PART II INDIVIDUALIZING APPROACHES TO CARDIOVASCULAR DISEASE

7 Applications of Genetics to Cardiovascular Medicine

PRADEEP NATARAJAN AND KIRAN MUSUNURU

KEY PRINCIPLES OF HUMAN GENETICS, 71
Central Dogma, 71
Heritability, 72
Genetic Architecture, 72
Genetic Variation, 72
Characterizing Human Genetic Variation, 73
GENE DISCOVERY, 75
Family-Based Studies, 75
Case-Control and Population-Based Studies, 75
CAUSAL INFERENCE OF EPIDEMIOLOGIC ASSOCIATIONS, 77
Mendelian Randomization Principles and Applications, 79
DISEASE RISK PREDICTION, 80
Pathogenicity Assessments and Monogenic Risk, 80
Polygenic Risk Scoring, 81
THERAPEUTIC RESPONSE PREDICTION, 81
Target Discovery and Clinical Trial Prediction, 81
On-Target Therapeutic Side Effect Prediction, 82
Precision Medicine, 82
NEXT-GENERATION TECHNOLOGIES AND THERAPEUTICS, 83
Somatic Genomics, 83
Epigenetics, 83
Single-Cell Ribonucleic Acid Sequencing, 84
Therapeutically Targeting the Genome, 84
FUTURE PERSPECTIVES, 85
REFERENCES, 85

Naturally occurring human genetic variation has served for decades to elucidate the root causes of disease, including cardiovascular disease. Exponential technologic advances in computation, data science, and assay development have recently enabled population-based analyses, broad clinical profiling, and direct-to-consumer genetic testing in millions of people. Because germline genetic variation is established at conception and persists for the lifetime, genetics offers a robust tool for causal inference for broader preventive and therapeutic insights.

This chapter reviews key principles in genetics, gene discovery approaches, and diverse applications of genetic association study findings toward clinical translation (Table 7.1). The molecular structure of deoxyribonucleic acid (DNA) was described approximately 70 years ago, and the Human Genome Project completed the first draft of the human genome sequence approximately 20 years ago at an estimated cost of US$2.7 billion. Over a remarkably short period of time, human genetic data have become increasingly pervasive, and their connection to disease is increasingly understood, thereby rapidly expanding their relevance to the practice of cardiovascular medicine. To highlight the diverse and emerging applications of genetics to cardiovascular medicine, we primarily focus on coronary artery disease (CAD), the leading cause of death worldwide.[1]

KEY PRINCIPLES OF HUMAN GENETICS

Central Dogma

Genes are encoded in DNA, a polymeric molecule with two intertwining strands of a deoxyribose-phosphate backbone surrounding a ladder of paired purine and pyrimidine bases in a double helical configuration. The purine nucleotides are adenine (A) and guanine (G), and the pyrimidine nucleotides are thymine (T) and cytosine (C). Purines and pyrimidines link complementarily by hydrogen bonds across opposing strands: A-T, T-A, C-G, and G-C.

The linear DNA sequence represents its primary structure, and the base-paired double helix represents its secondary structure. Geometric and steric constraints leading to differences in orientation and shape lead to the tertiary structure. Lastly, denser packing of DNA molecules around protein anchors, known as histones, into chromatin provides the quaternary structure. Further chromatin condensation and packing yields the 22 pairs of autosomal chromosomes and one pair of sex chromosomes.

The "central dogma" of molecular biology refers to the flow of information from DNA to ribonucleic acid (RNA) to proteins. Traditionally, a gene is a DNA sequence that encodes a functional protein, and roughly 20,000 genes leading to distinct proteins have been described. *Transcription* copies the information in the DNA sequence into a single-stranded coding RNA, also known as a messenger RNA (mRNA). This polymer is structurally similar to DNA but uses uracil (U) in place of thymine (T). Of the 6.4 billion base pairs in the human genome, just over 1% represent *exons*, or DNA regions that directly encode mRNA. Subsequently, *translation* copies the information in an mRNA into a sequence of amino acids that make up a protein, which can service in a variety of roles (e.g., structural elements, enzymes, hormones, gene expression regulation). Variation in DNA sequence, or *genotype*, may influence protein function or abundance directly through alteration of the amino acid sequence when occurring within exons or indirectly when occurring in noncoding regions, including effects on splicing or mRNA transcript abundance. Such effects on a protein may lead to variation in an observable characteristic, or *phenotype*.

Epigenetics refers to phenotypic changes caused by factors beyond the DNA base pair sequence that influence the process of transcription. The most common such modification is methylation of cytosine bases, typically those in CpG dinucleotides, which generally results in reduced transcription or "silencing" of a gene. Post-translational modification of histone proteins, such as acetylation of lysine residues, can influence the accessibility of DNA sequence to the transcriptional machinery. Additionally, expressed RNA molecules that do not code for proteins, termed noncoding RNAs (ncRNAs), can yield phenotypic changes. For example, long ncRNAs can regulate transcription through several mechanisms, including interactions with the cell's transcriptional machinery and with histone-modifying enzymes; this is the mechanism for X chromosome inactivation in mammals. Additionally, microRNAs, another form of ncRNA, physically bind to complementary sequences in mRNA molecules and result in either suppression of mRNA translation or degradation of the mRNAs.

TABLE 7.1 Translating Genetics to Cardiovascular Medicine

BENCH	BEDSIDE
Identify causal factors that influence disease Test epidemiologic associations for causal inference	Biomarkers titratable to disease risk
Penetrance estimation	Disease risk prediction
Therapeutic target prioritization	Novel therapeutic targets
Therapeutic response prediction	Maximization of therapeutic benefit
Discover and characterize the range of phenotypic consequences of therapeutic traits	Minimization of therapeutic side effects
Diverse targeting strategies	Novel medicines

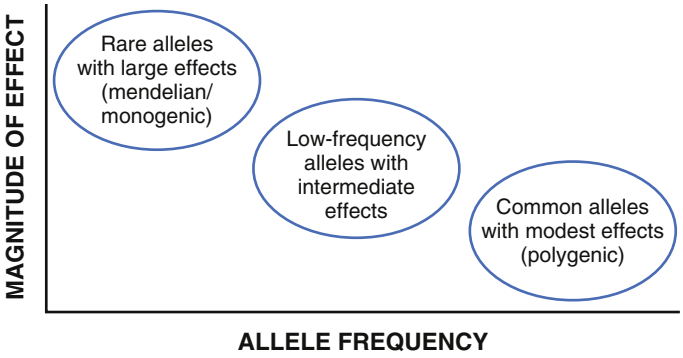

FIGURE 7.1 Relationship between allele frequency and effect magnitude of associated variants. Genome-wide assay studies, typically conducted with genome-wide genotyping arrays, typically identify common alleles with modest effects. Array coverage and imputation better enable the detection of lower frequency variants with intermediate effects. Rare alleles with larger effects are only detectable through genetic sequencing. Whole exome sequencing will detect the full allelic spectrum in coding regions, and whole genome sequencing will detect the full allelic spectrum across the genome.

Heritability

Many cardiovascular diseases, including CAD, aggregate within families. When disease occurs early, shared genetic factors may play a strong role. For example, a family history of premature CAD in a parent confers a nearly twofold risk for CAD.

Heritability refers to the fraction of interindividual variability in risk for disease attributable to additive genetic variation. Heritability is a population-based construct without clear meaning for individuals. Among individuals, 99.9% of the 6.4 billion base pairs are the same; genetic analyses leverage the 0.1% differences to understand trait or disease variation. It is estimated that CAD is 40% to 60% heritable, based on the aforementioned family-based methods or statistical genetics approaches. For common traits studied to date, heritability is typically in the 20% to 80% range. Traits with higher degrees of heritability are more suitable for gene discovery studies and genetic risk prediction. Remaining contributors to disease risk variability include environmental influences, nonadditive genetic influences (epistasis), nonadditive genotype/environment effects, errors in estimations of relatedness or disease, and random chance.

Genetic Architecture

The "genetic architecture" of a disease refers to the number and magnitude of genetic risk factors that exist in each patient and in the population, as well as their frequencies and interactions. For a given individual, diseases can result from genetic variation at a single gene (*monogenic*), few genes (*oligogenic*), or several genes (*polygenic*). In scenarios where a single gene defect is necessary to yield sufficiently large risk for disease, the condition is termed a *mendelian* disorder because it will obey classical modes of inheritance.

Typical mendelian modes of inheritance include autosomal dominant, autosomal recessive, or X-linked. In autosomal dominant disorders, a single defective copy of a gene (with most genes having two copies, one inherited from the mother and one from the father) suffices to cause the phenotype. Autosomal recessive disorders require both copies to be defective to lead to the phenotype. Familial hypercholesterolemia (FH), characterized by severely elevated blood cholesterol values and markedly increased risk for premature CAD, typically occurs due to single genetic variants in low-*LDLR*, *PCSK9*, or *APOB*. However, if both gene copies are disrupted, a more severe phenotype occurs, and thus the inheritance pattern is termed incomplete dominance. In X-linked disorders, the defective gene resides on the X chromosome. Given that men have only one X chromosome and women have two X chromosomes, men who carry the defective copy are affected with the disorder whereas women tend to be unaffected carriers, with some exceptions. Fabry disease, a lysosomal storage disease sometimes manifesting as cardiomyopathy due to disruptive mutations in *GLA* on the X chromosome, is typically more severe in hemizygous men (due to there being one X chromosome, and thus one *GLA* copy) than heterozygous women (due to there being two *GLA* copies). Thus, Fabry is not classically X-linked recessive and is generally simply termed X-linked.

Mendelian disorders imply that the presence of a pathogenic monogenic variant is deterministic for disease. However, genetic profiling in large datasets enables unbiased estimates of penetrance—the likelihood of a person with a pathogenic variant having disease—and expressivity—variation in severity of disease.[2,3]

Genetic Variation

Genetic architecture and phenotype largely dictate the diagnostic yield of genetic testing strategies (Fig. 7.1). Humans share the vast majority of DNA sequence, but variation in both coding and noncoding DNA sequences contributes to distinguishing characteristics between individuals. Due to natural selection over many generations, common genetic variation tends to link to modest phenotypic effects, whereas rarer genetic variation, arising relatively more recently in human history, can lead to larger phenotypic effects. Common genetic variation influencing phenotypes tends to occur within noncoding regulatory elements.[4] Coding sequence is less tolerant of genetic variation, and single base pair changes may lead to substantial phenotypic changes.

Current clinical cardiovascular genetics practice largely focuses on the detection of coding variants predisposing to large phenotypic changes (Fig. 7.2). DNA variation within coding sequence may not necessarily directly impact a protein's amino acid sequence. Degeneracy, or redundancy, in the genetic code refers to the observation that multiple codons (groups of three bases, the basis of the three-letter code) may yield the same amino acid. For example, variation at a G-C-A codon to G-C-G will lead to an alanine in both scenarios; such coding DNA sequence variants without impact on amino acid sequence are termed *synonymous* variants and tend to not have phenotypic consequences. Other coding variants can cause a variety of alterations in a protein—substitution of a single amino acid with another (*missense*), premature introduction of a stop codon (*nonsense*), scrambling of the amino acid sequence past the variant site (*frameshift*), or insertion or deletion of amino acids. These *nonsynonymous* variants may have a range of phenotypic effects from negligible to profound. Nonsense and frameshift variants tend to yield greater phenotypic effects than missense variants. Also, sequence variants at splice sites (the first and second bases after the end of each exon and before the beginning of each exon) can lead to a severely disrupted protein missing a domain encoded by an entire exon. Predicted loss-of-function, or protein-truncating, variants refer to nonsense, frameshift, or splice site variants; of note, such variants that occur near the downstream end of the DNA sequence may not have a significant phenotypic effect.[2,5] In silico prediction algorithms, largely weighted by assessments of evolutionary

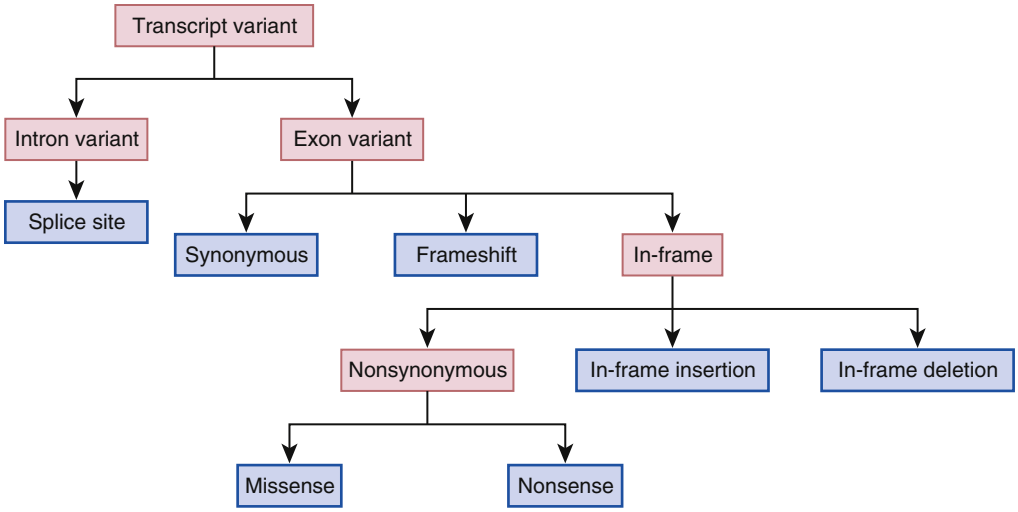

FIGURE 7.2 Protein-altering variant ontology. Key genetic variants expected to have direct impact on amino acid sequence, and therefore overall protein function, and their relationships are depicted.

conservation of DNA sequence across gene families and across species, may help to prioritize missense variants more likely to have larger phenotypic effects.[6]

Noncoding variants, although they do not directly affect the amino acid sequences of proteins, can cause phenotypic changes in other ways. A noncoding variant within regulatory elements, such as promoters or transcriptional enhancers, may result in a decreased amount of the protein product. Noncoding variants can affect the processing of RNA in other ways; for example, a noncoding variant that falls in the midst of a microRNA sequence might impair or enhance the microRNA's ability to interact with specific mRNAs. Large-scale research efforts are cross-referencing human genetic variation with diverse regulatory and intermediate effector molecule changes across tissues to help identify mechanistic links between noncoding DNA variation and phenotypes.[7]

Although most genetic variation is a single base pair change, larger DNA sequence changes may also yield phenotypic impacts. Viable aneuploidies (e.g., Down syndrome caused by trisomy 21) or chromosomal abnormalities can yield varied substantial effects. Copy number variants (CNVs) involve a variable number of repeats of a long DNA sequence (>1000 base pairs), whereas variable nucleotide tandem repeats refer to variation involving shorter nucleotide motifs. CNVs have been linked to congenital heart diseases as well as variation in atherosclerotic cardiovascular disease biomarkers, such as lipoprotein(a) [Lp(a)].

