

Understanding cardiovascular disease: a journey through the genome (and what we found there)

Enrique Lara-Pezzi^{1,*,#}, Ana Dopazo^{2,*,#} and Miguel Manzanares^{1,*,#}

Cardiovascular disease (CVD) is a major cause of mortality and hospitalization worldwide. Several risk factors have been identified that are strongly associated with the development of CVD. However, these explain only a fraction of cases, and the focus of research into the causes underlying the unexplained risk has shifted first to genetics and more recently to genomics. A genetic contribution to CVD has long been recognized; however, with the exception of certain conditions that show Mendelian inheritance, it has proved more challenging than anticipated to identify the precise genomic components responsible for the development of CVD. Genome-wide association studies (GWAS) have provided information about specific genetic variations associated with disease, but these are only now beginning to reveal the underlying molecular mechanisms. To fully understand the biological implications of these associations, we need to relate them to the exquisite, multilayered regulation of protein expression, which includes chromatin remodeling, regulatory elements, microRNAs and alternative splicing. Understanding how the information contained in the DNA relates to the operation of these regulatory layers will allow us not only to better predict the development of CVD but also to develop more effective therapies.

“We shall not cease from exploration
And the end of all our exploring
Will be to arrive where we started
And know the place for the first time.”

Little Gidding by T. S. Eliot (1942)

Introduction

Cardiovascular disease (CVD) is the leading cause of morbidity and mortality worldwide, responsible for an estimated 17.5 million deaths in 2005, representing 30% of all deaths (Mudd and Kass, 2008). In Europe alone, CVD causes over 4.3 million deaths each year, and is the leading cause of death (48%) and disease burden (23%). Despite substantial advances in medical management, the prognosis of CVD remains poor, and the identification of mechanisms and potential therapeutic approaches are still a priority of considerable importance.

CVD has a very clear environmental component; however, the risk factors defined in epidemiological studies (e.g. hypertension, high cholesterol, smoking) explain only a fraction of events (Thanassoulis and Vasan, 2010). Much hope has been set on the potential of genetics and genomics to reveal the molecular mechanisms responsible for the development of CVD and to explain cases that are not obviously

correlated with known risk factors. The strong familial component of CVD has long been recognized and is evident from the large-scale population studies that helped to define the classical set of CVD risk factors. However, studies of CVD heritability are confounded by the fact that several other risk factors, such as blood pressure, lipid levels and diabetes, are themselves under genetic control (North et al., 2003). Nonetheless, several studies have noted that family history is an independent risk factor (Shea et al., 1984; Myers et al., 1990), so that CVD heritability does not merely reflect the genetic component of classical risk factors. Unfortunately, evidence of a high degree of heritability of a given trait does not mean that identification of the underlying genes will be straightforward. The analysis of common multifactorial diseases such as CVD is hindered by the interdependence of genetic and environmental factors and the difficulties that are inherent in separating the influence of individual factors.

In order to fully grasp the contribution of genetics to CVD, we must go beyond classical Mendelian genetics, and explore the multiple interacting layers that regulate the genome. This will include the analysis of not only the protein-coding sequences of the genome, but the vast non-coding regions as well. An unbiased view of the whole genome is now possible, thanks to the development of large-scale approaches led by next generation sequencing (NGS) technologies. (For a Glossary of terms used, see Box 1.) In this review, we discuss how studying the many layers regulating the genome – including chromatin structure, cis-regulatory elements, microRNAs and alternative splicing – is beginning to provide us with a fresh view of the development of and therapeutic strategies for CVD.

The past: understanding the heritability of CVD

To date, most knowledge about genes causing CVD comes from the study of congenital heart defects and rare forms of familial

¹Cardiovascular Development and Repair Department, and ²Genomics Unit, Centro Nacional de Investigaciones Cardiovasculares (CNIC), Melchor Fernández Almagro 3, 28029 Madrid, Spain

*Authors for correspondence (elara@cnic.es; adopazo@cnic.es; mmanzanares@cnic.es)

#All authors contributed equally to this work

© 2012. Published by The Company of Biologists Ltd
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial Share Alike License (<http://creativecommons.org/licenses/by-nc-sa/3.0>), which permits unrestricted non-commercial use, distribution and reproduction in any medium provided that the original work is properly cited and all further distributions of the work or adaptation are subject to the same Creative Commons License terms.

Box 1. Glossary: approaches for studying layers of genomic regulation

ChIP-seq: combines chromatin immunoprecipitation (ChIP) with next generation sequencing to identify the binding sites of DNA-associated proteins. It can be used to map global binding sites precisely for any protein of interest. Previously, ChIP-on-chip was the most common technique used to study these protein-DNA relationships.

Chromatin immunoprecipitation (ChIP): a technique used to identify potential regulatory sequences by isolating soluble DNA chromatin extracts (complexes of DNA and protein) using antibodies that recognize specific DNA-binding proteins.

DNase I hypersensitivity assay: a method for detecting sites throughout the genome that are more easily cleaved by DNase I owing to an open chromatin configuration; these sites are usually associated with active regulatory elements.

Exome sequencing: an efficient strategy to selectively sequence the exome – the subset of a genome that is protein coding – as a cheaper but still effective alternative to whole genome sequencing. Exons are short, functionally important sequences of DNA that represent the regions in genes that are translated into protein.

Formaldehyde-assisted isolation of regulatory elements (FAIRE): an approach used to identify regulatory genomic regions based on their open conformation and their loose association with structural proteins. Sheared chromatin is subjected to phenol-chloroform extraction; protein-free chromatin remains in the aqueous solution, whereas tightly packed chromatin is retained in the organic phase. FAIRE-enriched chromatin can then be sequenced by next generation DNA sequencing (FAIRE-seq).

Genome-wide association studies (GWAS): studies that search for a population association between a phenotype and a particular allele by screening loci (most commonly by genotyping SNPs) across the entire genome.

Next generation DNA sequencing (NGS): highly parallel DNA sequencing technology that produces many hundreds of thousands or millions of short reads (25-500 bp) at low cost and in a short time. Currently, most established sequencing platforms include the Illumina/Solexa Analyzers, Roche/454 Genome Sequencer and Applied Biosystems SOLiD platforms.

inherited CVD. Congenital heart defects can be non-syndromic or syndromic (such as Holt-Oram and DiGeorge syndromes), and usually affect genes involved in the early phases of heart development; these genes can encode transcription factors (such as Gata4, Nkx2.5 or Tbx factors), signaling pathway components (such as those of the Notch or the Ras-MAPK pathways) or structural proteins (such as the α -myosin heavy chain encoded by the *MYH6* gene) (Srivastava, 2006). Cardiac developmental defects are the most common human birth defects (Hoffman and Kaplan, 2002) and their study remains of great importance. Furthermore, a still-uncharted territory is that of developmental defects that do not cause an overt fetal phenotype but that subtly impair proper cardiac function or stress responses and are an underlying morphological substrate for later-onset CVD.

Familial conditions are usually caused by autosomal dominant mutations that show a variable degree of expressivity. Identification of the culprit genes underlying these inherited CVDs has allowed genetic testing in families with affected members, therefore allowing counseling and prevention in asymptomatic carriers (Tester and Ackerman, 2011). Inherited heart diseases are broadly classified into cardiomyopathies and channelopathies. In cardiomyopathies, such as hypertrophic cardiomyopathy or arrhythmogenic right ventricular cardiomyopathy, there is a structural alteration in the heart (Watkins et al., 2011). By contrast,

the hearts of individuals with channelopathies are structurally normal, but these conditions, such as long QT or Brugada syndromes, mainly cause arrhythmias and can result in sudden cardiac death (Bastiaenen and Behr, 2011).

Although information gained from the study of these and other conditions has been extraordinarily useful in defining the genetic basis of cardiovascular morphology and physiology, they only account for a very small fraction of heritable CVDs. The existence of these conditions has also prompted numerous studies in which candidate genes, identified because of their developmental function in animal models or in tissue culture assays, have been screened for disease association in affected families or small case-control studies. However, many associations identified in this way have not been replicated, showing the inherent limitations of this approach (Ioannidis et al., 2001; Morgan et al., 2007).

The present: GWAS, eQTLs and exome sequencing Genome-wide association studies

The sequencing of the human genome, together with an ever expanding catalog of variations mapped as single nucleotide polymorphisms (SNPs), has allowed large-scale genome-wide association studies (GWAS), which have the potential to fill the gap in our understanding of the genetic basis of CVD and other common diseases (Hirschhorn and Daly, 2005). Because they examine the entire genome in an unbiased fashion, GWAS approaches can identify any genomic region involved in a given disease and should therefore be free of any kind of ascertainment bias, which is the main problem with candidate gene approaches. Overall, there is no doubt that GWAS have provided interesting new biological insights into many conditions. Surprising findings include the discovery of SNPs in genes that were originally not thought to have a role in disease and the identification of loci shared by diseases, including CVDs, that were previously thought to be unrelated (Pandey, 2010; O'Donnell and Nabel, 2011).

