The Role of Genetics in Endocrinology, 49
Principles of Genetics, 49
Genetics of Endocrine Diseases, 54
Considerations for Clinical Use of Genetic Information and Sequencing in Endocrinology, 62

KEY POINTS

• The genetic basis of each heritable endocrine disease/trait is quantified by its genetic architecture: (1) the number of genetic variants/genes, (2) their frequency in the population, and (3) their respective contributions to disease risk/phenotypic variation.

• Mendelian endocrine disorders are caused by few variants in few genes, found rarely in the population, and each has a large individual effect on disease risk.

• Common endocrine diseases/traits such as stature, type 2 diabetes, and serum lipids are polygenic—the result of combined, simultaneous effects of many variants in many genes, found frequently in the general population, and with each variant contributing a small individual effect.

• Genetic information enables endocrinologists to personalize therapy for patients.

• Comprehensive genetic testing (i.e., genome sequencing) can be standardized and automated, but drawing valid and clinically useful conclusions requires integration with patient history, physical examination, and other laboratory examinations.

• Genetic information is most likely to be of direct clinical use in patients with suspected mendelian syndromes.

THE ROLE OF GENETICS IN ENDOCRINOLOGY

The sequencing of the human genome has ushered in an era of genomic medicine. The catalog of protein-coding genes in humans is essentially complete, and the number of associations between genes and specific diseases is growing rapidly. Moreover, it is now feasible to identify nearly every genetic variant in an individual’s protein-coding genes (whole-exome sequencing) or in his or her entire genome (whole-genome sequencing). The ability to interpret this variation is less advanced but is improving, as databases of variants and their clinical associations increase in both size and accuracy.

With the expanding reach of precision medicine—individualized diagnosis and therapy informed by genetics—we anticipate that increasing numbers of patients will have clinical indications for exome or genome sequencing, and others will come to clinical encounters with their sequences already in hand. Clinicians will be asked to interpret these genetic data to shed light on an individual’s risk of developing disease, on diagnosis and prognosis for those already affected, on implications to family members, and on individualization of therapy. As such, it is critical that clinicians be able to draw valid and clinically useful connections between DNA sequence variation and human traits and diseases. Perhaps even more important, it is critical that clinicians understand the limits of such information.

In this chapter we present a guide to help clinicians appreciate and critically interpret the relationship between DNA sequence variation and human traits and diseases. Perhaps even more important, it is critical that clinicians understand the limits of such information.

In this chapter we present a guide to help clinicians appreciate and critically interpret the relationship between DNA sequence variation and human traits and diseases. We first discuss principles of genetics to provide the framework for understanding and interpreting DNA variation in patients. We then focus on endocrine disorders, providing an overview of the genetics of endocrine diseases, with illustrative examples from both mendelian disorders (caused by mutations in single genes) and polygenic disorders (in which variation in many genes influences disease risk). Finally, we examine scenarios for clinical uses of genetic information in endocrinology and provide recommendations.

Most diseases, including endocrine disorders, are heritable, meaning that genetic variation contributes to disease risk in a population. These diseases range across the spectrum of rare, single-gene disorders, such as multiple endocrine neoplasia (Chapter 39), Carney complex (Chapter 15), and congenital adrenal hyperplasia (CAH) (Chapter 23), to polygenic diseases, such as type 2 diabetes (Chapter 31), Graves disease (Chapter 12), and osteoporosis (Chapter 28). The detailed discussions of the genetics of these and other disorders can be found throughout this textbook; this chapter will provide illustrative examples that illuminate key concepts and will refer the reader to those appropriate chapters for additional detail.

PRINCIPLES OF GENETICS

A Brief Historical Perspective

In Western conception, the relationship between inheritance and physical characteristics (disease and nonpathologic) has been recognized since the time of Aristotle (323 BC). But it was not until 1865 that the Austrian abbot
Gregor Mendel, after decades of careful experimentation in pea plants, posited and provided evidence for the modern genetic concept of genes (as coined by the botanist Wilhelm Johannsen in 1909). Mendel deduced certain rules governing the passage of genotype (the collective versions of multiple genes in an individual) from parent to offspring, enabling the prediction of the resulting physical characteristics (phenotype) of the offspring. It was recognized in the early 20th century that certain human phenotypes, including diseases, were inherited according to the same rules that Mendel had described; these diseases are called mendelian.

Over the course of the next century, numerous breakthroughs established that genes were composed of DNA, physically connected on chromosomes, and encoded proteins. The first description of the molecular basis of a mendelian disease was made for sickle cell anemia, which involved a mutation in a single gene. In the 1970s, the ability to sequence DNA revealed natural and heritable sequence variation (genetic polymorphisms) in any given gene among different individuals. It was appreciated that the molecular basis of variation in the genotypes of individuals resulted from DNA sequence polymorphisms, which in turn effected alterations in phenotype. By tracing the transmission of these polymorphisms in families, it became possible to identify genes causing mendelian human disorders (those caused by altered function in a single gene and that consequently show distinctive patterns of inheritance in families).

However, most human diseases and phenotypes are not mendelian. Biometricians had appreciated in the early 1900s that most continuous and commonly varying traits (such as height and blood pressure) did not follow mendelian patterns of inheritance. In 1918, R.A. Fisher provided a general framework explaining continually varying traits as the consequence of polygenic inheritance; that is, polygenic phenotypes are a result of combined, small, and additive effects of variation in many genes simultaneously. In this framework, monogenic/mendelian traits were a special case. Despite this recognition, only a few genetic variants were convincingly connected with polygenic diseases/trait over the next 80 years. It would take a series of technologic advances, including the sequencing of the human genome over the next 80 years. It would take a series of technologic advances, including the sequencing of the human genome (Human Genome Project 1990-2003) and the systematic cataloging of DNA sequence polymorphisms across diverse human populations (International HapMap Project 2002-2005 Phase I), to systematically identify the genetic causes for common polygenic diseases.

## Table 4-1

<table>
<thead>
<tr>
<th>Heritable Endocrine Traits and Diseases</th>
<th>Heritability</th>
<th>Reference</th>
<th>Selected Mendelian Forms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type 1 diabetes</td>
<td>80%</td>
<td>117</td>
<td>KCNJ1, ABCC8 (permanent neonatal diabetes)</td>
</tr>
<tr>
<td>Type 2 diabetes</td>
<td>40-80%</td>
<td>29, 34, 118</td>
<td>AGPAT2 (congenital generalized lipodystrophy), LMNA (familial partial lipodystrophy 1)</td>
</tr>
<tr>
<td>Obesity</td>
<td>40-70%</td>
<td>119, 120</td>
<td>MEN1, RET (MEN2A/B), VHL, SCN9A (Liddle syndrome)</td>
</tr>
<tr>
<td>Hypertension</td>
<td>30-70%</td>
<td>121</td>
<td>HHT1 (achondroplasia), SMOX1 (Lillic-Home syndrome), FBN1 (Marfan syndrome)</td>
</tr>
<tr>
<td>Hyperlipidemia</td>
<td>50-80%</td>
<td>122</td>
<td>GH1, FGF21 (achondroplasia), SMOX1 (Lillic-Home syndrome), FBN1 (Marfan syndrome)</td>
</tr>
<tr>
<td>Graves disease</td>
<td>50-80%</td>
<td>123</td>
<td>KAL1, KISS1R, FGF11 (hypogonadotropic hypogonadism)</td>
</tr>
<tr>
<td>Hypothyroidism</td>
<td>60-70%</td>
<td>124</td>
<td>TSHR (familial nonautonomous hyperthyroidism)</td>
</tr>
<tr>
<td>Osteoporosis</td>
<td>50-85%</td>
<td>125, 126</td>
<td>SCL3A8, TG, TPO, and TSHB (congenital hypothyroidism)</td>
</tr>
<tr>
<td>Serum calcium</td>
<td>40%</td>
<td>127, 128</td>
<td>COL1A1, COL1A2, IFITM5 (ostogenesis imperfects)</td>
</tr>
<tr>
<td>Lipids</td>
<td>40-60%</td>
<td>81, 82</td>
<td>CASR (familial hypocalciuric hypercalciemia), HRPT2 (hyperparathyroid jaw-tumor syndrome)</td>
</tr>
<tr>
<td>Kidney stones</td>
<td>56%</td>
<td>129</td>
<td>LDL: LDLR (familial hypercholesterolemia), HDL: CETP</td>
</tr>
</tbody>
</table>

| Triglycerides: APOE (familial dysbetalipoproteinemia) |
| CLCN8 (X-linked recessive nephrolithiasis), NKCC2 (Bartert syndrome) |

*Numbers in this column indicate references listed at the end of the chapter.*
example of the importance of history can be drawn by examining type 1 diabetes rates across the Scandinavian region of Karelia. In 1940 this region was divided between Finland and the former Soviet Union with little contact between the two sections over the next 60 years. Finnish Karelians have a sixfold increased rate of type 1 diabetes compared to Russian Karelians.\(^5\) As a result, heritability for type 1 diabetes will be different when estimated in the combined Karelian populations than when estimated in Finnish or Russian Karelians alone. The difference in diabetes rate is likely due to environmental factors, because both Karelian populations recently originated from a common ancestry and therefore likely have similar genetic risk factors for type 1 diabetes.\(^6\)

**Human DNA Sequence Variation: Molecular Forms and Biologic Effects**

Each human has two versions of his or her genome (one from each parent); each version consists of a sequence of approximately 3 billion DNA bases. When comparing two versions of the human genome, either within the same person or between two different people, about 1/1000 of these bases vary (that is, 99.9% of them are the same) (Table 4-2). There are many possible ways in which DNA sequences can vary; several specific types of DNA sequence variants are frequently observed (Fig. 4-1).

The most frequent form of variation, the single nucleotide polymorphism (SNP), refers to the situation in which a single base in the sequence of one individual is different from the base seen at the same position in the sequence of another individual. SNPs can exert a wide range of biologic effects, depending on where the variant occurs and whether it alters the function of the DNA sequence. Some SNPs occur within the portions of genes that are transcribed into RNA and then translated into proteins (protein-coding regions). Synonymous SNPs occur in the protein-coding portion of DNA but both versions (alleles) of the SNP encode the same amino acid, and so this sort of variation usually does not affect function. SNPs can be missense changes (alteration of a single amino acid in a protein-coding gene) as is the case of the C282Y mutation in the HFE gene responsible for autosomal recessive hereditary hemochromatosis (Chapter 19). Some missense SNPs greatly alter function, whereas others appear to have no consequences. SNPs can also alter splice sites, disrupting the structure of the mRNA that is transcribed from the DNA during gene expression. For example, the most common cause of autosomal dominant isolated growth hormone (GH) deficiency is single-base mutations that inactivate a splice donor site of intron 3 in the GH1 gene, causing skipping of exon 3 in GH1 (Chapter 24). SNPs can also introduce stop codons, leading to premature termination of translation and a truncated protein product. These nonsense variants typically dramatically impair or eliminate the function of the protein.

