

The Relevance of Molecular Biology to Clinical Practice

Key learning points

1. The inheritance of genetic disorders in humans may be mendelian or more complex.
2. The human genome contains around 30,000–40,000 genes, and its complexity is further increased by the use of alternative promoters, epigenetic phenomena and alternative splicing.
3. Further complexity is imparted by disorders of imprinting, mitochondrial disorders, mosaicism, digenic/ oligogenic inheritance, sex-influenced phenotypes and variability of penetrance and expressivity.
4. Novel genetic techniques such as array-CGH and next-generation sequencing will revolutionize molecular diagnostics over the next few years.
5. Molecular diagnosis is increasingly important for optimal medical management of a number of human disorders.

An understanding of and a facility with the molecular basis of many endocrine and non-endocrine disorders is essential to the practice of pediatric endocrinology in the 21st century and knowledge is accumulating rapidly (see Online Mendelian Inheritance in Man [OMIM], a comprehensive catalog of human genes and genetic disorders in order to keep pace with this field: www.ncbi.nlm.nih.gov/sites/entrez?db=omim) (see also Web links box). With the exception of simple trauma, every disease has a genetic component.

In monogenic disorders, for example, congenital adrenal hyperplasia (CAH), the genetic component is the major etiological factor. Multiple genes operating in conjunction with environmental and lifestyle factors contribute to the pathogenesis of polygenic or multifactorial disorders. Genetic factors also influence the manifestation of disease directly through the genetic defect or indirectly by defining the host's susceptibility and resistance to an environmental disease such as infection.

Genetics (Fig. 1.1) is the science of heredity and variation and has heretofore focused on chromosomal abnormalities and inborn errors of metabolism. Analysis of the transmission of human traits and disease within families has led to understanding many monogenic disorders. Diabetes mellitus type 2, obesity, hypertension, heart disease, asthma and mental illnesses are complex and the genetic susceptibilities to these disorders are influenced by exogenous factors interacting with genetic susceptibilities.

Phenotype can also be influenced by genetic and environmental modifiers in monogenic disorders. For example, the expression of the phenotype in monogenic forms of diabetes mellitus due to mutations in the maturity onset of diabetes in the young (MODY) genes is influenced by factors such as diet and physical activity.

The term *genome*, introduced before the recognition that DNA is the genetic material, designates the totality of all genes on all chromosomes in the nucleus of a cell. Genomics refers to the discipline of mapping, sequencing

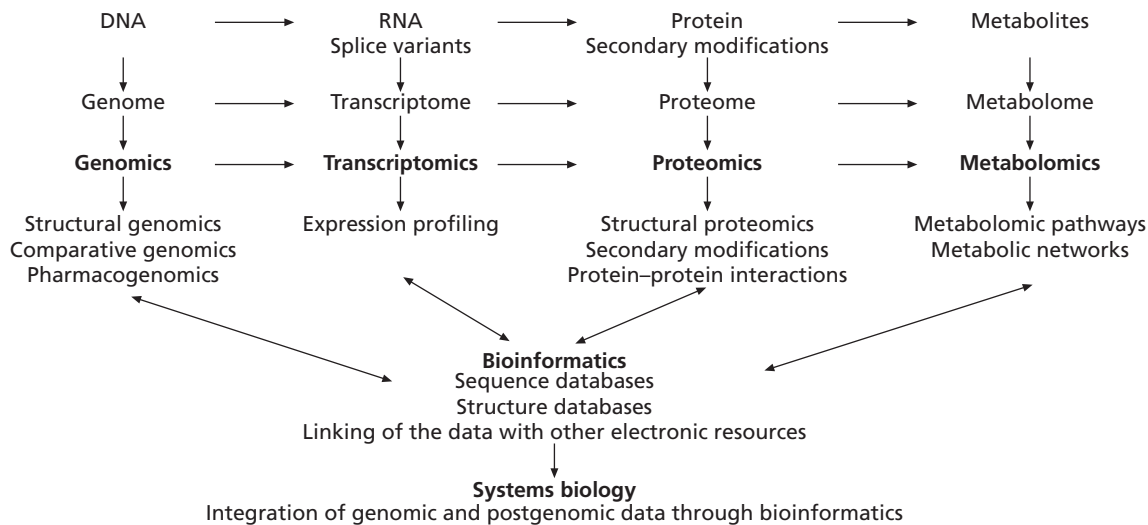


Figure 1.1 Overview of genomics, transcriptomics, proteomics, metabolomics, bioinformatics and systems biology.

and analyzing genomes. Genome analysis can be divided into structural and functional genomics. The analysis of differences among genomes of individuals of a given species is the focus of comparative genomics. The complement of messenger RNAs (mRNAs) transcribed by the cellular genome is called the transcriptome and the generation of mRNA expression profiles is referred to as transcriptomics. Epigenetic alterations and chemical modifications of DNA or chromatin proteins influence gene transcription. The sum of all epigenetic information defines the epigenome, which, in contrast to the genome, is highly variable between cells and changes within a single cell over time. Epigenetic modifications lead to phenotypic changes without alteration of DNA sequence, and may be heritable.

The term *proteome* describes all the proteins expressed and modified following expression by the entire genome in the lifetime of a cell. Proteomics refers to the study of the proteome using techniques of large-scale protein separation and identification. The field of metabolomics aims at determining the composition and alterations of the metabolome, the complement of low molecular weight molecules. The relevance of these analyses lies in the fact that proteins and metabolites function in modular networks rather than linear pathways. Hence, any physiological or pathological alteration may have many effects on the proteome and metabolome.

Pharmacogenomics, which involves the analysis of the genetic factors determining the response of an individual to a particular therapeutic agent, is emerging as a major new field. Genetic polymorphisms can not only influence the effects of medications but also result in variable absorption, distribution, metabolism, and excretion of a drug.

The growth of biological information has required computerized databases to store, organize, annotate and index the data which has led to the development of bioinformatics, the application of informatics to (molecular) biology. Computational and mathematical tools are essential for the management of nucleotide and protein sequences, the prediction and modeling of secondary and tertiary structures, the analysis of gene and protein expression and the modeling of molecular pathways, interactions and networks.

The integration of data generated by transcriptomic, proteomic, epigenomic and metabolomic analyses through informatics is an emerging discipline aimed at understanding phenotypic variations and creating comprehensive models of cellular organization and function. These efforts are based on the expectation that an understanding of the complex and dynamic changes in a biological system may provide insights into pathogenic processes and the development of novel therapeutic strategies and compounds.

