Abstract
Cardiovascular disease (CVD) is a leading cause of mortality and morbidity in the Western world. Large genome-wide association studies (GWASs) of coronary artery disease, myocardial infarction, stroke and dilated cardiomyopathy have identified a number of common genetic variants with modest effects on disease risk. Similarly, studies of important modifiable risk factors of CVD have identified a large number of predominantly common variant associations, for example, with blood pressure and blood lipid levels. In each case, despite the often large numbers of loci identified, only a small proportion of the phenotypic variance is explained. It has been hypothesised that rare variants with large effects may account for some of the missing variance but large-scale studies of rare variation are in their infancy for cardiovascular traits and have yet to produce fruitful results. Studies of monogenic CVDs, inherited disorders believed to be entirely driven by individual rare mutations, have highlighted genes that play a key role in disease aetiology. In this review, we discuss how findings from studies of rare variants in monogenic disease and GWAS of predominantly common variants are converging to provide further insight into biological disease mechanisms.

Keywords: cardiovascular; genetics; rare variants; monogenic disease; genome-wide association studies; complex traits

INTRODUCTION
Cardiovascular disease (CVD), which includes coronary artery disease (CAD; also known as coronary heart disease, ischemic heart disease and atherosclerotic heart disease), heart failure, stroke and hypertension, is a leading cause of mortality and morbidity in Western countries. Despite improved treatment of CAD complications such as myocardial infarction (MI), prevalence of CVD is projected to increase over the next 20 years [1] and with this comes an increasing economic burden. Alongside the non-modifiable risk factors of CAD (age and sex), modifiable risk factors include high blood pressure, high circulating plasma cholesterol, diabetes mellitus, obesity, inactivity and smoking.

Identification of genetic variants associated with cardiovascular health and disease can yield insights into the underlying biological pathways, improve prediction of risk of CVD and enable development of personalised approaches to treatment and therapy. CAD is the most common form of heart disease and twin studies suggest that genetic variation may account for up to 60% of the variation in risk of CAD [2]. Genome-wide association studies (GWAS) have so far identified 47 loci with significant evidence of association with risk of CAD [3–5]. These studies have mainly (though not exclusively) identified common variants (minor allele frequency >5%) with modest effects which collectively only explain ~10% of the additive genetic variance of CAD [3]. The story is similar for ischemic stroke (and subtypes) [6–9], idiopathic dilated cardiomyopathy (DCM; a leading cause of heart failure) [10–12] and MI [13–15]. Common genetic determinants of hypertension and dyslipidaemia, major risk factors for CVD, have also been identified through the study of blood pressure [16–24] and blood lipid levels [25, 26] as quantitative traits in the general population.

To date, these large GWASs have primarily focused on common single-nucleotide polymorphism (SNPs) which are well covered by commercial genotyping platforms and imputation reference
panels although lower frequency variants have shown evidence of association. For example, low-frequency (1–5% minor allele frequency) missense and rare (minor allele frequency <1%) non-sense SNPs in \textit{LP4} \cite{25, 26} and \textit{PCSK9} \cite{28, 29}, respectively, have been shown to have large effects on low-density lipoprotein (LDL) cholesterol and risk of CAD. Next-generation genotyping arrays, which target rare coding SNPs (’exome chip’ arrays), and next-generation re-sequencing approaches which assay all genome-wide or exome-wide variation in an individual, open up opportunities for novel discovery of low-frequency and rare variants associated with disease. At the time of writing this review, few large-scale studies of cardiovascular phenotypes based on exome chip or re-sequencing data had been published. However, there is a wealth of literature describing rare variant discovery for monogenic CVD and these studies have provided valuable mechanistic insight into cardiovascular health and disease.

In this review, we will briefly summarise the literature on genetic loci associated with CAD and overlap with the genetic determinants of blood pressure and blood lipid levels, with special emphasis on low-frequency and rare variants identified through GWAS. We will discuss the discovery of rare variants underlying familial and monogenic CVD and how this has informed the understanding of the underlying biology. In addition, we will describe the overlap of loci identified through rare variant discovery and GWAS screens of common variants.