However, GWAS have, in the view of many, failed to deliver as expected. The main criticisms are that the identified alleles explain only a small fraction of the heritability of common diseases and traits (Manolio et al., 2009) and have a low predictive value compared with classical risk factors (Thanassoulis and Vasan, 2010). Possible reasons for this apparent failure are that initial expectations were too high or that the wrong questions were posed. In addition, because GWAS are designed to identify common variants, the existence of rare variants with a large effect has not been addressed. Consequently, initial hopes that the identification of a small number of SNPs would explain the inheritance of CVD and other diseases and serve as predictors, similarly to classical Mendelian inheritance, have not been realized, because this is not the aim of GWAS.

GWAS methodology has improved and possible explanations for missing heritability are currently being explored (Zuk et al., 2012); moreover, ongoing investigation of GWAS results continues to increase the number of variants associated with common traits. A major strategy developed to overcome the limitations of GWAS is a brute-force attack by means of meta-analysis, in which information from multiple GWAS analyses is pooled to increase the number of cases and controls analyzed (Cantor et al., 2010; Pandey, 2010). Another strategy to refine and complement GWAS is targeted re-sequencing of disease-associated loci to identify rare or causal variants. This approach has successfully identified rare

variants associated with hypertriglyceridemia (Johansen et al., 2010), although exactly how the rare and common variants interact and associate with the disease is still unclear.

Despite the limitations of GWAS analyses, their unbiased searching of the whole genome for associations has identified previously unknown genetic components of CVD. Some of the discovered loci were completely unexpected, opening new avenues of research into the pathways and processes underlying disease. For example, variants discovered at the 9p21 region were the first common genetic variants identified as genetic risk factors for coronary artery disease and other forms of CVDs independent of classical risk factors (Burton et al., 2007; Helgadottir et al., 2007; McPherson et al., 2007; Samani et al., 2007). The SNP variants in the 9p21 gene desert region constitute an early and elegant example of how disease association can be related to molecular phenomena. Different mechanisms for this association have been proposed, such as disruption of regulatory elements involved in the interferon- γ signaling response (Harismendy et al., 2011) or of a non-coding RNA (Pasmant et al., 2011). Another unexpected association is that of *SORT1*, encoding the multi-ligand sorting protein sortilin 1, with plasma low-density-lipoprotein cholesterol levels and myocardial infarction; this association has uncovered a newly identified pathway involved in lipid metabolism and CVD, thus offering new possibilities for therapeutic intervention (Dube et al., 2011).

It is also notable that many disease-associated variants identified by GWAS are located in loci that have previously been linked to Mendelian diseases (Lupski et al., 2011). For example, mutations in *KCNE1*, encoding a potassium channel subunit, are responsible for congenital long QT syndrome (Splawski et al., 2000) and Jervell and Lange-Nielsen syndrome, and GWAS-identified variants near this gene are associated with QT interval duration (Newton-Cheh et al., 2009). A similar finding has been reported for *SCN5A*, which is associated with Brugada syndrome (Ruan et al., 2009). Therefore, classical Mendelian genetics and some GWAS converge on a similar range of genomic loci, the only difference being the nature and effect of the variants (Lupski et al., 2011).

Gene expression and eQTLs

Gene expression in heart disease has been studied for a long time. Some genes, such as *NPPA* [which encodes atrial natriuretic factor (ANF)], *NPPB* [brain natriuretic peptide (BNP)], *ACTA1* (α -skeletal actin) and *MYH7* (β -myosin heavy chain), are known to be induced in stressed cardiomyocytes in infarcted, hypertrophic and dilated hearts. Indeed, some of these genes have been used for diagnosis and prognosis in heart disease.

Gene expression in the heart is strongly heritable (Petretto et al., 2006), which suggests that it is genetically controlled. By combining microarray studies with genetic linkage analysis we can identify control points in the genome that regulate gene expression. These are considered expression quantitative trait loci (eQTL) and can help reveal genetic associations for certain traits for which GWAS has not provided a clear answer. One such trait is left ventricular mass, a risk factor for heart failure and a predictor of all-cause mortality. By using microarray and linkage analysis, an association was found, both in rats and humans, between increased left ventricular mass and elevated expression of the extracellular matrix protein osteoglycin (Petretto et al., 2008). After hypertrophic stimulation, osteoglycin knockout mice show reduced ventricular

mass compared with wild-type mice, establishing a causal relationship between osteoglycin expression and increased left ventricular mass. The same group recently established an association between ventricular mass and endonuclease G (*Endog*), a mitochondrial nuclease that controls mitochondrial mass (McDermott-Roe et al., 2011). Loss-of-function mutations in *Endog* result in increased left ventricular mass and a decline in cardiac function in rats. Deletion of the *Endog* gene in mice induces mitochondrial depletion and dysfunction, increased oxidative stress, and cardiac steatosis and hypertrophy. eQTL analysis has also unveiled an inflammatory network, driven by interferon regulatory factor 7 (*Irf7*), that is mainly expressed by monocytes and macrophages and is associated with susceptibility to type 1 diabetes (Heinig et al., 2010). Expression of the genes in this network correlates with genetic variation in Epstein-Barr virus-induced gene 2 (*Ebi2*), suggesting that the network is controlled by this gene. These studies reveal how eQTL analysis can improve our knowledge of the biological basis of CVD.

Exome sequencing

With the advent of NGS technologies, exome sequencing – the targeted sequencing of the portion of the human genome that is protein coding – has become a widely used tool for discovering rare alleles underlying Mendelian phenotypes and complex traits. Although recent examples abound of the successful use of exome sequencing to identify genes underlying monogenic disorders (Bamshad et al., 2011), this approach has yet to make a major contribution to our understanding of complex diseases, and is currently the focus of a major debate. Large-scale sequencing efforts have shown that putative loss-of-function variants in protein-coding genes are abundant in apparently healthy individuals (MacArthur and Tyler-Smith, 2010; Conrad et al., 2011), which calls for caution when establishing causal relations between mutations identified by exome sequencing and disease etiology. In the cardiovascular field, exome sequencing has identified several previously unknown candidate causal genes for Mendelian disorders, including *SHROOM3* for heterotaxy (Tariq et al., 2011) and *ANGPTL3* in cases of familial combined hypolipidemia (Musunuru et al., 2010). In other cases in which exome sequencing failed to identify the causal variant, complementary approaches such as copy-number variation analysis did so; for example, variation in *BAG3* causes familial dilated cardiomyopathy (Norton et al., 2011). Exome sequencing and what the future holds for this technology have recently been reviewed (Bamshad et al., 2011). However, it is likely that this approach will soon be superseded by whole genome sequencing, which has the advantage of being free of the ascertainment bias inherent to exome sequencing, which explores just 1% of the genome.

Layers of regulation in CVD

Genomic analysis has so far fallen short of expectations in terms of establishing the underlying genetic cause of CVD. Although associations between certain loci and disease have been found, their biological impact is not immediately obvious because we remain largely ignorant of the role of genomic elements located outside coding regions.

Whole individual genomes will be fully sequenced in the near future, and comparison of these sequences might identify new

