DISEASE MECHANISMS

Genetics of coronary artery disease: discovery, biology and clinical translation

Amit V. Khera and Sekar Kathiresan

Abstract | Coronary artery disease is the leading global cause of mortality. Long recognized to be heritable, recent advances have started to unravel the genetic architecture of the disease. Common variant association studies have linked approximately 60 genetic loci to coronary risk. Large-scale gene sequencing efforts and functional studies have facilitated a better understanding of causal risk factors, elucidated underlying biology and informed the development of new therapeutics. Moving forwards, genetic testing could enable precision medicine approaches by identifying subgroups of patients at increased risk of coronary artery disease or those with a specific driving pathophysiology in whom a therapeutic or preventive approach would be most useful.

Heritable
Capable of being transmitted from parent to offspring via genetic variation.

Genetic architecture
The full spectrum of common and rare genetic variation that contributes to a trait of interest.

Linkage analysis
Systematic localization of a genetic region that is co-inherited with a trait of interest in members of a family.

Division of Cardiology, Department of Medicine and Center for Genomic Medicine, Massachusetts General Hospital; Cardiovascular Disease Initiative, Broad Institute of Harvard and Massachusetts Institute of Technology, 185 Cambridge Street, CPZN 5.252, Boston, Massachusetts 02114, USA.

Correspondence to S.K.
skathiresan1@mgh.harvard.edu

doi:10.1038/nrg.2016.160
Published online 13 Mar 2017

Observational epidemiology and translational research efforts have led to significant progress in improving the understanding of the pathophysiology underlying coronary artery disease (CAD) (BOX 1). Prevention and treatment strategies developed on the basis of this knowledge led to a >50% decrease in age-adjusted CAD mortality rate in the United States between 1980 and 2000 (REF. 1). However, despite these advances, CAD remains the leading global cause of mortality[2]. Current predictions estimate that more than 900,000 individuals in the United States will suffer a myocardial infarction (heart attack) or die of CAD this year[3].

As with most complex diseases, an individual's risk of developing CAD is modulated by an interplay between genetic and lifestyle factors[4]. Clinical observations dating back to the 1950s have supported the notion that risk of CAD is heritable[5]. A study of more than 20,000 Swedish twins subsequently confirmed this finding of increased risk for CAD among close relatives, and estimated a heritability of ~50% for fatal CAD[6,7]. An analysis that quantified heritability using updated genome-wide approaches similarly estimated the heritability of CAD at 40–50%[8]. In the Framingham Heart Study, a family history of cardiovascular disease in a parent or sibling was a strong predictor of incident disease[9,10]. These seminal studies laid the foundation for the application of emerging human genetics tools to understand the underlying genetic architecture of CAD, to uncover novel biology and to translate these findings into clinical practice.

Substantial progress in our understanding of the genetic underpinnings of CAD has been made since the topic was last reviewed in this journal in 2006 (REF. 11). First reported in 2007, genetic association studies have identified ~60 genetic variants with robust links to risk of CAD. Sequencing studies of rare variants have highlighted pathways underlying CAD and, in several cases, directly catalysed the development of novel therapeutic strategies. Large-scale biobanks containing both genetic and clinical information have enabled the assessment of relationships between a given variant and a broad range of human disease states.

In this Review, we outline research efforts to understand the genetic drivers of CAD, the role of human genetics in catalysing CAD drug discovery efforts and the promises and challenges of integrating genetic information into routine clinical practice.

Gene discovery for CAD

Moving from a recognition of familial patterns to discovering the discrete genetic drivers of CAD has been the primary focus of gene discovery efforts for this disease. Sequencing of the human genome (which was completed in 2003), rapid declines in the costs of genotyping, and data sharing via multinational collaborations have each played a key part in these efforts.

Family-based studies. Detailed studies of families with a predisposition to early-onset CAD, as classically performed via linkage analysis, provided the first opportunity to gain insight into monogenic drivers of CAD. A hereditary pattern of increased cholesterol and premature CAD, now known as familial hypercholesterolaemia, was first described among six patients with xanthomata (skin nodules reflecting the deposition of excess

Monogenic drivers
Variations in a single gene dictates the observed variation in a trait of interest; also referred to as Mendelian disorders.

cholesterol) in 1938 (REF. 12). In 1985, a 5 kb deletion in *LDLR*, the gene encoding the low-density lipoprotein (LDL) receptor, was identified in a patient with familial hypercholesterolaemia and in his mother. This study was the first to demonstrate that a molecular defect in a single gene can drive CAD risk[13]. Impaired receptor-mediated hepatic uptake of LDL leads to substantially increased levels of circulating cholesterol and premature CAD. Family-based studies similarly identified mutations in *APOB* (which encodes apolipoprotein B) and gain-of-function mutations in *PCSK9* (which encodes proprotein convertase subtilisin/kexin type 9) as additional causes of familial hypercholesterolaemia; mutations in *APOB* and *PCSK9* prevent the binding of LDL particles to LDLRs for uptake and promote LDLR catabolism, respectively[14,15]. The genetic underpinnings of autosomal recessive hypercholesterolaemia were linked to null mutations in the genes encoding LDLR adaptor protein 1 (*LDLRAP1*) and ATP-binding cassette sub-family G member 5 (*ABCG5*) or *ABCG8* (REFS 16,17).

Beyond familial hypercholesterolaemia and related conditions, the use of family studies to identify monogenic drivers of CAD has proven challenging. A 2003 analysis of a large family implicated a locus at chromosome 15q26 as an autosomal dominant driver of CAD risk. Sequencing of *MEF2A*, a gene encoding the transcription factor myocyte enhancer factor 2A, which is expressed in the vasculature, identified a 21 bp deletion in each of the affected family members[18]. However, despite biological plausibility, subsequent efforts have failed to confirm an association between loss-of-function variants in *MEF2A* and CAD[19], suggesting that this initial report may have been related to a chance finding. Additional linkage analyses of families with early-onset CAD and metabolic risk factors have noted potentially causal missense mutations in the genes encoding LDLR-related protein 6 (*LRP6*) and dual specificity tyrosine phosphorylation regulated kinase 1B (*DYRK1B*)[20,21]. Owing to the rarity of these genotypes, definitive confirmation of these findings in the general population has not yet been possible.

The careful study of families with extreme phenotypes (for example, onset of CAD at a young age in the absence of traditional risk factors) may still prove useful in gene discovery. However, before assuming causality for any given finding, rigorous replication that places the genetic, bioinformatics and experimental results into a broader context is required[22].

Common variant association studies. Despite the tendency to cluster in families, CAD is a complex and common disorder. Genotyping chips designed to capture the majority of common inter-individual genetic variation provided the foundation for common variant association studies (CVAS), also termed genome-wide association studies (GWAS). Because common variants occur sufficiently often, it is practical to test each variant individually by comparing its frequency in disease cases and disease-free controls[23]. One operational definition of 'common' is variation with an allele frequency of ≥0.5% (one carrier per 100 individuals)[24].

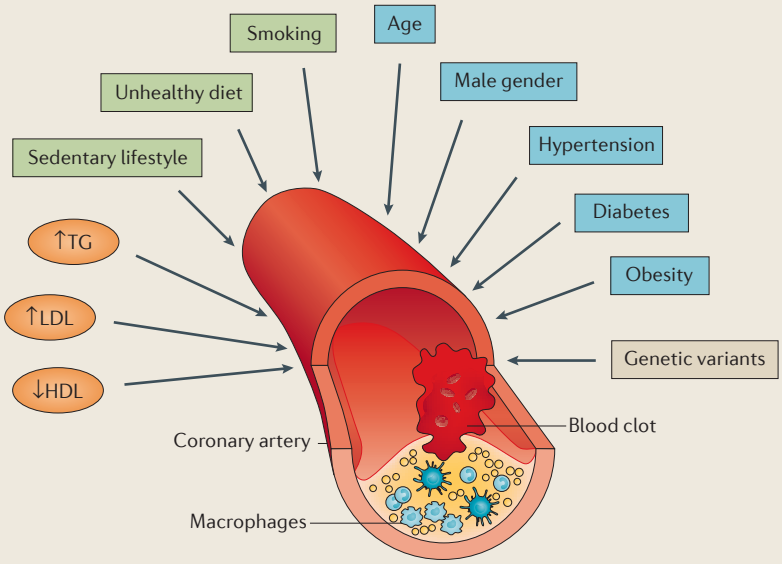
The first CVAS for CAD were published in 2007, when three independent groups reported common variants at the 9p21 locus that were associated with a ~30% increased risk of CAD per copy of the risk allele[25–27]. Subsequent studies have both replicated this finding and extended the association to other vascular phenotypes, including carotid atherosclerosis[28], peripheral arterial disease[29] and stroke[30]. Preliminary evidence suggests that the 9p21 risk variants alter the expression of the non-coding RNA *ANRIL*, thereby altering the activity of two nearby cyclin-dependent kinase inhibitors (CDKN2A and CDKN2B) that are involved in regulating the cell cycle

Box 1 | **The pathophysiology and treatment of coronary artery disease**

Coronary artery disease (CAD) refers to the build-up of atherosclerotic plaque in the blood vessels that supply oxygen and nutrients to the heart (reviewed in REF. 126). The complex process of atherosclerosis begins early in life and is thought to initiate with dysfunction of endothelial cells that line the coronary arteries; these cells are no longer able to appropriately regulate vascular tone (narrowing or constriction of the vessels) with nitric oxide signalling. Progressive infiltration of the vessel wall by lipoprotein particles carrying cholesterol propagates an inflammatory response by cholesterol-loaded macrophage 'foam cells'. Smooth muscle cells underlying the vessel wall proliferate and lead to remodelling of the vessel that can ultimately lead to a narrowing of the vessel that obstructs blood flow. A myocardial infarction (heart attack) is typically caused when a blood clot is incited by a rupture in the surface of the plaque; this process deprives the heart muscle downstream of the blood clot of adequate blood flow and leads to cell death.

Epidemiological studies of CAD demonstrated that age, male gender, smoking, elevated blood pressure, diabetes, obesity and a sedentary lifestyle each lead to an increased risk of suffering from a myocardial infarction (see the figure). Similarly, increased concentrations of circulating low-density lipoprotein (LDL) cholesterol, increased triglyceride (TG)-rich lipoproteins (a form of fat storage) or decreased high-density lipoprotein (HDL) cholesterol are associated with risk of CAD. In clinical practice, these risk factors can be combined to identify subsets of the population at increased risk of CAD who would most benefit from preventive therapies.

Efforts to prevent CAD (reviewed in REF. 127) begin with encouraging adherence to a healthy lifestyle — for example, not smoking, avoiding obesity, a healthy diet and regular exercise — in the population. Higher-risk individuals benefit from additional medications to reduce LDL cholesterol (for example, statins), lower blood pressure or help prevent formation of a blood clot (for example, aspirin). Should an individual ultimately suffer a myocardial infarction, blood flow can be restored via a procedure to place a stent in the narrowed vessel or bypass it via open-heart surgery. Because of a substantial risk of recurrence, medical therapy is intensified in these individuals.



and cellular proliferation[31,32]. Furthermore, inflammatory signalling mediated by interferon-γ may alter long-range DNA interactions, linking the 9p21 risk alleles to *CKDN2A* and *CDNK2B* expression[33]. However, despite scrutiny over the past 10 years, the precise mechanism underlying the 9p21 association remains elusive.

