

## REVIEW

# Genetics of congenital heart disease: the contribution of the noncoding regulatory genome

Alex V Postma<sup>1,2</sup>, Connie R Bezzina<sup>3</sup> and Vincent M Christoffels<sup>1</sup>

Congenital heart disease (CHD) is the most common type of birth defect. The advent of corrective cardiac surgery and the increase in knowledge concerning the longitudinal care of patients with CHD has led to a spectacular increase in life expectancy. Therefore, >90% of children with CHD, who survive the first year of life, will live into adulthood. The etiology of CHD is complex and is associated with both environmental and genetic causes. CHD is a genetically heterogeneous disease that is associated with long-recognized chromosomal abnormalities, as well as with mutation in numerous (developmental) genes. Nevertheless, the genetic factors underlying CHD have remained largely elusive, and it is important to realize that in the far majority of CHD patients no causal mutation or chromosomal abnormality is identified. However, new insights (alternative inheritance paradigms) and technology (next-generation sequencing) have become available that can greatly advance our understanding of the genetic factors that contribute to CHD; these will be discussed in this review. Moreover, we will focus on the discovery of regulatory regions of key (heart) developmental genes and the occurrence of variations and mutations within, in the setting of CHD.

*Journal of Human Genetics* (2016) 61, 13–19; doi:10.1038/jhg.2015.98; published online 30 July 2015

## INTRODUCTION

Congenital heart disease (CHD) is the most common type of birth defect, accounting for one-third of all major congenital anomalies. Worldwide, 1.35 million infants are born with CHD each year, with a worldwide occurrence of 7 per 1000 live births.<sup>1</sup> CHD comprises a group of structural abnormalities of the heart, including septal defects, valve defects and lesions affecting the outflow tract. The etiology of CHD is complex and is associated with both environmental and genetic causes (see below). The morphogenetic events that are disrupted during cardiogenesis and that lead to CHD have started to be unraveled, as are several of the molecular networks that guide normal heart development.<sup>2,3</sup> Before the advent of cardiac surgery, CHD was, together with infectious diseases, the most prominent cause of death in the neonatal and infant age group. Of all patients with a substantial CHD, a small minority—estimated <15%—reached adulthood. The advent of corrective cardiac surgery, introduced in the late 1950s, and the increase in knowledge concerning the longitudinal care of patients with CHD, has led to a spectacular increase in life expectancy. Nowadays, >90% of children with CHD that survive the first year of life will live into adulthood;<sup>4</sup> this has brought with it new challenges in their clinical management, as many patients develop late complications. It has become clear that the type and severity of long-term clinical outcome (for example, heart failure,<sup>5</sup> arrhythmia<sup>6</sup> and aneurysm<sup>7</sup>) that occur throughout life in CHD patients varies not only between patients with different types of CHD but also among

patients with the same type of CHD.<sup>8</sup> Some may have an uncomplicated clinical course, while others are disabled by many late complications. It has been proven difficult to predict—on clinical grounds alone—which patients will develop the more severe spectrum of late complications. As many CHD patients now live to reproductive age, offspring of CHD patients are also at risk of CHD.<sup>9</sup> Our current lack of knowledge about the genetic underpinnings of CHD constitutes a major obstacle in reproductive counseling of patients with CHD.

## GENETICS OF CHD

CHD is a genetically heterogeneous disease. Some forms of CHD are associated with long-recognized chromosomal abnormalities such as chromosome 22q11 deletion and trisomy of chromosome 21. Moreover, over 50 human disease genes have been implicated in CHD so far,<sup>10</sup> although, the bulk of CHD-associated mutations fall in a small set of developmental genes (for example, *NKX2-5*,<sup>11</sup> *GATA4*<sup>12</sup> and *NOTCH1*<sup>13</sup>). Studies in mice, with targeted gene deletion, have indicated that over 500 genes can induce heart defects when mutated.<sup>10,14</sup> It is likely that at least the same number of human CHD disease genes exist. However, it is important to realize that in the far majority of CHD patients no causal mutation or chromosomal abnormality is identified. Thus, despite uncovering causal CHD genes, the genetic factors underlying CHD remain largely elusive.<sup>15,16</sup> This is because classical genetic methodologies used in uncovering causal

<sup>1</sup>Department of Anatomy, Embryology & Physiology, Academic Medical Center, Amsterdam, The Netherlands; <sup>2</sup>Department of Clinical Genetics, Academic Medical Center, Amsterdam, The Netherlands and <sup>3</sup>Department of Clinical and Experimental Cardiology, Academic Medical Center, Amsterdam, The Netherlands  
Correspondence: Dr AV Postma, Department of Anatomy, Embryology & Physiology or Department of Clinical Genetics, Academic Medical Center, Heart Failure Research Center, Meibergdreef 15, Amsterdam 1105 AZ, The Netherlands.  
E-mail: a.v.postma@amc.uva.nl

Received 22 June 2015; revised 7 July 2015; accepted 8 July 2015; published online 30 July 2015

genes (for example, linkage analysis) have, until recently, restricted gene discovery efforts to familial forms of CHD with highly penetrant autosomal-dominant inheritance (or autosomal-recessive inheritance in consanguineous families). These efforts have, with exceptions,<sup>13,17,18</sup> been largely unsuccessful due to the lack of large families with multiple affected individuals (rare in CHD). Therefore, until recently, the availability of only small families, or the absence of familial aggregation of the disease (sporadic presentation), has led investigators to identify CHD genes by a candidate gene approach. Functional validation studies of identified genes in model systems have only been carried out in few instances. Similarly, identified gene/s have not been implicated in multiple independent cases. Thus the causality of many genes identified in this way has remained unclear.

### NOVEL WAYS TO IDENTIFY CAUSAL CHD GENES

In the last decade, technological developments, which allow us to look for genetic alterations in a genome-wide fashion in large series of patients, have given the search for causal (CHD) genes a remarkable boost. An early development was the availability of array-comparative genomic hybridization, later followed by genome-wide single-nucleotide polymorphism arrays. These led to the delineation of various microdeletion and duplication syndromes, which include CHD as a part of their clinical spectrum.<sup>10</sup> One such an example is the 1q21.1 deletion syndrome (which is associated with CHD in a subset of patients<sup>19</sup>), which has a minimal deleted region that contains *GJA5*. This gene encodes a cardiac gap junctional channel (connexin 40) that is present in atrial myocardium, bundle of histone and the ventricular conduction system.<sup>20,21</sup> Interestingly, several recent screens of tetralogy of fallot cohorts demonstrated that small duplications (together with deletions also termed copy number variations or CNVs) encompassing *GJA5* were highly overrepresented.<sup>22,23</sup> Moreover, *Gja5* knockout mice have various cardiac defects including tetralogy of fallot,<sup>24</sup> suggesting that haploinsufficiency of *GJA5* may underlie the cardiac defects in some 1q21.1 deletion carriers. Thus far, more than a dozen whole-genome CNV screening studies have been published, comprising more than 5000 patients (overview in Andersen *et al.*<sup>10</sup>). The overall conclusion is that a significant portion of CHD patients have pathogenic CNVs. Not surprising, the highest incidence of pathogenic CNVs are found in patients who have both CHD and extra-cardiac anomalies ( $\pm 15$ –20%).<sup>25–29</sup> In contrast, the incidence of pathogenic CNVs is significantly lower among patients with an isolated CHD, (between 4 and 14%).<sup>25,30–32</sup> However, these estimations should be interpreted with caution, as the various studies differ in their designs, and experimental and analytical setups. Importantly, many of these CHD loci do not contain known (cardiac) developmental genes, or, in some cases, no genes at all (intra-genic regions). Moreover, most CNV loci harbor multiple genes, where it is likely that only one gene is responsible for CHD in those patients. Overall, this can make it difficult to pinpoint the exact cause of the disease and uncover its pathophysiological mechanism.

