



**Cite this article:** Villar D, Frost S, Deloukas P, Tinker A. 2020 The contribution of non-coding regulatory elements to cardiovascular disease.

*Open Biol.* **10**: 200088.

<http://dx.doi.org/10.1098/rsob.200088>

Received: 2 April 2020

Accepted: 8 June 2020

**Subject Area:**

genomics/genetics/molecular biology

**Keywords:**

cardiovascular, gene regulation, GWAS, epigenetics, functional genomics

**Author for correspondence:**

Diego Villar


e-mail: [d.villarozano@qmul.ac.uk](mailto:d.villarozano@qmul.ac.uk)

# The contribution of non-coding regulatory elements to cardiovascular disease

Diego Villar<sup>1</sup>, Stephanie Frost<sup>1</sup>, Panos Deloukas<sup>2</sup> and Andrew Tinker<sup>2</sup>

<sup>1</sup>Blizard Institute, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, 4 Newark Street, London E1 2AT, UK

<sup>2</sup>William Harvey Research Institute, Heart Centre, Barts and the London School of Medicine and Dentistry, Queen Mary University of London, Charterhouse Square, London EC1M 6BQ, UK

 DV, 0000-0002-8614-6232; SF, 0000-0003-2174-0510; PD, 0000-0001-9251-070X; AT, 0000-0001-7703-4151

Cardiovascular disease collectively accounts for a quarter of deaths worldwide. Genome-wide association studies across a range of cardiovascular traits and pathologies have highlighted the prevalence of common non-coding genetic variants within candidate loci. Here, we review genetic, epigenomic and molecular approaches to investigate the contribution of non-coding regulatory elements in cardiovascular biology. We then discuss recent insights on the emerging role of non-coding variation in predisposition to cardiovascular disease, with a focus on novel mechanistic examples from functional genomics studies. Lastly, we consider the clinical significance of these findings at present, and some of the current challenges facing the field.

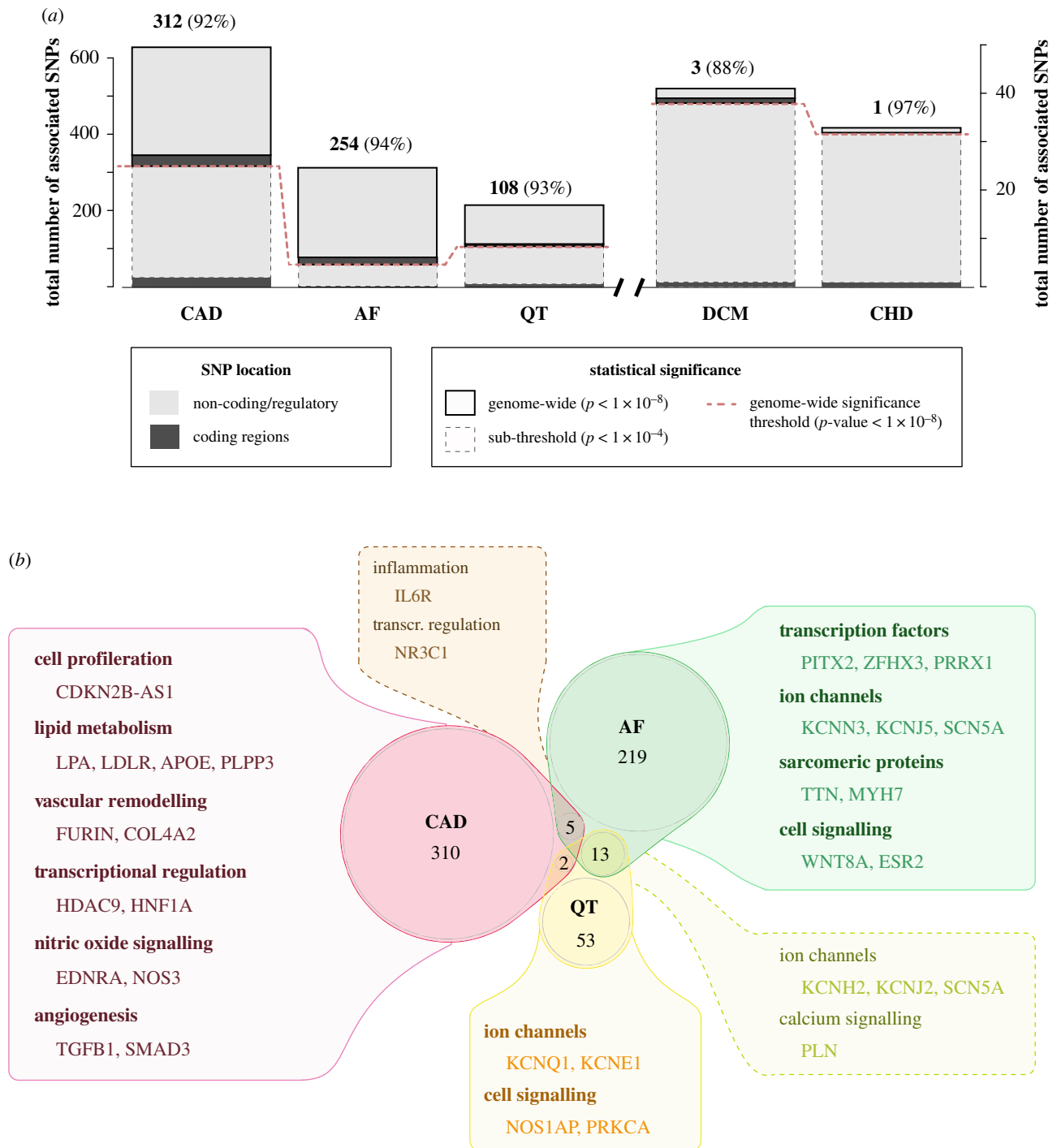
## 1. Introduction

Human individuals differ from each other in millions of DNA sequence variants, some of which contribute to phenotypic differences in physiology and disease predisposition. Cardiovascular disease (CVD) comprises a range of monogenic and complex diseases accounting for a quarter of deaths worldwide. The study of family pedigrees has revealed new molecular players in monogenic cardiovascular disorders such as familial forms of cardiomyopathy, yet the most prevalent cardiovascular conditions are complex diseases with a heritable component. Facilitated by technological advances in sequencing technology, genome-wide association studies (GWAS) have analysed human genetic variation across individuals presenting a range of cardiovascular traits and diseases [1–3]. These comprise physiological parameters of cardiovascular function, such as heart rate or electrocardiogram measurements, and common pathologies such as cardiac arrhythmias and coronary artery disease (CAD).

A general finding from cardiovascular GWAS is the high prevalence of common genetic associations in the non-coding fraction of the human genome. The majority of human loci associated with cardiovascular traits and diseases are not in linkage disequilibrium (LD) with coding exons [4], thus a large fraction of genetic associations must reflect the effect of variation in gene regulation. Here, we will review recent evidence from genetic studies into the potential role of non-coding DNA elements in cardiovascular physiology and disease, with discussion of mechanistic insights from molecular approaches evaluating the effect of non-coding genetic variation on gene regulation and cellular phenotypes.

## 2. Genetic association studies in cardiovascular disease

Population genetics in CVD has classically focused on the analysis of familial pedigrees of suspected monogenic disorders, and for the last decade on association studies investigating complex diseases genome-wide. For common traits and diseases, GWAS are typically conducted on hundreds to thousands of unrelated individuals that are genotyped on up to two million single nucleotide



**Figure 1.** Summary of genome-wide association signals and proximal genes for major cardiovascular diseases and traits. (a) Total number of associated single nucleotide polymorphisms (SNPs) for coronary artery disease (CAD), atrial fibrillation (AF), QT interval duration (QT), idiopathic dilated cardiomyopathy (DCM) and congenital heart disease (CHD). For each disease or trait, top solid-outline boxes correspond to SNPs at genome-wide significance with  $p$ -value lower than  $1 \times 10^{-8}$ , and bottom dashed-outline boxes to sub-threshold SNPs ( $p$ -value  $< 1 \times 10^{-4}$ ). The dashed red line denotes the genome-wide significance threshold. SNPs located in coding regions of the genome are represented as black bars, and with those in non-coding segments in light grey. Numbers on top of each bar indicate the total number of associated SNPs at a genome-wide level of significance ( $p$ -value  $< 1 \times 10^{-8}$ ), and percentages the non-coding fraction across all associated SNPs. Data from [5]. (b) Genes proximal to the associated SNPs in (a). ( $p$ -value  $< 1 \times 10^{-8}$ ) are represented as a Venn diagram for CAD, AF and QT (centre). Numbers indicate gene counts in each region of the plot, including genes associated with two traits/diseases. The boxes linked to each region summarize example genes proximal to GWAS signals, and the functional categories they belong to.

polymorphisms (SNPs) found in the population at high frequency (greater than 5% for minor alleles). Genotyped SNPs are then used to impute between 10 and 20 million additional variants including indels, and each of these variants is tested for association with a trait of interest, providing a global view of potential candidate disease loci across the whole genome. Recently, SNP arrays have been complemented by

including low-frequency (MAF 1–5%) and rare (MAF  $< 1\%$ ) non-synonymous coding variants. In this section, we discuss the findings of GWAS across key types of CVD, and their relevance to our understanding of how non-coding regions of the genome may contribute to disease risk (see also figure 1).

Association signals from GWAS are commonly reported for variants showing a  $p$ -value lower than  $1 \times 10^{-8}$ , which is

widely accepted as a genome-wide significance threshold (figure 1a, top bars with solid outline). More recently, several studies have also interrogated association signals at higher  $p$ -values (typically  $p < 1 \times 10^{-4}$ ), under the assumption that these may also include functionally relevant signals (figure 1a, bars with dash outline). Sub-threshold variants have been shown to include those at a false discovery rate lower than 5% [6], and to enrich for tissue-specific epigenetic signals [7,8]. Nevertheless, it is likely they are overall weaker in their functional significance when compared to variants at genome-wide significance, as suggested by their modest contribution to genetic risk scores [6]. For clarity, we focus our discussion of GWAS findings on reported variants at genome-wide significance ( $p < 1 \times 10^{-8}$ ), which to date account for most of those that have been functionally characterized.

### 2.1. Coronary artery disease and myocardial infarction

CAD due to atherosclerosis and potential myocardial infarction (MI), after occlusion of a coronary artery due to plaque rupture, is a major cause of mortality and morbidity. CAD is a paradigm of complex disease, with an estimated heritability of 50% [9] and multiple risk factors (e.g. high lipid levels, hypertension, diabetes and inflammation) which in turn appear to be under genetic control. Given its huge societal burden, CAD and MI were among the first complex diseases to be targeted in GWAS.

To date, genetic associations studies have collectively assessed over 70 000 CAD cases and hundreds of thousands of controls, revealing over 160 genomic loci associated with CAD and MI at a genome-wide level of significance (reviewed in [10]). CAD and MI share a very similar genetic architecture, with two notable exceptions (the *Abo* and the *Hdca9* loci) [11]. In line with the complex nature of CAD, genetically associated loci are involved in a range of biological processes, such as lipid metabolism, vascular tone, inflammation or vascular remodeling [6,12]. However, despite important biological insights into variants regulating cholesterol metabolism [13,14] and nitric oxide signalling [15], the precise molecular mechanisms underlying the majority of associated SNPs with CAD risk are largely unknown.

The complex genetic association data in CAD also exemplifies the importance of co-occurrence of common risk variants in a given individual, and their potential joint contribution to disease risk. Given the number and population frequency of genetic variants associated with CAD, an individual is expected to carry over a hundred of currently known risk alleles. In this regard, a recent evaluation of polygenic scores [16] estimated 8% of the population being at more than threefold increased risk for CAD—a prevalence 20 times higher than the carrier frequency of rare monogenic mutations conferring a comparable disease risk [17].

### 2.2. Congenital heart disease

Congenital heart disease (CHD) is the most common type of birth defect (affecting approx. 1% of live births) and comprises structural abnormalities affecting the heart septum, valves and the outflow tract. CHD is genetically heterogeneous, and even within the same abnormality, differences in clinical symptoms across patients are very common [18]. The post-natal survival of CHD patients has dramatically improved in the last decade due to the implementation of

corrective cardiac surgery. However, most CHD patients suffer cardiac complications in adulthood, with arrhythmias and heart failure being the most common.

Some forms of CHD are associated with well-known chromosomal abnormalities (such as trisomy of chromosome 21 [19]) and copy number variations, which account for approximately 25% of cases [20]. Studies to date have identified over 50 genes with point mutations associated with familial CHD, which largely correspond to cardiac developmental genes (e.g. transcription factors *GATA4* and *NKX2-5*). However, for the vast majority of CHD patients, no apparent chromosomal abnormality or causal mutation can be identified. In fact, familial studies have often failed to identify a single candidate genetic variant that completely segregates with the disease [21,22]. These findings collectively suggested a significant contribution of *de novo* mutations and polygenic inheritance in CHD.