Changing the protein sequence is not the only way that SNPs (and other types of genetic variations) can alter gene function. Most of the human genome does not code for proteins (see Table 4-2) and most genetic variation occurs in this noncoding portion of the genome. For example, noncoding variants can alter the level, timing, or location of gene expression, without changing the sequence of the encoded protein. Noncoding variants often result in more subtle biologic effects, and the mechanisms are still being uncovered. For example, some SNPs subtly influence type 1 diabetes risk and lie in enhancers (noncoding DNA segments that activate gene transcription at a distance) that appear to affect gene expression only in lymphoid cells.\(^9\)

Insertions and deletions (collectively called indels) refer respectively to the addition or removal of one or more bases in the DNA sequence. Indels in protein-coding sequences are called frameshift mutations, as long as the number of bases inserted or deleted is not a multiple of three. Because the genetic code is composed of triplets (every three bases encode one amino acid), a frameshift mutation alters how every subsequent base in the sequence is translated into a protein, resulting in profound molecular and clinical consequences. For example, classic salt-wasting CAH is often caused by frameshift deletions in the CYP21A2 gene that ablute its function (Chapter 23). Repeat polymorphisms (often referred to as copy number variants, or CNVs, if the repeats are large) are a special case of indels in which DNA sequences are repeated in tandem and the number of copies of the repeated sequence varies. For example, the AR gene (encoding the androgen receptor) contains a repeat polymorphism in which a CAG codon, encoding glutamine, is repeated 11 to 31 times (Chapter 23). Structural variation can include both insertions and deletions as well as rearrangement of large chunks of DNA sequence (translocations and other complex forms of genomic variation). Structural variation causes familial hyperaldosteronism type 1; the adrenocorticotrophic hormone (ACTH, corticotropin)-responsive promoter of the CYP11B1 gene is incorrectly located adjacent to the aldosterone synthase gene (CYP11B2), causing aldosterone to be produced by ACTH stimulation (Chapter 16).

**Factors Influencing the Biologic Impact of Genetic Variants in a Particular Gene**

As discussed previously, the impact of a genetic variant on gene function will depend on the type of variant and its location with respect to the gene. For example, frameshift deletions in the CYP21A2 gene completely eliminate 21-hydroxylase activity, whereas missense variants in CYP21A2 often retain partial 21-hydroxylase activity (Chapter 23). However, even a single, specific variant may

---

**TABLE 4-2**

**Characteristics of Human Genome Sequence Variation**

<table>
<thead>
<tr>
<th>Characteristic</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>Length of the human genome sequence (base pairs)</td>
<td>3 billion</td>
</tr>
<tr>
<td>Number of human genes (estimated)</td>
<td>20,000</td>
</tr>
<tr>
<td>Fraction of base pairs that differ between the genome sequence of a human and a chimpanzee</td>
<td>1.3% (1 in 80)</td>
</tr>
<tr>
<td>Fraction of base pairs that vary between the genome sequence of any two humans</td>
<td>0.1% (1 in 1000)</td>
</tr>
<tr>
<td>Fraction of coding region base pairs that vary in a manner that substantially alters the sequence of the encoded protein</td>
<td>0.02% (1 in 5000)</td>
</tr>
<tr>
<td>Number of sequence variants present in each individual as heterozygous sites</td>
<td>3 million</td>
</tr>
<tr>
<td>Number of amino acid-altering variants present in each individual as heterozygous sites</td>
<td>12,000</td>
</tr>
<tr>
<td>Number of sequence variants in any given human population with frequency of &gt;1%</td>
<td>10 million</td>
</tr>
<tr>
<td>Number of amino acid polymorphisms present in the human genome with a population frequency of &gt;1%</td>
<td>75,000</td>
</tr>
<tr>
<td>Fraction of all human heterozygosity attributable to variants with a frequency of &gt;1%</td>
<td>98%</td>
</tr>
</tbody>
</table>

patterns are called correlated. Although these six polymorphisms could theoretically occur in 26 possible patterns, only three patterns are observed (indicated by pink, orange, and green). These of variation shown here is typical for a 5-kb stretch of genome and is centered on a strong recombination hotspot. The 12 common variations include 10 single nucleotide polymorphisms because a hotspot of genetic recombination lies between them. In addition to the common polymorphisms, lower frequency polymorphisms occur in the human genome. Five rare SNPs are shown, with the variant nucleotide marked in red and the reference nucleotide not shown. In addition, on the second to last chromosome, a larger deletion variant is observed that removes several kilobases of DNA. Such larger deletion or duplication events (i.e., copy number variants [CNVs]) may be common and segregate as other DNA variants. (Redrawn from Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. 2008;322(5903):881-888.)

Figure 4-1 DNA sequence variation in the human genome. Common and rare genetic variation in 10 individuals, carrying 20 distinct copies of the human genome. The amount of variation shown here is typical for a 5-kb stretch of genome and is centered on a strong recombination hotspot. The 12 common variations include 10 single nucleotide polymorphisms (SNPs), an insertion-deletion polymorphism (indel), and a tetranucleotide repeat polymorphism. The six common polymorphisms on the left side are strongly correlated. Although these polymorphisms could theoretically occur in 26 possible patterns, only three patterns are observed (indicated by pink, orange, and green). These patterns are called haplotypes. Similarly, the six common polymorphisms on the right side are strongly correlated and reside on only two haplotypes (indicated by blue and purple). The haplotypes occur because there has not been much genetic recombination between the sites. By contrast, there is little correlation between the two groups of polymorphisms because a hotspot of genetic recombination lies between them. In addition to the common polymorphisms, lower frequency polymorphisms occur in the human genome. Five rare SNPs are shown, with the variant nucleotide marked in red and the reference nucleotide not shown. In addition, on the second to last chromosome, a larger deletion variant is observed that removes several kilobases of DNA. Such larger deletion or duplication events (i.e., copy number variants [CNVs]) may be common and segregate as other DNA variants. (Redrawn from Altshuler D, Daly MJ, Lander ES. Genetic mapping in human disease. Science. 2008;322(5903):881-888.)

have different effects in different individuals. The effect of any given genetic variant (genotype) on phenotype can be modified by variants in other genes (gene-gene interactions) or by environmental factors (gene-environment interactions) or by random chance. It is usually not possible to measure or quantify these factors in any one person, but their combined effect can be quantified on a population level as penetrance, the proportion of individuals carrying a genetic variant who exhibit the phenotype. The penetrance of a genetic variant is highly context-dependent with respect to phenotypic definition. For example, the hemochromatosis-associated C282Y allele in the HFE gene exhibits high penetrance for the biochemical phenotype of high ferritin (>60% of homozygous carriers manifest increased ferritin levels) but only 2% penetrance for the clinical phenotype of liver cirrhosis. Temporal context is also an important consideration, as disease incidence often increases with age. Carriers of mutations causing MEN1 have nearly 100% penetrance for parathyroid adenomas by age 40 but only 20% penetrance at age 20.

A common observation in members of a family carrying the same disease-causing genetic variant is that not all members of the family are equally affected. This range of phenotypic expression resulting from a particular genotype is referred to as variable expressivity and, as with penetrance, arises from the range of impacts of specific variants as well as modifying influences of genetic background
(gene-gene interactions), environment (gene-environment interactions), and random chance. For example, the same mutation in the androgen receptor (AR, encoding an $703G$ substitution) resulted in a spectrum of clinical androgen insensitivity such that some individuals were raised as 46,XY females and others as males; other mutations in AR have different ranges of phenotypic effects (Chapter 23).

Mosaicism, whereby cells within a single individual have different genotypes, is another mechanism that leads to variable expressivity. Most mutations known to influence disease are germ line mutations—inherit ed from the sperm or egg and present in every cell—but some diseases can be caused by somatic mutations that occur after fertilization and are present in only some cells, leading to mosaicism. In these cases, which tissues or organs carry the mutation will influence the clinical outcome. The most familiar class of disease caused in large part by somatic mutations is neoplasia, including endocrine tumor syndromes such as Conn syndrome and Cushing disease.

Another classic example from endocrinology is the McCune-Albright syndrome, in which the same activating mutation in the GNAS1 gene exhibits variable expressivity because of postzygotic mosaicism. The phenotype of patients with McCune-Albright syndrome depends on which tissues and what fraction of cells carry the GNAS1 mutation. A minority of affected individuals (24%) display the classic triad (minority of affected individuals (24%) display the classic triad) of postzygotic mosaicism. The mechanism of variable expressivity likely maps to the zygotic stage in which the mutation arose: a mutation earlier in embryogenesis is present in more tissue lineages. Because mutations in a mosaic individual are not present in every cell, they can be hard to detect in DNA isolated from a blood sample if the cell in which the mutation occurred does not give rise to blood leukocytes. The GNAS1 mutation responsible for the McCune-Albright syndrome is detected in only 8% to 46% of blood samples from affected individuals but is found in 90% of affected tissue sampled irrespective of clinical presentation (Chapter 25). Conversely, blood cells can contain somatic variation that is absent in other tissues or the germline.

It is important to remember that the base pair composition of a DNA sequence is not the only molecular determinant of phenotypic expression (Table 4-3). DNA is subject to other forms of modification besides sequence variation (termed epigenetic variation), such as cytosine methylation or packaging into nucleosomes with various biochemically modified histones, each of which can alter gene expression and function. Thus, the same molecular form of DNA sequence variation can vary in its cellular and phenotypic effect through epigenetic modifications. Indeed, epigenetic modification is a normal part of development and is the reason why different cells have different properties even though they share the identical DNA sequence. A striking example of the effect of epigenetics is imprinting, the expression of a genetic variant in a parent-of-origin specific manner. For paternally imprinted genes, the copy that is inherited from the father is silenced, and only the mother’s copy is expressed in the offspring. Imprinting can affect the impact of disease-causing mutations. Inactivating mutations in SDHID cause familial paraganglioma type 1 (Chapter 16). SDHID is maternally imprinted, so the mutation does not cause disease when inherited from the father but is highly penetrant when inherited from the mother. Imprinting can also be tissue-specific. A paternally inherited inactivating mutation in GNAS1 causes Albright hereditary osteodystrophy (AHO, pseudopseudohypopara-

The type of genetic variant (missense, frameshift, noncoding, etc.) provides clues to its possible consequences; in addition, the population frequency of a variant, whether it is common or rare, can also provide information about its likely impact on phenotype. The relative balance between common and rare genetic variation is strongly influenced by evolution and human demographic history. Modern humans likely originated from a small population residing in Africa that had been evolving over millions of years. Within the past 50,000 years, members of this ancestral population migrated “out of Africa,” settled the globe, and only recently, over the past 5000 to 10,000 years, multiplied exponentially. As a consequence of this demographic history, most of the 3 million genetic variants an individual inherits from his or her parents are common (typically >1% frequency in the population), can be traced back to the ancient African population, and are shared in many unrelated individuals in the population. Individuals also inherit thousands of genetic variants unique to themselves and their relatives. These rare genetic variants arose more recently from spontaneous mutation in the past 10 millennia, after the migration of many humans out of Africa, and are typically observed infrequently (<0.1% of all chromosomes) in the population.