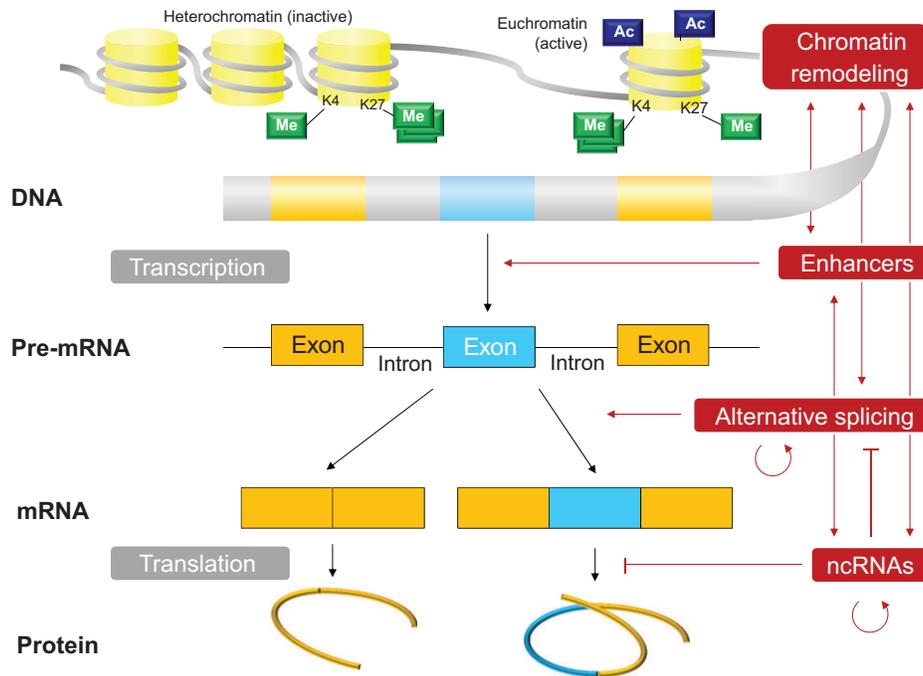


Fig. 1. Gene activity is controlled by multiple layers of regulation. For RNA to be transcribed, chromatin needs to be unwound so that genes can be accessed by the transcription machinery. Chromatin remodeling is controlled by methylation (Me) and acetylation (Ac) marks on histones that determine how open or closed the chromatin is. Activation of gene transcription is also controlled by enhancer regulatory elements that can sometimes be several kilobases away. Once transcribed, the immature mRNA (pre-mRNA) undergoes splicing to remove introns. The inclusion or exclusion of an exon is subject to variation (alternative splicing), which can generate multiple protein isoforms from a single gene. Splicing regulators are often alternatively spliced themselves (represented by curved arrow). mRNA translation is regulated by a range of non-coding RNAs (ncRNAs), including microRNAs. Some ncRNAs, such as ceRNAs, can regulate the activity of microRNAs (represented by curved arrow). These multiple layers of regulation interact with each other; for instance, chromatin modification and microRNAs both modulate alternative splicing, and ncRNAs can also act as transcriptional enhancers to promote transcription. The different regulatory mechanisms respond to pathological stimuli by changing both the amounts and the isoforms of the multiple proteins that are involved in tissue homeostasis and the development of CVD. Further investigations will determine the potential of each of these layers to be targeted therapeutically.

associations with CVD. However, it is arguable whether these data will yield biological information about the molecular mechanisms that underlie CVD. Even if this information is encoded in the DNA sequence and can be found, it is by no means certain that we would be able to understand it. The production of a particular protein in a cell from the information encoded in DNA is a process that is exquisitely controlled; intertwined layers of regulation provide both qualitative and quantitative control of gene expression. To understand the real contribution of the different elements in the genome to CVD, it is crucial to carry out a multilayered analysis of cells and tissues. Such an approach must include the analysis of chromatin modification and epigenetic regulation, activation of regulatory elements, gene transcription, alternative splicing and non-coding RNAs, among other levels of control (Fig. 1). We therefore need to look beyond the DNA sequence and integrate different layers of biological information so that we can return to the starting point, look again at the DNA sequence and, hopefully, understand for the first time what is written in it (as prefigured in Eliot's poem).

Chromatin modifications

DNA is wound around histones to form chromatin, which is in turn tightly packed to save space. For gene transcription to occur, chromatin needs to be locally unwound to make the locus accessible

to the transcription machinery. Different modifications along the chromatin thread determine whether it will be open or remain closed. These are mainly acetylation and methylation modifications of histones and DNA, which can be inherited from cell to cell through mitosis and constitute an epigenetic level of gene regulation. Epigenetic regulation is particularly sensitive to environmental changes and is thought to be a major mechanism by which external stimuli induce an inheritable response (Ordovas and Smith, 2010). In addition, differences in epigenetic modifications might explain changes in disease susceptibility in the absence of DNA sequence variation. Epigenetic changes are not written in the DNA sequence but can be detected at a global level through ChIP-seq, which combines immunoprecipitation of modified histones linked to DNA with NGS. A recent report describes distinct epigenetic modifications identified in end-stage heart failure patients (Movassagh et al., 2011).

Epigenetic modifications are carried out by families of histone acetyl transferases (HATs), histone deacetylases (HDACs), histone methyl transferases (HMTs) and histone demethylases (HDMs). These protein families play a major role in heart development and disease (Ohtani and Dimmeler, 2011). Histone-modifying enzymes do not bind to DNA but are recruited by transcription factors, coactivators and repressors that provide the sequence specificity that these enzymes lack. Histone acetylation occurs on lysine

residues and promotes DNA unwinding to facilitate transcription, whereas deacetylation has the opposite effect. The role of HDACs in heart disease is diverse and has been thoroughly studied in animal models (Haberland et al., 2009). HDAC9 and HDAC5 control cardiac growth, and knockout mice for *Hdac9* are more sensitive to hypertrophic stimuli (Zhang et al., 2002; Chang et al., 2004). Furthermore, variants in the vicinity of *HDAC9* are implicated in causing increased risk of large vessel ischemic stroke (Bellenguez et al., 2012). HDAC2 contributes to the development of cardiac hypertrophy and re-expression of the fetal gene program associated with cardiac stress (Trivedi et al., 2007). HDAC3 is necessary for proper heart metabolism. Cardiac-specific deletion of the gene encoding this protein results in deregulation of PPAR γ , lipid accumulation and massive cardiac hypertrophy. HATs also play roles in heart disease; for example, alterations in the levels of p300 can lead either to developmental cardiac defects or to hypertrophy and heart failure (Yao et al., 1998; Miyamoto et al., 2006; Wei et al., 2008). Although these and other studies highlight roles for HATs and HDACs in heart development and disease, it should be noted that these enzymes might have targets other than histones and that their effects might not necessarily be directly linked to chromatin modification.

Histones can be methylated on different lysine residues and so can undergo different degrees of methylation, resulting in distinct effects on transcription regulation (Ohtani and Dimmeler, 2011). Loss of K4 (Lys4) trimethylation on histone 3 (H3) deregulates expression of ion channels and cytoskeleton genes and results in altered contraction (Stein et al., 2011). Like HDACs, HDMs play a major role in heart development and disease. In the embryo, the HDM Jarid2 regulates cardiomyocyte proliferation and is necessary for proper heart development (Lee et al., 2000; Toyoda et al., 2003). Another HDM, UTX, acts on trimethylated K27 to de-repress transcription of cardiac genes and switch on cardiac differentiation (Lee et al., 2012). In the adult heart, expression of JMJD2A is increased in individuals with hypertrophic cardiomyopathy and plays a key role in the development of cardiac hypertrophy, as revealed through gain- and loss-of-function mouse models (Zhang et al., 2011). The role of HMTs in the heart is less well understood. A recent report shows that the HMT DotL1 controls dystrophin transcription and that DotL1 knockout mice develop dilated cardiomyopathy (Nguyen et al., 2011). For a full review of the role of these proteins in cardiovascular development, see Chang and Bruneau (Chang and Bruneau, 2012).

Chromatin is also remodeled by the Brg1/Brm-associated factor (BAF) complexes in an ATP-dependent manner. The BAF complex is composed of several proteins, some of which have been reported to play a major role in the heart. For example, Brg1 expression is induced in individuals with hypertrophic cardiomyopathy and is necessary for cardiac hypertrophy development in mice (Hang et al., 2010). Another BAF-complex component, Baf60c, plays a major role in cardiac differentiation and heart development (Lickert et al., 2004). Together, these reports highlight how chromatin-remodeling regulators control gene expression and susceptibility to CVD.

Cis-regulatory elements

The identification and characterization of the functional sequence elements that determine when, where and how much a gene is expressed is a major goal in biology (Birney et al., 2007; Myers et

al., 2011). Acquiring this knowledge will explain a large fraction of the genetic component of CVD. However, at present this is a daunting task because we still have not learned to recognize regulatory elements based on sequence alone (as can be done for coding regions), and their identification thus far relies on time-consuming small-throughput functional assays. Furthermore, there is still no clear estimate of how many functional elements are present in the genome, although it is expected that a large fraction of non-coding sequences (the vast majority of the genome) will contain such elements. The early observation (Dermitzakis et al., 2002) that sequence conservation of non-coding regions is comparable to that of coding regions further supports a functional role for non-coding regions, because they have been conserved throughout evolution and subjected to positive selection.

A major frustration with GWAS has been that the great majority of risk-associated SNPs are in the non-coding portion of the genome, and usually at a great distance from the nearest protein-coding gene or in different linkage disequilibrium blocks. This disappointment arises from the common view that coding variants will have a more profound effect than those in functional elements present in the non-coding regions (Cooper and Shendure, 2011). Nevertheless, there is little direct evidence supporting this, and we are only starting to glimpse the complexity of gene regulation through non-coding functional elements. Because these vastly outnumber coding exons, the abundance of non-coding risk variants identified by GWAS should come as no surprise.

The advent of new technologies such as ChIP-seq that allow genome-wide screening to identify regions bound by specific proteins (transcription factors) or that show certain epigenetic features (DNA methylation, nucleosome-free regions or histone modifications) has opened new avenues in the search for functional cis-regulatory elements. Recent work has shown how combinations of these epigenetic marks distinguish active from inactive chromatin regions, or proximal regulatory elements (promoters) from distal elements (Heintzman et al., 2007; Rada-Iglesias et al., 2011). Active regulatory elements have also been shown to possess an 'open' chromatin configuration characterized by depletion of nucleosomes, which can be identified by DNase I hypersensitivity or FAIRE (formaldehyde-assisted identification of regulatory elements) assays (Song et al., 2011). With these tools in hand, it is now possible to generate a catalog of candidate regulatory elements in distinct genomic regions.

