



Review article

Navigating the non-coding genome in heart development and Congenital Heart Disease



Gulrez Chahal^{a,b}, Sonika Tyagi^{c,d,**}, Mirana Ramialison^{a,b,*}

^a Australian Regenerative Medicine Institute (ARMI), 15 Innovation Walk, Monash University, Wellington Road, Clayton, 3800, VIC, Australia

^b Systems Biology Institute (SBI), Wellington Road, Clayton, 3800, VIC, Australia

^c School of Biological Sciences, Monash University, Wellington Road, Clayton, 3800, VIC, Australia

^d Australian Genome Research Facility, 305 Grattan Street, Melbourne, VIC, 3000, Australia

ARTICLE INFO

Keywords:

Epigenome

Non-coding genome

Regulatory elements

High-throughput sequencing

Congenital heart disease

Heart development

ABSTRACT

Congenital Heart Disease (CHD) is characterised by a wide range of cardiac defects, from mild to life-threatening, which occur in babies worldwide. To date, there is no cure to CHD, however, progress in surgery has reduced its mortality allowing children affected by CHD to reach adulthood. In an effort to understand its genetic basis, several studies involving whole-genome sequencing (WGS) of patients with CHD have been undertaken and generated a great wealth of information. The majority of putative causative mutations identified in WGS studies fall into the non-coding part of the genome. Unfortunately, due to the lack of understanding of the function of these non-coding mutations, it is challenging to establish a causal link between the non-coding mutation and the disease. Thus, here we review the state-of-the-art approaches to interpret non-coding mutations in the context of CHD and address the following questions:

What are the non-coding sequences important for cardiac function? Which technologies are used to identify them? Which resources are available to analyse them? What mutations are expected in these non-coding sequences? Learning from developmental process, what is their expected role in CHD?

1. Introduction

Congenital heart disease (CHD) refers to the group of abnormalities observed in the heart which disrupt its normal function and development. It is one of the most common birth defects observed in newborns in various parts of the world (Tennant et al., 2010), (Mitchell et al., 1971). The incidence of CHD-related defects, albeit variable, remains high in densely populated countries with high fertility rate, increasing the need of efficient diagnosis and clinical management (Hoffman, 2013), (van der Linde et al., 2011).

CHD has two main causes, environmental and/or genetic (Vecoli et al., 2014).

Environmental exposures during fetal heart development, specifically fetal placental-maternal interaction between weeks 2 through 7 of gestation, play a crucial role in predicting CHD occurrence (Vecoli et al., 2014), (Jenkins et al., 2007). These exposures include teratogens (such as pesticides), and maternal therapeutics (such as barbituric acid or chemotherapeutic agents). Smoking, alcohol consumption, and

gestational hypoxia, have also been associated with risk of CHD (Vecoli et al., 2014), (Jenkins et al., 2007), (Shi et al., 2016), (Patel and Burns, 2013).

Decades of extensive research on the molecular control of cardiac development has led to a deeper understanding of the genetic basis of CHD. In particular, the ability to generate *in vivo* and *in vitro* genetic models to explore cardiac phenotypes, enabled us to identify genes and developmental pathways essential for cardiogenesis and CHD pathogenesis (Moon, 2008). Several studies have established links between genetic variations in cardiac genes and CHD. Single gene variations in cardiac developmental genes were identified as a cause for about 3–5% of the disease. More recent evidence points out to the polygenic (or complex) nature of CHD (Blue et al., 2012).

Whether monogenic or complex, CHD occurs in sporadic as well as familial forms. Massively parallel sequencing technologies have indicated that *de novo* mutations contribute to ~10% of its sporadic form (Zaidi et al., 2013). These include deleterious mutations such as non-sense, frameshifts and splicing variants. Additionally, genetic screening

* Corresponding author. Australian Regenerative Medicine Institute (ARMI), 15 Innovation Walk, Monash University, Wellington Road, Clayton, 3800, VIC, Australia.

** Corresponding author. School of Biological Sciences, Monash University, Wellington Road, Clayton, 3800, VIC, Australia.

E-mail addresses: sonika.tyagi@monash.edu (S. Tyagi), mirana.ramialison@monash.edu (M. Ramialison).

