3

Genetics of Endocrinology

ROBERT L. THOMAS, JOEL N. HIRSCHHORN, AND AMIT R. MAJITHIA

CHAPTER OUTLINE

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 variants found rarely in the population, usually from a relatively small number of genes, and each variant has a large individual effect on disease risk so that in any one individual, most of the disease risk is explained by variants in a single gene. Mendelian variants can be highly penetrant, but this is not always the case.
- 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, often found frequently in the general population, and with each variant contributing a small individual effect so that the phenotype in any one individual results from variants in many different genes.
- 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 [WES]) or in his or her entire genome (whole-genome sequencing) due to revolutionary advances in sequencing technologies (collectively referred to as next generation sequencing [NGS]). 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. Founded in 1982, the National Center for Biotechnology Information's GenBank now contains more than 200 million sequences and includes worldwide contributions from the DNA Data Bank of Japan and the European Nucleotide Archive.[1] Ongoing projects include the Genome Aggregation Database, Exome Aggregation Consortium (ExAC), United Kingdom 100,000 Genomes Project, and the National Center for Biotechnology Information's Single Nucleotide Polymorphism Database. Available sequencing data are growing exponentially, and the volume of information places increasing the demand for methods to distinguish pathologic variants from benign ones.

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 are 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 a DNA sequence (genotype) and an individual's clinical presentation (phenotype). 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 (see Chapter 42), Carney complex (see Chapter 15), and congenital adrenal hyperplasia (CAH) (see Chapter 24) to polygenic diseases, such as type 2 diabetes (see Chapter 34 and 35), Graves disease (see Chapter 12), and osteoporosis (see Chapter 29). The detailed discussions of the genetics of these and other disorders can be found throughout this textbook; this chapter provides illustrative examples that illuminate key concepts and refers the reader to the 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).[2] 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).[3]

However, most human diseases and phenotypes are not mendelian. Biometricians had appreciated in the early 1900s that most continuous and commonly varying traits (e.g., height and blood pressure) did not follow mendelian patterns of inheritance. In 1919, R.A. Fisher[4] 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/traits over the next 80 years. It would take a series of technological 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.[5]

Heritability: An Estimate of the Importance of Genetic Factors to Disease Causation

Relatives resemble each other in many ways. Resemblance with respect to traits such as height or to diseases such as multiple endocrine neoplasia type 1 (MEN1) could be explained by shared genotypes passed down through generations, shared environments, and nonlinear interactions between genes and environment. Heritability quantifies, as a proportion, how much of this familial resemblance is due to genetic factors. A trait that has no genetic influence would have a heritability of 0%; a trait that is completely determined by inherited factors would have a heritability of 100%. Most clinically important traits have heritabilities ranging from 20% to 80% (Table 3.1). Appreciating the heritability of a trait is important when interpreting the contribution of genetic risk factors in disease: genetic factors are less influential for traits with low heritability and are likely to have more predictive or explanatory power for traits with high heritability.

In the past, the gold standard for heritability estimation was the comparison of monozygotic and dizygotic twin concordance rates for diseases/traits. Such studies relied on the rationale that an excess of disease correlation between genetically identical individuals (monozygotic twin pairs) compared with those who share only 50% of their genes (dizygotic twin pairs) pointed to the role of genetic factors. However, the validity of comparing twin concordance rates across different families relied on the assumption that the effect of environment was the same for the twin pairs, regardless of whether they were monozygotic or dizygotic twins. More recent methods for heritability estimation can overcome some of these limitations by leveraging subtle fluctuations in genetic similarities between sibling pairs.[6]

Regardless of the methodology employed, it is critical to appreciate that heritability is not a fixed property of a disease/trait. The heritability estimate from any study must be interpreted in the context of the population in which it is being measured, including the historical period, and variability in environmental factors such as socioeconomic status and nutrition. These factors likely explain the wide range of heritability estimates for type 2 diabetes, ranging from 40% in Finland[7] to 80% in Japan.[8] An illustrative 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 with Russian Karelians.[9] 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.[10]

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 a human genome, either within the same person or between two different people, about 1/1000 of these bases vary (i.e., 99.9% of them are the same) (Table 3.2). There are many possible ways in which DNA sequences can vary; several specific types of DNA sequence variants are frequently observed (Fig. 3.1).