Investigators have begun to use such approaches to determine the regulatory basis of heart development and function, with the expectation that such knowledge will yield useful insights into the genetic component of CVD. Analysis of the binding of the co-activator protein p300 to chromatin has identified thousands of putative heart enhancers in mice (Blow et al., 2010) and humans (May et al., 2011). More importantly, functional testing in mouse transgenic assays has validated the predicted enhancer function of a large proportion of these predicted elements. Other studies using a combination of bioinformatic, conservation and functional validation have also predicted an extensive list of genomic locations with putative regulatory activity in the heart (Narlikar et al., 2010).

In other cases, the genome-wide binding profiles of cardiac-specific transcription factors (including Gata4, Nkx2.5, Tbx5, Srf and Mef2a) have been defined using in vitro mouse tissue culture systems (He et al., 2011; Schlesinger et al., 2011). These studies showed that

these factors often bind to the same genomic regions, identifying cis-regulatory modules, in a similar manner to that described in other systems such as embryonic stem cells (Chambers and Tomlinson, 2009). The integration of these data with maps of histone modification, DNA methylation, co-activator and nucleosome exclusion in cardiovascular tissues will provide a comprehensive view of how the genome is dynamically regulated in health and disease.

Nevertheless, the generation of complete catalogs of cis-regulatory elements presents some problems. First, functional validation of predictions based on genome-wide scans will be only as good as the assays used. At present, robust assays for enhancer function are available (both for tissue culture and in vivo transgenesis). However, novel assays to test for silencer, insulator or other functions need to be developed and applied in a systematic fashion to classify predicted elements. Second, once these elements have been found, it is necessary to identify which genes they act on. Although common sense would predict that the nearest gene will be the target, this is not necessarily the case. To address this issue, high-throughput chromatin interaction maps can be built to show the genomic architecture in a tissue-specific manner (de Wit and de Laat, 2012). And last but not least, a major problem facing studies of tissue-specific patterns of gene regulation is the availability of samples, both from patients and controls. This is particularly the case for the study of CVD, because biopsies are rare and in many circumstances the etiology of the diseased tissue is complex and multifactorial. Attempts have been made to correlate changes in gene expression and regulation by using another easily accessible tissue as a proxy (such as blood). However, careful analysis shows that gene expression patterns are not correlated even between closely related samples (Powell et al., 2011).

Alternative splicing

After the completion of the human genome project, researchers were somewhat puzzled, and fairly disappointed, by the low number of protein-coding genes in the genome. The 21,000 genes in the genome must suffice to make the ~1 million proteins estimated to be present in the human proteome. Protein diversity originates not so much in DNA, but in RNA. As genes are being transcribed, introns are removed from the immature mRNA and exons are linked together in a process known as mRNA splicing. Rather than being a rigid process, splicing allows alternative combinations of exons that result in the production of different proteins from a single gene. In fact, alternative splicing (AS) is considered the main factor underlying protein diversity.

Virtually all genes (94%) with more than one exon undergo AS. AS events include exon inclusion or exclusion, mutually exclusive exons, usage of alternative 5' or 3' splice sites and intron retention. These changes result in alterations in the final mRNA product that can lead to shifts in the open reading frame, generation of premature stop codons (nonsense-mediated decay) or changes in protein domains. Between 50% and 80% of splicing events are regulated in a tissue-specific manner (Wang and Burge, 2008). A recent study using exon microarrays showed reduced splicing efficiency and altered AS of sarcomeric genes in heart-failure patients (Kong et al., 2010). However, our knowledge about the involvement of AS in heart pathophysiology is mainly limited to individual genes whose different isoforms play diverse roles during heart failure. For example, AS variants of the troponin I gene reduce

contraction efficiency, and AS-generated titin isoforms alter cardiac stiffness in individuals with dilated cardiomyopathy (Makarenko et al., 2004; Feng and Jin, 2010). AS of the sodium channel *SCN5A*, which mediates cardioprotection by ischemic preconditioning, generates two non-functional variants during heart failure (Shang et al., 2007). Similarly, a splicing variant of cell-cycle-regulated kinase (*CCRK*), which promotes cardiomyocyte growth and survival, is downregulated in heart failure (Qiu et al., 2008). We have also observed that the *CnAβ1* AS variant of the phosphatase calcineurin improves cardiac function after infarction instead of inducing maladaptive hypertrophy like other calcineurin isoforms (Felkin et al., 2011).

AS is regulated by cis-regulatory genomic sequences that are present in the alternatively spliced exon and its flanking introns, and by trans-regulatory factors that recognize these sequences. The role of some of these trans-regulatory factors has been studied in animal models. Knockout mice for *SF2/ASF*, a splicing enhancer of the serine and arginine-rich (SR) protein family, show contraction defects that lead to dilated cardiomyopathy and death within the first 8 weeks of life due to postnatal developmental defects (Xu et al., 2005). Similarly, knockout of the SR protein *SC35* leads to dilated cardiomyopathy 3-5 weeks after birth, although *SC35* itself is not necessary for cardiac development (Ding et al., 2004). Elevated levels of the CELF trans-regulatory factor *CUGBP1* are associated with myotonic dystrophy, and overexpression of a dominant-negative CELF protein results in cardiac hypertrophy, fibrosis and dilated cardiomyopathy (Ladd et al., 2005; Wang et al., 2007). Interestingly, *Rbm20* knockout mice were recently reported to develop dilated cardiomyopathy due to defects in titin splicing (Guo et al., 2012). These defects are also detected in humans with *RBM20* mutations. Although these reports highlight the general importance of AS trans-regulatory factors in embryonic and postnatal development, their role in the response of the heart to pathological stimuli and their therapeutic potential has barely been explored.

Non-coding RNAs

Much of what was once considered – somewhat presumptuously – ‘junk’ DNA has turned out to be actively transcribed to produce RNAs that do not encode protein information [non-coding RNAs (ncRNAs)] but instead act as regulators of other RNAs. Depending on their size, origin or function, ncRNAs are classified as microRNAs, long non-coding RNAs (lncRNAs) and so on. Although these elements are transcribed from intergenic or intronic regions of the genome, they might hold the key to understanding how genetic variation can impact protein expression.

microRNAs are endogenous small non-coding RNAs of ~22 nucleotides that constitute the predominant form of double-stranded RNA (dsRNA) in mammalian cells. Unlike other interference RNAs, most microRNAs do not induce mRNA cleavage, but instead silence protein expression through post-transcriptional mechanisms, either preventing mRNA translation or regulating its stability. Rather than an on-off switch, microRNAs can be viewed as biological rheostats that fine-tune the expression of a specific protein (Bartel and Chen, 2004; Baek et al., 2008). MicroRNA-encoding genes are usually found in intergenic, intronic or polycistronic regions and lack a canonical TATA box and introns (Rana, 2007). MicroRNAs are transcribed in the nucleus mainly by type II RNA polymerases and are processed first by

Drosha and then by Dicer into a mature microRNA (Rana, 2007). Depending on its degree of complementarity with the microRNA, the target mRNA will either be prevented from undergoing translation or will be degraded (Rana, 2007). Target mRNAs often carry more than one microRNA-recognition sequence in their 3'UTR and can sometimes be regulated by two different microRNAs. In addition, each microRNA can target up to hundreds of different mRNAs, the efficiency depending on the complementarity between microRNA and mRNA (Baek et al., 2008). MicroRNAs play a major role in CVD, which has been discussed extensively elsewhere (Latronico and Condorelli, 2009).

lncRNAs play diverse – and somewhat unexpected – regulatory roles (Wilusz et al., 2009). Long intergenic non-coding RNAs (lincRNAs) are transcribed from enhancer elements and act as transcriptional enhancers for neighboring genes (De Santa et al., 2010; Ørom et al., 2010), a role also ascribed to some short ncRNAs called eRNAs (Kim et al., 2010). The mechanism of action of these ncRNAs is not entirely clear, although they are speculated to facilitate the assembly of histone-modifying complexes (Ong and Corces, 2011). An additional category of lincRNAs was recently described, called competing endogenous RNAs (ceRNAs) (Franco-Zorrilla et al., 2007; Tay et al., 2011). ceRNAs mimic mRNA target sequences for microRNAs, resulting in competition between ceRNA and mRNA for the microRNA. In this way, ceRNAs act as microRNA 'sponges', regulating the availability of microRNAs and indirectly controlling the amount of protein that will be translated from the mRNA target. Although nothing is known about ceRNAs in the heart, the ceRNA linc-MD1 was recently identified as a regulator of muscle differentiation by sequestering miR-133 and miR-135, which results in increased expression of their targets (MAML1 and MEF2C) and activation of muscle-specific transcription (Cesana et al., 2011). More than 100 lincRNAs were shown to be differentially expressed in hypertrophic mouse hearts in a recent RNA-seq analysis (Lee et al., 2011). However, their general function and specific roles in heart disease remain unknown.

