Decoding the complex genetic causes of heart diseases using systems biology

Djordje Djordjevic · Vinita Deshpande · Tomasz Szczesnik · Andrian Yang ·
David T. Humphreys · Eleni Giannoulitou · Joshua W. K. Ho

Received: 9 September 2014 / Accepted: 10 November 2014 / Published online: 10 December 2014
© International Union for Pure and Applied Biophysics (IUPAB) and Springer-Verlag Berlin Heidelberg 2014

Abstract The pace of disease gene discovery is still much slower than expected, even with the use of cost-effective DNA sequencing and genotyping technologies. It is increasingly clear that many inherited heart diseases have a more complex polygenic aetiology than previously thought. Understanding the role of gene–gene interactions, epigenetics, and non-coding regulatory regions is becoming increasingly critical in predicting the functional consequences of genetic mutations identified by genome-wide association studies and whole-genome or exome sequencing. A systems biology approach is now being widely employed to systematically discover genes that are involved in heart diseases in humans or relevant animal models through bioinformatics. The overarching premise is that the integration of high-quality causal gene regulatory networks (GRNs), genomics, epigenomics, transcriptomics and other genome-wide data will greatly accelerate the discovery of the complex genetic causes of congenital and complex heart diseases. This review summarises state-of-the-art genomic and bioinformatics techniques that are used in accelerating the pace of disease gene discovery in heart diseases. Accompanying this review, we provide an interactive web-resource for systems biology analysis of mammalian

heart development and diseases, *CardiacCode* (<http://CardiacCode.victorchang.edu.au/>). CardiacCode features a dataset of over 700 pieces of manually curated genetic or molecular perturbation data, which enables the inference of a cardiac-specific GRN of 280 regulatory relationships between 33 regulator genes and 129 target genes. We believe this growing resource will fill an urgent unmet need to fully realise the true potential of *predictive* and *personalised* genomic medicine in tackling human heart disease.

Keywords Cardiac gene regulatory network · Whole-genome sequencing · Epigenomics · Congenital heart disease · Cardiomyopathy · Gene prioritisation

Introduction

Many forms of congenital and acquired cardiovascular diseases show familial aggregation, indicating that genetic factors play a role in disease aetiology. Advances in genomic technology have accelerated the transition from single gene assessment in rare Mendelian disorders, such as different types of congenital heart disease (CHD), to studies of more common complex disorders such as cardiomyopathies, ischemic heart disease and atrial fibrillation (AF) (Gelb and Chung 2014). In the latter group of disorders, many different additive genetic and environmental risk factors, each of relatively small effect size, are hypothesised to be contributing to the disease risk. In recent years, there has been a significant advancement in understanding the genetic contribution to different types of cardiac disease.

CHD is a large collection of structural and functional deficits that arise during cardiac embryogenesis, and exhibit large genetic heterogeneity (Fahed et al. 2013; Gelb and Chung 2014). CHDs are the most common form of birth defect, affecting up to 6–8 in 1,000 live-born babies (Blue

“Special Issue: Biophysics of Human Heart Failure”

Electronic supplementary material The online version of this article (doi:10.1007/s12551-014-0145-3) contains supplementary material, which is available to authorized users.

D. Djordjevic · V. Deshpande · T. Szczesnik · A. Yang ·
D. T. Humphreys · E. Giannoulitou · J. W. K. Ho (✉)
Victor Chang Cardiac Research Institute, Darlinghurst, NSW 2010,
Australia
e-mail: j.ho@victorchang.edu.au

D. Djordjevic · T. Szczesnik · D. T. Humphreys · E. Giannoulitou ·
J. W. K. Ho
St. Vincent's Clinical School, The University of New South Wales,
Darlinghurst, NSW 2010, Australia

et al. 2012). Familial CHD can be caused by single-gene mutations, but population prevalence of CHD can also indicate multifactorial aetiology following the common disease–common variant hypothesis. Currently, many known disease-causing genes affect developmental signalling pathways involved in cardiogenesis. Examples of such genes include *NKX2-5*, *NKX2-6*, *GATA4*, *GATA5*, *GATA6*, *IRX4*, *TBX20*, *ZIC3*, *NOTCH1*, *NOTCH2* and *JAG1* (Hershberger et al. 2013; Fahed et al. 2013). Mutations in these genes can repress or enhance gene transcription, or affect protein structure and function, which can lead to disruption of developmental signalling pathways. In addition, de novo mutations in histone-modifying genes were found to contribute to approximately 10 % of CHD cases (Zaidi et al. 2013), indicative of a strong epigenetic component. Following multiple gene discoveries based on highly penetrant gene mutations with Mendelian segregation in CHD, current methods that utilise state-of-the-art genomic techniques aim to give further insight into the genetic complexity of CHD. On the other hand, a different set of genes are involved in other inherited cardiovascular pathologies such as hypertrophic or dilated cardiomyopathy (HCM and DCM respectively), long QT syndrome (LQTS), or Marfan syndrome. These mainly arise from mutations in sarcomeric proteins or ion channel components with specific functions in cardiac biology such as *TTN*, *MYH11*, *MYH6*, *MYH7*, *SCN5A* and *KCNH2* among others (Hershberger et al. 2013; Fahed et al. 2013).

Traditional gene discovery methods such as linkage and candidate gene studies were successful in identifying causal genes in cardiac diseases such as HCM (Kimura et al. 1997), idiopathic DCM (Krajinovic et al. 1995; Messina et al. 1997), AF (Chen et al. 2003) and LQTS (Keating et al. 1991), among others. For more complex cardiac diseases, a hypothesis-free genome-wide approach was needed to assess their polygenic cause. The advent of Genome-Wide Association Studies (GWAS) facilitated the high-throughput screening of single nucleotide polymorphisms (SNPs) for the identification of

marker alleles or genotypes that are more frequent in diseased individuals compared to healthy control individuals.

In the past 8 years, GWAS have identified multiple genetic variants associated with many complex diseases including cardiac diseases (Cappola et al. 2010; Gudbjartsson et al. 2009). Although our understanding of the genetic architecture of many complex diseases has been improved, the findings so far confer small increments in risk, explaining a small proportion of familial clustering (Manolio et al. 2009). This has led to scepticism regarding the potential clinical applicability of these findings, while the question of what may contribute to the “missing heritability” remains open. The current hypothesis is that this may include contributions of rare inherited or de novo variants and epistatic effects, among others (Eichler et al. 2010; Manolio et al. 2009).

To explore the impact of GWAS on identifying genetic variants that are associated with heart diseases or traits, we searched the GWAS catalogue provided by the National Human Genome Research Institute of the USA (<http://www.genome.gov/gwastudies/>). As of 6 August 2014, this catalogue contains results from 1,960 publications, and 14,001 reported SNP associations. Using the keywords “heart OR cardiac OR cardio OR fibrillation OR coronary OR arrhythmia”, we found 75 publications satisfying this criterion, resulting in 600 reported SNP associations, which map to about 300 genes (Table S1). These publications were published from 2007 to 2013. As shown in Fig. 1, many SNPs have been reported to be associated with coronary diseases and heart rate/arrhythmia, but relatively little is known about cardiomyopathy and other heart muscle conditions (Fig. 1a). More importantly, the majority of the SNPs are located in intronic or intergenic regions (Fig. 1b). Currently, many of these SNPs are assigned to their closest genes, yet the mechanistic relationships between the genetic variant and the aetiology are in most cases unknown. Clearly, a lot more work needs to be performed to further dissect the genetic functions of these SNPs.

Fig. 1 Summary of the 600 SNPs that have been reported to be associated with a heart disease or trait. These SNPs are associated with about 300 genes. Data were downloaded from the NHGRI GWAS catalogue

The advent of next-generation sequencing (NGS) has facilitated the identification of causal genetic variants given its potential to discover the entire spectrum of sequence variation within a sample. It has revolutionised genomic and genetic analyses and is now becoming the primary discovery tool in human genetics (Cirulli and Goldstein 2010). Sequencing individuals within families, apart from identifying genetic causes of Mendelian disorders, can improve power to detect rare variants that are associated with common diseases, since predisposing variants will be present at a much higher frequency in affected relatives of a proband (Roach et al. 2010). In addition, family-based studies can allow the detection of de novo mutations, parent-of-origin effects as well as gene–gene interactions when affected relatives share two nearby epistatic loci in linkage disequilibrium.

The discovery of extensive copy number variation in the genomes of normal and diseased individuals provided new hypotheses to account for the phenotypic variability among inherited disorders and new leads for the detection of the molecular basis of common complex disorders (Beckmann et al. 2007). Moreover, since most of the single nucleotide variants associated with diseases confer relatively small increments in risk, copy number variation was thought to account for some of the remaining ‘missing’ heritability (Manolio et al. 2009). To address this, the Wellcome Trust Case Control Consortium undertook a large genome-wide study of association between common copy number variants (CNVs) and eight common human diseases (Wellcome Trust Case Control Consortium et al. 2010). CNVs were found to be well tagged by SNPs, while the CNV association analysis was shown to be susceptible to a range of artefacts, which can lead to false positive associations. As a conclusion, common CNVs, which are typed using existing technologies, were found to be unlikely to contribute greatly to the genetic basis of common human diseases. On the other hand, rare CNVs, detected by either array or sequencing technologies, have often been found to contribute to the risk of rare or common cardiac diseases such as CHD (Fakhro et al. 2011; Soemedi et al. 2012), specifically Tetralogy of Fallot (TOF) (Greenway et al. 2009) and early-onset myocardial infarction (Myocardial Infarction Genetics Consortium et al. 2009).

Remarkable progress has been made in understanding the genetic basis of many cardiac diseases, including rare CHD as well as more common complex cardiac diseases. However, advances in technology have now made it possible to interrogate the genetic causes of these diseases at a genome-wide level using the concepts of systems biology.

From molecular biology to systems medicine

The molecular approach to genetics is built on the premise that physiological and cellular phenotype can be explained by the

action of one or more genes. This rationale leads to a strong focus on associating genes with a particular phenotype (e.g., hypertension), biological process (e.g., ageing), or disease (e.g., cardiomyopathy). Commonly, once a candidate gene is identified, it is characterised (e.g., sequenced) and perturbed (e.g., through gene knock-out) to understand its association with the phenotype of interest. Additional experiments might be performed to assess specific functional properties of the gene and the results are then interpreted in the context of existing knowledge about the biological pathways and their relationship with the phenotype of interest.

This idea of associating a disease phenotype to one or several ‘disease genes’ has dominated human genetics for many years. Linus Pauling and colleagues published their seminal paper, ‘Sickle Cell Anemia, a Molecular Disease’, over half a century ago (Pauling and Itano 1949), in which they identified biophysical differences in haemoglobin molecules between healthy individuals and those suffering from sickle cell anaemia. This raised the concept that human disease can be explained by the biophysical properties and action of individual macromolecules (Strasser 1999). As a result of this discovery, the focus of human disease studies shifted towards molecular alterations, an approach that has had an enormous impact on virtually all specialities in medicine (Trent 2012).

Such a gene-centric approach has its merits since it allows us to focus our efforts on the detailed molecular mechanisms involving a small number of genes that may have a direct functional impact on a biological process or disease. However, this gene-centric view of biological processes has its limitations and is potentially misleading, especially in studying complex human diseases where both genes and environmental factors play important roles. A newspaper headline such as ‘scientists found heart disease gene’ is potentially misleading, since it implies that heart disease (which itself is a heterogeneous class of diseases) can be understood by the actions of a gene or a small number of genes alone. One practical limitation of focusing on a small number of genes is the restricted range of questions one can ask about the cellular or organ-level systems properties—such as feedback control, feedforward amplification, redundancy, robustness, hierarchy, and self-organisation. The advances in genome-wide ‘omic’ technologies and computational modelling have enabled us to interrogate a large portion of the genome, transcriptome, and proteome in a much more exploratory manner. This has fuelled many interesting biological questions that we could not previously attempt to answer by reductionist approaches, leading to the emergence of systems biology (Ehrenberg 2003; Ideker et al. 2001; Westerhoff and Palsson 2004).