CHD gene discovery was further boosted by the advent of exome and whole-genome sequencing using the so-called next-generation sequencing (NGS). These technologies bypass steps that define CHD chromosomal loci, and instead identify CHD mutations by direct sequencing of the whole genome or exome, with the exome defined as the genomic sequence that codes for proteins (~50 Mb). Such an approach identifies tens of thousands of single-nucleotide variants (SNVs) per exome that differ from the reference sequence<sup>33</sup> and multiple-fold more SNVs per genome. Given that most of these SNVs will be unrelated to CHD, extensive post-sequencing filters are used to focus on novel or rare SNVs that are predicted to have deleterious

functional consequences.<sup>15</sup> Still in a typical experiment, this can yield hundreds of candidate variants. One way around this is to establish the mode of inheritance of the disease. In general there are three options: (1) mutations can be inherited (autosomal recessive, X-linked recessive or autosomal dominant), (2) mutations are *de novo* dominant or (3) mutations occur in a mosaic form.<sup>34</sup> When there is familial recurrence or parental consanguinity, the likelihood that a disease is monogenic is high. The mode of inheritance influences the selection and number of individuals to sequence, as well as the analytical approach used.<sup>34</sup> In general, exome sequencing has been most successful in the identification of *de novo* and homozygous mutations, both inheritance models that allow for a substantial reduction of candidate variants.<sup>35</sup> In such inheritance models, the (healthy) parents and the affected offspring are sequenced, and only the variants that are *de novo* or homozygous in the child are then taken into consideration. By contrast, genes responsible for autosomal-dominant disorders are more difficult to identify by exome sequencing,<sup>34</sup> as the number of heterozygous gene variants shared among affected relatives typically yields a considerable number of variants, particularly in small families in whom co-segregation analysis cannot be effectively performed.

As the assumption is that CHD is under strong selective pressure, *de novo* mutations might account for the sporadic pattern of occurrence among many CHD cases (see also below). The study by Zaidi *et al.*<sup>36</sup> on the incidence of *de novo* mutations in 362 severe CHD cases and 264 controls, highlights the successes of this approach. These investigators identified a marked excess of *de novo* mutations in genes involved in the production, removal or reading of histone 3 lysine 4 methylation.<sup>36</sup> Likewise, Al Turki *et al.*<sup>37</sup> performed whole-exome sequencing (WES) on 13 parent-offspring trios and 112 unrelated individuals with nonsyndromic atrio ventricular septum defects (AVSDs). They identified rare *de novo* missense variants in the highly conserved gene *NR2F2* that encodes a pleiotropic developmental transcription factor. Functional mouse studies revealed that *NR2F2* determines atrial identity during heart development<sup>38</sup> providing support for the relevance of this finding. They went on to screen *NR2F2* in familial CHD and identified three CHD-affected families with other variants in *NR2F2*, including a *de novo*-balanced chromosomal translocation, a *de novo* substitution disrupting a splice donor site and a 3 bp duplication that co-segregated in a multiplex family.

Similarly, WES has also been used successfully in cases of recessive inheritance. For instance, Chetaille *et al.*<sup>39</sup> identified 17 subjects in whom sick sinus syndrome and chronic intestinal pseudo-obstruction co-occurred during the first four decades of life. Distinctive cardiac clinical features included atrial dysrhythmias, sick sinus syndrome and valve anomalies. Family evaluation suggested the inheritance in an autosomal recessive manner. WES of three unrelated probands identified one homozygous pathogenic variant shared by all probands, in Shugoshin-like 1 (*SGOL1*), predicted to encode a damaging change at a highly conserved residue. Morpholino-induced knockdown of *sgol1* in zebrafish recapitulated the abnormalities seen in humans.

Although challenging, successful identification by WES of genetic defects with an autosomal-dominant transmission has also been reported. For instance, Bainbridge *et al.*<sup>40</sup> identified causative mutations that segregated with left ventricular noncompaction in five small families using WES. They identified mutations in known left ventricular noncompaction genes such as *MYH7* and *TPM1*, but also identified a frameshift mutation in nicotinamide nucleotide transhydrogenase (*NNT*), a nuclear encoded mitochondrial protein, not implicated previously in human cardiomyopathies. Resequencing of *NNT* in additional left ventricular noncompaction families identified a

second likely pathogenic missense allele. They went on to show using *in vivo* complementation studies in zebrafish that mutant human NNT failed to rescue NNT morpholino-induced heart dysfunction, indicating a probable haploinsufficiency mechanism. Taken together, the above examples show that WES can be a good way to identify causative CHD mutations, and that its success depends partly on the mode of inheritance of the phenotype.