**GWASS OF CVD AND RELATED RISK FACTORS**

Since 2007, GWAS of CAD have amassed sample sizes of up to 63,000 cases (and more than 130,600 controls) and identified a total of 47 loci with genome-wide or array-wide significant evidence of association with risk of CAD \cite{3–5}. For the majority of these loci, the most strongly associated (sentinel) SNPs are common with modest effects. High blood pressure and circulating plasma levels of low-density lipoprotein and high-density lipoprotein (HDL) cholesterol, total cholesterol and triglycerides are modifiable risk factors for CAD and are known to have a genetic component. Studies based on sample sizes of up to 200,000 individuals have so far identified more than 40 loci associated with blood pressure \cite{16–24, 30} and 157 associated with blood lipid levels \cite{25, 26}. Five CAD loci (\textit{CYP7A1-NT5C2, SH2B3, GUCY1A3-GUCY1B3, FURIN-FES} and \textit{ZC3H11}) are shared with blood pressure traits and 20 CAD loci are shared with blood lipid loci. This suggests that there are multiple underlying mechanistic pathways which influence the risk of CAD. A recent study of rare coding variants (utilizing exome chip data) identified four additional loci associated with HDL cholesterol or triglycerides although these variants did not show association with risk of CAD \cite{31}. Among the loci in common between CAD and blood lipid loci, two contain low-frequency variants with large effects on risk of CAD.

\textit{LP4} encodes apolipoprotein(a) (apo(a)), a component of lipoprotein(a). Various isoforms of apo(a) exist with variable numbers of kringle 4 type 2 repeats; cysteine-rich sequence that form loop-like structures. A low-frequency missense SNP in \textit{LP4}, rs3798220 (minor allele frequency 2%), is associated with LDL cholesterol levels \cite{25} and, of the 47 variants associated with risk of CAD, shows the biggest effect with an increase in risk of 51% \cite{4, 27}. This SNP is correlated with short isoforms of apo(a) \cite{32}, and low copy number of these repeats is associated with the risk of CAD \cite{32} and MI \cite{33} through an effect on lipoprotein(a) levels \cite{34}.

Genetic variants in \textit{PCSK9} show association with lipid levels \cite{26, 31, 35, 36} and CAD \cite{14}. \textit{PCSK9} plays a role in cholesterol regulation and encodes a protein which leads to degradation of the LDL receptor; loss of function of \textit{PCSK9} therefore leads to increased metabolism of LDL cholesterol by the liver and decreases circulating LDL cholesterol. Rare nonsense variants in \textit{PCSK9} in an African American ancestry population (minor allele frequencies 0.8–1.8%) are associated with a marked reduction in LDL cholesterol and up to 88% reduced risk of CAD \cite{28, 29, 31}. This illustrates the potential use of interventions which inhibit PCSK9 in lowering LDL cholesterol and risk of CAD. A recent phase 2 clinical trial of a monoclonal antibody to PCSK9 has reported decreased circulating LDL cholesterol \cite{37} and phase 3 trials have begun.

Although large numbers of loci are now known to show association with CAD and/or blood pressure and lipid levels, the effects of these common variants still collectively explain only a very modest proportion of the genetic variance. Obesity is another modifiable risk factor for which large GWASs have identified a number of genetic loci \cite{38–40}. Similarly, for Type 1 and 2 Diabetes, a total of
~100 genomic loci have now shown evidence of association [41–43]. These studies have highlighted many potentially interesting genes, many of which have roles which could plausibly have an effect on the trait with which they are associated. However, a large proportion of the genes identified do not have an obvious link with previous biological knowledge of the trait and many are of unknown function. Lower frequency and rare variants identified through GWAS have tended to have larger effect sizes and present more obvious choices of candidates for follow-up, as for the examples described above.

FAMILIAL AND MONOGENIC CVD
While cardiovascular phenotypes such as blood pressure and CAD are known to be complex traits which are driven by multiple genetic and environmental factors, there are forms of inherited disease which have been demonstrated to be caused by specific rare mutations which segregate with disease within families. Although many common genetic variants associated with cardiovascular traits have been identified, identification of causal rare variants in familial and monogenic forms of CVD has more directly contributed to understanding of the underlying biology. These rare variants have tended to directly affect protein structure and function, whereas most common variants which have been identified are intronic or intergenic and are likely to have regulatory roles.

As re-sequencing has become affordable for small family sizes, candidate gene or exome re-sequencing within affected families has largely superseded linkage as the method of choice to identify putatively causative mutations for rare disorders exhibiting Mendelian inheritance. Where there is already information about which gene or genes are most likely to contain the causative mutation based on previous studies, genotyping of known recurrent mutations or complete gene re-sequencing to identify new mutations is undertaken. If no novel mutation is identified which satisfies the criteria for causality, whole-exome re-sequencing can be employed. The challenges, advantages and disadvantages of this approach are beyond the scope of this review.