Since 2007, progressively larger sample sizes have been used to interrogate the genetic architecture of CAD, yielding ~60 distinct genetic loci for CAD[34–39]. The results of this cumulative experience permit several conclusions. First, the vast majority of these variants have a minor allele frequency of >5% in the population, are associated with modest increases in CAD risk (for example, <20% change in risk per allele) and cumulatively explain 30–40% of CAD heritability[39,40]. By contrast, 15 low-frequency variants (minor allele frequency <5%) identified using a false-discovery rate threshold explained only 2% of CAD heritability[39]. This pattern of results for CAD is similar to that of other complex diseases, including type 2 diabetes mellitus[41] and schizophrenia[42].

Second, most of the variants identified to date are located outside protein-coding regions. Using a genotyping chip designed to capture the vast majority of coding variants in individuals of European ancestry, a recent CVAS identified robust associations for coding variants at only four loci[43]. Rather, a significant enrichment of variants in regulatory regions has highlighted a predominant impact of CAD risk variants on altering gene expression[39,44].

Third, the genetic loci identified to date have highlighted the biology underlying CAD risk (FIG. 1). Approximately 20% of the loci are located near genes with known roles in metabolism of LDLs, triglyceride-rich lipoproteins (TRLs) or lipoprotein(a) (a modified LDL particle encoded by *LPA*), reinforcing key roles for these pathways in the development of CAD and providing internal validation of CVAS findings. An additional 5–10% of the loci relate to blood pressure, a known and modifiable causal risk factor for CAD. For example, guanylate cyclase 1, soluble, alpha 3 (GUCY1A3) and nitric oxide synthase 3 (NOS3) are key regulators of vascular tone and platelet aggregation. Common DNA sequence variants at the *GUCY1A3* and *NOS3* loci have been associated with both blood pressure and CAD[39,45]. Furthermore, loss-of-function mutations in *GUCY1A3* were associated with risk of myocardial infarction in a large family enriched for premature CAD[46]. The equivalent mutations in mice were shown to accelerate thrombus formation in the microcirculation following local trauma[46]. These data highlight a role of nitric oxide signalling in protection against CAD.

Rare variant association studies. A rapid decline in cost has facilitated large-scale genetic sequencing efforts that enable rare variant association studies (RVAS). Such variants are not typically included on genotyping chips (which are designed to capture variation that is common in the population). Furthermore, rare variants do not occur with sufficient frequency to enable association tests of individual variants. Rather, rare variants in a given gene are aggregated into sets enabling a comparison of the aggregate frequency across disease strata[47]. Ongoing and future studies that perform whole-exome or whole-genome sequencing in a large number of individuals will provide additional statistical power for RVAS.

Studies to date have identified at least nine genes for which an aggregation of rare mutations alters the risk of CAD (TABLE 1). A whole-exome sequencing study comparing ~5,000 cases with early-onset CAD to CAD-free controls was used to scan each gene for association with CAD in an unbiased fashion[48]. Unsurprisingly, the strongest signal was noted for damaging mutations in *LDLR*, which was associated with a fourfold increase in risk of CAD; ~2% of individuals with early-onset CAD harboured a mutation. A second finding was related to inactivating mutations in *PCSK9*. In contrast to the gain-of-function mutations causative for familial hypercholesterolaemia discussed above, gene sequencing revealed two damaging *PCSK9* mutations with an aggregate frequency of 2% of individuals with African ancestry[49]. Carriers of either of these two mutations had substantially lower LDL cholesterol and risk of CAD[50]. Most recently, a whole-genome sequencing study of Icelandic individuals identified a 12 bp deletion that leads to inactivation of *ASGR1* (which encodes asialoglycoprotein receptor)[51]. Heterozygous carriers of this mutation had decreased levels of LDL cholesterol and triglycerides, translating into a decreased risk of CAD.

RVAS have also provided evidence linking genes related to the metabolism of TRLs, in particular those in the lipoprotein lipase (LPL) pathway, with risk of CAD. The enzymatic activity of LPL serves as the rate-determining step in the clearance of dietary fat from the circulation (FIG. 2). Individuals harbouring a heterozygous damaging mutation in *LPL* have increased levels of circulating triglycerides as well as risk of CAD[52]. The activity of LPL is regulated by the protein products of multiple additional genes, several of which have been similarly associated with CAD. Damaging mutations in *APOA5*, which encodes a protein that enhances LPL activity, are associated with increased CAD risk[48]. By contrast, rare mutations in *APOC3* and *ANGPTL4*, the protein products of which inhibit LPL, are associated with decreased CAD risk[43,53–55].

Biological underpinnings of CAD

From association to mechanism. The numerous genetic loci robustly associated with CAD using the CVAS approach have exposed new mechanisms leading to atherosclerosis. For example, a systematic series of experiments for a novel genetic locus at 1p13, which is associated with increased LDL cholesterol and risk of CAD, identified the causal variant that affected *SORT1* expression[34,56,57]. Functional studies confirmed a role for sortilin, the protein product of *SORT1*, in both the secretion of APOB (the major protein component of LDL particles) and LDL catabolism[58,59]. Similar lines of work hold substantial promise for the identification of novel therapeutic targets, particularly for loci that are not linked to pathways (for example, LDL metabolism) that are currently targeted by available drugs.

Allele frequency
The relative frequency of an allele (specific genetic variant) in the population; typically reported as the proportion of all chromosomes in the population that carry an allele.

Inactivating mutations
Variants that disrupt the ability of a given gene to produce its protein product, that is, due to premature truncation, scrambling of the amino acid code or disrupting gene splicing.

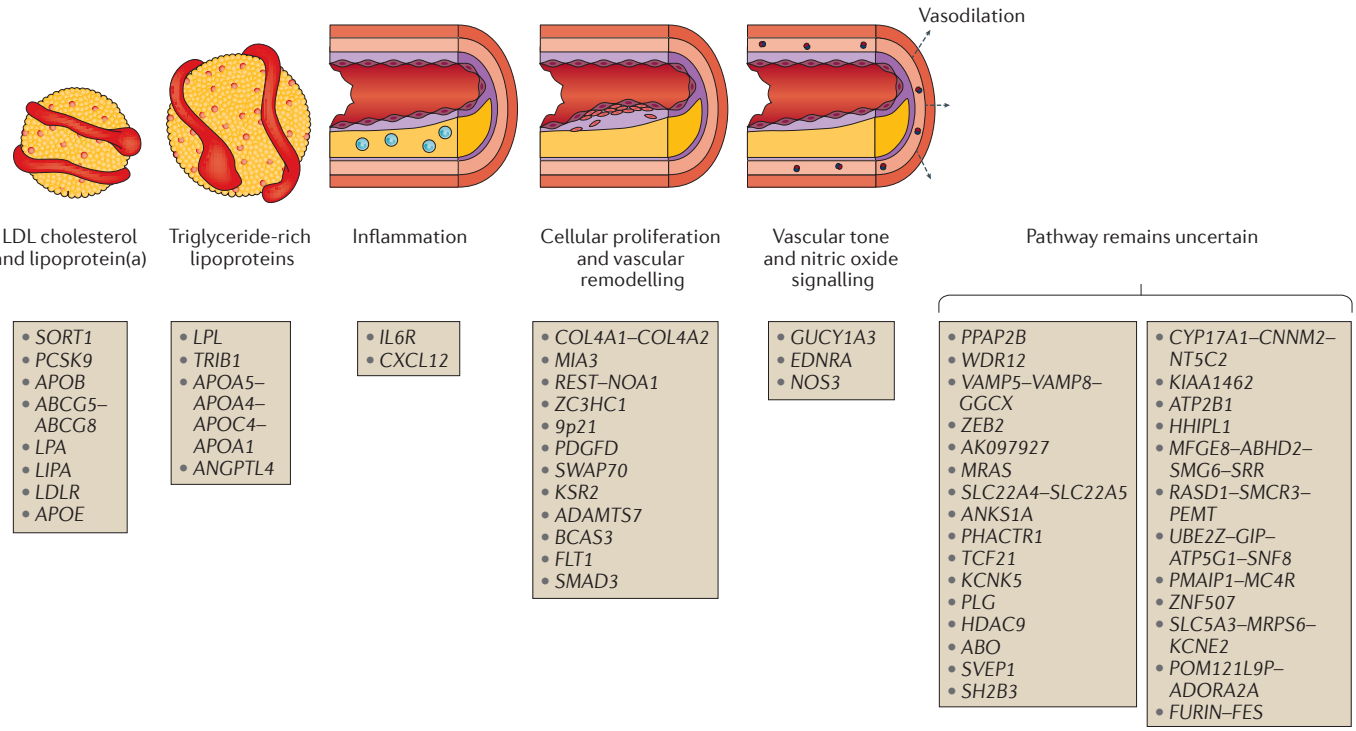


Figure 1 | **Physiological pathways related to genetic loci associated with coronary artery disease.** Genetic loci identified to date are displayed along with the presumed relationship to the causal pathway. Loci are labelled based on the nearest genes because the causal genes and variants have not been definitively identified for most loci; this commonly used form of annotation may prove incorrect in some cases. Adapted from REF. 128.

ADAMTS7* — *a nonlipid cause of CAD identified by human genetics studies. A 2010 CVAS identified a variant located in an intronic region of a disintegrin and metalloproteinase with thrombospondin motifs 7 (*ADAMTS7*); this variant was associated with a 19% increase in risk of obstructive CAD (>50% narrowing of a major coronary vessel)[60]. This finding led to substantial interest in elucidating the underlying mechanism, particularly because the variant was not associated with blood lipid levels or other known CAD risk factors. ADAMTS7 belongs to a family of proteins involved in proteolysis (the cleavage of substrate proteins) and remodelling of blood vessel walls. The risk variant at *ADAMTS7* seems to be associated with increased gene expression and proteolytic activity, and the ability to promote vascular smooth muscle migration *in vitro*[61]. *Adamts7* knockout mice were noted to have significantly reduced atherosclerosis burden[62]. Furthermore, mice deficient in *Adamts7* exhibited decreased cellular proliferation and enhanced endothelial cell repair in response to vascular injury[62,63]. These results suggest that a pharmacological strategy of ADAMTS7 inhibition could prove useful in attenuating the cellular proliferation that plays a key part in CAD progression.

Catalysing drug development for CAD

The development of novel, efficacious and safe therapeutic strategies for CAD remains a major public health need. However, the cost of developing a new drug has continued to increase; a recent analysis estimated direct costs of US$1.4 billion per new compound approved[64]. Much of this cost relates to the vast majority (>95%) of compounds that fail in clinical development, largely owing to inadequate therapeutic efficacy or unanticipated toxicity[65]. Human genetics may provide a key opportunity to improve drug development efforts by confirming the physiological and causal relevance of a given target in humans and anticipating the full range of efficacy and safety consequences of pharmacological modulation[66] (FIG. 3).