TABLE 3.1 Heritable Endocrine Traits and Diseases

Common Form	Heritability	Reference[a]	Selected Mendelian Forms
Type 1 diabetes	80%	132	*KCNJ11, ABCC8* (permanent neonatal diabetes)
Type 2 diabetes	40–80%	23, 28, 133	*AGPAT2* (congenital generalized lipodystrophy), *LMNA* (familial partial lipodystrophy 1) *HNF4A, GCK, HNF1A* (MODY1-3)
Obesity	40–70%	134, 135	*MC4R, POMC*
Hypertension	30–70%	136	*MEN1, RET* (MEN2A/B), *VHL, SCNN1A* (Liddle syndrome), *CYP17A1* (17OHD), *HSD11B2* (AME)
Height	80%	24, 72	*GH1, FGFR3* (achondroplasia), *SHOX1* (Ullrich-Turner syndrome), *FBN1* (Marfan syndrome)
Pubertal timing	50–80%	137	*KAL1, KISS1R, FGFR1* (hypogonadotropic hypogonadism)
Hyperthyroidism	80%	138	*TSHR* (familial nonautoimmune hyperthyroidism)
Hypothyroidism	67%	139	*TSHR, SLC5A5, TG, TPO,* and *TSHB* (congenital hypothyroidism)
Osteoporosis	50–85%	140, 141	*COL1A1, COL1A2, IFITM5* (osteogenesis imperfecta)
Serum calcium	40%	142, 143	*CASR* (familial hypocalciuric hypercalcemia), *HRPT2* (hyperparathyroid jaw-tumor syndrome)
Lipids	40–60%	85, 86	*Low-density lipoprotein:* LDLR (familial hypercholesterolemia)
			High-density lipoprotein: CETP
			Triglycerides: APOE (familial dysbetalipoproteinemia)
Kidney stones	56%	144, 145	*CLCN5* (X-linked recessive nephrolithiasis), *NKCC2* (Bartter syndrome)

[a]Numbers in this column indicate references listed at the end of the chapter.

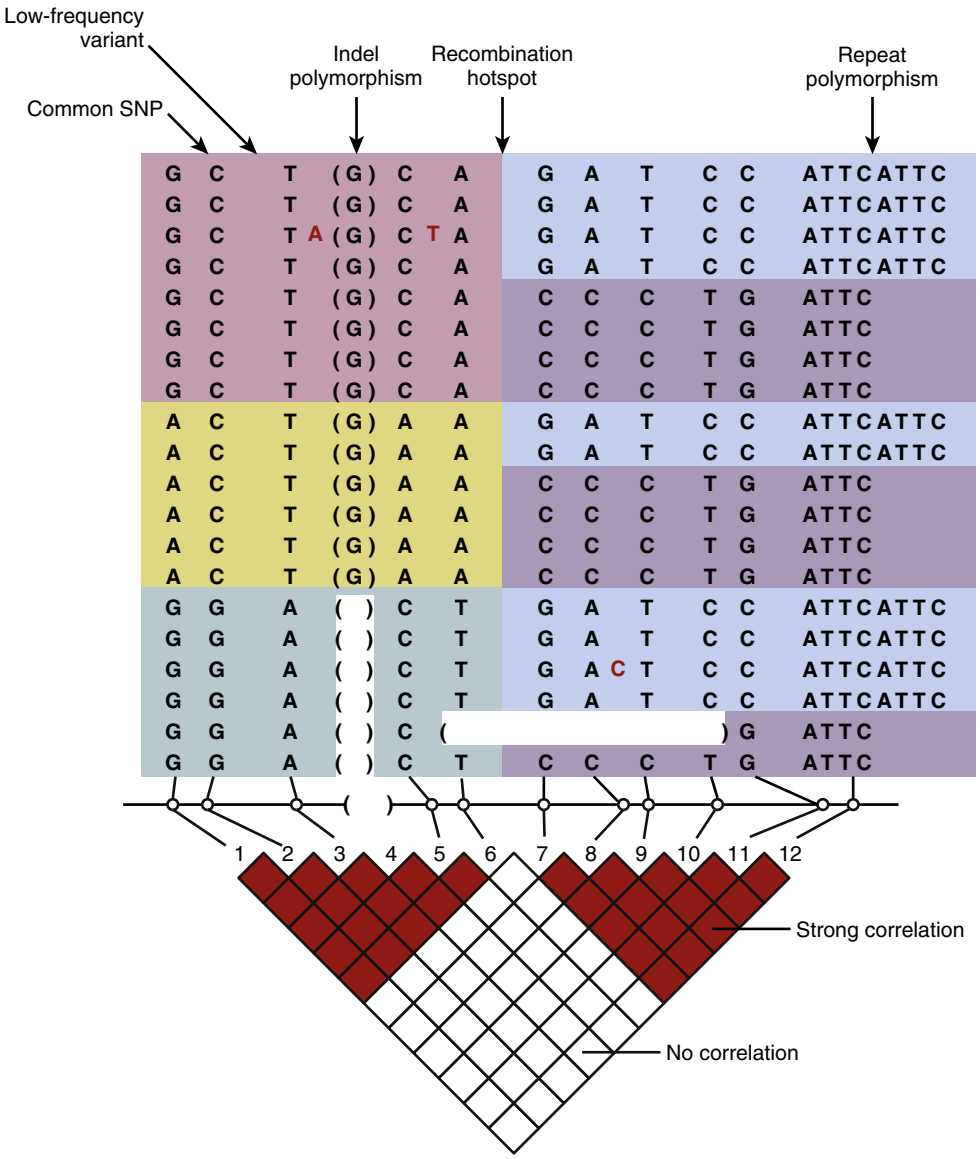
TABLE 3.2 Characteristics of Human Genome Sequence Variation

