisjunction causes errors in chromosome segregation and leads to aneuploidies. Multiple congenital anomalies, growth restriction, and intellectual disability are the most common phenotypes of these trisomies.

Nevertheless, each has a reasonably unique neonatal phenotype that is recognizable by an experienced clinician. Trisomy 13 and 18 are both less common than trisomy 21, and survival beyond the first year is rare for trisomy 13 and 18. In contrast, individuals with Down syndrome have a life expectancy of over 50 years. Most other autosomal trisomies result in early pregnancy loss, with trisomy 16 being a particularly common finding in first-trimester spontaneous miscarriages.

Trisomy 21, which causes Down syndrome, is the most common autosomal aneuploidy seen among live births. The most common features include intellectual disability, hypotonia, gastrointestinal obstruction, congenital heart defects, and dysmorphic facial features. In approximately 90% of cases, the third chromosome 21 is of maternal origin. About 95% of Down syndrome cases are caused by nondisjunction, and Robertsonian translocations cause most of the remaining cases. Mosaicism is seen in 2% to 4% of Down syndrome cases and is often associated with a milder phenotype. The most frequent cause of mosaicism in trisomies is a trisomic conception followed by loss of the extra chromosome during mitosis in some embryonic cells (trisomic rescue).

Trisomy 13 and 18 are sometimes compatible with survival to term, although 95% or more of affected fetuses are spontaneously aborted. These trisomies produce more severe disease than trisomy 21, with 90% to 95% mortality during the first year of life.

Turner syndrome is most commonly associated with 45, X. Although this disorder is common at conception, it is relatively rare among live births, reflecting a high rate of spontaneous abortion. Mosaicism, including confined placental mosaicism, appears to increase the probability of survival to term. Klinefelter syndrome occurs in men who receive two X chromosomes in addition to the Y chromosome. The presence of an extra sex chromosome (X or Y) has only mild phenotypic effects.

GENOMIC DISORDERS (MICRODELETION AND DUPLICATION SYNDROMES, STRUCTURAL VARIATIONS, COPY NUMBER VARIATIONS)

Chromosomal deletions and duplications are an important cause of human malformations and intellectual disability. Those caused by submicroscopic changes were not easily detectable with a standard karyotype (with a resolution of approximately 5 to 10 Mb). With the advent of high-resolution banding, it has become feasible to identify smaller deletions. Advances in molecular genetics, mainly the FISH and chromosome microarray techniques, have permitted the detection of deletions that are often too small to be observed microscopically (i.e., <5 Mb).

Because these syndromes generally involve the deletion or duplication of a series of adjacent genes, it is sometimes referred to as a *contiguous gene syndrome*. Recent studies show that this is caused by the presence of flanking repeat sequences, termed *low-copy repeats* (also termed *segmental duplications*), at the deletion borders. These repeat sequences favor unequal crossing-over, which then produces duplications and deletions of the region bounded by the repeat elements. These disorders are collectively called *genomic disorders*.[13]

Some of well-known genomic disorders and their associated clinical features are shown in Table 1.1.

GENOMIC IMPRINTING

There are regions of the genome with parent-of-origin effects as a result of genomic imprinting. *Genomic imprinting* is an epigenetic term describing monoallelic gene expression according to parental origin. This epigenetic "mark" or imprint affects the chromatin structures and silences expression of the gene/gene(s) that are imprinted. The imprint is maintained throughout the life of the organism, in virtually all tissues; however, germ cells erase and then reset imprints for transmission to the next generation. Imprinting disorders can be caused by:

(a) sequence mutation of the relevant gene (*UBE3A* for Angelman syndrome),
(b) deletion or duplication of imprinted genes,
(c) UPD, or
(d) epigenetic errors in imprinting centers causing faulty imprinting.

Prader-Willi syndrome, Angelman syndrome, and Beckwith-Wiedemann syndrome are the best-studied examples of the role of genomic imprinting in human disease.

Table 1.1 Examples of Microdeletion Syndromes.

Syndrome	Chromosomal Locus	Major Clinical Features
Deletion 1p36 Syndrome	1p36	Intellectual disability, seizures, hearing loss, heart defects, growth failure, behavioral symptoms
Wolf-Hirschhorn Syndrome	4p16.3	Pre- and postnatal growth failure, iris coloboma, seizure, microcephaly
Cri-du-chat Syndrome	5p	High-pitched cat-like cry, microcephaly, hypotonia, developmental delay, low-set ears
Williams-Beuren Syndrome	7q11.23	Supravalvular aortic stenosis, hypercalcemia, periorbital fullness, thick lips, friendly personality
WAGR Syndrome	11p13	Wilms tumor, aniridia, genitourinary abnormalities, intellectual disability, obesity
Prader-Willi Syndrome	15q11–q13	Intellectual disability, short stature, obesity, hypotonia, characteristic facies, small feet
Angelman Syndrome	15q11–q13	Intellectual disability, ataxia, behavioral abnormalities, seizures, hypotonia, wide-based gait
Rubinstein-Taybi Syndrome	16p13.3	Intellectual disability, broad thumbs and great toes, vertebral and sternal abnormalities, heart defects
Smith-Magenis Syndrome	17p11.2	Intellectual disability, hyperactivity, dysmorphic features, self-destructive behavior
Miller-Dieker Syndrome	17p13.3	Lissencephaly, microcephaly, seizures, growth retardation, facial dysmorphism, early death
DiGeorge/VCF Syndrome	22q11.2	Characteristic facies, cleft palate, heart defects, hypocalcemia, thymic hypoplasia