Intertwined regulation

It is important to note that all of the layers of biological regulation discussed above also control each other (Fig. 1). For instance, both chromatin modification and microRNAs modulate AS (Kalsotra et al., 2010; Luco et al., 2010). This regulation is bidirectional: splicing activation can promote the recruitment of HMTs, and changes in AS regulatory sequences result in the repositioning of epigenetic marks (de Almeida et al., 2011; Kim et al., 2011). Like protein-coding genes, ncRNAs are epigenetically regulated and some of them can in turn modulate chromatin modification (Szulwach et al., 2010). As explained above, ncRNAs can also act as transcriptional enhancers. Together, these reports clearly indicate that the expression of a protein in a given cell and during a given time frame is very delicately regulated. If we are to infer information from DNA sequences, we need to integrate all available data on the genome together with correlative studies with other -omics analyses (such as transcriptomics, metabolomics, exposomics), which will provide a broader conceptual framework within which to understand the heritability of common diseases. The characterization of the majority of these genomic features will require functional assays in tissue culture or in vivo that will have to be carried out in animal models of CVD (e.g. mouse, rat, dog,

Fig. 1). Another approach that holds great promise is the generation of patient-specific induced pluripotent stem (iPS) cells, which can then be differentiated towards a cardiovascular phenotype. Such 'disease-in-a-dish' models have begun to be used for the study of Mendelian CVD (Carvajal-Vergara et al., 2010; Itzhaki et al., 2011; Yazawa et al., 2011) and population-scale GWAS analysis in iPS-cell-derived cardiac cells is not far away.

The future

With sequencing costs dropping rapidly, we will be soon shift from analysis of exomes to genomes, enabling us to interrogate many more variants that might be important for controlling gene transcription, splicing or even translation. Technical challenges remain, but work to surmount these difficulties is progressing rapidly. For instance, the limitations inherent in the use of human samples (e.g. myocardium) are being overcome, at least in part, through the use of improved animal and cellular models that more closely resemble human physiology, including pigs and other large animals, and patient-derived iPS cells. There is no doubt that the discovery of variants that underlie Mendelian and complex traits will lead to a much deeper understanding of disease mechanisms, and these insights should, in turn, facilitate the development of better diagnostics, prevention strategies and targeted therapeutics. With the advent of whole-genome sequencing on a population scale,

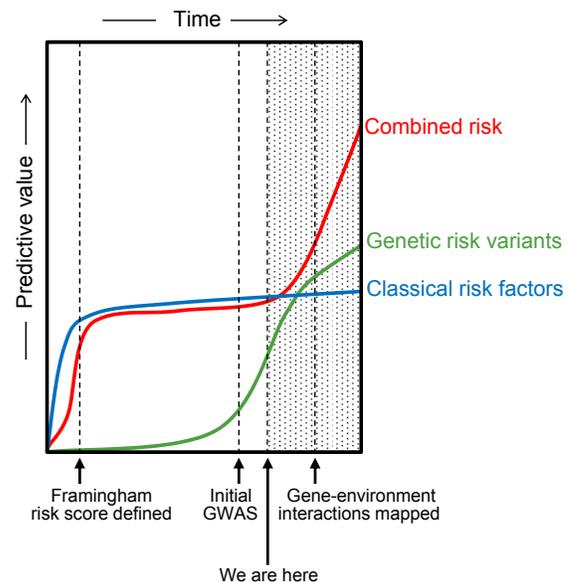


Fig. 2. Future trends in predicting the risk of CVD. The need to predict the risk of CVD is undeniable. The Framingham Heart Study defined a set of risk factors (represented by blue line) that has been the gold standard since the 1960s. However, these classical risk factors only predict a fraction of CVD. Although large-scale genetic analyses have identified hundreds of risk-associated common sequence variants (green line), these have added little or no additional predictive value to the classical risk score. However, this trend is set to change in the near future, as additional common variants and novel rare variants are identified. Furthermore, once we gain a better understanding of how genes and gene variants interact with each other and with environmental and behavioral factors, a synergistic effect can be expected, and thus our ability to predict CVD before the appearance of symptoms will greatly increase (red line).

the next challenge will be to identify rare variants with possibly larger effects. For this we will need to interpret non-coding variation in a similar way to how we predict the effect of coding variants on the protein product of genes. At present, this is not possible, but as our understanding of the regulatory genome grows, we believe this capability will be a reality in the near future. So far, the predictive value of identified variants associated with CVD is poor, and barely improves current scores based on classical risk factors (Thanassoulis et al., 2012). However, novel advances in the field, together with a better understanding of gene-environment interactions, will surely result in an increased predictive score for CVD risk (Fig. 2), and therefore a brighter outlook for early prevention and diagnosis.

The question that still lingers is how much of our susceptibility to developing CVD is written in the genome. According to a deterministic view, the probability of developing CVD and our susceptibility to environmental factors would be fully coded in the genome, and consequently our ability to prevent disease would be conditioned by the depth of our knowledge of this code. The reality is that we do not know to what extent information contained within the DNA code predicts disease, and our efforts to discover more have so far been somewhat naive. On the bright side, identifying the full predictive potential of the genome is a finite task and as such can be expected to be completed eventually, even if it takes decades or centuries. A combination of full genome sequencing, more powerful questionnaires to identify additional environmental risk factors, new animal models to study cause-effect relationships and increased knowledge of regulatory layers will provide the information needed to establish new correlations with DNA sequence. Researchers will use these blocks to build regulatory codes that will better explain the biological role of each piece of the genome so that we can use genetic variation to predict and treat disease. With the development of faster computers and increased bioinformatic expertise, we will one day be in a position to answer the question of how much information our genome really contains. Until then, we will have to take things one step at a time.

ACKNOWLEDGEMENTS

We thank Simon Bartlett for critical reading of the manuscript and English editing.

COMPETING INTERESTS

The authors declare that they do not have any competing or financial interests.

FUNDING

E.L.-P. was supported by grants from the European Union [ERG-239158, ITN-289600], the Spanish Ministry of Science and Innovation [BFU2009-10016, CP08/00144] and the Regional Government of Madrid [S2010/BMD-2321]; A.D. by a grant from the Spanish Ministry of Science and Innovation [FIS PI10/01124]; and M.M. by grants from the Spanish Ministry of Science and Innovation [BFU2011-23083], the Regional Government of Madrid [S2010/BMD-2315] and a CNIC Translational Grant [CNIC-08-2009]. The CNIC is supported by the Spanish Government and the Pro-CNIC Foundation.