The human genome

Genes were identified because they conferred specific traits transmitted from one generation to the next. They are functional units regulated by transcription and encode messenger (m)RNA which is subsequently translated into protein. Biological roles have been described for other RNAs, such as transfer (t)RNA, ribosomal (r)RNA and micro (mi)RNA. The latter are small non-coding RNAs that regulate gene expression by targeting mRNAs of protein coding genes or non-coding RNA transcripts. Micro RNAs also have an important role in developmental and physiological processes and can act as tumour suppressors or oncogenes in the ontogenesis of cancers.

The Human Genome Project led to the complete publication of the human genome DNA sequence in 2003, and revealed that 30,000–40,000 genes account for 10–15% of genomic DNA. Much of the remaining DNA consists of highly repetitive sequences, the function of which remains incompletely understood. Genes are unevenly distributed along the various chromosomes, and range in size from a few hundred base pairs to more than 2 million base pairs. The number of genes identified was surprisingly small, which suggests that the use of various promoters, alternative splicing of genes, and epigenetic phenomena account for the rich phenotypic diversity observed in humans.

Human genetics aims at understanding the role of common genetic variants in susceptibility to common disorders by identifying, cataloging and characterizing gene variants which include short repetitive sequences in regulatory or coding regions and single nucleotide polymorphisms (SNPs), changes in which a single base in the DNA differs from the usual base at that position. SNPs occur roughly every 300 base pairs and most are found outside coding regions.

Single nucleotide polymorphisms within a coding sequence can be synonymous (i.e. not altering the amino acid code) or non-synonymous. There are roughly 3 million differences between the DNA sequences of any two copies of the human genome. SNPs that are in close proximity are inherited together as blocks referred to as haplotypes, and this forms the basis of the International HapMap project. The identification of approximately 10 million SNPs that occur commonly in the human genome through the International HapMap Project is of great relevance for genome-wide association studies.

Structure and function of genes

The structure of a typical gene consists of regulatory regions followed by exons and introns and downstream untranslated regions (Fig. 1.2). Regulatory sequences controlling gene expression characteristically lie upstream (5') of the transcription start site, although a number of regulatory regions have now been identified in the introns or downstream (3') of the coding region of the gene.

The regulatory DNA sequences of a gene upstream of the coding region are referred to as the promoter region, which contains specific sequences called response elements that bind transcription factors. Some are ubiquitous. Others are cell-specific. Gene expression is controlled by additional regulatory elements, such as enhancers and locus control regions which may be located far away from the promoter region. The transcription factors that bind to the promoter and enhancer sequences provide a code for regulating transcription that is dependent on developmental state, cell type and endogenous and exogenous stimuli. Transcription factors interact with other nuclear proteins and generate large regulatory complexes that ultimately activate or repress transcription.

Transcription factors account for about 30% of all expressed genes. Mutations in them cause a number of endocrine and non-endocrine genetic disorders. Because a given set of transcription factors may be expressed in various tissues, it is not uncommon to observe a syndromic phenotype. The mechanism by which transcription factor defects cause disease often involves haploinsufficiency, a situation in which a single copy of the normal gene is incapable of providing sufficient protein production to assure normal function. Biallelic mutations in such a gene may result in a more pronounced phenotype. For example, monoallelic mutations in the transcription factor *HESX1* result in various constellations of pituitary hormone deficiencies and the phenotype is variably penetrant among family members with the same mutation. Inactivating mutations of both alleles of *HESX1* cause familial septo-optic dysplasia and/or combined pituitary hormone deficiency (de Morsier syndrome).

Gene expression can be influenced by epigenetic events, such as X-inactivation and imprinting, i.e. a marking of genes that results in monoallelic expression depending on their parental origin. In this situation,

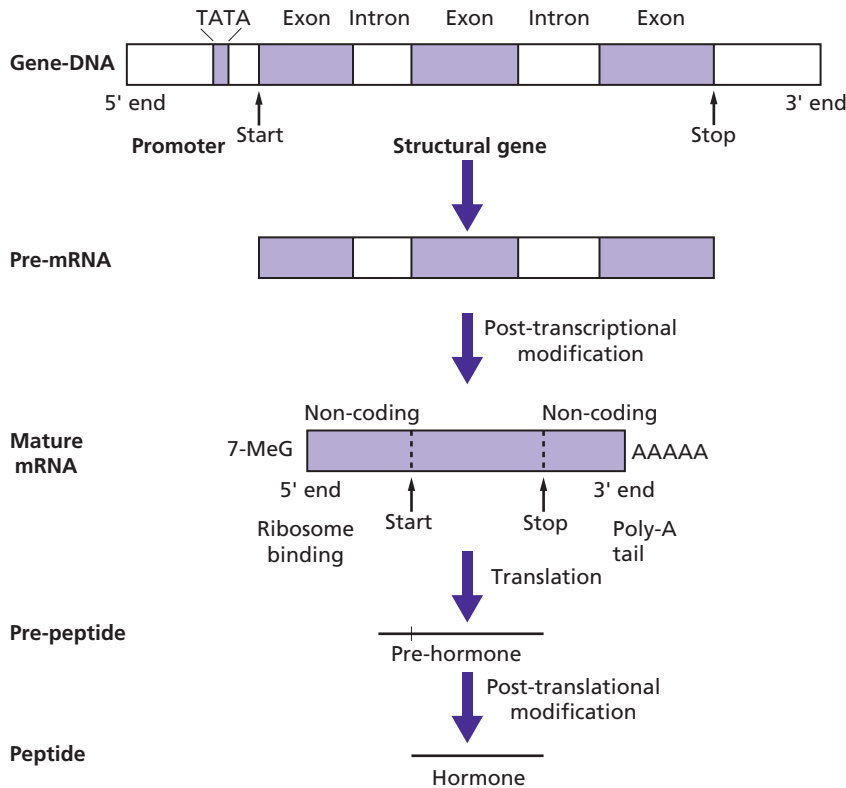


Figure 1.2 Schematic representation of a gene and of the events leading to the synthesis of a peptide hormone.

DNA methylation leads to silencing, i.e. suppression of gene expression on one of the chromosomes. Genomic imprinting has an important role in the pathogenesis of several genetic disorders such as Prader–Willi or Silver–Russell syndromes and Albright hereditary osteodystrophy.

Molecular biology in diagnosis

Analysis of chromosomes, DNA and RNA

Analyses of large alterations in the genome are possible using cytogenetics, fluorescence *in situ* hybridization (FISH), Southern blotting, high-throughput genotyping and sequencing. More discrete sequence alterations rely heavily on the use of the polymerase chain reaction (PCR), which permits rapid genetic testing and

mutational analysis with small amounts of DNA extracted from solid tissues, nucleated blood cells, leukocytes, buccal cells or hair roots. Reverse transcription PCR (RT-PCR) transcribes RNA into a complementary DNA strand, which can then be amplified by PCR. RT-PCR can be used for sequence analyses of the coding regions and to detect absent or reduced levels of mRNA expression resulting from a mutated allele.