In a systems biology approach, the first objective is to define the system being studied (i.e., the heart), followed by characterisation of the components of the system. The next step is to examine how these components interact, before inferring the emergent properties of the complex network.

This can uncover hidden relationships and establish global principles (MacLellan et al. 2012).

Systems biology of heart development and disease

The vertebral heart is a complex adaptive biological system, with attributes and organisational features spanning multiple orders of magnitude of space, time and energy requirements. It undoubtedly has many characteristics of complex adaptive systems, including adaptive capacity, inter-component communication, homeostasis, spatiotemporal self-organisation and specialisation, emergent properties and cascading failures.

At the cellular level, cardiomyocytes are studied with heart specific voltage-gated and mechano-sensitive ion channels resulting in a unique electrophysiological profile. Cardiomyocytes are packed full of long chains of myofibrils which provide the fundamental contractile forces needed for the heart to function. These tubular cardiomyocytes are connected to each other via intercalated disks - porous regions containing electrical and mechanical connections that enable the individual cells to communicate. Bundles of aligned cardiomyocytes constitute the extensive muscular mass of the heart, with altered 3D organisation and gene expression signatures depending on location and function (Barth et al. 2005).

The electrical signal that sparks the heartbeat originates at the sinoatrial node, a group of specialised cardiomyocytes. The impulse propagates through the right and left atria causing them to contract and leading to activation of the atrioventricular node. The impulse then spreads via specialised cardiomyocytes called the Purkinje fibres, which act as electrical “superhighways”, coordinating muscular contraction of the ventricles, the main blood pumping chambers. It is this highly organised biomechanical and electrically coupled propagation system that generates the coordinated contraction of the billions of cardiomyocytes that constitute the four human heart chambers robustly over a lifetime.

The vertebral heart is known to be adaptive in response to external stimuli during all stages of life. During heart development, the formation of a complex multi-chambered 3D organ from a collection of multi-potent cells is tightly regulated, but is also easily influenced by external factors, as is evident from the vast literature on cardiac developmental perturbations from multiple species. In the reverse role, the maternal human heart has been observed to undergo eccentric hypertrophy, changes in sphericity and decreased left ventricular strain during pregnancy (Savu et al. 2012). Extensive mechanical and electrical remodelling of the heart is also observed in many types of adult onset heart disease, and indeed reverse remodelling with treatment of disease has been observed (De Jong et al. 2010; Glukhov et al. 2012; Merlo et al. 2011).

In terms of system failures, the healthy heart is in general fairly robust to physiological levels of strain in the short term. However, many pathological environmental factors and diseases will eventually lead to a cascading set of failures, ultimately resulting in a collapse of the system. This often results in a class of clinical phenotypes known as heart failure (HF), characterised by very poor functional output of the heart, shortness of breath and an inability to adapt to exercise. As such, HF is not necessarily the result of an underlying molecular problem, but an inability to maintain homeostasis of the heart’s emergent properties. While risk factors including age, previous myocardial infarcts, smoking, exercise and diet play a strong role in determining the occurrence and severity of HF, clear genetic links between familial occurrence of cardiomyopathies, AF and HF are emerging (MacRae 2010; Hershberger et al. 2013; Hookana 2012).

Perhaps the strongest example of this is sudden cardiac death, which is believed to be caused when a transient event, like intense exercise, meets an underlying disease substrate that may not have previously displayed symptoms. Mendelian inheritance of some cases of disease aetiologies including HCM, DCM, AF, LQTS and CHD is well known (Hershberger et al. 2013; MacRae 2010; Kamisago et al. 2000), and several studies have suggested similar inheritance patterns for sudden cardiac death (Myerburg 2001; Hookana 2012; Deo and Albert 2012). In a post-mortem screening of 173 cases of sudden unexplained death, 26 % showed novel putatively pathogenic mutations in a targeted screen of 5 LQTS disease genes and *RYR2* (Tester et al. 2012). Considering the high hit rate from such a limited screening of only 6 genes, this finding suggests that mutations in cardiac electrical genes are potentially a strong precursor to catastrophic systems failure in the heart.

The final goal of systems biology is often a complete and comprehensive computational model of the complex biology being studied. Such a complete model theoretically allows the functional consequences of any perturbation to the system to be simulated, tracked and characterised. Multi-scale modelling of the heart is an extremely exciting field that has developed immensely over half a century, from simple mathematical models of blood flow volume through the heart, to computationally solving millions of equations simultaneously to model the mechanical forces during a ventricular contraction (Bernabeu Llinares 2011). These nested models include the contribution of ion channels to the action potential within individual cardiomyocytes, the subsequent mechanical forces at the sarcomeric, cellular, tissue and chamber levels, and the propagation of electrical stimuli and contraction throughout the heart (Smaill and Hunter 2010). While computational cardiac models can be personalised from clinical imaging technologies (Wang et al. 2013a), and have been successfully used to recapitulate the contribution of faulty proteins to heart disease phenotypes (Sadrieh et al. 2014), they are still far from

an exhaustive genome-wide model that could be applied as a high-throughput variant prioritisation framework.

With these limitations in mind, our review will focus on systems biology approaches for discovering the genetic causes of CHD and complex heart diseases, from the overwhelming pool of candidate genetic variants identified through whole-genome sequencing and whole-exome sequencing of patients. Although complex heart phenotypes with later onset and milder progression have distinguishably different genetic aetiology compared to rare congenital heart phenotypes, the difficulties of discovering disease-causing variants can be very similar. For example, in family-based NGS studies, large lists of coding variants can be found to be segregating with disease. In such cases, variant prioritisation is necessary to help us identify the disease-causing genes. Conversely, many families will show incomplete penetrance or missing heritability, resulting in a lack of obvious pathogenic mutations after standard filtering. We focus on using systems biology resources and approaches that exist today and are increasingly based on genome-wide ‘omics’ data. We discuss how these methods can be applied for prioritising variants and elucidating the mechanism by which mutations in both coding and non-coding regions affect the whole system and ultimately lead to disease.

Network approaches to decode the genetic causes of heart diseases

It is possible that many inherited heart diseases are a result of allelic variants or damaging mutations, the effects of which are propagated through multiple interconnected molecular levels, rather than only affecting a single gene. While genes associated with disease can be identified, the underlying molecular mechanism in cardiac development and pathogenesis remains largely elusive (He et al. 2011; Lin et al. 2010). Such situations, which entail a large number of molecules spanning across several biological pathways that potentially contribute to pathogenicity, demand a global, system-level view of the functional genetic architecture (Lage et al. 2010) and how its dysregulation leads to disease. Molecular networks are a valuable means of assembling this information, in the form of nodes representing molecules, connected by edges representing interactions, which can be directed or undirected. Networks unify what may appear as disparate biological pathways from single experiments, driving the formulation of novel hypotheses and models to explain the mechanisms of pathogenesis. They can also be applied in the discovery of novel disease ‘network biomarkers’ (Chen and VanBuren 2014) and inform new therapeutic treatments, conceptualised by the idea of ‘network medicine’ (Barabasi et al. 2011).

Many biological networks have been constructed and analysed in the context of heart diseases (Table 1). Protein–protein interaction (PPI) networks utilise proteomics data to

model physical interactions between proteins as undirected edges. Gene co-expression networks are based on the premise that genes that have similar expression profiles across a number of samples are likely to have similar biological functions. These networks are generated from transcriptomic data such as microarrays, where it is expected that co-expressed genes will cluster together. A unifying feature of bioinformatics analysis of these networks is the identification of sub-networks of molecules, commonly referred to as ‘modules’. Investigations of various biological networks have demonstrated that hub molecules (i.e., those that have many incoming or outgoing network connections) are often the most critical for normal cellular function, are closely related to disease, and perturbing them often results in embryonic lethality or pathogenic phenotypes (Dickerson et al. 2011). Further studies have shown that common disease causing genes often cluster together, representing important protein complexes or biological pathways, where the failure of any single component can lead to a similar disease phenotype (Goh et al. 2007). These findings validate the general approach of disease module identification (Barabasi et al. 2011).

Systems biology approaches have previously been used in the study of cardiac development and diseases, although not as prevalently as one might expect (Sperling 2011). Berger et al. (2010) used PPI data to investigate the functional neighbourhood of 13 known LQTS genes. By building and analysing the PPI network around known disease genes, they were able to rank molecules, diseases and drugs in relation to LQTS. They found that the LQTS neighbourhood represented the convergence of cardiac channelopathies in the diseasome. They also found that the gene targets of drugs designed as QT-prolonging and drugs that have an undesired QT-related side effect were enriched in the LQTS neighbourhood. Furthermore, they were able to use the ranked LQTS neighbourhood proteins to classify drugs with adverse QT-related side effects with reasonable accuracy, based on the FDA Adverse Event Reporting System database.

Lage et al. (2010) used phenotype annotations from gene knock-out mouse models to annotate 255 genes into 19 spatio-temporal cardiac developmental phenotype modules (Lage et al. 2010). Based on PPI data, they identified 49 statistically significant novel heart development candidate genes linked to one or more modules. Using immunohistochemistry, they were able to validate that 11 of 12 tested novel candidates were in fact expressed in the specific heart tissue and at the developmental time predicted by their assigned phenotype module.

Modules form the basic functional unit of most networks; however, the methods used to derive these differ substantially (Table 1). These methods generally restrict a gene to be part of one module only, which may not be a true reflection of its activity if it performs multiple functions in multiple pathways. A novel clustering algorithm, recently developed to construct

Table 1 A summary of published heart disease studies that involve construction and analyses of protein–protein interaction networks and gene co-expression networks

Disease	Data used to construct network	Network construction method	Network analyses	Reference
19 heart phenotypes / diseases	PPI, cardiac developmental genes	Phenotype specific networks, literature based module annotation	Complexity measure, temporal phenotype ordering, experimental validation of novel candidate genes	Lage et al. 2010
CHD	PPI, gene sets for CHD genetic and environmental risk factors and responses	Networks for separate anatomical structures as described in Lage et al. 2010	Network permutation tests	Lage et al. 2012
DCM	PPI, DCM gene expression profiles, GO annotations	Pearson's correlation between co-expressions and GO annotations to identify DCM and non-DCM modules	Identification of hub genes, topological measures, functional enrichment of modules, performance as disease classifier	Lin et al. 2010
AF	Human microarray profiles grouped by disease state	Weighted gene co-expression network analysis and connectivity to derive modules	Module preservation, PCA of modules to identify “eigengenes”, identification of hub genes, functional enrichment using IPA	Tan et al. 2013
Cardiac hypertrophy and heart failure	Microarrays of mouse myocardium	Weighted gene co-expression network analysis and topological overlap to derive modules	Identification of hub genes, topology, PCA of modules to identify “eigengenes”, GO annotation, module preservation	Dewey et al. 2011
MI	Microarrays from mice with MI	Co-expression and topological modules	GO annotation, hub gene (<i>Col5a2</i>) as disease classifier	Azuaje et al. 2013
LQTS	PPI, disease genes	Immediate neighbours	Ranking genes, diseases and drugs based on LQTS neighbourhood	Berger et al. 2010

CHD Congenital heart disease, *DCM* dilated cardiomyopathy, *AF* atrial fibrillation, *MI* myocardial infarction, *LQTS* long QT syndrome, *PPI* protein–protein interaction network, *PCA* principal component analysis, *GO* gene ontology, *IPA* ingenuity pathway analysis

a transcriptional regulatory network of mouse heart development, permits genes to participate in multiple modules (Chen and VanBuren 2014). Other limitations of network-based approaches include heterogeneity of available datasets and indeed changes in the underlying biology across different tissues and organisms, and the generalisability of animal models to humans (Dewey et al. 2011; Tan et al. 2013). Furthermore, predictions of novel disease genes are made through computational analysis of networks, and thus need to be experimentally validated in animal models.