UNIPARENTAL DISOMY

Uniparental disomy (UPD) refers to the presence of a disomic cell line containing two chromosomes that are inherited from only one parent, rather than one chromosome being inherited from the mother and the other from the father. If the disomic chromosomes are received from identical sister chromatids, it is called *isodisomy* (the same copy of two chromosomes); if both homologs come from one parent, the situation is *heterodisomy*.

Trisomy rescue is the mechanism of loss of a chromosome that restores a disomic state and escape from trisomy. If it happens, the resulting cell might show UPD.

If UPD occurs and includes a chromosome with an imprinted region such as chromosome 15, this may cause disease. For example, Angelman syndrome is due to mutations/deletions in the maternally expressed gene *UBE3A* or paternal UPD 15, such that there is no functional UBE3A allele, as the maternal allele is missing. Angelman syndrome can also be due to epigenetic modifications of the imprinting center on chromosome 15q11, which results in the loss of expression of UBE3A.

SEX CHROMOSOME ABNORMALITIES

Sex chromosome aneuploidies are common and may not be diagnosed for decades, if ever. Males with Klinefelter syndrome (47, XXY) are taller than average, may have slightly reduced IQ, and are usually infertile. The 47, XXX and 47, XYY karyotypes are present in about 1/1000 female and male births, respectively. Each involves a slight reduction in IQ and potentially behavioral issues, including attention deficit hyperactivity disorder, but are associated with few physical or medical problems.

X INACTIVATION

The principle of *X inactivation* is that in somatic cells in normal females (but not in normal males), one X chromosome is inactivated early in development, thus equalizing the expression of X-linked genes in the two sexes. In normal female development, because the choice of which X chromosome is to be inactivated is random and then clonally maintained, females are mosaic with respect to X-linked gene expression. In patients with extra X chromosomes (whether male or female), additional X chromosomes are randomly inactivated, leaving only one active X per cell. Thus all diploid somatic cells in both males and females have a single active X chromosome, regardless of the total number of X or Y chromosomes present.

The X chromosome contains approximately 1000 genes, but not all of them are subject to inactivation. Notably, the genes that continue to be expressed from the inactive X are not distributed randomly along the X chromosome. Many of the genes that "escape" inactivation are on distal Xp (as many as 50%) rather than on Xq (just a few percent). This finding has important implications for genetic counseling in cases of a partial X chromosome aneuploidy, because imbalance for genes on Xp

may have greater clinical significance than imbalance for genes on Xq, where the effect is mostly normalized by X inactivation.

SINGLE-GENE DISORDERS

A trait or disorder that is determined by a gene on an autosome is said to show *autosomal inheritance*, whereas a trait or disorder determined by a gene on one of the sex chromosomes is said to show *sex-linked inheritance*. For autosomal loci, the genotype of a person at a locus consists of both of the alleles occupying that locus on the two homologous chromosomes. If two alleles are the same for a particular locus, the genotype is *homozygous*. When the alleles are different, and one of the alleles is the reference allele (common variant in the population), it is called *heterozygous*. If the two alleles are different and neither is the reference allele, it is called *compound heterozygote*. The term *hemizygous* refers to one abnormal allele located on the X chromosome in a male. Mitochondrial loci are present in thousands of copies per cell; thus the terms mentioned herein are not used to describe mitochondrial genotypes.

There are two types of genetic heterogeneity. *Allelic heterogeneity* is when many different mutations in one gene cause similar phenotypes. *Locus heterogeneity* is when mutations in several different genes all cause a similar phenotype. *Phenotype* is the expression of genotype as a morphologic, clinical, cellular, or biochemical trait, which may be clinically observable or may only be detected by blood or tissue testing. The phenotype can be discrete or can be a measured quantity, such as body mass index or blood glucose levels.