Because congenital defects are expected to be under strong selective pressure, *de novo* mutations may account for a significant fraction of sporadic CHD occurrences. Recent studies analysing hundreds of disease cases and controls have shown an increased burden of *de novo* mutations in genes involved in transcriptional regulation, including enzymes controlling the reading and removal of histone 3 lysine 4 methylation [23,24] and the pleiotropic transcription factor *NR2F2* [25]. *De novo* mutations may account for up to 10% of CHD cases [26], with an increased prevalence among patients where CHD is accompanied by extra-cardiac abnormalities such as neurodevelopmental disorders [24]. More recently, exome sequencing of approximately 2500 trios suggested distinct gene pathways associated with *de novo* and recessive forms of CHD, with mutations in chromatin-modifying genes being enriched in *de novo* sporadic cases [27].

The potential role of polygenic inheritance in CHD has been highlighted by recent GWAS studies focused on common genetic variation, which to date have identified a relatively low number of associations at a genome-wide level of significance (figure 1a). Two strong associations were identified in tetralogy of Fallot patients, one of which falls in a large intronic region with potential regulatory function [28]. Across thousands of CHD cases encompassing all major clinical types, genetic associations were most significant when patients were segregated by CHD subtype, and the strongest association was found in an intergenic region proximal to the homeobox transcription factor *MSX1* [29].

These studies, therefore, suggest a key role for gene regulatory processes in mediating congenital heart disease. Focusing on regulatory regions of the developmental transcription factor *TBX5*, Smemo *et al.* [30] were among the first to provide molecular and mechanistic evidence for the role of transcriptional enhancers in CHD—more recent examples will be the focus of later sections of this review.

### 2.3. Long QT syndrome

The continuous and coordinated contraction of the heart relies on tightly regulated opening and closing of specialized ion channels in heart muscle cells. Cardiac arrhythmias result from disruptions in myocyte electrical function and are a leading cause of sudden cardiac death, stroke and heart failure. The genetics of a wide spectrum of cardiac rhythm disorders has been reviewed elsewhere [31]—here, we will focus on two common arrhythmias, namely long QT syndrome (LQTS)

and atrial fibrillation (AF), which have been the focus of recent genetic studies.

Cardiac arrhythmias have been classically studied within family pedigrees, such as in the pioneering work that led to the mapping of the chromosomal loci underlying hereditary LQTS [32]. This form of inherited arrhythmia leads to a prolonged duration of ventricular action potentials, and thus an extended QT interval on the electrocardiogram. Affected individuals have a propensity to a particular form of ventricular tachycardia known as torsade-de-pointes, and this can degenerate into ventricular fibrillation leading to sudden death if untreated. Familial studies identified three major causative genes in two potassium and the cardiac sodium channel genes, which gave rise to the classification of patients into three main LQTS types. Later work extended the number of genes linked to hereditary LQTS to 15, although most of these additional mutations are very rare (reviewed in [33]). In most cases, LQTS mutations lead to defective channel function or trafficking, as shown in a number of experimental systems—including engineered mouse models [34] and human patient-derived induced pluripotent stem cells (iPSCs) [35].

The QT interval is heritable in the wider population and GWAS have identified common genetic variants associated with this electrocardiographic trait. The first study [36] identified a strong association signal in the *NOS1AP* locus, a potential regulator of ion channel function [37]. Subsequent studies have identified genetic variants associated with electrocardiographic traits [38,39]. These are proximal to many of the same ion channel genes that harbour familial LQTS mutations, in addition to some developmental transcription factors (e.g. *MEIS1* and *TBX5*) [40]. A recent GWAS of QT interval duration across over 70 000 individuals identified loci associated with cardiomyocyte calcium handling, highlighting a potential role of repolarization pathways in LQTS susceptibility [41] (see also figure 1).

The variable clinical manifestations in LQTS patients with the same gene mutations [42] have led to the investigation of potential modifier genes contributing to incomplete penetrance or variable expressivity (reviewed in [43])—including non-coding genetic variants). An elegant study focusing on a South African founder population for the *KCNQ1*-A341V mutation [44] demonstrated that two common non-coding *NOS1AP* variants associated with increased risk of life-threatening arrhythmia in LQTS patients. In fact, variants in this locus have also been linked to drug-induced LQTS risk [45]. Conversely, a case-control study in LQTS duos identified an intronic *KCNQ1* variant associated with lower arrhythmic risk [46]. Recently, iPSCs were generated from several members of a family with type 2 LQTS, including severely and mildly affected individuals [47]. By combining exome sequencing and electrophysiological measurements, the authors revealed a protective gain-of-function mutation in *KCNK17* and an aggravating variant in *REM2*. In sum, these studies have provided proof-of-principle evidence for the existence of predisposing and protective modifier genes in LQTS. Recent prospective analyses assessing the value of multi-locus genetic scores [48] illustrate the potential for variants involved in disease predisposition to inform clinical management of LQTS patients.

## 2.4. Atrial fibrillation

AF is the most common cardiac arrhythmia (affecting up to 2% of the general population). It occurs due to abnormal electrical

impulses firing in the atria that override the natural pacemaker of the heart, and lead to an uncontrolled and highly irregular heart rhythm. Although AF can develop secondarily to chronic cardiac stress and other heart diseases, it has a substantial genetic heritability estimated at approximately 60% [49]. Initial investigations into the genetic causes of AF focused on candidate gene screening in cohorts of early-onset lone AF patients or families showing Mendelian disease segregation. Although these classical studies discovered a number of rare coding variants in ion channel genes responsible for generating the atrial action potential [50], recent population-based approaches suggest coding mutations account for a relatively small proportion of genetically determined AF [51].

By contrast, analyses of common and rare variation across AF patients and controls suggest an important role for non-coding genetic variants in mediating disease susceptibility. The first major AF GWAS identified a significantly associated locus in an intergenic non-coding region on chromosome 4q25, adjacent to the *PITX2* gene [52]. The AFGen Consortium [53] and subsequent meta-analyses [54,55] have extended these GWAS to hundreds of thousands of AF patients and controls. To date, these efforts have collectively identified over 100 genomic regions associated with AF, the majority of which lie within introns or intergenic regions.

Genetic loci associated with AF are often within the non-coding regions of ion channel genes such as *KCNQ1*. These variants may predispose to AF by shortening or lengthening atrial action potential leading to changes in atrial tissue refractoriness, and can be carried in multiple combinations of rare and common variants with epistatic effects [56]. The second set of associated GWAS loci are proximal to transcription factors such as *PITX2*, *ZFHX3*, *GATA4*, *NKX2-5* and *TBX5*—mostly with previously known roles in heart development. Four AF-associated regions at 4q25 are closely associated with the *PITX2* gene, and thus deficiency of the major heart isoform *PITX2c* was thought to mechanistically contribute to AF. In agreement, haploinsufficient *PITX2* mice show a propensity to atrial arrhythmias and AF [57], albeit coding mutations in *PITX2* are uncommon in AF patients [58]. The role of non-coding SNPs corresponding to the four association signals in the 4q25 locus is still under investigation, and recent studies have demonstrated the interaction of enhancer elements in this region with *PITX2*, with significant differences in *PITX2c* expression levels between allele variants [59]. Intriguingly, evaluation of *PITX2c* expression in human AF patients and controls have been inconclusive in observing significant differences [60], suggesting the effects of *PITX2* on AF susceptibility may occur during development or in response to cardiac stress in the adult. In this regard, *PITX2* is known to interact with other transcription factors whose loci also harbour AF-associated non-coding variants (e.g. *ZFHX3* and *TBX5*)—suggesting an altered atrial regulatory network may play a role in AF pathogenesis (reviewed in [61]). Thirdly, GWAS loci associated with AF are also found close to structural heart muscle genes, such as those encoding for sarcomeric MYH6/7, TTN and associated regulatory proteins (e.g. CAMK2D). Variants in these loci potentially contribute to an atrial cardiomyopathy that may be exacerbated by chronic cardiac stress. In this regard, MYH7 expression has been shown to be elevated in human patients and animal models of chronic AF [62].

Despite the plethora of genetic associations in AF patients (figure 1), genetic testing is not recommended in standard

clinical diagnosis. The complexity of GWAS associated genes in AF and a limited knowledge on the causative role of most candidate genes probably underlie the poor performance of polygenic scores based on GWAS [63].

In sum, genetic studies suggest a role for transcriptional enhancers and non-coding genetic variation in contributing to disease penetrance in LQTSs, as well as for regulatory networks of cardiac development in mediating onset and progression of AF. Mechanistic examples of specific loci will be discussed in §4 of this review.

## 2.5. Cardiomyopathies

Cardiomyopathies are a heterogeneous group of diseases characterized by impairment of heart muscle function. Although they can be secondary to factors such as ischaemia or hypertension, cardiomyopathies are often heritable. Although left ventricular hypertrophy and systolic heart failure often associate with cardiomyopathies, we will not cover either here due to their very diverse aetiologies [64]. Instead, here we focus on the two major clinical types of primary cardiomyopathies: hypertrophic cardiomyopathy (HCM) and dilated cardiomyopathy (DCM). HCM is a major cause of sudden death and heart failure, and functionally defined by a characteristic pattern of ventricular hypertrophy and diastolic dysfunction; whereas DCM manifests with ventricular dilation and systolic dysfunction.

More than half of HCM patients have a family history of the disease, with a 20–35% figure in DCM. Therefore, genetic studies have classically focused on linkage analysis and candidate gene approaches, revealing rare variants associated with familial cardiomyopathy. HCM is mainly linked to mutations in genes for sarcomeric proteins, especially MYH7 [65,66]. Linkage analysis in HCM families not linked to the MYH7 locus has identified mutations in additional genes encoding contractile elements of the sarcomere, such as MYCBP3 or TTN2. Further sequencing of candidate loci found HCM-associated mutations in genes for other components of the contractile machinery, including MYL2, MYL3 and TNNC1. Overall, contractile element mutations are found in nearly half of patients with familial HCM, leading to the notion that this cardiomyopathy is primarily a disease of the cardiac sarcomere [67].

By contrast, DCM patients display substantial genetic heterogeneity. More than 50 genes have been linked to familial DCM, some of which overlap with those associated with HCM. A major DCM-associated gene is *TTN*, encoding for the giant protein Titin—mutations in *TTN* are associated with 20% of DCM cases [68]. However, current genetic testing of DCM patients, even with large panels of genes, has only approximately 50% sensitivity in detecting mutations. Even within the same family, there is substantial heterogeneity of symptoms for individuals carrying the same rare variant. Moreover, the frequency of known pathogenic mutations in HCM and DCM is much lower than the prevalence of these diseases, suggesting up to 3–5% of the population carries genetic variants associated with cardiomyopathy (i.e. a significant fraction of patients have no characterized pathogenic variants). In line with these findings, recent large-scale exome sequencing of DCM patients suggested a robust disease association for only 12 of 56 genes commonly implicated in DCM [69].

GWAS studies of cardiomyopathy patients have detected few associated non-coding variants at a high level of

significance (figure 1a). GWAS of DCM [70,71] identified a number of susceptibility loci, corresponding to non-coding variants proximal to *HLA-C* and *HSP7*, and common coding variants in the *BAG3* gene, for which rare variants were also found in familial DCM [72]. More recent studies in specific populations have reported non-coding associated variants in the *CACNB4* cardiac channel locus [73] (in an African American population), and rare coding variants in the developmental cardiac TF *NKX2-5* and the cytoskeletal protein *FLNC* in Icelanders [74]. Overall, the low number of associated loci identified may be related to idiopathic DCM often being a consequence of pre-existing conditions (such as congenital heart disease). It is therefore possible that genetic heterogeneity of DCM GWAS cohorts is substantially higher than that of other CVD GWAS, as suggested by the low replicability of association signals across studies.

The above results also contrast with reports quantifying the degree of gene expression changes in DCM patients, which have documented over 250 genes as differentially expressed [75]. Among these were many DCM-associated genes in GWAS, including the cardiac transcription factor *TBX20* [76] and many of its target genes. Focusing on a carefully phenotyped cohort, a recent study has also revealed structural genetic variants (e.g. insertions, deletions and duplications) associated with DCM [77], most of which reside in non-coding regions. Nevertheless, the mechanistic role of the few non-coding genetic variants associated with cardiomyopathies remains poorly understood.

## 3. Experimental approaches for functional annotation of non-coding variation

By design, the GWAS studies discussed above identify broad genomic regions (loci) associated with CVD, which are often proximal to several candidate genes. This is because common genetic variants in the human genome are pervasive (occurring on average every 300 bp) and typically co-segregate in blocks of genomic DNA. This co-segregation of alleles is termed LD, and is often the case that GWAS lead variants will be in high LD ( $r^2 > 0.8$ ) with many other sequence variants within a locus. Moreover, comprehensive catalogues of common sequence variation such as the 1000 Genomes Project allow consideration of all candidate variants. These factors complicate the identification of the individual variant(s) underlying a GWAS association signal, and have prompted the development of statistical methods to identify credible subsets more likely to contain causal variants (reviewed in [78]).

In parallel, functional interpretation of GWAS results has prompted detailed and careful annotation of association intervals, spanning variants in high LD with a lead variant and sub-threshold genomic variants. In this section, we will review some of the current experimental approaches aimed at the functional annotation of GWAS variants and associated loci, with a focus on epigenomic approaches and the complementary evidence they provide for an investigation of non-coding regions associated with CVD. There are two related areas we will not cover here. First, *in vivo* transgenic reporter assays provide rich information on the tissue and developmental specificity of candidate regulatory elements, and are often used in combination with the strategies we discuss [79]. Second, we will not discuss gene expression annotation of GWAS in detail, such as the use of expression quantitative

trait loci (eQTLs) for variant interpretation. A substantial fraction of non-coding GWAS signals corresponds to eQTLs, and their use for the annotation of cardiovascular GWAS has been reviewed elsewhere [80].