Here we compare the findings from studies of monogenic disease and GWAS for three examples: hypertension, DCM and Brugada syndrome. These examples illustrate how identification of rare causative mutations segregating in families, and identification of common variants associated with cardiovascular phenotypes in the general population, are converging to give insight into the genetic architecture of these traits.

Monogenic hypertension and hypotension
Blood pressure is a complex trait that is known to be influenced by multiple genetic and environmental factors. Clinical diagnosis of hypertension (high blood pressure) is traditionally based on a clinic systolic blood pressure of more than 140 mm Hg and diastolic blood pressure more than 90 mm Hg. However, there are forms of hypertension and hypotension, which are usually early-onset and with additional specific clinical characteristics, which are known to be caused by single mutations in single genes. At least 17 genes have so far been identified that harbour rare mutations which cause these monogenic forms of hyper- and hypotension, and this has furthered understanding of the mechanisms underlying blood pressure regulation [44, 45]. In particular, these disorders have advanced understanding of the role of sodium transportation in the distal nephron of the kidney [45].

Familial hyperkalaemic hypertension (FHHt; also known as Gordon syndrome or pseudohypoaldosteronism type II) is characterised by hypertension and elevated potassium levels in the blood despite normal glomerular filtration [46, 47] and can be treated with thiazide diuretics and dietary sodium restriction. Autosomal dominant and recessive modes of inheritance have been recorded. Linkage analyses identified loci at 12p13.3 [48] and 17q21 [49] and rare mutations in the kinase WNK1 and WNK4 genes were identified [49]. Subsequently, re-sequencing in families with FHHt but without mutations in WNK1 and WNK4 identified causative rare mutations in KLHL3 and CUL3 [50–52]. WNK4 plays a regulatory role in renal salt reabsorption and potassium secretion and KLHL3 and CUL3 are components of an E3 ubiquitin ligase complex which target WNK1 and WNK4 for ubiquitination and degradation. Mutations in WNK4, KLHL3 or CUL3 impair these interactions and lead to increased levels of WNK4. This in turn leads to increased salt retention in the kidney through activation of the Na–Cl (NCC)/Na–K–2Cl (NKCC2) ion co-transporters and inhibition of potassium secretion through the renal apical potassium channel.
ROMK [53–55], leading to hypertension and hyperkalaemia.

Bartter syndrome (of which there are now known to be at least five subtypes) and Gitelman syndrome are normotensive or hypotensive disorders characterised by hypokalaemic alkalosis, hypocalciuria and renal salt-wasting. Despite similar clinical presentations, these syndromes are clearly distinguishable by their underlying rare genetic mutations in genes encoding proteins with functions relating to solute transport in the kidney nephron. Gitelman’s syndrome and Bartter’s syndrome Type 1 are the result of mutations in genes encoding the NCC (in the distal convoluted tubule) and NKCC2 (in the thick ascending limb of the loop of Henle), respectively [56–59]. Bartter’s syndrome Type 2 is characterised by mutations in the gene encoding the ROMK potassium channel. Bartter’s syndrome Type 3 is caused by mutations in the gene encoding the major component of one of the major chloride channels in the kidney and Type 4 by mutations in the gene encoding an essential chloride channel subunit. Bartter’s syndrome Type 5 is caused by mutations in a calcium-sensing receptor which is expressed in the renal tubule.

In general, large GWASs of blood pressure have identified different genes to those implicated in monogenic disorders, with few exceptions. Rare mutations in CYP17A1 can cause 17-alpha-hydroxylase deficiency which leads to hypertension and is implicated in a small proportion of cases of congenital adrenal hyperplasia [60]. Common variants near CYP17A1, identified through GWAS, have shown strong association with blood pressure in European and South and East Asian populations [17, 20–23]. Although, there has been some evidence of association of common WNK1 variants with blood pressure and hypertension through haplotype analysis [61–64], individual common variants in WNK1 have not shown association with blood pressure in the general populations in large GWAS [17]. Given that GWASs, focused primarily on common variation, have only explained up to 3% of the variance of blood pressure, it has been hypothesised that rare variants in genes which underlie monogenic disorders may also contribute to blood pressure variation in the general population. Analyses of the individual and combined effects of rare and common variants within genes implicated in monogenic hypertension have shown some weak evidence of association with blood pressure in the general population [65–67]. However, it is clear that, to date, these studies have lacked statistical power to fully explore the role of rare variants in these genes in blood pressure variation in the general population.