Selection of therapeutic targets by Mendelian randomization. Human genetics can complement existing preclinical prioritization strategies by providing confidence that a given target is a root cause of CAD in humans. Mendelian randomization studies have emerged as a technique grounded in human genetics to help infer causality between a given biomarker and risk of CAD (FIG. 4). For example, the large number of common genetic variants linked to lipid levels (>150) provided an opportunity to test the causality of lipid biomarkers using Mendelian randomization[67]. As expected, variants affecting LDL cholesterol levels had robust associations with risk of CAD in a concordant fashion[68]. A similar finding was noted for *LPA* variants affecting the levels of lipoprotein(a)[69]. In combination with the sequencing-based analyses of genes related to the LPL pathway discussed above, Mendelian randomization studies have supported a causal relationship between TRLs and CAD, and have reinvigorated interest in

Mendelian randomization
A human genetics tool that leverages the random assortment of genetic variants at time of conception to assess causality of observed associations.

targeting triglyceride metabolism for therapeutic gain[70]. By contrast, no consistent relationship of variants affecting high-density lipoprotein (HDL) cholesterol levels and CAD was observed, raising the possibility that the well-established inverse association between HDL cholesterol levels and risk of CAD is not reflective of a causal relationship[68].

Early examples suggest that Mendelian randomization studies may predict the outcome of large clinical trials. For example, inactivating mutations in *NPC1L1* were associated with decreased LDL cholesterol levels and risk of CAD[71]. A subsequent report from a randomized controlled trial of the drug ezetimibe, which inhibits the protein product of *NPC1L1* to decrease cholesterol absorption, confirmed its efficacy in reducing cardiovascular events[72]. By contrast, despite promising observational epidemiology and animal model evidence[73,74], lipoprotein-associated phospholipase A_2 (Lp-PLA2; also known as platelet-activating factor acetylhydrolase) inhibitors proved ineffective in multiple trials involving more than 28,000 patients[75,76]. Genetic studies of both common and rare variants in the gene encoding Lp-PLA2 (*PLA2G7*) were shown to have no impact on the risk of CAD, raising the possibility that the null clinical trial results might have been foreseen[77,78].

Developing therapeutics to mimic protective variants. The examples of *NPC1L1* and *PLA2G7* were based on targets with drug development programmes already in place. However, genetic evidence indicating that individuals with inactivating mutations in *PCSK9* have decreased levels of circulating LDL cholesterol and reduced risk of CAD fostered intense interest in the development of PCSK9 inhibitors[50]. In 2015, a mere 12 years after the initial discovery of PCSK9 (REF. 15), two monoclonal antibodies that inhibit PCSK9 were approved by the US Food and Drug Administration. Each of the drugs led to a ~50% decrease in circulating LDL cholesterol in initial clinical trials and seemed to decrease risk of cardiovascular events; ongoing large clinical trials are seeking to confirm this preliminary signal for efficacy[79,80]. Similarly, antisense oligonucleotides designed to mimic the protective mutations noted in *APOC3* or *LPA* demonstrated a ~70% reduction in triglyceride levels and 80% reduction in circulating lipoprotein(a) levels, respectively, in early-phase studies[81,82]. Future studies will seek to demonstrate favourable effect on clinical outcomes.

Phenome-wide association studies. Population-based analyses of variants related to a putative drug target can anticipate the full range of phenotypic consequences that might be expected by pharmacological modulation. For example, a common variant in the gene encoding HMG-coenzyme A reductase (*HMGCR*), the target of statin therapy, was associated with decreased LDL cholesterol levels but an increased risk of type 2 diabetes[83]. This finding is well aligned with an adverse effect of statins in increasing the risk of diabetes, which was discovered only after decades of clinical trial experience[84]. Furthermore, an ongoing trial is seeking to determine whether anacetrapib, a cholesteryl ester transfer protein

Table 1 | **Summary of results from gene sequencing studies for CAD**

Gene	Carrier frequency	Intermediate phenotype	CAD risk	Therapy to mimic protective variants	Refs
Inactivating mutations confer increased risk					
LDLR	1 in 221 (0.5%)	↑LDL cholesterol	↑320%	Not applicable	48
LPL	1 in 249 (0.4%)	↑Triglyceride-rich lipoproteins	↑84%	Not applicable	52
APOA5	1 in 216 (0.5%)	↑Triglyceride-rich lipoproteins	↑120%	Not applicable	48
Inactivating mutations confer decreased risk					
PCSK9	1 in 50 (2%)*	↓LDL cholesterol	↓88%	Alirocumab, evolucumab (approved by the FDA and EMA)	50
NPC1L1	1 in 650 (0.2%)	↓LDL cholesterol	↓53%	Ezetimibe (approved by the FDA and EMA)	71
ASGR1	1 in 120 (0.8%)	↓LDL cholesterol ↓Triglyceride-rich lipoproteins	↓34%	None	51
APOC3	1 in 150 (0.7%)	↓Triglyceride-rich lipoproteins	↓40%	Volanesorsen (formerly known as ISIS-APOCIII$_{Rx}$; phase III trials)	53,54
ANGPTL4	1 in 360 (0.3%)	↓Triglyceride-rich lipoproteins	↓53%	REGN1001 (preclinical development)	43,55
LPA	1 in 285 (0.4%)	↓Lipoprotein(a)	↓24%	AKCEA-APO(a)-L$_{Rx}$ (phase II trials)	90

ANGPTL4, angiopoietin like 4; *APO*, apolipoprotein; *ASGR1*, asialoglycoprotein receptor; CAD, coronary artery disease; EMA, European Medicines Agency; FDA, US Food and Drug Administration; *LDLR*, low-density lipoprotein receptor; *LPA*, lipoprotein(a); *LPL*, lipoprotein lipase; *NPC1L1*, Niemann–Pick C1-like intracellular cholesterol transporter 1; *PCSK9*, proprotein convertase subtilisin/kexin type 9. Damaging mutations in at least nine genes have been robustly associated with risk of coronary artery disease; in each case, identified genes disrupt pathways related to low-density lipoprotein (LDL) cholesterol, triglyceride-rich lipoproteins or lipoprotein(a) metabolism. Pharmacological therapies are in current use or development to mimic the protective variants for five of the six genes in which inhibition of the related protein would be predicted to reduce risk. *Prevalence estimate based on individuals of African ancestry. Carrier frequency substantially lower in other racial and ethnic groups.

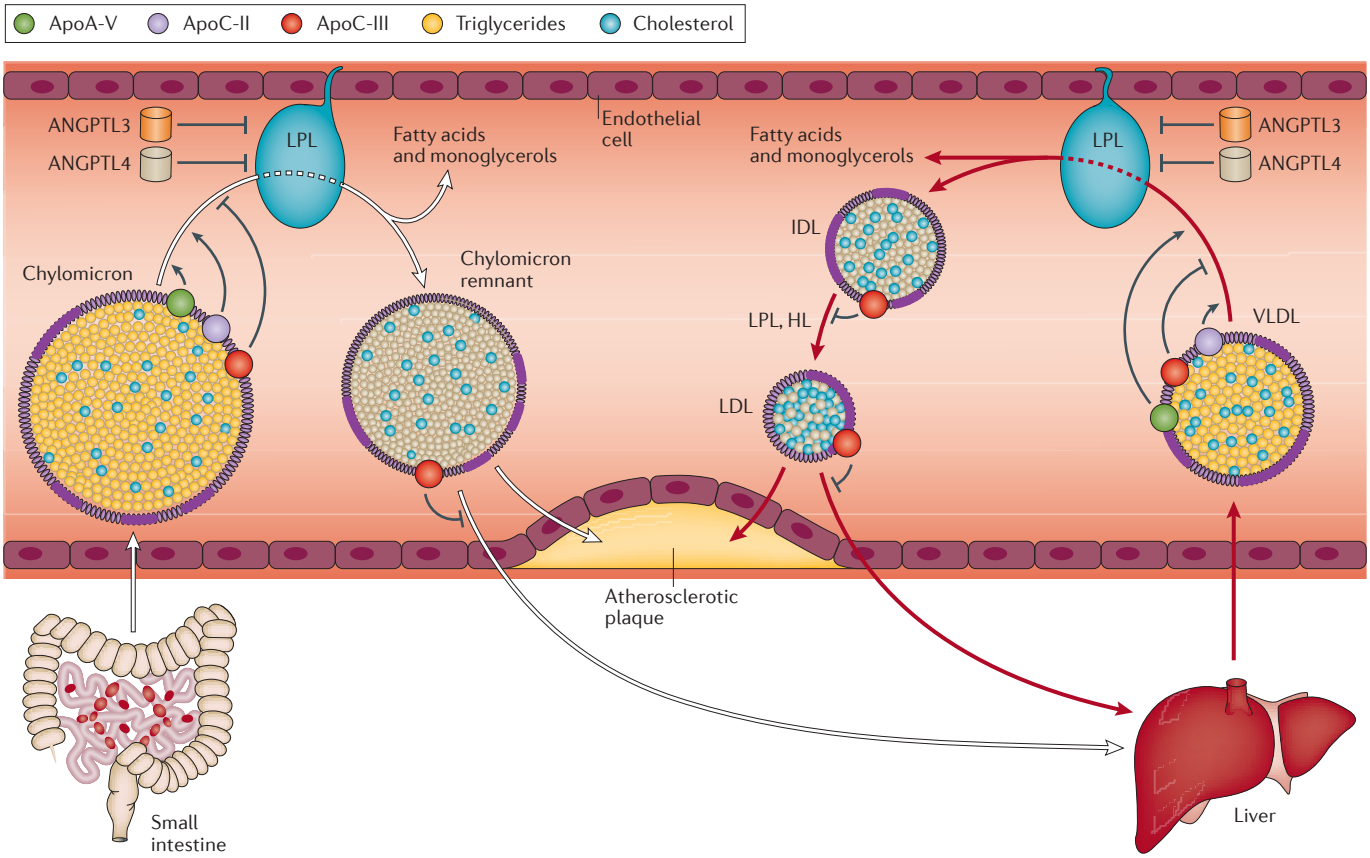


Figure 2 | **The LPL pathway.** Lipoprotein lipase (LPL) is an enzyme anchored to the endothelial cells lining blood capillaries. Dietary fat is absorbed by the small intestine and enters the bloodstream as triglyceride-rich lipoprotein particles known as chylomicrons. These chylomicrons are hydrolysed by LPL to provide free fatty acids (used for energy by muscle tissue or deposited into fat stores) and chylomicron remnant particles. LPL has an additional role in hydrolysing very-low-density lipoprotein (VLDL) particles secreted by the liver to produce intermediate-density lipoprotein (IDL), subsequently degraded into low-density lipoprotein (LDL) particles by hepatic lipase (HL). Both chylomicron remnants and LDL can penetrate the vessel wall and propagate atherosclerotic plaque. LPL activity is the rate-determining step in clearance of dietary fat from the circulation and is highly regulated in the body: apolipoprotein A-V (ApoA-V) and apolipoprotein C-II (ApoC-II) activate LPL, whereas apolipoprotein C-III (ApoC-III), angiopoietin-like 4 (ANGPTL4) and ANGPTL3 each inhibit LPL activity. Rare variant association studies have supported a link between several of the proteins involved in the LPL pathway and coronary artery disease (CAD), including LPL itself, ApoC-III, ApoA-V and ANGPTL4 (TABLE 1).