### 3.1. Epigenetic annotation of non-coding elements

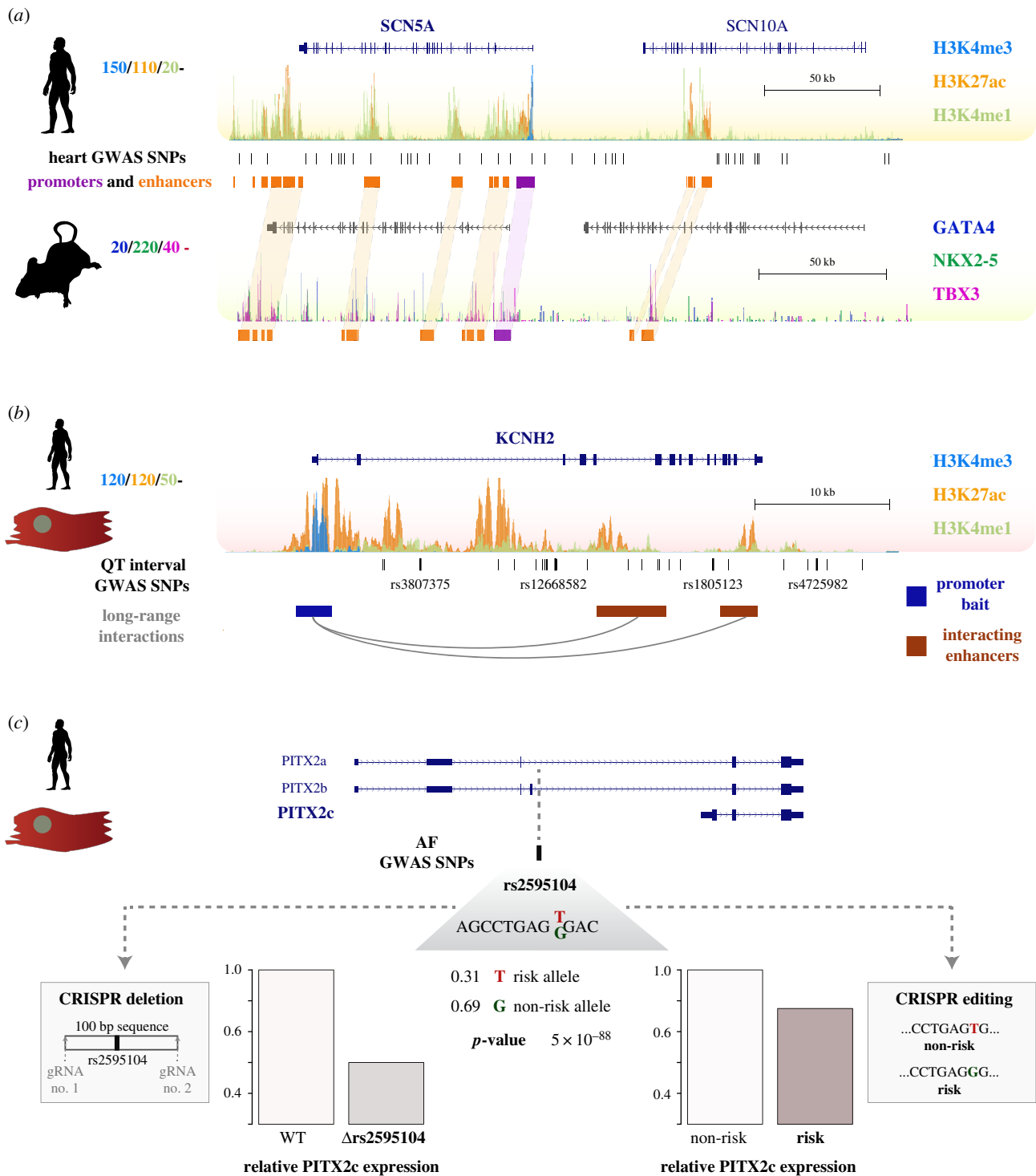
When combined with specific experimental protocols, sequencing technologies enable genome-wide profiling of hallmarks of regulatory activity across the genome. Transcription factor binding can be assessed genome-wide to identify all genomic elements bound by a protein in a tissue or cell line of interest, using chromatin immunoprecipitation coupled to DNA sequencing (ChIP-seq). This strategy can also be applied to histone modifications to identify DNA regions harbouring epigenetic marks characteristic of active or inactive regulatory elements. By leveraging these approaches, international consortia such as the ENCODE and Roadmap Epigenomics have discovered the unexpectedly high proportion of human non-coding DNA containing biochemical signatures of regulatory activity [81]. These efforts have profiled the binding of over a hundred transcription factors in human cells and tissues, as well as enrichment of histone modifications associated with the regulatory activity (e.g. promoters and enhancers). Their results suggest over 35% of the human genome can be considered as functional based on its potential to regulate tissue-specific expression patterns. Moreover, similar genome-wide approaches to profile RNA transcripts have indicated the human genome is also pervasively transcribed [81].

These genome-wide profiles provide invaluable information to interrogate genomic regions from GWAS, for example, to examine whether associated variants localize within regions with regulatory activity in a disease-relevant tissue. In fact, a recent epigenomic analysis of over a hundred human samples encompassing major tissue types [82] has highlighted the tissue-specific nature of a good fraction of genetic association signals. Non-coding GWAS signals were strongly enriched among genomic elements harbouring marks of regulatory activity, most notably H3K4me1 and H3K27ac (both characteristic of enhancer elements). This enrichment was markedly tissue-specific: across CVDs and traits, GWAS signals were most enriched in heart, vasculature and liver epigenomes. These results highlight the importance of investigating GWAS genomic regions in relevant tissues and cellular models, and in the context of dynamic regulatory environments [83]. Of note, recent work from the mouse ENCODE consortium reported significant conservation of human and mouse regulatory landscapes [84], supporting the utility of comparing human and mouse epigenome annotations.

In this regard, recent work has begun dissecting genome-wide changes in epigenetic marks, transcription factor binding and chromatin accessibility during cardiac development and in response to stress (reviewed in [85]). Applying functional genomics to null mouse models for the cardiogenic transcription factors NKX2-5 and TBX5, Luna-Zurita *et al.* [86] documented extensive cooperativity between these transcription factors and GATA4, which could often be explained by specific binding motif arrangements. Moreover, genetic ablation of either factor alone or in double knock-outs showed interdependent binding between the three factors, including significant ectopic binding of the remaining two TFs upon ablation of either one—suggesting mutations in these factors found in CHD patients can act both by loss of TF binding and abnormal ectopic

binding of cooperating TFs. Using a similar approach in an inducible TBX3 transgenic mouse, van den Boogard *et al.* [87] mapped genome-wide binding of the transcription factors TBX3, NKX2-5 and GATA4 in the developing heart, also documenting significant co-binding by the three TFs. Comparing gene expression changes upon TBX3 induction with regions bound by TBX3, the authors defined a core-set of ten ion channel genes as TBX3 targets, including the *SCN5A* cardiac sodium channel locus. Using their TBX3 ChIP-seq data, the authors identified candidate TBX3-bound enhancers in this locus and evaluated the heart-specific activity of orthologous human sequences in a mouse reporter assay. One of the functionally conserved human enhancers contained a SNP associated with cardiac arrhythmias, directly overlapping a TBX3 ChIP-seq peak and corresponding to a highly conserved position in the T-box binding consensus sequence. Accordingly, a reporter construct with the minor SNP allele did not respond to TBX3 overexpression, and associated with reduced activity by 60% in a zebrafish reporter assay. These two reports exemplify the power of epigenetic annotation approaches to inform transcriptional mechanisms in CVD, and the function of disease-associated non-coding variants (see also figure 2*a*).

Cardiac stressors such as mechanical overload play a well-documented role in heart disease. Under pressure overload, the heart has to contract and eject blood against excessive pressure, for example, caused by systemic hypertension or aortic stenosis. Over time, this form of stress results in pathological hypertrophy of the heart muscle, a process in which transcription factors active in the fetal heart are thought to be re-activated [90]. Several recent studies have employed functional genomics approaches to assess how extensive a ‘fetal gene programme’ is in cardiac hypertrophy. He *et al.* [91] focused on transgenic and knock-out *GATA4* mouse models to investigate this cardiac TF in mouse fetal and adult hearts, and in response to pressure overload. Their results showed *GATA4* binding differed markedly between fetal and adult samples: while two-thirds of adult *GATA4*-bound regions were also bound in E12.5 mouse hearts, 80% of fetal *GATA4* peaks were not bound in the adult. Comparison with H3K27ac data and genes regulated downstream of *GATA4* indicated an important functional role for regions showing both *GATA4* binding and H3K27ac signal, suggesting binding of this transcription factor mediates H3K27ac deposition and subsequent transcriptional changes. Upon pressure overload in adult mice, upregulated genes were significantly associated with fetal-specific *GATA4*-bound regions. However, cardiac stress-induced *GATA4* binding across a much larger set of new, stress-specific genomic regions. An earlier study [92] also found widespread changes in gene expression and the level of activating (e.g. H3K27ac) and repressive histone modifications under pressure overload. More recently, histone modifications associated with active transcriptional enhancers and downstream gene expression were profiled in left ventricle tissue from 18 idiopathic DCM patients and healthy controls [93]. Analyses of these data found over 7000 disease-associated regions showing gains or losses of regulatory activity. By comparison with data from the human fetal heart, the authors found around half of the disease-associated regions shifted towards a fetal-like epigenome and transcriptome. While consistent with the reactivation of fetal enhancers under cardiac stress and disease, these results also suggest a significant activation of new stress-responsive enhancers—which are not active during cardiac development but may also contribute to heart disease [94].



**Figure 2.** Experimental approaches for functional investigation of non-coding elements associated with cardiovascular disease. (a) Epigenomic annotation of the *SCN5A*/*SCN10A* cardiac disease locus in human and mouse left ventricle. Human genome tracks show epigenome signals for three histone marks associated with regulatory activity (H3K4me3 in blue; H3K27ac in orange and H3K4me1 in green) [82]; GWAS lead SNPs in this locus (black bars); and putative promoters (purple) and enhancers (orange). Mouse genome tracks below show ChIP-seq data for cardiac transcription factors GATA4 (blue), NKX2-5 (green) and TBX3 (purple) [88]. Orthologous promoters and enhancers in the human and mouse loci are connected by light purple and orange guides, respectively. (b) Epigenomic annotation of the *KCNH2* QT interval locus in the human left ventricle (epigenome signals as in (a)). Bottom tracks show genetic variants associated with QT interval duration (black bars), and long range interactions between the *KCNH2* promoter (blue) and enhancer elements (red) [89]. (c) Genomic location of the lead genetic variant rs2595104, associated with atrial fibrillation (AF) and located upstream of the *PITX2c* transcript annotation. The grey inset shows the sequence context of the variant, the minor risk allele (0.31 frequency), the major protective allele (0.69 frequency) and the *p*-value of the AF association. The regulatory effect of this variant on *PITX2c* expression was analysed in iPSC-derived cardiomyocytes by CRISPR/Cas9 deletion of a 100 bp sequence encompassing rs2595104 (left barplot, 54% reduction); and also by CRISPR genetic editing producing isogenic cardiomyocytes carrying the major non-risk allele and the minor risk allele (right barplot, 27% difference). Adapted with permission from [59].

### 3.2. Long-range chromatin interactions

A key property of non-coding regulatory elements such as transcriptional enhancers is long-range gene regulation: target genes of regulatory elements can be located hundreds of kilobases

away in the linear DNA sequence, complicating efforts to assign effector genes of disease-associated genetic variants located in the non-coding genome. However, a range of experimental approaches can be used to detect interactions between transcriptional enhancers and the promoter regions of distant

genes—thereby addressing an important limitation of the epigenetic annotation methods discussed above. The most common of these techniques are *in situ* fluorescent hybridization (FISH) or experiments based on proximity ligation of cross-linked chromatin fragments (known as chromosome conformation capture) [95]. Although both techniques have specific advantages and disadvantages, FISH is intrinsically a low-throughput approach, and typically favoured to study specific loci and interactions, and to independently validate findings from proximity ligation methods (see [96] for a detailed discussion).

We focus below on chromosome conformation capture approaches, and their application in the cardiovascular area. A critical aspect in designing these experiments for the study of gene regulation is tissue-specificity, both for three-dimensional chromatin interactions and (especially) the activity of interacting regulatory elements. This factor underlies the importance of assaying the right cell type and developmental stage for meaningful analyses, as exemplified in the studies discussed in this section.