TABLE 7.2 Factors Influencing Linkage Disequilibrium

FACTORS	MECHANISMS
Variable recombination rates	LD extent is inversely proportional to the recombination rate, and certain regions of the genome have higher rates of recombination than others.
Variable mutation rates	Some regions, such as CpG dinucleotides, may have high mutation rates and show little LD.
Gene conversion	During meiosis, homologous recombination between heterozygous sites may result in correction of mismatched alleles effectively copying DNA sequence.
Natural selection	Haplotypes containing favorable alleles may be quickly swept to high frequency.
Population structure	Population subdivisions promote LD patterns in humans.
Admixture	Subsequent generations after gene flow can newly establish LD between nearby markers.
Genetic drift	Random sampling of gametes in each generation can lead to allele frequency changes, more pronounced in smaller populations

DNA, Deoxyribonucleic acid; *LD*, linkage disequilibrium.

Characterizing Human Genetic Variation

In most cases, a person has two copies of each DNA sequence because of the presence of paired chromosomes, and the two copies are known as *alleles*. Exceptions are for DNA sequences on the X or Y chromosomes in men, the two sex chromosomes being quite different, and for DNA sequences in the mitochondria which are exclusively maternally inherited. For a DNA variant, the *genotype* is the identity of the two alleles at the site of the variant. The two alleles may be identical (homozygous) or different (heterozygous).

A series of genetic variants that occur together is termed a *haplotype*. After the completion of the Human Genome Project, the International HapMap Consortium performed dense sequencing of large genomic segments in hundreds of individuals and identified regions of the genome (loci) where single base pair changes, or *single nucleotide polymorphisms* (SNPs), commonly occur across individuals. Nearby common variants are often found to be inherited together and exist in a state called *linkage disequilibrium* (LD) (Table 7.2). Because the haplotype is located on a single region of the chromosome, it tends to retain the linked genotypes as it passes from parents to offspring.

Genotyping technologies directly ascertain the genotype at prespecified variant sites. A common approach to interrogate the presence of a single variant is the polymerase chain reaction-based TaqMan assay; probes are designed to specific SNP alleles, each with a different 5′ fluorophore color that is detected during amplification. More commonly, prespecified variants are assayed in multiplex through array "chips" with the capacity to assess up to 2 million variants at once. Arrays are designed based on LD patterns detected in reference sequencing studies to ensure adequate coverage of haplotypes via "tagging" SNPs across the genome. This technology is used in conventional genome-wide assay studies and in most direct-to-consumer genetic testing services. Imputation, or statistical inference of nondirectly assayed genotypes using data from reference sequencing studies, can infer several million additional genotypes.[8] The imputed allele dosage (0 to 2 on a continuous scale) for each variant with frequency greater than 0.5% in the population is probabilistically assigned based on the combination of genotypes directly assayed on the array.

Sequencing technologies directly identify the order of base pairs in DNA (Fig. 7.3).[9] Sanger sequencing, first described in the 1970s and still in routine use, uses DNA polymerase to synthesize new DNA chains, using the DNA under study as a copy template, with trace amounts of fluorescently labeled chain-terminating nucleotides (four different colors for the four bases) to yield fragments of differing lengths that identify the base in each position by its color. Shotgun

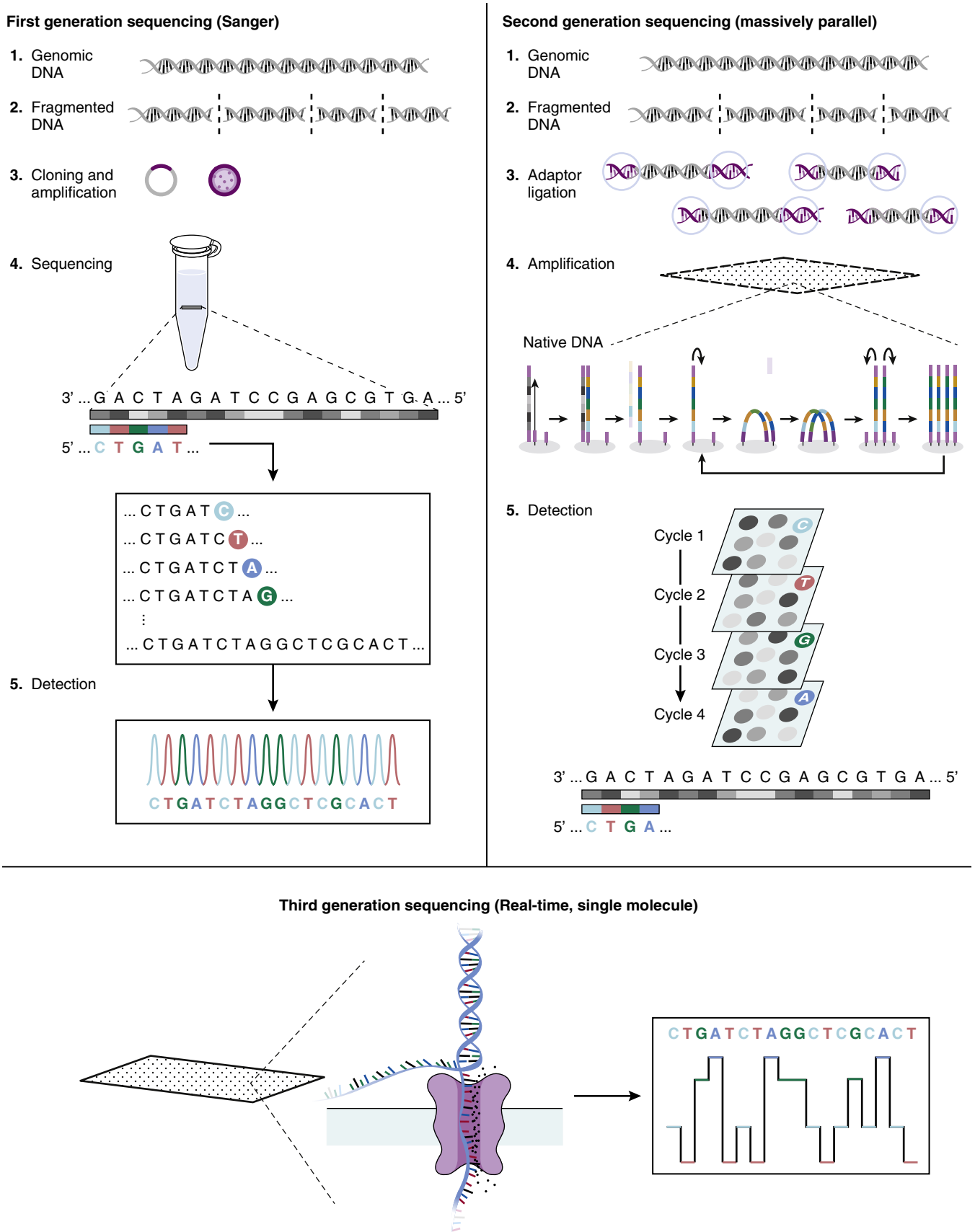

FIGURE 7.3 Schematic of DNA sequencing technologies. Second generation sequencing is also referred to as next generation sequencing. (Adapted from Shendure J, Balasubramanian S, Church GM, et al. DNA sequencing at 40: past, present and future. *Nature*. 2017;550:345–53.)

sequencing, involving the sequencing of random fragments of DNA with subsequent assembly of the sequences via overlaps between the fragments, was used for the Human Genome Project. Massively parallel "next-generation sequencing" (NGS) was developed in the late 1990s through early 2000s. In NGS, fixed DNA libraries provide templates for "sequencing-by-synthesis" in multiplex fashion. NGS can enumerate base pair changes across all 6.4 billion base pairs of the human genome (whole genome sequencing) or exclusively the protein-coding regions (whole exome sequencing). Both whole exome and whole genome sequencing are increasingly applied in population-based research analyses as well as clinical applications. To minimize biases introduced from templates (such as copying errors and sequence-dependent amplification biases) used for NGS, novel approaches such as real-time, single-molecule sequencing platforms for long-read de novo sequencing are being explored but have not yet been applied at similar scale as NGS.[10]

GENE DISCOVERY

Family-Based Studies

Conditions that occur prematurely and aggregate in families suggest important contribution from genetic variation. When classic mendelian inheritance patterns are observed for a suspected mendelian condition, genetic analyses to confirm the presence of a monogenic factor will have greater diagnostic yield than when such inheritance patterns are absent. For adult-onset conditions with strong genetic and nongenetic determinants, general familial enrichment may also result from polygenic or environmental factors. *Phenocopy* refers to a phenotype consistent with genetic predisposition but largely caused by environmental conditions for a given individual.

For novel syndromes or phenotypes without a known genetic basis or with nondiagnostic conventional genetic testing, family-based analyses may serve to discover novel implicated genes. Recruitment of multiple family members both with and without the phenotype allows for elimination of genotypes inconsistent with mendelian segregation. Both phenocopy and reduced penetrance may lead to deviation from expected inheritance patterns, and thus analyses of large extended pedigrees aid such analyses.

Previously, linkage studies were used to prioritize genomic regions that tended to cosegregate with the presence of a phenotype rather than the absence of the phenotype. Classic approaches, prior to widespread use of NGS, involved genotyping hundreds of genetic markers across the genome. Cosegregation of a marker with disease in pedigrees suggested that the causal disease mutation lay within several megabases of the marker, a region that often encompasses numerous candidate genes. Positional cloning would further narrow down the region by genotyping more markers, with subsequent sequencing used to identify the causal gene.

NGS is often now used upfront for broad gene sequencing, particularly whole exome sequencing, for family-based analyses. Variants annotated to disrupt protein function are prioritized if they are consistently observed among affected family members but not present among unaffected family members. The advent of large publicly available reference multi-ethnic whole exome and whole genome sequence databases of allele frequencies now allow for the verification of the absence of a suspected disease-causing variant among unrelated healthy individuals.[2] Once the rare genetic variant thought most likely to be the causal mutation is selected, it can be confirmed by sequencing the gene in unrelated individuals who have the same disorder. If some of these individuals have variants in the same gene (either the same or, more likely, different variants), it strongly argues that the gene is responsible for the disease.

Hypercholesterolemia and Coronary Artery Disease (see also Chapter 27)

FH afflicts approximately 1 in 300 individuals, manifesting as severely elevated blood cholesterol levels and increased risk for early-onset myocardial infarction (Fig. 7.4). Work in the 1970s and 1980s demonstrated that most cases of FH result from mutations in the *LDLR* gene, and subsequent studies implicated mutations in the gene for apolipoprotein B (*APOB*) at domains that interact with the LDL receptor.[11]

In the early 2000s, various studies identified families with apparent incompletely dominant FH but without *LDLR* or *APOB* variants. Linkage analyses and subsequent positional cloning identified *PCSK9* as the causal gene. Sequencing studies and subsequent functional work identified two different rare gain-of-function *PCSK9* variants in different families. PCSK9 increases blood cholesterol by binding to the LDL receptor and reducing the availability of the LDL receptor at the cell surface for cholesterol clearance from blood.

Also in the early 2000s, linkage and cloning analyses of families with autosomal recessive FH prioritized a large region on chromosome 1. Ultimately, homozygous mutations in *LDLRAP1* (previously known as ARH, autosomal recessive hypercholesterolemia) were implicated in several families of Sardinian origin. *LDLRAP1* encodes LDL receptor adaptor protein 1, which is required for endocytosis of the LDL receptor.

Metabolic Syndrome and Coronary Artery Disease

In 2007, linkage analysis of an extended family of Iranian ancestry with premature CAD and features of the metabolic syndrome resulted in the identification of a causal missense variant in *LRP6*. In vitro analyses indicated that the *LRP6* missense variant disrupts Wnt signaling. More recently, the same investigators used linkage analyses in three large families of Iranian ancestry with cosegregation of premature CAD and the metabolic syndrome to prioritize a region in chromosome 19.[12] Whole exome sequencing and focused analysis within the prioritized region identified a perfectly cosegregating missense variant in *DYRK1B* in all three families. Screening of morbidly obese individuals of European descent with CAD and multiple metabolic phenotypes identified a family with cosegregation of a different missense variant in *DYRK1B*. Functional analysis indicated that the variants were gain-of-function, promoting the expression of the gene encoding glucose-6-phosphatase.

Case-Control and Population-Based Studies

The technologic advances described earlier in this chapter allow unbiased assessments of the effects of genome-wide genetic variation on cardiovascular traits in large cohorts. Family-based analyses continue to be an efficient study design for families with apparently mendelian conditions with nondiagnostic genetic panel testing. However, heritability assessments for common conditions, such as CAD, indicate that naturally occurring common genetic variation may contribute to CAD risk broadly and not just in such exceptional families.