**TESTING ALTERNATIVE INHERITANCE PARADIGMS**

Notwithstanding the above successful examples, data from NGS studies in other complex disease traits indicates that WES alone will not fully determine the causes of CHD; at least not with the prevailing analysis paradigm that WES will always detect causal coding variants.<sup>41</sup> For all the successes, there are examples in which a single causative variant cannot be identified in familial forms of CHD, especially autosomal-dominant ones. For instance, Arrington *et al.*<sup>42</sup> performed WES in a large family with an autosomal-dominant inheritance pattern with 11 affected family members (diverse CHD including septal defects and co-arcuation of the aorta). The WES data were filtered to eliminate common variants and sequencing artifacts and were prioritized based on the predicted effect of the variant and on gene function. After filtering and prioritization, they reduced the number of variants potentially involved in CHD to 18. However, none of the variants completely segregated with CHD in the family. The best candidate was a variant in *MYH6* previously identified in a patient with tricuspid atresia and large secundum atrial septal defect. However, this variant was identified in all but one affected family member. Likewise, Martin *et al.*<sup>43</sup> failed to identify the causal variant in a multiplex family with bicuspid aortic valve and other cardiovascular malformations using WES. They could not find a single likely-causal coding variant shared by all affected. These examples demonstrate that we should keep in mind that only a minority of CHD follows clear Mendelian transmission patterns. The low recurrence risk (probability of occurrence within a given family when one family member is affected) for CHD,<sup>8,44</sup> is compatible with the often-sporadic presentation of the disease. Although sporadic disease presentation may be at least in part accounted for by the contribution of non-genetic (environmental) factors, some of which are recognized,<sup>15</sup> sporadic presentation of CHD may very well be the manifestation of a number of different inheritance patterns. Alongside the inheritance patterns discussed above (autosomal-dominant *de novo*, autosomal or X-linked recessive or compound heterozygous inheritance), these may entail digenic or polygenic inheritance. Ultimately, it is likely that all these genetic models contribute to (various forms of) CHD.<sup>16</sup> Evidence that polygenic inheritance has a role in CHD is the string of recent publications that suggest that common variations in (regulatory) DNA can attribute to CHD (Table 1). The first genome-wide association study (GWAS) on CHD by Cordell *et al.*<sup>45</sup> in 2013 focused on cases of Western descent with tetralogy of fallot and identified two loci, respectively, on 12q24 and 13q32, strongly associated with the disease. This provided evidence that common genetic variation influences the risk of tetralogy of fallot.<sup>45</sup> Although some CHD candidate genes were present in these loci, the association signals indicated that variants in noncoding regions were likely the cause. Likewise, a similar approach by the same group identified one locus on 4p16 highly associated with secundum atrial septal defect that likely involves noncoding variants that lower the expression of an unknown gene (*LOC100507266*).<sup>46</sup> An interesting conclusion from that study was that lumping all CHDs together provided less power than focusing on specific lesions. This suggests that future investigations may be more powerful if performed

**Table 1 Congenital heart disease genome-wide association studies**

Phenotype	Ethnicity	Accession	Location	Closest gene	-logp	Effect size	Risk allele	Reference
Atrial Septal Defect Type 2	Western descent	rs870142	4p16.2	<i>STX18-AS1</i>	9.523	OR:1.46 (95% CI:1.19–1.65)	A (freq:0.228)	Cordell <i>et al.</i> <sup>46</sup>
Congenital heart malformation	Han Chinese	rs1531070	4q31.1	<i>MAML3</i>	11.301	OR:1.4 (95% CI:1.27–1.54)	A (freq:0.11)	Hu <i>et al.</i> <sup>47</sup>
Congenital heart malformation	Han Chinese	rs2474937	1p12	<i>SPAG17</i>	9.097	OR:1.4 (95% CI:1.26–1.56)	C (freq:0.08)	Hu <i>et al.</i> <sup>47</sup>
Tetralogy of Fallot	Western descent	rs11065987	12q24.12	<i>BRAP</i>	10.097	OR:1.34 (95% CI:1.208–1.496)	G (freq:0.418)	Cordell <i>et al.</i> <sup>45</sup>
Tetralogy of Fallot	Western descent	rs7982677	13q31.3	<i>GPC5</i>	8.523	OR:1.29 (95% CI:1.152–1.441)	A (freq:0.281)	Cordell <i>et al.</i> <sup>45</sup>
Tetralogy of Fallot	Western descent	rs2388896	10p14	<i>LOC105755953</i>	7.046	OR:1.28 (95% CI:1.15–1.43)	G (freq:0.599)	Cordell <i>et al.</i> <sup>45</sup>
Tetralogy of Fallot	Western descent	rs2228638	10p11.22	<i>NRPI</i>	6.699	OR:1.45 (95% CI:1.239–1.69)	A (freq:0.096)	Cordell <i>et al.</i> <sup>45</sup>
Tetralogy of Fallot	Western descent	rs6499100	16q12.2	<i>CASC16</i>	6	OR:1.26 (95% CI:1.14–1.4)	A (freq:0.519)	Cordell <i>et al.</i> <sup>45</sup>
Tetralogy of Fallot	Western descent	rs7646881	3q25.32	<i>MFSD1</i>	5.699	OR:1.39 (95% CI:1.224–1.581)	A (freq:0.168)	Cordell <i>et al.</i> <sup>45</sup>

Abbreviations: CI, confidence interval; OR, odds ratio.

on homogeneous CHD populations. The third CHD GWAS study identified two additional risk loci (1p12 and 4q31.1) for congenital heart malformations in the Han Chinese population, again either in intronic or intragenic regions.<sup>47</sup> The fact that most of these CHD risk-associated loci map to regions outside coding genes indicates that noncoding variation can contribute to CHD and that variation in gene expression/dosage is involved in CHD pathogenesis.

### BEYOND THE EXOME, IDENTIFYING THE REGULOME

Historically, the search for mutations contributing to human disease was mainly limited to exons and their immediate surrounding areas. The functional genomic elements that are not transcribed and translated remained uninvestigated. This has also changed dramatically due to the new NGS techniques and protocols. Indeed, one of the key discoveries of genome sequencing has been the unexpectedly high proportion of noncoding functional DNA. Data from the ENCODE project suggest that 37% of the human genome might have a function and probably affects regulatory and tissue-specific expression patterns.<sup>48</sup> Given the above, a key component in the modern CHD cause-discovery pipeline would therefore be the identification of regulatory elements controlling a cell's transcription and translation process. This is known as the cellular regulome. These are the enhancing and repressing elements in the genome that bind key regulatory transcription factor complexes that can be anywhere within or around gene sequences, and which are themselves subject to regulation via histone modifications and structural conformation of the direct and surrounding chromatin (for reviews see De Laat and Duboule,<sup>49</sup> Shlyueva *et al.*<sup>50</sup>). Importantly, as discussed above, these elements are just as susceptible to variability as the rest of the genome and can cause disease.

A major breakthrough in uncovering the regulome was the upscaling of sequencing efforts to the genomic level. Transcription factor-binding studies, typically carried out *in vitro*, can now be performed genome wide, permitting one to identify all elements bound by a transcription factor in a single experiment using a technique called ChIP-seq (chromatin immunoprecipitation-sequencing).<sup>51</sup> Similarly, this technique can be applied to histone modifications, allowing the identification of activating or repressive histone complexes highlighting active or inactive regulatory elements during a tissue's or even a cell's development. Performing multiple ChIP-seq experiments with different factors and examining the overlap of the thus-obtained data have become a powerful tool for locating regulatory elements over the entire genome. However, information from such experiments gives an overview of all possible regulatory elements, not necessarily the ones that are active. One way around this is to also incorporate data from directly measuring chromatin accessibility by techniques that isolate and sequence the accessible locations of a genome.<sup>52</sup> Historically, open chromatin has been identified by the hypersensitivity of genomic sites to the non-specific double-strand endonuclease DNase I.<sup>53</sup> DNase I can liberate accessible chromatin by preferentially cutting within nucleosome-free genomic regions, characterized as DNase I hypersensitive sites. Subsequently, with the advent of NGS, this gave rise to DNase sequencing. DNase sequencing allows the genome-wide identification of DNase I hypersensitive sites with specificity and sensitivity in a single reaction. This technique was extensively used by the ENCODE consortium<sup>48</sup> to unveil cell-specific chromatin accessibility and its relation to differential gene expression in various cell lines.<sup>54–57</sup> One drawback of DNase sequencing is that it requires large amounts of cells and the involvement of numerous sample preparation and enzyme titration steps. Very recently, a new technique was developed that circumvented