<https://doi.org/10.1016/j.diff.2019.05.001>

Received 3 October 2018; Received in revised form 14 January 2019; Accepted 6 May 2019

Available online 08 May 2019

0301-4681/© 2019 The Authors. Published by Elsevier B.V. on behalf of International Society of Differentiation. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of patients with familial CHD using gene panels identified variants in ~31–46% of the cohorts, while the cause remained unknown for the rest of the cohort (Blue et al., 2017), (Jia et al., 2015), (Blue et al., 2014). Furthermore, CHD may occur in association with extra-cardiac phenotypes, such as Down syndrome. Chromosomal anomalies associated with these syndromes contribute to 8–10% of the CHD cases (Blue et al., 2012).

Despite the large body of information available (McDermott et al., 2005), (Zahavich et al., 2017), (Granados-Riveron et al., 2012), the number of known genetic factors contributing to CHD remains limited. For instance, genetic screening platforms dedicated to diagnose CHD contain less than 100 genes (Cowan and Ware, 2015), and the majority of patients with familial CHD still lack a genetic diagnosis.

While our understanding of the etiology of CHD has slowly progressed, CHD management has drastically improved. Indeed, with advances in surgical procedures, prenatal and neonatal diagnosis, and improved guidelines for management of the disease, mortality rate has decreased (Khoshnood et al., 2005), (Baumgartner et al., 2010). Consequently, patients treated with surgery and carrying gene mutations linked with CHD may transmit the mutation to their offspring through the next generation, leading to an increase in incidence of the disease (Richards, 2010). It is therefore imperative to develop comprehensive genetic tests allowing patients to have access to complete information regarding the risk of CHD when planning a family. Hence, there is a fundamental need to identify genetic factors controlling to heart development in order to provide novel insights in the genetic bases of CHD.

One of the most important reasons for lack of our understanding of the genetic basis of CHD is limiting the focus on protein-coding regions of the genome. Most of the efforts in studying the human genome have been concentrating on screening for alterations in protein-coding genes. However, these compose only ~2% of our genome. The greater part of the genome comprises of non-coding regions which harbour functional non-coding DNA regulatory elements such as promoters and enhancers, and genomic regions encoding for small and long non-coding RNAs (ncRNA) (Perthea and Salzberg, 2010).

Mutations in the non-coding regions of the genome have been associated with a wide range of diseases such as cancer and various congenital anomalies (Scacheri and Scacheri, 2015), (Hornshoj et al., 2018). However, their role in CHD has been largely unexplored. Hence, studying non-coding mutations in the context of heart development is a very effective approach to identify novel genetic causes for CHD. This review focuses on our current knowledge of the cardiac non-coding genome and techniques and resources that have been recently made available to explore its role in CHD.

2. cis-Regulatory elements (cREs) in heart development and how to find them

2.1. Definition of cis-regulatory elements (cREs)

The machinery involved in spatio-temporal regulation of gene expression includes transcription factors (TFs) binding DNA at sequence-specific binding sites within cis-regulatory elements (cREs) (Maston et al., 2006), (Narlikar and Ovcharenko, 2009). Thus, cREs such as promoters, enhancers, silencers and insulators are stretches of open chromatin regions accessible to TFs to regulate the expression of their target genes. The binding of TFs to their respective transcription factor binding sites (TFBS) may occur in a cell type- or tissue-specific manner (Narlikar and Ovcharenko, 2009), (Spitz and Furlong, 2012). The composition of TFBS within the regulatory elements (also called cis-regulatory module) contributes to the tight spatio-temporal control of gene expression (Weingarten-Gabbay and Segal, 2014), (Zinzen et al., 2009).

There are a limited number of studies which explore the role of cREs in CHD etiology. Since CHD is characterised by impaired heart

formation, there is an urgent need to decipher the cis-regulatory mechanisms vital for cardiogenesis in order to identify novel cis-regulatory pathways contributing to CHD. In a bid to chart the regulatory blueprint of the developing heart, several studies have been undertaken in cardiac cells at different cardiac developmental stages as well as the whole heart (Martinez et al., 2015), (Freire-Pritchett et al., 2017), (Wamstad et al., 2014).

In this section we discuss the current principles, technologies and approaches to identify cREs in heart development, based on their characteristic features as TF binding sequences and regions of open chromatin. Furthermore, we also discuss the resources made available in recent years to explore their implications in CHD.

2.2. Detection of cis-regulatory elements (cREs) as targets of cardiac transcription factors (TFs)

Unlike protein-coding genes, to date there is no known “regulatory code” allowing direct detection of cREs in the genome. Since cREs harbour transcription factor binding sites, identifying cardiac-specific TFs and designing experiments to recognise their target regions genome-wide provides the first step to discover cardiac-specific cREs.