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

Adapted from Altshuler D. The inherited basis of common diseases. In Goldman L, Schafer AI, eds. *Goldman's Cecil Medicine*, 24th ed. Philadelphia, PA: WB Saunders; 2012.

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 (see 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* (see Chapter 25). 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 3.2), 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



• **Fig. 3.1** 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 (*SNP*s), an insertion-deletion polymorphism (indel), and a tetranucleotide repeat polymorphism. The six common polymorphisms on the left side are strongly correlated. Although these six polymorphisms could theoretically occur in 26 possible patterns, only 3 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 single-nucleotide polymorphisms 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) 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.)

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.[11]

Insertions and deletions (collectively called *indels*) respectively refer 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 ablate its function (see Chapter 15 and 24). Repeat polymorphisms (often referred to as copy number variants [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 [*AR*]) contains a repeat polymorphism in which a CAG codon, encoding glutamine, is repeated 11 to 31 times (see Chapter 24). 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 adrenocorticotropic 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 (see 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 eliminate 21-hydroxylase activity, whereas missense variants in *CYP21A2* often retain partial 21-hydroxylase activity (see Chapter 24). However, even a single, specific variant may have different effects in different individuals. The effect of any given genetic variant (genotype) on the 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 *AR* (encoding an S703G 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 (see Chapter 24).

TABLE 3.3 Origins of DNA Sequence Variation in Human Populations: Common Versus Rare Variants

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.[127] 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.

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 *germline* mutations—inherited 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 *GNAS1* 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 of café au lait spots, polyostotic fibrous dysplasia, and gonadotropin-releasing hormone (GnRH)-independent precocious puberty; the majority express two or fewer features of the classic triad (see Chapter 26). 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 (see Chapter 26). Conversely, blood cells can contain somatic variation that is absent in other tissues or the germline.[12]

It is important to remember that the base pair composition of a DNA sequence is not the only molecular determinant of phenotypic expression (Table 3.3). 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. 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 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 *SDHD* cause familial paraganglioma type 1. *SDHD* is maternally imprinted, so the mutation does not cause disease when inherited from the mother but is highly penetrant when inherited from the father. Imprinting can also be tissue specific. A paternally inherited inactivating mutation in *GNAS1* causes Albright hereditary osteodystrophy (pseudopseudohypoparathyroidism; see Chapter 29). The same mutation, when maternally inherited, manifests not only with Albright hereditary osteodystrophy but also with hypocalcemia secondary to parathyroid hormone resistance (pseudohypoparathyroidism type 1a [PHP1a]), because only the maternal copy of *GNAS1* is expressed in renal proximal tubules.