Some of the key genetic loci associated with cardiac disease have been analysed with chromosome conformation capture approaches. Natriuretic peptide A and B (encoded by the neighbouring *NPPA* and *NPPB* genes) become expressed in the atrial and ventricular myocardium during development, and are strongly induced in the ventricles upon cardiac stress and heart failure. Because of their paradigmatic roles in heart development and disease, their regulatory landscapes and three-dimensional organization were analysed with an adaptation of chromatin conformation capture (4C-seq) [97]—in which locus-specific interrogation of promoter–enhancer interactions is achieved by sequencing all DNA fragments interacting with a chosen genomic region. By choosing viewpoints on each side of the *NPPA/NPPB* locus, the authors showed three-dimensional interaction profiles that were largely maintained in the adult and fetal heart, as well as under cardiac stress induced by pressure overload. These interactions were even present in the liver, a tissue where the two genes are not expressed, suggesting the three-dimensional conformation of the locus is permissive, yet leads to gene expression specifically in the heart. Employing an elegant combination of the 4C-seq approach, H3K27ac ChIP-seq profiles and transgenic reporter assays, the study also found adult and fetal enhancers were largely distinct, with developmental enhancers being located more distally. However, and in line with results discussed in the previous section [91], stress-responsive enhancers for *NPPA/NPPB* were largely distinct from those regulating its developmental expression—thus pointing to specific regulatory mechanisms under cardiac stress. Similar locus-specific studies for the *SCN5A* and *KCNH2* cardiac arrhythmia loci have combined profiling of regulatory landscapes and three-dimensional interactions with experimental perturbations [88,98]—these will be discussed in the next section (see also figure 2*a* and *b*).

More recently, chromatin conformation capture experiments have been adapted to interrogate three-dimensional DNA interactions across the whole genome, an approach termed Hi-C. Such chromatin interaction maps are incredibly complex, and interactions can be selected for specific genomic regions such as gene promoters. In an impressive study aimed at annotating non-coding genetic variants associated with CVD, Montefiori *et al.* [89] used this promoter–capture adaptation of Hi-C to map genomic interactions with the promoters of approximately 22 000 human genes. By applying this approach to human-

iPSCs and derived cardiomyocytes (iPSC-CMs), this study provided a detailed comparison of three-dimensional genome architecture in a chief experimental system of heart muscle function. Their results revealed approximately 400 000 enhancer–promoter interactions in these cell types, over half of which were shared between iPSCs and their derived CMs. Leveraging matched H3K27ac and H3K4me1 data, the authors showed chromatin interactions specific to either iPSCs or CMs were enriched in regulatory elements with active epigenetic marks in the matched cell-type. Moreover, integration of these results with GWAS SNPs across a wide range of CVDs showed approximately 20% of LD SNPs located within distal genomic elements interacting with promoters of their likely target genes. In fact, more than 90% of SNP–target interactions did not involve the nearest gene, highlighting the importance of long-range regulation for functional interpretation of non-coding genetic variants. Of note, an earlier study employing an enhancer-centric Hi-C approach on human coronary artery smooth muscle cells (HCASMCs) [99] reached similar conclusions for GWAS variants associated with CAD—89% of SNP–promoter contacts in HCASMCs skipped at least one gene, and 64% were mapped to more than a single gene target. More recently, an integrative analysis of genetic variants associated with AF has also made use of epigenome, transcriptome and promoter–capture Hi-C data to score and prioritize genomic loci for functional validation [8].

### 3.3. Perturbation of non-coding regions by genome and epigenome editing

A classic limitation in functional genomics studies has been the reliance on correlative evidence to inform genome function. As such, the above approaches annotate putative regulatory elements based on their epigenetic annotation (§3.1), and their long-range interactions across the genome (§3.2). However, these strategies do not address the impact of candidate promoters and enhancers (and the genetic variants they harbour) in cellular and organismal functions. Thanks to major developments in genome-editing technologies in the last decade, it is now becoming feasible to functionally perturb non-coding genomic elements. These technologies have been extensively reviewed recently [100], and here we will focus on illustrative applications of genome-editing applied to non-coding elements in the cardiovascular area. A good fraction of such recent work employs the CRISPR/Cas9 technique, an approach based on the expression of an RNA-directed and sequence-specific bacterial nuclease, which can be employed for genetic or epigenetic editing of specific non-coding elements and genetic variants. However, classical genetics based on homologous recombination has also been successfully employed to target regulatory elements [101].

Recent work by Dickel *et al.* [102] provided a powerful example of genetic deletion of candidate genomic elements associated with CVD. Leveraging a range of epigenetic datasets in adult and fetal heart tissue from mouse and humans, the authors built a compendium of approximately 10 000 genomic elements with evidence of enhancer function in the heart [102]. By comparing these regions with current catalogues of human genetic variants associated with cardiac traits and diseases [5], Dickel *et al.* [102] found approximately 900 variants in high-scoring heart enhancers. In elegant proof-of-



principle experiments, genetic deletion of two predicted enhancers proximal to the sarcomeric genes *MYL2* and *MYH7* led to measurable cardiomyopathy phenotypes in the mouse, such as cardiomyocyte disarray and reduced cardiac output. Overall, these experiments provide compelling evidence for the functional impact of GWAS variants in cardiac disease phenotypes *in vivo*. More recently, a similar approach was taken to delete large genomic regions encompassing non-coding genetic variants associated with AF [8].

*In vitro* genetic editing in human models represents an exciting complementary approach to investigate the function of disease-associated variants and the genomic regions they occur in. For example, iPSC models of cardiomyocyte function can be used to interrogate non-coding regions associated with myocardial pathologies in which marker gene expression substantially differs between humans and mice [103]. A recent study focusing on the p.R453C-*MYH7* pathogenic mutation in HCM [104] highlights the usefulness of these models for detailed characterization of myocardial phenotypes associated with discrete genetic perturbations. By careful CRISPR/Cas9 engineering of this mutation across three independent iPSC lines (for wild-type, heterozygous and homozygous mutants), the authors demonstrate the dose-dependent contribution of this coding mutation to several hallmarks of HCM, such as increased cardiomyocyte size, sarcomeric disarray and arrhythmogenic alterations in calcium handling. This study illustrates the potential of human *in vitro* models for functional investigation of non-coding disease-associated variants and the regulatory elements they reside in—recent studies in this direction are discussed in the next section (see also figure 2c).

CRISPR/Cas9 approaches have also been adapted to modify the epigenome, although their application to cardiovascular loci has been limited to date. By combining a dead Cas9 protein with transcriptional activators or repressors, CRISPR activation and CRISPR inhibition (CRISPRi) approaches allow targeted perturbation of regulatory elements without direct genetic modification. A recent high-throughput study used CRISPRi for functional perturbation of regulatory epigenomes [105]. The study used an elegant approach with CRISPRi of a gene of interest in conjunction with FISH labelling of single cells (according to their RNA expression for the targeted gene). By flow-sorting perturbed cell populations based on the RNA FISH signal, the method allows the binning of cells according to their expression level for the gene of interest, which is associated with abundance of guide RNAs in that bin by high-throughput sequencing. The authors applied this strategy to the analyses of over 3000 enhancer-gene connections across 30 genes—including complex connections involving enhancers regulating up to five genes, and individual genes regulated by up to 14 distal elements. Consistent with the high degree of redundancy and robustness in mammalian regulatory landscapes, a relatively small 141 of the tested interactions between regulatory elements and genes led to significant effects on gene expression, which were of modest effect size on average (median 22%).

## 4. Mechanistic insights into the role of non-coding regulatory variation in cardiovascular disease

This section will focus on selected examples of the application of complementary experimental approaches (§3) to

the investigation of genetic regions associated with CVD, with a focus on non-coding variants and the larger promoter and enhancer elements they occur in. For completeness, we summarize selected and additional studies in table 1. Finally, we discuss the mechanistic insights these studies are starting to provide on the role of regulatory elements in diseases of the cardiovascular system.

### 4.1. Disruption of transcription factor binding sites

Genetic variants associated with CVD have long been postulated to result in altered binding of transcription factors, such as those directing cardiac development. Pioneering work by Smemo *et al.* [30] integrated functional genomics and targeted sequencing of CHD patients to identify a mutation within an enhancer element for the *TBX5* gene segregating with the disease. *In silico* analyses suggested the mutation alters a binding site for the TAL1 transcription factor, and disrupted the activity of the enhancer *in vitro* and in transgenic mouse and zebrafish models. Similarly, Kapoor *et al.* [116] carried out a detailed mapping of GWAS variants associated with LQTS in the *NOS1AP* locus, including integration with available functional genomics data. By interrogating 12 variants in reporter assays, the authors revealed a genetic variant in an enhancer element proximal to *NOS1AP*. This SNP clearly affected enhancer activity *in vitro*, although the identity of the bound protein remains unknown.

More recently, non-coding variants associated with AF in the *PITX2c* locus have been analysed using a diverse experimental toolkit [59] (see also figure 2c). This elegant study focused on variant annotation within the LD block most proximal to *PITX2c*, out of four showing statistical association in this locus. Through the integration of genomic data and reporter assays, the authors identified a potentially functional enhancer variant. Generation of a CRISPR/Cas9 model of the risk allele in hESC-derived CMs showed a 29% decrease in enhancer RNAs from this region, suggesting the risk allele disrupts enhancer activity. The authors combined *in silico* sequence analysis and *in vitro* assays to show the risk variant affects a binding site for the TFAP2A transcription factor which was confirmed through reduced *in vivo* binding of this protein to the risk allele in hESC-CMs. Moreover, downstream effects on *PITX2c* expression were analysed upon deletion of the associated SNP (30–54% reduction), as well as by comparing risk versus non-risk alleles (27% reduction). Lastly, knock-down experiments using interference RNA demonstrated the effects of the regulatory variant being dependent on the expression of TFAP2A and its cofactor CITED2. In a similar vein, a recent functional genomics approach focusing on the analysis of CAD variants in HCASMs after perturbation by pro- or de-differentiation growth factors [109] selected candidate genetic variants based on the integration of epigenetic profiles for chromatin accessibility, enhancer histone marks and transcription factor binding. The authors focused on a candidate intronic variant in the *SMAD3* locus predicted to disrupt a canonical AP1 motif. By combining allele-specific ChIP in HCASMs from individuals heterozygous for this SNP with overexpression and knock-down of AP1 members, this study provided convincing evidence for disruption of AP1 binding by the minor, protective allele—additionally suggesting that, compared to other candidate variants, the greater effect size of this SNP may be due its disruption of an AP1 motif. In sum, these works illustrate the integration of

**Table 1.** Summary of experimental studies of regulatory elements and trait-associated variants in cardiovascular disease. Abbreviations: ECs, endothelial cells; SMCs, smooth muscle cells; VSMCs, vascular smooth muscle cells; iPSCs, induced pluripotent stem cells; iPSC-CMs, cardiomyocytes derived from induced pluripotent stem cells; ESC-CMs, cardiomyocytes derived from embryonic stem cells; HCFs, human cardiac fibroblasts.

category	species	model	locus	variant(s)	experimental perturbation	experimental evidence	ref.
<b>coronary artery disease</b>							
non-coding interval	mouse	whole organism	CDKN2A/B	rs12555547 rs10757274 rs2383206 rs10965244	homozygous deletion	increased mortality reduced CDKN2A/B expression increased proliferation/reduced senescence in VSMCs	[101]
distal enhancer	human	LCLs HUVECs	CDKN2B-AS1	rs10811656 rs10757278	siRNA/natural genetic variation	STAT1 binding disrupted by risk haplotype long-range interaction between enhancer and CDKN2A/B locus STAT1 inhibits CDKN2BAS expression	[106]
non-coding interval	human	VSMCs	CDKN2A/B CDKN2B-AS1	rs1333049	natural genetic variation	risk genotype associated with reduced gene expression and increased proliferation	[107]
intronic interval	human	coronary artery ECs	PHACTR1	rs9349379	heterozygous deletion	MEF2 binding disrupted by risk allele	[108]
intronic interval	human	coronary artery SMCs	SMAD3	rs17293632	natural genetic variation	34 bp deletion leads to 35% reduction in PHACTR1 expression SNP in open chromatin region, induced in TGF $\beta$ - or PDGF-BB-treated cells	[109]
distal enhancer	human	iPSC-ECs and VSMCs	PHACTR1	rs9349379	90 bp enhancer deletion homozygous major/minor biallelic clones	AP-1/TCF21 binding disrupted by protective minor allele enhancer deletion increased EDN1 expression	[110]
intronic enhancer	human	coronary artery ECs	PLPP3	rs17114036	enhancer deletion CRISPR inhibition	minor risk allele associates with 20% increase in EDN1 expression SNP in endothelial enhancer induced by haemodynamics enhancer deletion/repression leads to reduced PLPP3 expression	[111]
non-coding interval	human	iPSC-VSMCs	CDKN2B-AS1	risk/non-risk haplotype	60 kb deletion of risk and non-risk haplotypes	KLZF2 binding disrupted by risk allele deletion of risk haplotype leads to non-risk like expression profile in iPSC-VSMCs	[112]
<b>congenital heart disease</b>							
distal enhancer	mouse	LoxZ transgene	TBX5	rare variant	transgene expression	rare variant disrupts heart-specific expression and nuclear protein binding <i>in vitro</i>	[30]
intronic enhancer	human rat	heart tissue HEK293 cells H9C2 cells	MTRR	rs326119	natural genetic variation	minor allele disrupts CEBPA binding and associates with elevated plasma homocysteine levels	[113]
distal enhancer	human	normal iPSCs	GATA4	rs6601627 rs118065347	GATA4 small deletion by CRISPR	non-coding variants genetically associated with aortic bicuspid valve GATA4 deficiency impairs EMT in iPSCs	[114]
missense variant	mouse	knock-in mouse model	PBX1	<i>de novo</i> mutation	CRISPR genetic editing	multiple developmental defects in knock-in mice, including ventricular septal defects and persistent truncus arteriosus	[115]

(Continued.)