The design of studies using large cohorts focuses on maximizing power (likelihood of detecting true associations) to test hypotheses while minimizing the risk of detecting false associations. Power for genetic association analyses is determined by: (1) exposure (allele) frequency, (2) total sample size, particularly case count, (3) true effect of the exposure, and (4) threshold for statistical significance. Because there are approximately 1 million independent sites of common genetic variation in the human genome, a Bonferroni-corrected alpha threshold of 5×10^{-8} (0.05 divided by 1 million) for statistical significance is typically applied to genome-wide studies. Despite stringent thresholds for statistical significance used to mitigate false-positives in a single discovery cohort, putative novel associations should undergo independent replication in a validation cohort.[13] Both population stratification (systematic allele frequency differences between subpopulations) and cryptic relatedness (greater degree of relatedness among individuals in a cohort than is assumed) may lead to spurious associations. The use of genome-wide genotyping data to adjust for ancestry and genetic relatedness may mitigate such confounding.

Two broad analytic approaches are used—the common variant association study (CVAS) and the rare variant association study (RVAS).[14] CVAS is also termed genome-wide association study (GWAS). In a GWAS, genetic variants are sufficiently prevalent to estimate the relative difference between cases and controls or incremental change in a continuous outcome. In contrast, RVAS aims to test the collective contribution of individually rare variants to a phenotype, requiring the aggregation of rare variants into a statistical exposure unit for effect estimation.

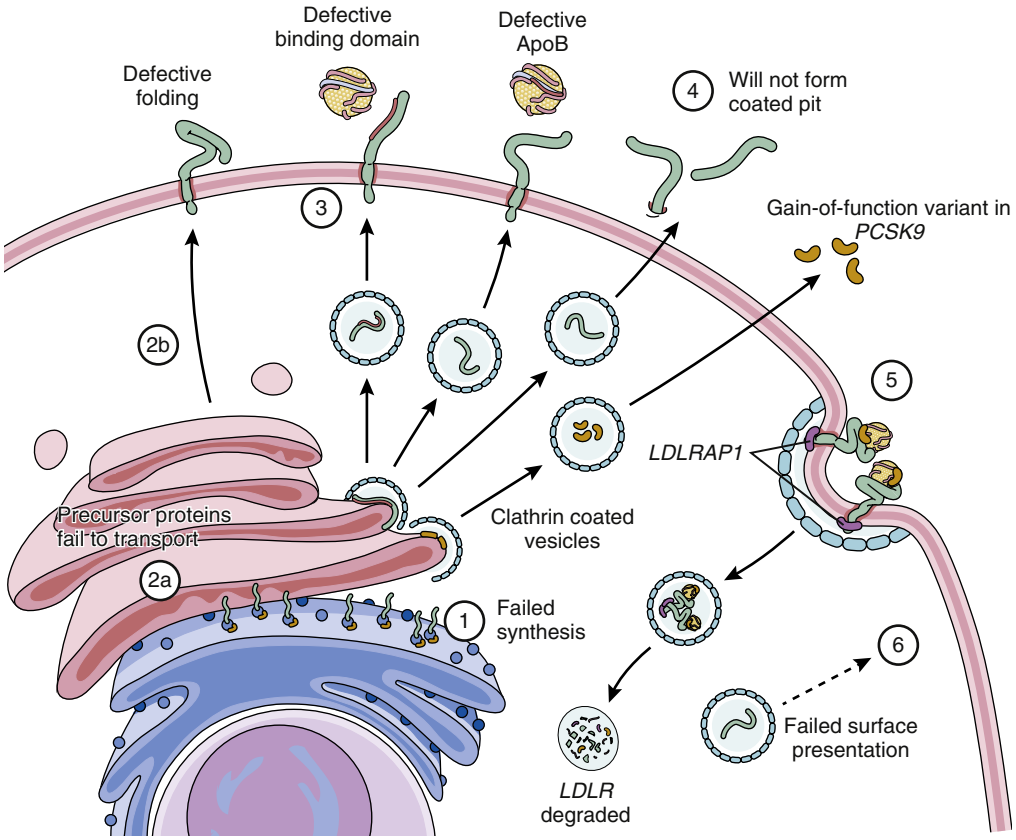

FIGURE 7.4 Mechanisms of *LDLR* dysfunction leading to familial hypercholesterolemia. Numbers refer to classes of *LDLR* variants: (1) synthesis of receptor or precursor protein is absent, (2) absent [2a] or impaired [2b] formation of receptor protein, (3) normal synthesis of receptor protein, abnormal low-density lipoprotein binding, (4) clustering in coated pits, internalization of the receptor complex does not take place, (5) receptors are not recycled and are rapidly degraded, and (6) receptors fail to be targeted in the basolateral membrane. *ApoB*, apolipoprotein B; *LDLR*, low-density lipoprotein receptor; *LDLRAP1*, low-density lipoprotein receptor associated protein 1; *PCSK9*, proprotein convertase subtilisin/kexin type 9. (Adapted from Gidding SS, Champagne MA, de Ferranti SD, et al. The Agenda for familial hypercholesterolemia: a scientific statement from the American Heart Association. *Circulation*. 2015;132:2167–92.)

GWASs use arrays comprising prespecified genetic variants, typically up to 2 million. Reference datasets may be used to impute 10 to 30 million additional variants depending on ethnicity and panel. Conventional statistical models use multivariable regression frameworks to compare each variant's allele frequency between cases or controls or with graded effect on a continuous outcome. Case-control cross-sectional study designs have a lower risk of confounding in GWAS versus in observational epidemiologic studies because putative confounders are unlikely to influence the random allocation of alleles at birth. Because case count strongly influences statistical power, case-control experimental designs are frequently used in GWASs. As broadly phenotyped mega-biobanks become increasingly available, new computationally efficient mixed model approaches to analyze unbalanced case-control phenotypes are often used.[15,16] Conventional methods ignore putative genetic interactions between loci, or epistasis; to address this omission, emerging methods aim to use multidimensional genetic architecture into genetic discovery. A novel discovery from a GWAS, typically a common noncoding SNP, represents just the first step in characterizing the biologic and clinical relevance of the genomic locus marked by the SNP, because the locus will often contain numerous candidate genes, any one of which could be causal. Follow-up efforts include comprehensive in silico and functional dissection to prioritize causal variants and genes toward understanding how the SNP genotype leads to the phenotype.

RVASs, which typically interrogate rare disruptive protein-coding variants, allow for more robust prioritization of causal genes, because any identified variants nominate the genes in which they reside. Given the infrequency of each individual variant, and the corresponding lack of statistical power, variants in the same gene are collapsed into a single statistical unit for association. Because approximately 20,000 protein-coding genes have been described in the human genome, the Bonferroni-corrected alpha threshold (i.e., corrected for multiple comparisons) for exome-wide significance is 2.5×10^{-6}. Because disruptive variants within the same gene may have bidirectional functional effects (loss-of-function variants versus gain-of-function variants), as is the case with *PCSK9* and *APOB*, specialized methods accounting for this phenomenon, such as the sequence kernel association test, are preferred.

Genome-Wide Association Studies for Lipids

Starting in 2007, GWASs have been performed on cohorts of individuals of European descent to identify SNPs associated with blood low-density lipoprotein cholesterol (LDL-C), high-density lipoprotein cholesterol (HDL-C), triglycerides, or total cholesterol. With each successive year, an increasing number of variants and loci are newly discovered due to: (1) increased sample sizes, (2) improved coverage from successive genotyping arrays, (3) incorporation of diverse ethnicities, and (4) improvements in genotype imputation. These advances also permit the characterization and association of uncommon alleles with larger effect sizes (Fig. 7.5). To date, over 350 distinct regions of the genome have been identified to be significantly associated with blood lipids.[17]

Imputation and the analysis of diverse ethnicities have enabled the detection of so-called Goldilocks alleles. Such variants represent large-effect disruptive mutations with sufficiently high allele frequencies to have statistical power in population-based studies. The analysis of founder or bottlenecked populations are well suited to identifying large-effect uncommon alleles. For example, a study of nearly 120,000 adults living in Iceland used array-derived genotypes imputed to 25.3 million variants from reference genomes from Iceland.[18] A novel rare (allele frequency 0.4%) Northern European–specific 12-base pair deletion in the fourth intron of *ASGR1* (a receptor on hepatocytes for a class of glycoproteins) was found to be associated with both reduced non–HDL-C and reduced risk for CAD. Polymerase chain reaction (PCR)-based and direct sequence analyses indicated that the intronic variant disrupted *ASGR1* mRNA splicing, leading to a truncated ASGR1 protein.

In addition to imputation, genotyping arrays enriched for exonic variant coverage ("exome chips") also identify large-effect uncommon disruptive variants. Such an approach was recently applied to lipids across diverse ethnicities, with several novel associations.[19,20] A new observation was the association of *A1CF* p.Gly398Ser with increased

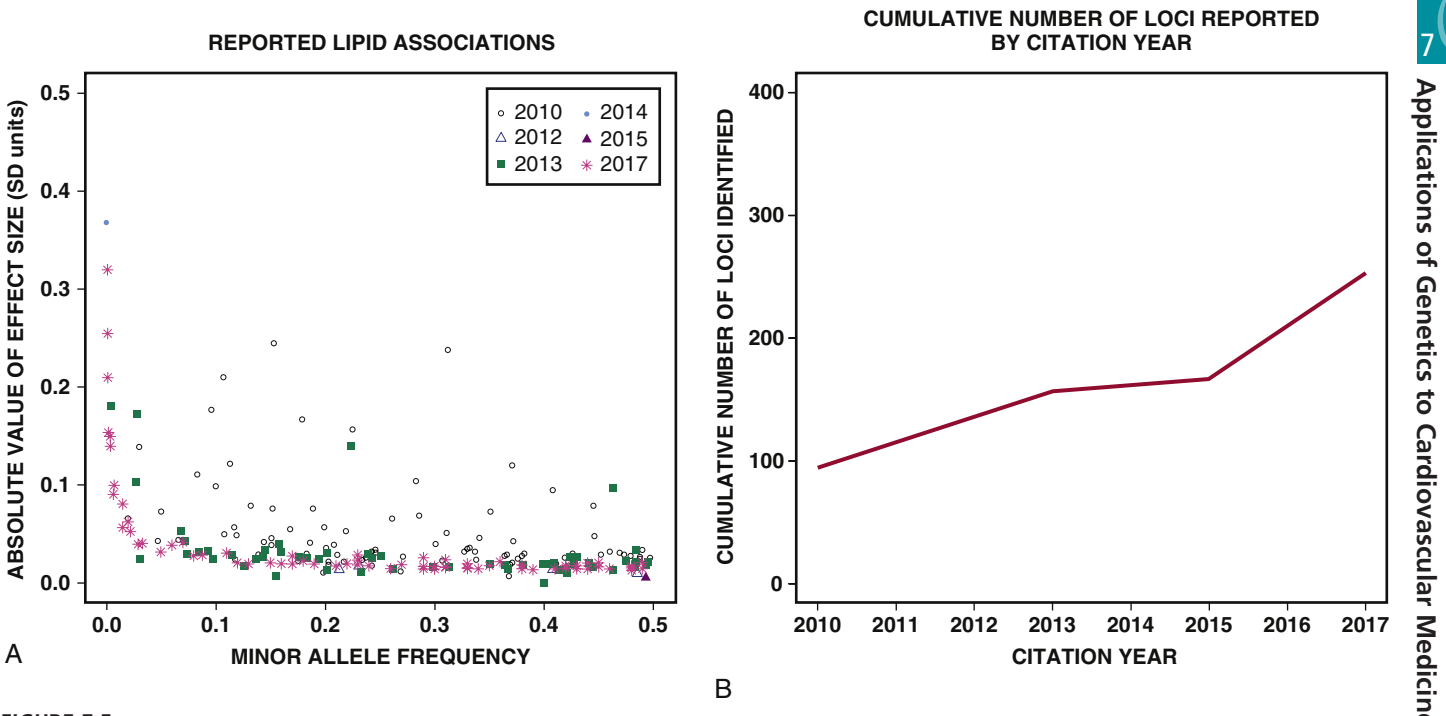

FIGURE 7.5 Identified lipid associations through genome-wide scans for plasma lipids. **A,** Compared with earlier studies, newer studies with denser arrays (including arrays enriched for coding variation), improved imputation, and larger sample sizes enable detection of variants across the allelic spectrum with more modest effects as well as lower frequency and rare variants with larger effects. **B,** Due to denser arrays, improved imputation, and larger sample sizes, genetic association studies for lipids continue to identify novel genomic loci associated with lipids. (Adapted from Peloso GM, Natarajan P. Insights from population-based analyses of plasma lipids across the allele frequency spectrum. *Curr Opin Genet Dev.* 2018;50:1–6.)

triglyceride and total cholesterol concentrations, as well as nominal association with increased risk for CAD.[19] Consistent with this observation, knock-in mice with the equivalent of the *A1CF* p.Gly398Ser mutation had increased triglycerides. A1CF is an RNA-binding protein that alters the splicing of messages that encode enzymes involved in carbohydrate metabolism.

Genome-Wide Association Studies for Coronary Artery Disease

The first GWASs for CAD were reported in 2007, all identifying a 58-kilobase interval in chromosome 9p21 not previously recognized to be relevant to CAD and not containing any protein-coding genes (a so-called gene desert). Despite intensive efforts since the discovery of this 9p21 locus, the mechanisms by which variants in the locus influence CAD risk remain unclear, highlighting how functional interrogation of disease-associated variants in genomic regions without robust pathophysiologic hypotheses remains a formidable challenge. The list of loci associated with CAD continues to expand, with 163 loci identified to date (Fig. 7.6).[21] Based on observed pleiotropy and prior biologic hypotheses, many loci may contribute to CAD risk through various established risk factors, and many other loci, including the 9p21 locus, may act through currently undiscovered pathways.