(CETP) inhibitor known to increase HDL cholesterol, can decrease the risk of cardiovascular disease[85]. Genetic variants within *CETP* that are associated with increased HDL cholesterol were recently linked to an increased risk of age-related macular degeneration, a leading cause of blindness[86–88]. Had this genetic finding been discovered earlier, trial sponsors or regulatory agencies may have requested inclusion of a surveillance programme to determine whether pharmacological inhibition would similarly lead to the toxicity predicted by human genetics.

These examples focus on relationships between genetic variants and candidate phenotypes selected *a priori* for investigation. However, this concept can be generalized using phenome-wide association studies (PheWAS), which enable the unbiased interrogation of relationships between a given variant and a broad range of human disease states[89]. The utility of this approach is likely to increase in coming years, as large biobanks link genotype information with electronic health records. For example, our group recently leveraged data from >100,000 participants of the UK Biobank to assess the phenotypic consequences of genetically lowered lipoprotein(a) levels on a range of 36 different disease states[90]. Beyond confirming an expected association with decreased risk of CAD and calcific aortic valve stenosis, this PheWAS uncovered relationships with reduced risk of stroke, peripheral vascular disease, congestive heart failure and chronic kidney disease. By contrast, genetic analyses did not replicate toxicity concerns based on observational epidemiology that linked reduced lipoprotein(a) levels to an increased risk of type 2 diabetes or cancer[91,92].

Recall by genotype studies. Deep phenotyping of individuals with a genetic variant related to a drug target provides an opportunity to test specific hypotheses in a rigorous fashion. Human genetics has traditionally

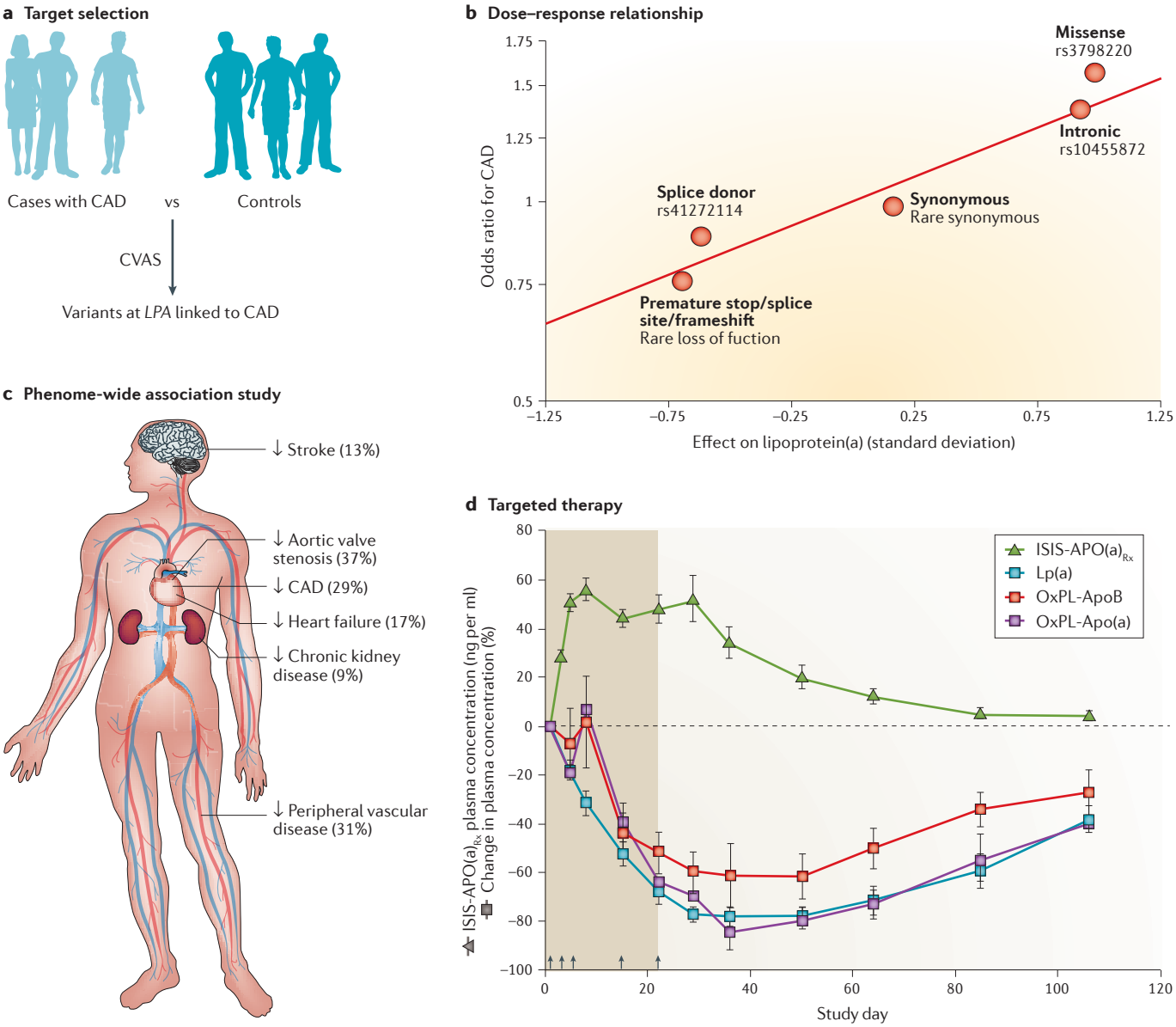
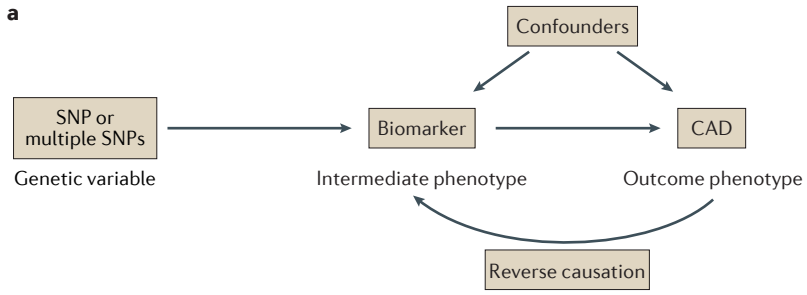
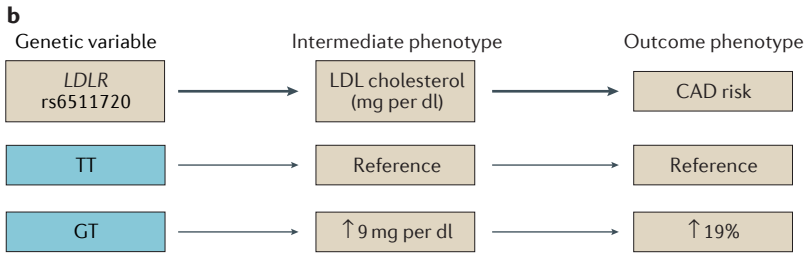


Figure 3 | **Human genetics to facilitate drug development: lipoprotein(a).** Human genetics data serves as the foundation for ongoing efforts to develop therapies to reduce lipoprotein(a) (Lp(a); encoded by *LPA*) levels, a causal risk factor for coronary artery disease (CAD). **a** | The selection of Lp(a) as a therapeutic target was supported by a 2009 common variant association study (CVAS) comparing 1,145 CAD cases to 3,352 controls, noting a robust association between variants near *LPA*, levels of circulating Lp(a) and risk of CAD[69]. **b** | A dose–response relationship was noted, such that the impact of a given variant on circulating Lp(a) levels was predictive of the association with CAD. **c** | To anticipate the full spectrum of phenotypic consequences of Lp(a) reduction, a phenome-wide association study was performed in participants of the UK Biobank. A genetically mediated one standard deviation decrease in levels of Lp(a) was associated with a reduced risk of six distinct diseases[90]. **d** | An antisense oligonucleotide (ISIS-APO(a)-L_{Rx}; now known as AKCEA-APO(a)-L_{Rx}) targeting hepatic production of Lp(a) was associated with a >80% decrease in circulating levels, providing proof of principle that this causal pathway may be targeted in a highly specific fashion. Similar reductions in OxPL-ApoB (oxidized phospholipids on apolipoprotein B) and OxPL-Apo(a) were noted. Part **b** is reproduced with permission from REF. 90. This image was published in *J. Am. Coll. Cardiol.*, **68**, Emdin, C. A. *et al.*, Phenotypic characterization of genetically lowered human lipoprotein(a) levels, 2761–2772, © Elsevier (2016). Part **d** is reprinted from *The Lancet*, **386**, Tsimikas, S. *et al.*, Antisense therapy targeting apolipoprotein(a): a randomised, double-blind, placebo-controlled phase 1 study, 1472–1483, © (2015), with permission from Elsevier.

utilized a 'phenotype first' paradigm, in which participants or families were ascertained on the basis of a phenotype of interest (for example, early-onset CAD) and the genetic underpinnings determined. The availability of genetic information in large populations enables a complementary 'genotype first' approach, in which individuals with a genotype of interest (for example, a large-effect inactivating mutation in a CAD-related gene) are called back for additional hypothesis-driven phenotyping.





c

Mendelian randomization supports causal relationship	Mendelian randomization does not support causal relationship
• LDL cholesterol	• C-reactive protein
• Lipoprotein(a)	• HDL cholesterol
• Triglyceride-rich lipoproteins	• Lipoprotein-associated and secretory phospholipase A_2
• Interleukin-6 signalling	• Urate
• Systolic blood pressure	• Cystatin C
• Body mass index	• Fibrinogen

Figure 4 | **Mendelian randomization to identify causal biomarkers for risk of CAD.** **a** | Mendelian randomization analyses require the following parameters: a genetic variant (or group of variants) is robustly associated with the biomarker of interest; the genetic variant is independent of confounders that influence the biomarker or risk of coronary artery disease (CAD); and any impact of the genetic variant on risk of CAD is mediated by the biomarker (as opposed to other pleiotropic genetic effects). Because genetic variants are assorted within the population at the time of conception and largely at random, these analyses are less susceptible to the issues of confounding or reverse causality that commonly limit causal inference from observational epidemiology. An important limitation of such a study is that it requires a robust relationship between the genetic variant and a biomarker: a small effect size mandates a large number of individuals with CAD to achieve adequate power. **b** | For example, rs6511720 is an intronic variant in the low-density lipoprotein (LDL) receptor (*LDLR*) gene. Each copy of the T allele is associated with lower LDL cholesterol levels and a decreased risk of CAD[68]. **c** | Among biomarkers linked with CAD in the population, studies performed to date have supported a causal relationship for some (for example, LDL cholesterol, triglyceride-rich lipoproteins and lipoprotein(a)) and a non-causal relationship for others (for example, high-density lipoprotein (HDL) cholesterol and C-reactive protein).