Whilst networks based on PPI and gene co-expression data are useful for identifying genes associated with diseases, to understand the underlying mechanism we need to model causality using directed networks such as gene regulatory networks (GRNs). Causal relationships can be robustly inferred through perturbation of molecules, allowing nodes to assume roles of ‘regulator’ and ‘target’ depending on the direction of the interaction. This type of data can come from large-scale ‘omic’ studies, but they could also come from the

vast collection of previously published molecular biology data probing the effect of perturbing individual genes and proteins. In fact, one effective way to gain an understanding of a causal GRN is by piecing together many previously experimentally identified molecular interactions. In other words, we want to discover the truth based on many previous discoveries, as epitomised by Sir Isaac Newton’s famous quote - “If I have seen further, it is by standing on the shoulders of giants”.

CardiacCode: a manually-curated resource for cardiac-specific GRN analysis

There is a vast amount of high quality genetic or molecular perturbation data in the published literature that largely remains computationally inaccessible—mostly hidden in figures, tables or text in developmental biology papers. We manually collected over 700 pieces of genetic perturbation evidence from 43 published primary research papers on

in vivo mouse cardiac development, from embryonic day E6.5–E13.5 heart tissues. We integrated this data into a causal GRN using a statistical model. The network consisted of 280 unique edges between 33 regulators and 129 target molecules. The majority (59 %) of these edges represent activating regulatory relationships, 13 % are inhibitory and 28 % appear to have no-causal effect. We complemented this with 86 microarray expression profiles from the GEO database. The curated perturbation dataset, the assembled microarray data, and the inferred cardiac development network can be accessed through our newly developed interactive web resource, CardiacCode (<http://CardiacCode.victorchang.edu.au/>; Fig. 2a). The resource is built on an SQL database and interfaces with Javascript and HTML5 through PHP. The network visualisation and analysis tools are supported by the cytoscape.js plugin and Python scripts running on the server.

The CardiacCode network represents the core knowledge about the causal gene regulatory interactions that drive mammalian cardiac development. Key cardiac markers and transcription factors (TFs), including *Nkx2-5*, *Gata4/6*, *Tbx1/5/20*, *Isl1* and *Mef2c*, are the hub genes in the network, potentially reflecting their important roles in the network or the widespread studies of these genes. Interactions between many of these key TFs have been shown to be critical for cardiac development and foetal viability. *Nkx2-5* is a key regulator of heart morphogenesis (Lyons et al. 1995) and differentiation of cardiomyocytes (Tanaka et al. 1999). *Gata* TFs are required for the formation of the primitive heart tube (Molkentin et al. 1997) and myocardium maturation (Peterkin et al. 2003). *Tbx* TFs are crucial for morphogenesis of the outflow tract and aorto-pulmonary septum (Xu et al. 2004) and chamber differentiation (Stennard et al. 2005). *Isl1* is required for outflow tract and right ventricle formation (Cai et al. 2003), and *Mef2c* plays a role in heart tube looping (Lin et al. 1997). It is therefore not surprising that the breakdown of any link in this core regulatory network has been shown to cause CHD (McCulley and Black 2012; Schlesinger et al. 2011).

Many critical signalling pathways are also represented as molecules from the WNT, BMP, SHH, NOTCH, FGF, TGF-beta and other growth factor families that have been linked to CHD, as well as their receptors (Li et al. 2014). CardiacCode also encodes regulatory knowledge about known CHD genes, including the key TFs and markers mentioned before, as well as *Myocd*, *Foxh1*, *Myh6*, *Actc1*, *Nodal* and *Notch1*.

The CardiacCode website has a growing suite of inbuilt analysis options. Apart from node position, three node attributes can be customised using uploaded datasets: node colour, node border colour and node shape. By default, the known CHD genes are coloured red. These colour and

shape features allow the user to simultaneously visualise continuous values such as gene expression levels and rankings, or discrete values such as module membership, gene ontology (GO) annotation, disease association and candidate disease gene status. Therefore, CardiacCode can be very useful in facilitating disease gene prioritisation in NGS studies where a large number of candidate genes are found.

Integrative analysis for disease gene prioritisation

A whole-exome sequencing experiment returns possibly hundreds of predicted deleterious variants, and thus there is an urgent need to intelligently prioritise these variants for functional investigation in the laboratory. In the past, gene prioritisation was done manually and relied mainly on individual, domain-specific expertise. This approach is clearly hard to scale up to the genome-wide level. Most current variant prioritisation pipelines will compare the results to a disease database such as OMIM, and try to focus the search at the gene level. Unfortunately, this heavily biases the results towards what is already known about the disease, and many cases remain unsolved, with no obvious candidates. Systems biology-based variant prioritisation holds a lot of promise to improve this situation. By harnessing genome-scale biological knowledge, we can perform unbiased analyses and potentially identify new mechanisms underlying disease. Emerging forms of systems-level prioritisation are based on a combination of network analysis, gene or protein expression data, and literature curated databases (Fig. 2b).

Networks can be built from a variety of biological information, such as PPI (Chatr-aryamontri et al. 2012; Franceschini et al. 2012), gene expression and other functional relationships (Franceschini et al. 2012) and metabolic pathway information (Kanehisa et al. 2013; Kelder et al. 2011). The majority of network based approaches for gene prioritisation stem from the ‘guilt-by-association’ principle (Fig. 2c). This involves ranking candidate genes by their ‘closeness’ to a set of seed genes, commonly disease-causing genes or disease modules, using measures such as shortest path (Krauthammer et al. 2004), random-walk and diffusion kernel (Kohler et al. 2008). Several tool kits are freely available to perform these topological analyses, including the *igraph* package (Csardi and Nepusz 2006) in R and *NetworkPrioritiser* (Kacprowski et al. 2013) in Cytoscape.

Expression-based prioritisation approaches generally rank genes according to differential gene or protein expression between diseased and non-diseased tissues. Apart from the standard differential expression (DE) approach, several alternative expression-based methods exist. Differential variability analysis captures disease-induced changes in regulatory control that might be missed by DE analysis (Ho et al. 2008), and

Fig. 2 Gene prioritisation methods using multiple sources of biological information. **a** A screenshot of a cardiac-specific GRN from the CardiacCode website. Nodes are coloured based on differential expression of mouse E10.5 heart against whole embryo body (WB). Node shape is related to whether a gene is known to cause congenital heart disease. **b** Gene set enrichment of the 200 most highly specific (or expressed) genes at E10.5 and E11.5 for whole heart and at E12.5–E18.5 for ventricles and atrial chambers, with (or without) normalisation against WB. The gene set enrichment results demonstrate that the normalisation of cardiac-specific

gene expression patterns from WB can increase the enrichment of cardiac-specific gene sets, and decrease the enrichment of housekeeping gene sets. This suggests that WB normalisation may be a simple method for enriching for potential cardiac disease genes. **c** A PPI network showing the differential clustering of AF GWAS genes and known CHD genes in the human protein interactome. Candidate disease genes were prioritised based on proximity to all known disease genes for each of the two diseases. The sub-network containing the connections between the combined top 100 ranked candidates from each disease is shown

may yield important information about the connectivity or other characteristics of the underlying GRN (Mar et al. 2011; Padovan-Merhar and Raj 2013). The exact nature of the relationship between gene expression variability and the

dynamics of the underlying GRN is still being actively investigated, but we expect up-and-coming single-cell genome-wide technologies can further clarify the relationships. Another approach is to ‘subtract away’ the baseline gene expression

in various tissues in the embryos by considering genes that are differentially expressed in a specific tissue compared to the whole embryo (Fig. 2b). This approach has been shown to be very useful in prioritising cataract genes using embryonic ocular lens gene expression data (Lachke et al. 2012). A new resource for murine heart development gene expression data from E7.5 all the way to adult tissues has recently been published (Li et al. 2014), which in combination with other published data will provide a foundation for expression-based prioritisation for human heart disease genes in the future.

More recently, several groups have developed gene prioritisation tools that combine multiple data types and analysis approaches in order to better rank candidate genes. Barriot et al. (2010) created an online collaborative wiki-based resource for studying CHD that implements a mixed data gene prioritisation analysis. They implement the ENDEAVOUR algorithm (Tranchevent et al. 2008) and allow the user to specify data sources including GO annotations, PPIs, cis-regulatory information, gene expression datasets, sequence information and text-mining data. ENDEAVOUR was applied to prioritise genes located in an 850-kb locus associated with CHD, using gene expression microarrays and gene homology data (Thienpont et al. 2010). Seven training sets of genes representing different cardiac developmental and genetic phenotypes were used as input to the algorithm, which identified *TAB2* as the top-ranking gene from all 105 gene candidates. The role of this gene in heart development was experimentally verified using knock-downs in zebrafish which led to heart defects, and mutation analysis in 402 patients with CHD revealed a translocation disrupting *TAB2* that cosegregated with familial CHD (Thienpont et al. 2010). ENDEAVOUR was also applied to study two patients with CHD, identifying *CRKL* and *MAPK1* as the most likely causal genes out of those lost in a large deleted region of the genome (Breckpot et al. 2012).

Another tool, Gentrepid, applies two algorithms to prioritise candidate disease genes (Ballouz et al. 2013). Common Pathway Scanning (CPS) utilises PPI and pathway data to determine relationships between candidate and disease genes based on membership in the same protein complex or pathway. Common Module Profiling (CMP) performs sequence comparisons of Pfam domains to assess functional similarity of candidates with disease genes. Gentrepid was applied to GWAS data on hypertension (Ballouz et al. 2013) and coronary artery disease (Ballouz et al. 2014), identifying disease genes that were consistent with those previously reported as well as several novel candidates.

Li et al. (2013) built three cardiomyopathy subtype-specific networks [DCM, Arrhythmogenic Right Ventricular Cardiomyopathy (ARVC), HCM] based on the PPI neighbours of known disease proteins from the OMIM database. They used the STRING PPI or functional link database which provides a confidence of interaction score, and combined this with gene

ontology similarity between proteins to weight each edge of the network. Based on the weights of a protein's immediate neighbourhood, along with the number of disease genes it neighboured, each protein in the network received a disease relevance score and hence a rank. From the top 50 DCM candidate proteins, 9 have been previously linked to DCM in the literature; 7 additional candidates were directly associated with other cardiomyopathies such as HCM, and 4 additional proteins were related to cardiac arrhythmias, cardiac dysfunction and cardiac cell damage. Similar results were achieved for the HCM and ARVC candidates.

A web resource (<http://www.esat.kuleuven.be/gpp>) developed by Tranchevent et al. (2010) contains detailed information about a large selection of gene prioritisation tools to help researchers decide which tool is most appropriate for their analysis. While existing tools are very useful in disease gene discovery, they are predominately gene-centric. Interactions between genes, microRNAs (miRNAs), non-coding RNA (ncRNAs), disruption of transcription factor binding sites and epigenomic modifications have all been found to play a role in heart development and disease, and emerging knowledge about these processes needs to be incorporated into future gene prioritisation efforts (Schlesinger et al. 2011; Klattenhoff et al. 2013; Matkovich et al. 2014; Smemo et al. 2012; Zaidi et al. 2013).