Screening for point mutations can be performed by numerous methods, such as sequencing of DNA fragments amplified by PCR, recognition of mismatches between nucleic acid duplexes or electrophoretic separation of single- or double-stranded DNA. Most traditional diagnostic methods focus on single genes. Novel techniques for the analysis of mutations, genetic mapping and mRNA expression profiles are evolving. Chip techniques allow hybridization of DNA or RNA to hundreds of thousands of probes simultaneously.

Microarrays are being used for mutational analysis of human disease genes, for the identification of viral sequence variations and for large-scale analyses of mRNA transcripts. Comprehensive genotyping of SNPs can be performed with microarray and beadarray technologies or mass spectrometry. Complete sequencing of genomes or sequencing of exons that encode proteins (exome sequencing) is now possible, and will lead to the elucidation of the etiology of a number of human diseases in the next few years.

The availability of individual sequence information is expected to have a significant impact on medical care and preventive strategies but it also raises ethical and legal concerns over how such information may be used by third parties, e.g. insurers and employers.

Genetic linkage and association

There are two primary strategies for mapping genes that cause or increase susceptibility to human disease: *linkage* and *association* studies.

Two principals are essential for understanding the concept of linkage. First, when two genes are close together on a chromosome, they are usually transmitted together, unless a recombination event separates them. Secondly, the odds of a crossover or recombination event between two linked genes are proportional to the distance that separates them. Genes that are further apart are more likely to undergo recombination than genes that are close together. The detection of chromosomal loci that segregate with a disease by linkage has been widely used to identify the gene responsible for the disease by *positional cloning*, a technique of isolating a gene from the knowledge of its map location. It has also been used to predict the odds of disease gene transmission in genetic counseling.

Polymorphisms are essential for linkage studies because they provide a means of distinguishing maternal and paternal chromosomes. On average, one of every 300 base pairs varies from one person to the next. Although this degree of variation seems low (99.9% identical), it means that more than 3 million sequence differences exist between any two unrelated individuals and the probability that the sequence at such loci will differ on the two homologous chromosomes is high (often more than 70–90%). These sequence variations include a variable number of tandem repeats (VNTRs), microsatellites (also referred to as short tandem repeats

[STRs]) and SNPs. Most microsatellite markers consist of di-, tri- or tetranucleotide repeats that can be measured readily using PCR and primers that reside on either side of the repeat sequences. Automated analysis of SNPs with microarrays, beadarrays or mass spectrometry is now the method of choice for determining genetic variation for linkage and association studies.

In order to identify a chromosomal locus that segregates with a disease, it is necessary to determine the genotype or *haplotype* of DNA samples from one or several pedigrees. A haplotype designates a group of alleles that are closely linked (in close proximity on a chromosome) and that are usually inherited as a unit. After characterizing the alleles, one can assess whether certain marker alleles co-segregate with the disease. Markers closest to the disease gene are less likely to undergo recombination events and therefore receive a higher linkage score. Linkage is expressed as a *logarithm of odds (lod) score*, the ratio of the probability that the disease and marker loci are linked rather than unlinked. Lod scores of 3 (1000:1) are generally accepted as supporting linkage.

Allelic association refers to a situation in which the frequency of an allele is significantly increased or decreased in a particular disease. Linkage and association differ in several respects.

- Genetic linkage is demonstrable in families or sibships.
- Association studies compare a population of affected individuals with a control population.
- Association studies are often performed as case-control studies that include unrelated affected individuals and matched controls or as family-based studies that compare the frequencies of alleles that are transmitted to affected children.
- Allelic association studies are useful for identifying susceptibility genes in complex disorders. When alleles at two loci occur more frequently in combination than would be predicted based on known allele frequencies and recombination fractions, they are said to be in *linkage disequilibrium*.

Mutations and human disease

Mutations cause genetic diversity as well as disease. Mutations, which are structurally diverse, can be defined as any change in the nucleotide sequence of DNA

regardless of its functional consequences. They can affect one or a few nucleotides, consist of gross numerical or structural alterations in individual genes or chromosomes or involve the entire genome. Mutations can occur in all domains of a given gene. Large deletions may affect a portion of a gene or an entire gene or, if several genes are involved, they may lead to a *contiguous gene syndrome*.

Occasionally, mispairing of homologous sequences leads to *unequal crossover*. This results in gene duplication on one of the chromosomes and gene deletion on the other chromosome. For example, a significant fraction of growth hormone (*GH*) gene deletions involves unequal crossing-over. The *GH* gene is a member of a large gene cluster that includes a growth hormone variant gene as well as several structurally related chorionic somatomammotropin genes and *pseudogenes*, which are highly homologous but functionally inactive relatives of a normal gene. Because such gene clusters contain multiple homologous DNA sequences arranged along the same chromosome, they are particularly prone to undergo recombination and, consequently, gene duplication or deletion.

Unequal crossing over between homologous genes can result in fusion gene mutations, as illustrated, for example, by glucocorticoid-remediable aldosteronism (GRA). GRA is caused by a rearrangement involving the genes that encode aldosterone synthase (*CYP11B2*) and steroid 11 β -hydroxylase (*CYP11B1*), which are normally arranged in tandem on chromosome 8q. Because these two genes are 95% identical, they are predisposed to undergo unequal recombination. The rearranged gene product contains the regulatory regions of 11 β -hydroxylase upstream to the coding sequence of aldosterone synthetase. The latter enzyme is then expressed in the adrenocorticotrophic hormone (ACTH)-dependent zona fasciculata of the adrenal gland, resulting in overproduction of mineralocorticoids and hypertension.

Gene conversion refers to a non-reciprocal exchange of homologous genetic information by which a recipient strand of DNA receives information from another strand having an allelic difference. The original allele on the recipient strand is converted to the new allele as a consequence of this event. These alterations may range from a few to several thousand nucleotides. Gene conversion often involves exchange of DNA between a gene and a related pseudogene. For example,

the 21-hydroxylase gene (*CYP21A*) is adjacent to a non-functional pseudogene. Many of the nucleotide substitutions found in the *CYP21A* gene of patients with CAH correspond to sequences present in the pseudogene, suggesting gene conversion as the underlying mechanism of mutagenesis.