AUTOSOMAL DOMINANT DISORDERS

Autosomal dominant (AD) disorders manifest in heterozygous individuals who have a single copy of the mutant allele. Many molecular mechanisms are associated with AD inheritance, including haploinsufficiency (loss of function of one amorphic allele), hypomorphic alleles with decreased function, hypermorphic alleles exhibiting gain of function (constitutive protein activity), neomorphic alleles that acquire a new function (such as altered substrate specificity), and toxic (antimorphic) effects of a protein, leading to dominant-negative (DN) mechanisms (Fig. 1.5A).

Principles of Autosomal Dominant Inheritance

- Males and females are affected equally.
- Males and females can both transmit the disorder.
- There is a 50% risk to offspring in any pregnancy in which one parent carries a mutation.
- The severity of the disorder in the offspring may vary, being similar, more severe, or less severe than in the parent.
- Many AD disorders are due to de novo mutations.

Dominant diseases show the same phenotype in either the heterozygous or homozygous state. The majority of AD disorders are incompletely dominant or semidominant, which means homozygous individuals have a more severe phenotype than heterozygous individuals. In codominant inheritance, both alleles are phenotypically expressed (e.g., ABO blood group).

Penetrance is the percentage of individuals who carry the relevant genotype and have signs or symptoms of the disorder. If all individuals who have a disease genotype show the disease phenotype, then the disease is said to be "fully penetrant," or to have a penetrance of 100%. Many dominant disorders show age-dependent penetrance. Some conditions show incomplete penetrance (i.e., not all mutation carriers will manifest the disease during a natural lifespan).

Expressivity is the difference in the severity of a disorder in individuals who have inherited the same disease alleles. Many genetic conditions show a striking difference between families (interfamilial variation) and also within families carrying the same mutation (intrafamilial variation). Pleiotropy is when a mutation in a single gene has effects on the body in more than one way (e.g., congenital heart disease and intellectual disability).

The proportion of cases resulting from new (de novo) mutations varies considerably between different AD conditions. In some disorders, the de novo mutation rate is high (nearly 100% of cases), whereas for some other conditions, a new mutation is unusual.

In many AD disorders, reproductive fitness is zero, (i.e., mutation carriers do not reproduce). Such a condition is maintained in the population entirely by new mutations, and the majority of cases are due to de novo mutations (although parental germline mosaicism may sometimes lead to recurrence in a sibling). Many other AD disorders have only modest effects on reproductive fitness.

AUTOSOMAL RECESSIVE DISORDERS

Autosomal recessive (AR) conditions occur when the mutant allele is present in both copies of the gene. Heterozygous individuals are said to be *carriers* for that condition and are asymptomatic (see Fig. 1.5B).

Principles of Autosomal Recessive Inheritance

- Disease is expressed only in homozygotes and compound heterozygotes.
- Parents are obligate carriers.
- Risk to carrier parents for having an affected child is 1 in 4.
- Healthy siblings of affected individuals have a two-thirds risk of being a carrier.
- Males and females are affected equally.

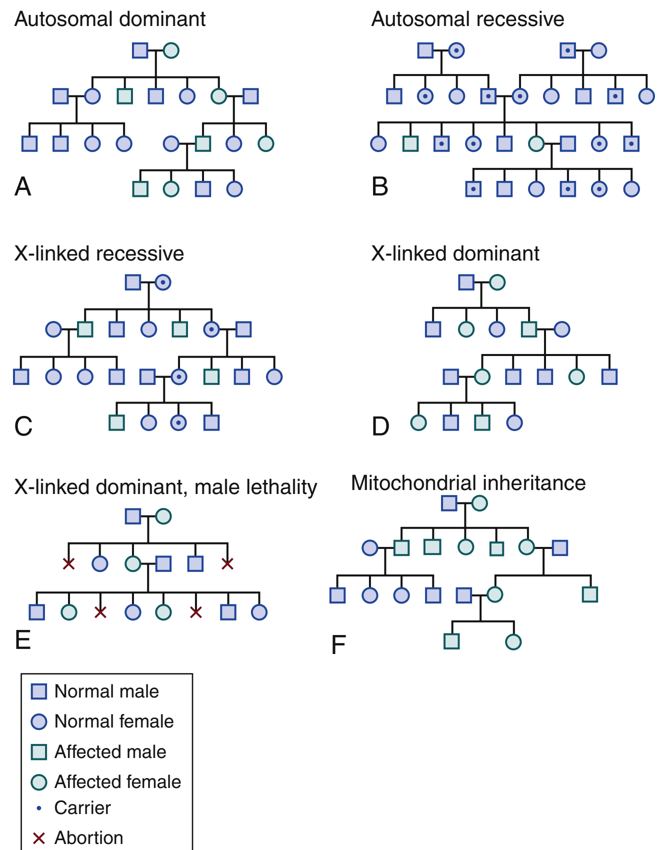


Fig. 1.5 (A–F) Pedigrees for disorders exhibiting the various mendelian and mitochondrial modes of inheritance. These are idealized pedigrees, assuming full penetrance and no new mutations.