Beyond protein-coding genes: Non-coding RNA in heart development and diseases

The widespread application of high-throughput RNA sequencing reveals the complexity and diversity of ncRNAs, both large and small. The best-studied small ncRNA class are miRNAs, which are processed in two successive steps from longer precursor transcripts by type III RNase enzymes called Drosha and Dicer. Conditional deletion of Dicer revealed that miRNAs are essential in heart development and disease (da Costa Martins et al. 2008; Saxena and Tabin 2010). Targeted studies have been able to identify specific miRNAs that can perturb cardiac development and disease. Intriguingly, these miRNAs were originally identified from their enriched expression in adult hearts and other genomic associations. For example, miR-1 and miR-133, essential for heart development (Liu et al. 2008; Zhao et al. 2007), were originally identified as being highly expressed in a cell model of skeletal muscle differentiation as well as adult skeletal and heart muscle (Chen et al. 2006). miR-208 and family members are differentially expressed in embryonic and adult tissues (Callis et al. 2009), and were first identified to be encoded in introns of myosin heavy chain cardiac-specific genes (van Rooij et al. 2007). A list of 139 (129 remain listed in miRBase v.21) active, functional cardiac miRNAs was collated from multiple studies (Matkovich et al. 2011), in which many were

identified from microarray studies where their expression correlated to a cardiac disease state in adult tissues. In light of this, they are not likely to be exhaustive for consideration in developmental cardiac analyses. Gene network analysis has proven to be a useful methodology that can identify miRNAs that are involved in cardiac function. For example, knock-down of *Srf* resulted in differential expression of mRNA of which ~10 % could be explained from ChIP–chip interaction data. Analysis of the targets of 42 differentially expressed miRNAs, however, could explain 45 % of the differentially expressed mRNAs (Schlesinger et al. 2011).

Long ncRNA (lncRNAs) are transcripts that have a length greater than 200 nucleotides and have no defined open reading frame. Interestingly, lncRNAs share processing properties of mRNAs in that many have a 5' cap, are spliced, and have a poly(A) tail. For this reason, there are no laboratory-based high-throughput sequencing protocols that enrich for lncRNAs. Instead, lncRNAs are analysed in either poly(A)-enriched or ribosome-depleted RNA-Seq datasets. Poly(A)-enriched libraries typically have fewer intronic sequence reads compared to ribo-depleted ones, and for this reason are more suitable for novel lncRNA identification by transcript assembly. However, it is now possible to analyse lncRNA transcripts by utilising curated annotated ncRNA databases. One well-maintained database that catalogues ncRNA is NONCODE, and version 4 lists 56,018 human lncRNA genes (Xie et al. 2014). Matkovich et al. (2014) curated a refined list of cardiac lncRNAs from such online resources and used this to identify 157 lncRNA differentially expressed between embryonic and adult hearts.

A significant challenge in integrating the analysis of lncRNAs is to define their function. Unlike miRNAs, many lncRNAs are not conserved between species. Recently, a number of nuclear localised lncRNA transcripts have been identified to have key functional roles in heart development. *Bvht* is an example of a non-conserved lncRNA transcript only identified as being expressed in the mouse. *Bvht* regulates SUZ12, which encodes a component of the Polycomb Repressive Complex 2 (PRC2), which in turn epigenetically regulates the transcription of key cardiac genes (*Mesp1*, *Gata6*, *Hand1*, *Hand2*, *Nkx2-5*) that are important for specifying cardiac progenitors (Klattenhoff et al. 2013). *Fendrr* is an example of a conserved lncRNA whose transcript has been identified in both human and mouse and is transcribed on the opposite strand ~1 kb downstream of *Foxf1*. *Fendrr* was found to be essential for heart and body wall development and reduced the occupancy of PRC2 at promoter regions of *Foxf1* and other transcription factors involved in the formation and differentiation of lateral mesoderm lineage (Grote et al. 2013). The modulating activity of PRC2 by lncRNAs is a recurring theme and makes them ideal candidates to be incorporated into gene regulatory network analysis. For example, the reciprocal expression of lncRNAs to adjacent coding

mRNAs and subsequent network analysis identified a potential role for lncRNAs in modulating NF κ B- and CREB1-regulated genes during embryonic heart growth (Matkovich et al. 2014). Our understanding of the regulatory potential of lncRNAs in gene networks is in its infancy and it is likely that major insights are soon to be gained in this field.

Epigenomics of heart diseases

Eukaryotic genomes are packaged into chromatin through interactions with histone and non-histone chromosomal proteins along the entire length of DNA. The structure of chromatin is highly dynamic and is regulated through covalent modifications of histones and other forms of chromatin remodelling during heart development and pathogenesis (Chang et al. 2004; Han et al. 2011), thus conferring an additional level of control of gene expression. Heart development involves intricate programs of gene expression that differ across different cell types and stages. This necessitates looking deeper into chromatin structure and regulatory elements such as enhancers to elucidate the regulatory mechanisms in cardiac development and disease. This urgency is exemplified by the discovery that heart diseases can arise from mutations in non-coding, regulatory regions (Smemo et al. 2012).

Histone modifications involve the addition of chemical groups (e.g., methyl, acetyl, phosphate) to the N-terminus of histone proteins. These modifications, individually or in combination, often correlate with specific genomic features such as promoters, enhancers, transcribed regions, Polycomb-associated domains, and heterochromatin. For example, active regions such as promoters are typically marked by trimethylation of lysine 4 of histone H3 (H3K4me3) and acetylation of lysine 27 of histone H3 (H3K27ac), while repressed regions such as heterochromatin are marked by H3K9me3. Histone modifications are catalysed by histone-modifying enzymes including methyltransferases and acetyltransferases, while other groups of proteins read and remove these modifications.

Zaidi et al. (2013) performed exome sequencing of children with severe CHD and their parents, identifying de novo missense, frameshift and truncation mutations in genes that deposit, read and remove H3K4 methylation. These mutations were located in the *MLL2* methyltransferase, *KDM5A* and *KDM5B* demethylases, the *H2BK120* ubiquitination complex and the *CHD7* helicase. Mutations were also found in *SMAD2*, which encodes a protein that binds chromatin and leads to downstream demethylation of H3K27, facilitating increased activation of normally repressed genes (Zaidi et al. 2013).

Several studies have implicated the *Ezh2* subunit of the methyltransferase PRC2, which trimethylates H3K27, in heart development and disease. In mouse E12.5 ventricles, ChIP-

qPCR revealed enrichment of *Ezh2* and H3K27me₃ at the promoters of the transcription factors *Six1*, *Pax6* and *Isl1* (He et al. 2012). This was concordant with derepression of these genes when *Ezh2* was conditionally knocked-out in mouse embryonic hearts. The inactivation of *Ezh2* also resulted in lethal congenital heart defects including myocardial hypoplasia and ventricular septal defect (He et al. 2012). In a similar study, *Ezh2* knock-out in adult mice hearts revealed a decrease in H3K27me₃ at the promoter of the transcription factor *Six1*, which was accompanied by an increase in PolII binding (Delgado-Olguin et al. 2012). This derepression of *Six1* was associated with activation of skeletal muscle genes and hypertrophy of cardiomyocytes. Thus, epigenetic repression by *Ezh2* is established during embryonic development and persists in late development (He et al. 2012) and postnatally to maintain the normal phenotype in adult hearts (Delgado-Olguin et al. 2012).

In mouse postnatal hearts with pathologically induced cardiac hypertrophy, overexpression of the *JMJD2A* demethylase resulted in increased expression of foetal cardiac genes such as *Myh7* and the disruption of cardiac function (Zhang et al. 2011). The re-occurrence of these results in human patients with cardiac hypertrophy, who also displayed increased levels of *JMJD2A*, is suggestive of its role in aggravating cardiac hypertrophy in pathological conditions (Zhang et al. 2011).

Histone acetyltransferases (HATs) and deacetylases (HDACs) have been shown to play critical roles in cardiac development and disease. Through the addition of acetyl groups to histone molecules, HATs mark regions associated with active transcriptional activity. The HAT activity of *CBP/p300* leads to a hypertrophic phenotype when overexpressed in cultured ventricular myocytes taken from neonatal rats (Gusterson et al. 2003). In transgenic adult mice subjected to myocardial infarction, overexpression of a mutant form of *p300* lacking the HAT domain prevented extensive left

ventricular remodelling, a process leading to heart failure, when compared to overexpression of the normal *p300* (Miyamoto et al. 2006). HDACs remove acetyl groups, and are therefore associated with repression of genes. Mice with deletion of *Hdac3* specifically in the heart displayed extensive and lethal cardiac hypertrophy (Montgomery et al. 2008). Single deletions of *Hdac5* and *Hdac9* led to an exacerbated hypertrophic response in mouse models, while double deletion of these genes increased susceptibility to lethal ventricular septal defects (Chang et al. 2004).

An additional valuable application of histone modification data is genome-wide ‘chromatin state’ annotation, which can be applied to discover candidate regulatory elements such as enhancers, especially those associated with a disease (Fig. 3). The concept of chromatin states arises from the observation that of the possible combinations of histone modifications, only a subset is confidently observed within a cell. Each unique combination of histone modifications becomes a chromatin state, which can then be used to functionally annotate the entire genome. Computational tools have been developed for this task, such as ChromHMM (Ernst and Kellis 2012) and Spectacle (Song and Chen 2014), which are based on Hidden Markov Models. We recently developed a new method, called hiHMM, which can jointly learn a chromatin state definition across multiple samples (Ho et al. 2014).

Chromatin remodelling involves reorganisation of nucleosomes and affects DNA accessibility, particularly the ability of transcriptional activators and repressors, and other proteins such as RNA polymerase II, to bind to DNA during transcription (Han et al. 2011). Deletion of *Brg1*, the major ATPase subunit of the BAF chromatin remodelling complex, in mouse myocardium, reduced the incidence of cardiac hypertrophy, thus making *Brg1* an attractive therapeutic target (Hang et al. 2010). Allelic imbalance between *Brg1* and cardiac

Fig. 3 Analysis of human genome-wide chromatin landscape of cardiac cells can be used to identify cardiac-specific enhancers that are associated with congenital heart disease. ChIP-seq of histone modifications in human left ventricular heart tissue (GEO Accession: GSE16256) shows the presence of enhancer-associated marks H3K27ac and H3K4me1

surrounding a SNP that is associated with a congenital heart disease. This SNP may exert its function through affecting this enhancer, which is approximately 90 kb downstream of the *TBX5* gene (Smemo et al. 2012). This example illustrates how disease-associated enhancers can be identified using ChIP-seq

transcription factors (*Nkx2-5*, *Tbx20*, *Tbx5*) also results in heart defects (Takeuchi et al. 2011).

Working in synergy with histone modifications and chromatin remodelling, DNA methylation also acts as a key epigenetic regulator of gene expression during development and disease. During cardiogenesis, differential DNA methylation is believed to regulate the expression of specific cardiac developmental genes (Chamberlain et al. 2014). In particular, the expression of *Has2*, required for proper heart valve formation, is predicted to be controlled by methylation of its enhancer by DNA methyltransferase 3b (Chamberlain et al. 2014). In patients with TOF, DNA methylation profiling revealed greater methylation of the promoters of genes associated with CHD, which corresponded to decreased mRNA expression levels (Sheng et al. 2014). Another key player believed to contribute to the pathogenesis of TOF is retinoid X receptor α (*RXR α*), a member of the retinoic acid signalling pathway (Zhang et al. 2014). Patients with TOF exhibited significantly less *RXR α* mRNA in their right ventricular outflow tract myocardium, which was likely due to an observed increase in methylation of the *RXR α* promoter (Zhang et al. 2014). Similarly, in paediatric patients with CHD, attenuated transcriptional activity was hypothesised to be caused by hypermethylation of the promoter of the *CITED2* gene, which encodes the transcriptional co-activators CBP/p300 (Xu et al. 2014). Regions of methylated DNA are read by proteins such as MeCP2, whose overexpression in mice was associated with hypertrophy of the septum in embryonic hearts, resulting in lethality around E14.5 (Alvarez-Saavedra et al. 2010).