Evolution influences the frequency of variants that affect human phenotypes (such as endocrine diseases) through the process of natural selection. Variants that greatly increase the risk of a disease that is deleterious from a reproductive standpoint are less likely to be passed on to offspring and will be rare in the population (unless they have a compensatory benefit, such as malaria resistance in carriers of sickle cell disease). If a disease is at mildly evolutionary deleterious, then the genetic variants associated with that disease will only modestly increase disease risk. This is because those common variants, if they had strongly increased disease risk, would have then been subject to strong negative evolutionary selection and never would have risen in frequency to become common in the first place. By contrast, it is more plausible for rare/recent variants to exert strong effects on phenotype and greatly increase disease risk.

Finally, the number of genes that contribute to disease in a single individual ( mendelian or polygenic disease) will be related to the strength of effect of any one variant on disease risk. By definition, variants that cause mendelian disorders have strong effects, whereas variants contributing to risk of polygenic diseases will typically have more modest effects. Thus, most variants with strong effects on disease will be rare, especially for those diseases that are clearly deleterious from an evolutionary standpoint (lethal before reproductive age). By contrast, common polygenic diseases and traits will have a much more substantial contribution from common genetic variants, although these considerations do not rule out an important role for rarer variants in polygenic phenotypes. As we will see later in this chapter, these patterns of genetic variation have important implications for identifying genetic variants that underlie disease and also for interpreting the impact of genetic variation on disease.

<table>
<thead>
<tr>
<th>Table 4-3</th>
<th>Origins of DNA Sequence Variation in Human Populations: Common Versus Rare Variants</th>
</tr>
</thead>
<tbody>
<tr>
<td>The type of genetic variant (missense, frameshift, noncoding, etc.) provides clues to its possible consequences; in addition, the population frequency of a variant, whether it is common or rare, can also provide information about its likely impact on phenotype. The relative balance between common and rare genetic variation is strongly influenced by evolution and human demographic history. Modern humans likely originated from a small population residing in Africa that had been evolving over millions of years. Within the past 50,000 years, members of this ancestral population migrated “out of Africa,” settled the globe, and only recently, over the past 5000 to 10,000 years, multiplied exponentially. As a consequence of this demographic history, most of the 3 million genetic variants an individual inherits from his or her parents are common (typically &gt;1% frequency in the population), can be traced back to the ancient African population, and are shared in many unrelated individuals in the population. Individuals also inherit thousands of genetic variants unique to themselves and their relatives. These rare genetic variants arose more recently from spontaneous mutation in the past 10 millennia, after the migration of many humans out of Africa, and are typically observed infrequently (&lt;0.1% of all chromosomes) in the population.</td>
<td></td>
</tr>
<tr>
<td>Evolution influences the frequency of variants that affect human phenotypes (such as endocrine diseases) through the process of natural selection. Variants that greatly increase the risk of a disease that is deleterious from a reproductive standpoint are less likely to be passed on to offspring and will be rare in the population (unless they have a compensatory benefit, such as malaria resistance in carriers of sickle cell disease). If a disease is at mildly evolutionary deleterious, then the genetic variants associated with that disease will only modestly increase disease risk. This is because those common variants, if they had strongly increased disease risk, would have then been subject to strong negative evolutionary selection and never would have risen in frequency to become common in the first place. By contrast, it is more plausible for rare/recent variants to exert strong effects on phenotype and greatly increase disease risk.</td>
<td></td>
</tr>
<tr>
<td>Finally, the number of genes that contribute to disease in a single individual ( mendelian or polygenic disease) will be related to the strength of effect of any one variant on disease risk. By definition, variants that cause mendelian disorders have strong effects, whereas variants contributing to risk of polygenic diseases will typically have more modest effects. Thus, most variants with strong effects on disease will be rare, especially for those diseases that are clearly deleterious from an evolutionary standpoint (lethal before reproductive age). By contrast, common polygenic diseases and traits will have a much more substantial contribution from common genetic variants, although these considerations do not rule out an important role for rarer variants in polygenic phenotypes. As we will see later in this chapter, these patterns of genetic variation have important implications for identifying genetic variants that underlie disease and also for interpreting the impact of genetic variation on disease.</td>
<td></td>
</tr>
</tbody>
</table>
To summarize this introductory section, we have briefly described a number of basic principles of genetics. Heritability describes the proportion of a disease/trait that can be explained by genetic factors; the heritability of most endocrine diseases ranges between 20% and 80% (see Table 4-1). Genetic variants can take many forms ranging from single-base changes (SNPs) to translocations of entire chromosomes (see Fig. 4-1). The biologic effect of these variants depends on the type of variant, where in the DNA they are located (e.g., within coding sequence, splice sites, enhancers), how severely the variant affects function, and for somatic mutations, the cells and tissues that carry the mutation. Biologic impact can also be modified by the presence of genetic variants in other genes (gene-gene interactions), the individual organism’s environment (gene-environment interactions), and random chance. The demographic history of modern human populations explains the presence of common and rare genetic variants in the human genome (see Table 4-2). Common variants are mostly ancient, and typically have relatively modest clinical effects, whereas rare variants tend to have arisen more recently and can exert larger clinical effects (Table 4-4).

### GENETICS OF ENDOCRINE DISEASES

As described earlier, heritable diseases and traits, including endocrine phenotypes, span a range of genetic architectures ranging from single-gene mendelian disorders to common, polygenic diseases and traits. Mendelian and polygenic disorders represent two ends of a spectrum (see Fig. 4-2) of genetic architectures. While we distinguish between these two extremes of genetic architecture, it is important to appreciate that many disorders lie between these two extremes: rare variants of moderate effect can affect the common form of the disease, and genetic and nongenetic modifiers can strongly influence the outcome of mendelian disorders. Furthermore, many polygenic endocrine disorders also have rare mendelian forms (see Table 4-1).

The genes for a wide range of mendelian endocrine diseases have been mapped, revealing great mechanistic insight. Although mendelian diseases have offered valuable insights into pathophysiology, not all insights gained from mendelian forms of disease translate directly to the common forms of disease. For example, mendelian obesity caused by recessive inactivating mutations in the leptin receptor could be well treated by exogenous leptin, but this clinical insight did not apply to most obese individuals who actually demonstrate elevated leptin levels and do not respond to exogenous therapy with leptin (Chapter 36). Obesity as a common trait is highly heritable (heritability 40-80%), and genome-wide association studies (GWAS) analysis has begun to identify risk variants for the common form. Although some of the risk variants overlap with those causing mendelian syndromes (as is also true for other diseases), GWAS have pointed to additional genetic contributions outside the mendelian genes. And, of course, the variants that have strong effects on quite rare genetic syndromes do not explain much, if any, of the risk of the common forms of disease. Thus, genetics of both mendelian forms and common polygenic forms will have important, often complementary impacts on our understanding of disease and on patient care.

The sections that follow discuss representative examples of mendelian and polygenic endocrine disorders that illustrate important concepts in gene discovery, understanding of the impact of genetic variation on disease, and implications for clinical care and insights into new biology. We discuss several classes of mendelian diseases and highlight three polygenic endocrine diseases/trait: (1) type 2 diabetes, (2) stature, and (3) serum lipids. In each section, we discuss what is known about the underlying genetic

---

### TABLE 4-4

**Performing and Interpreting Genetic Studies**

For any heritable disease, the success of genetic mapping efforts, the strategy employed, and the clinical utility of any resulting genotype-phenotype map depend on its genetic architecture: (1) the number of genetic variants/genes, (2) their frequency in the population, and (3) their respective contributions to risk (i.e., penetrance). On one end of the spectrum lie mendelian diseases, such as multiple endocrine neoplasia type 1 (MEN1), characterized by (1) few variants often in a single gene, (2) extremely rare frequency in the population (<1:1,000), and (3) high penetrance (>50-fold risk). On the other end of the spectrum lie the so-called common diseases, such as type 2 diabetes, characterized by (1) many variants in many genes (polygenic), (2) frequency in the population (>1:20), and (3) low penetrance (<1:5-fold risk) (Fig. 4-2).

Owing to their simple genetic architectures, mendelian endocrine disorders were ideally suited for genetic mapping using the techniques of familial linkage mapping developed in the 1980s. Because they are rare and have strong effects on phenotype, mendelian variants were typically identified in families. As a result, the genotype-phenotype correlations for these variants could not be generalized to the population at large. For example, penetrance of mendelian variants could be accurately estimated only if these variants were ascertained in the general population, rather than in selected families with a specific genetic background. Large-scale sequencing studies in the general population, which can identify all variants, rare and common, are now enabling such estimates. Such studies have found that, when ascertained in the general population, the so-called mendelian variants are less penetrant that was estimated from family-based studies. By contrast, the variants for common, polygenic disorders have been identified through genetic association studies in the general population. Genetic association studies do not require the identification of rare families segregating disease because they simply compared the frequency of a given genetic variant in disease cases and controls. Thus, they can be applied to identify genetic factors underlying diseases occurring in a population of unrelated individuals (i.e., common diseases). Unlike clinical risk factors/biomarkers association studies, correlation in genetic association studies implies causation because genotype always precedes phenotype. Through the 1980s, genetic association studies were performed using single nucleotide polymorphisms (SNPs) at candidate genes selected by educated guessing. Such studies yielded a number of common disease associations but were poorly reproducible and confounded by false-positive results arising from population stratification. The development of modern sequencing and genotyping technologies along with the cataloging of over 10 million common variants (the International HapMap project) enabled genome-wide association studies (GWAS), a systematic approach to simultaneously test all genes for associations that could account for population-based confounding. GWAS have yielded a large number of reproducible genetic associations for diverse common/polygenic diseases/trait, yielding insight into disease biology and genetic architecture.

When interpreting a result from any genetic study, it is important to bear in mind that the actual variant (usually a SNP) tested in the study marks a haplotype (a combination of genetic variants inherited together) that can span millions of bases. The causal variant, in the sense that it is molecularly responsible for alteration in gene function leading to cellular and disease phenotype, may lie anywhere on this haplotype. As with the chromosomal linkage studies of the past, identifying the causal variants/genes on a haplotype necessitates a combination of further association analysis (fine-mapping) and functional experimentation in model systems.
contributors, the impact of genetics on our understanding of disease biology, and the translation into clinical care in the short and long term.