REFERENCES

- Baek, D., Villén, J., Shin, C., Camargo, F. D., Gygi, S. P. and Bartel, D. P. (2008). The impact of microRNAs on protein output. *Nature* **455**, 64-71.
- Bamshad, M. J., Ng, S. B., Bigham, A. W., Tabor, H. K., Emond, M. J., Nickerson, D. A. and Shendure, J. (2011). Exome sequencing as a tool for Mendelian disease gene discovery. *Nat. Rev. Genet.* **12**, 745-755.
- Bartel, D. P. and Chen, C. Z. (2004). Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs. *Nat. Rev. Genet.* **5**, 396-400.
- Bastiaenen, R. and Behr, E. R. (2011). Sudden death and ion channel disease: pathophysiology and implications for management. *Heart* **97**, 1365-1372.
- Bellenguez, C., Bevan, S., Gschwendtner, A., Spencer, C. C., Burgess, A. I., Pirinen, M., Jackson, C. A., Traylor, M., Strange, A., Su, Z. et al. (2012). Genome-wide association study identifies a variant in HDAC9 associated with large vessel ischemic stroke. *Nat. Genet.* **44**, 328-333.
- Birney, E., Stamatoyannopoulos, J. A., Dutta, A., Guigo, R., Gingeras, T. R., Margulies, E. H., Weng, Z., Snyder, M., Dermitzakis, E. T., Thurman, R. E. et al. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. *Nature* **447**, 799-816.
- Blow, M. J., McCulley, D. J., Li, Z., Zhang, T., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F. et al. (2010). ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* **42**, 806-810.
- Burton, P. R., Clayton, C. D., Cardon, L. R., Craddock, N., Deloukas, P., Duncanson, A., Kwiatkowski, D. P., McCarthy, M. I., Ouwehand, W. H., Samani, N. J. et al. (2007). Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661-678.
- Cantor, R. M., Lange, K. and Sinsheimer, J. S. (2010). Prioritizing GWAS results: A review of statistical methods and recommendations for their application. *Am. J. Hum. Genet.* **86**, 6-22.
- Carvajal-Vergara, X., Sevilla, A., D'Souza, S. L., Ang, Y. S., Schaniel, C., Lee, D. F., Yang, L., Kaplan, A. D., Adler, E. D., Rozov, R. et al. (2010). Patient-specific induced pluripotent stem-cell-derived models of LEOPARD syndrome. *Nature* **465**, 808-812.
- Cesana, M., Cacchiarelli, D., Legnini, I., Santini, T., Sthandier, O., Chinappi, M., Tramontano, A. and Bozzoni, I. (2011). A long noncoding RNA controls muscle differentiation by functioning as a competing endogenous RNA. *Cell* **147**, 358-369.
- Chambers, I. and Tomlinson, S. R. (2009). The transcriptional foundation of pluripotency. *Development* **136**, 2311-2322.
- Chang, C.-P. and Bruneau, B. (2012). Epigenetics and cardiovascular development. *Annu. Rev. Physiol.* **74**, 41-68.
- Chang, S., McKinsey, T. A., Zhang, C. L., Richardson, J. A., Hill, J. A. and Olson, E. N. (2004). Histone deacetylases 5 and 9 govern responsiveness of the heart to a subset of stress signals and play redundant roles in heart development. *Mol. Cell. Biol.* **24**, 8467-8476.
- Conrad, D. F., Keebler, J. E., DePristo, M. A., Lindsay, S. J., Zhang, Y., Casals, F., Idaghdour, Y., Hartl, C. L., Torroja, C., Garimella, K. V. et al. (2011). Variation in genome-wide mutation rates within and between human families. *Nat. Genet.* **43**, 712-714.
- Cooper, G. M. and Shendure, J. (2011). Needles in stacks of needles: finding disease-causal variants in a wealth of genomic data. *Nat. Rev. Genet.* **12**, 628-640.
- de Almeida, S. F., Grosso, A. R., Koch, F., Fenouil, R., Carvalho, S., Andrade, J., Levezinho, H., Gut, M., Eick, D., Gut, I. et al. (2011). Splicing enhances recruitment of methyltransferase HYPB/Setd2 and methylation of histone H3 Lys36. *Nat. Struct. Mol. Biol.* **18**, 977-983.
- De Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B. K., Müller, H., Ragoussis, J., Wei, C.-L. and Natoli, G. (2010). A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. *PLoS Biol.* **8**, e1000384.
- de Wit, E. and de Laat, W. (2012). A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11-24.
- Dermitzakis, E. T., Reymond, A., Lyle, R., Scamuffa, N., Ucla, C., Deutsch, S., Stevenson, B. J., Flegel, V., Bucher, P., Jongeneel, C. V. et al. (2002). Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* **420**, 578-582.
- Ding, J. H., Xu, X., Yang, D., Chu, P. H., Dalton, N. D., Ye, Z., Yeakley, J. M., Cheng, H., Xiao, R. P., Ross, J. et al. (2004). Dilated cardiomyopathy caused by tissue-specific ablation of SC35 in the heart. *EMBO J.* **23**, 885-896.
- Dube, J. B., Johansen, C. T. and Hegele, R. A. (2011). Sortilin: an unusual suspect in cholesterol metabolism: from GWAS identification to in vivo biochemical analyses, sortilin has been identified as a novel mediator of human lipoprotein metabolism. *BioEssays* **33**, 430-437.
- Felkin, L. E., Narita, T., Germack, R., Shintani, Y., Takahashi, K., Sarathchandra, P., López-Olañeta, M. M., Gómez-Salineró, J. M., Suzuki, K., Barton, P. J. R. et al. (2011). Calcineurin splicing variant CnAβ1 improves cardiac function after myocardial infarction without inducing hypertrophy. *Circulation* **123**, 2838-2847.
- Feng, H.-Z. and Jin, J. P. (2010). Coexistence of cardiac troponin T variants reduces heart efficiency. *Am. J. Physiol. Heart Circ. Physiol.* **299**, H97-H105.
- Franco-Zorrilla, J. M., Valli, A., Todesco, M., Mateos, I., Puga, M. I., Rubio-Somoza, I., Leyva, A., Weigel, D., Garcia, J. A. and Paz-Ares, J. (2007). Target mimicry provides a new mechanism for regulation of microRNA activity. *Nat. Genet.* **39**, 1033-1037.
- Guo, W., Schafer, S., Greaser, M. L., Radke, M. H., Liss, M., Govindarajan, T., Maatz, H., Schulz, H., Li, S., Parrish, A. M. et al. (2012). RBM20, a gene for hereditary cardiomyopathy, regulates titin splicing. *Nat. Med.* [Epub ahead of print] Doi: 10.1038/nm2693.
- Haberland, M., Montgomery, R. L. and Olson, E. N. (2009). The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat. Rev. Genet.* **10**, 32-42.