For *ANGPTL4*, a CAD target validated through human genetics, the identification of a potential toxicity in animal models has raised concern about whether it is an appropriate pharmacological target. A monoclonal antibody-based approach designed to inhibit ANGPTL4, thus mimicking the variants protective against CAD, led to a substantial decrease in TRLs in both mice and non-human primates[55]. However, these compounds also led to abnormal lipid accumulation in abdominal lymph nodes in these model systems, a finding previously observed in both a murine model of ANGPTL4 inhibition and in *Angptl4* knockout mice[93,94]. Phenotyping humans with a lifelong genetic deficiency in ANGPTL4 would provide one way to assess whether this toxicity also has relevance in humans. For example, magnetic resonance imaging can be used to characterize the size and tissue characteristics of abdominal lymph nodes. A comparison of such features between wild-type controls and those homozygous for an inactivating mutation in *ANGPTL4* may prove informative.

Genomic medicine

The substantial progress gained in understanding the genetic underpinnings of CAD in the population has laid the groundwork for the integration of genetic data into routine clinical practice; that is, 'genomic medicine'. If health care providers had widespread access to their patients' genetic data, which may well be the case in coming years, what are the potential implications for the prevention and treatment of CAD?

A longstanding debate in the field of complex traits genetics pits the 'common disease–common variant' and 'common disease–rare variant' hypotheses against one another[95]. In the latter, the manifestation of CAD may be reflective of a different monogenic driver in each individual[96]. Rare but large-effect variation could be used to stratify CAD into multiple disease subtypes and treatment directed accordingly. However, although likely to be the case for rare monogenic conditions such as familial hypercholesterolaemia, studies to date have instead suggested that the majority of heritable risk for CAD risk is predicated on common variants. These variants, in combination with environmental factors, lead to a quantitative blend of multiple driving processes in each individual.

Here, we describe the potential utility in identifying individuals with either monogenic or polygenic increased risk for CAD (FIG. 5). Recent reviews have highlighted an additional potential use of predicting drug efficacy or toxicity based on human genetics, often termed pharmacogenetics[97,98].

Lessons learned from familial hypercholesterolaemia. In principle, familial hypercholesterolaemia, the primary monogenic driver of CAD identified to date, represents a major opportunity to use human genetics to improve cardiovascular health. For example, a national programme initiated in 1994 in the Netherlands involves the genetic testing of individuals with features suggestive of familial hypercholesterolaemia, including family history, physical examination findings and severely elevated LDL cholesterol levels. To date, this programme has identified and provided treatment to thousands of individuals[99]. In practice, the implementation of programmes to systematically identify and treat individuals with familial hypercholesterolaemia at the population scale has proven difficult. Although familial hypercholesterolaemia remains both underdiagnosed and undertreated, efforts at both the national and institutional level to improve recognition are ongoing[100].

Largely owing to cost and logistical concerns, few previous efforts have sought to combine gene-sequencing data and observed LDL cholesterol levels

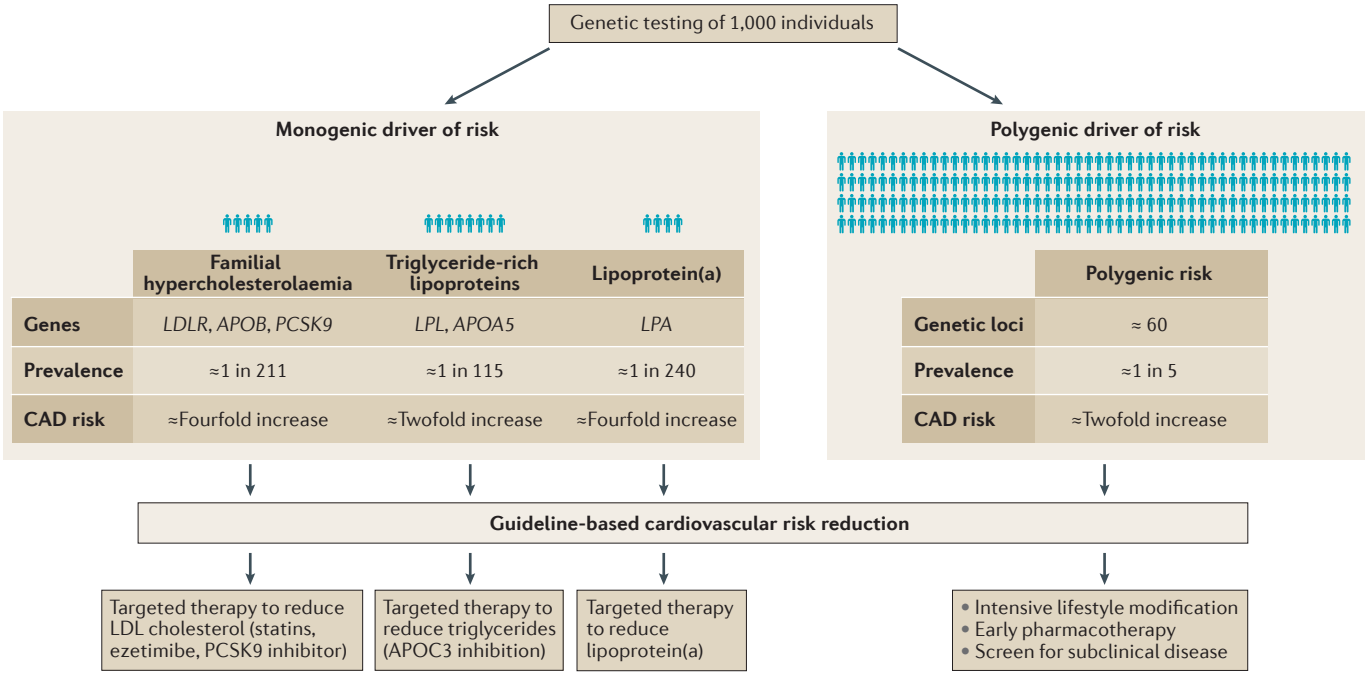


Figure 5 | **Precision medicine for CAD guided by human genetics.** Genetic testing of 1,000 people would identify 1–2% of individuals with a monogenic driver of risk. For example, individuals with familial hypercholesterolaemia[101] may derive particular benefit from lowering of low-density lipoprotein (LDL) cholesterol levels with proprotein convertase subtilisin/kexin type 9 (PCSK9) inhibitors. Second, damaging mutations in genes encoding lipoprotein lipase (*LPL*)[52] or apolipoprotein A5 (*APOA5*)[48] increase coronary artery disease (CAD) risk by preventing the clearance of dietary fat and triglyceride-rich lipoproteins from the bloodstream; this risk might be directly attenuated through inhibition of APOC3. Third, those with a genetic predisposition to increased lipoprotein(a) (encoded by *LPA*), that is, at least two risk variants (at SNP sites rs10455872 or rs3798220) as previously characterized[69], might be offered an antisense oligonucleotide to decrease circulating levels. Importantly, any such therapies would likely be incorporated in addition to guideline-based cardiovascular risk reduction therapeutics. Furthermore, confirmation that targeted PCSK9 inhibition, APOC3 inhibition, or lipoprotein(a) reduction leads to reduction in CAD events in clinical trials is needed before widespread implementation. By contrast, 20% of individuals tested would be in the top quintile of a polygenic risk score, and this population might be targeted for aggressive lifestyle or pharmacological therapies beyond current treatment guidelines. *LDLR*, LDL receptor.

in the prediction of CAD. We recently found that for any given observed LDL cholesterol, the risk of CAD is substantially higher among those who harbour a familial hypercholesterolaemia mutation compared with those who do not[101]. A direct comparison of serial LDL cholesterol levels between individuals with versus without a familial hypercholesterolaemia mutation confirmed higher cumulative exposure among those with a mutation. This observation supports the hypothesis that an increased lifelong exposure to high LDL cholesterol levels is a key driver of CAD risk and provides evidence for the potential clinical utility of knowing the mutation status of an individual.

The study of familial hypercholesterolaemia has also highlighted concepts of broad relevance to CAD population genetics; that is, penetrance, phenocopy and variant annotation. Incomplete penetrance refers to the finding that not all individuals with a familial hypercholesterolaemia mutation manifest severely elevated cholesterol levels. For example, despite a significant increase in average LDL cholesterol among carriers of a familial hypercholesterolaemia mutation, 27% of these individuals had a normal concentration, reflecting the significant heterogeneity both within and across specific mutations[101]. Second, the majority of cases of severely elevated LDL cholesterol cannot be explained by familial hypercholesterolaemia; a phenocopy of severe hypercholesterolaemia can also occur due to lifestyle or polygenic causes[102,103]. Among individuals with severe hypercholesterolaemia (defined as LDL cholesterol >190 mg per dl), only 2% harboured a familial hypercholesterolaemia mutation[101]. Third, assessing the functional impact of rare missense variants (variant annotation) on the function of a protein is not always straightforward, particularly if a variant has not been seen before. Efforts to centralize annotation based on previous observations and to refine computer prediction algorithms as well as high-throughput functional screening of identified variants are likely to improve efficiency and consistency of this process in coming years.

Human genetics to guide precision medicine therapeutics. For a subset of the genes with validated associations with CAD, pharmacological therapies are available or in development to specifically target a pathway and provide an additional opportunity for risk reduction beyond standard lifestyle and pharmacotherapy recommendations (FIG. 5). Importantly, although these precision

medicine therapeutics are likely to be used initially in populations with abnormalities in the pathway, they may well also prove useful for the population at large. For example, the efficacy and safety of statin therapy was initially demonstrated in 1994 in a clinical trial of participants with markedly elevated cholesterol levels and manifest CAD[104]. Numerous trials in ensuing years have demonstrated clinical benefit across a broad range of lower baseline cholesterol and CAD risk, albeit with lower absolute risk reductions[105].

However, only a small proportion of individuals harbour a rare, large-effect mutation in one of the genes discussed above. A more common aetiology of an increased genetic risk is related to a polygenic cause; specifically, the inheritance of a large number of smaller impact, more common risk alleles. For any given individual, a weighted score of risk variants can be calculated to provide a continuous and quantitative measure of genetic risk[106]. This approach was first implemented for CAD in 2008 using DNA sequence variants associated with circulating cholesterol levels and extended in 2009 to nine validated CAD risk variants[34,107]. Compared with those in the lowest quintile of this polygenic risk score, those in the top quintile were at a more than twofold increased risk of early-onset CAD[34]. This finding has subsequently been confirmed by multiple groups using progressively more variants[108–111].

As DNA is stable over the course of the lifetime, genetic risk can be ascertained from birth. Polygenic risk scores may therefore be particularly useful in risk prediction among younger patients (for example, children or young adults) in whom the cumulative impact of lifestyle factors is less pronounced. This early knowledge may facilitate an intervention to attenuate this risk. For example, adherence to a healthy lifestyle may be particularly important in these high-risk individuals. A recent analysis of individuals at high genetic risk noted a 46% attenuation of risk of incident CAD in those with a favourable versus unfavourable lifestyle[4]. Similarly, among those with a high genetic risk, statin therapy was associated with a nearly 50% reduction in CAD risk[111].