Epigenetic mechanisms are heritable and affected by nutritional and environmental factors (Udali et al. 2013). It is the ability of epigenetic mechanisms to be stably transmitted across cell divisions and generations of offspring that creates the diversity of multi-cellular complex organisms. However, epigenetic inheritance also introduces potential points of system dysregulation, leading to disease. Differential methylation of heart development genes *EGFR* and *GATA4* in mothers can result in CHD in their offspring (Chowdhury et al. 2011). Furthermore, chronic treatment of pregnant rats with hypoxia directly resulted in methylation and subsequent repression of the *PKC-epsilon* gene promoter (Patterson et al. 2010). This was associated with greater susceptibility to cardiac ischemia and injury in adult offspring (Patterson et al. 2010). Evidence for trans-generational epigenetic inheritance has also been demonstrated in mice, where mutations in the *Mtrr* (methionine synthase reductase) enzyme in maternal grandparents led to detrimental uterine environmental conditions in their wild-type daughters (Padmanabhan et al. 2013). This in turn resulted in growth deficiencies and congenital defects of various organs, including the heart, in grandprogeny, that was independent of the maternal genotype (Padmanabhan et al. 2013). Together, this alludes to the idea that heart diseases can occur as a result of both genetic and environmental risk factors that

have been inherited from previous generations. Thus, a complete understanding of all potential pathogenic mechanisms is warranted to develop therapeutic strategies and reduce the incidence of heart disease in current and future generations.

We have compiled a summary of published epigenomic datasets related to heart development and disease which is available at <http://cardiaccode.victorchang.edu.au/download.php>.

Cis-regulatory modules of transcription factor binding sites

While the genomic sequence is largely constant throughout a multicellular organism, each cell interprets its genome differently. This is mediated by the different ways in which a combination of transcription factors (TFs) binds to the genome at TF-specific sequence motifs in non-coding DNA regulatory elements, termed cis-regulatory modules (CRMs). Elucidation of the location, state, and function of all CRMs throughout the genome would be a great boon to research. In particular, knowledge of the CRMs that regulate gene expression during heart development would provide a basis for understanding how genetic mutations in non-coding DNA contribute to heart diseases. As an example, ChIP-seq for cardiac TFs has identified putative heart CRMs, upstream of *Sox9*, that are deleted in CHD (Sanchez-Castro et al. 2013).

A particularly fruitful approach for identification of enhancers is performing ChIP-seq to identify binding sites for the enhancer-associated coactivators p300 or CBP. When carried out on a specific tissue or homogeneous cellular population, this method enables the direct observation of (presumably) active CRMs genome wide. Blow et al. (2010) utilised this approach to catalogue enhancers from murine forebrain, midbrain, limb, and heart, and, particularly in the case of heart CRMs, demonstrated that a reasonable proportion (~20 %) showed no detectable evolutionary constraint. Similarly, May et al. (2012) identified human cardiac CRMs from foetal and adult heart tissues. Predictably, both studies showed enrichment of CRMs near genes that are active within the respective tissue. A total of 65 predicted CRMs were selected based on high p300/CBP binding and validated in vivo using transgenic mouse models, in which 43 resulted in observable reporter expression just in the heart. Surprisingly, this result was irrespective of sequence constraint and p300/CBP binding in mice, indicating that this approach is unlikely to characterise all active enhancers. Notably, a study of histone marks in cardiac hypertrophy identified a set of candidate enhancers, based on H3K27ac, that was distinct from p300 (Papait et al. 2013). A study by He et al. (2011) provides clearer evidence that p300/CBP is not a ubiquitous marker for all active enhancers in mouse cardiomyocytes. In

their study, ChIP-seq of cardiac transcription factors Nkx2-5, Gata4, Tbx5, Srf, and Mef2A, identified a large set of CRMs (1,715), of which only ~20 % overlapped with a similarly sized set identified from the binding profile of p300. The CRMs showed some functional segregation with type: those found by multiple TF binding tended to be more proximal to transcription start sites than those bound by p300, and were enriched for different GO terms (He et al. 2011). Interestingly, some combinations of TFs are much more common than others, suggesting that they are more preferentially used to confer cell-type specificity (Fig. 4). In a similar study, ChIP pull-down of cardiac TFs in *Drosophila melanogaster* found that CRMs tend to either be occupied by one or two cardiac transcription factors, or by all of them, suggesting that combinatorial TF interactions are particularly important in affecting CRM function (Junion et al. 2012).

In addition to experimental techniques, several computational methods for predicting enhancers have also been developed. Narlikar et al. (2010) utilised validated heart enhancers compiled from the literature to train a linear regression classifier based on the presence of known TF binding sites and de novo prediction of motifs. This approach led to the identification of 41,930 potential human heart enhancers, of which 16 out of the 26 tested were experimentally verified as heart enhancers in transgenic zebrafish (Narlikar et al. 2010).

The previous study predicted novel enhancers based on sequence features such as transcription factor binding sites. Erwin et al. (2014) instead demonstrate the importance of combining multiple and diverse datasets obtained from evolutionary conservation, epigenetic landscape (chromatin and histone modifications), transcription factor binding sites and sequence motifs across different tissues and cell types. The greater power and confidence of predictions enabled by this

integrative approach is harnessed in EnhancerFinder, a classifier based on multiple kernel learning, which predicts enhancers in the context of their specific tissue (Erwin et al. 2014). It was observed that heart enhancers were predicted with greater ease compared to enhancers from other tissues, which may be attributed to higher GC content, greater proximity to transcriptional start sites of nearest genes and less evolutionary conservation (Erwin et al. 2014). Such machine learning approaches utilise the presence of features for classification of enhancers. While useful to a certain extent, many enhancers have been observed to vary significantly between species despite sharing conserved activity. This suggests that there may be a higher order organisation or complexity to TF binding arrangement that may not be effectively captured by a static, feature-based analysis. A recent, novel approach for CRM identification looks at conserved TF binding modules throughout evolution, rather than direct sequence conservation to identify regulatory regions associated with diabetes (Claussnitzer et al. 2014).

Several other recent studies (Alekseyenko et al. 2012; Sherwood et al. 2014; Smith et al. 2013; White et al. 2013) suggest that the particular combination of TF motif arrangement, along with chromatin environment, histone modification and local GC content, determines cell-type-specific TF binding, which in turn affects cell-type-specific gene expression. These observations are suggestive of a ‘grammar’, or a set of rules, that defines cell-type-specific TF binding (Fig. 4). State-of-the-art probabilistic grammar modelling methods can be applied to ‘decode’ this grammar, thus allowing us to exploit a wide range of tools originally developed for computational linguistics or machine learning applications. One major advantage of this linguistic approach is that, once a grammar is learned, we can use it to generate new meaningful ‘sentences’—new DNA regulatory sequences. This gives rise

Fig. 4 The observed/expected (obs/exp) ratio of different co-occupancy patterns of cardiac transcription factors (green) in mouse cardiomyocyte HL-1 cell line (He et al. 2011). There is a strong enrichment (red, obs/exp > 2) for co-occupancy of all 5 TFs, and several combinations of TFs,

while other TF combinations are strongly depleted (blue, obs/exp < 0.5). The presence of cooperative binding of specific combinations of cardiac TFs suggests that a cell-type specific combinatorial TF interaction, or a ‘TF grammar’, may be involved in cardiac gene regulation

to an attractive prospect of designing and synthesising de novo regulatory elements from first principles. This novel approach has immense implications for the rational design of regulatory DNA sequences for gene therapy in the future.

Future directions

The advancement of microarray and NGS technologies underlies the widespread adoption of GWAS and whole-genome sequencing for disease gene discovery. These technologies have fundamentally changed the way we do research and have enabled us to address systems-level biological questions at a genome-wide scale. More recently, two innovative technologies are promising to transform the landscape of biomedical research: single cell genomic analysis, and CRISPR-Cas9-based genomic editing.

Single-cell genomic technology was named the Method of the Year in 2013 by the journal *Nature Methods*. This innovation means that we can now perform genome-wide genomic and transcriptomic sequencing at the level of individual cells, instead of averaging the signals from a heterogeneous cell population. This has facilitated the fine-scale identification of gene expression status during cellular differentiation (Trapnell et al. 2014). Highly parallel single-cell qPCR has found that there is greater variability in gene expression between individual B lymphocyte cells than between different human donors, and that this variation has been largely masked by lower experimental resolution in previous studies (Wills et al. 2013).

Existing microarray and NGS profiling techniques are applicable to millions of cells. Therefore, it is important to recognise the potential effect of averaging molecular profiles from heterogeneous cells, which is potentially a problem in analysing primary heart tissues. It has been suggested that a single-cell approach would be beneficial for the study of heart development and pathology (Sperling 2011). The human heart consists of a mixture of cell types; the most prominent including fibroblasts, myocytes, endothelial cells, and epicardial cells. These cells interact with one another during development, and each play a key role in cardiac development, repair, and pathogenesis (Deb 2014; de la Pompa and Epstein 2012; Runyan and Markwald 1983). Although the application of this technology to the heart is in its infancy, in an organ where defects in a single cell can trigger a life-threatening arrhythmia, the ability to investigate cardiac cells at this resolution might prove to be the driver of the next major breakthrough.

Clustered regularly interspaced short palindromic repeats (CRISPR)-associated (Cas) (CRISPR-Cas) is an exciting new technology that exploits a bacterial defence mechanism against viruses and plasmids to perform genome editing with unprecedented precision and efficiency (Ran et al. 2013). The

CRISPR-Cas mechanism relies on small RNAs to act as sequence-specific templates and recognition factors to cleave and silence undesired nucleotides. Recently, the CRISPR mechanism has been demonstrated as a viable alternative to siRNA to knock-down expression of almost any gene in many model organisms important in heart disease research, including zebrafish, mouse, goat and human (Wu et al. 2013). Jinek et al. (2012) discovered that a single chimeric short-guide RNA (sgRNA) molecule can program the Cas9 protein to cleave double-stranded DNA with single base-pair accuracy. Because it directly modifies the DNA of the cell, the effect of the gene knock-out persists through cell division, unlike siRNA interventions. Since then, CRISPR-Cas9 has also been used to knock-in an inducible protein knock-down, by inserting a *Shield1* conditional destabilisation domain into a target protein (Park et al. 2014).

The ability to edit DNA with single base pair accuracy means that CRISPR-Cas9 can be used to systematically investigate functional effects of mutations anywhere in the genome, including non-coding regulatory regions, in a much more efficient workflow than cloning and breeding approaches (Dickel et al. 2014). This should greatly improve the pace of discovery of functionally active cell-type-specific regulatory elements in mammalian systems. By creating libraries of the sgRNAs that direct the Cas9 protein, genome-scale knock-down screenings in human cell lines have systematically perturbed thousands of genes in a robust and lasting way with minimal off-target effects (Shalem et al. 2014; Wang et al. 2014). The recently demonstrated ability to introduce multiple mutations in vivo in a single step will further facilitate research into complex polygenic diseases including heart diseases (Wang et al. 2013b). This possibility opens up new avenues for large-scale genetic perturbation experiments that will greatly enable inference of tissue-specific causal GRNs.