**Mendelian Endocrine Diseases**

**Genetic Architecture**

Mendelian diseases represent one extreme of a spectrum of possible genetic architectures (see Fig. 4-2). The alleles causing mendelian diseases are found in a small number of genes, are typically rare (<1:1000), are highly penetrant, and follow simple patterns of dominant and recessive inheritance. They are considered monogenic in that a mutation in a single gene causes disease in an individual or family. But as different families segregating the same mendelian disease are identified and the causal genetic variants mapped, genetic heterogeneity is often observed: different alleles in different genes can cause the same disease. Some mendelian disorders (e.g., MEN2) demonstrate recurrent mutations in the same gene, but of different molecular nature. In other disorders (e.g., familial paraganglioma) multiple genes across different chromosomes are implicated, each causing the same/similar disease in different individuals. This phenomenon, variants in different genes causing the same disease, is termed locus heterogeneity. It is important to bear in mind that locus heterogeneity is intrinsically tied to the precision of disease definition. For example, CAH can be caused by defects in multiple genes encoding steroid biosynthetic enzymes (CYP21A2, CYP11B1, CYP17A1, HSD3B2, POR, StAR; Chapter 15). However, the CAH phenotype is refined to include biochemical measurements (mineralocorticoid, sex hormone, and electrolyte levels), individual subtypes emerge, each of which possesses a simpler genetic architecture (i.e., decreased locus heterogeneity).

When contrasted with common polygenic diseases, mendelian disorders exhibit relatively less locus heterogeneity. In other words, an appreciable fraction of mendelian disease cases can be largely explained by mutations in one or a few genes. For example, recurrent mutations in a single gene (the eponymous MEN1) account for 70% of families segregating the MEN1 clinical syndrome. Even in this classic mendelian case, however, the genetic architecture remains incompletely defined, as 30% of cases have no mutation in MEN1. Thus, much of the genetic architecture of mendelian diseases remains uncharted territory for genetic mapping. Modern sequencing technologies have facilitated a renaissance in mendelian disease gene mapping and will help improve our understanding of the genetic basis of mendelian disorders. For example, by exome sequencing two individuals in a kindred with familial combined hypolipidemia (Chapter 37), investigators identified two nonsense mutations in ANGPTL3 that segregated with low serum lipoproteins when genotyped in other family members. These mutations and the ANGPTL3 gene were contained in the region identified by traditional linkage mapping and could be quickly identified because the sequence of all exons in that region had been determined.

**Disease Biology**

Every endocrine organ ranging from the pituitary to the adrenal is affected by well-described and less-described mendelian disorders. Mechanistic insight into disease biology has been gained from discovering the identities of the genes that lead to disease. When mutations in a number of different genes can all cause a disease (locus heterogeneity), additional mechanistic insight into molecular pathophysiology becomes possible. This makes intuitive sense in the context of a molecular understanding of genes as encoding proteins that act in concert to accomplish cellular functions. For example, Noonan syndrome (characterized endocrinologically by variable short stature, delayed puberty, and cryptorchidism in the setting of dysmorphic features and variable cardiac defects; Chapter 23) is typically caused by activating the RAS-MAPK (mitogen-activated protein kinase) signaling pathway. Dominant gain-of-function mutations in multiple pathway members (PTPN11, SOS1, KRAS, RAF1, BRAF, NRAS) have all been shown to cause Noonan syndrome. For other disorders, a more complex picture emerges in which multiple molecular pathways are implicated. For example, Kallmann syndrome (Chapter 25), which arises from failure of migration of GnRH neurons during fetal development, demonstrates X-linked (KAL1), autosomal dominant (FGFR1), and autosomal recessive (PROK2) inheritance. The gene product of KAL1, a secreted protein called anosmin, is thought to interact with the fibroblast growth factor (FGF) receptor,
whereas the gene product of PROK2, the secreted protein prokineticin 2, interacts with a different receptor. Both signaling pathways are required for GnRH neuronal migration.

At the level of a single gene/locus, genotype-phenotype correlations mapping allelic heterogeneity to phenotypic heterogeneity can provide detailed insight into how alterations in gene function affect disease severity. CAH caused by CYP21A2 deficiency is a classic case. A number of genetic variants including frameshifting deletions, splice site alterations, and missense mutations have been identified (Chapter 23) in CYP21A2. This spectrum of alleles has been mapped to a biochemical spectrum of 21-hydroxylase enzyme activity, which in turn maps to a spectrum of clinical features along the axes of mineralocorticoid sufficiency, androgen excess, and ACTH elevation (Chapter 23). In this disorder, it is possible to make predictions about clinical phenotype (categorized as salt-wasting, simple virilizing, and nonclassic) based on genotype. Notably, the positive predictive value (PPV, the strength of the genotype-phenotype correlation) is strongest for variants that severely affect CYP21A2 gene function and are predicted to cause severe disease (salt-wasting, PPV ~100%). Predictive power is weaker for genetic variants that are expected to have more moderate effects on gene function and therefore result in milder disease (nonclassic, PPV ~60%). Some of this complexity is due to the potentially compensatory 21-hydroxylase enzyme activity of CYP22C9 and CYP3A4, a form of genotype interaction.

Genotype-phenotype correlations must be established empirically and are not possible in many cases. Even when mutations of varying molecular severity are identified, they may not predictably affect phenotype. For example, many individuals with variant genetics in the SRD5A2 gene (encoding 5α-reductase) of differing molecular severity and location have been identified (Chapter 23), but no correlation between the genotype and the clinical degree of virilization is apparent.

The genetic causes of many mendelian disorders remain unknown, but advances in sequencing technology have accelerated the pace of discovery. The identification of gain-of-function mutations in the KCNJ5 and CACNA1D genes is an example of a gain-of-function variant affecting the potassium channel, KCNJ5, which is involved in the regulation of the electrochemical gradient for potassium ions and the maintenance of the resting membrane potential. This protein is thought to play a role in the control of cell proliferation and differentiation. The KCNJ5 gene is involved in the development of various disorders such as congenital hypothyroidism, neonatal diabetes, and neonatal severe hyperparathyroidism. The KCNJ5 gene is located on chromosome 6q23.3 and is composed of 18 exons. The protein encoded by KCNJ5 is a voltage-gated potassium channel subunit, and mutations in this gene lead to a range of clinical manifestations, including hypothyroidism, diabetes, and hyperparathyroidism. The KCNJ5 gene is involved in the development of various disorders such as congenital hypothyroidism, neonatal diabetes, and neonatal severe hyperparathyroidism. The KCNJ5 gene is located on chromosome 6q23.3 and is composed of 18 exons. The protein encoded by KCNJ5 is a voltage-gated potassium channel subunit, and mutations in this gene lead to a range of clinical manifestations, including hypothyroidism, diabetes, and hyperparathyroidism.

In the case of congenital hyperinsulinism, autosomal recessive mutations in ABCC8 and KCNJ11 genes have been identified. These genes encode theSURF1 and SURF2 proteins, respectively, which are subunits of the SURF1/SURF2 complex, a large ATP-binding cassette (ABC) transporter complex. Mutations in these genes lead to a disruption in the function of the SURF1/SURF2 complex, resulting in increased insulin secretion from pancreatic beta cells and the development of hyperinsulinism.

Clinical Translation

Target discovery, risk prediction, and the tailoring of pharmacotherapy based on genotype are potential clinical applications of genotype-phenotype correlation. The existence of loss-of-function and gain-of-function variants in an allelic series and their concordance with opposing phenotypes can provide a rationale for therapeutically modulating gene function. For example, inactivating mutations in the KISS1R receptor cause hypogonadotropic hypogonadism, whereas an Arg386Pro missense variant in KISS1R is associated with central precocious puberty. Kisspeptin, the agonist ligand of the KISS1R receptor, has shown promise as a fertility treatment. Genotype-phenotype correlation can be used to predict risk of disease in asymptomatic carriers. Prior to the identification and cloning of the RET proto-oncogene, MEN2 kindreds were monitored for evidence of medullary thyroid cancer (MTC) by calcitonin stimulation tests. Once mutations in RET were established as causing MEN2A/B and familial MTC, it became apparent that specific mutations could be mapped to the different syndromes. The RET gene product encodes a cell surface receptor tyrosine kinase. Mutations in the extracellular domain predispose to MEN2A (characterized by MTC, pheochromocytomas, and hyperparathyroidism), whereas mutations in the intracellular tyrosine kinase domain predispose to MEN2B (characterized by MTC, pheochromocytomas, and mucosal neuromas). The clinical aggressiveness of MTC, the sine qua non of all three syndromes, is greatest in MEN2B, then less in MEN2A, with familial MTC demonstrating the least propensity to grow and metastasize. A well-defined genotype-phenotype correlation between specific RET mutations and clinical aggressiveness of MTC now dictates the timing of lifesaving prophylactic thyroidectomy in carriers of RET mutations. The key to establishing clinically robust risk prediction based on genotype-phenotype correlations is a well-differentiated allelic series derived from multiple individuals/families. The consensus genotype-phenotype correlation for prophylactic thyroidectomy in RET mutation carriers was derived from analysis of over 200 individuals from over 100 families (Chapter 39).

A genetic diagnosis in a number of mendelian disorders can also directly inform pharmacotherapies. A classic example includes obesity caused by leptin deficiency (Chapter 36), which can be treated by exogenous leptin injections. Other examples include HNF1A MODY (maturity-onset diabetes of the young) and neonatal diabetes (discussed in detail later) caused by genes whose properties predict excellent response to sulfonylureas. In the case of congenital hyperinsulinism, autosomal recessive mutations in ABCC8 and KCNJ11 genes have been identified. These genes encode the SURF1 and SURF2 proteins, respectively, which are subunits of the SURF1/SURF2 complex, a large ATP-binding cassette (ABC) transporter complex. Mutations in these genes lead to a disruption in the function of the SURF1/SURF2 complex, resulting in increased insulin secretion from pancreatic beta cells and the development of hyperinsulinism.
or KCNJ11 correlate with diffuse disease on spectroscopic imaging and lack of responsiveness to medical therapy (diazoxide); such individuals require near-total pancreatectomy for control of hyperglycemia.26

**Type 2 Diabetes**

**Genetic Architecture**

Type 2 diabetes (T2D) is a multifactorial, polygenic disorder for which manifestation depends on multiple interacting genetic and environmental risk factors. Heritability estimates show strong evidence of familial clustering, ranging from 49% to 80%.27 Approximately 5% of diabetes cases that may be classified as type 2 (nonautoimmune) arise from a single gene disorder, follow mendelian patterns of inheritance, and cluster into clinically defined syndromes. These mendelian diabetes syndromes include neonatal diabetes (Chapter 31), MODY (Chapter 31), and congenital lipodystrophies (Chapter 31).27 To date, familial linkage studies have successfully implicated approximately 30 genes as monogenic causes of diabetes.