some of these limitations. This technique is termed assay for transposase-accessible chromatin sequencing. It is based on the ability of a hyperactive Tn5 transposase<sup>58,59</sup> to fragment DNA and integrate into active regulatory regions *in vivo*.<sup>60</sup> In a typical experiment, 500–50000 unfixed nuclei are tagged by Tn5 transposase with specific sequencing adapters. However, due to steric hindrance, the majority of these adapters can only integrate into regions of accessible chromatin, which is subsequently sequenced using paired-end NGS. Assay for transposase-accessible chromatin sequencing has been used to uncover open chromatin, nucleosome positioning and transcription factor footprints on genome-wide scale.<sup>60</sup> The sensitivity and specificity of assay for transposase-accessible chromatin sequencing is similar to DNase sequencing. However, it requires three to five orders of magnitude-fewer cells.

However, once potential enhancers are identified, they should be tested for their ability to regulate transcription. The most straightforward way to do this is by testing whether the DNA element can enhance transcription from a minimal core promoter. Several different forms of functional testing of enhancers have been developed, for a review see Shlyueva *et al.*<sup>50</sup> Two main classes of tests are currently available based on *in vivo* or *in vitro* models. The *in vivo* tests are performed in developing embryos (for example, in flies, zebrafish or mice<sup>61,62</sup>). These tests have readouts based on images and use staining patterns and intensity as a measure. However, the functionality of enhancers can also be tested *in vitro* using cell culture. Such tests either use luciferase or directly measure the abundance of reporter transcripts (by deep sequencing<sup>63</sup>). The advantage of the *in vitro* methods is a much higher throughput, at the cost of losing developmental information. Moreover, the model cell does not fully represent the tissue of interest *in vivo*. In general, these functional enhancer tests are extremely useful, however, their results should always be interpreted with caution, as currently these tests can only investigate the ability of elements to enhance transcription, and cannot yet validate other important regulatory elements such as repressors, organizers and boundaries. Furthermore, the elements are tested outside their genomic context, whereas their function is highly dependent on the topology of the chromatin, modifications of histones and the presence of other elements providing combinatorial activity.

An important limitation of the above methods is that the structural context of the genome is ignored. For transcription factors to function they must be juxtaposed to the gene promoter on which they act. Those positioned on regulatory elements, thousands or even hundreds of thousands of base pairs away, must therefore be physically folded to within the promoter's vicinity. This architectural design can now be explored using variants of the chromatin conformation capture technique (4C, 5C and HiC<sup>64</sup>). These techniques demonstrated that regulatory elements act at finite distances from their promoters (typically up to 1 Mb) and can form transcriptionally active domains constituting one or several genes.<sup>65</sup>

Taken together, these recent advances make it now possible to identify active regulatory regions of the genome and re-sequence them to determine whether they cause disease. However, as mentioned earlier, this will likely result in a large amount of variants per regulome/genome, some of which can be filtered out based on conservation, frequency or co-segregation. Still, to investigate which promoter/enhancer variants cause disease (or alter transcription), one would need to generate a large number of targeted promoter/enhancer sequences and quantify the effects of each variation on function. Recently, new techniques were published that alleviate this problem using massively parallel, high-throughput *in vivo* investigations of variation on regulatory elements (reviewed in Haberle and Lenhard<sup>66</sup>).



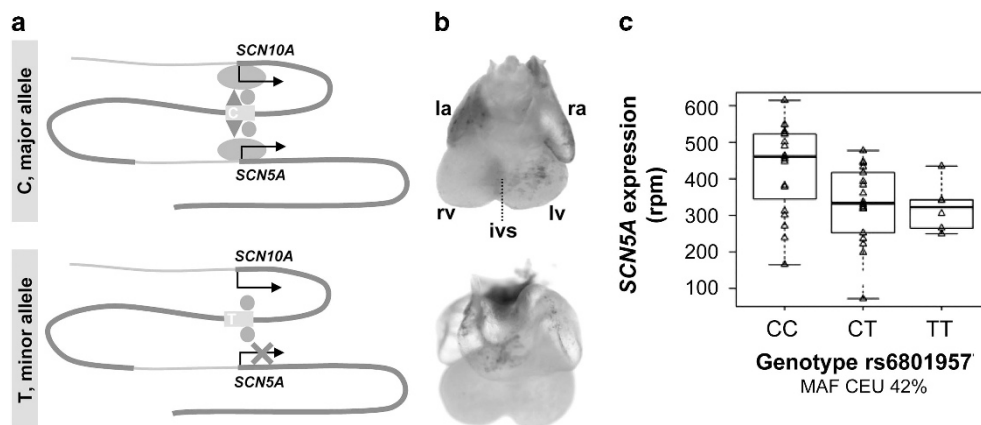
These techniques produce high-volume, complex outputs that rely on computation and statistical modeling to determine the effects of individual changes on transcriptional output. Such high-throughput, empirical measurements of the functional impact of SNVs in enhancers will substantially facilitate the analysis of noncoding variants from GWASs, regulome and genome-wide sequencing.

### ENHANCER VARIATION, MUTATION AND DISEASE TRAITS

Given that regulatory regions can now be identified, and the fact that many transcription factors have a crucial role in (ab)normal heart development, it is perhaps not surprising that investigators started to be interested in the regulatory regions that are bound by specific transcription factors in the search for causal disease mechanisms. The hypothesis is that variations/mutations in regulatory regions can predispose/cause disease by altering transcription factor binding, either directly or indirectly. As mentioned earlier, a group of core cardiac transcription factors controls heart development. These include, amongst others, the homeodomain protein NKX2-5 and T-box factors such as TBX3 and TBX5.<sup>2</sup> Cardiac expression of most of these factors continues throughout the development and into adult life. Not surprisingly, mutations in these crucial transcription factors have been identified in patients with CHD.<sup>10,14,67</sup> In humans, dominant mutations in *NKX2-5* cause a variety of cardiac anomalies, as well as atrioventricular conduction disease, even in the absence of cardiac structural malformations.<sup>11</sup> This suggests that NKX2-5 has a function in conduction system development that is independent of its role in cardiac morphogenesis. *TBX3* and *TBX5* belong to the TBX2 subfamily of T-box transcription factors and are involved in the development and maintenance of the cardiac conduction system.<sup>68</sup> Heterozygous mutations in *TBX3* cause ulnar-mammary syndrome in humans, which sometimes includes CHD.<sup>69,70</sup> Likewise, loss-of-function mutation of *TBX5* causes the Holt-Oram syndrome, a heart-hand defect, which can be associated with conduction abnormalities.<sup>71</sup> Interestingly, a reduction in the amount of *TBX3* in the developing mouse heart causes QRS prolongation and a spectrum of conduction defects<sup>72</sup> and loss of *TBX5* in the mature murine ventricular conduction system results in the loss of fast conduction,