2.2.1. cREs discovered from known cardiac TFs

Multiple transcription factor networks have been identified in cardiac tissues, which when altered, govern the development and etiology of heart disorders, including CHD. Table 3 presents a list of known TFs with a key role in heart development and CHD, collated from different studies (McCulley and Black, 2012), (Herrmann et al., 2012), (May et al., 2011). Notably, NKX2-5, GATA4, and TBX5 have been extensively studied in heart development as well as in CHD. The Nkx2-5 gene, encoding a protein from the homeodomain family, has been demonstrated to have a central role throughout heart development, from proliferation of cardiac precursors to defining chamber identity in murine models (Prall et al., 2007). For these various processes, it interacts with several developmental regulators such as Isl1, Bmp2, Smad1, Irx4 and Mef2c; indicating the complexity of transcriptional control in the heart (Prall et al., 2007), (Liberatore et al., 2002), (Vincentz et al., 2008), (Barth et al., 2010), (Schwartz and Olson, 1999). GATA4, encodes for a zinc finger protein which is involved in cardiac morphogenesis and proliferation of cardiomyocytes. It also interacts with several signaling pathway components including cyclin D2 and Cdk4, Tbx5, and MEF2 proteins (McCulley and Black, 2012), (Morin et al., 2000). Another critical and well explored TF, TBX5, is involved in cardiac morphogenesis and development, and has been commonly associated with cardiac defects occurring in Holt-Oram syndrome (Bruneau et al., 2001).

Genome-wide mapping assays can be performed to identify the target of these known cardiac TFs, thus indirectly identifying cardiac cREs. The most widely used technique to identify the binding sites of TFs is Chromatin immunoprecipitation (ChIP) (Table 1). Since the TF machinery is recruited at these accessible DNA sequences, an antibody specific to the TF of interest can be used to identify the bound DNA regions within the genome. These binding sites are deciphered using high-throughput sequencing (ChIP-seq) (Raha et al., 2010). In ChIP-seq, the DNA sequences obtained as “reads” are filtered based on various quality parameters and mapped to a reference genome. The enriched TF-specific precipitated regions are compared with the regions obtained by fragmenting DNA (or input DNA) to identify differential peaks representing binding regions for these TFs. In addition, an enzyme technology-based method, DamID (DNA adenine methyltransferase identification), has also been used to identify TF binding sites in the cardiac context (van Steensel and Henikoff, 2000), (Bouveret et al., 2015), (Ramialison et al., 2017). *Escherichia coli* DNA adenine methyltransferase (Dam) is fused with a TF of interest. This Dam fusion binds to DNA and locally methylates the surrounding adenine, thereby capturing its target region (Bouveret et al., 2015), (Ramialison et al., 2017).

Table 1
Current high-throughput technologies, their principle and genomic information they capture as outcome.

Sequencing technology	Methodology	Outcome
ChIP-seq	Sequencing of specific isolated protein-DNA interaction. This protein can be transcription factors, co-factors or histones.	Putative enhancers, TFBS, respective histone marks
RNA-seq	Evaluates mRNA expressions for a given sample	Expression pattern of translated protein
ATAC-seq	Identifies open chromatin regions in the genome using hyperactive Tn5 transposase which inserts adapter sequences in the accessible regions.	Open chromatin regions and nucleosome position
DNase-seq	Identifies loosely packed DNase I hypersensitive sites in the chromatin by subjecting them to DNase enzyme digestion. These sites in the genome have been correlated to frequently contain regulatory elements and can be sequenced genome wide.	TF binding sites, regulatory elements
MNase-seq	Identifies nucleosome positioning using micrococcal nuclease digestion followed by high throughput sequencing	Regulatory elements, nucleosomes, DNA-binding proteins
FAIRE-seq	Identifies nucleosome-depleted regions by using formaldehyde to cross-link the nucleosome-rich regions. The nucleosome-depleted regions can be separated by further chemical treatment and fluorescent labelling, followed by sequencing	DNaseI hypersensitive sites, transcriptional start sites, and active promoters

These genome-wide mapping assays have been used to detect cREs upstream and downstream of cardiac genes (He et al., 2011, 2014), that are crucial for cardiac development. For instance, regulatory elements guiding spatial expression of *Nkx2-5* in the cardiac crescent have been identified in mouse models (Reecy et al., 1999). Examples of cardiac cREs mutated in disease have also emerged such as the identification of mutation in an enhancer region in proximity of the *TBX5* gene associated with the Holt–Oram syndrome (Smemo et al., 2012), (Bruneau et al., 2001), Another study reported an interaction between the enhancer of *Scn10A*, a gene involved in Brugada syndrome in humans, with the promoter in *Scn5A*, to govern the expression of *Scn5A* in mice (van den Boogaard et al., 2014).