Table 1. (Continued.)

category	species	model	locus	variant(s)	experimental perturbation	experimental evidence	ref.
<b>long QT syndrome</b>							
distal enhancer	human	HL-1 cells	<i>MOST1AP</i>	rs7539120	<i>in vitro</i> deletion	variant deletion reduces <i>in vitro</i> enhancer activity	[116]
	rat	zebrafish embryo			<i>NOST1AP</i> overexpression	risk haplotype increases transgenic reporter activity and nuclear protein binding <i>in vitro</i>	
	zebrafish	RNVMS					
missense variant	human	patient-specific iPSCs	<i>KCNH2</i>	novel missense variant	correction of variant in patient cells/ homozygous knock-in in healthy cells	prolongation of action potential in patient iPSC-CMs and reduced <i>I<sub>Kr</sub></i> normalized phenotype in corrected cells	[117]
promoter region	human	human left ventricle	<i>PRKCA</i>	rs9892651	natural genetic variation	minor protective allele associated with decreased PRKCA expression	[118]
		HEK293 cells		rs9909004		insert containing rs9909004 shows enhancer activity in HEK293 cells	
				rs7210446			
<b>atrial fibrillation</b>							
distal enhancer	human	ESC-CMs	<i>PITX2c</i>	rs2595104	knock-in of risk allele	30–54% reduced <i>PITX2c</i> expression upon SNP deletion; 27% reduction by risk allele	[59]
					approximately 100 bp enhancer deletion		
					siRNA knock-down	effect dependent on TFAP2a/CITED2 preferential binding to non-risk allele	
distal enhancer	mouse	whole organism	<i>GJA1</i>	many in each	40 kb homozygous deletion	reduced <i>GJA1</i> expression in atria	[8]
promoter/exon				genomic interval	85 kb homozygous deletion	loss of <i>KCNH3</i> expression in atria and ventricles	
					33 kb homozygous deletion	no significant expression changes	
distal enhancer	human	HL-1 cells	<i>PPRX1</i>	rs577676	<i>PPRX1</i> gene knock-out	interaction between enhancer and <i>PPRX1</i> promoter in cardiac fibroblasts	[119]
		HCFs				variant required for enhancer activity	
		iPSC-CMs				<i>PPRX1</i> suppression leads to shortening of atrial action potential	
<b>Brugada syndrome</b>							
distal enhancer	mouse	cardiac tissue	<i>SCN10A</i>	rs6801957	BAC transgenic mouse	genetic variant interacts with <i>SCN5A</i> promoter	[88]
	human					engineered enhancer deletion essential for cardiac <i>SCN5A</i> expression	
						risk allele associated with reduced enhancer activity	
3' UTR	human	blood samples	<i>SCN5A</i>	rs45592631	natural genetic variation	3'UTR variants identified in Tunisian family	[120]
						effect hypothesized to be dependent on miR-1270	
<b>cardiomyopathies</b>							
distal enhancers	mouse	whole organism	<i>MYL2</i>	proximal to coding	homozygous deletions	cardiac abnormalities upon deletion of both enhancers (reduced fractional shortening, cardiomegaly and cardiomyocyte disarray)	[102]
				mutations			

(Continued.)

**Table 1.** (Continued.)

category	species	model	locus	variant(s)	experimental perturbation	experimental evidence	ref.
missense variant	human	patient-specific iPSCs	<i>SCN5A</i>	R1898H	genetic editing to correct mutation	36% reduction in peak sodium current in patient iPSC-CMs	[121]
missense variants	human	patient-specific iPSCs	<i>MYH7</i>	R453C	heterozygous and homozygous genetic editing	dose-dependent cardiomyopathy phenotypes (size, sarcomeric disarray, altered calcium handling)	[104]
promoter	human	IHKE cells	<i>DSG2</i>	rare variant	risk and non-risk alleles <i>in vitro</i>	reduced <i>DSG2</i> transcriptional activity associated with minor allele preferential AP-1 binding to non-risk allele	[122]
<b>obesity</b>							
distal enhancer	human	brain tissue	<i>FTO/IRX3</i>	rs9930506	<i>IRX3</i> knock-out	robust interaction between intronic <i>FTO</i> variants and <i>IRX3</i> promoter in brain tissue	[123]
	mouse	whole organism				<i>IRX3</i> -deficient mice show phenotypes consistent with role in control of body mass	

complementary strategies to investigate non-coding genetic variants, and provide compelling mechanistic evidence for their role in disrupting the tissue-specific binding of transcription factors and downstream gene expression (see also table 1 for additional examples).

## 4.2. Long-range gene regulation

As discussed in previous sections, a key challenge in functionally interrogating non-coding genetic variants associated with human disease is the ability of regulatory elements to target genes at very long distances. Although this inherent property of transcriptional enhancers complicates assigning non-coding variants to their likely effector genes, several recent studies provide mechanistic examples for long-range gene regulation mediated by regulatory elements harbouring genetic variants associated with specific forms of CVD, or cardiovascular risk factors such as obesity.

The *SCN5A/SCN10A* locus is genetically associated with congenital heart disease and cardiac arrhythmias such as Brugada syndrome. Van den Boogaard *et al.* [88] reported the first detailed analysis of non-coding variants in this cardiac disease locus, through a combination of functional genomics approaches and reporter assays. Employing the 4C assay for targeted interrogation of three-dimensional chromatin interactions (see §3.2), the authors showed that the most likely functional variant in this locus—which had been ascribed to the *SCN10A* gene based on genomic distance—in fact physically contacted the *SCN5A* promoter, and was associated with allele-specific *SCN5A* expression in humans. In line with studies discussed in the previous section, the risk allele was predicted to disrupt a T-box binding motif, as suggested by *in vitro* reporter assays. Using a similar approach, a recent study characterized variants associated with AF in the *PRRX1* locus [119]. Combining interaction frequency data from published Hi-C studies and locus-specific validation in cardiac fibroblasts, the authors demonstrated a physical interaction between an enhancer element harbouring a disease-associated variant and the *PRRX1* promoter. Moreover, suppression of *PRRX1* expression led to shortening of the atrial action potential in iPSC-derived cardiomyocytes—an electrophysiological hallmark of AF.

Obesity is a major risk factor for CVD, and the genetic basis for its heritability has been the focus of several association studies [124,125]. One of the lead genomic signals associated with obesity in humans corresponds to intronic variants within the *FTO* locus, and follow-up studies have suggested this gene is implicated in the control of body mass [126]. Smemo *et al.* [123] employed a series of elegant experiments to revisit the molecular genetics of this locus. Using 4C-seq to map genomic interactions between the *FTO* intronic region and surrounding promoters, they showed a lack of interaction with the *FTO* promoter in the adult mouse brain. Instead, the authors detected robust interactions between the obesity-associated interval and the *IRX3* gene (located half a megabase downstream). Moreover, obesity-associated SNPs correlated with *IRX3* expression in the human brain, but not with that of *FTO*—suggesting that the association signal overlaps an enhancer element involved in long-range regulation of *IRX3* in the brain. To support this conclusion, the authors performed detailed phenotyping of mice deficient for *IRX3*, which support a global and hypothalamus-specific role for this gene in control of body mass. Although these results are also

compatible with the proposed role of FTO in modulating obesity, they powerfully highlight the importance of long-range gene regulation as a mechanism mediating the effect of genetic variants associated with CVD.

### 4.3. Gene regulation by long non-coding RNAs

A key finding from genomic consortia such as the ENCODE project [81] has been the pervasive transcription of the human genome. Current estimates suggest that up to 80% of our genome may be actively transcribed, including a large variety of non-coding RNA species such as microRNAs, enhancer RNAs and long non-coding RNAs (lncRNAs). In this section, we focus on the emerging roles of lncRNAs in cardiovascular biology (reviewed in [127]), and on recent examples of CVD-associated genetic variants that have been mechanistically linked to lncRNAs.

The potential relevance of lncRNAs on cardiac development was highlighted by the discovery of *FENDRR*, a lncRNA highly expressed in the posterior mesoderm, and whose genetic deletion leads to impaired heart development [128]. Recent work has also implicated lncRNAs in the gene expression response to cardiac stress [129,130]. Focusing on the transcriptional response to pressure overload in mice, Wang *et al.* [131] recently reported and functionally characterized the cardiomyocyte-specific lncRNA *CHAER1*. Employing knock-out mice and perturbation experiments *in vitro*, the authors proposed this lncRNA is necessary and sufficient for cardiomyocyte hypertrophy, which suggested its role in regulating gene expression under cardiac stress. Functional genomics experiments in normal and *CHAER1*-deficient cardiomyocytes indicated negative regulation of H3K27 methylation by *CHAER1*, and further *in vitro* work showed this effect may be mechanistically due to interaction between *CHAER1* and the chromatin repressive protein PRC2—which appears to be transiently enhanced early in response to hypertrophic signals, and precedes a decrease in H3K27me3 and induction of hypertrophic gene expression.

One of the first indications of a putative role of lncRNAs in heart disease came from genetic studies identifying a susceptibility locus for MI [132], encoding a lncRNA termed MI associated transcript (*MIAT*). However, transcriptomic studies of mouse brain development suggest *MIAT* may have pleiotropic effects on gene expression, including on cardiovascular-relevant pathways such as WNT signalling [133]. In addition, this lncRNA has also been linked to fibrosis in HCM and disease prognosis [134]. More recently, detailed experimental annotation of the *KNCH2* locus (associated with QT interval duration in humans) identified regulatory elements in the mouse controlling expression of its two isoforms *KCNH2A* and *B*, one of which controls a lncRNA likely involved in *KCNH2* regulation [98]. Lastly, genetic association studies of CAD independently identified the 9p21 susceptibility locus, which is adjacent to the lncRNA *CDKN2B-AS1* (also known as ANRIL), and hundreds of kilobases from the protein-coding genes *CDKN2A* and *CDKN2B*. Although the biological function of ANRIL and the nearby disease-associated variants is incompletely understood, recent studies have provided evidence for its role in CVD [101,106,107,112]—mechanistic insights into such complex genetic loci will be the focus of the next section.

### 4.4. Mechanistic insights on complex genetic loci

The variety of genetic, epigenomic and molecular approaches currently available for the study of non-coding genetic variants is beginning to allow very detailed investigation of complex genetic loci, such as those harbouring multiple susceptibility variants, or associated with diverse cardiovascular phenotypes. In this section, we discuss two recent studies that exemplify the power of integrative experimental approaches to provide mechanistic insights into complex disease genetics.

Gupta and colleagues [110] recently conducted a detailed study of the 6p24 cardiovascular risk locus, overlapping the *PHACTR1* gene and associated with susceptibility to several CVDs, such as CAD, migraine headache and hypertension. Using imputation-based fine mapping and probabilistic estimation of causal SNPs in the locus, the authors focused on a lead variant showing association of its minor allele with increased risk for CAD, but decreased risk of migraine headache or hypertension—suggesting a role in endothelial function. The study next employed available epigenomic datasets to show the likely causal SNP overlaps an enhancer element with strong H3K27ac signal specifically in aortic tissue. Using an *in vitro* system of iPSC-derived endothelial and vascular smooth muscle cells, the authors conducted elegant gene editing experiments for biallelic deletion of an approximately 90 bp region containing the enhancer variant, which led to higher expression of the distal endothelin-1 (*EDN1*) gene. These results were combined with similar experiments comparing iPSC clones edited to harbour homozygous major (A/A) and homozygous minor (G/G) versions of the lead SNP. Gene expression profiling of these clones showed over 200 differentially expressed genes, which included known regulators of vascular function (e.g. *COL4A1* or *FN1*). Among differentially expressed genes, *EDN1* was the closest to the lead variant, with the minor allele driving an approximately 20% increase in *EDN1* expression. Although clearly consistent with the enhancer variant acting through the regulation of *EDN1*, chromatin-conformation capture experiments failed to show a robust interaction between this variant and the *EDN1* promoter.

Based on these observations, the authors proposed a model where the differing associations of the 6p24 locus with risk for CVDs can be explained by context-dependent effects of *EDN1* on the vasculature. First, increased endothelial production of *EDN1* (associated with the risk allele) and binding to the ET<sub>A</sub> receptor can promote atherosclerosis. Second, increased *EDN1* binding to the same receptor in VSMCs leads to vasoconstriction (and reduced susceptibility to migraine). Lastly, the association of the minor allele with lower systolic blood pressure may be due to *EDN1* binding to the ET<sub>B</sub> receptor in the endothelial cells of large systemic arteries, leading to systemic vasodilation. Overall, these results show the potential of integrating genetic, epigenetic and gene editing approaches for functional investigation of risk variants associated with multiple diseases and phenotypes.