Evolution influences the frequency of variants that affect human phenotypes (e.g., 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 like malaria resistance in carriers of sickle cell disease). If a disease is at least mildly evolutionarily deleterious, then most common 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 for interpreting the impact of genetic variation on disease.

Summary

To summarize this introductory section, we have briefly described several 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 3.1). Genetic variants can take many forms ranging from single-base changes (SNPs) to translocations of entire chromosomes (see Fig. 3.1). 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 3.2). 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 3.4).

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 (Fig. 3.2) of genetic architectures. Although 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 3.1).

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 (see Chapter 40). Obesity as a common trait is highly heritable (heritability 40–80%), and genome-wide association studies (GWASs) analysis has begun to identify risk variants for the common form.[13] Although some of the risk variants overlap with those causing mendelian syndromes (as is also true for other diseases), GWASs 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 and 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/traits: (1) type 2 diabetes, (2) stature, and (3) serum lipids. In each section, we discuss what is known about the underlying genetic contributors, the impact of genetics on our understanding of disease biology, and the translation into clinical care in the short and long term.

TABLE 3.4 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, characterized by (1) few variants often in a single gene, (2) extremely rare frequency in the population (<1:1000), and (3) potentially 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) high frequency in the population (>1:20), and (3) often low penetrance (<1.5-fold risk) (see Fig. 3-2).

Because of their simple genetic architectures, mendelian endocrine disorders were ideally suited for genetic mapping using the techniques of familial linkage mapping developed in the 1980s.[3] 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 than was estimated from family-based studies.[146]

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 compare 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 at candidate genes selected by educated guessing. Such studies yielded several 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 more than 10 million common variants (the International HapMap project[147]) enabled genome-wide association studies (GWASs), a systematic approach to simultaneously test all genes for associations that could account for population-based confounding.[5] GWASs have yielded a large number of reproducible genetic associations for diverse common/polygenic diseases/traits,[148] 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 single-nucleotide polymorphism) 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)[149,150] and functional experimentation in model systems.[151]

Mendelian Endocrine Diseases

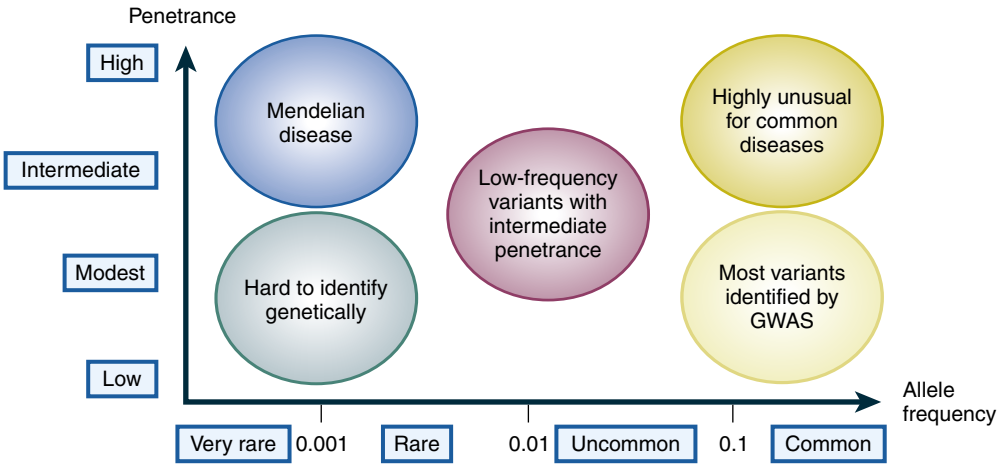
Genetic Architecture

Mendelian diseases represent one extreme of a spectrum of possible genetic architectures (see Fig. 3.2). The alleles causing mendelian diseases are found in a *small number of genes*, are typically *rare* (<1:1000), can be (although not always) *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 forms and locations, a phenomenon termed *allelic heterogeneity*. For 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*; see Chapter 15). However, if 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. By exome sequencing two individuals in a kindred with familial combined hypolipidemia (see Chapter 41), investigators identified two nonsense mutations in *ANGPTL3* that segregated with low serum lipoproteins when genotyped in other family members.[14] These mutations and the *ANGPTL3* gene were contained in the region identified by traditional linkage mapping[14] 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 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 several 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