cardiac arrhythmias and sudden cardiac death.<sup>73</sup> Given their important roles in heart development and especially the development and maintenance of the cardiac conduction system, it is, therefore, perhaps not surprising that these core transcription factors have been implicated by GWAS to associate with electrocardiogram parameters such as PR and QRS interval.<sup>74-76</sup> Furthermore, genome-wide mapping of the binding profile for NKX2-5, TBX3 and TBX5 in murine hearts uncovered that many cardiac enhancers co-localize with ion channel genes.<sup>77</sup> Case in point, a meta-analysis of a dozen GWASs implicated an intronic region of *SCN10A* to associate with prolongation of the QRS duration.<sup>76</sup> Recent work by van den Boogaard *et al.*<sup>78</sup> using high-resolution 4C-seq analysis, demonstrated that this region in *Scn10a* actually interacts with the promoter of *Scn5a* (Figure 1a). Analysis of BAC transgenic mouse strains harboring an engineered deletion of this region revealed that it was essential for *Scn5a* expression in cardiac tissue and functions as an enhancer. Moreover, introduction of the main GWAS single-nucleotide polymorphism associated with QRS prolongation (rs6801957) into the BAC reporter construct resulted in severe depression of *Scn5a* reporter expression in the compact myocardium and the cardiac conduction system (Figure 1b). They went on to show that in humans this common *SCN10A* single-nucleotide polymorphism actually associates with a reduced *SCN5A* expression (Figure 1c). Interestingly, earlier work by the same group demonstrated that this single-nucleotide polymorphism is located in a TBX-binding element within the intronic *SCN10A* enhancer, and actually affects TBX3 and TBX5 binding, by which it can diminish the enhancer function.<sup>77</sup> Taken together, these observations establish a genomic mechanism for how common genetic variation can influence cardiac physiology and predisposes to disease (Figure 1).

Similarly, it is perhaps not surprising that mutations in enhancers can also cause disease. This was first established by the work of Lettice *et al.*<sup>79</sup> who identified an enhancer region 1 Mb from its target gene *SHH*. They went on to identify point mutations in this enhancer that segregated with disease (polydactyly) in four unrelated families.<sup>79</sup> Another more recent example of enhancer variation causing disease is the work of Weedon *et al.*<sup>80</sup> who used whole-genome sequencing in



**Figure 1** (a) Schematic overview depicting how single-nucleotide variation in an enhancer located far away from its target promoters, can impact binding of transcription factors (red triangle, blue circle) and consequently the expression. In this example, presence of both transcription factors is needed for target gene expression of *SCN5A*, but not for *SCN10A*. The red triangle illustrates a T-box protein that is not able to bind the DNA when the minor allele is present. Loss of this T-box binding leads to alteration of *SCN5a* expression. (b) Transient transgenic embryos at E14.5 harboring a human enhancer with either the major (top) or minor (bottom) allele of SNP rs6801957. Presence of the major allele is sufficient to drive cardiac conduction system expression, which the minor allele cannot. (c) *SCN5A* expression in human hearts differs according to the SNP rs6801957 genotype. Figure adapted from van den Boogaard *et al.*<sup>78</sup> iva, inter ventricular septum; la, left atrium; lv, left ventricle; ra, right atrium, r.p.m., reads per million; rv, right ventricle; SNP, single-nucleotide polymorphism. A full color version of this figure is available at the *Journal of Human Genetics* journal online.

families with isolated pancreatic agenesis. They identified six segregating mutations in a 400-bp noncoding region of DNA that is located 25 kb upstream of the promoter of *PTF1A*. This 400-bp region acts as a developmental enhancer of *PTF1A*, which codes for the pancreas-specific transcription factor 1a. They went on to prove that the identified mutations abolish the enhancer's activity. Thus, point mutations residing in long-range regulatory elements are capable of causing congenital abnormalities. This phenomenon also has a role in CHD; a recent study by Smemo *et al.*<sup>81</sup> identified a homozygous mutation in an enhancer of *TBX5*, a key heart developmental gene, in a patient with a ventricular septal defect. They demonstrated that this single-basepair mutation in a *TBX5* enhancer, 90 kbp downstream of the gene, abrogates the ability of the enhancer to drive *TBX5* expression within the heart *in vivo*.<sup>81</sup> Taken together, the above examples demonstrate that disease-causing mutations, and disease predisposing variants (GWAS), need not reside in the exome *per se*, and that the search for causal variants should be expanded to include the whole genome, or at least the regulatory regions of the genome.

### FUTURE PERSPECTIVE

As discussed, new insights and technology have become available that can greatly advance our understanding of the genetic factors that contribute to CHD. In familial cases of CHD, identification of disease genes and/or loci will benefit the genetic counseling process for CHD-afflicted families. However, the overall goal will also be to provide actual therapeutic opportunities. Work by Tian *et al.*<sup>82</sup> proved that this need not be science fiction, as they showed that Wnt2 knockout mice, which have a heart malformation that resembles a complete atrioventricular canal defect in humans, can actually be rescued by the transient pharmacological activation of the Wnt signaling network using LiCl. Knowledge of such networks, and how variants in these networks affect their outcome, is crucial in understanding how to manipulate them for therapeutic opportunities. Furthermore, variations in regulatory regions of key (heart) developmental genes can cause disease, most likely by a change in expression level (and perhaps pattern) of target genes. This in turn, will affect particular developmental networks with a time and place-specific output. However, presently, therapeutic opportunities are limited and much more knowledge about the molecular genetics and the molecular pathology of CHD is needed before these can become reality.

### CONFLICT OF INTEREST

The authors declare no conflict of interest.

### ACKNOWLEDGEMENTS

We are funded by The Netherlands CardioVascular Research Initiative (Dutch Heart Foundation, Dutch Federation of University Medical Centres, the Netherlands Organization for Health Research and Development, and the Royal Netherlands Academy of Sciences) CONCOR-genes (AVP, CRB and VMC) and PREDICT (CRB) projects, and The Netherlands Heart Foundation grant 2013T091 'COBRA3' (VMC).