In a similar vein, a recent study focused on the analysis of the 9p21.3 CVD locus [112] (§4.3), an approximately 60 kb region associated with CAD risk, and containing over a hundred genetic variants [11] that commonly occur in one of two haplotypes ('risk' and 'non-risk'). The authors used an elegant set of experiments to functionally analyse the aggregate effect of variants in each haplotype. First, the authors compared previously generated iPSCs homozygous for the non-risk

(NN) and risk haplotypes (RR) with homozygous deletion clones of either haplotype. Although these deletions did not seem to impact iPSC function, differentiation of edited clones towards VSMCs showed faster proliferation of mesoderm precursors in RR cells. Deletion of the risk haplotype reverted proliferation rates to those of NN mesodermal progenitors, and resulted in differential expression of over a hundred genes. In fully differentiated VSMCs this effect was amplified, with deletion of the risk haplotype associating with thousands of differentially expressed genes. Deletion of the risk haplotype also seemed to restore VSMCs to a gene expression pattern resembling that of the non-risk haplotype, and accordingly measures of VSMC function (such as contractile force) were reduced in cells harbouring the risk haplotype, and restored to non-risk levels upon deletion of the 60 kb region.

To inform the mechanistic basis of the gene expression and functional changes associated with the risk haplotype, the authors focused on the complex splicing of the *CDKN2B-AS1* lncRNA (the terminal exons of which are contained in the risk interval). Short isoforms were specifically detected in VSMCs harbouring the risk haplotype, and their overexpression in non-risk VSMCs induced morphological and functional changes resembling those in RR VSMCs, such as reduced contractile force. However, gene expression changes upon expression of short *CDKN2B-AS1* isoforms did not recapitulate most of those associated with the risk haplotype—suggesting that the functional effects of risk variants in this locus are not solely mediated by altered expression of *CDKN2B-AS1*. In sum, this study provides a powerful example of combining genetic editing and phenotyping approaches for the detailed characterization of complex disease-associated haplotypes.

## 5. Clinical significance of cardiovascular non-coding variants

The results of cardiovascular GWAS studies (§2) have undoubtedly succeeded in characterizing the genetic architecture of CVDs—especially for prevalent conditions such as CAD and AF, for which large numbers of patients and controls have been analysed. Nevertheless, in terms of clinical applicability, the ultimate goal is to understand cardiovascular pathophysiological mechanisms, establish improved risk prediction methods and develop effective therapies. Functional annotation of GWAS findings (§3) and mechanistic studies of individual loci (§4) have contributed important insights into pathophysiological pathways, yet many challenges remain in leveraging cardiovascular non-coding variation for clinical application.

One area of substantial development is the use of non-coding GWAS signals in risk prediction. Most CVDs are polygenic, and perturbation experiments of individual non-coding variants are consistent with small individual effects on disease susceptibility—which can be aggregated in genetic scores predictive of individual disease risk or treatment response. This strategy has been recently applied to common polygenic diseases, and suggested polygenic risk scores can identify individuals at a similar risk than that conferred by rare monogenic mutations [16]. Because genetic risk scores can be assessed at birth, they may be particularly useful in risk prediction among younger patients with a reduced cumulative impact of lifestyle factors [135]. This caveat may explain

why genetic risk scores only modestly improve the cardiology-based prognosis for diseases such as AF [63]. Nevertheless, increased knowledge of the mechanistic significance of GWAS loci will be key in improving risk prediction from non-coding cardiovascular variants.

A second challenge is the substantial missing heritability in current studies. A good example not covered above is blood pressure GWAS (reviewed in [2]). Familial studies of blood pressure estimate up to 50% heritability, yet the collective effect of all GWAS loci only explains approximately 10%. Significant gene–gene and gene–environment interactions, structural variation, rare variants and epigenetic mechanisms likely contribute to this missing heritability. In this regard, resources such as the UK Biobank [136] are instrumental in the development of experimental designs and analytical approaches to systematically evaluate the interaction between disease genetics and the environment. It is also possible that substantial missing heritability remains hidden in GWAS datasets as sub-threshold variants, corresponding to context-dependent risk effects that may be dependent on environmental influences found in particular stages of the disease.

Lastly, the ultimate challenge of CVD genetics is the development of novel treatment strategies based on genetically supported targets [137]. Although much progress is needed in mechanistic understanding of non-coding GWAS variants, these already point to novel therapeutic pathways. First, non-coding variants influence traditional risk factors, such as lipid metabolism, that may be treatment targets themselves. An example is the *PCSK9* CAD risk locus. *PCSK9* functions in LDL cholesterol metabolism and several compounds are in development to inhibit its effect in the circulation. Second, non-coding GWAS variants are also key in suggesting disease-driving molecular processes that could be exploited therapeutically, such as an altered atrial gene regulatory network in AF. Lastly, non-coding genetic variation can also be used to develop individualized treatment strategies. Although the success of genetic testing for patient stratification has been limited to date, individuals with high genetic risk scores appear to have distinct therapeutic responses, such as a larger benefit from statin treatment [138]. Nevertheless, the clinical application of polygenic risk scores for therapeutic targeting is currently hampered by relatively low precision [139].

## 6. Conclusion and future perspectives

Facilitated by key technological developments in sequencing technology, genetic editing and integrative analyses, CVD genetics has made tremendous progress in identifying and prioritizing human genomic intervals associated with disease predisposition. Genetic association studies now involve global efforts to comprehensively characterize putative susceptibility loci. Critically, complementary experimental approaches allow very detailed epigenetic annotation of disease-associated loci, and are providing important insights into the genomic and molecular complexity of gene regulatory programmes in disease-relevant contexts. Lastly, integrative strategies leveraging multi-level approaches are beginning to inform the role of genetic variants and risk haplotypes in cellular and complex phenotypes.

This progress notwithstanding, a number of key challenges remain. First, mechanistic insights for the vast majority of loci associated with CVD are yet to be determined, posing a

daunting task that calls for novel strategies to prioritize loci for functional testing. Second, many disease-associated regions harbour complex haplotypes with dozens to hundreds of variants, and determining the contribution of individual SNPs in complex genetic loci will likely require new experimental approaches. Third, most CVDs with a genetic component also occur in the context of multiple environmental factors, such as metabolic disease and chronic cardiac stress. Untangling the contributions of genetic and environmental cues to complex CVD constitutes a primary challenge in the field that are likely to require improved integration of genetic findings and experimental modelling of environmental risk factors.

A number of emerging approaches and models have begun to address these. Data integration strategies incorporating multiple experimental datasets and genetic GWAS findings are increasingly being used to rank non-coding genetic variants for functional experiments [8,102]. Further development of *in vitro* modelling of subtype-specific CVD [140] will be needed to fully realize the potential of human ‘disease in a dish’

studies. The complex nature of many CVDs is also calling for improved strategies to evaluate tissue-specific contributions experimentally, such as the development of engineered human tissues [141]. Lastly, novel high-throughput approaches to screen gene expression contributions [105] and tissue-specificity [142] of regulatory elements (and the non-coding variants within) should be pivotal in gleaning mechanistic insights into how gene regulation contributes to onset and progression of CVDs.

**Data accessibility.** This article has no additional data.

**Authors’ contributions.** D.V. designed and wrote the review, with contributions from S.F., P.D. and A.T. All authors reviewed and approved the final manuscript.

**Competing interests.** We declare we have no competing interests.

**Acknowledgements.** This work has been funded by the British Heart Foundation (Fellowship FS/18/39/33684 to D.V.) and Queen Mary University of London. The authors apologize to colleagues whose work could not be covered due to space limitations.

## References

- Adriaens ME, Bezzina CR. 2018 Genomic approaches for the elucidation of genes and gene networks underlying cardiovascular traits. *Biophys. Rev.* **10**, 1053–1060. (doi:10.1007/s12551-018-0435-2)
- Ng FL, Warren HR, Caulfield MJ. 2018 Hypertension genomics and cardiovascular prevention. *Ann. Transl. Med.* **6**, 291. (doi:10.21037/atm.2018.06.34)
- Schunkert H, von Scheidt M, Kessler T, Stiller B, Zeng L, Vilne B. 2018 Genetics of coronary artery disease in the light of genome-wide association studies. *Clin. Res. Cardiol.* **107**, 2–9. (doi:10.1007/s00392-018-1324-1)
- Maurano MT *et al.* 2012 Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195. (doi:10.1126/science.1222794)
- Buniello A *et al.* 2019 The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res.* **47**, D1005–D1012. (doi:10.1093/nar/gky1120)
- Nelson CP *et al.* 2017 Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat. Genet.* **49**, 1385–1391. (doi:10.1038/ng.3913)
- Wang X *et al.* 2016 Discovery and validation of sub-threshold genome-wide association study loci using epigenomic signatures. *Elife* **5**, e10557. (doi:10.7554/eLife.10557)
- van Ouwkerk AF *et al.* 2019 Identification of atrial fibrillation associated genes and functional non-coding variants. *Nat. Commun.* **10**, 4755. (doi:10.1038/s41467-019-12721-5)
- Zdravkovic S, Wienke A, Pedersen NL, Marenberg ME, Yashin AI, De Faire U. 2002 Heritability of death from coronary heart disease: a 36-year follow-up of 20 966 Swedish twins. *J. Intern. Med.* **252**, 247–254. (doi:10.1046/j.1365-2796.2002.01029.x)
- Erdmann J, Kessler T, Munoz Venegas L, Schunkert H. 2018 A decade of genome-wide association studies for coronary artery disease: the challenges ahead. *Cardiovasc. Res.* **114**, 1241–1257. (doi:10.1093/cvr/cvy084)
- Nikpay M *et al.* 2015 A comprehensive 1,000 genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat. Genet.* **47**, 1121–1130. (doi:10.1038/ng.3396)
- Klarin D *et al.* 2017 Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease. *Nat. Genet.* **49**, 1392–1397. (doi:10.1038/ng.3914)
- Musunuru K *et al.* 2010 From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* **466**, 714–719. (doi:10.1038/nature09266)
- Khera AV *et al.* 2017 Association of rare and common variation in the lipoprotein lipase gene with coronary artery disease. *J. Am. Med. Assoc.* **317**, 937–946. (doi:10.1001/jama.2017.0972)
- Kessler T *et al.* 2017 Functional characterization of the GUCY1A3 coronary artery disease risk locus. *Circulation* **136**, 476–489. (doi:10.1161/CIRCULATIONAHA.116.024152)
- Khera AV *et al.* 2018 Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224. (doi:10.1038/s41588-018-0183-z)
- Abul-Husn NS *et al.* 2016 Genetic identification of familial hypercholesterolemia within a single U.S. health care system. *Science* **354**, aaf7000. (doi:10.1126/science.aaf7000)
- van der Bom T, Zomer AC, Zwiderman AH, Meijboom FJ, Bouma BJ, Mulder BJ. 2011 The changing epidemiology of congenital heart disease. *Nat. Rev. Cardiol.* **8**, 50–60. (doi:10.1038/nrcardio.2010.166)
- Lana-Elola E, Watson-Scales S, Slender A, Gibbins D, Martineau A, Douglas C, Mohun T, Fisher EM, Tybulewicz V. 2016 Genetic dissection of Down syndrome-associated congenital heart defects using a new mouse mapping panel. *Elife* **5**, e11614. (doi:10.7554/eLife.11614)
- Zaidi S, Brueckner M. 2017 Genetics and genomics of congenital heart disease. *Circ. Res.* **120**, 923–940. (doi:10.1161/CIRCRESAHA.116.309140)
- Arrington CB *et al.* 2012 Exome analysis of a family with pleiotropic congenital heart disease. *Circ. Cardiovasc. Genet.* **5**, 175–182. (doi:10.1161/CIRCGENETICS.111.961797)
- Martin LJ, Piliipenko V, Kaufman KM, Cripe L, Kottyan LC, Keddache M, Dexheimer P, Weirauch MT, Benson DW. 2014 Whole exome sequencing for familial bicuspid aortic valve identifies putative variants. *Circ. Cardiovasc. Genet.* **7**, 677–683. (doi:10.1161/CIRCGENETICS.114.000526)
- Zaidi S *et al.* 2013 De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223. (doi:10.1038/nature12141)
- Homsy J *et al.* 2015 De novo mutations in congenital heart disease with neurodevelopmental and other congenital anomalies. *Science* **350**, 1262–1266. (doi:10.1126/science.aac9396)
- Al Turki S *et al.* 2014 Rare variants in NR2F2 cause congenital heart defects in humans. *Am. J. Hum. Genet.* **94**, 574–585. (doi:10.1016/j.ajhg.2014.03.007)
- Jin SC *et al.* 2017 Contribution of rare inherited and de novo variants in 2871 congenital heart disease probands. *Nat. Genet.* **49**, 1593–1601. (doi:10.1038/ng.3970)
- Watkins WS *et al.* 2019 De novo and recessive forms of congenital heart disease have distinct genetic and phenotypic landscapes. *Nat. Commun.* **10**, 4722. (doi:10.1038/s41467-019-12582-y)
- Cordell HJ *et al.* 2013 Genome-wide association study identifies loci on 12q24 and 13q32 associated