- Hang, C. T., Yang, J., Han, P., Cheng, H.-L., Shang, C., Ashley, E., Zhou, B. and Chang, C.-P. (2010). Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature* **466**, 62-67.
- Harismendy, O., Notani, D., Song, X., Rahim, N. G., Tanasa, B., Heintzman, N., Ren, B., Fu, X. D., Topol, E. J., Rosenfeld, M. G. et al. (2011). 9p21 DNA variants associated with coronary artery disease impair interferon-gamma signalling response. *Nature* **470**, 264-268.
- He, A., Kong, S. W., Ma, Q. and Pu, W. T. (2011). Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc. Natl. Acad. Sci. USA* **108**, 5632-5637.
- Heinig, M., Petretto, E., Wallace, C., Bottolo, L., Rotival, M., Lu, H., Li, Y., Sarwar, R., Langley, S. R., Bauerfeind, A. et al. (2010). A trans-acting locus regulates an antiviral expression network and type 1 diabetes risk. *Nature* **467**, 460-464.
- Heintzman, N. D., Stuart, R. K., Hon, G., Fu, Y., Ching, C. W., Hawkins, R. D., Barrera, L. O., Van Calcar, S., Qu, C., Ching, K. A. et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* **39**, 311-318.
- Helgadottir, A., Thorleifsson, G., Manolescu, A., Gretarsdottir, S., Blondal, T., Jonasdottir, A., Sigurdsson, A., Baker, A., Palsson, A., Masson, G. et al. (2007). A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491-1493.
- Hirschhorn, J. N. and Daly, M. J. (2005). Genome-wide association studies for common diseases and complex traits. *Nat. Rev. Genet.* **6**, 95-108.
- Hoffman, J. I. and Kaplan, S. (2002). The incidence of congenital heart disease. *J. Am. Coll. Cardiol.* **39**, 1890-1900.
- Ioannidis, J. P., Ntzani, E. E., Trikalinos, T. A. and Contopoulos-Ioannidis, D. G. (2001). Replication validity of genetic association studies. *Nat. Genet.* **29**, 306-309.
- Itzhaki, I., Maizels, L., Huber, I., Zwi-Dantsis, L., Caspi, O., Winterstern, A., Feldman, O., Gepstein, A., Arbel, G., Hammerman, H. et al. (2011). Modelling the long QT syndrome with induced pluripotent stem cells. *Nature* **471**, 225-229.
- Johansen, C. T., Wang, J., Lanktree, M. B., Cao, H., McIntyre, A. D., Ban, M. R., Martins, R. A., Kennedy, B. A., Hassell, R. G., Visser, M. E. et al. (2010). Excess of rare variants in genes identified by genome-wide association study of hypertriglyceridemia. *Nat. Genet.* **42**, 684-687.
- Kalsotra, A., Wang, K., Li, P.-F. and Cooper, T. A. (2010). MicroRNAs coordinate an alternative splicing network during mouse postnatal heart development. *Genes Dev.* **24**, 653-658.
- Kim, S., Kim, H., Fong, N., Erickson, B. and Bentley, D. L. (2011). Pre-mRNA splicing is a determinant of histone H3K36 methylation. *Proc. Natl. Acad. Sci. USA* **108**, 13564-13569.
- Kim, T.-K., Hemberg, M., Gray, J. M., Costa, A. M., Bear, D. M., Wu, J., Harmin, D. A., Laptewicz, M., Barbara-Haley, K., Kuersten, S. et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. *Nature* **465**, 182-187.
- Kong, S. W., Hu, Y. W., Ho, J. W. K., Ikeda, S., Polster, S., John, R., Hall, J. L., Bisping, E., Pieske, B., dos Remedios, C. G. et al. (2010). Heart failure-associated changes in RNA splicing of sarcomere genes. *Circ. Cardiovasc. Genet.* **3**, 138-146.
- Ladd, A. N., Taffet, G., Hartley, C., Kearney, D. L. and Cooper, T. A. (2005). Cardiac tissue-specific repression of CELF activity disrupts alternative splicing and causes cardiomyopathy. *Mol. Cell. Biol.* **25**, 6267-6278.
- Latronico, M. V. G. and Condorelli, G. (2009). MicroRNAs and cardiac pathology. *Nat. Rev. Cardiol.* **6**, 418-429.
- Lee, J.-H., Gao, C., Peng, G., Greer, C., Ren, S., Wang, Y. and Xiao, X. (2011). Analysis of transcriptome complexity through RNA sequencing in normal and failing murine hearts / novelty and significance. *Circ. Res.* **109**, 1332-1341.
- Lee, S., Lee, J. W. and Lee, S.-K. (2012). UTX, a histone H3-lysine 27 demethylase, acts as a critical switch to activate the cardiac developmental program. *Dev. Cell* **22**, 25-37.
- Lee, Y., Song, A. J., Baker, R., Micales, B., Conway, S. J. and Lyons, G. E. (2000). Jumonji, a nuclear protein that is necessary for normal heart development. *Circ. Res.* **86**, 932-938.
- Lickert, H., Takeuchi, J. K., von Both, I., Walls, J. R., McAuliffe, F., Lee Adamson, S., Mark Henkelman, R., Wrana, J. L., Rossant, J. and Bruneau, B. G. (2004). Baf60c is essential for function of BAF chromatin remodelling complexes in heart development. *Nature* **432**, 107-112.
- Luco, R. F., Pan, Q., Tominaga, K., Blencowe, B. J., Pereira-Smith, O. M. and Misteli, T. (2010). Regulation of alternative splicing by histone modifications. *Science* **327**, 996-1000.
- Lupski, J. R., Belmont, J. W., Boerwinkle, E. and Gibbs, R. A. (2011). Clan genomics and the complex architecture of human disease. *Cell* **147**, 32-43.
- MacArthur, D. G. and Tyler-Smith, C. (2010). Loss-of-function variants in the genomes of healthy humans. *Hum. Mol. Genet.* **19**, R125-R130.
- Makarenko, I., Opitz, C. A., Leake, M. C., Neagoe, C., Kulke, M., Gwathmey, J. K., del Monte, F., Hajjar, R. J. and Linke, W. A. (2004). Passive stiffness changes caused by upregulation of compliant titin isoforms in human dilated cardiomyopathy hearts. *Circ. Res.* **95**, 708-716.
- Manolio, T. A., Collins, F. S., Cox, N. J., Goldstein, D. B., Hindorf, L. A., Hunter, D. J., McCarthy, M. I., Ramos, E. M., Cardon, L. R., Chakravarti, A. et al. (2009). Finding the missing heritability of complex diseases. *Nature* **461**, 747-753.
- May, D., Blow, M. J., Kaplan, T., McCulley, D. J., Jensen, B. C., Akiyama, J. A., Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C. et al. (2011). Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* **44**, 89-93.
- McDermott-Roe, C., Ye, J., Ahmed, R., Sun, X.-M., Serafin, A., Ware, J., Bottolo, L., Muckett, P., Canas, X., Zhang, J. et al. (2011). Endonuclease G is a novel determinant of cardiac hypertrophy and mitochondrial function. *Nature* **478**, 114-118.
- McPherson, R., Pertsemidis, A., Kavaslar, N., Stewart, A., Roberts, R., Cox, D. R., Hinds, D. A., Pennacchio, L. A., Tybjaerg-Hansen, A., Folsom, A. R. et al. (2007). A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488-1491.
- Miyamoto, S., Kawamura, T., Morimoto, T., Ono, K., Wada, H., Kawase, Y., Matsumori, A., Nishio, R., Kita, T. and Hasegawa, K. (2006). Histone acetyltransferase activity of p300 is required for the promotion of left ventricular remodeling after myocardial infarction in adult mice in vivo. *Circulation* **113**, 679-690.
- Morgan, T. M., Krumholz, H. M., Lifton, R. P. and Spertus, J. A. (2007). Nonvalidation of reported genetic risk factors for acute coronary syndrome in a large-scale replication study. *JAMA* **297**, 1551-1561.
- Movassagh, M., Choy, M.-K., Knowles, D. A., Cordeddu, L., Haider, S., Down, T., Siggins, L., Vujic, A., Simeoni, I., Penkett, C. et al. (2011). Distinct epigenomic features in end-stage failing human hearts. *Circulation* **124**, 2411-2422.
- Mudd, J. O. and Kass, D. A. (2008). Tackling heart failure in the twenty-first century. *Nature* **451**, 919-928.
- Musunuru, K., Pirruccello, J. P., Do, R., Peloso, G. M., Guiducci, C., Sougnez, C., Garimella, K. V., Fisher, S., Abreu, J., Barry, A. J. et al. (2010). Exome sequencing, ANGPL3 mutations, and familial combined hypolipidemia. *N. Engl. J. Med.* **363**, 2220-2227.
- Myers, R. H., Kiely, D. K., Cupples, L. A. and Kannel, W. B. (1990). Parental history is an independent risk factor for coronary artery disease: the Framingham Study. *Am. Heart J.* **120**, 963-969.
- Myers, R. M., Stamatoyannopoulos, J., Snyder, M., Dunham, I., Hardison, R. C., Bernstein, B. E., Gingeras, T. R., Kent, W. J., Birney, E., Wold, B. et al. (2011). A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* **9**, e1001046.
- Narlikar, L., Sakabe, N. J., Blanski, A. A., Arimura, F. E., Westlund, J. M., Nobrega, M. A. and Ovcharenko, I. (2010). Genome-wide discovery of human heart enhancers. *Genome Res.* **20**, 381-392.
- Newton-Cheh, C., Eijgelsheim, M., Rice, K. M., de Bakker, P. I., Yin, X., Estrada, K., Bis, J. C., Marcic, K., Rivadeneira, F., Nosenworthy, P. A. et al. (2009). Common variants at ten loci influence QT interval duration in the QTGEN Study. *Nat. Genet.* **41**, 399-406.
- Nguyen, A. T., Xiao, B., Neppi, R. L., Kallin, E. M., Li, J., Chen, T., Wang, D.-Z., Xiao, X. and Zhang, Y. (2011). DOT1L regulates dystrophin expression and is critical for cardiac function. *Genes Dev.* **25**, 263-274.
- North, K. E., Howard, B. V., Welty, T. K., Best, L. G., Lee, E. T., Yeh, J. L., Fabsitz, R. R., Roman, M. J. and MacCluer, J. W. (2003). Genetic and environmental contributions to cardiovascular disease risk in American Indians: the strong heart family study. *Am. J. Epidemiol.* **157**, 303-314.
- Norton, N., Li, D., Rieder, M. J., Siegfried, J. D., Rampersaud, E., Zuchner, S., Mangos, S., Gonzalez-Quintana, J., Wang, L., McGee, S. 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.
- O'Donnell, C. J. and Nabel, E. G. (2011). Genomics of cardiovascular disease. *N. Engl. J. Med.* **365**, 2098-2109.
- Ohtani, K. and Dimmeler, S. (2011). Epigenetic regulation of cardiovascular differentiation. *Cardiovasc. Res.* **90**, 404-412.
- Ong, C.-T. and Corces, V. G. (2011). Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat. Rev. Genet.* **12**, 283-293.
- Ordovas, J. M. and Smith, C. E. (2010). Epigenetics and cardiovascular disease. *Nat. Rev. Cardiol.* **7**, 510-519.
- Ørom, U. A., Derrien, T., Beringer, M., Gumireddy, K., Gardini, A., Bussotti, G., Lai, F., Zytnicki, M., Notredame, C., Huang, Q. et al. (2010). Long noncoding RNAs with enhancer-like function in human cells. *Cell* **143**, 46-58.
- Pandey, J. P. (2010). Genomewide association studies and assessment of risk of disease. *N. Engl. J. Med.* **363**, 2076-2077.
- Pasmant, E., Sabbagh, A., Vidaud, M. and Bieche, I. (2011). ANRIL, a long, noncoding RNA, is an unexpected major hotspot in GWAS. *FASEB J.* **25**, 444-448.
- Petretto, E., Mangion, J., Dickens, N. J., Cook, S. A., Kumaran, M. K., Lu, H., Fischer, J., Maatz, H., Kren, V., Pravenec, M. et al. (2006). Heritability and tissue specificity

