Introduction to Genetics

1

Janell Kierstein

Contents

J. Kierstein (✉)
Clinical Genetics and Metabolism, Children's Hospital Colorado, University of Colorado Denver – Anschutz Medical Campus, Aurora, CO, USA
e-mail: Janell.Kierstein@childrenscolorado.org

L. E. Bernstein et al. (eds.), *Nutrition Management of Inherited Metabolic Diseases*,
https://doi.org/10.1007/978-3-030-94510-7_1

Core Messages

- Genes are short segments of DNA that carry information to make cellular proteins necessary for life.
- Mutations are heritable changes in the DNA nucleotide sequence of a gene.
- There are three broad categories of mutations: substitutions, insertions, and deletions.
- The most frequent consequence of a mutation is loss of protein function.
- Most inherited metabolic diseases are transmitted as an autosomal-recessive trait.

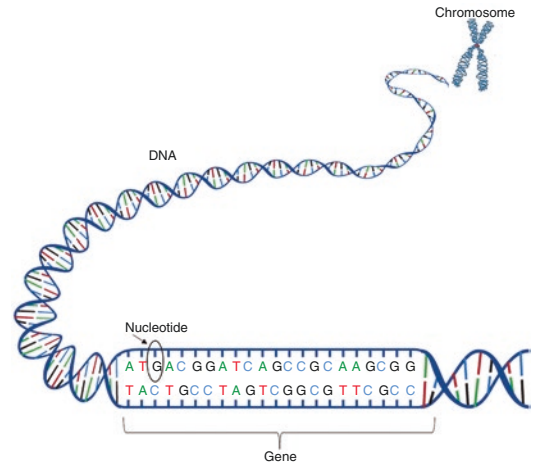

Fig. 1.1 Chromosomes are structures made of tightly coiled DNA

1.1 Background

The total number of cells in a human body is estimated to be 37.2 trillion [1]. In the nucleus of each of these cells are 46 chromosomes. One set of 23 chromosomes is inherited from an individual's mother and a second set of 23 chromosomes is inherited from their father. In every typical cell, there are 22 pairs of autosomes, which are the same in males and females. There is also one pair of sex chromosomes. Typically, females have two X chromosomes and males have one X chromosome (inherited from their mother) and one Y chromosome (inherited from their father). Each gamete (egg or sperm cell) contains only one member of each autosomal pair and one sex chromosome. When an egg cell is fertilized by a sperm cell, the result is a complete set of 46 chromosomes and unique individual.

Each chromosome is made up of tightly coiled strands of deoxyribonucleic acid (DNA). DNA is composed of long strings of nucleotides (Fig. 1.1). Each nucleotide includes a phosphate and sugar (deoxyribose) backbone attached to one of four nitrogenous bases: adenine (A), cytosine (C), guanine (G), and thymine (T) (Fig. 1.2). These bases form the chemical alphabet of DNA. The bases are complementary such that A always bonds with T on the opposing strand of DNA and C always bonds with G. Together, these matched nucleotides are referred to as base pairs. The bonds effectively create a ladder with rungs of bonded nucleotides and sides of sugar + phosphate backbones. The ladder is twisted into a double helical shape.

The vast majority of human DNA (approximately 98–99%) is considered "noncoding" DNA; its purpose is still enigmatic [2]. The remaining 1–2% of the human genome contains approximately 20,000 unique genes. Each gene represents a sequence of DNA that serves as a blueprint or code for production of a specific protein. Within each gene are segments of DNA called exons that actively encode protein production. Interspersed with exons are introns, which are composed of noncoding DNA. Genes also contain a variety of regulatory elements, which mark where a specific gene starts and ends and control in what tissues, at what point(s) in development, and how much of the final protein is produced (Fig. 1.3).

1.2 From Genes to Proteins

The process by which genes lead to production of proteins involves two key steps: transcription and translation. Transcription occurs when one DNA strand of a gene is used as the template to make a complementary strand of ribonucleic acid (RNA).

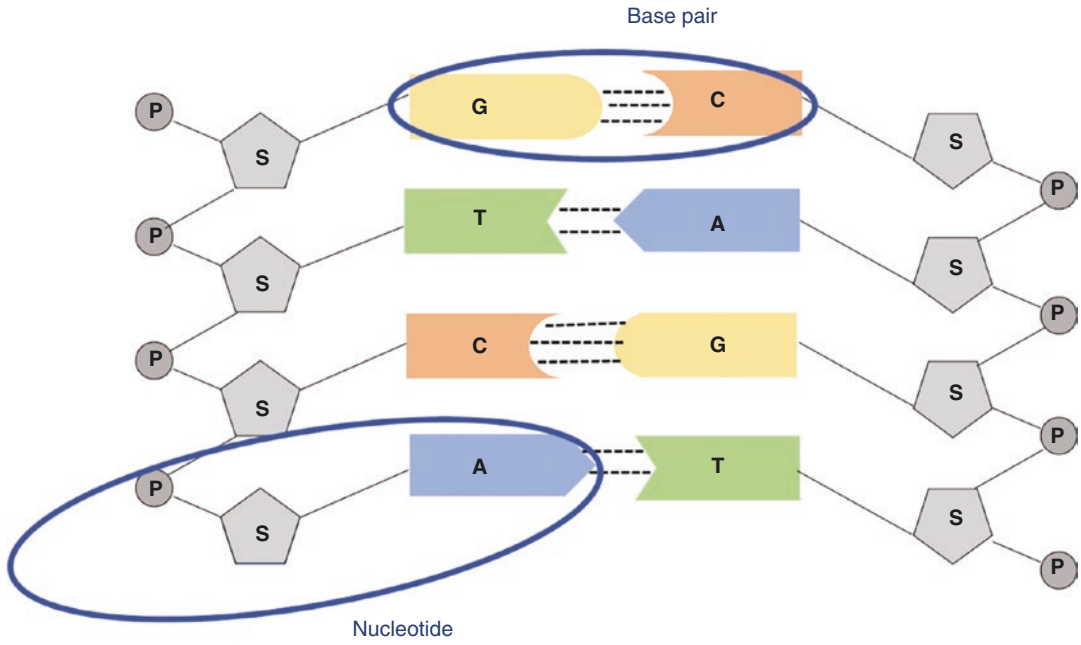

Fig. 1.2 DNA is composed of nucleotides, each of which includes a phosphate and sugar backbone and a nitrogenous base (A, C, G, or T). The bases form bonds with complementary bases on the opposite strand of DNA. The structure is twisted into a double helix

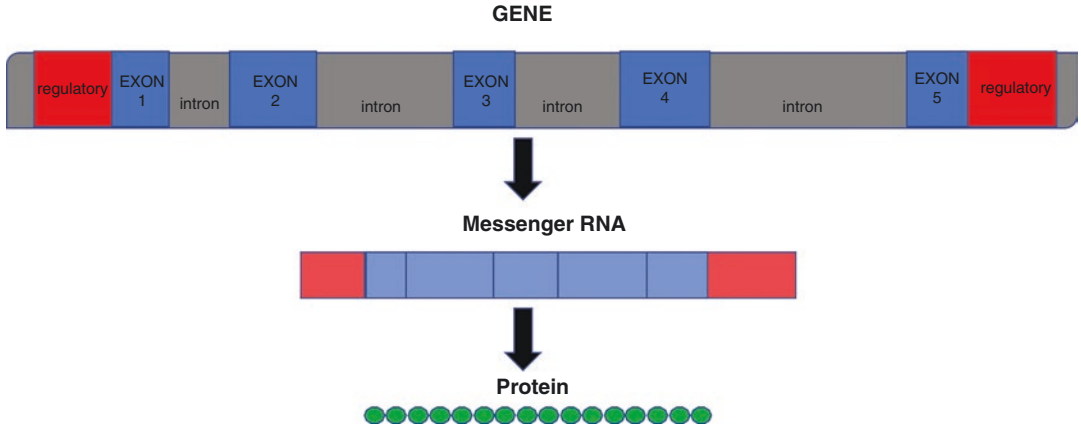

Fig. 1.3 Genes are composed of exons (protein coding regions), introns, and regulatory elements (nonprotein coding regions)

Similar to the structure of DNA, RNA is made of nucleotides composed of a phosphate and sugar (ribose) backbone, each attached to a nitrogenous base: adenine (A), cytosine (C), guanine (G), and uracil (U). The messenger RNA strand is able to leave the nucleus and attach to a ribosome where it is used as the template to assemble a protein. RNA reads three bases at a time, in a unit called a codon. Each codon corresponds to a specific amino acid. Most amino acids, however, can be encoded by more than one codon. This is described as a redundant or degenerate code. Note that there are also three termination or "stop" codons, which do not correspond to an

amino acid but rather signal the end of a coding region (Table 1.1). The process by which RNA is used to assemble a specific string of amino acids into a protein is called translation (Fig. 1.4).