- with tetralogy of Fallot. *Hum. Mol. Genet.* **22**, 1473–1481. (doi:10.1093/hmg/dd552)
29. Cordell HJ *et al.* 2013 Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat. Genet.* **45**, 822–824. (doi:10.1038/ng.2637)
  30. Smemo S, Campos LC, Moskowitz IP, Krieger JE, Pereira AC, Nobrega MA. 2012 Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21**, 3255–3263. (doi:10.1093/hmg/dd5165)
  31. Gray B, Behr ER. 2016 New insights into the genetic basis of inherited arrhythmia syndromes. *Circ. Cardiovasc. Genet.* **9**, 569–577. (doi:10.1161/CIRCGENETICS.116.001571)
  32. Keating M, Atkinson D, Dunn C, Timothy K, Vincent GM, Leppert M. 1991 Linkage of a cardiac arrhythmia, the long QT syndrome, and the Harvey ras-1 gene. *Science* **252**, 704–706. (doi:10.1126/science.1673802)
  33. Harmer SC, Tinker A. 2016 The impact of recent advances in genetics in understanding disease mechanisms underlying the long QT syndromes. *Biol. Chem.* **397**, 679–693. (doi:10.1515/hsz-2015-0306)
  34. Head CE, Balasubramanian R, Thomas G, Goddard CA, Lei M, Colledge WH, Grace AA, Huang CL. 2005 Paced electrogram fractionation analysis of arrhythmogenic tendency in DeltaKQP Scn5a mice. *J. Cardiovasc. Electrophysiol.* **16**, 1329–1340. (doi:10.1111/j.1540-8167.2005.00200.x)
  35. Moretti A *et al.* 2010 Patient-specific induced pluripotent stem-cell models for long-QT syndrome. *N. Engl. J. Med.* **363**, 1397–1409. (doi:10.1056/NEJMoa0908679)
  36. Arking DE *et al.* 2006 A common genetic variant in the NOS1 regulator NOS1AP modulates cardiac repolarization. *Nat. Genet.* **38**, 644–651. (doi:10.1038/ng1790)
  37. Chang KC, Barth AS, Sasano T, Kizana E, Kashiwakura Y, Zhang Y, Foster DB, Marban E. 2008 CAPON modulates cardiac repolarization via neuronal nitric oxide synthase signaling in the heart. *Proc. Natl Acad. Sci. USA* **105**, 4477–4482. (doi:10.1073/pnas.0709118105)
  38. van den Berg ME *et al.* 2017 Discovery of novel heart rate-associated loci using the Exome Chip. *Hum. Mol. Genet.* **26**, 2346–2363. (doi:10.1093/hmg/ddx113)
  39. van Setten J *et al.* 2019 Genome-wide association meta-analysis of 30,000 samples identifies seven novel loci for quantitative ECG traits. *Eur. J. Hum. Genet.* **27**, 952–962. (doi:10.1038/s41431-018-0295-z)
  40. Smith JG *et al.* 2011 Genome-wide association studies of the PR interval in African Americans. *PLoS Genet.* **7**, e1001304. (doi:10.1371/journal.pgen.1001304)
  41. Arking DE *et al.* 2014 Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization. *Nat. Genet.* **46**, 826–836. (doi:10.1038/ng.3014)
  42. Priori SG, Napolitano C, Schwartz PJ. 1999 Low penetrance in the long-QT syndrome: clinical impact. *Circulation* **99**, 529–533. (doi:10.1161/01.CIR.99.4.529)
  43. Schwartz PJ, Crotti L, George Jr AL. 2018 Modifier genes for sudden cardiac death. *Eur. Heart. J.* **39**, 3925–3931. (doi:10.1093/eurheartj/ehy502)
  44. Crotti L, Monti MC, Insolia R, Peljto A, Goosen A, Brink PA, Greenberg DA, Schwartz PJ, George Jr AL. 2009 NOS1AP is a genetic modifier of the long-QT syndrome. *Circulation* **120**, 1657–1663. (doi:10.1161/CIRCULATIONAHA.109.879643)
  45. Jamshidi Y *et al.* 2012 Common variation in the NOS1AP gene is associated with drug-induced QT prolongation and ventricular arrhythmia. *J. Am. Coll. Cardiol.* **60**, 841–850. (doi:10.1016/j.jacc.2012.03.031)
  46. Duchatelet S *et al.* 2013 Identification of a KCNQ1 polymorphism acting as a protective modifier against arrhythmic risk in long-QT syndrome. *Circ. Cardiovasc. Genet.* **6**, 354–361. (doi:10.1161/CIRCGENETICS.113.000023)
  47. Chai S, Wan X, Ramirez-Navarro A, Tesar PJ, Kaufman ES, Ficker E, George Jr AL, Deschenes I. 2018 Physiological genomics identifies genetic modifiers of long QT syndrome type 2 severity. *J. Clin. Investig.* **128**, 1043–1056. (doi:10.1172/JCI94996)
  48. Strauss DG *et al.* 2017 Common genetic variant risk score is associated with drug-induced QT prolongation and torsades de pointes risk: a pilot study. *Circulation* **135**, 1300–1310. (doi:10.1161/CIRCULATIONAHA.116.023980)
  49. Christophersen IE, Ravn LS, Budtz-Joergensen E, Skytthe A, Haunsoe S, Svendsen JH, Christensen K. 2009 Familial aggregation of atrial fibrillation: a study in Danish twins. *Circ. Arrhythm. Electrophysiol.* **2**, 378–383. (doi:10.1161/CIRCEP.108.786665)
  50. Christophersen IE, Ellinor PT. 2016 Genetics of atrial fibrillation: from families to genomes. *J. Hum. Genet.* **61**, 61–70. (doi:10.1038/jhg.2015.44)
  51. Lubitz SA *et al.* 2016 Whole exome sequencing in atrial fibrillation. *PLoS Genet.* **12**, e1006284. (doi:10.1371/journal.pgen.1006284)
  52. Gudbjartsson DF *et al.* 2007 Variants conferring risk of atrial fibrillation on chromosome 4q25. *Nature* **448**, 353–357. (doi:10.1038/nature06007)
  53. Ellinor PT *et al.* 2012 Meta-analysis identifies six new susceptibility loci for atrial fibrillation. *Nat. Genet.* **44**, 670–675. (doi:10.1038/ng.2261)
  54. Christophersen IE *et al.* 2017 Large-scale analyses of common and rare variants identify 12 new loci associated with atrial fibrillation. *Nat. Genet.* **49**, 946–952. (doi:10.1038/ng.3843)
  55. Nielsen JB *et al.* 2018 Biobank-driven genomic discovery yields new insight into atrial fibrillation biology. *Nat. Genet.* **50**, 1234–1239. (doi:10.1038/s41588-018-0171-3)
  56. Mann SA *et al.* 2012 Epistatic effects of potassium channel variation on cardiac repolarization and atrial fibrillation risk. *J. Am. Coll. Cardiol.* **59**, 1017–1025. (doi:10.1016/j.jacc.2011.11.039)
  57. Tao Y, Zhang M, Li L, Bai Y, Zhou Y, Moon AM, Kaminski HJ, Martin JF. 2014 Pitx2, an atrial fibrillation predisposition gene, directly regulates ion transport and intercalated disc genes. *Circ. Cardiovasc. Genet.* **7**, 23–32. (doi:10.1161/CIRCGENETICS.113.000259)
  58. Boldt LH *et al.* 2010 Mutational analysis of the PITX2 and NKX2-5 genes in patients with idiopathic atrial fibrillation. *Int. J. Cardiol.* **145**, 316–317. (doi:10.1016/j.ijcard.2009.11.023)
  59. Ye J, Tucker NR, Weng LC, Claus S, Lubitz SA, Ellinor PT. 2016 A functional variant associated with atrial fibrillation regulates PITX2c expression through TFAP2a. *Am. J. Hum. Genet.* **99**, 1281–1291. (doi:10.1016/j.ajhg.2016.10.001)
  60. Gore-Panter SR *et al.* 2014 Atrial fibrillation associated chromosome 4q25 variants are not associated with PITX2c expression in human adult left atrial appendages. *PLoS ONE* **9**, e86245. (doi:10.1371/journal.pone.0086245)
  61. Mikhailov AT, Torrado M. 2018 Interplay between cardiac transcription factors and non-coding RNAs in predisposing to atrial fibrillation. *J. Mol. Med. (Berl)* **96**, 601–610. (doi:10.1007/s00109-018-1647-4)
  62. Canon S *et al.* 2016 miR-208b upregulation interferes with calcium handling in HL-1 atrial myocytes: implications in human chronic atrial fibrillation. *J. Mol. Cell. Cardiol.* **99**, 162–173. (doi:10.1016/j.yjmcc.2016.08.012)
  63. Lubitz SA *et al.* 2017 Genetic risk prediction of atrial fibrillation. *Circulation* **135**, 1311–1320. (doi:10.1161/CIRCULATIONAHA.116.024143)
  64. Shah S *et al.* 2020 Genome-wide association and Mendelian randomisation analysis provide insights into the pathogenesis of heart failure. *Nat. Commun.* **11**, 163. (doi:10.1038/s41467-019-13690-5)
  65. Geisterfer-Lowrance AA, Kass S, Tanigawa G, Vosberg HP, McKenna W, Seidman CE, Seidman JG. 1990 A molecular basis for familial hypertrophic cardiomyopathy: a beta cardiac myosin heavy chain gene missense mutation. *Cell* **62**, 999–1006. (doi:10.1016/0092-8674(90)90274-1)
  66. Tanigawa G, Jarcho JA, Kass S, Solomon SD, Vosberg HP, Seidman JG, Seidman CE. 1990 A molecular basis for familial hypertrophic cardiomyopathy: an alpha/beta cardiac myosin heavy chain hybrid gene. *Cell* **62**, 991–998. (doi:10.1016/0092-8674(90)90273-H)
  67. Otsuka H *et al.* 2012 Prevalence and distribution of sarcomeric gene mutations in Japanese patients with familial hypertrophic cardiomyopathy. *Circ. J.* **76**, 453–461. (doi:10.1253/circj.CJ-11-0876)
  68. Herman DS *et al.* 2012 Truncations of titin causing dilated cardiomyopathy. *N. Engl. J. Med.* **366**, 619–628. (doi:10.1056/NEJMoa1110186)
  69. Mazzarotto F *et al.* 2020 Reevaluating the genetic contribution of monogenic dilated cardiomyopathy. *Circulation* **141**, 387–398. (doi:10.1161/CIRCULATIONAHA.119.037661)
  70. Stark K *et al.* 2010 Genetic association study identifies HSPB7 as a risk gene for idiopathic dilated