The genes underlying the majority of T2D cases (95% of cases) fit a polygenic model; genetic variants in multiple genes independently contribute to disease risk, each with a modest effect. The partial elucidation of these genetic risk factors required the advent of genetic association studies/GWAS and the assembly of cohorts of hundreds of cases and control subjects.32 As of 2014 about 70 loci have been identified from aggregated analysis on about 150,000 case controls. Taken together, these loci account for about 6% of heritability for T2D.33 Of these loci, an SNP at TCF7L2 (with the risk-increasing allele present at a frequency of ~30%) has the largest overall effect on risk, conferring a 1.4-fold increase in risk per allele.33 By contrast, type 1 diabetes shows a somewhat different genetic architecture, with common loci of large effect (a variant at the HLA locus found in 61% of the population confers a fivefold increase in risk,36 and a common variant at the insulin gene confers a threefold increase in risk). This finding was consistent with prior studies from the 1980s estimating that 50% of the heritability of type 1 diabetes was explained by common haplotypes at the HLA loci.37

Notably, the genes implicated in monogenic causes of diabetes also contribute to polygenic forms, but through distinct genetic variants. Genes associated with mendelian diabetes syndromes, such as KCNJ11 (neonatal diabetes, Chapter 31), HNF1A (MODY2, Chapter 31), and PPARG (familial partial lipodystrophy 3),33 were found to harbor common variants that conferred risk for common T2D.27,34 Conversely, genes first found associated with diabetes through GWAS have subsequently been identified as having rare, highly penetrant alleles. For example, noncoding common variants pointed to the MTNR1B gene (encoding the melatonin receptor) as a T2D-associated locus (1.15-fold risk).35 Subsequently, large-scale resequencing studies identified multiple, rare coding variants of the same gene (present in <1:1000 individuals) that conferred a greater than fivefold increased risk of T2D.36 Mapping studies performed across various populations reveal both similarities and differences in the genetic risk factors contributing to diabetes among different ancestral groups. T2D GWAS performed in multiple populations/ancestries (European, South Asian, East Asian, Latino, African American)37 reveal that many common variants are shared across populations with equivalent effects on disease risk, regardless of ancestry. This pattern is consistent with the origin of most common variants in an ancestral African population (see Table 4-3), but remarkable ancestry-specific effects have also been identified. A T2D GWAS performed in individuals of Latino and Mexican ancestry identified a common SNP at a locus containing the genes SLC16A11/13 that confers a 1.25-fold increased risk of diabetes.38 The same locus was identified by another T2D GWAS performed in Japanese individuals.39 Because the associated SNPs were rare in Europeans, the locus had not been detected in GWAS in European-ancestry populations. Similarly, a common variant in TBC1D4 in individuals from Greenland (present in 17% of Greenland’s population) strongly increases risk of T2D (10-fold increased risk).40 This variant, which causes a premature truncation and associates with elevated muscle insulin resistance, is extremely rare in continental Europe and likely became common in Greenland because it was present in founding ancestors of Greenland’s current population.

In summary, genetic mapping studies over the past 3 decades have revealed a genetic architecture for T2D with widespread locus and allelic heterogeneity. With regard to effect size and allele frequency, T2D genetic architecture so far consists of some very rare variants of large effect, some common variants with modest effects, and a large number of common variants with even more modest effects on disease risk, with rare and common genetic variants spread out across multiple loci in the genome. This genetic architecture has proved to be typical for other common diseases41 (see Fig. 4-2) and reflects both the underlying genetic architecture of the disease and the reality of large GWAS to more readily detect common variants of modest effect.

**Disease Biology**

The past three decades of genetic discoveries in T2D have nucleated a molecular understanding of disease mechanisms, highlighted differences between glycemia and T2D, and implicated previously unknown physiology in disease pathogenesis.

Supporting the current physiologic conception of T2D as a disorder of decreased insulin production as well as decreased insulin sensitivity, genetic mapping has pointed to a molecular basis for both axes. Prediabetic individuals harboring T2D-associated variants (SLC30A8, HNF1A) and cell survival genes (CDKL1) manifest with decreased insulin secretion (homeostatic model assessment [HOMA]-B; Chapter 31).42 On the other hand, prediabetic individuals harboring T2D-associated variants in adipocyte genes (PPARG, KLF14) tend toward increases in insulin resistance (HOMA-IR; Chapter 31).42 About 30% of T2D-associated SNPs point to insulin secretion/beta-cell function, and 15% point to insulin resistance.43 Interestingly, the SNPs associating with insulin secretion in prediabetic individuals predict incident T2D but the SNPs associating with insulin resistance do not.41 These findings from genetic epidemiology are consistent with beta-cell failure being a final common pathway for manifestation of hyperglycemia and diagnosis of T2D. Importantly, over half of associated SNPs and the genes they point to cannot be connected with either insulin secretion or sensitivity. Their pathogenic mechanisms remain to be elucidated by physiologic and functional investigation.

Even without a full understanding of their molecular/cellular mechanism of causation, the large number of T2D-associated loci (~70 as of 2014) have been deployed to refine disease classification. By examining quantitative glycemic traits (insulin production, sensitivity, processing, and fasting glucose) in nondiabetic individuals genotyped for 37 T2D-associated common genetic variants, investigators were able cluster genes with glycemic traits to define
unique diabetes subtypes. For example, individuals harboring variants in MNT1B and GCK manifested a combination of fasting hyperglycemia and decreased insulin secretion, whereas those harboring variants in SLC30A8, CDKN2A/B, TCF7L2, and other genes manifested primarily with decreased insulin secretion. Notably, many genes did not cluster with predefined glycemic traits, again suggesting that the current physiologic description of T2D remains incomplete.

Genetic mapping has also corroborated the epidemiologically identified intersection between T2D and obesity. A SNP in the second intron of the FTO gene was identified in parallel in GWAS for T2D as well as obesity. The association signal for T2D entirely disappeared with correction for body mass index (BMI), indicating that this SNP increased T2D risk by increasing BMI. Interestingly, this locus illustrates some of the difficulties in proceeding from GWAS signal to function. Although this SNP was initially thought to exert its effect on BMI by affecting FTO gene function, detailed mechanistic studies have revealed that it may function by altering expression levels of RX3, a gene present in million bases away. Although initial studies in mice showed that increasing Fto gene dosage increased food intake leading to increased fat mass, no connection has been found between the disease-associated SNPs and Fto expression level or function.

Whereas T2D is diagnosed on the basis of hyperglycemia, genetic mapping has revealed that the genes that determine fasting glucose are partly distinct from those that are associated with T2D. Comparison of GWAS performed for blood glucose levels in nondiabetics versus in T2D case-control studies reveals that glycaemia and T2D have distinct genetic associations. Some genes harbor variants that increase blood glucose levels and T2D risk, but others alter blood glucose levels but do not confer T2D risk. Thus, the two phenotypes have both common and distinct biology. Additionally, it is important to bear in mind that the genetic basis for surrogate measures of glycaemia do not always point to genes specifically altering glycemicy physiology. A particularly salient example is the association of hexokinase 1 (HK1) with hemoglobin A1c levels but not with fasting or dynamic glycaemia. It is that the genetic variant in HK1 alters hemoglobins A1c levels as a result of alteration of red blood cell life span and anaemia.

Clinical Translation

The principle of genetics pointing to important therapeutic targets in T2D is well validated. Both rare and common genetic variants link PPAR γ, the drug target of thiazolidinediones to syndromic and common T2D. Similarly, rare variants in the sulfonylurea receptor (encoded by ABCC8) cause neonatal diabetes. Although these oral hypoglycemics were identified in a pregenetic era, they point to an optimistic future of genetically guided drug discovery, one that will require detailed mechanistic understanding of the genes mapped by studies to date. A particularly attractive target nominated by genetics is SLC30A8, a gene that encodes a zinc transporter expressed almost exclusively in the endocrine pancreas. The common R325W missense variant in the protein encoded by the SLC30A8 gene (present in ~1.5% individuals in most continental populations) was found to associate with protection from T2D (1.18-fold decreased risk). Rare, protein-truncating variants in SLC30A8 (present in ~2:1000 individuals) have also been associated with protection from T2D with a larger effect size (2.6-fold decreased risk). The finding of human heterozygous knockouts for SLC30A8 who are protected from T2D and have no other deleterious phenotype offers a tantalizing therapeutic hypothesis that a chemical or antibody inhibitor of SLC30A8 could treat diabetes and minimize side effects.

In the arena of risk prediction, genetics has not yet made a major impact on T2D because existing clinical risk factors already predict disease well. Common genetic variants, by virtue of their relatively high frequency, can explain a large part of heritability for a trait/disease but have small effects on the individual. For example, the common P12A variant in PPAR γ is associated with 1.25-fold risk of T2D. Based on the high population frequency of the risk variant (85% P), if one were to theoretically substitute every P to A in the population, 20% of diabetes would be eliminated. Despite this incredible population-attributable risk, any given individual carrying the P variant only has a 25% increased risk of diabetes when compared to someone carrying the A variant. Given the high population frequency of common variants, many disease susceptibility variants are found in the same individual, each of which confers modest increases in risk. Thus, progressive accumulation of genome-wide association signals is feasible as it is possible to ascertain all known risk-conferring variants in an individual at once and combine these for potentially more clinically meaningful risk prediction. Investigators have attempted to combine common variants into a genetic risk score with modest success. For T2D, a risk score combining 18 common variants (including PPAR P12A) demonstrated 2.6-fold elevated risk in the high-risk versus low-risk group. By contrast, simply reporting a family history of diabetes increases risk by threefold.

Tailoring pharmacotherapy based on genotype has been successful in monogenic diabetes; the genome-sequencing era promises to bring this benefit to a wider group of individuals. The classic example of genotype guiding pharmacotherapy is that of individuals with MODY2 caused by autosomal dominant mutations in the GCK gene. Such individuals meet diagnostic criteria for diabetes but are able to regulate glycaemia at a higher set-point, thus avoiding all secondary complications (Chapter 31). Thus, a genetic diagnosis of GCK-related diabetes can allow such individuals to avoid pharmacotherapy. Individuals with permanent neonatal diabetes cause by ABCC8 or KCNJ11 mutations can be safely treated with high-dose sulfonylureas in place of insulin. It is likely that individuals with functional mutations in ABCC8 or KCNJ11 will be found who do not present with the classic neonatal syndrome and are classified as common T2D but who may still respond preferentially to sulfonylureas. Proving this will require identifying such individuals and performing prospective clinical trials. Prospective genotype-based intervention trials have been performed in individuals with MODY3 caused by HNF1A mutations and have shown the superiority of sulfonylureas over metformin. Interestingly, exome sequencing of Latin American T2D case-control groups revealed the HNF1A E508K variant, previously annotated as MODY3, as conferring a fivefold risk of T2D in approximately 1:1000 individuals. These data demonstrate that the E508K is not fully penetrant; nevertheless, the individuals who carry it may still benefit preferentially from sulfonylurea treatment, as do their counterparts with clinically defined MODY3.

Short Stature

Genetic Architecture

Adult height is a polygenic quantitative trait with a heritability of 80%. Many mendelian syndromes manifest...
(Chapter 24) large phenotypic variations in stature, and more than 150 genes are associated with monogenic short stature or overgrowth. The effect sizes for these rare, highly penetrant alleles are typically large, up to 300 mm (3 standard deviations [SD]). The genetic factors underlying most of human height variation are polygenic. GWAS aggregating more than 250,000 European samples have identified over 400 independent loci associated with height through common variant association. Common variants account for about 60% of heritability. The effect sizes for these common alleles are in the range of 1 to 15 mm (0.01 to 0.15 SD). Taken together, these genetic mapping studies show that the genetics of height consists of the additive effects of common alleles with modest individual effects except in very short individuals (>2 SD below the mean). 