Perhaps the most exciting application of CRISPR-Cas9 is its ability to edit germline DNA and remove disease-causing mutations from future generations (Lokody 2014; Wu et al. 2013). Wu et al. (2013) injected the CRISPR-Cas9-sgRNA complex into the zygote of a cataract mouse model, designing the sgRNA such that the site of the cataract-causing 1-bp deletion in the *Crygc* gene was precisely cut. 6 of 22 of the resulting pups did not display a cataract phenotype and were healthy. This tantalising result demonstrates the potential to use genome editing as a remediation of congenital diseases, such as CHD.

The playing field of cardiac disease gene discovery has improved substantially in the last few years thanks to the rapid advancement of experimental genomic technologies. Generation of genomic data is no longer a bottleneck for decoding the genetic cause of heart disease. We can sequence a patient's entire genome with roughly \$1,000 in 1–2 weeks. The real bottleneck is our ability to unravel

how changes in one gene can propagate through the entire system, and how genetic variation in coding and non-coding regions can contribute to a cardiac phenotype. When this wealth of genomic data is integrated with other ‘omic’ and clinical information through state-of-the-art analytical and modelling techniques, we can begin to truly decode the complex genetic causes of heart disease. In this article, we have reviewed many recent developments in cardiac systems biology in the context of decoding the genetic causes of heart diseases. We are sure to see new advancements in the coming few years.

Compliance with ethical standards

Funding This study was supported by the Victor Chang Cardiac Research Institute, and a grant by the Human Frontier Science Program (RGY0084/2014).

Conflict of interest All authors (D.D., V.D., T.S., A.Y., D.T.H., E.G., and J.W.K.H.) declare that they do not have any conflict of interest.

Ethical approval This article does not contain any studies with human or animal subjects performed by any of the authors.

References

- Alekseyenko AA, Ho JWK, Peng S, Gelbart M, Tolstorukov MY, Plachetka A, Kharchenko PV, Jung YL, Gorchakov AA, Larschan E et al (2012) Sequence-specific targeting of dosage compensation in drosophila favors an active chromatin context. *PLoS Genet* 8: e1002646
- Alvarez-Saavedra M, Carrasco L, Sura-Trueba S, Demarchi Aiello V, Walz K, Neto JX, Young JI (2010) Elevated expression of MeCP2 in cardiac and skeletal tissues is detrimental for normal development. *Hum Mol Genet* 19:2177–2190
- Azuaje F, Zhang L, Jeanty C, Puhl S-L, Rodius S, Wagner DR (2013) Analysis of a gene co-expression network establishes robust association between Col5a2 and ischemic heart disease. *BMC Med Genomics* 6:13
- Ballouz S, Liu JY, George RA, Bains N, Liu A, Oti M, Gaeta B, Fatkin D, Wouters MA (2013) Gentrepid V2.0: a web server for candidate disease gene prediction. *BMC Bioinformatics* 14:249
- Ballouz S, Liu JY, Oti M, Gaeta B, Fatkin D, Bahlo M, Wouters MA (2014) Candidate disease gene prediction using gentrepid: application to a genome-wide association study on coronary artery disease. *Mol Genet Genomic Med* 2:44–57
- Barabasi A-L, Gulbahce N, Loscalzo J (2011) Network medicine: a network-based approach to human disease. *Nat Rev Genet* 12:56–68
- Barriot R, Breckpot J, Thienpont B, Brohée S, Van Vooren S, Coessens B, Tranchevent L-C, Van Loo P, Gewillig M, Devriendt K et al (2010) Collaboratively charting the gene-to-phenotype network of human congenital heart defects. *Genome Med* 2:16
- Barth AS, Merk S, Arnoldi E, Zwermann L, Kloos P, Gebauer M, Steinmeyer K, Bleich M, Kääh S, Hinterseer M, Kartmann H, Kreuzer E, Dugas M, Steinbeck G, Nabauer M (2005) Reprogramming of the human atrial transcriptome in permanent atrial fibrillation: expression of a ventricular-like genomic signature. *Circ Res* 96:1022–1029
- Beckmann JS, Estivill X, Antonarakis SE (2007) Copy number variants and genetic traits: closer to the resolution of phenotypic to genotypic variability. *Nat Rev Genet* 8:639–646
- Berger SI, Ma’ayan A, Iyengar R (2010) Systems pharmacology of arrhythmias. *Sci Signal* 3:ra30
- Bernabeu Llinares MO (2011) An open source HPC-enabled model of cardiac defibrillation of the human heart. Doctoral dissertation, Oxford University
- Blow MJ, McCulley DJ, Li Z, Zhang T, Akiyama JA, 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
- Blue GM, Kirk EP, Sholler GF, Harvey RP, Winlaw DS (2012) Congenital heart disease: current knowledge about causes and inheritance. *Med J Aust* 197:155–159
- Breckpot J, Thienpont B, Bauters M, Tranchevent L-C, Gewillig M, Allegaert K, Vermeesch JR, Moreau Y, Devriendt K (2012) Congenital heart defects in a novel recurrent 22q11.2 deletion harboring the genes CRKL and MAPK1. *Am J Med Genet A* 158A: 574–580
- Cai CL, Liang X, Shi Y, Chu PH, Pfaff SL, Chen J, Evans S (2003) Isl1 identifies a cardiac progenitor population that proliferates prior to differentiation and contributes a majority of cells to the heart. *Dev Cell* 5:877–889
- Callis TE, Pandya K, Seok HY, Tang R-H, Tatsuguchi M, Huang Z-P, Chen J-F, Deng Z, Gunn B, Shumate J et al (2009) MicroRNA-208a is a regulator of cardiac hypertrophy and conduction in mice. *J Clin Invest* 119:2772–2786
- Cappola TP, Li M, He J, Ky B, Gilmore J, Qu L, Keating B, Reilly M, Kim CE, Glessner J et al (2010) Common variants in HSPB7 and FRMD4B associated with advanced heart failure. *Circ Cardiovasc Genet* 3:147–154
- Chamberlain AA, Lin M, Lister RL, Maslov AA, Wang Y, Suzuki M, Wu B, Grealis JM, Zheng D, Zhou B (2014) DNA methylation is developmentally regulated for genes essential for cardiogenesis. *J Am Heart Assoc* 3:e000976
- Chang S, McKinsey TA, Zhang CL, Richardson JA, Hill JA, Olson EN (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
- Chatr-aryamontri A, Breitkreutz B-J, Heinicke S, Boucher L, Winter A, Stark C, Nixon J, Ramage L, Kolas N, O’Donnell L et al (2012) The BioGRID interaction database: 2013 update. *Nucleic Acids Res* 41: D816–D823
- Chen H, VanBuren V (2014) A provisional gene regulatory atlas for mouse heart development. *PLoS ONE* 9:e83364
- Chen Y-H, Xu S-J, Bendahhou S, Wang X-L, Wang Y, Xu W-Y, Jin H-W, Sun H, Su X-Y, Zhuang Q-N et al (2003) KCNQ1 gain-of-function mutation in familial atrial fibrillation. *Science* 299:251–254
- Chen J-F, Mandel EM, Thomson JM, Wu Q, Callis TE, Hammond SM, Conlon FL, Wang D-Z (2006) The role of microRNA-1 and microRNA-133 in skeletal muscle proliferation and differentiation. *Nat Genet* 38:228–233
- Chowdhury S, Erickson SW, MacLeod SL, Cleves MA, Hu P, Karim MA, Hobbs CA (2011) Maternal genome-wide DNA methylation patterns and congenital heart defects. *PLoS ONE* 6:e16506
- Cirulli ET, Goldstein DB (2010) Uncovering the roles of rare variants in common disease through whole-genome sequencing. *Nat Rev Genet* 11:415–425
- Claussnitzer M, Dankel SN, Klocke B, Grallert H, Glunk V, Berulava T, Lee H, Oskolkov N, Fadista J, Ehlers K et al (2014) Leveraging cross-species transcription factor binding site patterns: from diabetes risk loci to disease mechanisms. *Cell* 156:343–358
- Csardi G, Nepusz T (2006) The igraph software package for complex network research. *InterJournal Complex Systems* 1695