The order and chemical properties of amino acids within a protein (as dictated by the nucleotide sequence of the gene) determines the final shape and function of the protein. Envision a three-dimensional structure in which hydrophobic amino acids pull their part of the chain toward the center of the structure to avoid contact with water. Hydrophilic amino acids move outward in search of water. In doing so, a hydrophobic central core is created (Fig. 1.5). Positively charged amino acids then seek negatively charged amino acids, covalent bonds are formed, and accommodations are made for variances in the amino acid shapes and sizes. Subject to these intermolecular forces, the linear amino acid chain folds into a compact structure.

The folding process is complex in the crowded cellular environment; hence, molecular chaperones assist. Chaperones are specialized molecules with cell housekeeping duties. They interact with a newly synthesized unfolded or partially folded amino acid chain and promote folding and stabilization of the protein structure. The resultant, folded structure is genetically designed to serve a specific function in the cell. The final protein may work autonomously in the cell, or it may join with other proteins to form a functional unit.

Table 1.1 Amino acid abbreviations and associated DNA codons

GENETIC CODE			
Amino acid (AA)	AA ABBREVIATIONS		DNA CODONS
	3-letter	1-letter	
Alanine	Ala	A	GCT, GCC, GCA, GCG
Arginine	Arg	R	CGT, CGC, CGA, CGG, AGA, AGG
Asparagine	Asn	N	AAT, AAC
Aspartic acid	Asp	D	GAT, GAC
Cysteine	Cys	C	TGT, TGC
Glutamine	Gln	Q	CAA, CAG
Glutamate	Glu	E	GAA, GAG
Glycine	Gly	G	GGT, GGC, GGA, GGG
Histidine	His	H	CAT, CAC
Isoleucine	Ile	I	ATT, ATC, ATA
Leucine	Leu	L	CTT, CTC, CTA, CTG, TTA, TTG
Lysine	Lys	K	AAA, AAG
Methionine	Met	M	ATG
Phenylalanine	Phe	F	TTT, TTC
Proline	Pro	P	CCT, CCC, CCA, CCG
Serine	Ser	S	TCT, TCC, TCA, TCG, AGT, AGC
Threonine	Thr	T	ACT, ACC, ACA, ACG
Tryptophan	Trp	W	TGG
Tyrosine	Tyr	Y	TAT, TAC
Valine	Val	V	GTT, GTC, GTA, GTG
Termination (stop) codons			TAA, TAG, TGA

1.3 Genetic Variants

A variant (or mutation) is a permanent, heritable change in the nucleotide sequence of a gene. Variants can be classified based on the effect they have on the structure of the gene, the sequence of the amino acids, or the ultimate protein.

1.3.1 Variant Effects on Gene Structure

Some variants affect only one nucleotide in the DNA sequence of a gene. Whether that nucleotide is deleted or substituted for a different nucleotide, or if there is an insertion of an additional nucleotide, this single nucleotide change can be referred to as a point mutation. Other variants may involve several nucleotides or perhaps even an entire gene. In some cases, there can be deletion or translocation of multiple neighboring genes. Such variants are referred to as large-scale variants.

A substitution occurs when one or more nucleotides in the gene are mistakenly replaced with others (Table 1.2). Depending on the loca-

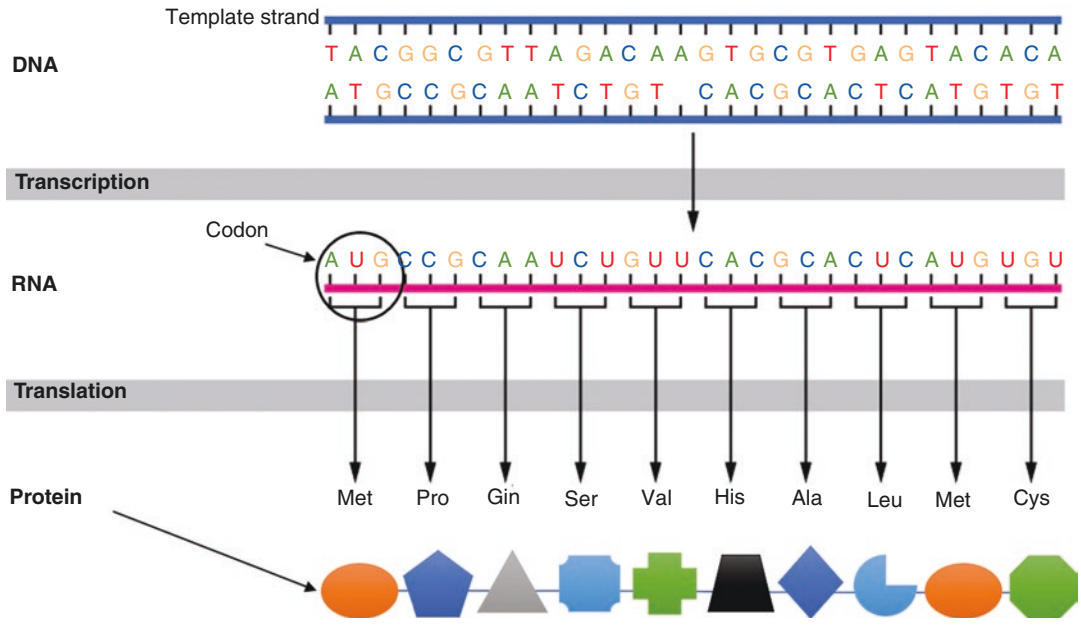

Fig. 1.4 Genes are transcribed into messenger RNA and translated into amino acids, which combine to form proteins

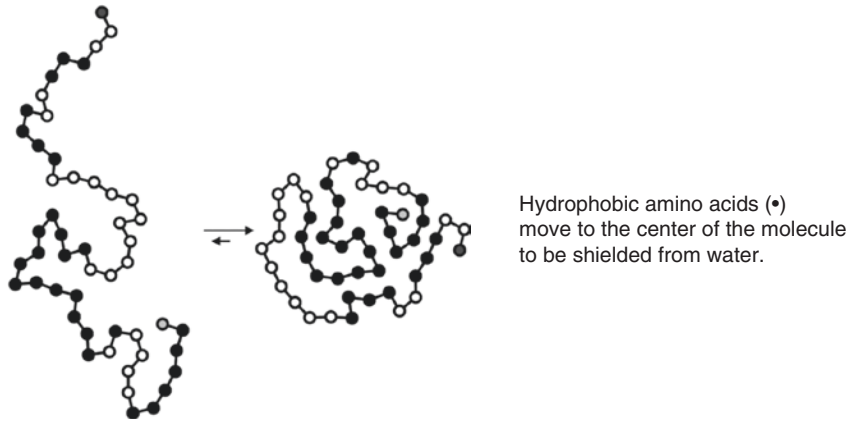

Fig. 1.5 The linear chain of amino acids (left side) assumes a compact, folded shape (right side) as hydrophobic amino acids (dark dots) gravitate toward the center core, promoting protein folding

tion of the substitution and how many nucleotides are involved, this may affect one or more codons and thus one or more amino acids in the protein sequence.

An insertion involves the addition of one or more extra nucleotides that do not belong in the typical DNA sequence of the gene. Recall that DNA is translated to RNA and then transcribed one codon (3 nucleotides) at a time. An insertion, unless it happens to occur between established codons and involve a number of nucleotides divisible by three, is likely to disrupt the codon reading frame. This variant leads to an unintended sequence of amino acids from that point on in the newly synthesized protein. The result is referred to as a frameshift variant. When the codon reading frame is shifted, it is also likely that one of the newly created codons will be a stop codon, creating a premature truncation of the protein.