Consanguinity is a significant risk factor for AR disorders. The *founder effect* is often responsible for a higher prevalence of a recessive genetic disorder in an isolated population, because multiple individuals of the population are descended from a common ancestor who carried a disease-causing mutation.

X-LINKED RECESSIVE DISORDERS

Principles of X-Linked Recessive Inheritance

See Fig. 1.5C.

- Males carrying the mutation are more severely affected; females carrying the mutation are generally either unaffected or more mildly affected than males, and severity in females depends on the X-inactivation pattern.
- When a carrier female conceives, there are four possible outcomes, each equally likely: a normal daughter, a carrier daughter, a normal son, and an affected son.
- The pedigree shows no male-to-male transmission.
- An affected male transmits the mutation to all his daughters.

X-LINKED DOMINANT DISORDERS

Principles of X-Linked Dominant Inheritance

See Fig. 1.5D.

- Males carrying the mutation are severely affected, often leading to spontaneous pregnancy loss or neonatal death of affected males (see Fig. 1.5E).
- Female heterozygotes are affected but have less severe features than males. X-inactivation patterns determine the degree to which females express the disorder.
- When a heterozygous affected female conceives, there are four genetic possibilities at conception, each equally likely: a normal daughter, an affected daughter, a normal son, and an affected son.
- The pedigree shows no male-to-male transmission.
- When an affected male has a child, all of his daughters will inherit the mutation, and none of his sons will be affected (unless their mother is a carrier).
- Males who are born with features of a severe and usually lethal XLD condition might have Klinefelter syndrome or an additional X chromosome disorder.
- Females with unusually severe features of an XLD disorder may have this as a consequence of:
  - unfavorably X-inactivation,
  - Turner syndrome (45, X), or
  - X-autosome translocation.

MITOCHONDRIAL DISORDERS

PRINCIPLES OF MITOCHONDRIAL INHERITANCE

See Fig. 1.5F

- The condition can only be transmitted by females in the maternal line.
- Males do not transmit mitochondrial inherited disorders, with extremely rare exceptions.
- A mitochondrial inherited condition can affect either sex.
- If the mother is heteroplasmic for a mutation, the proportion of mutant mtDNA in her offspring can vary considerably between offspring and between tissues in the same offspring.

MULTIFACTORIAL DISORDERS

If mutations in two different genes are necessary to cause a phenotype, the inheritance is *digenic*. If a disease is the result of contributions on a few genes, it is *oligogenic*. If a disease is the result of the combination of many genes, it is *polygenic*. If a disease is the result of the combination of many genes and the environment, it is *multifactorial*.

GENETIC DIAGNOSIS

VALUE OF A DIAGNOSIS AND NEED FOR A DIAGNOSIS

Genetic disorders can be classified as Mendelian (monogenic), oligogenic, chromosomal, and multifactorial. In this section, we will limit ourselves to rare diseases. Rare diseases are described as those that affect fewer than 200,000 people in the United States. There are more than 7,000 rare diseases, approximately 80% of which are thought to have a genetic cause.[10] The majority (50% to 75%) of rare diseases affect children.[14] Making an accurate diagnosis is fundamental to good patient care. Molecular diagnoses for patients with rare disorders are important for patients and their families to refine prognosis, management, recurrence risk, and reproductive options; identify other families with the same condition; and participate in research for new sources of support and therapies.[15] It is important to diagnose rare diseases early in the disease course, when the medical and financial burdens to families are fairly minimal, and then proceed with treatment.[16] Increasingly, genomic diagnostic methods are more comprehensive and support more facile diagnoses of genetic conditions.

CLINICAL APPROACH TO CONGENITAL ANOMALIES AND THE DYSMORPHIC CHILD

Dysmorphology is the study of congenital anomalies and developmental disorders. The term is used to describe visible malformations or distinctive structural features of the face or other parts of the body.

A clinical geneticist is often consulted for one or more of the following reasons:

- to examine the patient, gather the clinical and familial information, and plan the genetic tests to order and interpret to make a diagnosis;
- to interpret an existing genetic test result, particularly one that is not definitive;
- to discuss the genetic basis of the condition with the family;
- to recommend studies to evaluate for disease manifestations once a genetic diagnosis is made;
- to consult on the prognosis, management, and therapeutic options;
- to discuss the risk of recurrence and reproductive options;
- to coordinate prenatal/preimplantation genetic testing; and
- to discuss genetic test results with other members of the extended family and arrange for genetic testing for relatives.

PHYSICAL EXAMINATION, FAMILY HISTORY, AND PEDIGREE INFORMATION

A genetic consultation differs in the ability to examine the patient in the prenatal versus neonatal timeframe. In this section, we will focus on how a clinical geneticist evaluates a child. Although the same strategy applies to the prenatal setting, there is less clinical data available in the prenatal setting because the physical examination is limited to ultrasound and magnetic resonance images, and because blood samples cannot be readily obtained.