**Dilated cardiomyopathy**

DCM is a cause of heart failure and is one of the most common reasons for heart transplant. Both familial and idiopathic forms of DCM exist although it is estimated that around a third of idiopathic cases of DCM are found to be familial when family members are clinically screened [68]. Much of the genetic discovery has emerged from studies of families with DCM which have identified rare variants in more than 33 genes [69, 70]. These genes predominantly encode proteins with a role in muscle function (sarcomere, Z-disc and cytoskeleton proteins) with the largest proportion being found in the titin (connectin) gene [71]. Many of these rare variants appear to be private mutations having only been observed in a single family and, despite the large number of genes identified, these still only explain around 35% of DCM. In addition, there is uncertainty about whether a small proportion of the published familial DCM rare variants are truly causative [72]. A complexity of the relationship between the thus far identified genetic causes of familial DCM and the aetiology of disease is that even within families, individuals carrying the putative causal rare variant can exhibit heterogeneity of symptoms. This suggests that although segregation of these variants with disease within families enables screening of other family members for DCM, the relationship between genotype and phenotype is still not well understood. GWASs of DCM have identified a number of susceptibility loci [10–12]. Interestingly, a gene identified through a common SNP association in a genome-wide scan, BAG3, was subsequently shown to harbour rare variants, including a deletion, believed to be the underlying cause of cases of familial DCM [12, 73].

**Brugada syndrome**

Rare sequence variants in genes encoding cardiac sodium and potassium ion channels, essential for correct electrical conduction in the heart, have also been shown to cause a variety of rare and familial monogenic cardiac arrhythmias and conduction disorders, for example, Brugada syndrome, long QT syndrome and inherited cardiac conduction disease [74–77]. Brugada syndrome, characterised by an abnormal
CONCLUSIONS AND FUTURE DIRECTIONS

The study of rare monogenic cardiovascular disorders, and genome-wide hypothesis-free screens for association of common variants with cardiovascular traits and related risk factors, has fuelled understanding of the underlying biology and informed approaches to treatment. GWASs have yielded a large number of novel loci with a putative effect on cardiovascular health which, with further functional investigation, could uncover alternative relevant biological pathways. To date, the overwhelming majority of these loci have been identified through association with common variants with minor allele frequencies >5%, largely reflecting the predominantly common content of the genotyping micro-arrays and imputation panels utilised. Large sample sizes of individuals with and without disease, and of healthy and diseased individuals with measurements of quantitative traits, are now available. Strategically designed arrays and imputation panels, along with increasingly affordable re-sequencing options, will lead to more novel discovery at the rare end of the minor allele frequency distribution. Most large studies to date have comprised mainly European ancestry individuals. However, large consortia bringing together Asian and African ancestry samples are increasing in size and trans-ethnic analyses, which exploit allele frequency differences between ancestries, have the potential to yield novel discovery and additional insight into previously identified loci associated with cardiovascular traits [81].

Gene-based approaches, where association is tested in one step across multiple low-frequency and rare variants in a gene, may provide further insight into the role of rare variants in cardiovascular health in the general population. These approaches are best suited to re-sequencing studies and ‘exome chip’ analyses which are designed for measurement of rare non-synonymous variants.

In summary, GWAS and studies of monogenic disease have provided complementary approaches to the study of the genetics of cardiovascular health and disease. As studies of rare variants in large sample sizes are undertaken, it is expected that findings from these approaches will merge to provide a complete picture of the genetic architecture of this important group of traits and diseases.

Key points
- Although many loci showing association with cardiovascular disease and traits, including common modifiable risk factors, have been identified through genome-wide association studies, these collectively only explain a small proportion of the phenotypic variance (typically <10%).
- Inherited monogenic disease, for example, monogenic forms of hypertension, is caused by rare mutations in key genes and identification of these mutations has greatly informed understanding of the molecular mechanisms of disease.
- To date, there has been limited overlap of the findings for rare and common variants but advances in technology, particularly next-generation sequencing, will enable study of the effects of rare variants on cardiovascular disease risk and health in the general population.
- Combining insight from studies of common and rare variants will lead to a more complete understanding of the molecular mechanisms of cardiovascular health and disease which will ultimately inform approaches to treatment.

References