Furthermore, these genome-wide mapping assays revealed generic properties of cardiac transcription networks. For instance, ChIP-seq experiments with antibodies specific to Gata4, Mef2a, Nkx2-5 and Srf performed in mouse cardiomyocytes investigated the TFBSs lying 2.5 kb–10 kb upstream of the transcription start sites (Schlesinger et al., 2011). They revealed the combinatorial nature of cardiac TFs binding at their enhancer target sequences. The cardiac TF GATA4 in particular is shown to interact with histone acetyltransferase p300 and aids its recruitment to the chromatin (He et al., 2014). DamID has been used to identify cREs that are targets of NKX2-5, and novel binding sites bound by NKX2-5 mutants associated with CHD (Bouveret et al., 2015; Ramialison et al., 2017). These studies revealed that mutant TFs target a very distinct set of novel cREs, that are not the endogenous targets of wild-type TFs.

From this great wealth of information on cardiac TF binding, machine learning approaches can be used to infer whether there is a specific combinatorial TF signature that correlates with gene expression or cell-type specificity (Wang et al., 2018). One of the earliest applications of this technology used Support Vector machines to train on *Drosophila melanogaster* cardiac enhancers with known TFBS composition (obtained from ChIP datasets) and known activity (obtained from

in situ hybridization on neighbouring genes). This training phase identified precise TFBS combinations that, not only were shared by enhancers driving similar gene expression, but also could be used to predict gene expression driven by an unknown enhancer (Zinzen et al., 2009). Further development of methods based on artificial intelligence combined with the accumulation of sequencing information will fast-forward our understanding of the regulatory code.

2.2.2. New cardiac TF prediction

Out of the ~3000 TFs predicted in human genome (Brivanlou and Darnell, 2002), to date only 33 are known to play a key role in heart development and disease (Table 3). Hence, there are likely to be many more TFs having a role in the heart yet to be elucidated. Identifying these TFs and their genomic targets will greatly improve the detection of novel cREs essential for the regulation of cardiac development. Here we describe two methods to identify novel TFs.

2.2.2.1. New cardiac TF prediction by database mining. The restricted expression of a TF in a tissue or organ-specific manner could indicate that this TF has a role in defining the identity of that particular tissue or organ. Following this rationale, traditionally TFs were selected to be studied in the context of heart development if they displayed cardiac-specific expression (Schott et al., 1998), (Garg et al., 2003). Nowadays, identification of novel cardiac TFs based on the same principle can be facilitated by database mining. Indeed, several databases exist that contain comprehensive spatio-temporal transcriptional information from different stages of heart development in mouse, human and pluripotent stem cells (Liu et al., 2008), (Wu et al., 2009), (Papatheodorou et al., 2018) (Table 2). Identification of novel cardiac-restricted genes can be achieved by performing differential gene expression analysis, not only from wild-type or healthy heart samples, but also from diseased cardiac datasets such as coronary artery disease, congestive heart failure and cardiomyopathy (Wu et al., 2009).

Table 2
Databases storing information on non-coding elements.

Database Type			
Enhancer/promoter/histone modification	Transcription Factor Binding Site (TFBS)	ncRNA	Non-coding variants
Ensembl Regulation	Ensembl Regulation	RNAcentral	GWAS Central
FANTOM5	JASPAR	LncBase	NHGRI-EBI GWAS catalogue
VISTA Enhancer Browser	TRANSFAC	LNCipedia	GWASdb
TRANSFAC	HOCOMOCO	LncRNAWiki	ExAC
Expression Atlas	International Human Epigenome Consortium (IHEC)	miRBase	
ENCODE		miRTarBase	
NIH Roadmap Epigenomics Mapping Consortium		circNet	
International Human Epigenome Consortium (IHEC)		PolymiRTS	
ExAC		FANTOM5	
		Expression Atlas	

Table 3

Key transcription factors specific to heart development and CHD, and the corresponding model organism for which DNA binding site information is available in JASPAR.