Table 1.2 Variant effects on the structure of a gene

Normal DNA reference sequence	TAC GGC GTT AAA CAA GTG CGT ACG TAC ACA
Substitution	TAC AGC GTT AAA CAA GTG CGT ACG TAC ACA
Insertion Leading to frameshift	TAC GGC GTT AAA CAT AGT GCG TAC GTA CAC A
Deletion Of 3 nucleotides-no change in reading frame	TAC GGC GTT AAA CAA GTG ~~CGT~~ ACG TAC ACA

Table 1.3 Variant effects on the sequence of amino acids in a protein

Normal DNA reference sequence Messenger RNA Protein	TAC GGC GTT AAA CAA GTG CGT ACG TAC ACA AUG CCG CAA UUU GUU CAC GCA UCG AUG UGU Met Pro Gln Lys Val His Ala Ser Met Cys
substitution leading to **silent variant** The amino acid is not changed	TAC GGT GTA AAA CAA GTG CGT ACG TAC ACA AUG CCA CAA UUU GUU CAC GCA UCG AUG UGU Met Pro Gln Lys Val His Ala Ser Met Cys
substitution leading to **neutral variant** Lysine is replaced with arginine which is chemically similar	TAC GGC GTT AGA CCA AGT GCG ACG GTA CAC A AUG CCG CAA UCU GUU CAC GCA UCG AUG UGU Met Pro Gln Arg Val His Ala Ser Met Cys
substitution leading to **missense variant** Alanine is replaced with threonine which has very different chemical properties	TAC GGC GTT AAA CAA GTG TGT ACG TAC ACA AUG CCG CAA UUU GUU CAC ACA UCG AUG UGU Met Pro Gln Lys Val His Thr Ser Met Cys
substitution leading to **nonsense variant**	TAC GGC GTT AAA CAA GTG CGT ATC TAC ACA AUG CCG CAA UUU GUU CAC GCA UAG Met Pro Gln Lys Val His Ala STOP
deletion leading to **frameshift and premature protein truncation**	TAC G~~GC~~ GT TAA ACA AGT GCG TGA GTA CAC A AUG CCA AUU UGU UCA CGC ACU CAU GUG U Met Gly STOP

Deletions occur when one or more nucleotides are removed from the gene. As with insertions, deletions can disrupt the codon reading sequence of the gene, resulting in frameshift and/or protein truncation.

1.3.2 Variant Effects on Amino Acid Sequence

As discussed above, changing one or more nucleotides can lead to a change in the corresponding sequence of amino acids. Genetic variants can also be classified based on the functional effect of these changes (Table 1.3).

The most benign change is referred to as a silent variant. Silent variants occur when there is a change in the DNA, which does not lead to a change in the resulting codon or amino acid. For instance, a substitution that changes the sequence "GCT" to "GCC" results in a different codon, with the same reading frame, which also codes for alanine, the same amino acid encoded by the original sequence. Silent variants have no effect on the final protein.

Neutral variants involve a change in the DNA sequence that leads to a different but chemically similar amino acid at the designated place in the protein. For instance, changing an arginine to a lysine in the amino acid sequence will have little effect on the protein since both of these amino acids share similar chemical properties.

A missense variant changes the DNA in a way that leads to one or more amino acids with very different chemical properties from the intended sequence. For instance, a substitution that changes the sequence "CGT" to TGT" changes the corresponding amino acid from alanine to threonine. In this case, the intended amino acid, which is hydrophobic, is changed to a hydrophilic molecule, which could lead to changes in the conformation and thus the function of the ultimate protein. Of note, missense variants are typically most amenable to chemical chaperone therapy, which can help to restore a more normal protein conformation.

Variants that change the DNA, leading to a premature stop codon and, therefore, a truncated protein, are called nonsense variants. If a nonsense variant occurs toward the beginning of a gene, there may not be any protein made at all. If it occurs toward the end of the gene, there may or may not be a significant effect on the function of the truncated protein produced.

1.3.3 Variant Effects on the Protein

A variant's effect on health and well-being is dependent upon the gene involved and the effect of the variant on the protein it encodes. Possible effects include loss of protein function, gain of protein function, or no effect.

Variants that result in loss of function of the intended protein are the most common cause of inherited metabolic diseases. Loss of function may be due to an alteration of DNA sequences critical to the protein's activity or function, such as the catalytic properties of an enzyme. Loss of function may also be due to variants that drastically decrease the abundance of the protein in the cell. This includes variants that alter DNA sequences, which are critical to protein folding. Improper or misfolded proteins are unstable and may be flagged for destruction in the cell. Variants that result in the loss of protein expression, RNA degradation, or changes in the localization and targeting of the protein in the cell are other causes of decreased protein abundance.

Some variants alter the gene in such a way that the ultimate protein created takes on unintended functions. It may also cause the protein to be expressed in unintended cell types. This type of variant is referred to as a gain of function. Gain-of-function variants are typically associated with dominantly inherited diseases.

Variants that fully abolish the function of a protein are referred to as null variants. Null variants generally result in severe clinical disease, whereas variants that reduce, but do not abolish, protein function result in relatively less severe disease.

1.4 Variant Nomenclature

Variant nomenclature has evolved as we have learned more about the human genome and standard methods of reporting mutations have been developed [3, 4]. Describing a specific variant begins with identifying the reference sequence being used. A variant beginning with "c." refers to a change in coding DNA. Likewise, a variant beginning with "p." describes a change in the protein sequence. Other reference sequences include "g." for genomic DNA, "n." for noncoding DNA, "m." for mitochondrial DNA, and "r." for RNA.

For variants in coding DNA, "c." is followed by notation of which nucleotide(s) in the coding sequence are involved and what specific change occurred as compared to the typical or "wild type" reference sequence. The following are examples of basic variants in the *PAH* gene known to be associated with phenylketonuria (PKU) (Box 1.1):

Box 1.1: Variants in Coding DNA

Type of variant	Examples	Description
Substitution	c.1222C > T	At nucleotide number 1222, a cytosine has been replaced with a thymine.
Deletion	c.184delC	At nucleotide number 184, a cytosine has been deleted.
	c.163_165delTTT	Including nucleotides number 163, 164, and 165, there are three thymines deleted.
Duplication	c.111dupG	At nucleotide number 111, there are two guanine nucleotides (instead of one).
	c.185_188dupTGAC	Including nucleotides number 185 through 188, there is a duplication of 4 nucleotides: thymine, guanine, adenine, and cytosine.
Insertion	c.266_267insG	Between nucleotides 266 and 267, a guanine nucleotide was inserted.
	c.43_44insAG	Between nucleotides 43 and 44, an adenine and a guanine were inserted.

Source: PAHvdb http://www.biopku.org [5]

Box 1.2: Variants in the Amino Acid Sequence of a Protein

Change in gene	Resulting change in protein	Description
c.1222C > T	p. Arg408Trp	At amino acid number 408 in the protein, an arginine is changed to a tryptophan.
	p.R408W	
c.184delC	p.Leu62Ter	At amino acid number 62 in the protein, a leucine has been replaced with a termination (stop) codon.
	p.L62* or p.L62X	
c.163_165delTTT	p.Phe55del	At amino acid number 55 in the protein, a phenylalanine is deleted from the sequence.
c.111dupG	p.Ile38AspfsTer19	At amino acid number 38, an isoleucine was changed to an aspartate, the length of the frameshift is 19 amino acids including the termination (stop) codon.

Variants in the amino acid sequence of a protein are denoted as "p." followed by a description of the amino acid change. It is typically expressed using three letter abbreviations for the amino acids; however, former nomenclature used single letter abbreviations (Table 1.1). These examples correlate with some of the changes in coding DNA listed above (Box 1.2).

1.5 Genetic Testing

1.5.1 Genetic Testing Technologies

The origin of genetic testing dates back to the 1950s, when it was established that humans typically have 46 chromosomes in every body cell and it became possible to stain and count chromosomes in leukocyte cultures [6, 7]. By stopping cell division when the long strands of DNA are becoming most compacted into chromosomes and then applying special stains, it is possible to count the number and type of chromosomes in a cell. Areas of the DNA that are rich in adenine and thymine nucleotides stain differently than those with more cytosine and guanine nucleotides, creating "banding patterns" or darker stripes of stain along the chromosomes. For analysis, cytogeneticists line up the chromosomes based on size and match up the chromosome pairs based on banding patterns; the result is called a karyotype (Fig. 1.6). Initially, karyotypes

were limited to testing for an abnormal number of chromosomes. As the resolution of chromosome analysis improved, it also became possible to detect deletions, additions, or rearrangements of relatively large segments of DNA within a chromosome or exchanged between two or more chromosomes.

In the 1980s, molecular cytogenetic technologies were developed. By using the same culture techniques as karyotyping but employing a specific nucleotide sequence designed to attach to a targeted region of a specific chromosome, fluorescent in situ hybridization (FISH) is able to identify the presence or absence and the location of that specific sequence of DNA [8]. This enables testing for specific regions of a chromosome too small to be visible by karyotype.