Several diseases are associated with an increase in the number of trinucleotide repeats above a certain threshold. In some instances, the repeats are located within the coding region of the genes. For example, an expansion in a CAG repeat in the androgen receptor leads to the X-linked form of spinal and bulbar muscular atrophy (SBMA, Kennedy syndrome) and can be associated with partial androgen insensitivity. Similarly, an expansion in the *huntingtin* (*HD*) gene is the cause of Huntington disease.

In other instances, the repeats are located in regulatory sequences. If an expansion is present, the DNA fragment is unstable and tends to expand further during cell division; hence the designation *dynamic mutation*. The length of the nucleotide repeat often correlates with the severity of the disease. When repeat length increases from one generation to the next, disease manifestations may worsen or appear at an earlier age, a phenomenon referred to as *anticipation*. In Huntington disease, for example, there is a correlation between age of onset and length of the triplet codon expansion.

Mutations involving single nucleotides are referred to as *point mutations*. Substitutions are called *transitions* if a purine is replaced by another purine base (A to G) or if a pyrimidine is replaced by another pyrimidine (C to T). Changes from a purine to a pyrimidine or vice versa are referred to as *transversions*. Certain DNA sequences, such as successive pyrimidines or CG dinucleotides, are particularly susceptible to mutagenesis. Therefore, certain types of mutations (C to T or G to A) are relatively common. Moreover, the nature of the genetic code results in overrepresentation of certain amino acid substitutions. If the DNA sequence change occurs in a coding region and alters an amino acid, it is called a *missense mutation*. Depending on the functional consequences of such a missense mutation, amino acid substitutions in different regions of the protein can lead to distinct phenotypes. Occasionally, a single point mutation will result in a stop codon which generates a truncated protein. This is called a *nonsense mutation*.

Small deletions and insertions alter the reading frame if they do not represent a multiple of three bases. Such

“frameshift” mutations lead to an entirely altered carboxy-terminus. Mutations may also be found in the regulatory sequences of genes and result in reduced gene transcription. Mutations in intronic sequences or in exon junctions may destroy or create splice donor or splice acceptor sites.

Some mutations are lethal, some have less deleterious yet recognizable consequences and some confer evolutionary advantage. Mutations occurring in germ cells can be transmitted to the progeny. Alternatively, mutations can occur during embryogenesis or in somatic tissues. Mutations that occur during development lead to *mosaicism*, a situation in which tissues are composed of cells with different genetic constitutions, as illustrated by Turner or McCune–Albright syndromes. If the germline is mosaic, a mutation can be transmitted to some progeny but not others, which sometimes leads to confusion in assessing the pattern of inheritance. Other somatic mutations are associated with neoplasia because they confer a growth advantage to cells by activating (proto)oncogenes or inactivating tumor suppressor genes. Epigenetic events, heritable changes that do not involve changes in gene sequence (e.g. altered DNA methylation), may also influence gene expression or facilitate genetic damage.

Polymorphisms are sequence variations that have a frequency of at least 1% and do not usually result in an overt phenotype. Often they consist of single base pair substitutions that do not alter the protein coding sequence because of the degenerate nature of the genetic code, although some might alter mRNA stability, translation or the amino acid sequence. Silent base substitutions and SNPs are encountered frequently during genetic testing and must be distinguished from true mutations that alter protein expression or function. However, some SNPs or combinations of SNPs may have a pathogenic role in complex disorders by conferring susceptibility for the development of the disease.

Functional consequences of mutations

Mutations result in either gain or loss of function. The consequences of an altered protein sequence often need experimental evaluation *in vitro* to determine that the mutation alters protein function, although it is important to bear in mind that *in vitro* studies may not accurately reflect *in vivo* consequences.

Gain-of-function mutations are typically dominant and result in phenotypic alterations when a single allele is

affected. Inactivating mutations are usually recessive and an affected individual is homozygous or compound heterozygous (i.e. carrying two different mutant alleles) for the disease-causing mutations. Mutation in a single allele can result in *haploinsufficiency*, a situation in which one normal allele is not sufficient to maintain a normal phenotype. Haploinsufficiency is a commonly observed mechanism in diseases associated with mutations in transcription factors. For example, monoallelic mutations in the transcription factor TTF1 (NKX2.1) are associated with transient congenital hypothyroidism, respiratory distress and ataxia.

The clinical features among patients with an identical mutation in a transcription factor often vary. One mechanism underlying this variability consists of the influence of modifying genes. Haploinsufficiency can affect the expression of rate-limiting enzymes. For example, in MODY2, heterozygous glucokinase mutations result in haploinsufficiency with a higher threshold for glucose-dependent insulin release and mild hyperglycemia.

Mutation of a single allele can result in loss of function due to a dominant-negative effect. In this case, the mutated allele interferes with the function of the normal gene product by several different mechanisms. The mutant protein may interfere with the function of a multimeric protein complex, as illustrated by Liddle syndrome, which is caused by mutations in the β - or γ -subunit of the renal sodium channel. In thyroid hormone resistance, mutations in the thyroid hormone receptor β lead to impaired T3 binding; the receptors cannot release co-repressors and they silence transcription of target genes. The mutant protein can be cytotoxic, as in autosomal-dominant neurohypophyseal diabetes insipidus, in which abnormal folding leads to retention in the endoplasmic reticulum and degeneration of neurons secreting arginine vasopressin (AVP). A similar mechanism may lead to autosomal dominant or type II growth hormone deficiency (GHD).

An increase in dosage of a gene product may also result in disease. For example, duplication of the *DAX1* (*NR0B1*) gene results in dosage-sensitive sex reversal and *SOX3* duplications are associated with X-linked hypopituitarism.

Variable expressivity and incomplete penetrance

Penetrance and expressivity are two different yet related concepts which are often confused. *Penetrance* is a

qualitative notion designating whether a phenotype is expressed for a particular genotype. *Expressivity* is a quantitative concept describing the degree to which a phenotype is expressed. It is used to describe the phenotypic spectrum in individuals with a particular disorder. Thus, expressivity is dependent on penetrance.

Penetrance is complete if all carriers of a mutation express the phenotype, whereas it is incomplete if some individuals do not have any features of the phenotype. Dominant conditions with incomplete penetrance are characterized by skipping of generations with unaffected carriers transmitting the mutant gene. For example, hypertrophic obstructive cardiomyopathy (HOCM) caused by mutations in the *myosin-binding protein C* gene is a dominant disorder with clinical features in only a subset of patients who carry the mutation.

Incomplete penetrance in some individuals can confound pedigree analysis. In many conditions with postnatal onset, the proportion of gene carriers affected varies with age. Therefore, it is important to specify age when describing penetrance. Variable expressivity is used to describe the phenotypic spectrum in individuals with a particular disorder.