A genetic consultation includes a review of medical records; interviews with the family to review prenatal exposures; a three-generation family history, including ancestry and any history of consanguinity; and physical examination of the child and, in some cases, the parents. The geneticist will develop a differential diagnosis and order appropriate genetic tests to evaluate the possible conditions. This usually requires a biologic sample from the fetus/neonate and may require samples from the parents for comparison. The genetic testing may yield a definitive diagnosis or, in some cases, will suggest a possible diagnosis for which the geneticist must reassess the patient and determine whether the diagnosis fits the phenotype. If the test does not yield a diagnosis,

continued follow-up is important because additional clinical features may become apparent over time and aid in making the diagnosis; in addition, clinical diagnostic lab methods may continue to evolve and improve.

DIAGNOSTIC METHODS IN CLINICAL GENETICS

There are many methods for clinical genetic testing, each of which has been developed for particular clinical scenarios, including carrier screening, non-invasive prenatal testing, newborn screening, and diagnostic testing, with chromosome analyses to detect copy number variants and cytogenetic rearrangement and sequence-based tests.[17]

Many types of chromosomal abnormalities have been clinically and cytogenetically described and are diagnosed using conventional karyotyping, FISH, and chromosomal microarray (CMA). CMA is routinely performed as part of the comprehensive diagnostic testing for patients with unexplained developmental delay/intellectual disability, autism spectrum disorders, or multiple congenital anomalies.[18] CMA is now routinely performed as oligoarrays, with single nucleotide polymorphism (SNP) probes to provide high resolution for copy number variants and identify UPD and long stretches of homozygosity in families with consanguinity.[19]

Sanger sequencing was the primary genetic test for the diagnosis for monogenic disorders due to sequence variants; however, the decreased cost and increased throughput of massive parallel next-generation sequencing has significantly increased the number of conditions that can simultaneously be tested and can now include an entire genome. For conditions that are genetically heterogeneous, rather than selecting a single gene to test, it is now routine to test for a panel of genes causing a particular phenotype/disease. In addition, whole exome sequencing (WES), of all coding segments of almost all genes, and whole genome sequencing (WGS) are feasible and can even be performed within 1 to 2 weeks. As the number of genes assessed increases, the number of genetic variants that could be pathogenic also increases. The laboratory may issue a report with several variants in several genes as possible diagnoses and then rely on the geneticist to further assess the likelihood of the possibilities. Additionally, not all variants are detectable by exome/genome sequencing. Triplet repeats and somatic mutations can be particularly difficult or impossible to detect based upon current sequencing methods and read depth.

As the amount of genetic data generated increases with genetic testing, such as WES and WGS, there is a chance of identifying gene variants of clinical relevance that were not related to the primary indication for testing (incidental findings such as mutations for hereditary cancer or hereditary causes of sudden cardiac death). When the laboratory systematically and intentionally looks for variants in a prespecified set of genes unrelated to the primary indication, these are termed *secondary findings*. The consent process is important to determine which findings the patient would like to receive. Thus the generally accepted approach for incidental findings is to examine the exome/genome data for pathogenic/likely pathogenic variants in 59 genes.[20]

A summary comparing the different clinical genetic testing methods is provided in Table 1.2.

NEED FOR VARIANT REINTERPRETATION OVER TIME

A significant challenge associated with the clinical implementation of next-generation sequencing for large panels and exomes/genomes is the large number of variants identified. Distinguishing which of these variants is pathogenic is difficult since many of the variants identified are rare or novel and little is known about them. In addition, because not all disease genes have yet been identified, a diagnosis may be missed, despite comprehensive exome/genome sequencing, because the condition has not yet been scientifically recognized. Re-evaluation of sequence data over time may clarify the pathogenicity of variants and yield additional diagnoses as scientific understanding of genetic variants and additional genetic conditions advances. Thus reanalyzing and reinterpreting clinical sequence data is inevitable. The ordering healthcare provider, clinical geneticist, clinical laboratory, and patient/family each may have a role regarding reinterpretation of genetic results.[21,22] These expectations should be clearly outlined as part of the informed consent process.

FUTURE DIRECTIONS IN CLINICAL GENETICS

EARLIER DIAGNOSIS, INCREASINGLY PRENATAL

Making a genetic diagnosis earlier in life has a greater impact on medical care and may afford more effective treatment opportunities and minimize harm by decreasing the number of unnecessary diagnostic procedures or ineffective interventions. Rapid diagnosis in the neonatal or even prenatal period allows providers and parents to make more informed decisions about care, obtain more accurate prognostic information, and draw upon experience with the genetic condition. Rapid diagnosis of acutely ill patients in neonatal intensive care unit is increasingly common and decreases costs and length of stay.[23] One of the most common uses of NGS is non-invasive prenatal screening

Table 1.2 Comparison of Clinical Genetic Testing Methods.