• **Fig. 3.2** Genetic architecture of common and mendelian diseases. On one end of the spectrum are mendelian diseases caused by few variants in few genes, each with a large individual effect on disease risk. On the other end of the spectrum are common diseases and traits caused by the combined effects of many variants, observed frequently in the population, each with a modest individual effect. *GWAS*, genome-wide association. (Redrawn from McCarthy MI, Abecasis GR, Cardon LR, et al. Genome-wide association studies for complex traits: consensus, uncertainty and challenges. *Nat Rev Genet.* 2008;9[5]:356-369.)

setting of dysmorphic features and variable cardiac defects [see Chapter 24]) 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 (see Chapter 26), 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. Several genetic variants including frameshifting deletions, splice site alterations, and missense mutations have been identified (see Chapter 24) 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 (see Chapter 24). 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 *CYP2C19* and *CYP3A4*, a form of gene-gene 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 genetic variants in the *SRD5A2* gene (encoding 5α-reductase type 2) of differing molecular severity and location have been identified (see Chapter 24), but no correlation between the genotype and the clinical degree of virilization is apparent.

A second example of synergy between NGS and classical genetic study design revealed gain-of-function mutations in the *KCNJ5* and *CACNA1D* genes as causes of hyperaldosteronism.[15,16] By sequencing exomes in a series of aldosterone-producing adenomas from individuals with primary hyperaldosteronism (Conn syndrome), investigators identified missense mutations in the *KCNJ5* gene in about one-third of tumors from unrelated individuals. They also identified a separate *KCNJ5* missense mutation in a mendelian family with hypertension, primary hyperaldosteronism, and massive adrenal hyperplasia (familial hyperaldosteronism type 3; see Chapter 16). Follow-up biochemical and electrophysiologic studies showed that this series of somatic and inherited missense mutations eliminated ion selectivity in the *KCNJ5* gene product, a potassium channel. The increased sodium conductance through these mutant channels caused membrane depolarization of adrenal cortical cells in the zona glomerulosa, stimulating aldosterone release and cell proliferation.

As demonstrated previously, modern genome sequencing technologies facilitates the identification of pathogenic mutations. Translation of mutations into a disease mechanism, however, requires relevant experimental model systems and close correlation with human phenotype. PHP1a (see Chapter 29) with gonadotropin-independent sexual precocity provides a classic example of the iterative relationship between mutations, human phenotypes, and laboratory experiments. Some individuals with PHP1a harbor the Ala366Ser missense mutation in a stimulatory G protein (G_s). This mutation destabilizes the protein, causing loss of function in most body tissues and thus hormone resistance. But these mutation carriers also exhibited paradoxical testotoxicosis consistent with gain of function in G_s. The paradox of the same mutation causing both loss and gain of function was

resolved with a series of experiments showing that the Ala366Ser mutation caused a temperature-sensitive effect physiologically relevant to the normal temperature for Leydig cells. In most body tissues maintained at 37° C, the mutant G_s protein was destabilized, whereas in Leydig cells (within the testes, maintained at a 3–5° C lower temperature), the mutant G_s demonstrated increased activity. Elucidating this mechanism required an appreciation of the discordant phenotype (testotoxicosis) and biochemical characterization in the relevant physiologic system (Leydig cells at a lower temperature).

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.[17] 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.[18]

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) (see Chapter 42). 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.[19] 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 more than 200 individuals from more than 100 families (see Chapter 42).