The mechanisms underlying the variability in phenotypic penetrance and expressivity remain largely unknown; they may reflect the effects of multiple genetic variations and/or interactions with environmental factors. Recently, mutations in more than one gene have been shown to contribute to the manifestation of the phenotype as well as its severity in some patients with hypogonadotropic hypogonadism (digenic inheritance).

Sex-influenced phenotypes

Certain mutations affect males and females differently, sometimes because the gene resides on the X or Y sex chromosomes. As a result, the phenotype of mutated X-linked genes will usually be expressed fully in males but variably in heterozygous females, depending on the degree of X-inactivation and the function of the gene. Because only males have a Y chromosome, mutations in genes such as *SRY* (which causes male-to-female sex reversal) or *DAZ* (*deleted in azoospermia*), which causes abnormalities of spermatogenesis, are unique to males.

Other diseases are expressed in a sex-limited manner because of the differential function of the gene product in males and females. Activating mutations in the luteinizing hormone receptor (LHR) cause dominant male-limited

precocious puberty in boys. The phenotype is unique to males because activation of the receptor induces testosterone production in the testis, whereas it is functionally silent in the immature ovary. Homozygous inactivating mutations of the follicle-stimulating hormone (FSH) receptor cause primary ovarian failure in females because the follicles do not develop in the absence of FSH. Affected males have a more subtle phenotype, because testosterone production is preserved and spermatogenesis is only partially impaired.

Heterozygous mutations in the transcription factor NR5A1 (previously known as steroidogenic factor 1) may be associated with sex-limited phenotypes: males may manifest 46XY disorders of sex development with or without adrenal development, whereas females may have no phenotype or variable ovarian failure.

Chromosomal disorders

Chromosomal (cytogenetic) disorders are caused by numerical or structural aberrations in chromosomes. Errors in meiosis and early cleavage divisions occur frequently. Ten to 25% of all conceptions harbor chromosomal abnormalities, which often lead to spontaneous abortion in early pregnancy. Numerical abnormalities, especially trisomy, which is found in about 25% of spontaneous abortions and 0.3% of newborns, are more common than structural defects.

Autosomal monosomies are usually incompatible with life but 45X, present in 1–2% of all conceptuses, leads to spontaneous abortion in at least 99% of cases. Mosaicism (e.g. 45XO/45XX, 45XO/45XXX), partial deletions, isochromosomes and ring chromosomes can also cause Turner syndrome. Sex chromosome monosomy usually results from loss of the paternal sex chromosome.

Structural rearrangements involve breakage and reunion of chromosomes. Rearrangements between different chromosomes, *translocations*, can be reciprocal or Robertsonian. *Reciprocal* translocations involve exchanges between any of the chromosomes; *Robertsonian* rearrangements designate the fusion of the long arms of two acrocentric chromosomes.

Other structural defects include deletions, duplications, inversions and the formation of rings and isochromosomes. These may result in complex phenotypes or in highly specific disorders. Chromosomal breakpoints could occasionally lead to the identification of novel candidate genes for a specific disorder, as may

deletions or duplications. Rarely, there may be submicroscopic deletions at the breakpoints of apparently balanced chromosomal translocations, which may be detected by high-resolution array comparative genomic hybridization (CGH).

Because of the variable size of gene deletions in different patients, a systematic comparison of phenotypes and locations of deletion breakpoints allows the positions of particular genes to be mapped within the critical genomic region. Structural chromosome defects can be present in a “balanced” form without an abnormal phenotype. However, they can be transmitted in an “unbalanced” form to offspring and thus cause an hereditary form of chromosome abnormality.

Traditional karyotype analysis usually identifies chromosomal rearrangements and/or aberrations of 3–5 Mb and larger. Comparative genomic hybridization and other techniques now permit the detection of more subtle, submicroscopic chromosomal imbalances, such as copy number variations (CNVs). The clinical relevance of these alterations is not always known and many may be benign or polymorphic. *De novo* changes may offer insight into the etiology of some disorders. The Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER) (www.sanger.ac.uk/PostGenomics/decipher) catalogs pertinent genomic and clinical information of such patients and can assist in the interpretation of genome-wide, high-resolution tests.

Monogenic mendelian disorders

Monogenic human diseases are often called *mendelian disorders* because they obey the rules of genetic transmission defined by Gregor Mendel. The mode of inheritance for a given phenotype or disease is determined by pedigree analysis. About 65% of human monogenic disorders are autosomal dominant, 25% are autosomal recessive and 5% are X-linked. Genetic testing now available for many of these disorders has an increasingly important role in pediatric endocrinology.

Autosomal-dominant disorders

In autosomal-dominant disorders, mutations in a single allele are sufficient to cause the disease; recessive disorders are the consequence of biallelic loss-of-function mutations. Various disease mechanisms are involved in dominant disorders, which include gain-of-function, a dominant-negative effect, and haploinsufficiency. In

autosomal-dominant disorders, individuals are affected in successive generations and the disease does not occur in the offspring of unaffected individuals. Males and females are affected with equal frequency because the defective gene resides on one of the 22 autosomes. Because the alleles segregate randomly at meiosis, the probability that an offspring will be affected is 50%. Children with a normal genotype do not transmit the disorder. The clinician must be aware that an autosomal-dominant disorder can be caused by *de novo* germline mutations, which occur more frequently during later cell divisions in gametogenesis, explaining why siblings are rarely affected. New germline mutations occur more frequently in fathers of advanced age. The clinical manifestations of autosomal-dominant disorders may be variable as a result of differences in *penetrance* or *expressivity*. Because of these variations, it is sometimes difficult to determine the pattern of inheritance.

Autosomal-recessive disorders

The clinical expression of autosomal-recessive disorders is usually more uniform than in autosomal-dominant disorders. Most mutated alleles lead to a partial or complete loss of function. They frequently involve receptors, proteins in signaling cascades or enzymes in metabolic pathways. The affected individual, who can be of either sex, is homozygous or compound heterozygous for a single gene defect. In most instances, an affected individual is the offspring of heterozygous parents. In this situation, there is a 25% chance that the offspring will have a normal genotype, a 50% probability of a heterozygous state and a 25% risk of homozygosity or compound heterozygosity for the recessive alleles. In the case of one unaffected heterozygous and one affected homozygous parent, the probability of disease increases to 50% for each child. In this instance, the pedigree analysis mimics an autosomal-dominant mode of inheritance (*pseudodominance*). In contrast to autosomal-dominant disorders, new mutations in recessive alleles usually result in an asymptomatic carrier state without apparent clinical phenotype.