These examples provide key evidence that rather than operating in a deterministic fashion, high genetic risk is indeed modifiable. However, both lifestyle and statin therapy are safe and effective across a broad range of individual-level risk. The approach of targeting therapy according to genetic risk might be of more utility in the allocation of therapies with significant adverse effects (for example, antiplatelet therapies associated with bleeding) or cost (for example, antibody-based therapies).

A key goal of genomic medicine is to disclose genetic risk to individuals and their health care providers to enable behavioural changes or pharmacological therapy to attenuate risk before onset of disease. For example, the MI-GENES study demonstrated that the incorporation of a genetic risk score into shared decision-making sessions with patients and health care providers led to a modest increase in statin utilization and lower LDL cholesterol levels in those with high genetic risk[112]. Ongoing and future studies are needed to determine the optimal approach for genetic risk disclosure and assess whether this approach can improve clinical outcomes.

Future applications of human genetics findings

The study of the genetic architecture of human CAD has led to substantial progress in gene discovery, informing drug development through an improved understanding of human pathophysiology and laying the foundation for genomic medicine. Despite substantial progress, the following key questions regarding the genetics of CAD remain: can human genetics highlight a novel pathway, independent of LDL, TRL and lipoprotein(a) metabolism, that underlies CAD risk? Can genetics facilitate an understanding of the phenotypic consequences of each gene as they relate to CAD? Can the information gained about protective mutations be used to develop a therapy that would effectively cure atherosclerosis in this century?

Functional genomics to highlight novel causal pathways. The CVAS approach has been highly successful in validating a large number of genetic loci linked to CAD. However, for the majority of loci, key questions regarding the mechanisms underlying these associations remain unclear. In particular, the causal DNA variant, the causal gene under regulation, the mechanism by which the variant affects the gene and the mechanism by which the gene influences the risk of CAD. This line of work has proven challenging because the majority of variants lie in non-coding intergenic regions[39]; thus, even the initial step of identifying the gene of interest (that is, the effector transcript) is often not straightforward.

Successful approaches in functional genomics are likely to require close partnerships between computational and experimental biologists. Deep sequencing of the areas near CVAS loci and Bayesian approaches to prioritize variants can help to identify the causal variant[58,113]. Insights into the mechanism of gene regulation can be gained from publicly available databases, including the Encyclopedia of DNA Elements (ENCODE) project, which provides details on histone marks and transcription factor binding sites in the region[114]. The relationship of a given variant with tissue-specific expression of candidate genes (expression quantitative trait loci) can be evaluated using data from the Genotype-Tissue Expression (GTEx) project, thus facilitating integrative network analyses[115]. Epigenetic, metabolomic or proteomic data have become increasingly available and may similarly prove useful. From a functional perspective, more than 100 distinct genes have been shown to influence atherosclerosis in validated mouse models[116]. Collaborative efforts to demonstrate the relevance of such signals using human genetics analyses and, conversely, to use human genetic signals to prioritize candidate genes for animal model validation, represent a key opportunity to move the field forwards. Importantly, viral vectors that facilitate gene overexpression or knockdown in animal models are readily available. Substantial progress in gene editing approaches has dramatically decreased the time needed to engineer mouse models with a variant of interest. Similarly, induced pluripotent stem cells can be isolated from humans with a genotype of interest or engineered into existing cell lines for additional study.

Table 2 | **Current and emerging biobanks to facilitate genome–phenome interrogation**

Biobank	Web site	Enrolment locations	Initial enrolment	Enrolment to date	Target enrolment
Commercial funding					
deCODE Genetics (Amgen)	http://www.decode.com/	Iceland	1996	>200,000	Unknown
Geisinger MyCode Community Health (Regeneron Pharmaceuticals and others)	http://www.geisinger.org/for-researchers/partnering-with-patients/pages/mycode-health-initiative.html	Geisinger Health System (Danville, PA, USA)	2007	>50,000	Unknown
Government funding					
China Kadoorie Biobank	http://www.ckbiobank.org/site	China	2004	>500,000	Enrolment completed
UK Biobank	https://www.ukbiobank.ac.uk	UK	2006	>500,000	Enrolment completed
Electronic Medical Records and Genomics (eMERGE) Network	https://emerge.mc.vanderbilt.edu/about-emerge	United States Hospital Sites	2007	>50,000	Unknown
Million Veterans Program	http://www.research.va.gov/mvp	Veterans Affairs Hospitals, USA	2011	>500,000	~1,000,000
All of Us Research Program (part of the Precision Medicine Initiative)	https://www.nih.gov/research-training/allofus-research-program	USA	Early 2017	–	~1,000,000
Institutional funding					
BioVu Biorepository	https://victr.vanderbilt.edu/pub/biovu	Vanderbilt University Medical Center (Nashville, TN, USA)	2007	>215,000	Unknown
Kaiser Permanente Research Bank	http://researchbank.kaiserpermanente.org	USA	2016	>250,000	~500,000
Partners Healthcare Biobank	https://biobank.partners.org	Partners Health Care (Boston, MA, USA)	2010	>50,000	~100,000

Biobanks that link genetic and phenotypic information in a large number of individuals will provide a key resource for additional research in coming years. These biobanks will be most useful when linked with infrastructure to support hypothesis-based deep phenotyping among those with a genotype of interest. A representative example of such biobanks is provided.

This rapidly expanding functional genomics toolkit is likely to accelerate the process of moving from genomic localization to mechanistic discovery in coming years with the ultimate aim of identifying a novel pathway that can be therapeutically targeted[117].

Genetics at the population scale to understand genome–phenome relationships. A robust understanding of the phenotypic consequences of inactivating mutations in a given gene will require very large matrices that link an individual's genetic information with phenotypes (for example, phenotypes from electronic health records); several such resources are currently in use or in development (TABLE 2). Recent gene sequencing efforts have identified individuals homozygous for rare inactivating mutations (human knockouts) in >1,000 genes; this phenomenon seems to occur with substantially greater frequency in populations with higher rates of consanguinity[118,119]. Just as the targeted deletion of each gene in the mouse (Knockout Mouse Project) has been an invaluable resource to understand gene function, one might imagine a Human Knockout Project involving a systematic effort to phenotype humans who naturally lack a given gene.

As proof of concept, a recent, preliminary exome sequencing analysis identified an individual homozygous for a damaging mutation in *APOC3*, in whom levels of plasma APOC3 are virtually undetectable[119]. A subsequent call-back study of 27 additional family members confirmed the overall health status of human *APOC3* knockouts and a dramatic reduction in TRLs, particularly pronounced after a meal rich in fat. The identification of healthy individuals with an inactivating mutation in both copies of a given gene, effectively 'human knockouts,' can provide some reassurance that pharmacological inhibition of the gene's product would be tolerated in humans. Additional mechanistic and lipoprotein kinetics studies in these individuals are likely to have direct relevance to APOC3 inhibition as a therapeutic strategy for CAD.

Genome editing as curative therapy. Despite the demonstrated efficacy of currently available therapies, adherence and adverse effects limit their impact in clinical practice. For example, only 39% of individuals reported adherence to statin therapy in the year following a heart attack, even when the medicine was provided free of charge[120]. As noted above, rare genetic mutations in several genes have been noted to confer lifelong resistance to the development of CAD without detectable toxicity. A gene editing-based therapeutic that introduced such mutations via a one-time injection could extend these protective effects into the population at large.

Consanguinity
Production of offspring by related individuals (for example, second cousins or closer).

This approach would leverage one of the most significant scientific advances in recent decades, namely use of a CRISPR RNA-guided endonuclease system to cleave sequences of the human genome in a highly specific fashion[121,122]. A single injection of a viral vector designed to target hepatic PCSK9 using CRISPR–Cas9 led to mutations in ~50% of hepatocytes and a 40% reduction in cholesterol levels in a mouse model[123]. Similar therapeutics could be designed to decrease circulating TRLs, lipoprotein(a) or other causal risk factors for CAD identified in coming years.

Given that this gene editing approach would lead to irreversible change, an abundance of caution is needed before clinical implementation. Ongoing research is seeking to attenuate the immune reaction induced by the viral vectors used to introduce the therapeutic, to increase efficiency and specificity in targeting the desired gene, and to ensure that no other genes are affected (so-called off-target effects). With regard to target selection, careful phenotyping of 'human knockouts' may help ensure that there are no unanticipated side effects of gene inactivation. Gene editing to decrease the expression or function of a gene product is more tractable with current technology. However, for genes in which damaging mutations confer increased risk of CAD (for example, *LDLR*), future advances may enable upregulation or potentiation of gene activity. Furthermore, alteration of an individual's DNA raises a host of ethical and social questions that will need to be fully explored[124]. However, at present, the majority of individuals in the United States approve of gene editing with the express purpose of improving health or preventing disease[125].

Conclusions

Here, we reviewed the substantial progress made in understanding the genetic underpinnings of CAD. Moving forwards, the price of genetic sequencing will continue to decrease and an increased emphasis on variant interpretation, functional validation and integration with large-scale phenotyping efforts will become paramount. Over the next 10 years, we are hopeful that human genetics will prove useful in identifying novel root causes of CAD, guiding drug development efforts in anticipating the safety and efficacy profile of a given therapeutic, and providing patients and their health care providers with genetic data that will aid in CAD prevention and treatment.