Transcription Factors (human homologue)	Organism in JASPAR
ANKRD1	-
CITED2	-
ETS1	<i>Homo sapiens, Mus musculus</i>
FOG2/ZFPM2	-
FOXC1	<i>Homo sapiens</i>
FOXC2	<i>Homo sapiens</i>
FOXH1	<i>Homo sapiens</i>
GATA1	<i>Homo sapiens, Mus musculus, Arabidopsis thaliana</i>
GATA4	<i>Mus musculus</i>
GATA6	<i>Homo sapiens, Arabidopsis thaliana, Mus musculus</i>
HAND1	<i>Mus musculus</i>
HAND2	-
HOXA1	<i>Homo sapiens, Mus musculus</i>
IRX4	<i>Mus musculus</i>
ISL1	-
JAG1	-
MEF2C	<i>Homo sapiens</i>
MESP1	-
MESP2	-
MYOCD	-
NFATC1	<i>Meleagris gallopavo</i>
NKX2-5	<i>Mus musculus, Caenorhabditis elegans</i>
NKX2-6	<i>Mus musculus</i>
PITX2	<i>Mus musculus</i>
PPARG	<i>Homo sapiens, Mus musculus</i>
SALL4	-
TBX1	<i>Homo sapiens</i>
TBX2	<i>Homo sapiens</i>
TBX20	<i>Homo sapiens</i>
TBX5	<i>Homo sapiens</i>
TFAP2B	<i>Homo sapiens</i>
ZEB2	-
ZIC3	<i>Homo sapiens, Mus musculus</i>

For instance, Expression Atlas contains quality-controlled curated microarray and RNA-seq experiments from ArrayExpress or GEO (Kolesnikov et al., 2015). It allows baseline and differential analysis to identify gene products expressed in heart, and those that are differentially expressed in disease conditions. A meta-analysis study utilising 700 different mouse experiments downloaded from this resource predicted novel candidate genes with putative role in heart failure (Akavia and Benayahu, 2008). In summary, several databases are available to identify novel TFs by relying on their specificity of expression in the heart. However, this approach is limited as there is evidence that TFs that are broadly expressed can also play a specific role in heart (e.g. Elks TFs (Bouvet et al., 2015)). In addition, these databases may provide reliable *in silico* evidence of the role of new TFs in heart development, but this remains to be experimentally validated.

2.2.2.2. New cardiac TF prediction by motif discovery on targets of known TFs. Cardiac TFs bind to their target cREs in complexes, which include other transcriptional co-factors. Therefore, cREs contain multiple TF binding sites (TFBS) corresponding to the binding of the TFs in the complex. There are known TFs that are part of the regulatory complex, for example GATA4, TBX5, NKX2-5 and SRF have been shown to co-occupy the same cREs in HL1-cardiomyocytes (He et al., 2014). However, the regulatory complexes may contain other TFs that are not known to have a cardiac function. Thus, it is expected that the cREs targeted by these complexes will also contain the binding site of these novel TFs. With arduous experimental methods alone it is nearly impossible to identify these novel TFBS. Therefore, computational methods have been designed to predict these novel TFBS in the

genome. By using *in silico* methods, it is possible to identify these novel TFBS, and predict novel TFs involved in heart development and disease that are binding to these novel TFBSs. Here we present methods to predict new TFs involved in heart development from predicted novel TFBS in cardiac cREs.