Many autosomal-recessive diseases are rare and occur more frequently in isolated populations in the context of parental consanguinity. A few recessive disorders, such as sickle cell anemia, cystic fibrosis and thalassemia, are relatively frequent in certain populations, perhaps because the heterozygous state may confer a selective

biological advantage. Although heterozygous carriers of a defective allele are usually clinically normal, they may display subtle differences in phenotype that become apparent only with more precise testing or in the context of certain environmental influences.

X-linked disorders

Because males have only one X chromosome, a female individual always inherits her father's X chromosome in addition to one of the two X chromosomes of her mother. A son inherits the Y chromosome from his father and one maternal X chromosome. The characteristic features of X-linked inheritance are therefore the absence of father-to-son transmission and the fact that all daughters of an affected male are obligate carriers of the mutant allele. The risk of developing disease caused by a mutant X-chromosomal gene differs in the two sexes. Because males have only one X chromosome, they are hemizygous for the mutant allele. Consequently, they are more likely to develop the mutant phenotype, regardless of whether the mutation is dominant or recessive. A female may be either heterozygous or homozygous for the mutant allele, which may be dominant or recessive, and the terms *X-linked dominant* or *X-linked recessive* apply only to the expression of the mutant phenotype in women. In females, the expression of X-chromosomal genes is influenced by X chromosome inactivation. This can confound the assessment because skewed X-inactivation may lead to a partial phenotype in female carriers of an X-linked recessive defect, such as inactivating mutations of the AVPR2 receptor, the cause of X-linked nephrogenic diabetes insipidus.

Y-linked disorders

The Y chromosome harbors relatively few genes. Among them, the sex region-determining Y factor (*SRY*), which encodes the testis-determining factor (TDF), is essential for normal male development. Because the *SRY* region is closely adjacent to the pseudoautosomal region, a chromosomal segment on the X and Y chromosomes with a high degree of homology, crossing over can occasionally involve the *SRY* region. Translocations can result in XY females, with the Y chromosome lacking the *SRY* gene, or XX males harboring the *SRY* gene on one of the X chromosomes. Point mutations in the *SRY* gene may result in individuals with an XY genotype and an incomplete female phenotype. Men with oligospermia

or azoospermia frequently have microdeletions of the AZF (azoospermia factor) regions on the long arm of the Y chromosome, which contain several genes involved in the control of spermatogenesis. They may have point mutations in the transcription factor DAZ (deleted in azoospermia), which is located in this chromosomal region.

Exceptions to simple mendelian inheritance

Mitochondrial disorders

Mendelian inheritance refers to the transmission of genes encoded by DNA in nuclear chromosomes but each mitochondrion contains several copies of a circular chromosome. Mitochondrial DNA (mtDNA) is small (16.5 kb) and encodes transfer and ribosomal RNAs and 13 proteins that are part of the respiratory chain involved in oxidative phosphorylation and ATP generation. In contrast to the nuclear chromosomes, the mitochondrial genome does not recombine and is inherited through the maternal line because sperm does not contribute significant cytoplasmic components to the zygote. The D-loop, a non-coding region of the mitochondrial chromosome, is highly polymorphic. This property, together with the absence of recombination of mtDNA, makes it a helpful tool for studies tracing human migration and evolution and for specific forensic applications.

Inherited mitochondrial disorders are transmitted in a matrilineal fashion. All children from an affected mother inherit the disease but it will never be transmitted from an affected father to his offspring except by intracytoplasmic sperm injection (ICSI). Alterations in the mtDNA affecting enzymes required for oxidative phosphorylation lead to reduction of ATP supply, generation of free radicals and induction of apoptosis. Several syndromic disorders arising from mutations in the mitochondrial genome are known in humans and they affect both protein-coding and tRNA genes. The clinical spectrum often involves (cardio)myopathies and encephalopathies because of the high dependence of these tissues on oxidative phosphorylation.

Many may present with endocrine features. For example, the mitochondrial DIDMOAD syndrome consists of diabetes insipidus, diabetes mellitus, optic atrophy and deafness. The age of onset and the clinical course are variable because of the unusual mechanisms of mtDNA replication. mtDNA replicates independently from nuclear DNA and during cell replication, the

proportion of wild type and mutant mitochondria can drift among different cells and tissues. The resulting heterogeneity in the proportion of mitochondria with and without a mutation is referred to as *heteroplasmy* and underlies the phenotypic variability that is characteristic of mitochondrial diseases. Nuclear genes that encode proteins that are important for normal mitochondrial function can cause mitochondrial dysfunctions associated with autosomal-dominant or -recessive forms of inheritance.

Acquired somatic mutations in mitochondrial genes are thought to be involved in several age-dependent degenerative disorders involving muscle and the peripheral and central nervous systems. Because of the high degree of polymorphisms in mtDNA and the phenotypic variability of these disorders, it is difficult to establish that an mtDNA alteration is causal for a clinical phenotype.

Mosaicism

Mosaicism refers to the presence of two or more genetically distinct cell lines in the tissues of an individual. It results from a mutation that occurs during embryonic, fetal or extrauterine development. The developmental stage at which the mutation arises will determine whether germ cells and/or somatic cells are involved. Chromosomal mosaicism results from non-disjunction at an early embryonic mitotic division, leading to the persistence of more than one cell line, as exemplified by some patients with Turner syndrome. Somatic mosaicism is characterized by a patchy distribution of genetically altered somatic cells that occurs early in development. This is best illustrated by the McCune–Albright syndrome, which is caused by activating mutations in the *GNAS1* gene encoding the stimulatory G-protein α ($Gs\alpha$). The clinical phenotype varies depending on the tissue distribution of the mutation. Manifestations include ovarian cysts that secrete sex steroids and cause precocious puberty, polyostotic fibrous dysplasia, café-au-lait skin pigmentation, growth hormone-secreting pituitary adenomas and, among others, hypersecreting autonomous thyroid nodules and adrenal adenomas secreting cortisol and resulting in Cushing syndrome.

Epigenetic modifications, X-inactivation, imprinting and uniparental disomy

According to traditional mendelian principles, the parental origin of a mutant gene is irrelevant for the

expression of the phenotype, although there are important exceptions to this rule. X-inactivation prevents the expression of most genes on one of the two X chromosomes in every cell of a female (*lyonization*). Gene inactivation also occurs on selected chromosomal regions of autosomes. This phenomenon, *genomic imprinting*, leads to preferential expression of an allele depending on its parental origin. It is of importance in disorders in which the transmission of disease is dependent on the sex of the transmitting parent and has an important role in the expression of certain genetic disorders.