In 1992, comparative genomic hybridization (CGH) was first reported [9]. CGH involves combining DNA from one cell line (blood, tumor, etc.) with DNA from a normal or "wild type" reference sequence and attaching molecular labels or tags. The two sources of DNA are chemically induced to hybridize to each other where their DNA sequences match up. The molecular labels are fluorescent and can be measured at specific positions in the genome. By comparing fluorescence to a standard, CGH provides information on the relative copy number of sequences in the test cell line. By analyzing copy number variations (CNVs), it is possible to deduce whether there are small sequences of DNA that are missing or duplicated. Different versions of this technology have advanced to the chromosome microarray analysis (CMA) widely used today to detect microdeletions and microduplications, which may involve one or multiple genes.

The terms, "molecular testing" and "DNA testing," are often used interchangeably. Both

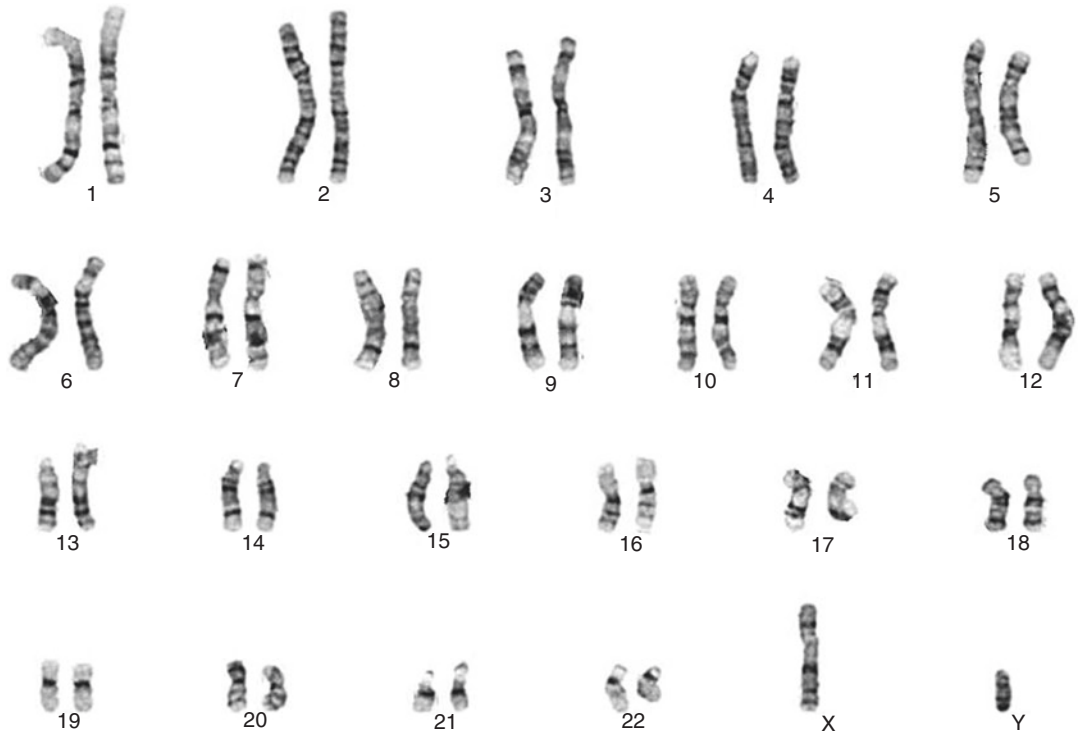

Fig. 1.6 A standard male karyotype with 46 total chromosomes including 22 autosomes and two sex chromosomes (XY)

refer to testing techniques that allow for identification of nucleotides in a gene. This allows for detection of DNA sequence variation due to substitutions, insertions, and deletions. A variety of molecular testing approaches and methods exist. For the purposes of this chapter, we will focus on two: Sanger sequencing and next-generation sequencing (NGS).

Sanger sequencing (developed by Frederic Sanger and colleagues in the 1970s) has long been a gold standard in DNA testing [10, 11]. It begins with isolating a specific gene and then using the DNA strand as a template to generate multiple copies through the use of polymerase chain reaction (PCR). As the name implies, sequencing involves reading nucleotide by nucleotide through the DNA sequence of a gene in an effort to identify changes from a normal or "wild type" reference sequence. Sanger sequencing is accurate and well suited for targeted molecular studies. Targeted variant analysis is used when looking for the presence of a specific variant in a gene. For instance, if there is a specific variant that was previously identified in another family member.

Next-generation sequencing (NGS) is a general term that involves several different technologies. NGS begins with fragmenting and amplifying millions of different DNA sequences at the same time. In this way, it can create multiple copies of hundreds to thousands of genes simultaneously. The millions of DNA fragments are then compared to a normal or "wild-type" reference sequence. Using an NGS platform allows for more automation and higher throughput. Accuracy of variant detection is also significantly improved owing to repeated sampling of the same DNA sequences.

NGS serves as the base for gene panel testing. Gene panels are a select group of genes that may be involved in a common differential diagnosis. Panel tests may be based on a specific disorder; for instance, a maple syrup urine disease (MSUD) panel that includes the different genes involved in encoding each subunit of the enzyme deficient in MSUD. Panel tests can also be built around a set of biochemical and/or clinical symptoms; for example, a panel of genes associated with ketotic hypoglycemia or developmental delay. A single panel can test for hundreds of genes, thereby eliminating the need to use Sanger sequencing to test one gene at a time.

NGS platforms also allow for broader scale testing including exome and genome sequencing. Whole exome sequencing (WES) allows for the analysis of the exons of most every gene. Collectively referred to as the exome, this is inclusive of the coding regions of most known genes. Although the exome represents <2% of the human genome, it contains an estimated 85% of known disease-causing variants [12]. Whole genome sequencing (WGS), by contrast, includes not only the exons, but also the introns and regulatory elements included in noncoding DNA. Since some disease-causing variants can occur in noncoding regions of a gene, WGS is the most comprehensive but is also the most complex to interpret. To aid in interpretation of possible variants, most WES and WGS analyses refer to DNA samples from both biological parents of an affected individual; thus, testing is often ordered as a "trio."

Importantly, both Sanger sequencing and NGS have a limited ability to detect relatively large deletions or duplications within a gene. They also cannot distinguish between two copies of the same variant (one on the maternal gene copy, one on the paternal gene copy) versus one copy of a variant paired with a deletion in that same area of the other gene copy. If a disease-causing variant is suspected in a specific gene but is not identified by sequencing, additional testing may be indicated. Microarray-based testing, multiplex ligation–dependent probe amplification (MLPA) analysis, and quantitative PCR (qPCR) analysis are all potential options to identify these variants (Fig. 1.7).

1.5.2 Interpretation of Genetic Testing

Dependent upon the testing method(s) used, there are four main categories of possible results:

Positive results are those that identify at least one specific genetic variant that accounts for all

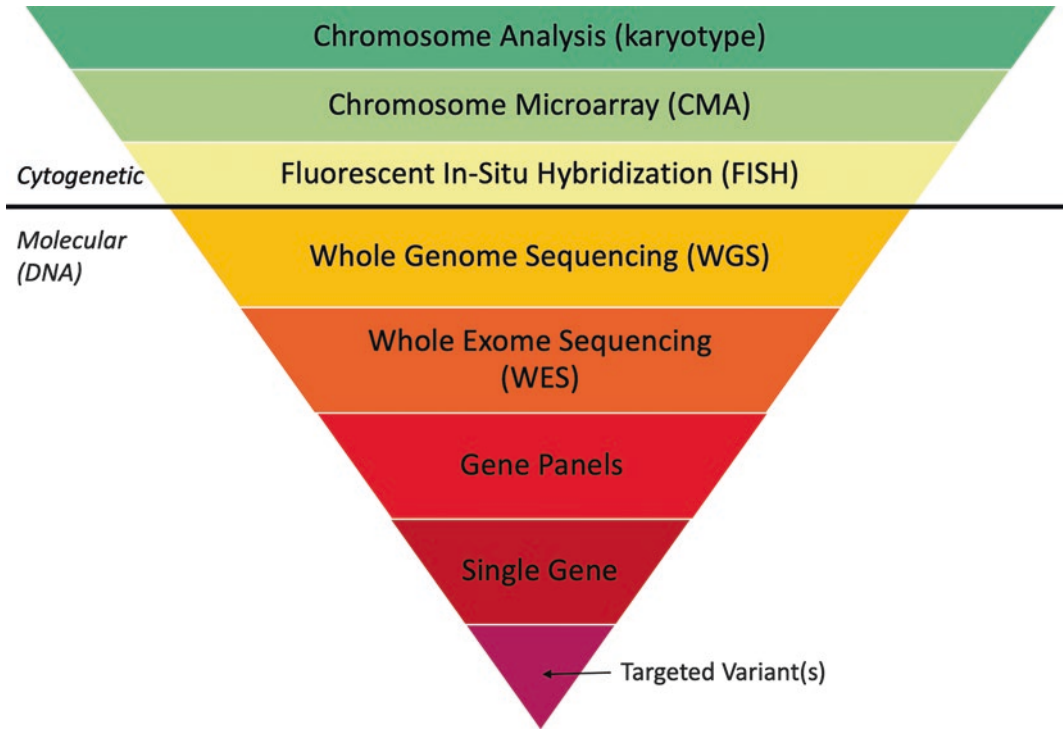

Fig. 1.7 Genetic testing methods organized from least (top) to most (bottom) specific

or part of the clinical and biochemical symptoms of the individual. Positive results involve at least one variant determined to be capable of causing disease, also referred to as a pathogenic variant. The specific variant may already be known in the scientific community and reported in association with clinical disease or, based on the type and location of the variant within the gene, it may be predicted to be disease-causing or likely disease-causing.