in silico methods to predict cREs in the genome initially relied on searching for DNA motifs with specific nucleotide composition (e.g. TATA-box like sequences and CpG dinucleotide abundance), usually in proximity to the transcription start site (TSS). However, this lead to a large number of false positives due to the abundance of such sequences with these simple signatures in the genome (Fickett and Hatzigeorgiou, 1997). Phylogenetic footprinting based algorithms were then developed to identify putative functional cREs. They are based on the principle that functional regions of the genome accumulate mutations more slowly compared to regions with no function. Thus, comparison of non-coding sequences across genomes may reveal highly conserved segments indicating a role in transcriptional regulation (Lenhard et al., 2003). However, it is non-trivial to identify stretches of highly conserved non-coding regions, in particular, cardiac enhancers have been shown to display less sequence conservation than brain enhancers for example (Blow et al., 2010). More recent methods are based on, *de novo* motif discovery to predict novel TFBS without any prior knowledge of TFBS composition. (Lenhard et al., 2003), (Ben-Gal et al., 2005), (Sandve and Drablos, 2006), (Narlikar et al., 2010), (Wong, 2017), (Hestand et al., 2008), (Mathelier et al., 2015), (Wei and Yu, 2007), (Dang et al., 2018), (Zhou and Troyanskaya, 2015), (Wang et al., 2018). These algorithms detect short DNA sequences (motifs) that are significantly enriched in a set of cREs (for instance obtained from ChIP-seq datasets), in comparison to random genomic sequences. In a ChIP-seq experiment, it is expected that the binding site of the targeted TF will appear as the highest-ranked predicted motif. Any additional enriched motifs may correspond to the binding sites of novel co-factors of that TF. To determine which TFs bind to these newly identified motifs, databases holding information about known TF binding sites can be interrogated. JASPAR is an open source collection of TF binding sites presented as position weight matrices (PWMs), where the JASPAR core version is a non-redundant collection of experimentally validated mammalian TFBSs (Vlieghe et al., 2006), (Khan et al., 2018). TRANSFAC is a commercially available database which covers a collection of experimentally determined TFBSs stored as PWMs (Vlieghe et al., 2006). The *Homo sapiens* comprehensive model collection (HOCOMOCO) and UniPROBE are other popular PWM-based resource which respectively integrate curated TFBSs from various experiments for over 400 human TFs (Kulakovskiy et al., 2013, 2018) or homeobox families of TFs (Newburger and Bulyk, 2009). However, the majority of transcription factors encoded in our genomes still have an unknown DNA binding site, due to the lack of experimentally validated TFBS in different organisms and disease models (Wasserman and Sandelin, 2004). Of the 3000 TFs predicted in the human genome, only a third has been captured in the databases above.

An example of cofactor identification using experimentally derived TF-associated sequence information is a discovery from DamID experiments targeting the NKX2-5 cardiac TF. For mutant NKX2-5 protein lacking the DNA-binding domain, a novel motif was enriched in the associated sequences that was not detectable when using the wild-type protein. Using TFBS databases, this motif was identified as the binding site of the Elk1/4 transcription factors. Hence, these studies revealed the role of Elk1/4 as novel co-factor of Nkx2-5 (Bruneau et al., 2001, Raha et al., 2010). Similar studies, based on *de novo* motif prediction on enhancers of taste receptor genes, revealed cardiac TFs regulating genes usually expressed in the tongue, to also be expressed in the heart (Foster et al., 2015).

2.3. Detection of cREs using epigenetic marks

Histone modifications are well known post-translational

modifications which regulate expression of genes by strengthening or weakening the interaction between DNA and histones. They regulate the binding of TFs to target regions based on strength of packaging of the chromatin, with the help of chromatin-modifying enzymes.

Histone modifications include histone (mono-, di- or tri-) methylation, acetylation, phosphorylation, ADP-ribosylation, ubiquitination and sumoylation. Histone H3 methylation (H3K4me3 and H3K27me3), and acetylation (H3K27ac) have been widely explored with respect to regulating gene expression, where the former is involved in both transcriptional activation or repression, the latter is associated with transcriptional activation (Vallaster et al., 2012). Various histone modifications govern the expression of different cardiac genes during early heart development and disease, including CHD (Zhou et al., 2011a), (Wamstad et al., 2012), (Heintzman et al., 2009), (Heintzman et al., 2007). ChIP-seq is a widely used technique to determine the location of histone with specific modifications (or histone marks) in the genome (O'Geen et al., 2011). One of the commonly used methods to identify enhancers is to perform ChIP-seq using antibodies against H3K4me1, H3K4me2, H3K27ac and the histone acetyltransferase p300 (Zhou et al., 2011b). Comparing and overlapping the resulting peaks of ChIP-seq experiments using different antibodies, allows a more accurate detection of cREs.

Another class of epigenetic marks associated with transcriptional regulation is DNA methylation. It involves covalent transfer of a methyl group to the C-5 position of the cytosine ring of DNA by DNA methyltransferases (DNMTs), at CG dinucleotides in the genome, resulting in epigenetic silencing of transcription. The CpG-rich regions within promoters and enhancers, tend to have a differential pattern of DNA methylation when compared to non-regulatory regions of the genome (Zhang et al., 2015), (Jin et al., 2011). Whole-genome bisulfite sequencing is the standard technique to identify differentially methylated cREs. Treatment of DNA with bisulfite mediates the conversion of cytosine into uracil, which is then read as thymine in the sequencing downstream. However, the methylated cytosine-5 are resistant to this conversion and are thus retained as cytosine. Thus, differentially methylated regulatory regions are identified by comparing the bisulfite treated with the untreated DNA sample (Kurdyukov and Bullock, 2016).