1. Ford, E. S. *et al.* Explaining the decrease in U.S. deaths from coronary disease, 1980–2000. *N. Engl. J. Med.* **356**, 2388–2398 (2007).
2. Lozano, R. *et al.* Global and regional mortality from 235 causes of death for 20 age groups in 1990 and 2010: a systematic analysis for the Global Burden of Disease Study 2010. *Lancet* **380**, 2095–2128 (2012).
3. Mozaffarian, D. *et al.* Heart disease and stroke statistics — 2016 update: a report from the American Heart Association. *Circulation* **133**, e38–e360 (2016).
4. Khera, A. V. *et al.* Genetic risk, adherence to a healthy lifestyle, and risk of coronary disease. *N. Engl. J. Med.* **375**, 2349–2358 (2016).
5. Gertler, M. M., Garn, S. M. & White, P. D. Young candidates for coronary heart disease. *JAMA* **147**, 621–625 (1951).
6. Marenberg, M. E., Risch, N., Berkman, L. F., Floderus, B. & de Faire, U. Genetic susceptibility to death from coronary heart disease in a study of twins. *N. Engl. J. Med.* **330**, 1041–1046 (1994).
The first large-scale prospective study of twins to confirm an increased risk of early-onset CAD among highly related individuals.
7. Zdravkovic, S. *et al.* Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J. Intern. Med.* **252**, 247–254 (2002).
8. Won, H. H. *et al.* Disproportionate contributions of select genomic compartments and cell types to genetic risk for coronary artery disease. *PLoS Genet.* **11**, e1005622 (2015).
9. Lloyd-Jones, D. M. *et al.* Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA* **291**, 2204–2211 (2004).
10. Murabito, J. M. *et al.* Sibling cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults. *JAMA* **294**, 3117–3123 (2005).
11. Watkins, H. & Farrall, M. Genetic susceptibility to coronary artery disease: from promise to progress. *Nat. Rev. Genet.* **7**, 163–173 (2006).
12. Müller, C. Xanthomata, hypercholesterolemia, angina pectoris. *J. Intern. Med.* **89**, 75–84 (1938).
13. Lehrman, M. A. *et al.* Mutation in LDL receptor: Alu-Alu recombination deletes exons encoding transmembrane and cytoplasmic domains. *Science* **227**, 140–146 (1985).
The first family-based study to identify a discrete mutation in a single gene predisposing to CAD (familial hypercholesterolaemia).
14. Soria, L. F. *et al.* Association between a specific apolipoprotein B mutation and familial defective apolipoprotein B-100. *Proc. Natl Acad. Sci. USA* **86**, 587–591 (1989).
15. Abifadel, M. *et al.* Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat. Genet.* **34**, 154–156 (2003).
16. Garcia, C. K. *et al.* Autosomal recessive hypercholesterolemia caused by mutations in a putative LDL receptor adaptor protein. *Science* **292**, 1394–1398 (2001).
17. Berge, K. E. *et al.* Accumulation of dietary cholesterol in sitosterolemia caused by mutations in adjacent ABC transporters. *Science* **290**, 1771–1775 (2000).
18. Wang, L., Fan, C., Topol, S. E., Topol, E. J. & Wang, Q. Mutation of *MEF2A* in an inherited disorder with features of coronary artery disease. *Science* **302**, 1578–1581 (2003).
19. Lieb, W. *et al.* Lack of association between the *MEF2A* gene and myocardial infarction. *Circulation* **117**, 185–191 (2008).
20. Mani, A. *et al. LRP6* mutation in a family with early coronary disease and metabolic risk factors. *Science* **315**, 1278–1282 (2007).
21. Keramati, A. R. *et al.* A form of the metabolic syndrome associated with mutations in *DYRK1B*. *N. Engl. J. Med.* **370**, 1909–1919 (2014).
22. MacArthur, D. G. *et al.* Guidelines for investigating causality of sequence variants in human disease. *Nature* **508**, 469–476 (2014).
Provides a framework for systematically assessing a potentially causal relationship between a given genetic variant and risk of human disease.
23. Altshuler, D., Daly, M. J. & Lander, E. S. Genetic mapping in human disease. *Science* **322**, 881–888 (2008).
24. Zuk, O. *et al.* Searching for missing heritability: designing rare variant association studies. *Proc. Natl Acad. Sci. USA* **111**, E455–E464 (2014).
25. Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443–453 (2007).
26. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
27. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
28. Ye, S., Willeit, J., Kronenberg, F., Xu, Q. & Kiechl, S. Association of genetic variation on chromosome 9p21 with susceptibility and progression of atherosclerosis: a population-based, prospective study. *J. Am. Coll. Cardiol.* **52**, 378–384 (2008).
29. Helgadottir, A. *et al.* The same sequence variant on9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat. Genet.* **40**, 217–224 (2008).
30. Smith, J. G. *et al.* Common genetic variants on chromosome 9p21 confers risk of ischemic stroke: a large-scale genetic association study. *Circ. Cardiovasc. Genet.* **2**, 159–164 (2009).
31. Jarinova, O. *et al.* Functional analysis of the chromosome 9p21.3 coronary artery disease risk locus. *Arterioscler. Thromb. Vasc. Biol.* **29**, 1671–1677 (2009).
32. Holdt, L. M. *et al.* ANRIL expression is associated with atherosclerosis risk at chromosome 9p21. *Arterioscler. Thromb. Vasc. Biol.* **30**, 620–627 (2010).
33. Harismendy, O. *et al.* 9p21 DNA variants associated with coronary artery disease impair interferon-γ signalling response. *Nature* **470**, 264–268 (2011).
34. Myocardial Infarction Genetics Consortium. Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat. Genet.* **41**, 334–341 (2010).
35. Erdmann, J. *et al.* New susceptibility locus for coronary artery disease on chromosome 3q22.3. *Nat. Genet.* **41**, 280–282 (2009).
36. Coronary Artery Disease Genetics (C4D) Consortium. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat. Genet.* **43**, 339–344 (2011).
37. IBC 50K CAD Consortium. Large-scale gene-centric analysis identifies novel variants for coronary artery disease. *PLoS Genet.* **7**, e1002260 (2011).
38. CARDIoGRAMplusC4D Consortium *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat. Genet.* **45**, 25–33 (2013).
39. Nikpay, M. *et al.* A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130 (2015).
40. So, H. C., Gui, A. H., Cherny, S. S. & Sham, P. C. Evaluating the heritability explained by known susceptibility variants: a survey of ten complex diseases. *Genet. Epidemiol.* **35**, 310–317 (2011).
41. Flannick, J. & Florez, J. C. Type 2 diabetes: genetic data sharing to advance complex disease research. *Nat. Rev. Genet.* **17**, 535–559 (2016).