The two classic examples are the Prader–Willi and Angelman syndromes. Prader–Willi syndrome (PWS) is characterized by diminished fetal activity, obesity, hypotonia, mental retardation, short stature and hypogonadotropic hypogonadism. Deletions in PWS occur exclusively on the paternal chromosome 15. Patients with Angelman syndrome present with mental retardation, seizures, ataxia and hypotonia and have deletions at the same site of chromosome 15 but they are located on the maternal chromosome 15. These two syndromes may also result from *uniparental disomy*, i.e. by the inheritance of either two maternal chromosomes 15 (PWS) or two paternal chromosomes (Angelman syndrome).

Another example of importance for pediatric endocrinology concerns the *GNAS1* gene encoding the $Gs\alpha$ subunit. Heterozygous loss-of-function mutations in the *GNAS1* gene lead to Albright hereditary osteodystrophy (AHO) with its characteristic features including short stature, obesity, round face, brachydactyly, subcutaneous ossifications and mental deficits. Paternal transmission of *GNAS1* mutations leads to the AHO phenotype alone (*pseudopseudohypoparathyroidism*), while maternal transmission leads to AHO in combination with resistance to several hormones such as parathyroid hormone (PTH), thyroid-stimulating hormone (TSH) and gonadotropins, which act through transmembrane receptors coupling to $Gs\alpha$ (*pseudohypoparathyroidism type IA*). These phenotypic differences result from a tissue-specific imprinting of *GNAS1*, which is expressed primarily from the maternal allele in tissues such as the proximal renal tubule and the thyroid. In most other tissues, however, it is expressed biallelically. Disrupting mutations in the maternal allele lead to loss of $Gs\alpha$ expression in proximal tubules and loss of PTH action in the kidney, while mutations in the paternal allele have little effect on PTH action. In patients with isolated renal resistance to PTH

(pseudohypoparathyroidism type IB), an imprinting defect of *GNAS1* leads to decreased $Gs\alpha$ expression in the proximal renal tubules.

Somatic mutations

Acquired mutations that occur in somatic rather than germ cells are called *somatic mutations*. This creates a chimeric situation and, if the cells proliferate, a neoplastic lesion. Therefore, cancer can be defined as a genetic disease at the cellular level. Cancers are monoclonal, indicating that they have arisen from a single precursor cell that has acquired one or several mutations in genes controlling growth and/or differentiation. These mutations are somatic, i.e. restricted to the tumor and its metastases, but not found in surrounding normal tissue. The molecular alterations include dominant gain-of-function mutations in oncogenes, recessive loss-of-function mutations in tumor suppressor genes and DNA repair genes, gene amplification and chromosome rearrangements. Rarely, a single mutation in certain genes may be sufficient to transform a normal cell into a malignant cell but the development of a malignant phenotype in most cancers requires several genetic alterations for the gradual progression from a normal to a cancerous cell, a process termed *multistep carcinogenesis*.

In many cancer syndromes, there is an inherited *predisposition* to tumor formation. In these instances, a germline mutation is inherited in an autosomal-dominant fashion. This germline alteration affects one allele of an autosomal tumor suppressor gene. If the second allele is inactivated by a somatic mutation in a given cell, this will lead to neoplastic growth (*Knudson two-hit model*). In this instance, the defective allele in the germline is transmitted in a dominant way, whereas the tumorigenic mechanism results from a recessive loss of the tumor suppressor gene in affected tissues.

The classic example to illustrate this phenomenon is retinoblastoma, which can occur as a sporadic or hereditary tumor. In sporadic retinoblastoma, both copies of the retinoblastoma (*RB*) gene are inactivated through two somatic events. In hereditary retinoblastoma, one mutated or deleted *RB* allele is inherited in an autosomal-dominant manner and the second allele is inactivated by a subsequent somatic mutation. This “two-hit” model applies to other inherited cancer syndromes, such as multiple endocrine neoplasia type 1 (MEN1), which is caused by mutations in the tumor suppressor gene *menin*.

Inherited defects in enzymes involved in DNA replication and repair can lead to a significant increase in mutations and are associated with several disorders predisposing to cancer.

Complex disorders

Many disorders have a complex etiology involving multiple genes (*polygenic disorders*), often in combination with environmental and lifestyle factors (*multifactorial disorders*). The major healthcare problems, cardiovascular disease, hypertension, diabetes, obesity, asthma and psychiatric disorders, fall into this category but it also includes certain developmental abnormalities, such as cleft palate, congenital heart defects and neural tube defects.

Compared with single gene defects, complex disorders have a low heritability and do not fit a mendelian pattern of inheritance. Twin studies are particularly helpful in demonstrating the importance of genetic and environmental factors. For example, first-degree relatives of patients with diabetes mellitus type 1 are about 15 times more likely to develop diabetes. The concordance rate for developing diabetes is about 50% in monozygotic twins and about 8% in dizygotic twins. The discordance rate in monozygotic twins illustrates the significant requirement for environmental factors. In addition, some of the susceptibility genes, a designation indicating that the carrier is susceptible to developing the disease, have a low penetrance. Susceptibility genes or loci can be mapped using several methods, including linkage analyses, association studies and affected sib-pair analyses. Current efforts aim to identify these genes by establishing correlations between SNPs or SNP haplotypes and complex disorders in large populations through genome-wide association studies (GWAS). The HapMap data are significantly facilitating this type of study because they allow genotyping a reduced number of tag SNPs reflecting certain haplotypes. The results of GWAS may, in part, depend on ethnicity and ascertainment criteria.

The study of rare monogenic diseases may also provide insights into genetic and molecular mechanisms important for the understanding of complex disorders. For example, the identification of the genetic defects underlying the various autosomal-dominant forms of MODY have defined them, in part, as candidate genes contributing to the pathogenesis of diabetes mellitus type 2.

Novel techniques – the future of molecular diagnosis

Advances in molecular biology have produced a number of new techniques which are likely to lead to major advances in understanding a number of disease processes. The use of array CGH has revolutionized karyotype analysis and is used increasingly to identify submicroscopic chromosomal deletions, duplications and rearrangements. The technique relies on the differential labeling with different fluorochromes of DNA isolated from a test and reference cell population and to co-hybridize the labeled samples to a metaphase spread from a reference cell. The relative hybridization intensity of the test and reference signals at a given location is then proportional to the relative copy number of those sequences in the test and reference genomes. If the reference genome is normal, increases and decreases in the intensity ratio directly indicate DNA copy number variation in the genome of the test cells. The ratio of the intensities of two fluorochromes reflects the copy number differences between the cells of interest and the control cells. This has led to the introduction to CGH of high-resolution hybrid arrays that integrate both SNP and CNV probes.