Negative results refer to testing that does not identify a disease-causing variant. Negative results do not necessarily rule out the possibility of the disease in question. It is possible that additional testing technologies may be required to identify the variant. It is also possible to have a gene that, in sequence, appears to be normal but is affected by a secondary genetic or epigenetic mechanism that leads to malfunction.

Genetic testing may also identify one or more variants of uncertain significance (VUS). A VUS is a variant about which we have no, limited, or conflicting information. Many DNA variants, whether inherited or unique to a given individual, may just be part of normal, benign variation and not associated with disease. Other variants may have disease-causing potential; however, the specific variant has never been reported before or has not been subject to functional studies to help clarify the potential significance. Some variants have been reported in both healthy individuals and in individuals with clinical symptoms and are therefore more difficult to interpret. In the case of a VUS, it is important to consider the clinical context of the individual. As our knowledge of genes and human disease continues to expand, VUSs are often reclassified as benign, likely benign, pathogenic, or likely pathogenic.

A fourth category of results is important to consider, especially when using more broad testing methods that can uncover unexpected results. In some cases, genetic testing can reveal that biological relationships are not as reported. For instance, it can identify misattributed paternity or consanguinity (a common ancestor as in siblings or cousins). Chromosome analysis and CMA may

identify a difference between the biological sex and phenotypic or apparent sex of an individual. Since panel testing, WES, and WGS analyze more than one gene, it is possible to uncover positive results in a gene that was not part of the primary reason for testing. This could include identification of carrier status for, susceptibility to, or presence of another disorder. For instance, exome analysis completed for the purpose of identifying a cause of developmental delay may also identify that the individual is a carrier for phenylketonuria (PKU) or at increased risk to develop hereditary cancer or cardiovascular disease. These results are also referred to as secondary findings. The identification of a secondary finding in the individual being tested can also inform the risk for other family members to have the same condition or predisposition. Complex results and unsought information can raise moral and ethical challenges [13–16]. Clear expectations regarding the benefits and limitations of testing, what information could be identified, and what information will or will not be shared should be established prior to testing [17]. Therefore, an informed consent process is generally required, and genetic counseling is strongly recommended.

1.5.3 Purposes of Genetic Testing

Genetic testing may be helpful in a number of different circumstances. It can establish a genetic diagnosis in an individual, an in vitro embryo, or a pregnancy. A known diagnosis, in turn, can allow for more comprehensive management, monitoring for other anticipated findings, and potentially targeted treatments. Genetic testing can determine an inheritance pattern and allow for identification of other family members who may be at risk of developing the condition. It can determine carrier status for a specific variant or disorder, or for a broad panel of disorders. It can identify individuals who are clinically asymptomatic or presymptomatic who are at increased risk to develop a later-onset condition. Even if an affected individual already has a clinical and/or biochemical diagnosis, knowledge of the specific genetic variant(s) may lead to additional information about the expected clinical course (genotype/phenotype correlation described in Sect. 1.6) or may open doors to therapies based on a specific variant or type of variant. For example, some variants are responsive to chemical chaperone therapy such as the use of sapropterin dihydrochloride (Kuvan®) in patients with PKU who have certain missense variants [18–20]. Stop codon read-through therapy is another example currently in development, which allows cells to ignore premature termination codons induced by nonsense variants in order to restore more normal protein production. Other investigational treatments and gene therapies hold promise for the future.

1.6 Genotype and Phenotype

DNA testing determines the genotype, or genetic constitution, of an individual. The genotype is typically expressed with molecular nomenclature noting the specific variant(s). Since most genes occur in pairs (with one copy inherited from the mother and the other from the father), there are typically two versions of each gene in an individual. Each unique version of a gene is referred to as an allele. If the alleles are identical, the person is described as being homozygous at that genetic location. If the two alleles are different, they are described as being heterozygous at that location. One can be heterozygous with one variant allele and one "normal" allele. Alternatively, one can be compound heterozygous with a variant on one allele and a *different* variant on the other allele (Box 1.3).

Box 1.3: Examples of Homozygous and Heterozygous States

State	*Galactose-1-phosphate uridyltransferase (GALT) gene*
Homozygous	
Not affected	Normal/Normal
Affected	p.Q188R/p.Q188R
Heterozygous	
Carrier, not affected	Normal /p.Q188R
Compound heterozygous, affected	p.Q188R/ p.H319Q

In contrast to an individual's genotype, their phenotype refers to their observable physical and biochemical characteristics. If an individual's genotype, as noted by specific variant(s) of a certain gene, predicts a specific phenotypic outcome, there is said to be a genotype/phenotype correlation. For example, an individual with a genotype of p.Q188R/p.Q188R in the *GALT* gene is expected to have classic galactosemia. Some genotypes can predict responsiveness to various cofactors or therapies or may predict more mild or more severe clinical courses. This information, if established, can help to suggest outcome and drive management of the disease.

The term phenocopy refers to an environmentally induced variation that closely resembles a genetically determined variation. For example, dietary vitamin B_{12} deficiency is a phenocopy of the inherited disease, methylmalonic acidemia and homocystinuria, due to cobalamin C disease. Both dietary vitamin B_{12} deficiency and cobalamin C disease have the same biochemical findings of elevated plasma methylmalonic acid and homocysteine. Awareness of phenocopies is important as they can provide an alternative explanation for clinical findings. For example, in the case of elevated methylmalonic acid and elevated homocystinuria, one may want to exclude maternal vitamin B_{12} deficiency as a possible cause of these abnormal labs prior to testing for cobalamin C disease or other possible metabolic etiologies.

1.7 Single Gene Inheritance Patterns and Pedigrees

There are three main patterns of single gene inheritance (also called Mendelian inheritance, discovered by Gregor Mendel in the late 1800s): autosomal recessive, autosomal dominant, and X-linked. In autosomal-dominant and autosomal-recessive inheritance, the gene responsible for the disorder is located on one of the 22 autosomal chromosomes. For X-linked conditions, the gene involved is located on the X chromosome. Y-linked, polygenic, and mitochondrial inheritance patterns exist as well but will not be addressed in this chapter.

1.7.1 Single Gene Inheritance Patterns

In order for an individual to be affected with an autosomal-recessive disorder, he or she must have inherited a disease-causing variant from both parents. Although the variants are recessive, an affected individual has no normal copy of the gene. Parents of an affected individual are considered to be obligate carriers of the disorder, having one variant gene copy and one presumably normal gene copy. Since the variant is recessive, carriers are able to use their normal gene copy as a template to make sufficient amounts of the encoded protein so that they are not affected with the disorder. For two carriers of an autosomal-recessive disorder, there is a 1 in 4 (25%) chance that *each* pregnancy they conceive will be affected with the disorder (Fig. 1.8). Children of two carriers, who are not affected themselves, have a 2 in 3 (66%) chance to also be carriers. In the general population (individuals without a family history of the disorder in question), the carrier frequency or chance that an individual would happen to be a carrier for a given disorder varies depending on the specific disorder, gene, and sometimes ethnic population. Autosomal-recessive inheritance is the most common inheritance pattern seen in inborn errors of metabolism.