In extremely short individuals, it is likely that rare alleles of large effect play a larger role. Many genes harboring rare alleles causing monogenic alteration in stature also harbor common alleles that contribute to polygenic stature variations. Up to 30% of the genes in height-associated loci identified by GWAS also contain rare, monogenic alleles. This overlap suggests that the genes in associated loci that have yet to demonstrate rare, large-effect alleles may be prime candidates for resequencing studies to discover new monogenic causes of alterations in stature. 

Short stature, but not tall stature, has been associated with deletions in the genome. By systematically comparing CNVs across the genomes of over 4000 individuals with developmental delay and congenital abnormalities with 7000 population-based control subjects, investigators observed that individuals with short stature harbored an excess number of low-frequency (found in <5% of the population) deletions. Individuals in the cohort clinically diagnosed with short stature demonstrated an average loss of 900,000 bp from their genomes.

In summary, the genetic architecture of height is consistent with classical theory and animal models for a polygenic, quantitative trait: thousands of genetic variants in hundreds of genes contribute to the genetic variability in height. Most of the genetic variation in height (97%) occurs through additive effects of common variants, each contributing a modest effect. Rare variants of large effect often cause mendelian syndromes cluster at the short extreme of stature have delineated multiple components of the GH/IGF-1 axis as a key endocrine pathway in human height regulation. A notable example is the microRNA clusters MIR17HG, which was also identified as a syndromic cause of short stature. By cross-referencing GWAS-associated genes with gene-expression microarrays from thousands of human tissue samples, investigators have found height-associated genes are most highly expressed in cartilage and the growth plate and to a lesser extent in bone and endocrine organs.

**Clinical Translation**

Gene mapping and functional characterization in monogenic stature disorders have motivated therapeutic advances for diseases of overgrowth and short stature. A classic example is that of Marfan syndrome, in which excess TGF-β signaling resulting from the effects of disruptive mutations in FBN1 has led to the development of TGF-β receptor blocker losartan, which exhibit TGF-β antagonist properties in vivo preclinical models, were begun in cohorts of Marfan syndrome patients. Results three years after treatment showed a decrease in aortic root diameter but not more so than conventional therapy with beta blockers. The identification of the activating G380N mutation in FGRF3 that causes 95% of achondroplasia has motivated the development of inhibitors of the tyrosine kinase activity of FGRF3 through small molecules as well as through analogues of C-type natriuretic peptide.

Diagnosis of short stature is an important application of genetic testing, particularly in pediatric populations. For example, the genotype-phenotype correlation of SHOX deficiency shows a clinical spectrum ranging from homozgyous loss-of-function mutations causing the severe syndrome Langer mesomelic dysplasia to heterozygous defects found in patients with milder syndromic disease (Léri-Weill dyschondrostosis, Ullrich-Turner syndrome) or idiopathic short stature. Genetic diagnosis qualifies these patients for GH therapy. Genetic diagnosis also enables directed screening for comorbid conditions. For example, males with the likely underdiagnosed 3-M syndrome (caused by mutations in CUL7, Obsl1, or CCDC8), clinically defined by severe postnatal growth retardation, characteristic facies, and radiographic findings of skeletal abnormalities, are at high risk of primary hypogonadism and need to...
be monitored.23 Similarly, individuals with short stature secondary to Noonan syndrome are also often underdiagnosed and are at higher risk for cardiac disease.26

Tailoring pharmacotherapy based on genotype is a useful adjunct to biochemical testing in stature disorders. Physiologic stimulation and serum biochemistry are the gold standards for assessing GH sensitivity and resistance guiding its pharmacologic use. But genetic information plays an important role in solidifying diagnoses suggested by biochemical testing and suggesting alternative pharmacologic therapy. For example, children with defects in the GH receptor or post-GH receptor signaling (STAT5B) will be candidates for treatment with biosynthetic IGF-1.27 Children with defects in IGFL1, a serum protein that stabilizes IGFs, respond poorly to both medications.28 GH therapy is used to treat short stature arising from certain genetic defects outside the GH/IGF-1 axis but is not indicated in many others. Given the variable expressivity in many syndromic stature disorders, clinically distinguishing among syndromes presenting with short stature can be challenging and imprecise. Genetic diagnosis can resolve ambiguity, particularly in children before all the features of a syndrome are present. For example, GH is contraindicated in chromosomal breakage disorders. Bloom syndrome (caused by loss-of-function mutations in BLM, which encodes a DNA helicase) is one such disorder presenting with short stature. In the absence of genetic diagnosis with only clinical and biochemical testing, there are case examples of children being treated for years with GH until their clinical presentation evolved and Bloom syndrome was diagnosed.29

Lipids and Coronary Artery Disease

Genetic Architecture

Serum lipid levels are a complex polygenic trait significantly influenced by environmental factors such as diet. Heritability estimates suggest a large role for genetic factors: approximately 40% to 60% for high-density lipoprotein cholesterol (HDL-C), about 40% to 50% for low-density lipoprotein cholesterol (LDL-C), and about 35% to 48% for triglycerides (TG). Monogenic disorders causing extremes of dyslipidemia have been associated with around 20 genes. Mendelian dyslipidemia syndromes can present with single or combined lipoprotein abnormalities (Chapter 37). On the hyperlipidemic side these abnormalities include syndromes of elevated LDL (familial hypercholesterolemia, sitosterolemia), elevated TG (LPL deficiency, APOCIII deficiency), elevated HDL (CETP deficiency, and combined LDL/TG elevation (familial combined hyperlipidemia, dysbeta lipoproteinemia). Monogenic disorders manifesting with extremely low lipid levels have also been identified and include low LDL (familial hypobetalipoproteinemia, PCSK9 mutations), low HDL (familial hypoalphalipoproteinemia, lecithin cholesterol acyltransferase [LCAT] deficiency, Tangier disease), and combined low cholesterol/TG (abetalipoproteinemia, chylomicron retention syndrome).

The genetic factors underlying most serum lipid varia-
tions are polygenic. GWAS aggregating approximately 200,000 multiethnic samples have identified more than 150 independent loci associated with serum lipids through common variant association accounting for about 15% of heritability.84 The effect sizes of common variants on lipid levels range from less than 1 mg/dL to about 15 mg/dL (SNP rs964184 at APOAI and TG).82 Among the loci identified, many alter one of the lipoproteins identified, a few (CETP, TRIB1, FADS1-2-3, APOAI) alter all lipoprotein levels, and a subset alters various combinations of lipoproteins.81 These overlaps are consistent with observations in mendelian disorders and corroborate the shared metabolism and lipoprotein constituents of LDL, HDL, and TG.

Many of the genes identified as monogenic causes of dyslipidemia also alter lipid levels in the general population through both common and rare variants. LDLR, encoding the LDL receptor, provides a case in point of how the allelic spectrum of genetic variation ranges from rare, mendelian alleles to common, small-effect alleles with effect sizes inversely proportional to variant frequency. Disruptive mutations (those that frameshift or prematurely terminate the protein) in LDLR like the ones found in familial hypercholesterolemia are found in 2:1000 to 7:1000 individuals and increase LDL levels by 150 to 200 mg/dL.83 Missense variants that decrease function of LDLR in cell models are found in about 1:100 individuals and increase LDL levels by about 100 mg/dL.84 A common, intronic SNP in LDLR found in 1:10 individuals decreases LDL levels by 7 mg/dL.82

In summary, the genetic architecture of serum lipids consists of a full spectrum of rare and common alleles, with hundreds of genes throughout the genome acting in a polygenic fashion. Different variants at the same locus, based on varying impact on gene function, can alter lipid levels in a broad range. Variant frequency in the population is inversely correlated with the magnitude of effect on serum lipid levels. Many loci have an impact on multiple lipoprotein levels simultaneously.

Disease Biology

Genetic mapping for lipid traits has a rich history of synergizing with biochemical and physiologic investigation to reveal the molecular mechanisms of lipoprotein metabolism and its relationship to cardiovascular disease in humans. Mendelian hyperlipidemia syndromes were the first to yield pathophysio logic insight, starting with Brown and Goldstein’s classic studies showing that LDL failed to suppress HMG-CoA (3-hydroxy-3-methyl-glutaryl coenzyme A) reductase activity in fibroblasts from subjects with familial hypercholesterolemia.85,86 Subsequent studies of families with severe hypercholesterolemia (LDLR, APOB, ABCG5, ABCG8, ARH, PCSK9) yielded insights into basic mechanisms of cholesterol absorption and biliary excretion as well as contributing to basic biologic understanding of receptor-mediated endocytosis, recycling, and feedback regulation.87–89 The association between high LDL cholesterol and increased rates of myocardial infarction (MI) was also noticed in these families and complemented by observations of low rates of coronary disease in families segregating unusually low LDL levels (familial hypobetalipoproteinemia: APOB, PCSK9, ANGPTL3).88 Epidemiologic association and the success of statin therapy in preventing coronary heart disease (CHD) extended the relationship of LDL and CHD to the general population, as have common LDL-associated SNPs identified by GWAS.82 The large number of associated loci for LDL, HDL, and TG combined with clinical outcome data on population-sized cohorts of genotyped individuals has enabled causality testing for epidemiologic associations with important public health consequences. Elevated HDL cholesterol levels have been associated with protection from MI, but the causality of this association is controversial. Should public health efforts be made to raise HDL levels in the population? Should drugs that increase HDL levels continue to be developed after early clinical failures? By leveraging common variants associated with HDL levels, the
causality of HDL in heart disease can be tested using an approach called mendelian randomization. As described earlier, genetic associations imply causality because genotype precedes phenotype, mitigating epidemiologic issues of confounding, bias, and reverse causation. Mendelian randomization can be conceived of as a clinical trial performed by nature in which the subjects are randomized at conception to genetic variants associated with a risk factor (e.g., SNPs increasing levels of HDL cholesterol). The subjects randomized to genotype are then assessed for the outcome (e.g., MI) and relative risk for the outcome compared in the treated and untreated arms. By utilizing a mendelian randomization approach with SNPs quantitatively associated with lipid levels, investigators found that genetic variants increasing HDL cholesterol were not protective for MI, whereas genetic variants decreasing LDL cholesterol were protective of MI. These findings are consistent with drug trials showing that LDL-lowering agents protect from MI, whereas multiple agents aimed at increasing HDL do not.