- cardiomyopathy. *PLoS Genet.* **6**, e1001167. (doi:10.1371/journal.pgen.1001167)
71. Meder B *et al.* 2014 A genome-wide association study identifies 6p21 as novel risk locus for dilated cardiomyopathy. *Eur. Heart J.* **35**, 1069–1077. (doi:10.1093/eurheartj/ehz251)
  72. Norton N *et al.* 2011 Genome-wide studies of copy number variation and exome sequencing identify rare variants in BAG3 as a cause of dilated cardiomyopathy. *Am. J. Hum. Genet.* **88**, 273–282. (doi:10.1016/j.ajhg.2011.01.016)
  73. Xu H *et al.* 2018 A genome-wide association study of idiopathic dilated cardiomyopathy in African Americans. *J. Pers. Med.* **8**, 11. (doi:10.3390/jpm8010011)
  74. Sveinbjornsson G *et al.* 2018 Variants in NKX2–5 and FLNC cause dilated cardiomyopathy and sudden cardiac death. *Circ. Genom. Precis. Med.* **11**, e002151. (doi:10.1161/CIRCGEN.117.002151)
  75. Heinig M *et al.* 2017 Natural genetic variation of the cardiac transcriptome in non-diseased donors and patients with dilated cardiomyopathy. *Genome Biol.* **18**, 170. (doi:10.1186/s13059-017-1286-z)
  76. Kirk EP *et al.* 2007 Mutations in cardiac T-box factor gene TBX20 are associated with diverse cardiac pathologies, including defects of septation and valvulogenesis and cardiomyopathy. *Am. J. Hum. Genet.* **81**, 280–291. (doi:10.1086/519530)
  77. Haas J *et al.* 2018 Genomic structural variations lead to dysregulation of important coding and non-coding RNA species in dilated cardiomyopathy. *EMBO Mol. Med.* **10**, 107–120. (doi:10.15252/emmm.201707838)
  78. Schaid DJ, Chen W, Larson NB. 2018 From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat. Rev. Genet.* **19**, 491–504. (doi:10.1038/s41576-018-0016-z)
  79. Osterwalder M *et al.* 2018 Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* **554**, 239–243. (doi:10.1038/nature25461)
  80. Heinig M. 2018 Using gene expression to annotate cardiovascular GWAS loci. *Front. Cardiovasc. Med.* **5**, 59. (doi:10.3389/fcvm.2018.00059)
  81. Consortium EP. 2012 An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74. (doi:10.1038/nature11247)
  82. Roadmap Epigenomics C *et al.* 2015 Integrative analysis of 111 reference human epigenomes. *Nature* **518**, 317–330. (doi:10.1038/nature14248)
  83. Paul DS *et al.* 2013 Maps of open chromatin highlight cell type-restricted patterns of regulatory sequence variation at hematological trait loci. *Genome Res.* **23**, 1130–1141. (doi:10.1101/gr.155127.113)
  84. Cheng Y *et al.* 2014 Principles of regulatory information conservation between mouse and human. *Nature* **515**, 371–375. (doi:10.1038/nature13985)
  85. Slagle CE, Conlon FL. 2016 Emerging field of cardiomics: high-throughput investigations into transcriptional regulation of cardiovascular development and disease. *Trends Genet.* **32**, 707–716. (doi:10.1016/j.tig.2016.09.002)
  86. Luna-Zurita L *et al.* 2016 Complex interdependence regulates heterotypic transcription factor distribution and coordinates cardiogenesis. *Cell* **164**, 999–1014. (doi:10.1016/j.cell.2016.01.004)
  87. van den Boogaard M *et al.* 2012 Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *J. Clin. Investig.* **122**, 2519–2530. (doi:10.1172/JCI62613)
  88. van den Boogaard M *et al.* 2014 A common genetic variant within SCN10A modulates cardiac SCN5A expression. *J. Clin. Investig.* **124**, 1844–1852. (doi:10.1172/JCI73140)
  89. Montefiori LE *et al.* 2018 A promoter interaction map for cardiovascular disease genetics. *Elife* **7**, e35788. (doi:10.7554/eLife.35788)
  90. Oka T, Xu J, Molkentin JD. 2007 Re-employment of developmental transcription factors in adult heart disease. *Semin. Cell Dev. Biol.* **18**, 117–131. (doi:10.1016/j.semcdb.2006.11.012)
  91. He A, Gu F, Hu Y, Ma Q, Ye LY, Akiyama JA, Visel A, Pennacchio LA, Pu WT. 2014 Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat. Commun.* **5**, 4907. (doi:10.1038/ncomms5907)
  92. Papat R *et al.* 2013 Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy. *Proc. Natl Acad. Sci. USA* **110**, 20164–20169. (doi:10.1073/pnas.1315155110)
  93. Spurrell CH *et al.* 2019 Genome-wide fetalization of enhancer architecture in heart disease. *bioRxiv.* (doi:10.1101/591362)
  94. van der Harst P *et al.* 2016 52 Genetic loci influencing myocardial mass. *J. Am. Coll. Cardiol.* **68**, 1435–1448. (doi:10.1016/j.jacc.2016.07.729)
  95. Fraser J, Williamson I, Bickmore WA, Dostie J. 2015 An overview of genome organization and how we got there: from FISH to Hi-C. *Microbiol. Mol. Biol. Rev.* **79**, 347–372. (doi:10.1128/MMBR.00006-15)
  96. Giorgetti L, Heard E. 2016 Closing the loop: 3C versus DNA FISH. *Genome Biol.* **17**, 215. (doi:10.1186/s13059-016-1081-2)
  97. Sergeeva IA, Hooijkaas IB, Ruijter JM, van der Made I, de Groot NE, van de Werken HJ, Creemers EE, Christoffels VM. 2016 Identification of a regulatory domain controlling the Nppa-Nppb gene cluster during heart development and stress. *Development* **143**, 2135–2146. (doi:10.1242/dev.132019)
  98. van den Boogaard M *et al.* 2019 Identification and characterization of a transcribed distal enhancer involved in cardiac Kcnh2 regulation. *Cell Rep.* **28**, 2704–2714. (doi:10.1016/j.celrep.2019.08.007)
  99. Mumbach MR *et al.* 2017 Enhancer connectome in primary human cells identifies target genes of disease-associated DNA elements. *Nat. Genet.* **49**, 1602–1612. (doi:10.1038/ng.3963)
  100. Smith AJP, Deloukas P, Munroe PB. 2018 Emerging applications of genome-editing technology to examine functionality of GWAS-associated variants for complex traits. *Physiol. Genom.* **50**, 510–522. (doi:10.1152/physiolgenomics.00028.2018)
  101. Visel A *et al.* 2010 Targeted deletion of the 9p21 non-coding coronary artery disease risk interval in mice. *Nature* **464**, 409–412. (doi:10.1038/nature08801)
  102. Dickel DE *et al.* 2016 Genome-wide compendium and functional assessment of *in vivo* heart enhancers. *Nat. Commun.* **7**, 12923. (doi:10.1038/ncomms12923)
  103. Mummery CL. 2018 Perspectives on the use of human induced pluripotent stem cell-derived cardiomyocytes in biomedical research. *Stem Cell Rep.* **11**, 1306–1311. (doi:10.1016/j.stemcr.2018.11.011)
  104. Mosqueira D *et al.* 2018 CRISPR/Cas9 editing in human pluripotent stem cell-cardiomyocytes highlights arrhythmias, hypocontractility, and energy depletion as potential therapeutic targets for hypertrophic cardiomyopathy. *Eur. Heart J.* **39**, 3879–3892. (doi:10.1093/eurheartj/ehy249)
  105. Fulco CP *et al.* 2019 Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations. *Nat. Genet.* **51**, 1664–1669. (doi:10.1038/s41588-019-0538-0)
  106. Harismendy O *et al.* 2011 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* **470**, 264–268. (doi:10.1038/nature09753)
  107. Motterle A *et al.* 2012 Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Hum. Mol. Genet.* **21**, 4021–4029. (doi:10.1093/hmg/ddz224)
  108. Beaudoin M *et al.* 2015 Myocardial infarction-associated SNP at 6p24 interferes with MEF2 binding and associates with PHACTR1 expression levels in human coronary arteries. *Arterioscler. Thromb. Vasc. Biol.* **35**, 1472–1479. (doi:10.1161/ATVBAHA.115.305534)
  109. Miller CL *et al.* 2016 Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci. *Nat. Commun.* **7**, 12092. (doi:10.1038/ncomms12092)
  110. Gupta RM *et al.* 2017 A genetic variant associated with five vascular diseases is a distal regulator of endothelin-1 gene expression. *Cell* **170**, 522–533. (doi:10.1016/j.cell.2017.06.049)
  111. Krause MD *et al.* 2018 Genetic variant at coronary artery disease and ischemic stroke locus 1p32.2 regulates endothelial responses to hemodynamics. *Proc. Natl Acad. Sci. USA* **115**, E11349–E11358. (doi:10.1073/pnas.1810568115)
  112. Lo Sardo V *et al.* 2018 Unveiling the role of the most impactful cardiovascular risk locus through haplotype editing. *Cell* **175**, 1796–1810. (doi:10.1016/j.cell.2018.11.014)
  113. Zhao JY *et al.* 2012 Functional variant in methionine synthase reductase intron-1 significantly increases the risk of congenital heart disease in the Han Chinese population. *Circulation* **125**, 482–490. (doi:10.1161/CIRCULATIONAHA.111.050245)
  114. Yang B *et al.* 2017 Protein-altering and regulatory genetic variants near GATA4 implicated in bicuspid

- aortic valve. *Nat. Commun.* **8**, 15481. (doi:10.1038/ncomms15481)
115. Alankarage D *et al.* 2019 Functional characterisation of a novel PBX1 de novo missense variant identified in a patient with syndromic congenital heart disease. *Hum. Mol. Genet.* **29**, 1068–1082. (doi:10.1093/hmg/ddz231)
116. Kapoor A *et al.* 2014 An enhancer polymorphism at the cardiomyocyte intercalated disc protein NOS1AP locus is a major regulator of the QT interval. *Am. J. Hum. Genet.* **94**, 854–869. (doi:10.1016/j.ajhg.2014.05.001)
117. Garg P *et al.* 2018 Genome editing of induced pluripotent stem cells to decipher cardiac channelopathy variant. *J. Am. Coll. Cardiol.* **72**, 62–75. (doi:10.1016/j.jacc.2018.04.041)
118. Li L, Zhang L, Binkley PF, Sadee W, Wang D. 2017 Regulatory variants modulate protein kinase C alpha (PRKCA) gene expression in human heart. *Pharm. Res.* **34**, 1648–1657. (doi:10.1007/s11095-017-2102-x)
119. Tucker NR *et al.* 2017 Diminished PRRX1 expression is associated with increased risk of atrial fibrillation and shortening of the cardiac action potential. *Circ. Cardiovasc. Genet.* **10**, e001902. (doi:10.1161/CIRCGENETICS.117.001902)
120. Daimi H, Khelil AH, Neji A, Ben Hamda K, Maaoui S, Aranega A, Be Chibani J, Franco D. 2019 Role of SCN5A coding and non-coding sequences in Brugada syndrome onset: what's behind the scenes? *Biomed. J.* **42**, 252–260. (doi:10.1016/j.bj.2019.03.003)
121. Te Riele AS *et al.* 2017 Multilevel analyses of SCN5A mutations in arrhythmogenic right ventricular dysplasia/cardiomyopathy suggest non-canonical mechanisms for disease pathogenesis. *Cardiovasc. Res.* **113**, 102–111. (doi:10.1093/cvr/cvw234)
122. Christensen AH, Andersen CB, Wassilew K, Svendsen JH, Bundgaard H, Brand SM, Schmitz B. 2019 Rare non-coding Desmoglein-2 variant contributes to Arrhythmogenic right ventricular cardiomyopathy. *J. Mol. Cell. Cardiol.* **131**, 164–170. (doi:10.1016/j.yjmcc.2019.04.029)
123. Smemo S *et al.* 2014 Obesity-associated variants within FTO form long-range functional connections with IRX3. *Nature* **507**, 371–375. (doi:10.1038/nature13138)
124. Frayling TM *et al.* 2007 A common variant in the FTO gene is associated with body mass index and predisposes to childhood and adult obesity. *Science* **316**, 889–894. (doi:10.1126/science.1141634)
125. Turcot V *et al.* 2018 Protein-altering variants associated with body mass index implicate pathways that control energy intake and expenditure in obesity. *Nat. Genet.* **50**, 26–41. (doi:10.1038/s41588-017-0011-x)
126. Fischer J, Koch L, Emmerling C, Vierkotten J, Peters T, Bruning JC, Ruther U. 2009 Inactivation of the Fto gene protects from obesity. *Nature* **458**, 894–898. (doi:10.1038/nature07848)
127. Uchida S, Dimmeler S. 2015 Long noncoding RNAs in cardiovascular diseases. *Circ. Res.* **116**, 737–750. (doi:10.1161/CIRCRESAHA.116.302521)
128. Grote P *et al.* 2013 The tissue-specific lncRNA Fendrr is an essential regulator of heart and body wall development in the mouse. *Dev. Cell* **24**, 206–214. (doi:10.1016/j.devcel.2012.12.012)
129. Yang KC, Yamada KA, Patel AY, Topkara VK, George I, Cheema FH, Ewald GA, Mann DL, Nerbonne JM. 2014 Deep RNA sequencing reveals dynamic regulation of myocardial noncoding RNAs in failing human heart and remodeling with mechanical circulatory support. *Circulation* **129**, 1009–1021. (doi:10.1161/CIRCULATIONAHA.113.003863)
130. Ounzain S *et al.* 2014 Functional importance of cardiac enhancer-associated noncoding RNAs in heart development and disease. *J. Mol. Cell. Cardiol.* **76**, 55–70. (doi:10.1016/j.yjmcc.2014.08.009)
131. Wang Z *et al.* 2016 The long noncoding RNA Chaer defines an epigenetic checkpoint in cardiac hypertrophy. *Nat. Med.* **22**, 1131–1139. (doi:10.1038/nm.4179)
132. Ishii N *et al.* 2006 Identification of a novel non-coding RNA, MIAT, that confers risk of myocardial infarction. *J. Hum. Genet.* **51**, 1087–1099. (doi:10.1007/s10038-006-0070-9)
133. Aprea J *et al.* 2013 Transcriptome sequencing during mouse brain development identifies long non-coding RNAs functionally involved in neurogenic commitment. *EMBO J.* **32**, 3145–3160. (doi:10.1038/emboj.2013.245)
134. Zhou J, Zhou Y, Wang CX. 2018 LncRNA-MIAT regulates fibrosis in hypertrophic cardiomyopathy (HCM) by mediating the expression of miR-29a-3p. *J. Cell. Biochem.* **120**, 7265–7275. (doi:10.1002/jcb.28001)
135. Wang Y, Wang JG. 2019 Genome-wide association studies of hypertension and several other cardiovascular diseases. *Pulse (Basel)* **6**, 169–186. (doi:10.1159/000496150)
136. Aung N *et al.* 2018 Association between ambient air pollution and cardiac morpho-functional phenotypes: insights from the UK Biobank population imaging study. *Circulation* **138**, 2175–2186. (doi:10.1161/CIRCULATIONAHA.118.034856)
137. Kessler T, Vilne B, Schunkert H. 2016 The impact of genome-wide association studies on the pathophysiology and therapy of cardiovascular disease. *EMBO Mol. Med.* **8**, 688–701. (doi:10.15252/emmm.201506174)
138. Mega JL *et al.* 2015 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. (doi:10.1016/S0140-6736(14)61730-X)
139. Gibson G. 2019 On the utilization of polygenic risk scores for therapeutic targeting. *PLoS Genet.* **15**, e1008060. (doi:10.1371/journal.pgen.1008060)
140. Lee JH, Protze SI, Laksman Z, Backx PH, Keller GM. 2017 Human pluripotent stem cell-derived atrial and ventricular cardiomyocytes develop from distinct mesoderm populations. *Cell Stem Cell* **21**, 179–194. (doi:10.1016/j.stem.2017.07.003)
141. Zhao Y *et al.* 2019 A platform for generation of chamber-specific cardiac tissues and disease modeling. *Cell* **176**, 913–927. (doi:10.1016/j.cell.2018.11.042)
142. Kwon EZ *et al.* 2020 Comprehensive *in vivo* interrogation reveals phenotypic impact of human enhancer variants. *Cell* **180**, 1262–1271. (doi:10.1016/j.cell.2020.02.031)