42. Schizophrenia Working Group of the Psychiatric Genomics Consortium. Biological insights from 108 schizophrenia-associated genetic loci. *Nature.* **511**, 421–427 (2014).
43. Myocardial Infarction Genetics and CARDIoGRAM Exome Consortia Investigators. Coding variation in ANGPTL4, LPL, and SVEP1 and the risk of coronary disease. *N. Engl. J. Med.* **374**, 1134–1144 (2016).
44. Maurano, M. T. *et al.* Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195 (2012).
45. Ehret, G. B. *et al.* The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals. *Nat. Genet.* **48**, 1171–1184 (2016).
46. Erdmann, J. *et al.* Dysfunctional nitric oxide signalling increases risk of myocardial infarction. *Nature* **504**, 432–436 (2013).
47. Lee, S., Abecasis, G. R., Boehnke, M. & Lin, X. Rare-variant association analysis: study designs and statistical tests. *Am. J. Hum. Genet.* **95**, 5–23 (2014).
48. Do, R. *et al.* Exome sequencing identifies rare LDLR and APOA5 alleles conferring risk for myocardial infarction. *Nature* **518**, 102–106 (2015).
The first large study to use whole-exome sequencing to examine the relationship of rare variants in each gene with CAD.
49. Cohen, J. *et al.* Low LDL cholesterol in individuals of African descent resulting from frequent nonsense mutations in *PCSK9*. *Nat. Genet.* **37**, 161–165 (2005).
50. Cohen, J. C., Boerwinkle, E., Mosley, T. H. Jr & Hobbs, H. H. Sequence variations in *PCSK9*, low LDL, and protection against coronary heart disease. *N. Engl. J. Med.* **354**, 1264–1272 (2006).
Links inactivating mutations in *PCSK9* with significantly reduced LDL cholesterol and risk of incident CAD.
51. Nioi, P. *et al.* Variant ASGR1 associated with a reduced risk of coronary artery disease. *N. Engl. J. Med.* **374**, 2131–2141 (2016).
52. Khera, A. V. *et al.* Association of rare and common variation in the lipoprotein lipase gene with coronary artery disease. *JAMA* http://dx.doi.org/10.1001/jama.2017.0972 (2017).
53. Jørgensen, A. B., Frikke-Schmidt, R., Nordestgaard, B. G. & Tybjærg-Hansen, A. Loss-of-function mutations in *APOC3* and risk of ischemic vascular disease. *N. Engl. J. Med.* **371**, 32–41 (2014).
54. Crosby, J. *et al.* Loss-of-function mutations in *APOC3*, triglycerides, and coronary disease. *N. Engl. J. Med.* **371**, 22–31 (2014).
55. Dewey, F. E. *et al.* Inactivating variants in ANGPTL4 and risk of coronary artery disease. *N. Engl. J. Med.* **374**, 1123–1133 (2016).
56. Kathiresan, S. *et al.* Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. *Nat. Genet.* **40**, 189–197 (2008).
57. Willer, C. J. *et al.* Newly identified loci that influence lipid concentrations and risk of coronary artery disease. *Nat. Genet.* **40**, 161–169 (2008).
58. Musunuru, K. *et al.* From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–719 (2010).
The first paper to characterize the mechanism linking a non-coding variant with changes in gene regulation related to LDL cholesterol metabolism.
59. Strong, A. *et al.* Hepatic sortilin regulates both apolipoprotein B secretion and LDL catabolism. *J. Clin. Invest.* **122**, 2807–2816 (2012).
60. Reilly, M. P. *et al.* Identification of *ADAMTS7* as a novel locus for coronary atherosclerosis and association of ABO with myocardial infarction in the presence of coronary atherosclerosis: two genome-wide association studies. *Lancet* **377**, 383–392 (2011).
61. Pu, X. *et al.* ADAMTS7 cleavage and vascular smooth muscle cell migration is affected by a coronary-artery-disease-associated variant. *Am. J. Hum. Genet.* **92**, 366–374 (2013).
62. Bauer, R. C. *et al.* Knockout of *Adamts7*, a novel coronary artery disease locus in humans, reduces atherosclerosis in mice. *Circulation* **131**, 1202–1213 (2015).
63. Kessler, T. *et al.* ADAMTS-7 inhibits re-endothelialization of injured arteries and promotes vascular remodeling through cleavage of thrombospondin-1. *Circulation* **131**, 1191–1201 (2015).
64. DiMasi, J. A., Grabowski, H. G. & Hansen, R. W. Innovation in the pharmaceutical industry: new estimates of R&D costs. *J. Health Econ.* **47**, 20–33 (2016).
65. Paul, S. M. *et al.* How to improve R&D productivity: the pharmaceutical industry's grand challenge. *Nat. Rev. Drug Discov.* **9**, 203–214 (2010).
66. Plenge, R. M., Scolnick, E. M. & Altshuler, D. Validating therapeutic targets through human genetics. *Nat. Rev. Drug Discov.* **12**, 581–594 (2013).
This Review article describes the potential utility of human genetics to expedite drug development.
67. Global Lipids Genetics Consortium. Discovery and refinement of loci associated with lipid levels. *Nat. Genet.* **45**, 1274–1283 (2013).
68. Voight, B. F. Plasma HDL cholesterol and risk of myocardial infarction: a Mendelian randomisation study. *Lancet* **380**, 572–580 (2012).
69. Clarke, R. *et al.* Genetic variants associated with Lp(a) lipoprotein level and coronary disease. *N. Engl. J. Med.* **361**, 2518–2528 (2009).
70. Do, R. *et al.* Common variants associated with plasma triglycerides and risk for coronary artery disease. *Nat. Genet.* **45**, 1345–1352 (2013).
71. Myocardial Infarction Genetics Consortium Investigators. Inactivating mutations in *NPC1L1* and protection from coronary heart disease. *N. Engl. J. Med.* **371**, 2072–2082 (2014).
72. Cannon, C. P. *et al.* Ezetimibe added to statin therapy after acute coronary syndromes. *N. Engl. J. Med.* **372**, 2387–2397 (2015).
73. Lp-PLA(2) Studies Collaboration *et al.* Lipoprotein-associated phospholipase A_2 and risk of coronary disease, stroke, and mortality: collaborative analysis of 32 prospective studies. *Lancet* **375**, 1536–1544 (2010).
74. Wilensky, R. L. *et al.* Inhibition of lipoprotein-associated phospholipase A_2 reduces complex coronary atherosclerotic plaque development. *Nat. Med.* **14**, 1059–1066 (2008).
75. The STABILITY Investigators. Darapladib for preventing ischemic events in stable coronary heart disease. *N. Engl. J. Med.* **370**, 1702–1711 (2014).
76. O'Donoghue, M. L. *et al.* Effect of darapladib on major coronary events after an acute coronary syndrome: the SOLID-TIMI 52 randomized clinical trial. *JAMA* **312**, 1006–1015 (2014).
77. Polfus, L. M. *et al.* Coronary heart disease and genetic variants with low phospholipase A2 activity. *N. Engl. J. Med.* **372**, 295–296 (2015).
78. Casas, J. P. *et al. PLA2G7* genotype, lipoprotein-associated phospholipase A_2 activity, and coronary heart disease risk in 10 494 cases and 15 624 controls of European Ancestry. *Circulation* **121**, 2284–2293 (2010).
79. Robinson, J. G. *et al.* Efficacy and safety of alirocumab in reducing lipids and cardiovascular events. *N. Engl. J. Med.* **372**, 1489–1499 (2015).
80. Sabatine, M. S. Efficacy and safety of evolocumab in reducing lipids and cardiovascular events. *N. Engl. J. Med.* **372**, 1500–1509 (2015).
81. Gaudet, D. *et al.* Antisense inhibition of apolipoprotein C-III in patients with hypertriglyceridemia. *N. Engl. J. Med.* **373**, 438–447 (2015).
82. Tsimikas, S. *et al.* Antisense therapy targeting apolipoprotein(a): a randomised, double-blind, placebo-controlled phase 1 study. *Lancet* **386**, 1472–1483 (2015).
83. Swerdlow, D. I. *et al.* HMG-coenzyme A reductase inhibition, type 2 diabetes, and bodyweight: evidence from genetic analysis and randomised trials. *Lancet* **385**, 351–361 (2015).
84. Sattar, N. *et al.* Statins and risk of incident diabetes: a collaborative meta-analysis of randomised statin trials. *Lancet* **375**, 735–742 (2010).
85. US National Library of Medicine. *ClinicalTrials.gov* https://clinicaltrials.gov/ct2/show/NCT01252953 (2016).
86. Neale, B. M. *et al.* Genome-wide association study of advanced age-related macular degeneration identifies a role of the hepatic lipase gene (*LIPC*). *Proc. Natl Acad. Sci. USA* **107**, 7395–7400 (2010).
87. Chen, W. *et al.* Genetic variants near TIMP3 and high-density lipoprotein-associated loci influence susceptibility to age-related macular degeneration. *Proc. Natl Acad. Sci. USA* **107**, 7401–7406 (2010).
88. Cheng, C. Y. *et al.* New loci and coding variants confer risk for age-related macular degeneration in East Asians. *Nat. Commun.* **6**, 6063 (2015).
89. Bush, W. S., Oetjens, M. T. & Crawford, D. C. Unravelling the human genome–phenome relationship using phenome-wide association studies. *Nat. Rev. Genet.* **17**, 129–145 (2016).
90. Emdin, C. A. *et al.* Phenotypic characterization of genetically lowered human lipoprotein(a) levels. *J. Am. Coll. Cardiol.* **68**, 2761–2772 (2016).
91. Sawabe, M. *et al.* Low lipoprotein(a) concentration is associated with cancer and all-cause deaths: a population-based cohort study (the JMS cohort study). *PLoS ONE.* **7**, e31954 (2012).
92. Mora, S. *et al.* Lipoprotein(a) and risk of type 2 diabetes. *Clin. Chem.* **56**, 1252–1260 (2010).
93. Lichtenstein, L. *et al.* Angptl4 protects against severe proinflammatory effects of saturated fat by inhibiting fatty acid uptake into mesenteric lymph node macrophages. *Cell Metab.* **12**, 580–592 (2010).
94. Desai, U. *et al.* Lipid-lowering effects of anti-angiopoietin-like 4 antibody recapitulate the lipid phenotype found in angiopoietin-like 4 knockout mice. *Proc. Natl Acad. Sci. USA* **104**, 11766–11771 (2007).
95. Gibson, G. Rare and common variants: twenty arguments. *Nat. Rev. Genet.* **13**, 135–145 (2012).
96. McClellan, J. & King, M. C. Genetic heterogeneity in human disease. *Cell* **141**, 210–217 (2010).
97. Nelson, M. R. *et al.* The genetics of drug efficacy: opportunities and challenges. *Nat. Rev. Genet.* **17**, 197–206 (2016).
98. Paynter, N. P., Ridker, P. M. & Chasman, D. I. Are genetic tests for atherosclerosis ready for routine clinical use? *Circ. Res.* **118**, 607–619 (2016).
99. Umans-Eckenhausen, M. A., Defesche, J. C., Sijbrands, E. J., Scheerder, R. L. & Kastelein, J. J. Review of first 5 years of screening for familial hypercholesterolaemia in the Netherlands. *Lancet* **357**, 165–168 (2001).
100. Nordestgaard, B. G. Familial hypercholesterolaemia is underdiagnosed and undertreated in the general population: guidance for clinicians to prevent coronary heart disease: consensus statement of the European Atherosclerosis Society. *Eur. Heart J.* **34**, 3478–3490 (2013).
101. Khera, A. V. *et al.* Diagnostic yield and clinical utility of sequencing familial hypercholesterolemia genes in patients with severe hypercholesterolemia. *J. Am. Coll. Cardiol.* **67**, 2578–2589 (2016).
102. Teslovich, T. M. *et al.* Biological, clinical and population relevance of 95 loci for blood lipids. *Nature* **466**, 707–713 (2010).
103. Talmud, P. J. *et al.* Use of low-density lipoprotein cholesterol gene score to distinguish patients with polygenic and monogenic familial hypercholesterolaemia: a case–control study. *Lancet* **381**, 1293–1301 (2013).
104. Scandinavian Simvastatin Survival Study Group. Randomized trial of cholesterol lowering in 4444 patients with coronary heart disease: the Scandinavian Simvastatin Survival Study (4S). *Lancet* **344**, 1383–1138 (1994).
105. Cholesterol Treatment Trialists' (CTT) Collaborators. The effects of lowering LDL cholesterol with statin therapy in people at low risk of vascular disease: meta-analysis of individual data from 27 randomised trials. *Lancet* **380**, 581–590 (2012).
106. Chatterjee, N., Shi, J. & García-Closas, M. Developing and evaluating polygenic risk prediction models for stratified disease prevention. *Nat. Rev. Genet.* **17**, 392–406 (2016).
107. Kathiresan, S. *et al.* Polymorphisms associated with cholesterol and risk of cardiovascular events. *N. Engl. J. Med.* **358**, 1240–1249 (2008).
108. Ripatti, S. *et al.* A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet* **376**, 1393–1400 (2010).
109. Tada, H. *et al.* Risk prediction by genetic risk scores for coronary heart disease is independent of self-reported family history. *Eur. Heart J.* **37**, 561–567 (2016).
110. Abraham, G. *et al.* Genomic prediction of coronary heart disease. *Eur Heart J.* **37**, 3267–3278 (2016).
111. Mega, J. L. *et al.* Genetic risk, coronary heart disease events, and the clinical benefit of statin therapy: an analysis of primary and secondary prevention trials. *Lancet* **385**, 2264–2271 (2015).
112. Kullo, I. J. *et al.* Incorporating a genetic risk score into coronary heart disease risk estimates: effect on low-density lipoprotein cholesterol levels (the MI-GENES clinical trial). *Circulation* **133**, 1181–1188 (2016).
113. Wellcome Trust Case Control Consortium *et al.* Bayesian refinement of association signals for 14 loci in 3 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
114. ENCODE Project Consortium. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* **306**, 636–640 (2004).

115. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat. Genet.* **45**, 580–585 (2013).
116. Stylianou, I. M., Bauer, R. C., Reilly, M. P. & Rader, D. J. Genetic basis of atherosclerosis: insights from mice and humans. *Circ. Res.* **110**, 337–355 (2012).
117. Nurnberg, S. T. *et al.* From loci to biology: functional genomics of genome-wide association for coronary disease. *Circ. Res.* **118**, 586–606 (2016).
118. Narasimhan, V. M. *et al.* Health and population effects of rare gene knockouts in adult humans with related parents. *Science* **352**, 474–477 (2016).
119. Saleheen, D. *et al.* Human knockouts in a cohort with a high rate of consanguinity. Preprint at *bioRxiv* http://dx.doi.org/10.1101/031518 (2015).
Identifies humans with inactivating mutations ('knockouts') in ~1,000 genes and genotype-based call back to understand relevant physiology.
120. Choudhry, N. K. *et al.* Full coverage for preventive medications after myocardial infarction. *N. Engl. J. Med.* **365**, 2088–2097 (2011).
121. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
122. Mali, P. *et al.* RNA-guided human genome engineering via Cas9. *Science* **339**, 823–826 (2013).
123. Ding, Q. *et al.* Permanent alteration of PCSK9 with *in vivo* CRISPR-Cas9 genome editing. *Circ. Res.* **115**, 488–492 (2014).
Proof-of-concept study in a mouse model that permanent disruption of *PCSK9* using gene editing can decrease LDL cholesterol and atherosclerosis.
124. Lander, E. S. Brave new genome. *N. Engl. J. Med.* **373**, 5–8 (2015).
125. Blendon, R. J., Gorski, M. T. & Benson, J. M. The public and the gene-editing revolution. *N. Engl. J. Med.* **374**, 1406–1411 (2016).
126. Libby, P. in *Braunwald's Heart Disease: a Textbook of Cardiovascular Medicine* 10th edn (eds Bonow, R. O., Mann, D. L., Zipes, D. P. & Libby, P.) 873–890 (Saunders, 2014).
127. Ridker, P. M., Libby, P. & Buring, J. E. in *Braunwald's Heart Disease: a Textbook of Cardiovascular Medicine* 10th edn (eds Bonow, R. O., Mann, D. L., Zipes, D. P. & Libby, P.) 891–933 (Saunders, 2014).
128. Kessler, T., Vilne, B. & Schunkert, H. The impact of genome-wide association studies on the pathophysiology and therapy of cardiovascular disease. *EMBO Mol. Med.* **8**, 688–701 (2016).

Acknowledgements
A.V.K. is supported by a John S. LaDue Memorial Fellowship in Cardiology and a KL2/Catalyst Medical Research Investigator Training award (an appointed KL2 award) from Harvard Catalyst. S.K. is supported by an Ofer and Shelley Nemirovsky MGH Research Scholar Award and by grants HL127564 and UM1HG008895 from the US National Institutes of Health.

Competing interests statement
The authors declare competing interests: see Web version for details.

FURTHER INFORMATION
ENCODE: http://www.encodeproject.org/
GTEx: http://www.gtexportal.org/home
ALL LINKS ARE ACTIVE IN THE ONLINE PDF