A genetic diagnosis in several mendelian disorders can also directly inform pharmacotherapies. A classic example includes obesity caused by leptin deficiency (see Chapter 40), 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* 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 hypoglycemia.[20]

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 40%[7] to 80%.[8] Approximately 5% of diabetes cases that may be classified as 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, MODY, and congenital lipodystrophies.[21] To date, familial linkage studies have successfully implicated approximately 30 genes as monogenic causes of diabetes.[22]

The genetic factors 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/GWASs and the assembly of cohorts of thousands of cases and control subjects.[23] 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.[23] 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.[23]

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,[24] 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* locus.[25,26] 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), *HNF1A* (MODY2), and *PPARG* (familial partial lipodystrophy 3),[27] were found to harbor common variants that conferred risk for common T2D.[21,28] Conversely, genes first found associated with diabetes through GWASs 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).[29] 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.[30]

Mapping studies performed across various populations reveal both similarities and differences in the genetic risk factors contributing to diabetes among different ancestral groups. T2D GWASs performed in multiple populations/ancestries (European, South Asian, East Asian, Latino, African American)[31] 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 3.3), 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.[32] The same locus was identified by

another T2D GWAS performed in Japanese individuals.[33] Because the associated SNPs were rare in Europeans, the locus had not been detected in GWASs 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).[34] This variant, which causes a premature truncation and is associated with elevated muscle insulin resistance, is extremely rare in continental Europe and likely became common in Greenland because it was present in the founding ancestors of Greenland's current population.

In summary, genetic mapping studies over the past three 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 small to moderate effects (1.2- to 1.5-fold increased risk), and a larger 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 diseases[24] (see Fig. 3.2) and reflects both the underlying genetic architecture of the disease and the ability 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 in beta-cell genes (*SLC30A8*, *HNF1A*) and cell survival genes (*CDKAL1*) manifest with decreased insulin secretion (homeostatic model assessment B; see Chapter 34).[28] However, prediabetic individuals harboring T2D-associated variants in adipocyte genes (*PPARG*, *KLF14*) tend toward increases in homeostatic model assessment insulin resistance (see Chapter 34).[28] About 30% of T2D-associated SNPs point to insulin secretion/beta-cell function, and 15% point to insulin resistance.[22] Interestingly, the SNPs associated with insulin secretion in prediabetic individuals predict incident T2D, but the SNPs associated with insulin resistance do not.[35] 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, more than half of the 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.[36] For example, individuals harboring variants in *MNTR1B* 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. An SNP in the second intron of the *FTO* gene was identified in parallel in GWASs for T2D [37] and obesity.[38-40] 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 *IRX3*, a gene over a million bases away.[41] Although initial studies in mice showed that increasing *Fto* gene dosage increased food intake leading to increased fat mass,[42] no connection has been found between the disease-associated SNPs and *Fto* expression level or function.[43]

Even more recently, splice acceptor variants that generate premature stop codons and loss of function in *ADCY3* (a gene highly expressed in visceral adipose tissue) has been identified as causal for increased BMI and T2D risk. Concordant functional studies in mice with loss of function causing obesity, hyperphagia, and insulin resistance suggest that ADCY3 may be a new therapeutic target.[44]

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 GWASs performed for blood glucose levels in nondiabetics versus T2D case-control studies revealed that glycemia and T2D have distinct genetic associations.[45] Some genes harbor variants that increase blood glucose levels and T2D risk, whereas 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 glycemia do not always point to genes specifically altering glycemic physiology. A particularly salient example is the association of hexokinase 1 (*HK1*) with hemoglobin A_{1c} levels but not with fasting or dynamic glycemia.[46] It is thought that the genetic variant in *HK1* alters hemoglobin A_{1c} levels as a result of the alteration of the red blood cell life span and anemia.[46]

Clinical Translation

The principle of genetics pointing to important therapeutic targets in T2D is well validated. Both rare and common genetic variants link *PPARG*, the drug target of thiazolidines, to syndromic and common T2D.[27] Similarly, rare variants in the sulfonylurea receptor (encoded by *ABCC8*) cause neonatal diabetes.[47] 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 (ZnT8). The common R325W missense variant in the protein encoded by the *SLC30A8* gene (present in ~1:3 individuals in most continental populations) was found to associate with protection from T2D (1.18-fold decreased risk).[48] 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).[49] The finding of human