For an individual to be affected with an autosomal-dominant disorder, he or she must inherit one disease-causing variant. Since the variant is dominant, it will cause disease despite having another, normal copy of the gene. Especially with dominant variants, some variants can occur spontaneously, just by chance, and are not inherited from an affected parent. Such variants are referred to as de novo variants. Whether a dominant variant is inherited or de novo, an affected parent has a 50% chance to pass it to each of his/her children (Fig. 1.9). A child who inherits the variant will be affected with the disorder.

Sometimes, two individuals with the same variant may express different clinical features, severity, or age of onset, even within the same family. This is referred to as variable expressiv-

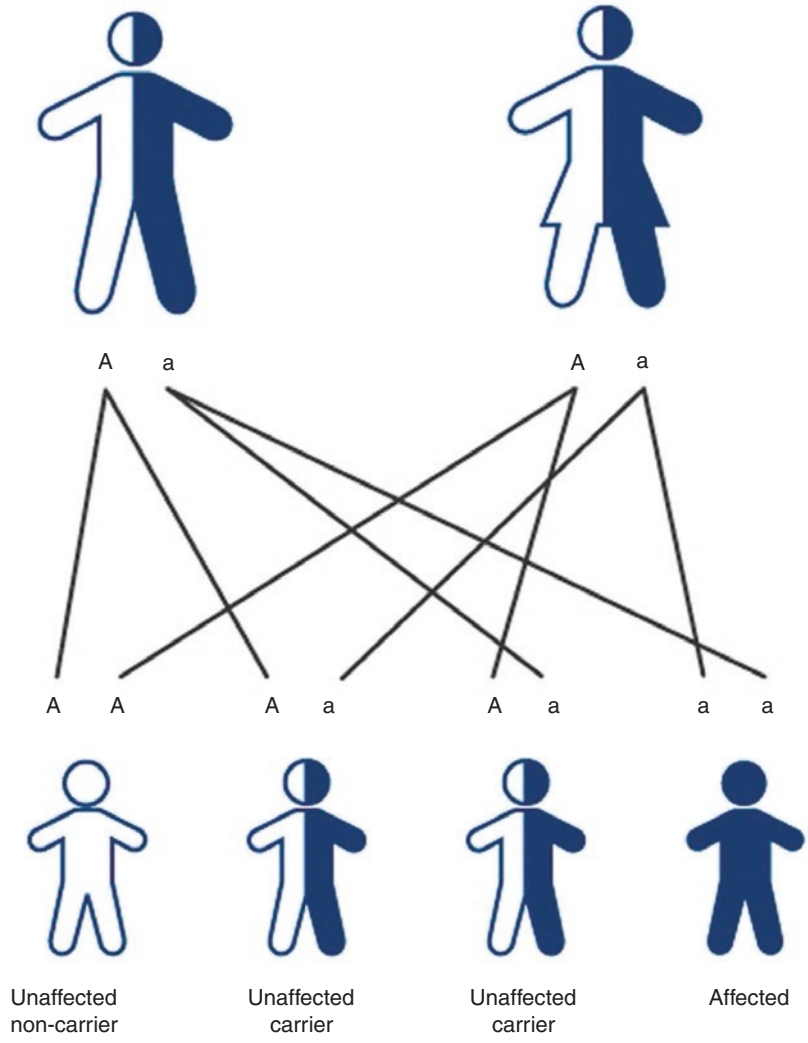

Fig. 1.8 Autosomal-recessive inheritance

ity. For some conditions, an individual with a confirmed disease-causing variant may not express any clinical symptoms at all through a phenomenon called incomplete penetrance**.** In that situation, the phenotype may appear to "skip a generation" but the genotype does not.

X-linked inheritance results from a variant that occurs in a gene located on the X chromosome. Recall that females have two X chromosomes (one inherited from their mother and the other from their father) and males have only one X chromosome (inherited from their mother) and one Y chromosome (inherited from their father). If a male inherits an X-linked variant, he will be affected with the disorder. Likewise, if he has a de novo, disease causing variant in an X-linked gene, he will also be affected. In either case, he will pass his variant to each daughter he has, but none of his sons (Fig. 1.10).

X-linked disorders in females are more complicated and typically variable in expression.

Since females have two X chromosomes, they also have two copies of all X-linked genes. As proven by males who only have one X chromosome, humans only need one copy of most of these genes. In order to account for this double dosage, one X chromosome in every female's cell is inactivated through a process called lyonization or X-inactivation. This process is typically random and occurs early in the embry-

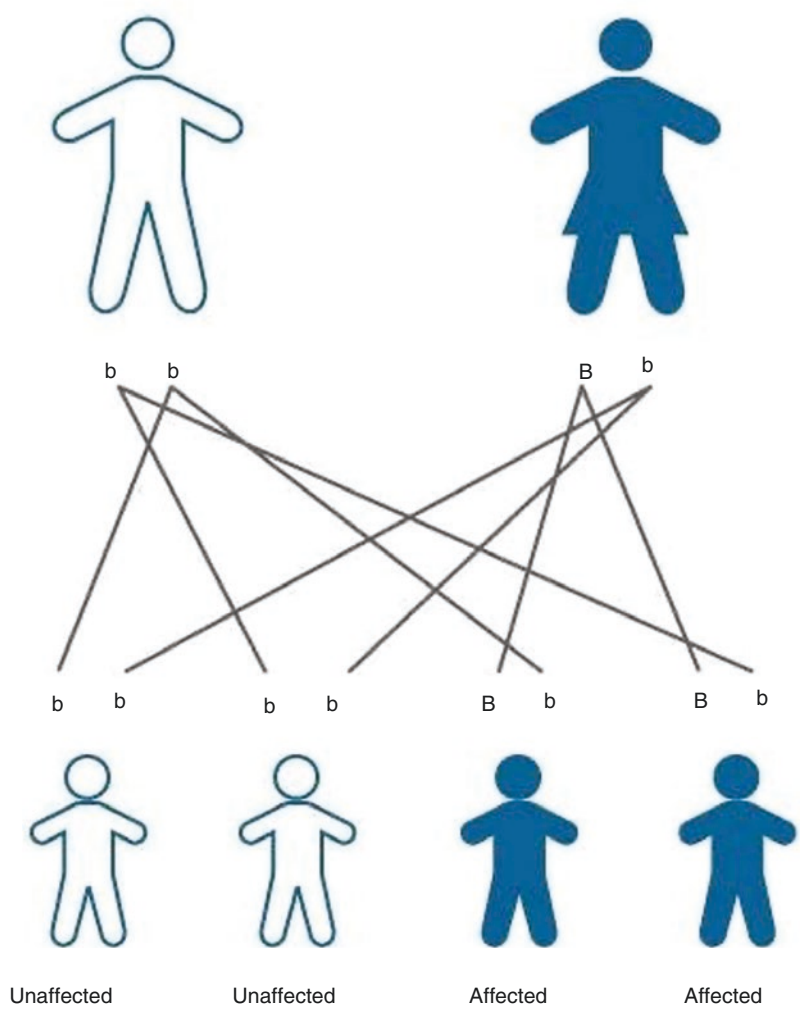

Fig. 1.9 Autosomal-dominant inheritance

onic development of a female, long before she is born. Once an X chromosome is inactivated, all new cells created from that cell line will have the same X chromosome inactivated. Inactivation does not change over time. Since this process happens early in embryonic development as different tissue types are beginning to differentiate, it is common that all cells in a given tissue will share the same X inactivation pattern. That pattern, however, can differ between tissue types. For instance, a female may have her maternally inherited X chromosome inactivated in her liver but her paternally inherited X chromosome inactivated in her brain. Occasionally, inactivation patterns may be skewed such that the vast majority of a female's cells have inactivated the same X chromosome. Skewed X-inactivation can be favorable or unfavorable if there happens to be a genetic variant on one of the X chromosomes.

If a female is affected with or is a carrier of an X-linked gene variant, she has a 50% chance to pass the variant to each child she has (Fig. 1.11). If the child who inherits the variant is a male, he will be affected with the disorder. If the child who inherits the variant is a female, her inactivation pattern will be determined after conception. Regardless of whether her mother is clinically symptomatic of the condition, a daughter who inherits the variant may be more mildly or more severely affected or, in some cases, may have no clinical symptoms at all.