	Karyotype	Chromosome SNP Microarray	FISH	Sanger Sequencing	Sequencing Panel	Exome Sequencing	Genome Sequencing
Single nucleotide variations (SNVs)				X	X	X	X
Copy number variations (CNVs)		X	X			+/−	X
Balanced chromosomal rearrangement	X		X				+/−
Identification of new disease genes						X	X
Incidental findings						X	X
Cost	Low	Low	Low	Modest	Modest	High	High

FISH, Fluorescence in situ hybridization; *SNP*, single nucleotide polymorphism.

(NIPS) to noninvasively identify pregnancies with a high likelihood of chromosomal aneuploidies using a maternal blood sample and enriching for and analyzing fragmented fetal DNA within the sample. With an amniocentesis or chorionic villus sample, karyotype, chromosome microarray, single/panel gene tests, WES, and WGS can be used to prenatally diagnose genetic diseases when there is an aberrant ultrasound finding or based upon a family history of a genetic condition, carrier screening, or NIPS result.

FUTURE OF NEWBORN SCREENING

Newborn screening (NBS) is carried out via various large public-health genetic-screening programs. The scope of NBS programs has increased with advances in technology. Current NBS programs in the United States consist of approximately 50 conditions and rely heavily on tandem mass spectrometry to detect inborn errors of metabolism. Molecular methods have been increasingly integrated as second-tier tests for cystic fibrosis or as first-tier tests for immunodeficiencies (TREC assay) and spinal muscular atrophy. In the future, it is possible that NBS will be expanded even further to screen for classes of genetic disorders that can only be diagnosed with sequencing methods such as glucose transport (GLUT1) deficiency syndrome or retinoblastoma. Recent pilot studies of sequencing-based methods are not as sensitive as biochemical screening. Therefore, the utilization of sequence-based methods will first require greater accuracy in variant interpretation in these genes across diverse ethnicities to maximize clinical sensitivity.

THERAPEUTIC APPROACHES FOR MONOGENIC DISEASES

There is an increased demand for therapies of monogenic diseases with the improvement of diagnostic methods described in previous sections. Most of the inborn errors of metabolism rely on dietary changes, enzyme replacement therapies (ERTs), or liver or bone marrow transplant. Transplantation also works for other genetic conditions beyond the inborn error of metabolism (e.g., bone marrow transplant for hematologic conditions and immunodeficiencies). However, therapies are only available for a fraction of genetic diseases.

Therapeutic options for monogenic diseases may significantly expand in the next decade. Gene editing has advanced, specifically with the use of Crispr-Cas9 for gene editing. There is no safe and efficient, approved gene editing for humans yet, but it is an active area of research. Gene therapy trials began in the 1990s but were limited by safety issues, including immunologic responses to viral vectors and genomic integration activating proto-oncogenes, leading to fatal leukemias. With the advent of safer viral vectors without genomic integration and with trophism for a greater range of tissues, there have been significant advances in gene therapy for monogenic diseases (Fig. 1.6). The European Medicines Agency (EMA) and the US Food and Drug Administration (FDA) have approved six gene therapy products since 2016,[24] including gene therapy for spinal muscular atrophy, a common genetic cause of death in infants. In addition to gene replacement or gene addition, there are improving technologies for somatic gene editing. Numerous gene-therapy clinical trials for monogenic diseases, including particularly hematologic, immunologic, and hepatic diseases, are ongoing (Table 1.3). Longer-term outcome studies of safety and durability are still needed, but gene therapy is likely to play a greater role in treatment in the future.

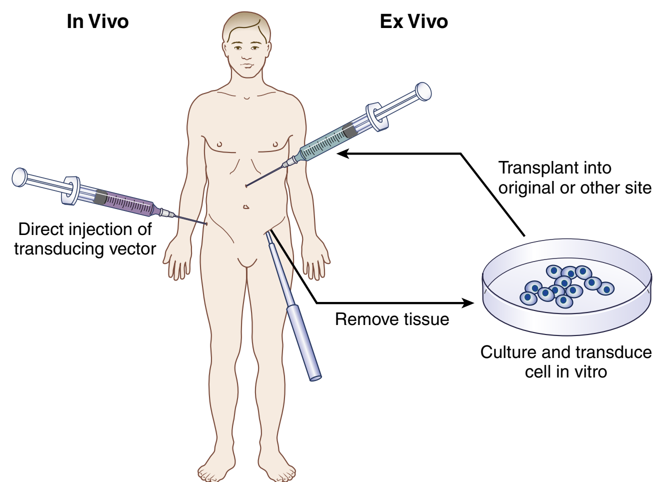


Fig. 1.6 Models for human gene therapy. In vivo gene therapy involves direct introduction of a transducing vector into the patient. Ex vivo gene therapy involves removal of tissue and transduction in vitro.