- 4 Gilboa, S. M., Salemi, J. L., Nembhard, W. N., Fixler, D. E. & Correa, A. Mortality resulting from congenital heart disease among children and adults in the United States, 1999 to 2006. *Circulation* **122**, 2254–2263 (2010).
- 5 Van der Bom, T., Winter, M. M., Bouma, B. J., Groenink, M., Vliegen, H. W., Pieper, P. G. *et al.* Effect of valsartan on systemic right ventricular function: a double-blind, randomized, placebo-controlled pilot trial. *Circulation* **127**, 322–330 (2013).
- 6 Koyak, Z., Harris, L., de Groot, J. R., Silversides, C. K., Oechslin, E. N., Bouma, B. J. *et al.* Sudden cardiac death in adult congenital heart disease. *Circulation* **126**, 1944–1954 (2012).
- 7 Van der Bom, T., van der Palen, R. L. F., Bouma, B. J., van Veldhuisen, S. L., Vliegen, H. W., Konings, T. C. *et al.* Persistent neo-aortic growth during adulthood in patients after an arterial switch operation. *Heart* **100**, 1360–1365 (2014).
- 8 Van der Bom, T., Zomer, A. C., Zwinderman, A. H., Meijboom, F. J., Bouma, B. J. & Mulder, B. J. M. The changing epidemiology of congenital heart disease. *Nat. Rev. Cardiol.* **8**, 50–60 (2011).
- 9 Peyvandi, S., Ingall, E., Woyciechowski, S., Garbarini, J., Mitchell, L. E. & Goldmuntz, E. Risk of congenital heart disease in relatives of probands with conotruncal cardiac defects: an evaluation of 1,620 families. *Am. J. Med. Genet. A* **164A**, 1490–1495 (2014).
- 10 Andersen, T. A., Troelsen, K. D. L. L. & Larsen, L. A. Of mice and men: molecular genetics of congenital heart disease. *Cell Mol. Life Sci.* **71**, 1327–1352 (2013).
- 11 Schott, J. J., Benson, D. W., Basson, C. T., Pease, W., Silberbach, G. M., Moak, J. P. *et al.* Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science* **281**, 108–111 (1998).
- 12 Garg, V., Kathiriyi, I. S., Barnes, R., Schluterman, M. K., King, I. N., Butler, C. A. *et al.* GATA4 mutations cause human congenital heart defects and reveal an interaction with *TBX5*. *Nature* **424**, 443–447 (2003).
- 13 Garg, V., Muth, A. N., Ransom, J. F., Schluterman, M. K., Barnes, R., King, I. N. *et al.* Mutations in *NOTCH1* cause aortic valve disease. *Nature* **437**, 270–274 (2005).
- 14 Lalani, S. R. & Belmont, J. W. Genetic basis of congenital cardiovascular malformations. *Eur. J. Med. Genet.* **57**, 402–413 (2014).
- 15 Fahed, A. C., Gelb, B. D., Seidman, J. G. & Seidman, C. E. Genetics of congenital heart disease: the glass half empty. *Circ. Res.* **112**, 707–720 (2013).
- 16 Barnett, P. & Postma, A. V. Genetics of congenital heart disease: Beyond half-measures. *Trends Cardiovasc. Med.* **25**, 302–304 (2014).
- 17 Ching, Y.-H., Ghosh, T. K., Cross, S. J., Packham, E. A., Honeyman, L., Loughna, S. *et al.* Mutation in myosin heavy chain 6 causes atrial septal defect. *Nat. Genet.* **37**, 423–428 (2005).
- 18 Postma, A. V., Van Engelen, K., van de Meerakker, J., Rahman, T., Probst, S., Baars, M. J. H. *et al.* Mutations in the sarcomere gene MYH7 in Ebstein anomaly. *Circ. Cardiovasc. Genet.* **4**, 43–50 (2011).
- 19 Mefford, H. C., Sharp, A. J., Baker, C., Itsara, A., Jiang, Z., Buysse, K. *et al.* Recurrent rearrangements of chromosome 1q21.1 and variable pediatric phenotypes. *N. Engl. J. Med.* **359**, 1685–1699 (2008).
- 20 Gourdie, R. G., Severs, N. J., Green, C. R., Rothery, S., Germroth, P. & Thompson, R. P. The spatial distribution and relative abundance of gap-junctional connexin40 and connexin43 correlate to functional properties of components of the cardiac atrioventricular conduction system. *J. Cell Sci.* **105**, 985–991 (1993).
- 21 Gros, D., Jarry-Guichard, T., Ten Velde, I., de Maziere, A., van Kempen, M. J., Davoust, J. *et al.* Restricted distribution of connexin40, a gap junctional protein, in mammalian heart. *Circ. Res.* **74**, 839–851 (1994).
- 22 Ou, Z., Berg, J. S., Yonath, H., Enciso, V. B., Miller, D. T., Picker, J. *et al.* Microduplications of 22q11.2 are frequently inherited and are associated with variable phenotypes. *Genet. Med.* **10**, 267–277 (2008).
- 23 Soemedi, R., Topf, A., Wilson, I. J., Darlay, R., Rahman, T., Glen, E. *et al.* Phenotype-specific effect of chromosome 1q21.1 rearrangements and GJA5 duplications in 2436 congenital heart disease patients and 6760 controls. *Hum. Mol. Genet.* **21**, 1513–1520 (2012).
- 24 Gu, H., Smith, F. C., Taffet, S. M. & Delmar, M. High incidence of cardiac malformations in connexin40-deficient mice. *Circ. Res.* **93**, 201–206 (2003).
- 25 Breckpot, J., Thienpont, B., Arens, Y., Tranchevent, L. C., Vermeesch, J. R., Moreau, Y. *et al.* Challenges of interpreting copy number variation in syndromic and non-syndromic congenital heart defects. *Cytogenet. Genome Res.* **135**, 251–259 (2011).
- 26 Goldmuntz, E., Paluru, P., Glessner, J., Hakonarson, H., Biegel, J. A., White, P. S. *et al.* Microdeletions and microduplications in patients with congenital heart disease and multiple congenital anomalies. *Congenit. Heart Dis.* **6**, 592–602.
- 27 Lalani, S. R., Shaw, C., Wang, X., Patel, A., Patterson, L. W., Kolodziejska, K. *et al.* Rare DNA copy number variants in cardiovascular malformations with extracardiac abnormalities. *Eur. J. Hum. Genet.* **21**, 173–181 (2013).
- 28 Thienpont, B., Mertens, L., de Ravel, T., Eyskens, B., Boshoff, D., Maas, N. *et al.* Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients. *Eur. Heart J.* **28**, 2778–2784 (2007).
- 29 Tomita-Mitchell, A., Mahnke, D. K., Struble, C. A., Tuffnell, M. E., Stamm, K. D., Hildebrand, M. *et al.* Human gene copy number spectra analysis in congenital heart malformations. *Physiol. Genomics* **44**, 518–541 (2012).
- 30 Greenway, S. C., Pereira, A. C., Lin, J. C., DePalma, S. R., Israel, S. J., Mesquita, S. M. *et al.* De novo copy number variants identify new genes and loci in isolated sporadic tetralogy of Fallot. *Nat. Genet.* **41**, 931–935 (2009).
- 31 Silversides, C. K., Lionel, A. C., Costain, G., Merico, D., Migita, O., Liu, B. *et al.* Rare copy number variations in adults with tetralogy of Fallot implicate novel risk gene pathways. *PLoS Genet.* **8**, e1002843 (2012).