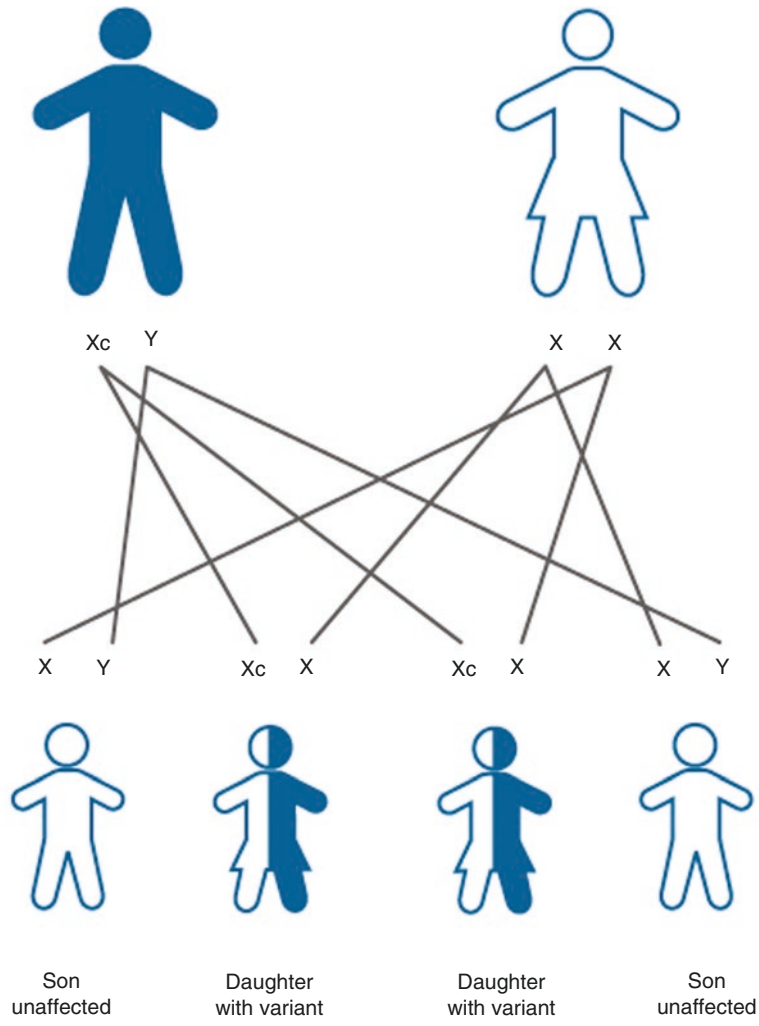

Fig. 1.10 X-linked inheritance, affected male. Note that carrier females may or may not show symptoms of the disorder depending on the disorder and their X inactivation pattern

1.7.2 Pedigrees

A genetic family history or pedigree, detailing genetic relationships and medical history of family members, can help determine the inheritance pattern of a disorder. Pedigrees can also identify individuals in the family who are at risk for developing disease or for passing on disease-causing variants in a gene. Standard symbols and terminology are used to identify individuals, relationships, and carrier or disease state (Fig. 1.12) [21]. The use of symbols allows for a concise, graphic representation of a family's genetic health history.

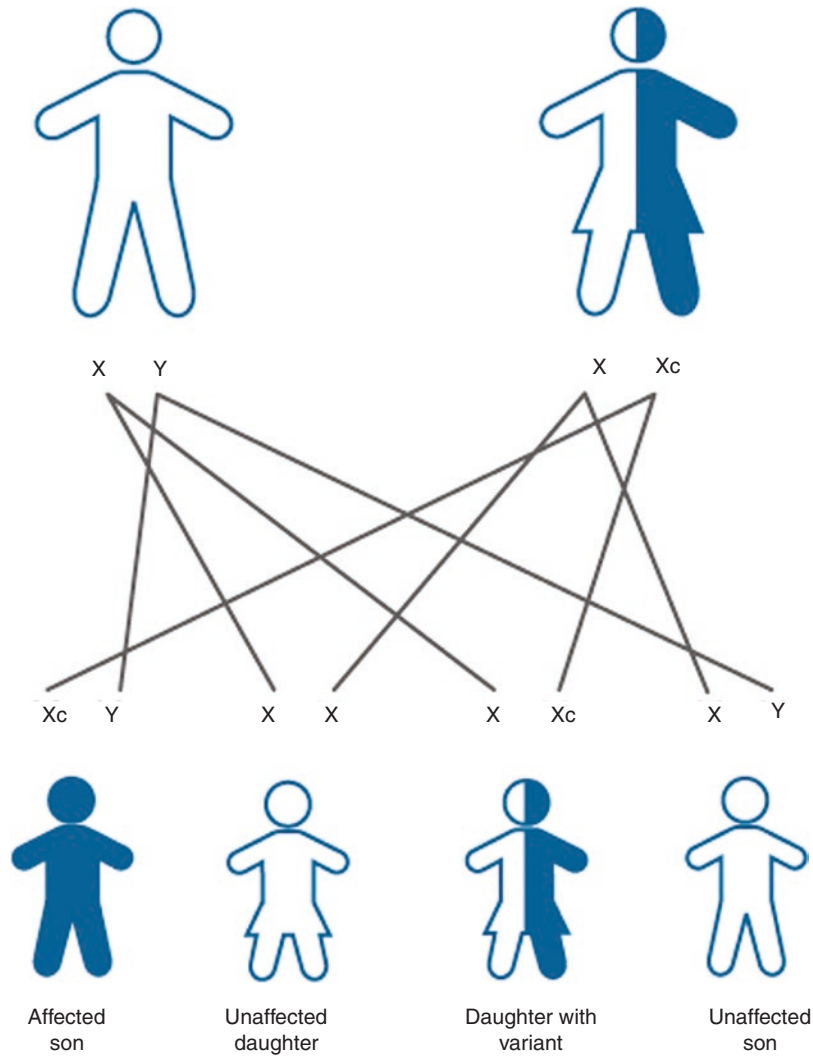

Fig. 1.11 X-linked inheritance, affected female. Due to the variability of X-linked disorders in females, a female who has the variant may or may not show symptoms of the disorder. True X-linked recessive disorders exist where carrier females are completely unaffected. X-linked dominant disorders also exist but are influenced by X inactivation patterns

The pedigree below is typical for an autosomal-recessive disorder (Fig. 1.13). Notice the carriers, all designated by the dot in the center of the square (males) or circle (females). In most cases, autosomal-recessive carrier status is not revealed until an affected individual is born into the family. Every child of an individual affected with an autosomal-recessive disorder will be an obligate carrier for the condition. They would only be at risk to be affected themselves if their other parent also happened to be a carrier of or affected with the disorder.