Analysis of low-frequency disruptive alleles for CAD using exome chips has also discovered newly implicated genes. *SVEP1* p.D2702G (allele frequency 3.6%) was recently found to be associated with increased risk for CAD.[22] SVEP1 encodes sushi, a cell-adhesion molecule. Interrogation of *SVEP1* p.D2702G with established CAD risk factors showed that it also led to increased blood pressure and increased risk for diabetes mellitus type 2. The CAD association appears outsized compared with the effects on blood pressure and diabetes mellitus, implicating potentially novel pathways that may contribute to CAD risk.

Evidence of association across an "allelic series"—multiple alleles with diverse frequencies (common and rare) and mechanisms (noncoding and coding) linked to the same gene—increases confidence in causal gene inference. Prior evidence strongly implicated the nitric oxide–cyclic GMP pathway in CAD risk, and CAD GWASs have detected several SNPs tagging key genes in the pathway, such as *NOS3*, *GUCY1A1* (formerly *GUCY1A3*, a guanylate cyclase subunit), *PDE5A*, *PDE3A*, and *MRVI1*. Luciferase assays for a CAD-associated noncoding variant near *GUCY1A1* show that it modulates *GUCY1A1* promoter activity.[23] Prior work linked loss-of-function mutations in *GUCY1A1* in an extended family with increased risk for premature CAD. Consistent with these findings, both common noncoding and rare coding disruptive alleles in *GUCY1A1* influence both blood pressure and CAD risk in population-based analyses.[24]

Population-Based Discovery of Rare Protein-Coding Variants Associated with Coronary Artery Disease

To date, few examples exist of aggregated rare variants significantly associated with CAD at exome-wide significance levels in RVASs. To date the only significantly associated gene to achieve exome-wide significance through this statistical procedure for CAD or myocardial infarction is *LDLR*.[25] Bolstered by strong evidence for association of *APOA5*, *APOC3*, *LPL*, *LPA*, *PCSK9*, *ANGPTL4*, *ANGPTL3*, and *NPC1L1* with atherogenic lipoproteins,[22,26–28] the observation of supportive albeit subsignificant associations of these genes with CAD at the population level supports likely causal involvement of these genes in CAD. As whole exome and whole genome datasets expand, power to detect rare variants through nonlipid pathways will improve. Focusing case-control ascertainment on extremes (early-onset cases and unaffected older control individuals) may prove to be a more efficient study design.[25]

CAUSAL INFERENCE OF EPIDEMIOLOGIC ASSOCIATIONS

Hypotheses concerning causal agents for complex diseases have often initially come from observational epidemiology. For example, seminal work in the 1960s in the Framingham Heart Study and other cohorts correlated blood cholesterol with future risk for CAD. Since then, studies have linked numerous soluble biomarkers with future risk for CAD (see also Chapter 10). How many of these biomarkers directly cause CAD, how many simply reflect other causal processes, and why is this

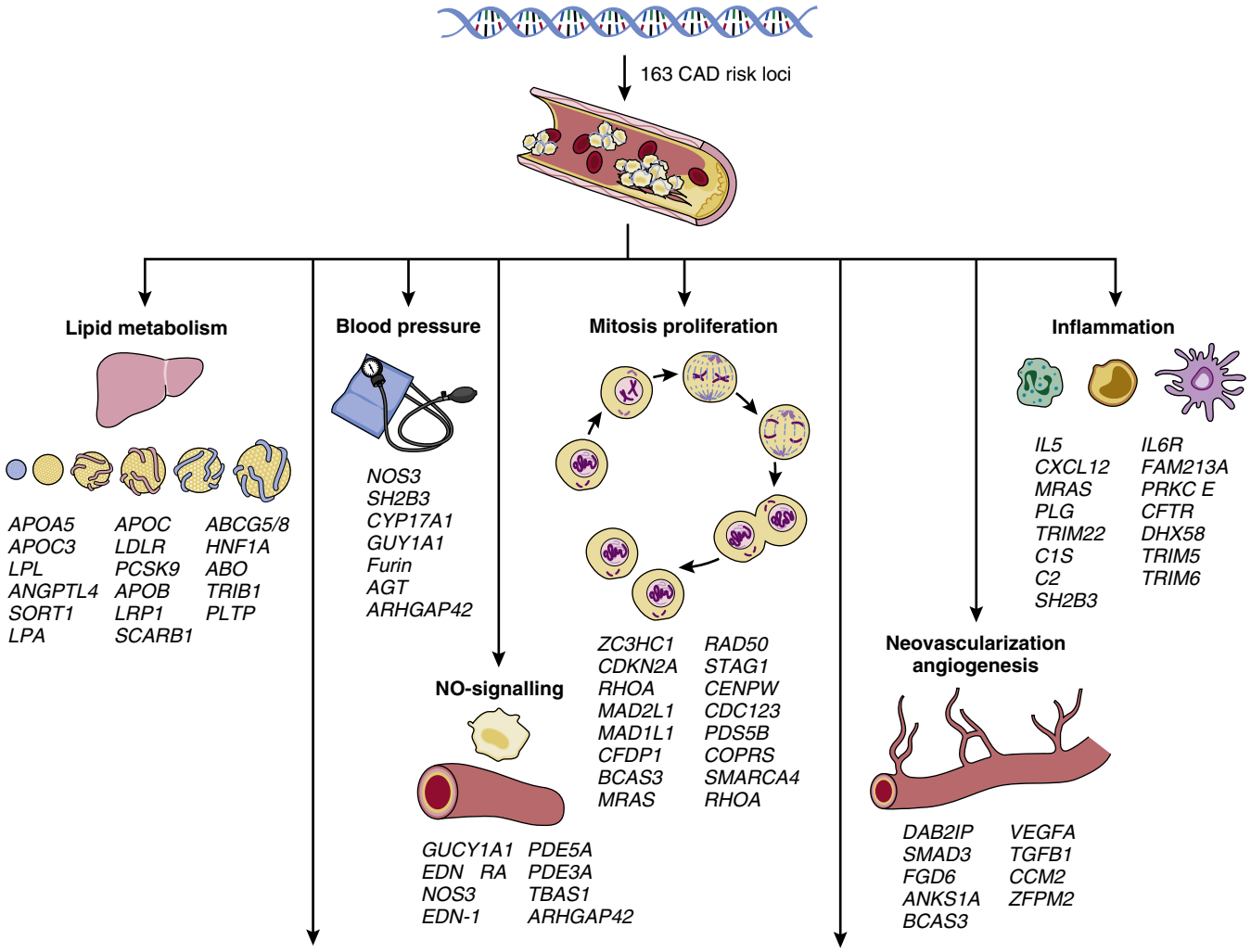

FIGURE 7.6 Genes mapped to known coronary artery disease loci from genome-wide association studies binned by atherosclerosis-related pathophysiologic pathways based on observed pleiotropy. (Adapted from Erdmann J, Kessler T, Munoz Venegas L, et al. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res.* 2018;114:1241–57.)

question important? Both causal and noncausal biomarkers may help predict risk for future disease, but only a causal biomarker may be appropriate as a target of therapy. A randomized controlled trial (RCT) testing whether a treatment that alters the biomarker will affect risk for disease is the ultimate test for causality in humans, but RCTs are expensive and time-consuming. However, supportive human genetic evidence increases the likelihood of RCT success.[29]

Mendelian Randomization Principles and Applications

A technique termed *mendelian randomization* (MR) uses DNA sequence variants to address the question of whether an epidemiologic association between a risk factor and disease actually reflects causality (Fig. 7.7). In principle, if a DNA sequence variant directly affects an intermediate phenotype (e.g., a variant in the promoter of a gene encoding a biomarker that alters its expression) and the intermediate phenotype truly contributes to the disease, the DNA variant should be associated with the disease to the extent predicted by (1) the size of the effect of the variant on the phenotype and (2) the size of the effect of the phenotype on the disease. If the predicted association between the variant and disease does not emerge from an adequately powered study, it would argue against a causal role for the intermediate phenotype in pathogenesis of the disease.

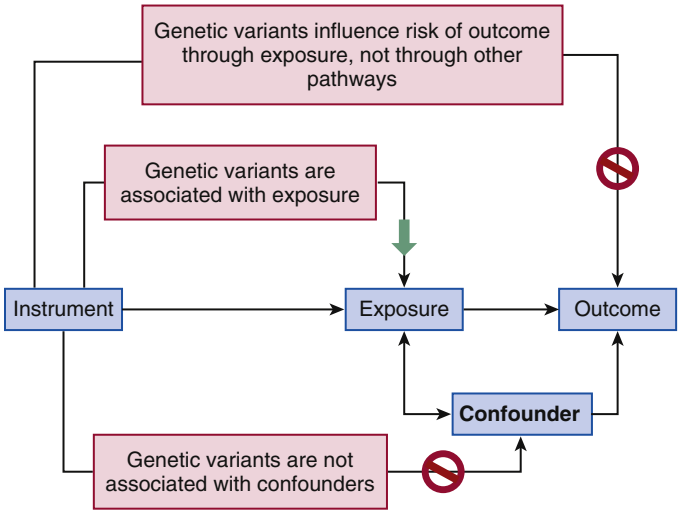

FIGURE 7.7 Mendelian randomization acyclic graph with assumptions.

The study design is akin to a prospective RCT in that randomization for each individual occurs at the moment of conception—genotypes of DNA variants are randomly "assigned" to gametes during meiosis, a process that avoids the typical confounders encountered in observational epidemiologic studies (Fig. 7.8). For example, a parent's disease status or socioeconomic status should not affect which of the parent's two alleles at a given SNP is passed to a child, with each allele having an equal (50%) chance of being transmitted by the gamete to the zygote. Thus, MR should mitigate confounding or reverse causation. MR has potential shortcomings, including that (1) the technique is only as reliable as the robustness of the estimates of the effect sizes of the variant on the intermediate and disease phenotypes, and (2) it assumes that the DNA variant does not influence the disease by other means (pleiotropy), which may not be true. In addition, a potential confounder of MR is that, in certain situations, a disease might cause the allele of a DNA variant passed from a parent to an offspring to be expressed in a different way (e.g., through epigenetic effects). Nevertheless, MR can prove informative for causal inference in observational human datasets.

Causal Inference for Lipoproteins (see also Chapter 27)

Numerous epidemiologic studies have positively correlated LDL-C and inversely correlated HDL-C with incident CAD risk. MR analyses support a causal relationship for LDL-C but not HDL-C. Consistently, multiple RCTs of different LDL-C–lowering medicines have demonstrated improved CAD outcomes, and multiple RCTs of different HDL-C–raising medicines have not noted any improvement in CAD outcomes. Consistent with meta-analyses implying that statins associate with an increased risk for diabetes mellitus, MR studies support a general causal inverse relationship between LDL-C and diabetes mellitus.[30] Although HDL-C is not a therapeutic target, it remains a robust biomarker for CAD risk prediction.[31]

Lp(a) is a circulating LDL-like particle covalently bound to apolipoprotein(a). Lp(a) is elevated in approximately one in five individuals and is independently associated with first and recurrent atherosclerotic cardiovascular disease events in multiple cohorts. Uniquely, Lp(a) is highly heritable across ethnicities, estimated at 85%, with associated genetic variation largely at the *LPA* locus.[32] MR studies indicate that Lp(a)-associated variants at *LPA* are also associated with CAD, supporting a causal relationship.[32,33] MR studies also extend this relationship to at least peripheral arterial disease and ischemic stroke.[34,35] Ongoing RCTs of medicines aimed at specifically lowering Lp(a) will test whether Lp(a) is causally associated with atherosclerotic cardiovascular disease. MR studies also indicate that Lp(a) is causal for aortic stenosis, for which a proven medical therapy has not yet been described.[34]

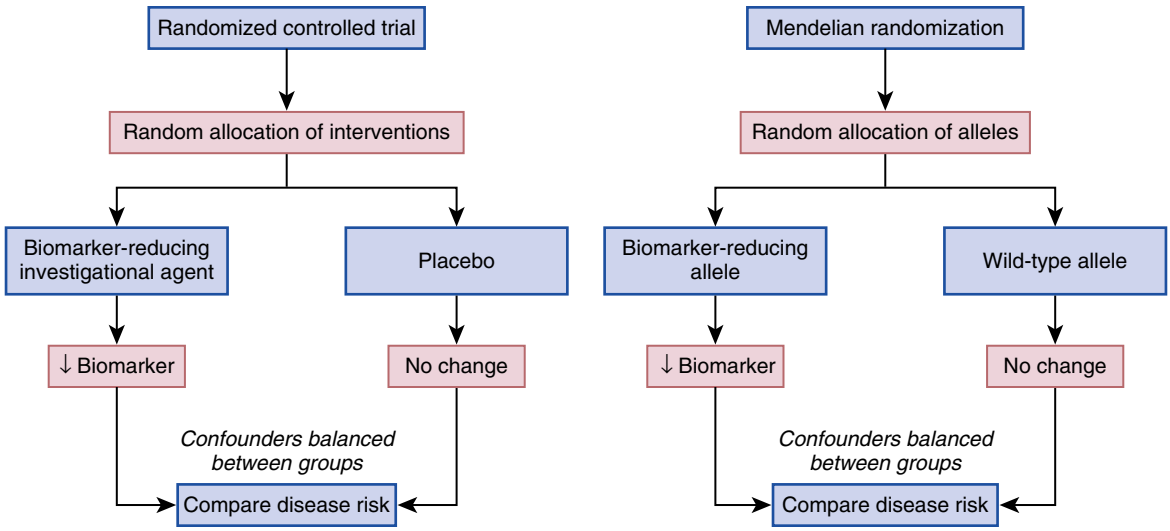

FIGURE 7.8 Parallel experimental designs between randomized controlled trials and mendelian randomization. Study randomization and the random allocation of alleles at birth facilitate the balance of putative confounders between exposure groups.