For TG, on the other hand, multiple lines of genetic evidence have supported a causal connection with CHD. First, rare, loss-of-function mutations in APOC3 are associated with low levels of TG and are also protective for ischemic cardiovascular disease in European and American cohorts. Second, a mendelian randomization study showed that SNPs that raised serum TG levels also increased rates of CHD. Finally, because of the interrelatedness of TG, LDL, and HDL, investigators have systematically examined all lipid-associated loci in cohorts phenotyped for CHD to dissect the contribution of TG to CHD risk apart from LDL and HDL. By constructing a statistical framework to account for the pleiotropic effects of SNPs on all three lipoprotein levels, they demonstrated that (1) SNPs that alter LDL and TG in the same direction of effect are associated with CHD risk, (2) SNPs that exclusively alter TG levels are also associated with CHD, and (3) the strength of a SNP’s effect on TG levels independently correlated with the magnitude of effect on CHD risk.

**Clinical Translation**

In the area of therapeutics, mapping of genes that affect lipid levels has exemplified a promising approach to drug target identification: genes that protect from disease when inactivated by nature might be useful pharmacologic targets. Statins, which inhibit HMG-CoA reductase (encoded by HMGCR), are among the most therapeutically successful drugs in lowering LDL levels and decreasing risk for CHD in both primary and secondary prevention settings. GWAS did identify HMGCR as a locus altering LDL levels (with an effect size of about 3 mg/dL), but how would this particular locus be prioritized as a therapeutic candidate among over 100 other associated loci? In some cases, experiments of nature (i.e., an allelic series) can be used to infer a dose-response curve of gene function that indicates how enhancement or suppression of the encoded protein’s activity raises or lowers disease risk. In the well-known case of PCSK9, for instance, loss-of-function mutations decrease LDL and cardiovascular disease risk, whereas gain-of-function mutations increase LDL and cardiovascular disease risk. Early clinical trials suggest that inhibiting the protein encoded by PCSK9 is a promising strategy for lowering LDL and preventing cardiovascular disease. Identifying genes that, when inactivated by nature, protect from disease offers several advantages for therapeutic targeting: (1) the target is already validated in humans, (2) designing inhibitors of gene/protein function is more tractable than increasing gene/protein function, and (3) nature has performed a lifelong clinical trial of inhibiting gene function, and the side effects of doing so are known. In the case of PCSK9, individuals with loss-of-function mutations demonstrated no phenotypic abnormalities other than low LDL levels and decreased risk of heart attack. In a similar example, naturally occurring mutations that disrupt NPC1L1 function, the inhibitory target of ezetimibe, were found to be associated with reduced plasma LDL cholesterol levels and a reduced risk of CHD.

Genetic risk predictors of CHD based on lipid-associated loci alone add little to the excellent risk prediction already provided by clinical risk factors, but when combined with CHD-associated loci they can provide clinically meaningful CHD risk prediction. An early study utilized nine SNPs at loci known to alter lipid metabolism (LDLR, APOB, HMGCR, PCSK9, CTP, LPL, ABCA1, APOE, and LIPC) to construct a genotype risk score and predict time to first cardiovascular event in a Swedish cohort. Even when the baseline lipid levels were corrected for the genotype, the risk score was statistically associated with cardiovascular disease, indicating that the score did capture the underlying genetic variants associated with CHD risk. This 13-SNP risk score also did not improve risk prediction over clinical risk factors alone. By contrast, a 27-SNP risk score consisting of both lipid and CHD-associated loci provided a threefold reduction in the number needed to treat with statins to prevent a CHD event as compared to clinical risk factors alone. This study was performed in high-risk primary prevention cohorts.

Given the broad indications for statins in primary and secondary CHD prevention, genetics can help to predict and elucidate side effects. GWAS has identified a few reproducible loci (APOE, LPA, SLC01B1, and SORT1/CERS2/PSRC1) for the trait of LDL response following statin therapy. SLC01B1 encodes the organic anion transporter, OATP1B1, which has been shown to regulate the hepatic uptake of statins. When exposed to a high dose of simvastatin, individuals carrying a missense variant in SLC01B1 (V174A), which causes loss-of-function, have up to a 2.5-fold increase in plasma levels of statin. A GWAS for statin-induced myopathy performed in a secondary prevention cohort receiving high-dose simvastatin identified the same genetic signal (via a noncoding SNP within the haptoglobin) at SLC01B1, which conferred a 4.5-fold risk of myopathy for one allele and a 16.9-fold risk of myopathy in homozygous individuals. The investigators estimated that 60% of the myopathy cases in their cohort were due to this variant at SLC01B1. They performed GWAS on only 200 case controls for myopathy, and it is likely that additional pharmacogenetically relevant loci like SLC01B1 will be identified in future studies with larger sample sizes. A number of observational and intervention studies have also associated statin therapy with increased T2D risk. An important question is whether this risk is mediated by an off-target effect of statins or an on-target effect via HMG-CoA reductase. An off-target effect suggests that new, more specific statins could be developed without this side effect, whereas an on-target effect implies that the development of more potent and specific statins would increase the risk of T2D. Genetics has begun to shed light on this question. Taking a mendelian randomization approach, investigators have found that SNPs at HMGCR that decrease LDL levels also increase BMI,
insulin resistance, and risk for T2D, suggesting an on-target mechanism for statin-mediated T2D risk.

**CONSIDERATIONS FOR CLINICAL USE OF GENETIC INFORMATION AND SEQUENCING IN ENDOCRINOLOGY**

As detailed previously, the clinical applications of genetic information include diagnosis, prognosis, risk prediction, and personalized therapy (e.g., pharmacogenetics). In the past, the cost of DNA sequencing placed limitations on the use of genetic testing. Revolutionary advances in sequencing technologies (collectively referred to as next-generation sequencing [NGS]) have made it feasible to sequence every gene in the genomes of all individuals at low cost. With the barriers of cost and feasibility diminishing every year, we believe that the use of genetic information will become commonplace in clinical practice. To maximize benefits to patients while minimizing the burden of false-positive and false-negative results, clinicians will have to select the right patients, deploy the appropriate genetic testing technology, and interpret the results. Specific clinical algorithms for patient/genetic test selection are being proposed for various endocrine diseases (e.g., short stature), but they will take time to validate. With this in mind, we present a series of broad patient scenarios ranging from those with no clinically apparent disease to affected individuals with clinically identifiable genetic syndromes, summarizing benefits and caveats of genetic testing in each scenario. We subsequently review issues related to targeted and genome-wide testing and provide some guidance for patient and test selection. Finally, we provide an overview of disease-relevant classification of genetic findings, examine the interpretation of genetic information as presented in a clinical laboratory report, and make suggestions for clinical decision making (Fig. 4-3).

**Genome Screening in the General Population**

Population-based screening for mendelian mutations known to cause disease would seem to be an advantageous application of genome sequencing. As mentioned earlier, certain mutations in the *RET* gene predispose to aggressive MTC with such high penetrance that the American Thyroid Association recommends prophylactic thyroidectomy in infants under 1 year of age. As genome sequencing becomes commonplace, it seems reasonable that genomes from the general population (e.g., as part of newborn screening) might be examined for *RET* mutations causing this rare but potentially fatal disease. Cases of cancer might be prevented and lives saved. However, one must consider that the clinical data on *RET* gene mutations have been obtained from kindreds affected by MEN2A/B and familial MTC (Chapter 39). Are genotype-phenotype correlations between *RET* mutations and MTC, or between other gene mutations and risk of other diseases, applicable to someone from the general population with no family history of disease?

An investigation of mendelian diabetes mutations ascertained in the general population suggests that genotype-phenotype correlations identified in mendelian disease families may not generally hold for the population at large. As a case in point, genome sequencing of a population-based U.S. cohort phenotyped longitudinally for diabetes identified 25 individuals with mutations previously known to cause autosomal dominant diabetes (MODY). Despite harboring mutations from a curated catalog of disease genes (Human Gene Mutation Database Professional) only one of these individuals met clinical criteria for MODY, and overall this group of MODY mutation carriers developed diabetes at a rate no different from that in the general population. As population-based sequencing becomes prevalent, it will become necessary to reexamine estimates of penetrance and genotype-phenotype correlations in the context of differing genetic backgrounds (e.g.,

---

**Figure 4-3** Suggested use of targeted and genome-wide genetic testing in individuals suspected of harboring a monogenic/mendelian disease. VUS, variant of uncertain significance (see text).
Males with this syndrome are at high risk for hypogonadism caused by mutations in the *stature, facial dysmorphism, and skeletal abnormalities* (example is the 3-M syndrome; clinically defined by short stature, genetic testing can have great prognostic value. Another mendelian disorders.

As with any medical test, the implication of genetic testing with regard to the likelihood of developing disease in the individual depends on a combination of inherent test characteristics (quantified by sensitivity/specificity) and the pretest likelihood of disease. For individuals with a family history of genetic disease the pretest probability of disease can be as high as 1 in 2 for a highly penetrant autosomal dominant disorder or 1 in 4 for a recessive disorder. Thus, for individuals with a family history of mendelian endocrine disease, genetic testing is often warranted (depending on the risks and benefits for the individual patient at hand, including psychosocial factors); this fits well with current clinical practice for asymptomatic individuals from families with mendelian disorders.

On the other hand, for individuals with no family history presenting with a mendelian-like genetic incident, the pretest likelihood of disease is that of the population (1:10,000 to 1:100,000). Even with a mutation conferring 50-fold increased risk, the individual is far more likely to remain free of disease. What reassurance can be offered to such individuals who carry mendelian mutations? Population-based sequencing surveys reveal that, on average, the genome of an apparently healthy individual contains approximately 100 mendelian-like disruptive mutations (i.e., frameshifting indels and SNPs resulting in premature stop codons) up to 20 of which are homozygously inactivated. Thus, even the protein-coding portion of the human genome (comprising only 1-2% of the total genome) contains unexpected redundancy that protects from disease. Depending on the severity of clinical consequences, and the estimate of penetrance in the general population, watchful waiting can be a prudent course of action.

### Asymptomatic Individuals

Endocrinologists may be referred individuals with no apparent symptoms in whom risk of disease may be increased: (1) those with a family history of known genetic disease who have not yet been tested and (2) those without family history who were tested and found with an apparently pathogenic mutation (i.e., a genetic incidentaloma). As with any medical test, the implication of genetic testing with regard to the likelihood of developing disease in the individual depends on a combination of inherent test characteristics (quantified by sensitivity/specificity) and the pretest likelihood of disease. For individuals with a family history of genetic disease the pretest probability of disease can be as high as 1 in 2 for a highly penetrant autosomal dominant disorder or 1 in 4 for a recessive disorder. Thus, for individuals with a family history of mendelian endocrine disease, genetic testing is often warranted (depending on the risks and benefits for the individual patient at hand, including psychosocial factors); this fits well with current clinical practice for asymptomatic individuals from families with mendelian disorders.