- of expression quantitative trait loci. *PLoS Genet.* **2**, e172.
- Petretto, E., Sarwar, R., Grieve, I., Lu, H., Kumaran, M. K., Muckett, P. J., Mangion, J., Schroen, B., Benson, M., Punjabi, P. P. et al.** (2008). Integrated genomic approaches implicate osteoglycin (Ogn) in the regulation of left ventricular mass. *Nat. Genet.* **40**, 546-552.
- Powell, J. E., Henders, A. K., McRae, A. F., Wright, M. J., Martin, N. G., Dermitzakis, E. T., Montgomery, G. W. and Visscher, P. M.** (2011). Genetic control of gene expression in whole blood and lymphoblastoid cell lines is largely independent. *Genome Res.* **22**, 456-466.
- Qiu, H., Dai, H., Jain, K., Shah, R., Hong, C., Pain, J., Tian, B., Vatner, D. E., Vatner, S. F. and Depre, C.** (2008). Characterization of a novel cardiac isoform of the cell cycle-related kinase that is regulated during heart failure. *J. Biol. Chem.* **283**, 22157-22165.
- Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S. A., Flynn, R. A. and Wysocka, J.** (2011). A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* **470**, 279-283.
- Rana, T. M.** (2007). Illuminating the silence: understanding the structure and function of small RNAs. *Nat. Rev. Mol. Cell. Biol.* **8**, 23-36.
- Ruan, Y., Liu, N. and Priori, S. G.** (2009). Sodium channel mutations and arrhythmias. *Nat. Rev. Cardiol.* **6**, 337-348.
- Samani, N. J., Erdmann, J., Hall, A. S., Hengstenberg, C., Mangino, M., Mayer, B., Dixon, R. J., Meitinger, T., Braund, P., Wichmann, H. E. et al.** (2007). Genomewide association analysis of coronary artery disease. *N. Engl. J. Med.* **357**, 443-453.
- Schlesinger, J., Schueler, M., Grunert, M., Fischer, J. J., Zhang, Q., Krueger, T., Lange, M., Tonjes, M., Dunkel, I. and Sperling, S. R.** (2011). The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. *PLoS Genet.* **7**, e1001313.
- Shang, L. L., Pfahnl, A. E., Sanyal, S., Jiao, Z., Allen, J., Banach, K., Fahrenbach, J., Weiss, D., Taylor, W. R., Zafari, A. M. et al.** (2007). Human heart failure is associated with abnormal C-terminal splicing variants in the cardiac sodium channel. *Circ. Res.* **101**, 1146-1154.
- Shea, S., Ottman, R., Gabrieli, C., Stein, Z. and Nichols, A.** (1984). Family history as an independent risk factor for coronary artery disease. *J. Am. Coll. Cardiol.* **4**, 793-801.
- Song, L., Zhang, Z., Grasfeder, L. L., Boyle, A. P., Giresi, P. G., Lee, B. K., Sheffield, N. C., Graf, S., Huss, M., Keefe, D. et al.** (2011). Open chromatin defined by DNaseI and FAIRE identifies regulatory elements that shape cell-type identity. *Genome Res.* **21**, 1757-1767.
- Splawski, I., Shen, J., Timothy, K. W., Lehmann, M. H., Priori, S., Robinson, J. L., Moss, A. J., Schwartz, P. J., Towbin, J. A., Vincent, G. M. et al.** (2000). Spectrum of mutations in long-QT syndrome genes. KVLQT1, HERG, SCN5A, KCNE1, and KCNE2. *Circulation* **102**, 1178-1185.
- Srivastava, D.** (2006). Making or breaking the heart: from lineage determination to morphogenesis. *Cell* **126**, 1037-1048.
- Stein, A. B., Jones, T. A., Herron, T. J., Patel, S. R., Day, S. M., Noujaim, S. F., Milstein, M. L., Klos, M., Furspan, P. B., Jalife, J. et al.** (2011). Loss of H3K4 methylation destabilizes gene expression patterns and physiological functions in adult murine cardiomyocytes. *J. Clin. Invest.* **121**, 2641-2650.
- Szulwach, K. E., Li, X., Smrt, R. D., Li, Y., Luo, Y., Lin, L., Santistevan, N. J., Li, W., Zhao, X. and Jin, P.** (2010). Cross talk between microRNA and epigenetic regulation in adult neurogenesis. *J. Cell Biol.* **189**, 127-141.
- Tariq, M., Belmont, J. W., Lalani, S., Smolarek, T. and Ware, S. M.** (2011). SHROOM3 is a novel candidate for heterotaxy identified by whole exome sequencing. *Genome Biol.* **12**, R91.
- Tay, Y., Kats, L., Salmena, L., Weiss, D., Tan Shen, M., Ala, U., Karreth, F., Poliseno, L., Provero, P., Di Cunto, F. et al.** (2011). Coding-independent regulation of the tumor suppressor PTEN by competing endogenous mRNAs. *Cell* **147**, 344-357.
- Tester, D. J. and Ackerman, M. J.** (2011). Genetic testing for potentially lethal, highly treatable inherited cardiomyopathies/channelopathies in clinical practice. *Circulation* **123**, 1021-1037.
- Thanassoulis, G. and Vasan, R. S.** (2010). Genetic cardiovascular risk prediction: will we get there? *Circulation* **122**, 2323-2334.
- Thanassoulis, G., Peloso, G. M., Pencina, M. J., Hoffmann, U., Fox, C. S., Cupples, L. A., Levy, D., D'Agostino, R. B., Sr, Hwang, S. J. and O'Donnell, C. J.** (2012). A genetic risk score is associated with incident cardiovascular disease and coronary artery calcium: the Framingham Heart Study. *Circ. Cardiovasc. Genet.* **5**, 113-121.
- Toyoda, M., Shirato, H., Nakajima, K., Kojima, M., Takahashi, M., Kubota, M., Suzuki-Migishima, R., Motegi, Y., Yokoyama, M. and Takeuchi, T.** (2003). Jumonji downregulates cardiac cell proliferation by repressing cyclin D1 expression. *Dev. Cell* **5**, 85-97.
- Trivedi, C. M., Luo, Y., Yin, Z., Zhang, M., Zhu, W., Wang, T., Floss, T., Goettlicher, M., Noppinger, P. R., Wurst, W. et al.** (2007). Hdac2 regulates the cardiac hypertrophic response by modulating Gsk3[β] activity. *Nat. Med.* **13**, 324-331.
- Wang, G. S., Kearney, D. L., Biasi, M. D., Taffet, G. and Cooper, T. A.** (2007). Elevation of RNA-binding protein CUGBP1 is an early event in an inducible heart-specific mouse model of myotonic dystrophy. *J. Clin. Invest.* **117**, 2802-2811.
- Wang, Z. and Burge, C. B.** (2008). Splicing regulation: from a parts list of regulatory elements to an integrated splicing code. *RNA* **14**, 802-813.
- Watkins, H., Ashrafian, H. and Redwood, C.** (2011). Inherited cardiomyopathies. *N. Engl. J. Med.* **364**, 1643-1656.
- Wei, J. Q., Shehadeh, L. A., Mitrani, J. M., Pessanha, M., Slepak, T. I., Webster, K. A. and Bishopric, N. H.** (2008). Quantitative control of adaptive cardiac hypertrophy by acetyltransferase p300. *Circulation* **118**, 934-946.
- Wilusz, J. E., Sunwoo, H. and Spector, D. L.** (2009). Long noncoding RNAs: functional surprises from the RNA world. *Genes Dev.* **23**, 1494-1504.
- Xu, X., Yang, D., Ding, J.-H., Wang, W., Chu, P.-H., Dalton, N. D., Wang, H.-Y., Bermingham, J. R., Jr, Ye, Z., Liu, F. et al.** (2005). ASF/SF2-regulated CaMKII δ alternative splicing temporally reprograms excitation-contraction coupling in cardiac muscle. *Cell* **120**, 59-72.
- Yao, T.-P., Oh, S. P., Fuchs, M., Zhou, N.-D., Ch'ng, L.-E., Newsome, D., Bronson, R. T., Li, E., Livingston, D. M. and Eckner, R.** (1998). Gene dosage-dependent embryonic development and proliferation defects in mice lacking the transcriptional integrator p300. *Cell* **93**, 361-372.
- Yazawa, M., Hsueh, B., Jia, X., Pasca, A. M., Bernstein, J. A., Hallmayer, J. and Dolmetsch, R. E.** (2011). Using induced pluripotent stem cells to investigate cardiac phenotypes in Timothy syndrome. *Nature* **471**, 230-234.
- Zhang, C. L., McKinsey, T. A., Chang, S., Antos, C. L., Hill, J. A. and Olson, E. N.** (2002). Class II histone deacetylases act as signal-responsive repressors of cardiac hypertrophy. *Cell* **110**, 479-488.
- Zhang, Q.-J., Chen, H.-Z., Wang, L., Liu, D.-P., Hill, J. A. and Liu, Z.-P.** (2011). The histone trimethyllysine demethylase JMJD2A promotes cardiac hypertrophy in response to hypertrophic stimuli in mice. *J. Clin. Invest.* **121**, 2447-2456.
- Zuk, O., Hechter, E., Sunyaev, S. R. and Lander, E. S.** (2012). The mystery of missing heritability: genetic interactions create phantom heritability. *Proc. Natl. Acad. Sci. USA* **109**, 1193-1198.