Causal Inference for Adiposity

Obesity has been correlated with diabetes mellitus type 2 and CAD risk but body fat distribution varies widely for a given body mass index (BMI). Waist-to-hip ratio (WHR) as a measure of abdominal adiposity associates independently with cardiometabolic risk in epidemiologic studies. However, reverse causation may lead to similar relationships; for example, individuals with CAD may be less prone to exercise, resulting in greater adiposity. A recent MR study, however, implied the relationship may be causal; genetic variants associated with WHR independent of BMI were strongly associated with both diabetes mellitus type 2 and CAD risks.[36] An increased WHR may occur either with increased abdominal adiposity or decreased gluteofemoral adiposity. Using dual-energy x-ray absorptiometry assessment and genotyping, MR studies indicate that abdominal adiposity and gluteofemoral adiposity may be separately deleterious and protective, respectively, for cardiometabolic disease.[37]

Mendelian Randomization Assumption Assessments

To buttress causal inferences from observational data, assumption assessments and sensitivity analyses are required. First, the hypothesis evaluated should have a strong scientific premise from observational epidemiology or experimental results from independent data sources. Second, a valid genetic instrument for the exposure of interest should be verified to mitigate weak instrument bias. After significant variants are selected and their corresponding exposure effects are tabulated from a discovery dataset, external verification of strong exposure association by examining effect estimate, model fit, and F statistic ensures validity.

After identification of a putative association, assessments of whether genetic variants influence the outcome via putative confounders correlated with the exposure of interest or independent pathways (horizontal pleiotropy) are pursued. Association of the genetic instrument with expected and measured confounders based on observational epidemiologic studies aids the assessment. The use of multiple significantly associated genetic variants in a composite score not only improves the instrument but also permits assessments of pleiotropy. For example, the more likely each variant's effect on the exposure is proportional to its effect on the outcome, the less likely pleiotropy is influencing the genetic association.[38,39] Some MR methods permit a non-zero intercept providing an estimate of unbalanced pleiotropy.[40] Novel methods now detect variant subsets that may exhibit horizontal pleiotropy to down-weight or remove outliers.[41,42]

DISEASE RISK PREDICTION

Current clinical practice focuses on the identification of monogenic variants among affected probands and asymptomatic family members. Genetic testing provides molecular confirmation for a clinical diagnosis and may inform treatments and surveillance. Because an increasing number of common genetic variants are found to be associated with cardiovascular diseases and risk factors, polygenic risk scores (PRSs) are being developed and evaluated for potential clinical application. As whole exome and genome sequencing becomes increasingly prevalent, both monogenic and polygenic factors may together improve disease risk prediction and preventive strategies.[3]

The liability threshold model of disease proposes a normal risk distribution for binary outcomes from numerous nongenetic and genetic factors, with a theoretical threshold above which a disease typically manifests. Knowledge of genotype-phenotype associations, even in the absence of identifying causal variants or genes, may inform phenotype prediction. Rare and common risk alleles as well as nongenetic risk factors, such as smoking, contribute to the overall liability of CAD risk.

Pathogenicity Assessments and Monogenic Risk

Clinical laboratories assess likelihood of disease risk on an ordinal scale using criteria put forward by the American College of Medical Genetics and Genomics (ACMG).[43] The five-tier terminology comprises: pathogenic, likely pathogenic, uncertain significance, likely benign, and benign. In addition to pathogenicity assertions, laboratories will assign inheritance patterns and associated conditions. Pathogenicity is interpreted according to: (1) scarcity in population-based datasets, (2) in silico assessments of deleteriousness, (3) functional assessments of deleteriousness, (4) cosegregation with disease in families, (5) de novo data in suitable pedigrees, (6) *trans* configuration with another pathogenic variant for autosomal recessive conditions, (7) curated reliable databases from external clinical laboratories, and (8) gene specificity for condition.[43]

The current classification system may inadvertently connote full penetrance for pathogenic variants and null risk for the remaining variants. As sequencing data are increasingly available in unselected populations, it is increasingly clear that pathogenic variants predisposing to adult-onset disease carry high disease risk but are not deterministic.[3] Approximately 1% of adults harbor a pathogenic variant for an "actionable" adult-onset condition, largely cardiovascular or oncologic. Currently, when secondarily detected in clinical testing, pathogenic variants for 59 such genes are recommended for return of results to patients according to the ACMG (Table 7.3).[44]

TABLE 7.3 American College of Medical Genetics and Genomics Cardiovascular Genes and Associated Phenotypes Recommended for Return of Secondary Findings in Clinical Sequencing

PHENOTYPE	GENE	INHERITANCE
Ehlers-Danlos syndrome, vascular type	*COL3A1*	AD
Marfan syndrome, Loeys-Dietz syndromes, and familial thoracic aortic aneurysms and dissections	*FBN1*	AD
	TGFBR1	AD
	TGFBR2	AD
	SMAD3	AD
	ACTA2	AD
	MYH11	AD
Hypertrophic cardiomyopathy, dilated cardiomyopathy	*MYBPC3*	AD
	MYH7	AD
	TNNT2	AD
	TNNI3	AD
	TPM1	AD
	MYL3	AD
	ACTC1	AD
	PRKAG2	AD
	GLA	XL
	MYL2	AD
	LMNA	AD
Catecholaminergic polymorphic ventricular tachycardia	*RYR2*	AD
Arrhythmogenic right ventricular cardiomyopathy	*PKP2*	AD
	DSP	AD
	DSC2	AD
	TMEM43	AD
	DSG2	AD
Romano-Ward long QT syndrome types 1, 2, and 3, Brugada syndrome	*KCNQ1*	AD
	KCNH2	AD
	SCN5A	AD
Familial hypercholesterolemia	*LDLR*	AD
	APOB	AD
	PCSK9	AD

Monogenic Coronary Artery Disease

Although FH is a well-established monogenic risk factor for CAD, the necessity of molecular confirmation after clinical diagnosis from routine lipid screening and history has been controversial.[11] Cascade testing, or screening family members of probands, has been touted as a key reason for genetic testing, but its value beyond lipid screening is less clear. Indeed, many relatives of those with FH have not had lipids assessed regardless of variant presence.[45] Two key factors have facilitated expansion of FH genetic testing in clinical practice: (1) large-scale NGS in population-based cohorts demonstrating incremental prognostic assessments[46,47] and (2) the availability of novel, expensive cholesterol-lowering medicines.

Beyond a single LDL-C value, the presence of an FH variant may yield added risk for CAD. Among individuals with severe hypercholesterolemia (LDL-C > 190 mg/dL), only 1 in 50 has an FH variant.[46,47] Among those with a clinical phenotype classified as "probable FH," 1 in 16 has an FH variant, whereas 1 in 4 with "definite" FH has an FH variant.[46] Compared with those without elevated LDL-C levels and without FH variants, severe hypercholesterolemia without an FH variant carried a 6-fold greater risk for CAD, but those with severe hypercholesterolemia and an FH variant had a 22-fold greater risk for CAD.[47] Per guidelines and current U.S. Food and Drug Administration labeling, in the primary prevention setting, more stringent LDL-C targets should be pursued in FH patients, with PCSK9 inhibitors as needed to attain those targets.[31]

Polygenic Risk Scoring

Individual common variants associated with a condition in GWASs may be leveraged for the construction of PRSs (Fig. 7.9). Early PRSs were simply a summation of uncorrelated, significantly associated risk alleles. As most genetic variants have unequal disease effects, weighting by the disease risk effect estimate improves model fit, and unweighted PRSs have largely now been abandoned for risk prediction. Because GWASs of increasing sizes detect increasing numbers of significantly associated variants, many variants not yet significantly associated with an outcome may still inform risk prediction. Novel methods focus on expanding the number of variants included in the model and reweighting to improve risk prediction, including for the creation of so-called genome-wide PRSs.

Due to their largely being informed by GWASs with individuals of European descent, the performance of contemporary PRSs continues to lag for non-European ancestries.[48] Ongoing efforts to close this gap to reduce the risk of exacerbating existing health disparities include increasingly large genetic studies in diverse ethnicities and the development of novel PRS methodologies.

Polygenic Coronary Artery Disease

Significant recent efforts have focused on polygenic risk scoring for CAD, given the prospect of facilitating earlier preventive strategies for the leading cause of death. Polygenic risk for CAD predisposes to the development of premature coronary atherosclerosis.[49,50] A CAD PRS was shown to predict future risk for CAD among individuals with or without a family history of premature CAD.[51] Recent implementation of genome-wide PRSs for CAD in the U.K. Biobank has shown improvement in the prediction of CAD beyond conventional risk factors.[52,53] Genome-wide PRSs may be particularly well suited to better identify individuals at markedly elevated risk for CAD at the distribution tails. For example, the top 5th percentile (1 in 20) carries similar odds for CAD as individuals with FH variants (1 in 300).[52] Although FH is typically readily detected by significantly elevated LDL-C concentrations, elevated CAD PRS is not readily detected by conventional risk factors. Among middle-aged adults, risk discrimination is similar to other cardiovascular disease risk factors.[54,55] A CAD PRS may be particularly helpful for middle-aged adults at intermediate cardiovascular disease risk.[56] However, a CAD PRS may prove more useful for guiding therapeutic intervention earlier in life even before the onset of conventional cardiovascular risk factors.[53,56]

THERAPEUTIC RESPONSE PREDICTION

In his 2015 State of the Union Address, President Obama launched the U.S. Precision Medicine Initiative. Although the notion of guiding preventive and treatment strategies by accounting for individual variability is not new, broad research and application are now feasible with large-scale human genetic and biologic databases, high-throughput molecular profiling, and computational advances.[57] Precision medicine aims to advance risk prediction to individualized therapies based on composite risk factors. Dense molecular and phenotyping profiling toward this goal also facilitates the discovery of broadly applicable novel therapies.

Target Discovery and Clinical Trial Prediction

Genetic variants that alter protein activity can provide robust inferences regarding the outcomes of pharmacologic manipulation before embarking on drug development. Furthermore, identification

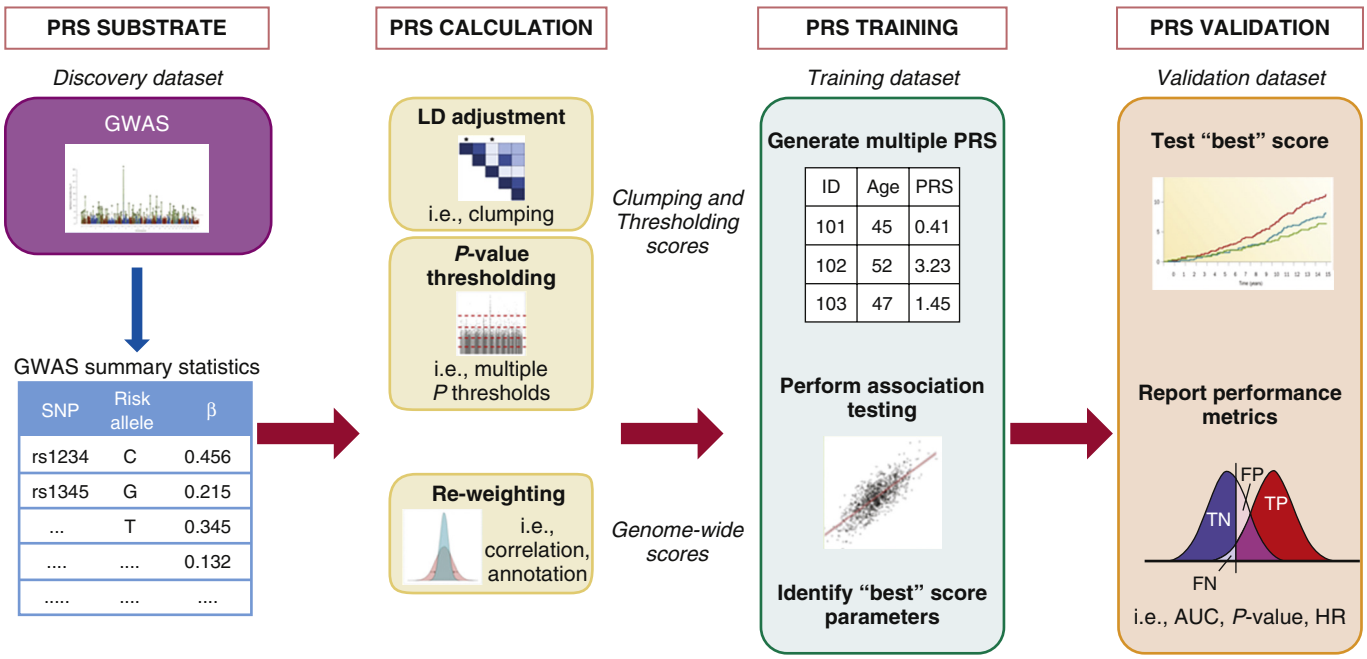

FIGURE 7.9 Development of polygenic risk scores. (Adapted from Aragam KG, Natarajan P. Polygenic scores to assess atherosclerotic cardiovascular disease risk: clinical perspectives and basic implications. *Circ Res.* 2020;126:1159–77.)