On the other hand, for individuals with no family history presenting with a mendelian-like genetic incident, the pretest likelihood of disease is that of the population (1:10,000 to 1:100,000). Even with a mutation conferring 50-fold increased risk, the individual is far more likely to remain free of disease. What reassurance can be offered to such individuals who carry mendelian mutations? Population-based sequencing surveys reveal that, on average, the genome of an apparently healthy individual contains approximately 100 mendelian-like disruptive mutations (i.e., frameshifting indels and SNPs resulting in premature stop codons) up to 20 of which are homozygously inactivated. Thus, even the protein-coding portion of the human genome (comprising only 1-2% of the total genome) contains unexpected redundancy that protects from disease. Depending on the severity of clinical consequences, and the estimate of penetrance in the general population, watchful waiting can be a prudent course of action.

### Symptomatic Individuals

Individuals with symptomatic disease can present with clinically defined or unknown syndromes. In both cases, a genetic diagnosis can be of psychological benefit, inform family planning, and sometimes can direct therapeutic screening/intervention. The genotype-phenotype correlation of endocrine tumor syndromes (MEN1, MEN2; see previous discussion) are a classic example of the benefits of genetic diagnosis for directed screening and prophylactic interventions in both affected individuals and close relatives. For example, the genotype-phenotype correlation of RET oncogene mutations dictates the urgency of prophylactic thyroidectomy (ranging from <1 year old to >5 years old) in children. Thus, even with a clinical diagnosis, genetic testing can have great prognostic value. Another example is the 3-M syndrome (clinically defined by short stature, facial dysmorphism, and skeletal abnormalities) caused by mutations in the *CUL7* gene (among others). Males with this syndrome are at high risk for hypogonad-
current versions of genome-wide sequencing, in which millions of genetic variants are identified simultaneously. There is often a tradeoff between the number of variants identified and the sensitivity/specificity for detecting and calling any individual variant. This analytic sensitivity/specificity relates to coverage, which is the depth of sequencing performed, or the number of independent times a particular nucleotide has been sequenced in a single test. Depth can vary greatly across the genome. For example, a clinical trial of whole-genome sequencing reported clinical grade genome sequencing at “30× coverage on average and at least 8× coverage for more than 95% of bases.” This means that of the 3 billion bases in the human genome, each base was sequenced independently 30 times on average, and more than 95% of bases were observed at least 8 times. However, 5% of the human genome was still poorly observed or missing. Thus, if clinical suspicion motivated examination of certain genes or genomic regions, a targeted approach might well have higher analytic sensitivity and specificity. The above-described clinical grade genome is sufficient for detecting variants as they are found with germline heterozygosity (on average 50% of the molecules sequenced would contain the variant base), but for disorders requiring detection below germline heterozygosity, such as somatic mutation testing in tumors, higher coverage would be required.

The molecular type of genetic variant (single-base changes or more complex variation such as insertions/delctions, duplications) strongly affects the sensitivity and specificity. The most commonly used gene-panel and genome-wide tests rely on NGS technology, which is currently well suited for detecting single or multiple base variations, but poorly optimized for the detection of structural variants (chromosomal rearrangements), triplet repeat expansions, and CNVs. For example, all false-positive genetic variants identified in a pilot clinical-grade genome sequencing study were indels or were near repetitive DNA stretches. As NGS technology and genome reconstruction algorithms improve, so will the ability to accurately detect these more complex genetic variants, reducing both false-negative and false-positive results. Currently, alternative modes of testing, such as chromosomal microarrays (array CGH), are used to identify large (>50 kb) structural and CNVs.

**Interpretation of Identified Genetic Variants**

Once genetic variants are identified (from sequencing or otherwise), they must be interpreted for their impact on health and disease. This interpretation requires the integration of population data (to know whether a variant is seen at higher frequency than expected for a pathogenic variant), computational predictions, experimental evidence, and familial comparisons. Curated databases are being established to begin to accurately catalogue this information and assist in interpretation. The ClinVar archive aggregates information about genomic variation and its relationship to human health.

Identified variants can be classified into three broad clinical categories: benign variants, pathogenic variants, and variants of uncertain significance (VUS), but intermediate categories have also been proposed. In single-gene or disease-targeted testing, the number of identified variants is small enough to allow for the individual assessment of all variants in each patient, once common benign variations are curated. However, exome sequencing identifies tens of thousands of variants, and genome sequencing identifies several million in each individual. Thus, automated filtering is necessary to point out a few pathogenic variants in a haystack of benign ones. Based on the assumption that the testing was done for a highly penetrant, mendelian variant, most genetic variants can be filtered as benign based on having been observed at a frequency greater than 1% (or even lower thresholds for rare disorders) in suitable reference populations. Computational analysis can support a benign classification by showing a variant is silent (e.g., encodes the same amino acid as the reference base). Experimental data can provide evidence that a missense variant does not alter protein function in disease-relevant bioassays. In addition, family-based data can be extremely powerful for interpretation; for example, the vast majority of variants for dominant diseases can be classified as benign if they are also observed in healthy relatives. Typically, the filtering of benign variants results in a 100-fold to 1000-fold reduction in the number of variants, requiring further analysis to 30 to 300 variants.

Among these, a similar hierarchy of evidence can support a pathogenic designation, or the evidence can remain inconclusive, resulting in a VUS designation. Computational analysis showing predicted gene disruption (premature stop or frameshift, etc.) can demonstrate deleteriousness of a variant, and family-based data can increase the conclusiveness of findings by showing that a proposed pathogenic variant segregates with disease or that it is absent from both parents in de novo cases of disease. Presence of the variant in databases such as the Human Gene Mutation Database (HGMD) can also provide supportive evidence, but the quality of evidence can vary widely among variants. Before final reporting, variants classified as pathogenic are typically validated by resequencing using traditional methods (i.e., Sanger sequencing), although this practice may shift depending on the state of sequencing technology. In a recent pilot case-control study of 20 cardiomyopathy patients and controls subjected to genome sequencing, 2 to 4 variants per individual (in both cases and controls) were classified as pathogenic, illustrating the challenges in interpreting genetic variation in isolation.

**Using a Genetics Laboratory Report to Make Clinical Decisions**

In the clinic, genetic testing is used to identify or confirm the cause of disease and to help the physicians make individualized treatment decisions (Table 4-5). Given the complexity of genetic testing, especially at the genomic (exome or genome) scale, physicians and clinical laboratories will have to work collaboratively to achieve useful results. For example, when a laboratory finds a rare or novel variant in the course of genomic sequencing, the director cannot assume it is relevant to a patient just because it is rare, or novel. The context of the patient’s history, physical examinations, and previous laboratory examinations are key to distinguishing among causal variants for the patient’s disorder, incidental findings, and benign variants.

How should identified variants be used in clinical decision making? Variant analysis contains uncertainties that are implicitly defined in the three-part classification described earlier: benign, pathogenic, and VUS. American College of Medical Genetics (ACMG) guidelines recommend that variants classified as pathogenic can be used in clinical decision making. However, genetic evidence should not be the only evidence of disease and should be used in conjunction with complementary clinical information when possible, especially as some pathogenic variants may be misclassified. Examples of corroborative clinical data
include enzyme assays, physical findings, and imaging studies. VUS should generally not be used in clinical decision making. Efforts to resolve the classification of the variant as pathogenic or benign should be undertaken, and interpretation in the context of the patient’s clinical scenario is critical. Variants classified as benign can usually be assumed not to cause the patient’s disorder.

Detection of pathogenic variants incidental to the diagnostic motivation for sequencing, but of potential clinical relevance, will be an inevitable consequence of genomic testing. This scenario is analogous to the inevitable detection of adrenal masses in computed tomography scans or thyroid nodules on physical examination. A clinician ordering genomic testing should be aware of laboratory policies and current ethical guidelines regarding such incidental secondary findings. Current recommendations are to offer the patient the option not to receive such incidental findings, and laboratories may vary in their reporting of such incidental findings. From both the clinician and patient perspective, incidental findings can also be specifically requested or declined. The laboratory should provide clear information about what constitutes a reportable incidental finding and how they may be requested or declined. Guidelines have been set forth in the ACMG recommendations for reporting of incidental findings in clinical exome and genome sequencing.115

Future Perspectives and Summary

In the future, we anticipate that a genome sequence will become a standard accompaniment to the medical chart; thus, the question Should we sequence? will transmute to What part of the sequence should we look at? A rational clinical approach will require the discipline to not look at all of it, or at least to rigorously interpret sequence data in the clinical context. As detailed previously, every human genome is littered with thousands of VUS and multiple variants classified as pathogenic; clinical suspicion is essential to help direct where to look and how to interpret genetic variation. This approach recapitulates current clinical algorithms for genetic testing. For example, familial paraganglioma shows locus heterogeneity as a mendelian disorder with cases attributable to mutations in multiple genes encoding suppressor proteins for the succinate dehydrogenase complex. The current genetic testing algorithm is hierarchical, starting with sequencing of the SDHD gene where the majority of causal mutations are found (Chapter 16). If no mutations in SDHD are found, other complex members are tested (e.g., SDHC). In a genome-sequencing era, a clinical algorithm might hierarchically look up mutations in the succinate dehydrogenase complex members from a sequenced individual with familial paraganglioma syndrome. No additional sequencing costs would be incurred as each gene is subsequently tested, but honing in on the appropriate and interpretable areas of the genome will reduce the clinical burden of false-positive results.

In summary, genetic information is most likely to be of clinical use in individuals with suspected mendelian syndromes (see 4S criteria enumerated earlier). For individuals with a clinically defined syndrome, for which targeted panels exist and are well validated, a targeted approach (single-gene or gene panel testing) is currently recommended as an initial approach. For example, genome-wide sequencing is likely not needed when MEN2B is suspected on clinical grounds; sequencing RET will usually make the diagnosis. If results from targeted genetic testing are uninformative, and the suspicion of a genetic disorder is high, exome or genome sequencing will make additional diagnoses in some patients (see Fig. 4-3). We recommend primary genome-wide approaches that assess both structural variation and sequence variation for individuals with clinically unclassifiable genetic syndromes or when targeted panels are not available or well validated. Depending on technological progress, this may simply be an unmasking of data that had not been reported back in a targeted test or may require new sequencing. The exome comprises 1% to 2% of the genome yet contains nearly 85% of known disease-causing mutations.116 Thus, given current technologies, exome sequencing supplemented with or following array CGH is a reasonable initial genome-wide approach. Many other best practices will improve the outcome of genetic diagnosis through sequencing. Ideally, DNA from both parents should be obtained, if possible, and DNA from additional relatives may also aid in interpretation. If unaffected and affected tissues are identified, paired affected-tissue-blood samples should be obtained when possible. Identified variants can be classified as benign, pathogenic, or VUS based on cross referencing with databases of diseases and nondiseased individuals as well as family members, computational analysis, and experimental evidence. Indeed, in order to maximize interpretability
of any genomic approach, it will be vital to interpret variation in the context of massive numbers of genomic sequences obtained in healthy individuals and in patients with disease. Finally, accurate classification will require physicians and clinical laboratories to work collaboratively, and the resulting genetic information should always be used in conjunction with complementary data (chemistry, imaging, etc.) for clinical decision making.

REFERENCES