Table 1.3 Therapeutic Methods for Genetic Diseases.

Method	Example
Dietary for errors of metabolism	PKU and some other inborn errors of metabolism
Enzyme replacement therapy (ERT)	Pompe, CLN2, MPS I, MPS IVA, MPS VIA
Pharmacologic chaperone therapy (PCT)	Pompe, Fabry, Gaucher
Recombinant peptide analog	Achondroplasia (Vosoritide [clinical trial continue] is a C-type natriuretic peptide analog that stimulates enchondral ossification and inhibits the FGFR3-mediated MAPK signaling pathway.)
Gene therapy	Retinal dystrophy (RPE65, AAV vector), SMA (Zolgensma, AAV vector)
Gene editing	Hemoglobinopathies, sickle cell, DMD, immuno-oncology (all in clinical trials)
Antisense and other therapeutic oligonucleotides	SMA (Nusinersen), hATTR amyloidosis (Tegsedi), trinucleotide repeat disorders
Monoclonal antibodies	X-linked hypophosphatemia (XLH) (Burosumab is a fully human monoclonal IgG1 antibody against FGF23 protein.)

CONCLUSION

The Human Genome Project has had a significant impact on medicine and especially neonatology. We routinely use genomic methods and data to diagnose genetic diseases in newborns, especially those associated with congenital anomalies, neurologic conditions, and early organ failure. We are entering the next exciting era, going beyond the diagnosis, toward the treatment of genetic diseases. We need to continue to understand the basic mechanisms of human genes to develop rational treatment strategies. Moreover, transformational platforms to deliver or alter genes may enable the treatment of whole classes of genetic diseases and offer new treatment opportunities to patients.

SUGGESTED READINGS

Nussbaum R, McInnes R, Willard H. *Thompson & Thompson Genetics In Medicine*. 8th ed. Elsevier; 2016.
Turnpenny P, Ellard S. *Emery's Elements of Medical Genetics*. 15th ed. Elsevier; 2017.
Jorde L, Carey J, Bamshad M. *Medical Genetics*. 6th ed. Elsevier; 2019.
Firth H, Hurst J. *Oxford Desk Reference Clinical Genetics and Genomics*. 2nd ed. Oxford University Press; 2017.

REFERENCES

1. Venter JC, Adams MD, Myers EW, et al. The sequence of the human genome. *Science*. 2001;291(5507):1304–1351. https://doi.org/10.1126/science.1058040.
2. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860–921. https://doi.org/10.1038/35057062.
3. Watson JD, Crick FH. Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*. 1953;171(4356):737–738.
4. Tukiainen T, Villani AC, Yen A, et al. Landscape of X chromosome inactivation across human tissues. *Nature*. 2017;550(7675):244–248. https://doi.org/10.1038/nature24265.
5. ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. *Nature*. 2012;489(7414):57–74. https://doi.org/10.1038/nature11247.
6. Carthew RW, Sontheimer EJ. Origins and mechanisms of miRNAs and siRNAs. *Cell*. 2009;136(4):642–655. https://doi.org/10.1016/j.cell.2009.01.035.
7. Kalousek DK, Vekemans M. Confined placental mosaicism. *J Med Genet.* 1996;33(7):529–533.
8. Soukarieh O, Gaildrat P, Hamieh M, et al. Exonic splicing mutations are more prevalent than currently estimated and can be predicted by using in silico tools. *PLoS Genet*. 2016;12(1):1–26. https://doi.org/10.1371/journal.pgen.1005756.
9. Centers for Mendelian Genomics. Mendelian traits by the numbers. http://mendelian.org/mendelian-traits-numbers. Accessed September 18, 2020.
10. Amberger JS, Bocchini CA, Scott AF, Hamosh A. OMIM.org: leveraging knowledge across phenotype-gene relationships. *Nucleic Acids Res*. 2019;47(D1):D1038–D1043. https://doi.org/10.1093/nar/gky1151.
11. Alkan C, Coe BP, Eichler EE. Genome structural variation discovery and genotyping. *Nat Rev Genet*. 2011;12(5):363–376. https://doi.org/10.1038/nrg2958.
12. Hsu L. *Prenatal Diagnosis of Chromosomal Abnormalities Through Amniocentesis in Genetic Disorders and the Fetus*. 3rd ed. Baltimore: Johns Hopkins University Press; 1992.
13. Lupski JR. 2018 Victor A. McKusick Leadership Award: Molecular Mechanisms for Genomic and Chromosomal Rearrangements. *Am J Hum Genet*. 2019;104(3):391–406. https://doi.org/10.1016/j.ajhg.2018.12.018.
14. European Organisation for Rare Diseases. Rare diseases: understanding this public health priority. *Eurordis*. 2005.
15. Wright CF, FitzPatrick DR, Firth HV. Paediatric genomics: diagnosing rare disease in children. *Nat Rev Genet*. 2018;19(5):253–268. https://doi.org/10.1038/nrg.2017.116.
16. Boycott KM, Rath A, Chong JX, et al. International cooperation to enable the diagnosis of all rare genetic diseases. *Am J Hum Genet*. 2017;100(5):695–705. https://doi.org/10.1016/j.ajhg.2017.04.003.
17. Katsanis SH, Katsanis N. Molecular genetic testing and the future of clinical genomics. *Genomic Precis Med Found Transl Implement Third Ed*. 2016;14(6):263–282. https://doi.org/10.1016/B978-0-12-800681-8.00018-9.
18. Miller DT, Adam MP, Aradhya S, et al. Consensus statement: chromosomal microarray is a first-tier clinical diagnostic test for individuals with developmental disabilities or congenital anomalies. *Am J Hum Genet*. 2010;86(5):749–764. https://doi.org/10.1016/j.ajhg.2010.04.006.
19. Harel T, Lupski JR. Genomic disorders 20 years on-mechanisms for clinical manifestations. *Clin Genet*. 2018;93(3):439–449. https://doi.org/10.1111/cge.13146.
20. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Genet Med*. 2017;19(2):249–255. https://doi.org/10.1038/gim.2016.190.
21. David KL, Best RG, Brenman LM, et al. Patient re-contact after revision of genomic test results: points to consider—a statement of the American College of Medical Genetics and Genomics (ACMG). *Genet Med*. 2018;0(0):1–3. https://doi.org/10.1038/s41436-018-0391-z.
22. Bombard Y, Brothers KB, Fitzgerald-Butt S, et al. The responsibility to recontact research participants after reinterpretation of genetic and genomic research results. *Am J Hum Genet*. 2019;104(4):578–595. https://doi.org/10.1016/j.ajhg.2019.02.025.
23. Farnaes L, Hildreth A, Sweeney NM, et al. Rapid whole-genome sequencing decreases infant morbidity and cost of hospitalization. *Npj Genomic Med*. 2018;3(1). https://doi.org/10.1038/s41525-018-0049-4.
24. High KA, Roncarolo MG. Gene therapy. *N Engl J Med*. 2019;381(5):455–464. https://doi.org/10.1056/NEJMra1706910.