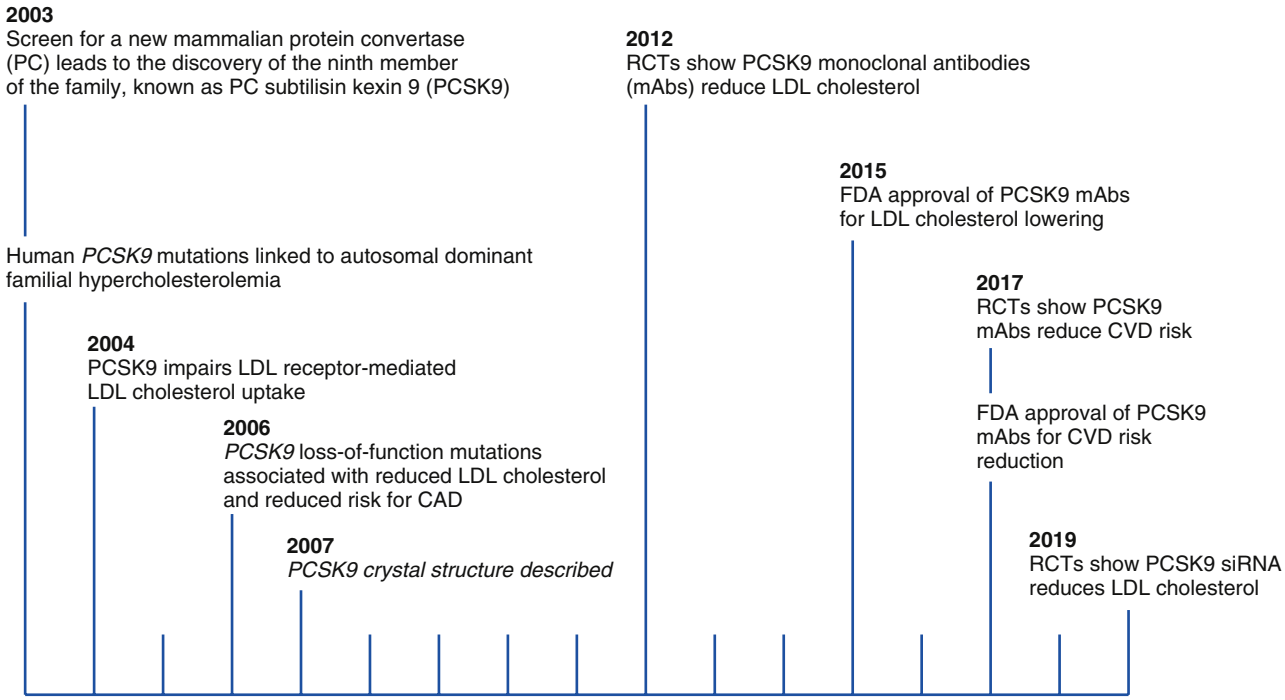

FIGURE 7.10 Timeline of PCSK9 discovery, evidence, and clinical implementation of monoclonal antibodies targeting PCSK9. (Adapted from Natarajan P, Kathiresan S. PCSK9 Inhibitors. *Cell*. 2016;165:1037.)

of putative causal biomarkers through MR with common genetic variants can also prioritize therapeutic targets. However, there are key distinctions between target modulation by genetic variation versus pharmacotherapies. First, for common diseases assessed in event-driven RCTs (often with 3 to 6 years of follow-up), reduction of events is typically assessed among individuals already with prevalent disease. In contrast, genetic variants model target modulation before the onset of disease. Second, related to the aforementioned concept, target modulation via genetic variants occurs at birth as opposed to middle age as in clinical trials. Third, target activity at the relevant tissue may be different for diverse pharmacotherapies and may not recapitulate relevant tissue-specific effects regulated by genetic alleles. Despite these caveats, a recent analysis indicated that prioritizing targets with human genetic validation may double the success rate of clinical development.[58]

PCSK9 (see also Chapter 27)

PCSK9 provides a prime example of successful therapeutic discovery from human genetics (Fig. 7.10). As discussed earlier, human genetic evidence for the relationship of *PCSK9* with blood cholesterol and CAD began with the identification of *PCSK9* gain-of-function variants in familial hypercholesterolemia families. Work described less than two decades ago identified that two nonsense *PCSK9* variants (p.Y142X and p.C679X) were particularly common (1% to 2%) specifically among individuals of African ancestry, and another disruptive missense *PCSK9* variant (p.R46L) was more common (3%) among those of European ancestry. African Americans with p.Y142X or p.C679X had 28% lower LDL-C concentration and 89% lower risk for CAD. European Americans with p.R46L had 15% lower LDL-C concentration and 50% lower risk for CAD. Additionally, rare individuals naturally carrying two nonsense *PCSK9* variants with lifelong genetic absence of PCSK9 ("human knockouts") and LDL-C levels of approximately 10 mg/dL appear to be healthy.

The aforementioned human genetic observations, as well as advances in understanding PCSK9 structure and function, spurred rapid therapeutic development. Over a relatively short period of time, PCSK9 monoclonal antibodies have come into widespread use to lower CAD risk. Two monoclonal antibodies targeting PCSK9 were shown to reduce LDL-C by approximately 50% and reduce the risk for major adverse cardiovascular events by approximately 85% over 3 years among individuals with atherosclerotic cardiovascular disease and LDL-C greater than 70 mg/dL in clinical trials.[59,60]

APOC3

Other genes, such as *APOC3*, have been similarly prioritized for CAD. Apolipoprotein C-III, encoded by *APOC3*, promotes the synthesis of and delays the clearance of triglyceride-rich lipoproteins. In 2008, genome-wide association analysis of a Lancaster Amish cohort with respect to fasting and postprandial triglycerides identified a large-effect common (5% carrier rate) noncoding variant near *APOC3*; sequencing indicated the sentinel SNP was tagging a nonsense variant in *APOC3* (p.R19X), which is now appreciated to be a founder variant in the Amish. In the Amish, presence of this variant was associated with a lower burden of subclinical coronary atherosclerosis. More recently, whole exome sequencing of European Americans and African Americans in the general population showed that *APOC3* p.R19X and other loss-of-function variants similarly reduced triglyceride concentrations and also were associated with reduced risk for CAD.[26] Among a cohort of adults living in Pakistan, where consanguinity is more common, several individuals homozygous for *APOC3* p.R19X were identified who have markedly reduced fasting and postprandial triglycerides.[5] Whether pharmacologic inhibition of apolipoprotein C-III leads to reduced CAD risk remains to be tested.

On-Target Therapeutic Side Effect Prediction

In addition to testing the association between genetic variants and primary outcomes for target efficacy assessment, one may evaluate their relationships with diverse clinical outcomes. In phenome-wide association studies (pheWASs), investigators can anticipate the beneficial and adverse consequences of modulating drug targets of interest. A systemic analysis indicates that drug side effects are more likely to occur when predicted from genetic association analyses.[61]

Among various research opportunities, contemporary densely phenotyped mega biobanks now provide the opportunity for large-scale pheWASs. The U.S. Precision Medicine Initiative led to the creation of the AllOfUs cohort, planned to comprise 1 million diverse Americans recruited through and outside of health care systems. Other cohorts of comparable size include the U.K. Biobank, Million Veterans Program, Biobank Japan, China Kadoorie Biobank, FinnGen, deCODE Genetics, and the eMERGE Network of health care system biobanks.

Precision Medicine

Pharmacogenomics refers to using genetics, in addition to clinical factors, to mitigate adverse drug reactions. Precision medicine aims to

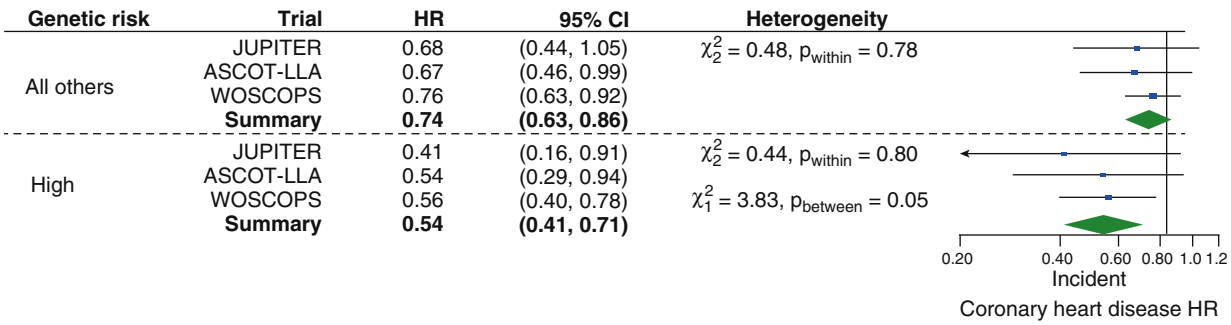

FIGURE 7.11 Forest plot of incident coronary artery disease risk from statin versus placebo by coronary artery disease polygenic risk group in three statin primary prevention trials. High coronary artery disease polygenic risk group refers to the top 20th percentile. For a given degree of low-density lipoprotein cholesterol lowering from statins, clinical benefit is greater among those with high coronary artery disease polygenic risk. (Adapted from Natarajan P, Young R, Stitziel NO, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation*. 2017;135[22]:2091–2101.)

extend this concept by using diverse factors, including genetics, to identify individuals more likely to benefit from preventive therapies. Currently, CAD-preventive pharmacotherapies are titrated to blood cholesterol, blood pressure, and glycemic indices, and therapies are further escalated for those with greater absolute intermediate-term risk.[31] Risk refinement from human genetics may guide therapeutic escalation. Although genetically ascertained RCTs are lacking; ongoing post-hoc analyses within completed clinical trials have led to promising hypotheses.

CYP2C19 (see also Chapter 38)

Along with the use of aspirin, inhibition of platelet P2Y purinoceptor 12 (P2Y12) receptors is standard-of-care therapy as an adjunct to percutaneous coronary intervention (PCI). Clopidogrel, the most widely prescribed P2Y12 inhibitor, is an inactive prodrug converted to its active form largely by cytochrome P-450 2C19 (CYP2C19) in the liver. Several *CYP2C19* genetic polymorphisms influencing enzymatic function have been described, with two relatively common loss-of-function variants (*CYP2C19**2, which disrupts splicing, and *CYP2C19**3, which is a nonsense variant). The allele frequency of *CYP2C19**2 is 30% in South Asians and East Asians, 17% in Europeans and Africans, and 10% in Latinos. The allele frequency of *CYP2C19**3 is 6% in East Asians.

Carriers of these alleles have reduced antiplatelet effects from clopidogrel. In RCTs of clopidogrel-treated patients undergoing PCI, carriers had a greater risk of adverse outcomes, leading to a U.S. Food and Drug Administration black box warning in 2010 recommending alternative antiplatelet agents for poor metabolizers of clopidogrel. Given the lack of prospective genotype-guided RCTs when they were written, guidelines in 2016 recommended against routine *CYP2C19* genotyping but noted that testing may be considered in patients at increased risk for poor clinical outcomes.[62]

More recently, a *CY2C19* genotype-guided strategy was assessed in a prospective RCT among patients undergoing PCI.[63] In the genotype-guided group, carriers of *CY2C19**2 or *CYP2C19**3 received ticagrelor or prasugrel, while noncarriers received clopidogrel. All participants in the standard-of-care group received ticagrelor or prasugrel. The genotype-guided group was noninferior to the standard-treatment group with respect to thrombotic events and had a 2.7% absolute risk reduction in bleeding events.

Familial Hypercholesterolemia

Retrospective observational analyses indicate a greater absolute and relative risk reduction in major adverse cardiovascular events from cholesterol lowering among those with FH variants compared with those without. Nonstatin cholesterol-lowering medicines, such as ezetimibe, PCSK9 monoclonal antibodies, and bempedoic acid, came to market with FH being the initial approved indication. In the primary prevention setting, guidelines recommend the use of additional nonstatin cholesterol-lowering medicines as needed to attain stricter LDL-C targets in patients with FH.[31]

Polygenic Coronary Artery Disease

Post hoc subgroup analyses within prospective clinical trials also indicate greater clinical cardiovascular benefit of cholesterol-lowering medicines among those with high polygenic CAD risk, even though LDL-C is typically only mildly elevated in this setting and LDL-C lowering is similar.[50,64–66] In the primary prevention setting, statin therapy versus placebo was associated with greater absolute and relative risk reduction for those with high CAD PRS versus all others (Fig. 7.11).[50,64] In the secondary prevention setting, a CAD PRS predicted recurrent events, and therapy with a PCSK9 monoclonal antibody versus placebo was associated with greater absolute and relative risk reduction for those with high CAD PRS versus all others.[65,66] A CAD PRS may help identify those more likely to clinically benefit from LDL-C-lowering therapies, which requires assessment in genotype-guided prospective clinical trials.

NEXT-GENERATION TECHNOLOGIES AND THERAPEUTICS

Human genetics, including its application to cardiovascular disease, continues to progress at a rapid rate. Novel experimental and analytic methods promise to expand our understanding of cardiovascular disease as well as develop new pharmacotherapies.

Somatic Genomics (see also Chapter 24)

Age remains the most important risk factor for CAD, but age-related factors causally contributing to CAD remain incompletely understood. Large-scale NGS of blood DNA showed that a large number of individuals (up to 1 in 10 adults older than 70 years) have clonal hematopoiesis of indeterminate potential (CHIP), an age-related phenomenon associated with the clonal selection of cancer-predisposing mutations (typically in *DNMT3A*, *TET2*, *ASXL1*, or *JAK2*) in the blood without cytopenia, dysplasia, or neoplasia. Although CHIP associates strongly with future risk of blood cancer, it has recently been linked to CAD in humans and murine models.[67–69] Inhibition of the NLRP3 inflammasome in mice with experimental atherosclerosis appears to reduce atherosclerosis burden to a greater degree in those with Tet2 loss of function versus without.[68] Applying principles of MR, investigators showed that germline genetic deficiency of *IL6R*, which encodes the interleukin (IL)-6 receptor in the NLRP3 inflammasome pathway, associated with a larger degree of reduced risk for CAD among those with CHIP versus without (Fig. 7.12).[69] These data indicate that modulation of this pathway may be particularly beneficial for those with CHIP.