- Da Costa Martins PA, Bourajraj M, Gladka M, Kortland M, van Oort RJ, Pinto YM, Molkentin JD, De Windt LJ (2008) Conditional dicer gene deletion in the postnatal myocardium provokes spontaneous cardiac remodeling. *Circulation* 118:1567–1576
- De Jong AM, Maass AH, Oberdorf-Maass SU, Van Veldhuisen DJ, Van Gilst WH, Van Gelder IC (2010) Mechanisms of atrial structural changes caused by stretch occurring before and during early atrial fibrillation. *Cardiovasc Res* 89:754–765
- De la Pompa JL, Epstein JA (2012) Coordinating tissue interactions: notch signaling in cardiac development and disease. *Dev Cell* 22:244–254
- Deb A (2014) Cell-cell interaction in the heart via Wnt/ β -catenin pathway after cardiac injury. *Cardiovasc Res* 102:214–223
- Delgado-Olguin P, Huang Y, Li X, Christodoulou D, Seidman CE, Seidman JG, Tarakhovsky A, Bruneau BG (2012) Epigenetic repression of cardiac progenitor gene expression by Ezh2 is required for postnatal cardiac homeostasis. *Nat Genet* 44:343–347
- Deo R, Albert CM (2012) Epidemiology and genetics of sudden cardiac death. *Circulation* 125:620–637
- Dewey FE, Perez MV, Wheeler MT, Watt C, Spin J, Langfelder P, Horvath S, Hannehalli S, Cappola TP, Ashley EA (2011) Gene coexpression network topology of cardiac development, hypertrophy, and failure. *Circ Cardiovasc Genet* 4:26–35
- Dickel DE, Zhu Y, Nord AS, Wylie JN, Akiyama JA, Afzal V, Plajzer-Frick I, Kirkpatrick A, Göttgens B, Bruneau BG et al (2014) Function-based identification of mammalian enhancers using site-specific integration. *Nat Methods* 11:566–571
- Dickerson JE, Zhu A, Robertson DL, Hentges KE (2011) Defining the role of essential genes in human disease. *PLoS ONE* 6:e27368
- Ehrenberg M (2003) Systems biology is taking off. *Genome Res* 13:2377–2380
- Eichler EE, Flint J, Gibson G, Kong A, Leal SM, Moore JH, Nadeau JH (2010) Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet* 11:446–450
- Ernst J, Kellis M (2012) ChromHMM: automating chromatin-state discovery and characterization. *Nat Methods* 9:215–216
- Erwin GD, Oksenberg N, Truty RM, Kostka D, Murphy KK, Ahituv N, Pollard KS, Capra JA (2014) Integrating diverse datasets improves developmental enhancer prediction. *PLoS Comput Biol* 10:e1003677
- Fahed AC, Gelb BD, Seidman JG, Seidman CE (2013) Genetics of congenital heart disease: the glass half empty. *Circ Res* 112:707–720
- Fakhro KA, Choi M, Ware SM, Belmont JW, Towbin JA, Lifton RP, Khokha MK, Brueckner M (2011) Rare copy number variations in congenital heart disease patients identify unique genes in left-right patterning. *Proc Natl Acad Sci USA* 108:2915–2920
- Franceschini A, Szklarczyk D, Frankild S, Kuhn M, Simonovic M, Roth A, Lin J, Minguez P, Bork P, von Mering C et al (2012) STRING v9.1: protein-protein interaction networks, with increased coverage and integration. *Nucleic Acids Res* 41:D808–D815
- Gelb BD, Chung WK (2014) Complex genetics and the etiology of human congenital heart disease. *Cold Spring Harb Perspect Med* 4:a013953
- Glukhov AV, Fedorov VV, Kalish PW, Ravikumar VK, Lou Q, Janks D, Schuessler RB, Moazami N, Efimov IR (2012) Conduction remodeling in human end-stage nonischemic left ventricular cardiomyopathy. *Circulation* 125:1835–1847
- Goh KI, Cusick ME, Valle D, Childs B, Vidal M, Barabasi AL (2007) The human disease network. *Proc Natl Acad Sci USA* 104:8685–8690
- Greenway SC, Pereira AC, Lin JC, DePalma SR, Israel SJ, Mesquita SM, Ergul E, Conta JH, Korn JM, McCarroll SA et al (2009) De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat Genet* 41:931–935
- Grote P, Wittler L, Hendrix D, Koch F, Währisch S, Beisaw A, Macura K, Bläss G, Kellis M, Werber M 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
- Gudbjartsson DF, Holm H, Gretarsdottir S, Thorleifsson G, Walters GB, Thorgeirsson G, Gulcher J, Mathiesen EB, Njølstad I, Nyrnes A et al (2009) A sequence variant in ZFXH3 on 16q22 associates with atrial fibrillation and ischemic stroke. *Nat Genet* 41:876–878
- Gusterson RJ, Jazrawi E, Adcock IM, Latchman DS (2003) The transcriptional co-activators CREB-binding protein (CBP) and p300 play a critical role in cardiac hypertrophy that is dependent on their histone acetyltransferase activity. *J Biol Chem* 278:6838–6847
- Han P, Hang CT, Yang J, Chang C-P (2011) Chromatin remodeling in cardiovascular development and physiology. *Circ Res* 108:378–396
- Hang CT, Yang J, Han P, Cheng H-L, Shang C, Ashley E, Zhou B, Chang C-P (2010) Chromatin regulation by Brg1 underlies heart muscle development and disease. *Nature* 466:62–67
- He A, Kong SW, Ma Q, Pu WT (2011) Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc Natl Acad Sci U S A* 108:5632–5637
- He A, Ma Q, Cao J, von Gise A, Zhou P, Xie H, Zhang B, Hsing M, Christodoulou DC, Cahan P et al (2012) Polycomb repressive complex 2 regulates normal development of the mouse heart. *Circ Res* 110:406–415
- Hershberger RE, Hedges DJ, Morales A (2013) Dilated cardiomyopathy: the complexity of a diverse genetic architecture. *Nat Rev Cardiol* 10:531–547
- Ho JWK, Stefani M, dos Remedios CG, Charleston MA (2008) Differential variability analysis of gene expression and its application to human diseases. *Bioinformatics* 24:i390–i398
- Ho JWK, Jung YL, Liu T, Alver BH, Lee S, Ikegami K, Sohn K-A, Minoda A, Tolstorukov MY, Appert A et al (2014) Comparative analysis of metazoan chromatin organization. *Nature* 512:449–452
- Hookana E (2012) Characteristics of victims of non-ischemic sudden cardiac death. Doctoral dissertation, University of Oulu
- Ideker T, Galitski T, Hood L (2001) A new approach to decoding life: systems biology. *Annu Rev Genomics Hum Genet* 2:343–372
- Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E (2012) A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 337:816–821
- Junion G, Spivakov M, Girardot C, Braun M, Gustafson EH, Birney E, Furlong EEM (2012) A transcription factor collective defines cardiac cell fate and reflects lineage history. *Cell* 148:473–486
- Kacprowski T, Doncheva NT, Albrecht M (2013) NetworkPrioritizer: a versatile tool for network-based prioritization of candidate disease genes or other molecules. *Bioinformatics* 29:1471–1473
- Kamisago M, Sharma SD, DePalma SR, Solomon S, Sharma P, McDonough B, Smoot L, Mullen MP, Woolf PK, Wigle ED, Seidman JG, Seidman CE (2000) Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. *N Engl J Med* 343:1688–1696
- Kanehisa M, Goto S, Sato Y, Kawashima M, Furumichi M, Tanabe M (2013) Data, information, knowledge and principle: back to metabolism in KEGG. *Nucleic Acids Res* 42:D199–D205
- 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
- Kelder T, van Iersel MP, Hanspers K, Kutmon M, Conklin BR, Evelo CT, Pico AR (2011) WikiPathways: building research communities on biological pathways. *Nucleic Acids Res* 40:D1301–D1307
- Kimura A, Harada H, Park JE, Nishi H, Satoh M, Takahashi M, Hiroi S, Sasaoka T, Ohbuchi N, Nakamura T et al (1997) Mutations in the cardiac troponin I gene associated with hypertrophic cardiomyopathy. *Nat Genet* 16:379–382

- Klattenhoff CA, Scheuermann JC, Surface LE, Bradley RK, Fields PA, Steinhauser ML, Ding H, Butty VL, Torrey L, Haas S et al (2013) Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 152:570–583
- Kohler S, Bauer S, Horn D, Robinson PN (2008) Walking the interactome for prioritization of candidate disease genes. *Am J Hum Genet* 82:949–958
- Krajinovic M, Pinamonti B, Sinagra G, Vatta M, Severini GM, Milasin J, Falaschi A, Camerini F, Giacca M, Mestroni L (1995) Linkage of familial dilated cardiomyopathy to chromosome 9. Heart muscle disease study group. *Am J Hum Genet* 57:846–852
- Krauthammer M, Kaufmann CA, Gilliam TC, Rzhetsky A (2004) Molecular triangulation: bridging linkage and molecular-network information for identifying candidate genes in Alzheimer's disease. *Proc Natl Acad Sci USA* 101:15148–15153
- Lachke SA, Ho JWK, Kryukov GV, O'Connell DJ, Aboukhalil A, Bulyk ML, Park PJ, Maas RL (2012) ISyTE: integrated systems tool for Eye gene discovery. *Invest Ophthalmol Vis Sci* 53:1617–1627
- Lage K, Möllgard K, Greenway S, Wakimoto H, Gorham JM, Workman CT, Bendtsen E, Hansen NT, Rigina O, Roque FS, et al. (2010) Dissecting spatio-temporal protein networks driving human heart development and related disorders. *Mol Syst Biol* 6:381
- Lage K, Greenway SC, Rosenfeld JA, Wakimoto H, Gorham JM, Segrè AV, Roberts AE, Smoot LB, Pu WT, Pereira AC et al (2012) Genetic and environmental risk factors in congenital heart disease functionally converge in protein networks driving heart development. *Proc Natl Acad Sci U S A* 109:14035–14040
- Li W, Chen L, He W, Li W, Qu X, Liang B, Gao Q, Feng C, Jia X, Lv Y et al (2013) Prioritizing disease candidate proteins in cardiomyopathy-specific protein-protein interaction networks based on “guilt by association?” analysis. *PLoS ONE* 8:e71191
- Li X, Martinez-Fernandez A, Hartjes KA, Kocher J-PA, Olson TM, Terzic A, Nelson TJ (2014) Transcriptional atlas of cardiogenesis maps congenital heart disease interactome. *Physiol Genomics* 46:482–495
- Lin Q, Schwarz J, Bucana C, Olson EN (1997) Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C. *Science* 276:1404–1407
- Lin C-C, Hsiang J-T, Wu C-Y, Oyang Y-J, Juan H-F, Huang H-C (2010) Dynamic functional modules in co-expressed protein interaction networks of dilated cardiomyopathy. *BMC Syst Biol* 4:138
- Liu N, Bezprozvannaya S, Williams AH, Qi X, Richardson JA, Bassel-Duby R, Olson EN (2008) microRNA-133a regulates cardiomyocyte proliferation and suppresses smooth muscle gene expression in the heart. *Genes Dev* 22:3242–3254
- Lokody I (2014) Genetic therapies: correcting genetic defects with CRISPR-Cas9. *Nat Rev Genet* 15:63
- Lyons I, Parsons LM, Hartley L, Li R, Andrews JE, Robb L, Harvey RP (1995) Myogenic and morphogenetic defects in the heart tubes of murine embryos lacking the homeo box gene Nkx2-5. *Genes Dev* 9:1654–1666
- MacLellan WR, Wang Y, Lusis AJ (2012) Systems-based approaches to cardiovascular disease. *Nat Rev Cardiol* 9:172–184
- MacRae CA (2010) The genetics of congestive heart failure. *Heart Fail Clin* 6:223–230
- Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorf LA, Hunter DJ, McCarthy MI, Ramos EM, Cardon LR, Chakravarti A et al (2009) Finding the missing heritability of complex diseases. *Nature* 461:747–753
- Mar JC, Matigian NA, Mackay-Sim A, Mellick GD, Sue CM, Silburn PA, McGrath JJ, Quackenbush J, Wells CA (2011) Variance of gene expression identifies altered network constraints in neurological disease. *PLoS Genet* 7:e1002207
- Matkovich SJ, Van Booven DJ, Eschenbacher WH, Dorn GW 2nd (2011) RISC RNA sequencing for context-specific identification of in vivo microRNA targets. *Circ Res* 108:18–26
- Matkovich SJ, Edwards JR, Grossenheider TC, de Guzman Strong C, Dorn GW 2nd (2014) Epigenetic coordination of embryonic heart transcription by dynamically regulated long noncoding RNAs. *Proc Natl Acad Sci USA* 111:12264–12269
- May D, Blow MJ, Kaplan T, McCulley DJ, Jensen BC, Akiyama JA, Holt A, Plajzer-Frick I, Shoukry M, Wright C et al (2012) Large-scale discovery of enhancers from human heart tissue. *Nat Genet* 44:89–93
- McCulley DJ, Black BL (2012) Transcription factor pathways and congenital heart disease. *Curr Top Dev Biol* 100:253–277
- Merlo M, Pyxaras SA, Pinamonti B, Barbati G, Di Lenarda A, Sinagra G (2011) Prevalence and prognostic significance of left ventricular reverse remodeling in dilated cardiomyopathy receiving tailored medical treatment. *J Am Coll Cardiol* 57:1468–1476
- Messina DN, Speer MC, Pericak-Vance MA, McNally EM (1997) Linkage of familial dilated cardiomyopathy with conduction defect and muscular dystrophy to chromosome 6q23. *Am J Hum Genet* 61:909–917
- Miyamoto S, Kawamura T, Morimoto T, Ono K, Wada H, Kawase Y, Matsumori A, Nishio R, Kita T, 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
- Molkentin JD, Lin Q, Duncan SA, Olson EN (1997) Requirement of the transcription factor GATA4 for heart tube formation and ventral morphogenesis. *Genes Dev* 11:1061–1072
- Montgomery RL, Potthoff MJ, Haberland M, Qi X, Matsuzaki S, Humphries KM, Richardson JA, Bassel-Duby R, Olson EN (2008) Maintenance of cardiac energy metabolism by histone deacetylase 3 in mice. *J Clin Invest* 118:3588–3597
- Myerburg RJ (2001) Sudden cardiac death: exploring the limits of our knowledge. *J Cardiovasc Electrophysiol* 12:369–381
- Myocardial Infarction Genetics Consortium, Kathiresan S, Voight BF, Purcell S, Musunuru K, Ardissono D, Mannucci PM, Anand S, Engert JC et al. (2009) Genome-wide association of early-onset myocardial infarction with single nucleotide polymorphisms and copy number variants. *Nat Genet* 41:334–341
- Narlikar L, Sakabe NJ, Blanski AA, Arimura FE, Westlund JM, Nobrega MA, Ovcharenko I (2010) Genome-wide discovery of human heart enhancers. *Genome Res* 20:381–392
- Padmanabhan N, Jia D, Geary-Joo C, Wu X, Ferguson-Smith AC, Fung E, Bieda MC, Snyder FF, Gravel RA, Cross JC et al (2013) Mutation in folate metabolism causes epigenetic instability and transgenerational effects on development. *Cell* 155:81–93
- Padovan-Merhar O, Raj A (2013) Using variability in gene expression as a tool for studying gene regulation. *Rev Syst Biol Med* 5:751–759
- Papait R, Cattaneo P, Kunderfranco P, Greco C, Carullo P, Guffanti A, Viganò V, Stirparo GG, Latronico MVG, Hasenfuss G et al (2013) Genome-wide analysis of histone marks identifying an epigenetic signature of promoters and enhancers underlying cardiac hypertrophy. *Proc Natl Acad Sci U S A* 110:20164–20169
- Park A, Won ST, Pentecost M, Bartkowski W, Lee B (2014) CRISPR/Cas9 allows efficient and complete knock-in of a destabilization domain-tagged essential protein in a human cell line, allowing rapid knockdown of protein function. *PLoS ONE* 9:e95101
- Patterson AJ, Chen M, Xue Q, Xiao D, Zhang L (2010) Chronic prenatal hypoxia induces epigenetic programming of PKCepsilon gene repression in rat hearts. *Circ Res* 107:365–373
- Pauling L, Itano HA (1949) Sickle cell anemia a molecular disease. *Science* 110:543–548