High-throughput, high-density sequencing using microarray technology potentially offers the option of rapid, accurate and relatively inexpensive sequencing of large portions of the genome. One such technique is oligohybridization sequencing, which relies on the differential hybridization of target DNA to an array of oligonucleotide probes. This technique is ideally suited to the analysis of DNA from patients with defined disorders, such as disorders of sex development and retinal disease, but suffers from a relatively high false-positive rate and failure to detect insertions and deletions.

The greatest strides have been made in the area of next-generation sequencing (NGS), which has made possible whole-genome sequence analysis. A number of technologies are available, with more being developed. Template preparation, preparation of the sequencing reaction, the sequencing itself and the data handling and analyses are critical steps in NGS. The costs involved and the data handling skills required currently restrict NGS to research laboratories. More targeted approaches such as exome capture, which analyzes the sequence of coding exons only, or targeting specific chromosomes, such as the X-chromosome in potentially X-linked conditions, are becoming more widely available. NGS technologies are

capable of identifying novel disease genes and structural chromosomal variations. For example, the molecular basis of recessive Charcot–Marie–Tooth neuropathy was identified in a pedigree using whole-genome sequencing.

Molecular biology and therapeutics

The understanding of the genetic basis of a disease is critical to optimizing treatment. Gene therapy (introduction of genetic material) requires replacement of a deficient gene or of a gene product, such as an enzyme, neither technique being simple. The product requires optimal delivery, sustained expression, avoidance of deleterious immunological or tissue-based responses and avoidance of secondary disease due to endogenous gene disruption at the site of gene insertion once vectors have integrated. These are all obstacles that need to be surmounted, thereby reducing the applicability of this technique to only the most severe and life-threatening conditions, such as severe combined immunodeficiency (SCID).

Advances in molecular biology have, however, resulted in the availability of new drugs which target or replace missing enzymes in lysosomal storage disorders as well as in some of the mucopolysaccharidoses. Advances have also been made in understanding rare disorders, such as Hutchinson–Gilford progeria syndrome and Marfan syndrome, with the hope of new therapies that will delay progression and prevent life-threatening complications.

The use of aminoglycosides to suppress the effects of a nonsense mutation and small interfering RNA (siRNA) offers promise of novel therapies in a range of conditions, such as cystic fibrosis and type II autosomal dominant idiopathic isolated GH deficiency (IGHD). Elucidation of the genetic basis of disease also allows more direct targeting of therapy. For instance, children with permanent neonatal-onset diabetes mellitus (PNDM) due to mutations in *SURI* or *KIR6.2* were previously treated with insulin but have now been shown to respond well to sulphonylureas, thereby allowing the cessation of insulin therapy.

Finally, we are now entering the era of pharmacogenetics when the response of an individual to various therapeutic agents may be determined by their genotype. For example, a polymorphism in the GH receptor that results in deletion of exon 3 may be associated with an improved response to GH. Thus the elucidation of the genetic basis of many disorders will aid their management.

What to look out for

1. The failure to detect genetic mutations in association with a disorder does not exclude a genetic basis to the disorder.
2. Autosomal-dominant mutations may be variably penetrant so that some carriers do not manifest the condition, yet are capable of transmitting the mutated gene to future progeny who may in turn be affected.
3. Disorders with variably penetrant autosomal-dominant inheritance may in fact be due to mutations in more than one gene.

Web links

- National Center for Biotechnology Information (NCBI): www.ncbi.nlm.nih.gov/
- Online Mendelian Inheritance in Man (OMIM): www.ncbi.nlm.nih.gov/omim/
- National Human Genome Research Institute: www.genome.gov/
- European Bioinformatics Institute (EBI): www.ebi.ac.uk
- DNA Database of Japan: www.ddbj.nig.ac.jp/
- University of California, Santa Cruz (UCSC) Genome Bioinformatics: <http://genome.ucsc.edu/>
- Swiss-Prot: www.ebi.ac.uk/swissprot/index.html
- Protein Structure Database: www.rcsb.org/pdb/home/home.do
- American College of Medical Genetics: www.acmg.net/
- Genecards: <http://bioinformatics.weizmann.ac.il/cards/>
- GeneTests·GeneClinics: www.genetests.org/
- Gene Ontology: www.geneontology.org/
- Chromosomal Variation in Man: www.wiley.com/legacy/products/subject/life/borgaonkar/access.html
- Database of Chromosomal Imbalance and Phenotype in Humans using Ensembl Resources (DECIPHER): www.sanger.ac.uk/PostGenomics/decipher
- Mitochondrial disorders, DNA repeat sequences and disease: <http://neuromuscular.wustl.edu/>
- National Organization for Rare Disorders: www.rarediseases.org/

Self-assessment

1. In the human genome:
 - A. More than 100,000 genes encode their relevant proteins.
 - B. Genes account for the majority of DNA.
 - C. Micro-RNAs play an important role in developmental and physiological processes, and may be implicated in the ontogeny of cancers.
 - D. Alternative splicing is extremely important to maintain phenotypic diversity.
 - E. Copy number variations always indicate pathology.
2. Array-CGH:
 - A. Can lead to the identification of novel disease-causing genes.
 - B. Can identify submicroscopic chromosomal deletions.
 - C. Can identify methylation defects.
 - D. Can identify genetic duplications.
 - E. Can identify missense mutations.
3. Which of the following are examples of imprinted disorders?
 - A. Kennedy syndrome
 - B. Pseudohypoparathyroidism type 1A
 - C. Silver–Russell syndrome
 - D. DSD due to mutation in *SRY*
 - E. Prader–Willi syndrome
4. Which of the following are examples of conditions where a molecular genetic diagnosis has influenced treatment?
 - A. Permanent neonatal diabetes mellitus
 - B. Huntington disease
 - C. Gaucher disease
 - D. Nephrogenic diabetes insipidus
 - E. Type 1 diabetes mellitus
5. Which of the following statements are false?
 - A. In an autosomal-recessive disorder, the parents are usually unaffected.
 - B. An X-linked disorder is characterized by transmission of the condition from father to son.
 - C. Autosomal-dominant conditions may skip generations.
 - D. Central diabetes insipidus and growth hormone deficiency are examples of dominant negative disorders.
 - E. With X-linked recessive disorders, carrier females are phenotypically normal.

Answers: 1. C, D. 2. A, B, D. 3. B, C, E. 4. A, C. 5. B, E.