Epigenetics

Epigenetics may contribute to CAD, because environmental factors associated with epigenetic changes such as altered histone acetylation and DNA methylation are also correlated with CAD risk and advanced atherosclerosis features.[70] For example, increased expression of histone deacetylase 3 has been observed at sites prone to atherogenesis, and increased expression of histone deacetylase 9 (HDAC9) associates with proinflammatory macrophage concentrations within atherosclerotic plaques. Common genetic variants near *HDAC9* associate with vascular calcification and myocardial infarction risk, and inhibition of

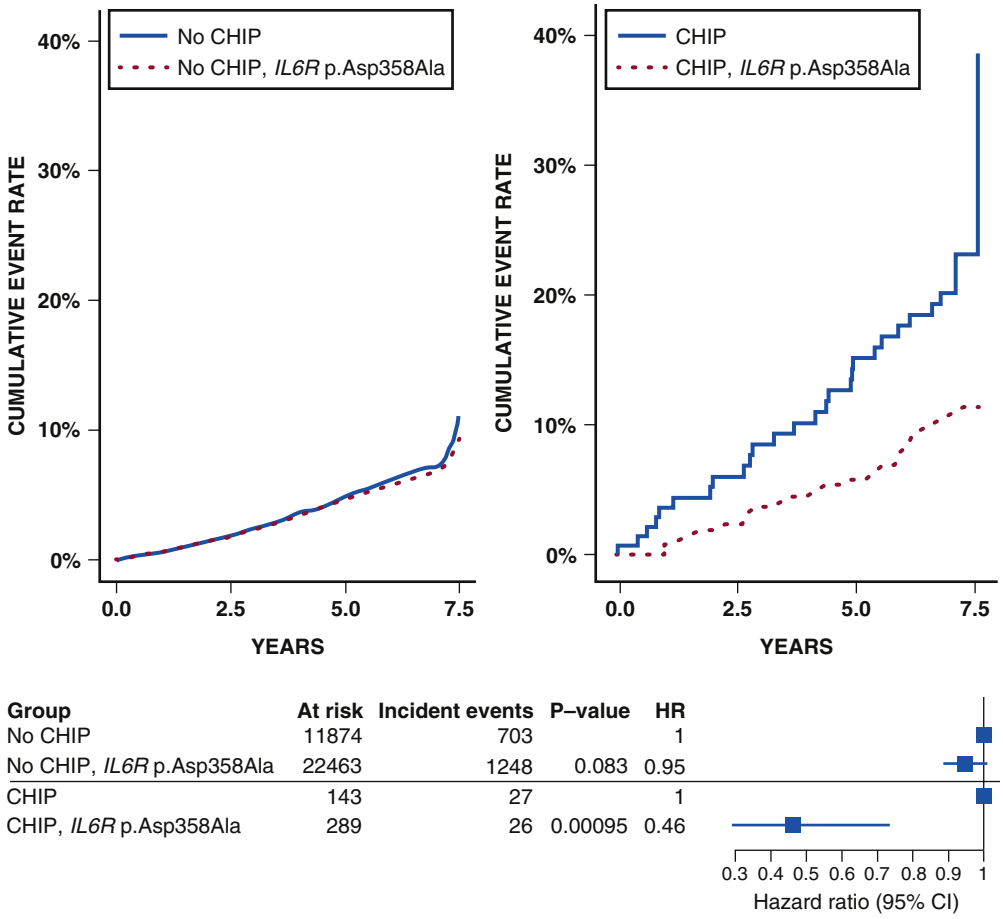

FIGURE 7.12 Cumulative incident cardiovascular risk associated with carrying *IL6R* p.Asp358Ala stratified by the presence of clonal hematopoiesis of indeterminate potential (CHIP). Cardiovascular risk reduction related to *IL6R* p.Asp358Ala is greater among those who develop CHIP with resultant cumulative risk similar to those without CHIP. (Adapted from Bick AG, Pirruccello JP, Griffin GK, et al. Genetic IL-6 signaling deficiency attenuates cardiovascular risk in clonal hematopoiesis. *Circulation*. 2019; 141[2]:124–31.)

HDAC9 in human aortic smooth muscle cells leads to reduced calcification in vitro.[71]

Methylation at specific genomic regions has been linked to CAD risk. A recent large-scale longitudinal analysis of 11,461 individuals for whom leukocyte genome-wide DNA methylation was interrogated with methylation arrays, 52 CpG methylation sites were associated with incident CAD risk.[72] MR analyses indicate that two of these CpG sites relate causally to CAD; both sites are in noncoding intergenic regions. Genetic variants associated with one of the CpG sites influence expression of *ITGA6* (which encodes integrin subunit alpha 6), and genetic variants for the other site influence expression of the long ncRNA RP4-555D20.2.

As noted in the aforementioned example, ncRNAs (microRNAs and long ncRNAs) have emerged as potential modulators of atherosclerosis. First, a high-throughput genome-wide in vitro screen for microRNAs regulating *LDLR* expression in hepatocytes prioritized miR-148a as a negative regulator.[73] In hypercholesterolemic mice, inhibition of miR-148a led to increased hepatic *LDLR* expression and a resultant decrease in LDL-C. Second, CAD risk alleles in the chromosome 9p21 GWAS locus influence the expression of the long ncRNA *ANRIL*. A linear form of *ANRIL* is enriched among those with atherosclerosis, but a circular form may control rRNA maturation in vascular smooth muscle cells and macrophages leading to apparent atheroprotection.[74]

Single-Cell Ribonucleic Acid Sequencing (see also Chapter 24)

NGS is applied to transcription profiling through RNA sequencing (RNA-seq) of tissue. Single-cell RNA sequencing (scRNA-seq) with complementary use of microfluidics may facilitate the characterization of diverse cell types in pathologic processes, discovery of novel cell populations, improved understanding of regulatory relationships between genes, and tracking of the development of specific cellular lineages. Compared with bulk RNA-seq, scRNA-seq has added technical challenges related to low starting material amount and added noise from stochastic or physiologic transcription variation. Experimental and computational tools are being developed and optimized to address these issues. Additionally, novel methods are being developed to (1) also transpose spatial information with gene expression relationships and (2) apply emerging unsupervised clustering and machine learning methods.

Immune cell profiling with atherosclerotic plaques has typically used immunostaining or fluorescence-activated cell sorting (FACS) but now frequently uses scRNA-seq technology. ScRNA-seq and single-cell proteomics analysis of carotid atherosclerotic plaques in patients with and without recent strokes identified novel activated macrophage and T cell subsets.[75] Using FACS to isolate murine vascular smooth muscle cells in atherosclerotic lesions, investigators then applied scRNA-seq to discover a new subpopulation of fibroblast-like cells termed "fibromyocytes"; scRNA-seq in human atherosclerotic lesions also identified fibromyocytes.[76] Knockout of mouse *Tcf21*, a gene prioritized from CAD GWASs, specifically in vascular smooth muscle cells led to fewer fibromyocytes.

Therapeutically Targeting the Genome

In addition to target discovery and prioritization, insights from human genetics have also led to novel approaches for therapeutic targeting. Conventional pharmacotherapies target proteins but newer classes of medicines target more proximal gene product, mRNAs. The two major RNA therapeutic approaches use (1) antisense oligonucleotides (ASOs) to inhibit mRNA translation and (2) oligonucleotides to activate RNA interference (RNAi) to inhibit mRNA translation. Even more proximally, gene therapy is used to circumvent genetically deficient gene products or augment cardioprotective genes. Lastly, emerging methods in gene editing are being explored to target genes or even correct pathogenic variants.

ASOs are synthetic single-stranded DNA sequences designed to bind and inactivate mRNAs produced by a specific gene. In addition to steric hindrance, the resultant RNA-DNA heteroduplex induces RNase H endonuclease activity that degrades in target mRNA and ultimately reduces target gene translation. ASOs are typically 20 base pairs in length and target either the initiation code or splice sites while minimizing polymorphic regions to enhance specificity. Phosphorothiolation of ASOs enables binding to plasma proteins to extend half-life, reduces renal excretion, and improve bioavailability but may lead to thrombocytopenia. Some ribose modifications to improve stability and affinity have been linked to hepatoxicity. In 2013, the FDA approved mipomersen, an ASO targeting *APOB* mRNAs, for homozygous FH but its use is limited by hepatotoxicity. ASOs have been developed for several lipid-related targets, including *APOC3*, *ANGPTL3*, and *LPA*, and are being assessed for effects on cardiovascular outcomes. Selective targeting of RNA therapeutics to hepatocytes with oligosaccharide ligands of the asialoglycoprotein receptor can reduce the doses needed and markedly minimize unwanted actions such as injection site reactions (see also Chapter 27).

RNAi is a naturally occurring eukaryotic innate immune response of sequence-specific mRNA degradation induced by foreign long double-stranded RNAs (dsRNAs). In mammalian cells, dsRNAs induce a strong interferon response, resulting in their processing to single-stranded approximately 22-base pair small interfering RNAs (siRNAs) by Dicer. For human therapeutics, synthetic siRNAs activate RNAi by incorporating in the RNA-induced silencing complex (RISC), leading to mRNA-specific translational repression and degradation. Inclisiran, a twice-annually administered siRNA targeting *PCSK9*, reduces LDL-C safely in RCTs.[77]

Gene-editing technologies leveraging bacterial immune systems now enjoy ubiquitous use for research and have received increasing attention as a therapeutic modality. Clustered regularly interspaced short palindromic repeat (CRISPR) RNAs and CRISPR-associated proteins (Cas, particularly Cas9) can be reprogrammed to target specific genomic DNA sequences. After introducing site-specific double-stranded DNA breaks, endogenous DNA repair mechanisms are activated. Typically, error-prone nonhomologous end joining (NHEJ) is activated and can be used to produce gene knockouts. In mice, using adenovirus vectors, CRISPR-Cas9 can efficiently introduce loss-of-function mutations in *Pcsk9* in the liver, with resultant reduced cholesterol levels.[78] Alternatively, a homologous repair template can stimulate the less error-prone homology-directed repair (HDR) to facilitate desired changes; emerging methods focus on maximizing HDR efficiency versus NHEJ-mediated repair. Base editing uses fusions of Cas proteins with deaminases to facilitate transition mutations (i.e., C→T and A→G conversions). In mice, using adenovirus vectors, base editing efficiently introduced loss-of-function point mutations in *Pcsk9* and *Angptl3* in the liver.[79]

FUTURE PERSPECTIVES

Technical advances in large-scale high-throughput genomic profiling continue to yield accelerating advances for cardiovascular disease, with diverse insights and applications for the most common form—CAD. Larger-scale application of whole genome sequencing in increasingly diverse datasets will: (1) better define both lipid and nonlipid genes responsible for CAD, (2) enable interpretation of the bulk of genomic variation (rare, noncoding) which is poorly understood, and (3) improve genetic risk prediction across diverse ethnicities. Genomic interpretation for diverse clinical outcomes, including within the context of ongoing clinical trials, may facilitate therapeutic paradigms maximizing efficacy and minimizing side effects for individual patients.

Driven by public interests and scientific observations, PRSs for CAD and other diseases will likely soon become broadly available. Given the widespread availability of genetic testing and democratized interpretation, PRSs may increasingly enter clinical practice. Ethical and confidentiality issues require due considerations. Broad availability of such testing will also enable molecularly targeted RCTs to evaluate specific strategies as well as enrich for events to improve trial efficiency. An intriguing prospect is whole genome sequencing earlier in life to inform baseline disease trajectories toward "primordial prevention."

The relationship between CHIP and cardiovascular disease implicates the prospect of novel, molecularly guided therapies. Unlike the current model of escalating therapies with the accumulation of risk factors, the presence of CHIP may prompt the orthogonal use of NLRP3/IL-1β/IL-6 inhibiting therapies.

Lastly, as the field continues to refine interpretation of the genome and prioritize therapeutic targets, advances in genome-based therapies offer the promise of durable molecular interventions to prevent and treat cardiovascular disease.