- Peterkin T, Gibson A, Patient R (2003) GATA-6 maintains BMP-4 and Nkx2 expression during cardiomyocyte precursor maturation. *EMBO J* 22:4260–4273
- Ran FA, Hsu PD, Lin C-Y, Gootenberg JS, Konermann S, Trevino AE, Scott DA, Inoue A, Matoba S, Zhang Y et al (2013) Double nicking by RNA-guided CRISPR Cas9 for enhanced genome editing specificity. *Cell* 154:1380–1389
- Roach JC, Glusman G, Smit AFA, Huff CD, Hubley R, Shannon PT, Rowen L, Pant KP, Goodman N, Bamshad M et al (2010) Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* 328:636–639
- Runyan RB, Markwald RR (1983) Invasion of mesenchyme into three-dimensional collagen gels: a regional and temporal analysis of interaction in embryonic heart tissue. *Dev Biol* 95:108–114
- Sadrieh A, Domanski L, Pitt-Francis J, Mann SA, Hodgkinson EC, Ng CA, Perry MD, Taylor JA, Gavaghan D, Subbiah RN, Vandenberg JJ, Hill AP (2014) Multiscale cardiac modelling reveals the origins of notched T waves in long QT syndrome type 2. *Nat Commun* 5: 5069
- Sanchez-Castro M, Gordon CT, Petit F, Nord AS, Callier P, Andrieux J, Guérin P, Pichon O, David A, Abadie V et al (2013) Congenital heart defects in patients with deletions upstream of SOX9. *Hum Mutat* 34:1628–1631
- Savu O, Jurcuț R, Giuscă S, van Mieghem T, Gussi I, Popescu BA, Ginghină C, Rademakers F, Deprest J, Voigt JU (2012) *Circ Cardiovasc Imaging* 5:289–297
- Saxena A, Tabin CJ (2010) miRNA-processing enzyme Dicer is necessary for cardiac outflow tract alignment and chamber septation. *Proc Natl Acad Sci U S A* 107:87–91
- Schlesinger, J., Schueler, M., Grunert, M., Fischer, J.J., Zhang, Q., Krueger, T., Lange, M., Tönjes, 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.
- Shalem O, Sanjana NE, Hartenian E, Shi X, Scott DA, Mikkelsen TS, Heckl D, Ebert BL, Root DE, Doench JG et al (2014) Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* 343:84–87
- Sheng W, Qian Y, Zhang P, Wu Y, Wang H, Ma X, Chen L, Ma D, Huang G (2014) Association of promoter methylation statuses of congenital heart defect candidate genes with tetralogy of fallot. *J Transl Med* 12:31
- Sherwood RI, Hashimoto T, O'Donnell CW, Lewis S, Barkal AA, van Hoff JP, Karun V, Jaakkola T, Gifford DK (2014) Discovery of directional and nondirectional pioneer transcription factors by modeling DNase profile magnitude and shape. *Nat Biotechnol* 32: 171–178
- Smaill BH, Hunter PJ (2010) Computer modeling of electrical activation: from cellular dynamics to the whole heart. In: Sigg DC, Iaizzo PA, Xiao Y-F, He B (eds) *Cardiac electrophysiology methods and models*. Springer, New York, pp 159–185
- 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
- Smith RP, Taher L, Patwardhan RP, Kim MJ, Inoue F, Shendure J, Ovcharenko I, Ahituv N (2013) Massively parallel decoding of mammalian regulatory sequences supports a flexible organizational model. *Nat Genet* 45:1021–1028
- Soemedi R, Wilson IJ, Bentham J, Darlay R, Töpf A, Zelenika D, Cosgrove C, Setchfield K, Thornborough C, Granados-Riveron J et al (2012) Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am J Hum Genet* 91:489–501
- Song J, Chen KC (2014) Spectacle: faster and more accurate chromatin state annotation using spectral learning. Cold Spring Harbor Laboratory, Cold Spring Harbor
- Sperling SR (2011) Systems biology approaches to heart development and congenital heart disease. *Cardiovasc Res* 91:269–278
- Stennard FA, Costa MW, Lai D, Biben C, Furtado MB, Solloway MJ, McCulley DJ, Leimena C, Preis JJ, Dunwoodie SL, Elliott DE, Prall OW, Black BL, Fatkin D, Harvey RP (2005) Murine T-box transcription factor Tbx20 acts as a repressor during heart development, and is essential for adult heart integrity, function and adaptation. *Development* 132: 2451–2462
- Strasser BJ (1999) Perspectives: molecular medicine. “Sickle cell anemia, a molecular disease”. *Science* 286:1488–1490
- Takeuchi JK, Lou X, Alexander JM, Sugizaki H, Delgado-Olguín P, Holloway AK, Mori AD, Wylie JN, Munson C, Zhu Y et al (2011) Chromatin remodelling complex dosage modulates transcription factor function in heart development. *Nat Commun* 2:187
- Tan N, Chung MK, Smith JD, Hsu J, Serre D, Newton DW, Castel L, Soltesz E, Pettersson G, Gillinov AM et al (2013) Weighted gene coexpression network analysis of human left atrial tissue identifies gene modules associated with atrial fibrillation. *Circ Cardiovasc Genet* 6:362–371
- Tanaka M, Chen Z, Bartunkova S, Yamasaki N, Izumo S (1999) The cardiac homeobox gene Csx/Nkx2.5 lies genetically upstream of multiple genes essential for heart development. *Development* 126: 1269–1280
- Tester DJ, Medeiros-Domingo A, Will ML, Haglund CM, Ackerman MJ (2012) Cardiac channel molecular autopsy: insights from 173 consecutive cases of autopsy-negative sudden unexplained death referred for postmortem genetic testing. *Mayo Clin Proc* 87:524–539
- Thienpont B, Zhang L, Postma AV, Breckpot J, Tranchevent L-C, Van Loo P, Möllgård K, Tommerup N, Bache I, Tümer Z, van Engelen K, Menten B, Mortier G, Waggoner D, Gewillig M, Moreau Y, Devriendt K, Larsen LA (2010) Haploinsufficiency of TAB2 causes congenital heart defects in humans. *Am J Hum Genet* 86:839–849
- Tranchevent L-C, Barriot R, Yu S, Van Vooren S, Van Loo P, Coessens B, De Moor B, Aerts S, Moreau Y (2008) ENDEAVOUR update: a web resource for gene prioritization in multiple species. *Nucleic Acids Res* 36:W377–W384
- Tranchevent L-C, Capdevila FB, Nitsch D, De Moor B, De Causmaecker P, Moreau Y (2010) A guide to web tools to prioritize candidate genes. *Brief Bioinform* 12:22–32
- Trapnell C, Cacchiarelli D, Grimsby J, Pokharel P, Li S, Morse M, Lennon NJ, Livak KJ, Mikkelsen TS, Rinn JL (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol* 32:381–386
- Trent RJ (2012) *Molecular and cellular therapies*. Elsevier, Amsterdam
- Udali S, Guarini P, Moruzzi S, Choi S-W, Friso S (2013) Cardiovascular epigenetics: from DNA methylation to microRNAs. *Mol Asp Med* 34:883–901
- Van Rooij E, Sutherland LB, Qi X, Richardson JA, Hill J, Olson EN (2007) Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* 316:575–579
- Wang H, Yang H, Shivalila CS, Dawlaty MM, Cheng AW, Zhang F, Jaenisch R (2013a) One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering. *Cell* 153:910–918
- Wang VY, Hoogendoorn C, Frangi AF, Cowan BR, Hunter PJ, Young AA, Nash MP (2013) Automated personalised human left ventricular FE models to investigate heart failure mechanics. *Proceedings of the third international conference on Statistical Atlases and Computational Models of the Heart: imaging and modelling challenges*; 10/2012, pp 307–316

- Wang T, Wei JJ, Sabatini DM, Lander ES (2014) Genetic screens in human cells using the CRISPR-Cas9 system. *Science* 343: 80–84
- Wellcome Trust Case Control Consortium, Craddock N, Hurles ME, Cardin N, Pearson RD, Plagnol V, Robson S, Vukcevic D, Barnes C, Conrad DF et al (2010) Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 464:713–720
- Westerhoff HV, Palsson BO (2004) The evolution of molecular biology into systems biology. *Nat Biotechnol* 22:1249–1252
- White MA, Myers CA, Corbo JC, Cohen BA (2013) Massively parallel in vivo enhancer assay reveals that highly local features determine the cis-regulatory function of ChIP-seq peaks. *Proc Natl Acad Sci U S A* 110:11952–11957
- Wills QF, Livak KJ, Tipping AJ, Enver T, Goldson AJ, Sexton DW, Holmes C (2013) Single-cell gene expression analysis reveals genetic associations masked in whole-tissue experiments. *Nat Biotechnol* 31:748–752
- Wu Y, Liang D, Wang Y, Bai M, Tang W, Bao S, Yan Z, Li D, Li J (2013) Correction of a genetic disease in mouse via use of CRISPR-Cas9. *Cell Stem Cell* 13:659–662
- Xie C, Yuan J, Li H, Li M, Zhao G, Bu D, Zhu W, Wu W, Chen R, Zhao Y (2014) NONCODEv4: exploring the world of long non-coding RNA genes. *Nucleic Acids Res* 42:D98–103
- Xu H, Morishima M, Wylie JN, Schwartz RJ, Bruneau BG, Lindsay EA, Baldini A (2004) Tbx1 has a dual role in the morphogenesis of the cardiac outflow tract. *Development* 131: 3217–3227
- Xu M, Wu X, Li Y, Yang X, Hu J, Zheng M, Tian J (2014) CITED2 mutation and methylation in children with congenital heart disease. *J Biomed Sci* 21:7
- Zaidi S, Choi M, Wakimoto H, Ma L, Jiang J, Overton JD, Romano-Adesman A, Bjornson RD, Breitbart RE, Brown KK et al (2013) De novo mutations in histone-modifying genes in congenital heart disease. *Nature* 498:220–223
- Zhang Q-J, Chen H-Z, Wang L, Liu D-P, Hill JA, 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
- Zhang J, Ma X, Wang H, Ma D, Huang G (2014) Elevated methylation of the RXRA promoter region may be responsible for its downregulated expression in the myocardium of patients with TOF. *Pediatr Res* 75:588–594
- Zhao Y, Ransom JF, Li A, Vedantham V, von Drehle M, Muth AN, Tsuchihashi T, McManus MT, Schwartz RJ, Srivastava D (2007) Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 129:303–317