1 Van der Bom, T., Bouma, B. J., Meijboom, F. J., Zwinderman, A. H. & Mulder, B. J. M. The prevalence of adult congenital heart disease, results from a systematic review and evidence based calculation. *Am. Heart J.* **164**, 568–575 (2012).

2 McCulley, D. J. & Black, B. L. Transcription factor pathways and congenital heart disease. *Curr. Top. Dev. Biol.* **100**, 253–277 (2012).

3 Bruneau, B. G. Signaling and transcriptional networks in heart development and regeneration. *Cold Spring Harb. Perspect. Biol.* **5**, a008292 (2013).

- 32 Soemedi, R., Wilson, I. J., Bentham, J., Darlay, R., Töpf, A., Zelenika, D. *et al*. Contribution of global rare copy-number variants to the risk of sporadic congenital heart disease. *Am. J. Hum. Genet.* **91**, 489–501 (2012).
- 33 Sanders, S. J., Murtha, M. T., Gupta, A. R., Murdoch, J. D., Raubeson, M. J., Willsey, A. J. *et al*. De novo mutations revealed by whole-exome sequencing are strongly associated with autism. *Nature* **485**, 237–241 (2012).
- 34 Boycott, K. M., Vanstone, M. R., Bulman, D. E. & MacKenzie, A. E. Rare-disease genetics in the era of next-generation sequencing: discovery to translation. *Nat. Rev. Genet.* **14**, 681–691 (2013).
- 35 Marsman, R. F., Tan, H. L. & Bezzina, C. R. Genetics of sudden cardiac death caused by ventricular arrhythmias. *Nat. Rev. Cardiol.* **11**, 96–111 (2014).
- 36 Zaidi, S., Choi, M., Wakimoto, H., Ma, L., Jiang, J., Overton, J. D. *et al*. De novo mutations in histone-modifying genes in congenital heart disease. *Nature* **498**, 220–223 (2013).
- 37 Al Turki, S., Manickaraj, A. K., Mercer, C. L., Gerety, S. S., Hitz, M.-P., Lindsay, S. *et al*. Rare variants in NR2F2 cause congenital heart defects in humans. *Am. J. Hum. Genet.* **94**, 574–585 (2014).
- 38 Wu, S., Cheng, C.-M., Lanz, R. B., Wang, T., Respress, J. L., Ather, S. *et al*. Atrial identity is determined by a COUP-TFII regulatory network. *Dev. Cell* **25**, 417–426 (2013).
- 39 Chetaille, P., Preuss, C., Burkhard, S., Côté, J.-M., Houde, C., Castilloux, J. *et al*. Mutations in SGOL1 cause a novel cohesinopathy affecting heart and gut rhythm. *Nat. Genet.* **46**, 1245–1249 (2014).
- 40 Bainbridge, M. N., Davis, E. E., Choi, W.-Y., Dickson, A., Martinez, H. R., Wang, M. *et al*. Loss of function mutations in NNT are associated with left ventricular noncompaction. *Circ. Cardiovasc. Genet.* (e-pub ahead of print 29 May 2015; doi:10.1161/CIRCGENETICS.115.001026).
- 41 Andelfinger, G. Next-generation sequencing in congenital heart disease: do new brooms sweep clean? *J. Am. Coll. Cardiol.* **64**, 2507–2509 (2014).
- 42 Arrington, C. B., Bleyl, S. B., Matsunami, N., Bonnell, G. D., Otterud, B. E. M., Nielsen, D. C. *et al*. Exome analysis of a family with pleiotropic congenital heart disease. *Circ. Cardiovasc. Genet.* **5**, 175–182 (2012).
- 43 Martin, L. J., Pilipenko, V., Kaufman, K. M., Cripe, L., Kottyan, L. C., Keddache, M. *et al*. Whole exome sequencing for familial bicuspid aortic valve identifies putative variants. *Circ. Cardiovasc. Genet.* **7**, 677–683 (2014).
- 44 Parrott, A. & Ware, S. M. The Role of the Geneticist and Genetic Counselor in an ACHD Clinic. *Prog. Pediatr. Cardiol.* **34**, 15–20 (2012).
- 45 Cordell, H. J., Topf, A., Mamasoula, C., Postma, A. V., Bentham, J., Zelenika, D. *et al*. Genome-wide association study identifies loci on 12q24 and 13q32 associated with Tetralogy of Fallot. *Hum. Mol. Genet.* **22**, 1473–1481 (2013).
- 46 Cordell, H. J., Bentham, J., Topf, A., Zelenika, D., Heath, S., Mamasoula, C. *et al*. Genome-wide association study of multiple congenital heart disease phenotypes identifies a susceptibility locus for atrial septal defect at chromosome 4p16. *Nat. Genet.* **45**, 822–824 (2013).
- 47 Hu, Z., Shi, Y., Mo, X., Xu, J., Zhao, B., Lin, Y. *et al*. A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nat. Genet.* **45**, 818–821 (2013).
- 48 Bernstein, B. E., Birney, E., Dunham, I., Green, E. D., Gunter, C. & Snyder, M. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
- 49 De Laat, W. & Duboule, D. Topology of mammalian developmental enhancers and their regulatory landscapes. *Nature* **502**, 499–506 (2013).
- 50 Shlyueva, D., Stampfel, G. & Stark, A. Transcriptional enhancers: from properties to genome-wide predictions. *Nat. Rev. Genet.* **15**, 272–286 (2014).
- 51 Robertson, G., Hirst, M., Bainbridge, M., Bilenky, M., Zhao, Y., Zeng, T. *et al*. Genome-wide profiles of STAT1 DNA association using chromatin immunoprecipitation and massively parallel sequencing. *Nat. Methods* **4**, 651–657 (2007).
- 52 Tsompana, M. & Buck, M. J. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* **7**, 33 (2014).
- 53 Weintraub, H. & Groudine, M. Chromosomal subunits in active genes have an altered conformation. *Science* **193**, 848–856 (1976).
- 54 John, S., Sabo, P. J., Thurman, R. E., Sung, M.-H., Biddie, S. C., Johnson, T. A. *et al*. Chromatin accessibility pre-determines glucocorticoid receptor binding patterns. *Nat. Genet.* **43**, 264–268 (2011).
- 55 Thurman, R. E., Rynes, E., Humbert, R., Vierstra, J., Maurano, M. T., Haugen, E. *et al*. The accessible chromatin landscape of the human genome. *Nature* **489**, 75–82 (2012).
- 56 Neph, S., Vierstra, J., Stergachis, A. B., Reynolds, A. P., Haugen, E., Vernot, B. *et al*. An expansive human regulatory lexicon encoded in transcription factor footprints. *Nature* **489**, 83–90 (2012).