These technologies have been widely used in the context of heart development and disease to identify relevant cREs. For instance, Gilsbach et al. provided a comprehensive resource of enhancers active in human cardiomyocytes in various developmental stages based on histone mark locations and differential CpG methylation during prenatal vs postnatal heart development. These enhancers regions were strongly associated with genes involved in cardiomyocyte development, highlighting their functional relevance in cardiogenesis. In addition, investigation of the activity of these enhancers in chronic heart failure revealed that dynamic changes to histone marks, including H3K27ac and H3K36me3, were concordant with disease-relevant gene expression, thereby supporting the role of enhancers in cardiac disease, implicating them as regulators of gene expression in cardiac development and disease (Gilsbach et al., 2018). Other genome-wide studies have expanded our knowledge of cardiac enhancers in adult and fetal human heart using this principle (Wamstad et al., 2014), (Blow et al., 2010), (van Duijvenboden et al., 2014), (May et al., 2011). A comprehensive study across more than 35 epigenetic data sets for pre- and post-natal mouse and human models predicted more than 80,000 putative enhancer regions (Dickel et al., 2016). May D et al. predicted ~6200 candidate enhancers enriched near genes involved in heart development and disorders including CHD, using the results from ChIP-seq of enhancer-associated proteins p300/CBP. The p300/CBP-bound regions were classified based on their vicinity to transcription start sites of heart-specific genes, including *KCNQ1*, *KCNE2*, *MYL2*, *ACTN2* and *TNNC1*, *GATA4*, *NKX2-5* and *TBX5* involved in CHD. Enhancer sequences located in the vicinity of embryonic heart developmental genes such as *GATA4*, *GATA6* and *MEF2C* were experimentally validated *in vivo* using transgenic mouse model enhancer assay (May et al., 2011).

Finally, there is evidence that *de novo* mutations in the genes contributing to histone modifications also contribute to CHD patients. Indeed mutations have been observed in genes such as *CHD7*, *KDM5A*, *KDM5B* and *RNF20*, which affect histone 3 lysine 4 (H3K4) methylation, and ubiquitination of H2BK120, required for H3K4 methylation (Zaidi et al., 2013), (Zhang and Liu, 2015). Another recent study demonstrated elevated catalytic activity of histone deacetylase (HDAC Class I, IIa, and IIb) in children with single ventricle heart disease of their right ventricle (Blakeslee et al., 2017). These few yet relevant studies further strengthen the evidence that histone modifications and associated regulatory pathways play a role in the pathogenesis of CHD.

2.4. Detecting cREs using chromatin accessibility and organisation

Chromatin state governs the accessibility of regulatory elements to TFs. Differential chromatin state (open or close) can indicate the presence and the activity of cREs in the DNA. Various high-throughput techniques that enable to directly profile the chromatin accessibility include DNase-seq, FAIRE-seq, ATAC-seq and MNase-seq (Table 1). One of the indirect ways of identifying TFBS is identification of loosely packed DNase I hypersensitive sites in the chromatin. These sites in the genome have been correlated to frequently contain regulatory elements and can be sequenced genome-wide (DNase-ChIP and DNase-seq) (Crawford et al., 2006), (Crawford et al., 2004), (Song and Crawford, 2010). Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq), is another method to identify open chromatin. It involves hyperactive Tn5 transposase which inserts adapter sequences in the accessible regions and is used to identify TF binding regions and nucleosome positions (Buenrostro et al., 2015). These accessible regions are captured as sequences which are then filtered for quality and adapter sequences, and are aligned to the reference genome. Peaks or collection of aligned reads in a genomic region indicate the region that is open, and therefore accessible for DNA-binding proteins such as TFs (Miskimen et al., 2017). The FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) sequencing method uses formaldehyde to crosslink nucleosome rich regions, and involves recovery of the non-nucleosome bound DNA which corresponds to regulatory components of the genome such as DNase hypersensitive sites, transcription start sites and active promoters (Giresi et al., 2007). Finally, MNase-seq is a method to identify nucleosome free regions of the genome, which involves micrococcal nuclease based digestion of nucleosome bound regions in the DNA, followed by sequencing (Rizzo et al., 2012). Of all these methods, ATAC-seq has been more often used to interrogate chromatin accessibility in the cardiac context and has provided a great wealth of information to complete the mapping of cardiac cREs, from bulk tissues or single cells (Yuan et al., 2018), (Jia et al., 2018).