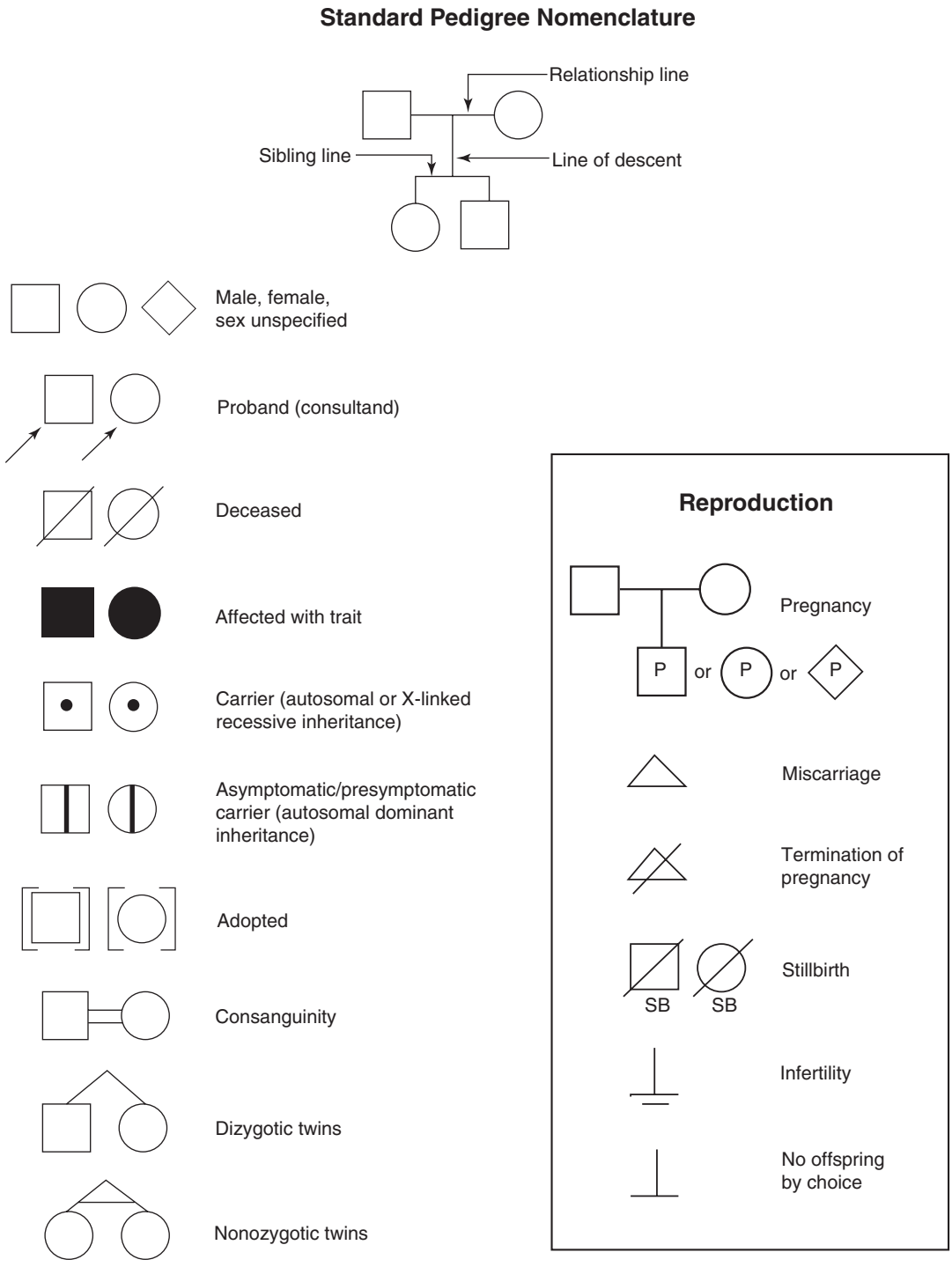

Fig. 1.12 Standard pedigree nomenclature. Common symbols are used to draw a pedigree (family tree). A pedigree shows relationships between family members and patterns of inheritance for certain traits and diseases

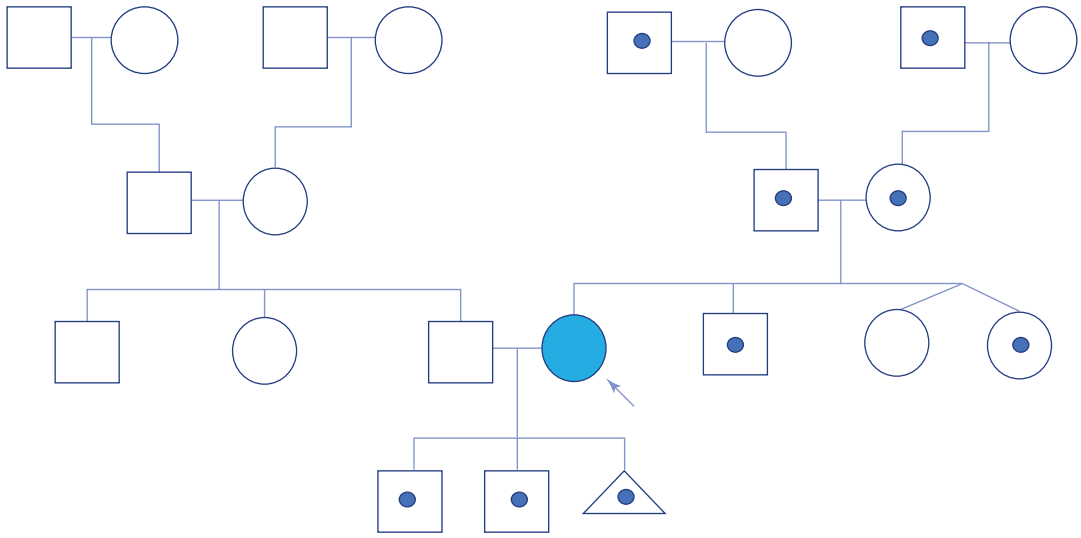

Fig. 1.13 An example of a pedigree for an autosomal-recessive disorder. The affected individual is shown as a shaded figure. Carriers are designated by a dot in the center of the figures

1.8 Summary

Our lives and well-being depend upon the function of thousands of unique proteins in each of our trillions of cells. Variants in the genes that encode these proteins disrupt cellular function and can lead to disease. The symptoms of the disease and severity are dependent upon the specific protein and the degree to which it is impacted. Genetic testing allows for the identification of the underlying genetic changes. Identifying specific variants can predict disease severity, provide risk assessment to other family members, suggest and inform on likely pathophysiology, and potentially provide strategies for intervention.

Acknowledgments Thank you to Cindy Freehauf of Children's Hospital of Colorado for her contributions to this chapter's content in the 1st Edition of Nutrition Management of Inherited Metabolic Diseases.

References

1. Bianconi E, Piovesan A, Facchin F, Beraudi A, Casadei R, Frabetti F, et al. An estimation of the number of cells in the human body. Ann Hum Biol. 2013;40(6):463–71.
2. Encode Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57–74.
3. Ogino S, Gulley ML, den Dunnen JT, Wilson RB, Association for Molecular Patholpogy T, Education C. Standard mutation nomenclature in molecular diagnostics: practical and educational challenges. J Mol Diagn. 2007;9(1):1–6.
4. den Dunnen JT, Dalgleish R, Maglott DR, Hart RK, Greenblatt MS, McGowan-Jordan J, et al. HGVS recommendations for the description of sequence variants: 2016 update. Hum Mutat. 2016;37(6):564–9.
5. PAHvdb: Phenylalanine Hydroxylase Gene Locus-Specific Database 2020. Available from: http://www.biopku.org/home/home.asp.
6. Tijo J, Levan A. The chromosome number of man. Hereditas. 2010;42(1–2):1–6.
7. Moorhead PS, Nowell PC, Mellman WJ, Battips DM, Hungerford DA. Chromosome preparations of leukocytes cultured from human peripheral blood. Exp Cell Res. 1960;20:613–6.
8. Langer-Safer PR, Levine M, Ward DC. Immunological method for mapping genes on Drosophila polytene chromosomes. Proc Natl Acad Sci U S A. 1982;79(14):4381–5.
9. Kallioniemi A, Kallioniemi OP, Sudar D, Rutovitz D, Gray JW, Waldman F, et al. Comparative genomic hybridization for molecular cytogenetic analysis of solid tumors. Science. 1992;258(5083):818–21.
10. Sanger F, Coulson AR. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. J Mol Biol. 1975;94(3):441–8.
11. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci U S A. 1977;74(12):5463–7.

12. Choi M, Scholl UI, Ji W, Liu T, Tikhonova IR, Zumbo P, et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing. Proc Natl Acad Sci U S A. 2009;106(45):19096–101.
13. Burke W, Trinidad SB, Clayton EW. Seeking genomic knowledge: the case for clinical restraint. Hastings Law J. 2013;64(6):1650–64.
14. Evans JP, Rothschild BB. Return of results: not that complicated? Genet Med. 2012;14(4):358–60.
15. Grove ME, Wolpert MN, Cho MK, Lee SS, Ormond KE. Views of genetics health professionals on the return of genomic results. J Genet Couns. 2014;23(4):531–8.
16. Yu JH, Harrell TM, Jamal SM, Tabor HK, Bamshad MJ. Attitudes of genetics professionals toward the return of incidental results from exome and whole-genome sequencing. Am J Hum Genet. 2014;95(1):77–84.
17. National Society of Genetic Counselors: NSGC Headquarters. 2013. Available from: nsgc@nsgc.org.
18. Erlandsen H, Pey AL, Gamez A, Perez B, Desviat LR, Aguado C, et al. Correction of kinetic and stability defects by tetrahydrobiopterin in phenylketonuria patients with certain phenylalanine hydroxylase mutations. Proc Natl Acad Sci U S A. 2004;101(48):16903–8.
19. Zurfluh MR, Zschocke J, Lindner M, Feillet F, Chery C, Burlina A, et al. Molecular genetics of tetrahydrobiopterin-responsive phenylalanine hydroxylase deficiency. Hum Mutat. 2008;29(1):167–75.
20. Farrugia R, Scerri CA, Montalto SA, Parascandolo R, Neville BG, Felice AE. Molecular genetics of tetrahydrobiopterin (BH4) deficiency in the Maltese population. Mol Genet Metab. 2007;90(3):277–83.
21. Bennett RL, French KS, Resta RG, Doyle DL. Standardized human pedigree nomenclature: update and assessment of the recommendations of the National Society of Genetic Counselors. J Genet Couns. 2008;17(5):424–33.