Epigenetics 2

William Schierding | Mark H. Vickers | Justin M. O'Sullivan | Wayne S. Cutfield

PRINCIPLES OF EPIGENETICS

INTRODUCTION TO EPIGENETICS

The publication of the majority of the human genome sequence in 2001[1,2] was the precursor to many important discoveries. However, the human genome sequence has not provided researchers with the codex to fully understand the genome's functionality or to predict its response to environmental cues (such as nutritional challenges). One reason why this is the case is that the human genome is more complicated than was originally postulated. Counterintuitively, this complexity partially arises from the finding that the human genome only has approximately one third of the predicted number of genes.[3] Fewer genes means that those genes that are present are more complex, producing multiple different messenger RNAs. As a result, the regulatory processes that control the expression of these genes are complex,[4,5] involving multiple layers of regulation, much of which still remains to be discovered and described.

Traditionally, it had been assumed that inherited genes control gene expression and, ultimately, phenotype. In the early 1940s Waddington introduced the concept of **epigenetics** ("on to top of" or "in addition to" genes) to describe the way in which genes interact with their surroundings to produce a phenotype during the differentiation of cells over the course of development (without a change in gene sequence).[6] Thus, environmental cues can lead to up- or down-regulation of gene activity. This definition leaves out the concept of inheritance, instead emphasizing the effect on the final cell type and how small nongenetic changes in development can lead to measurable differences in adult phenotype. Recently, epigenetics has been redefined, first by Riggs as "the study of mitotically and/or mitotically heritable changes in gene function that cannot be explained by changes in DNA sequence"[7] and more recently by Cavalli and Heard as "the study of molecules and mechanisms that can *perpetuate* alternative gene activity states in the context of the same DNA sequence."[8] Therefore, epigenetics is any element with permanent (or at least semi-permanent) changes in gene expression or cellular phenotype.[9] This encompasses transgenerational inheritance and the persistence of gene activity or chromatin states through extended periods of time. Throughout this chapter we will discuss epigenetics in the context of this more modern definition, but it should be noted that *epigenetics* is a term that has many different definitions, with "mitotically stable" and "epigenetic memory" being points of controversy.

FROM GENETICS TO EPIGENETICS

Double-stranded DNA is an efficient and reliable mechanism to pass information from one generation to another, given that it is stable and there are a number of repair systems that have evolved to maintain it. Thus, genetic changes tend to occur slowly, taking