REFERENCES

1. GBD 2017 Disease and Injury Incidence and Prevalence Collaborators. Global, regional, and national incidence, prevalence, and years lived with disability for 354 diseases and injuries for 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet*. 2018;392:1789–1858.
2. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*. 2016;536:285–291.
3. Natarajan P, Gold NB, Bick AG, et al. Aggregate penetrance of genomic variants for actionable disorders in European and African Americans. *Sci Transl Med*. 2016;8:364ra151.
4. Consortium GT, Laboratory DA, Coordinating Center -Analysis Working G, et al. Genetic effects on gene expression across human tissues. *Nature*. 2017;550:204–213.
5. Saleheen D, Natarajan P, Armean IM, et al. Human knockouts and phenotypic analysis in a cohort with a high rate of consanguinity. *Nature*. 2017;544:235–239.
6. Liu X, Wu C, Li C, Boerwinkle E. dbNSFP v3.0: a one-stop database of functional predictions and annotations for human nonsynonymous and splice-site SNVs. *Human Mutat*. 2016;37:235–241.
7. Roadmap Epigenomics Consortium, Kundaje A, Meuleman W, et al. Integrative analysis of 111 reference human epigenomes. *Nature*. 2015;518:317–330.
8. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat Genet*. 2016;48:1279–1283.
9. Shendure J, Balasubramanian S, Church GM, et al. DNA sequencing at 40: past, present and future. *Nature*. 2017;550:345–353.
10. Jain M, Koren S, Miga KH, et al. Nanopore sequencing and assembly of a human genome with ultra-long reads. *Nature Biotechnol*. 2018;36:338–345.
11. Gidding SS, Champagne MA, de Ferranti SD, et al. The agenda for familial hypercholesterolemia: a scientific statement from the American Heart Association. *Circulation*. 2015;132:2167–2192.
12. Keramati AR, Fathzadeh M, Go GW, et al. A form of the metabolic syndrome associated with mutations in DYRK1B. *New Engl J Med*. 2014;370:1909–1919.
13. Palmer C, Pe'er I. Statistical correction of the Winner's Curse explains replication variability in quantitative trait genome-wide association studies. *PLoS Genetics*. 2017;13:e1006916.
14. Musunuru K, Kathiresan S. Genetics of common, complex coronary artery disease. *Cell*. 2019;177:132–145.
15. Loh PR, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat Genet*. 2015;47:284–290.
16. Zhou W, Nielsen JB, Fritsche LG, et al. Efficiently controlling for case-control imbalance and sample relatedness in large-scale genetic association studies. *Nat Genet*. 2018;50:1335–1341.
17. Klarin D, Damrauer SM, Cho K, et al. Genetics of blood lipids among ~300,000 multi-ethnic participants of the Million Veteran Program. *Nat Genet*. 2018;50:1514–1523.
18. Nioi P, Sigurdsson A, Thorleifsson G, et al. Variant ASGR1 associated with a reduced risk of coronary artery disease. *New Engl J Med*. 2016;374:2131–2141.
19. Liu DJ, Peloso GM, Yu H, et al. Exome-wide association study of plasma lipids in >300,000 individuals. *Nat Genet*. 2017;49:1758–1766.
20. Lu X, Peloso GM, Liu DJ, et al. Exome chip meta-analysis identifies novel loci and East Asian-specific coding variants that contribute to lipid levels and coronary artery disease. *Nat Genet*. 2017;49:1722–1730.
21. Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc Res*. 2018;114:1241–1257.
22. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators, Stitziel NO, et al. Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *New Engl J Med*. 2016;374:1134–1144.
23. Kessler T, Wobst J, Wolf B, et al. Functional characterization of the GUCY1A3 coronary artery disease risk locus. *Circulation*. 2017;136:476–489.
24. Emdin CA, Khera AV, Klarin D, et al. Phenotypic consequences of a genetic predisposition to enhanced nitric oxide signaling. *Circulation*. 2018;137:222–232.
25. Do R, Stitziel NO, Won HH, et al. Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature*. 2015;518:102–106.
26. Crosby J, Peloso GM, Auer PL, et al. Loss-of-function mutations in APOC3, triglycerides, and coronary disease. *New Engl J Med*. 2014;371:22–31.
27. Myocardial Infarction Genetics Consortium Investigators, Stitziel NO, Won HH, et al. Inactivating mutations in NPC1L1 and protection from coronary heart disease. *New Engl J Med*. 2014;371:2072–2082.
28. Stitziel NO, Khera AV, Wang X, et al. ANGPTL3 deficiency and protection against coronary artery disease. *J Am Coll Cardiol*. 2017;69:2054–2063.
29. Hingorani AD, Kuan V, Finan C, et al. Improving the odds of drug development success through human genomics: modelling study. *Sci Rep*. 2019;9:18911.
30. White J, Swerdlow DI, Preiss D, et al. Association of lipid fractions with risks for coronary artery disease and diabetes. *JAMA Cardiol*. 2016;1:692–699.
31. Grundy SM, Stone NJ, Bailey AL, et al. 2018 AHA/ACC/AACVPR/AAPA/ABC/ACPM/ADA/AGS/APhA/ASPC/NLA/PCNA guideline on the management of blood cholesterol: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines. *J Am Coll Cardiol*. 2018.
32. Zekavat SM, Ruotsalainen S, Handsaker RE, et al. Deep coverage whole genome sequences and plasma lipoprotein(a) in individuals of European and African ancestries. *Nat Commun*. 2018;9:2606.
33. Nordestgaard BG, Langsted A. Lipoprotein (a) as a cause of cardiovascular disease: insights from epidemiology, genetics, and biology. *J Lipid Res*. 2016;57:1953–1975.
34. Emdin CA, Khera AV, Natarajan P, et al. Phenotypic characterization of genetically lowered human lipoprotein(a) levels. *J Am Coll Cardiol*. 2016;68:2761–2772.
35. Klarin D, Lynch J, Aragam K, et al. Genome-wide association study of peripheral artery disease in the Million Veteran Program. *Nat Med*. 2019;25:1274–1279.
36. Emdin CA, Khera AV, Natarajan P, et al. Genetic association of waist-to-hip ratio with cardiometabolic traits, type 2 diabetes, and coronary heart disease. *J Am Med Assoc*. 2017;317:626–634.
37. Lotta LA, Wittemans LBL, Zuber V, et al. Association of genetic variants related to gluteofemoral vs abdominal fat distribution with type 2 diabetes, coronary disease, and cardiovascular risk factors. *J Am Med Assoc*. 2018;320:2553–2563.

38. Burgess S, Dudbridge F, Thompson SG. Combining information on multiple instrumental variables in Mendelian randomization: comparison of allele score and summarized data methods. *Stat Med*. 2016;35:1880–1906.
39. Greco MF, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat Med*. 2015;34:2926–2940.
40. Bowden J, Del Greco MF, Minelli C, et al. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol*. 2016;45:1961–1974.
41. Verbanck M, Chen CY, Neale B, Do R. Detection of widespread horizontal pleiotropy in causal relationships inferred from Mendelian randomization between complex traits and diseases. *Nat Genet*. 2018;50:693–698.
42. Zhu Z, Zheng Z, Zhang F, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun*. 2018;9:224.
43. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Gen Med*. 2015;17(5):405–424.
44. Kalia SS, Adelman K, Bale SJ, et al. Recommendations for reporting of secondary findings in clinical exome and genome sequencing, 2016 update (ACMG SF v2.0): a policy statement of the American College of Medical Genetics and Genomics. *Gen Med*. 2017;19:249–255.
45. Alver M, Palover M, Saar A, et al. Recall by genotype and cascade screening for familial hypercholesterolemia in a population-based biobank from Estonia. *Gen Med*. 2018;21(5):1173–1180.
46. Benn M, Watts GF, Tybjaerg-Hansen A, Nordestgaard BG. Mutations causative of familial hypercholesterolaemia: screening of 98 098 individuals from the Copenhagen General Population Study estimated a prevalence of 1 in 217. *Eur Heart J*. 2016;37:1384–1394.
47. Khera AV, Won HH, Peloso GM, et al. Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J Am Coll Cardiol*. 2016;67:2578–2589.
48. Martin AR, Gignoux CR, Walters RK, et al. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100:635–649.
49. Khera AV, Emdin CA, Drake I, et al. Genetic risk, adherence to a healthy lifestyle, and coronary disease. *New Engl J Med*. 2016;375:2349–2358.
50. Natarajan P, Young R, Stitziel NO, et al. Polygenic risk score identifies subgroup with higher burden of atherosclerosis and greater relative benefit from statin therapy in the primary prevention setting. *Circulation*. 2017;135(22):2091–2101.
51. Tada H, Melander O, Louie JZ, et al. Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur Heart J*. 2016;37:561–567.
52. Khera AV, Chaffin M, Aragam KG, et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat Genet*. 2018;50:1219–1224.
53. Inouye M, Abraham G, Nelson CP, et al. Genomic risk prediction of coronary artery disease in 480,000 adults: Implications for primary prevention. *J Am Coll Cardiol*. 2018;72:1883–1893.
54. Elliott J, Bodinier B, Bond TA, et al. Predictive accuracy of a polygenic risk score-enhanced prediction model vs a clinical risk score for coronary artery disease. *J Am Med Assoc*. 2020;323:636–645.
55. Mosley JD, Gupta DK, Tan J, et al. Predictive accuracy of a polygenic risk score compared with a clinical risk score for incident coronary heart disease. *J Am Med Assoc*. 2020;323:627–635.
56. Aragam KG, Natarajan P. Polygenic scores to assess atherosclerotic cardiovascular disease risk: clinical perspectives and basic implications. *Circ Res*. 2020;126:1159–1177.
57. Collins FS, Varmus H. A new initiative on precision medicine. *New Engl J Med*. 2015;372:793–795.
58. Nelson MR, Tipney H, Painter JL, et al. The support of human genetic evidence for approved drug indications. *Nat Genet*. 2015;47:856–860.
59. Sabatine MS, Giugliano RP, Keech AC, et al. Evolocumab and clinical outcomes in patients with cardiovascular disease. *New Engl J Med*. 2017;376(18):1713–1722.
60. Schwartz GG, Steg PG, Szarek M, et al. Alirocumab and cardiovascular outcomes after acute coronary syndrome. *New Engl J Med*. 2018;379:2097–2107.
61. Nguyen PA, Born DA, Deaton AM, et al. Phenotypes associated with genes encoding drug targets are predictive of clinical trial side effects. *Nat Commun*. 2019;10:1579.
62. Levine GN, Bates ER, Bittl JA, et al. 2016 ACC/AHA guideline focused update on duration of dual antiplatelet therapy in patients with coronary artery disease: a report of the American College of Cardiology/American Heart Association task force on clinical practice guidelines: an update of the 2011 ACCF/AHA/SCAI guideline for percutaneous coronary intervention, 2011 ACCF/AHA guideline for coronary artery bypass graft surgery, 2012 ACC/AHA/ACP/AATS/PCNA/SCAI/STS guideline for the diagnosis and management of patients with stable ischemic heart disease, 2013 ACCF/AHA guideline for the management of ST-Elevation myocardial infarction, 2014 AHA/ACC guideline for the management of patients with non-ST-Elevation acute coronary syndromes, and 2014 ACC/AHA guideline on perioperative cardiovascular evaluation and management of patients undergoing noncardiac surgery. *Circulation*. 2016;134:e123–e155.
63. Claassens DMF, Vos GJA, Bergmeijer TO, et al. A genotype-guided strategy for oral P2Y12 inhibitors in primary PCI. *New Engl J Med*. 2019;381:1621–1631.
64. Mega JL, Stitziel NO, Smith JG, et al. Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet*. 2015;385:2264–2271.
65. Damask A, Steg PG, Schwartz GG, et al. Patients with high genome-wide polygenic risk scores for coronary artery disease may receive greater clinical benefit from alirocumab treatment in the ODYSSEY OUTCOMES trial. *Circulation*. 2020;141:624–636.
66. Marston NA, Gurmu Y, Melloni GEM, et al. The effect of PCSK9 (proprotein convertase subtilisin/kexin type 9) inhibition on the risk of venous thromboembolism. *Circulation*. 2020;141:1600–1607.
67. Jaiswal S, Natarajan P, Silver AJ, et al. Clonal hematopoiesis and risk of atherosclerotic cardiovascular disease. *New Engl J Med*. 2017.
68. Fuster JJ, MacLauchlan S, Zuriaga MA, et al. Clonal hematopoiesis associated with TET2 deficiency accelerates atherosclerosis development in mice. *Science (New York, NY)*. 2017;355:842–847.
69. Bick AG, Pirruccello JP, Griffin GK, et al. Genetic IL-6 signaling deficiency attenuates cardiovascular risk in clonal hematopoiesis. *Circulation*. 2019;141(2):124–131.
70. Rizzacasa B, Amati F, Romeo F, et al. Epigenetic modification in coronary atherosclerosis: JACC review topic of the week. *J Am Coll Cardiol*. 2019;74:1352–1365.
71. Malhotra R, Mauer AC, Lino Cardenas CL, et al. HDAC9 is implicated in atherosclerotic aortic calcification and affects vascular smooth muscle cell phenotype. *Nat Genet*. 2019;51:1580–1587.
72. Agha G, Mendelson MM, Ward-Caviness CK, et al. Blood leukocyte DNA methylation predicts risk of future myocardial infarction and coronary heart disease. *Circulation*. 2019;140:645–657.
73. Goedeke L, Rotllan N, Canfran-Duque A, et al. MicroRNA-148a regulates LDL receptor and ABCA1 expression to control circulating lipoprotein levels. *Nature Med*. 2015;21:1280–1289.
74. Holdt LM, Stahringer A, Sass K, et al. Circular non-coding RNA ANRIL modulates ribosomal RNA maturation and atherosclerosis in humans. *Nat Commun*. 2016;7:12429.
75. Fernandez DM, Rahman AH, Fernandez NF, et al. Single-cell immune landscape of human atherosclerotic plaques. *Nature Med*. 2019;25:1576–1588.
76. Wirka RC, Wagh D, Paik DT, et al. Atheroprotective roles of smooth muscle cell phenotypic modulation and the TCF21 disease gene as revealed by single-cell analysis. *Nat Med*. 2019;25:1280–1289.
77. Ray KK, Wright RS, Kallend D, et al. Two phase 3 trials of inclisiran in patients with elevated LDL cholesterol. *New Engl J Med*. 2020;382:1507–1519.
78. Ding Q, Strong A, Patel KM, et al. Permanent alteration of PCSK9 with in vivo CRISPR-Cas9 genome editing. *Circ Res*. 2014;115:488–492.
79. Chadwick AC, Evitt NH, Lv W, Musunuru K. Reduced blood lipid levels with in vivo CRISPR-cas9 base editing of ANGPTL3. *Circulation*. 2018;137:975–977.
80. Peloso GM, Natarajan P. Insights from population-based analyses of plasma lipids across the allele frequency spectrum. *Curr Opin Genet Dev*. 2018;50:1–6.
81. Natarajan P, Kathiresan S. PCSK9 inhibitors. *Cell*. 2016;165:1037.