- 57 Boyle, A. P., Davis, S., Shulha, H. P., Meltzer, P., Margulies, E. H., Weng, Z. *et al*. High-resolution mapping and characterization of open chromatin across the genome. *Cell* **132**, 311–322 (2008).
- 58 Goryshin, I. Y. & Reznikoff, W. S. Tn5 *in vitro* transposition. *J. Biol. Chem.* **273**, 7367–7374 (1998).
- 59 Adey, A., Morrison, H. G., Asan, Xun, X., Kitzman, J. O., Turner, E. H. *et al*. Rapid, low-input, low-bias construction of shotgun fragment libraries by high-density *in vitro* transposition. *Genome Biol.* **11**, R119 (2010).
- 60 Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y. & Greenleaf, W. J. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. *Nat. Methods* **10**, 1213–1218 (2013).
- 61 Manning, L., Heckscher, E. S., Purice, M. D., Roberts, J., Bennett, A. L., Kroll, J. R. *et al*. A resource for manipulating gene expression and analyzing cis-regulatory modules in the Drosophila CNS. *Cell Rep.* **2**, 1002–1013 (2012).
- 62 Visel, A., Taher, L., Girgis, H., May, D., Golonzhka, O., Hoch, R. V. *et al*. A high-resolution enhancer atlas of the developing telencephalon. *Cell* **152**, 895–908 (2013).
- 63 Arnold, C. D., Gerlach, D., Stelzer, C., Boryń ŁM, Rath, M. & Stark, A. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science* **339**, 1074–1077 (2013).
- 64 De Wit, E. & de Laat, W. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* **26**, 11–24 (2012).
- 65 Jin, F., Li, Y., Dixon, J. R., Selvaraj, S., Ye, Z., Lee, A. Y. *et al*. A high-resolution map of the three-dimensional chromatin interactome in human cells. *Nature* **503**, 290–294 (2013).
- 66 Haberle, V. & Lenhard, B. Dissecting genomic regulatory elements *in vivo*. *Nat. Biotechnol.* **30**, 504–506 (2012).
- 67 Lahm, H., Schön, P., Doppler, S., Dreßen, M., Cleuziou, J., Deutsch, M.-A. *et al*. Tetralogy of fallot and hypoplastic left heart syndrome—complex clinical phenotypes meet complex genetic networks. *Curr. Genomics* **16**, 141–158 (2015).
- 68 Christoffels, V. M., Smits, G. J., Kispert, A. & Moorman, A. F. Development of the pacemaker tissues of the heart. *Circ. Res.* **106**, 240–254 (2010).
- 69 Bamshad, M., Lin, R. C., Law, D. J., Watkins, W. C., Krakowiak, P. A., Moore, M. E. *et al*. Mutations in human TBX3 alter limb, apocrine and genital development in ulnar-mammary syndrome. *Nat. Genet.* **16**, 311–315 (1997).
- 70 Meneghini, V., Odent, S., Platonova, N., Egea, A. & Merlo, G. R. Novel TBX3 mutation data in families with Ulnar-Mammary syndrome indicate a genotype-phenotype relationship: mutations that do not disrupt the T-domain are associated with less severe limb defects. *Eur. J. Med. Genet.* **49**, 151–158 (2006).
- 71 Barnett, P. & Postma, A. V. Molecular genetics of Holt–Oram syndrome. *eLS* **256**, 799–804 (2014).
- 72 Frank, D. U., Emechebe, U., Thomas, K. R. & Moon, A. M. Mouse TBX3 mutants suggest novel molecular mechanisms for Ulnar-mammary syndrome. *PLoS One* **8**, e67841 (2013).
- 73 Arnolds, D. E., Liu, F., Fahrenbach, J. P., Kim, G. H., Schillinger, K. J., Smemo, S. *et al*. TBX5 drives Scn5a expression to regulate cardiac conduction system function. *J. Clin. Invest.* **122**, 2509–2518 (2012).
- 74 Holm, H., Gudbjartsson, D. F., Arnar, D. O., Thorleifsson, G., Thorgeirsson, G., Stefansdottir, H. *et al*. Several common variants modulate heart rate, PR interval and QRS duration. *Nat. Genet.* **42**, 117–122 (2010).
- 75 Pfeufer, A., van Noord, C., Marcicante, K. D., Arking, D. E., Larson, M. G., Smith, A. V. *et al*. Genome-wide association study of PR interval. *Nat. Genet.* **42**, 153–159 (2010).
- 76 Sotoodehnia, N., Isaacs, A., de Bakker, P. I., Dorr, M., Newton-Cheh, C., Nolte, I. M. *et al*. Common variants in 22 loci are associated with QRS duration and cardiac ventricular conduction. *Nat. Genet.* **42**, 1068–1076 (2010).
- 77 Van den Boogaard, M., Wong, L. Y. E., Tessadori, F., Bakker, M. L., Dreizehnter, L. K., Wakker, V. *et al*. Genetic variation in T-box binding element functionally affects SCN5A/SCN10A enhancer. *J. Clin. Invest.* **122**, 2519–2530 (2012).
- 78 Van den Boogaard, M., Smemo, S., Burnicka-Turek, O., Arnolds, D. E., van de Werken, H. J. G., Klous, P. *et al*. A common genetic variant within SCN10A modulates cardiac SCN5A expression. *J. Clin. Invest.* **124**, 1844–1852 (2014).
- 79 Lettice, L. A., Heaney, S. J. H., Purdie, L. A., Li, L., de Beer, P., Oostra, B. A. *et al*. A long-range Shh enhancer regulates expression in the developing limb and fin and is associated with preaxial polydactyly. *Hum. Mol. Genet.* **12**, 1725–1735 (2003).
- 80 Weedon, M. N., Cebola, I., Patch, A.-M., Flanagan, S. E., De Franco, E., Caswell, R. *et al*. Recessive mutations in a distal PTF1A enhancer cause isolated pancreatic agenesis. *Nat. Genet.* **46**, 61–64 (2014).
- 81 Smemo, S., Campos, L. C., Moskowitz, I. P., Krieger, J. E., Pereira, A. C. & Nobrega, M. A. Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* **21**, 3255–3263 (2012).
- 82 Tian, Y., Yuan, L., Goss, A. M., Wang, T., Yang, J., Lepore, J. J. *et al*. Characterization and *in vivo* pharmacological rescue of a Wnt2-Gata6 pathway required for cardiac inflow tract development. *Dev. Cell* **18**, 275–287 (2010).