The three-dimensional (3D) organisation of chromatin adds another layer of complexity in identifying cREs and their target genes. Distal enhancers have been shown to physically interact with gene promoters located thousands of base pairs apart. Thus, understanding the 3D organisation of chromatin helps to reveal additional REs by reducing the search space to interacting genomic regions. Chromatin conformation capture technologies (3C) and its derived methods (4C, 5C and HiC) are used to uncover three-dimensional loops of cREs resulting from the interaction between TFs bound to one cRE and their co-factors bound to another cRE. These methods involve chromatin fixation, followed by restriction enzyme-based cutting to produce DNA fragments with sticky ends. These sticky ends are re-ligated to capture long distance fragment interactions (Denker and de Laat, 2016), (de Wit and de Laat, 2012), (Tolhuis et al., 2002). HiC was used to map chromatin interactions in cardiac cells and tissues (Montefiori et al., 2018), (Schmitt et al., 2016) and has revealed the crucial role of chromatin remodeling in cardiac myocytes and hypertrophy (Rosa-Garrido et al., 2017). While to date there have not been any such studies in the specific context of human CHD, a 4C-seq study in mouse heart has revealed the interaction of a

cardiac enhancer located in *Scn10a* (encompassing *SCN10A* variant associated with slow conduction) with the promoter of *Scn5a* (van den Boogaard et al., 2014).

2.5. Available resources

Numerous online databases and web portals provide experimental datasets and computational analyses that can be applied to identification of cREs in cardiac systems (Table 2). Ensembl Regulation at the European Bioinformatics Institute (EBI) provides resources for studying gene expression and its regulation in human and mouse. Their database contains annotations for regulatory features such as promoters, enhancers, binding sites of CTCFs (ubiquitous chromatin structural proteins) and open chromatin regions and transcription start sites (TSSs), supported by ChIP-seq and DNase-seq (Zerbino et al., 2016). Ensembl provides the data mining tool BioMart to access and retrieve information from its databases, which can be accessed via web browser or programmatically (e.g. for R Bioconductor packages and other programming languages) through application programming interfaces (APIs) (Smedley et al., 2015). Data from VISTA Enhancer Browser containing TSS and enhancer data have also been implemented (Visel et al., 2007). It does a pairwise alignment of sequences with a sliding window across the sequences to generate the nucleotide identity plot (Wasserman and Sandelin, 2004). A comparison of VISTA-predicted heart-specific enhancers with those identified in human cardiomyocytes in a ChIP-seq study, revealed an overlap of 49% of enhancers. VISTA enhancers that were not identified in the cardiomyocytes were adjacent to genes with a role in artery morphogenesis and organisation of collagen. This overlap increased to 90% by narrowing the VISTA list to confirmed mouse embryonic heart-specific enhancers (Gilsbach et al., 2018). This study demonstrates the modest efficiency of cRE prediction, at the same time highlights the need for more experimentally validated datasets to improve its efficiency.

One of the breakthroughs in exploring the human epigenome has been the Encyclopedia of DNA Elements (ENCODE) project aimed at capturing genome wide functional elements (Table 2). This is a comprehensive catalogue of non-coding functional elements including TFBSs, CTCF-occupied sites, non-coding RNAs and also loci associated with a dozen of histone modifications. The project quantified that 80.4% of the genome is involved in at least one biochemical functional event in at least a single cell type, highlighting our limited understanding of the roles of non-coding parts of the genome (Consortium, 2012). The NIH Roadmap Epigenomics Mapping Consortium is an additional project aimed at mapping functional epigenomic elements from publicly accessible, tissue-specific sequencing data (such as ENCODE) (Romanoski et al., 2015). As per ENCODE Portal (2007-Present), out of 340 heart-related datasets from human and mouse in the data matrix, 258 from ENCODE and 82 from NIH Roadmap are currently available. Another major effort undertaken by the International Human Epigenome Consortium (IHEC) is the generation of more than 1000 epigenomes for normal vs diseased human cell types. In addition, ENCODE and NIH Roadmap data are accessible via the IHEC portal (Bae, 2013), (Albrecht et al., 2016). These resources integrate genome-wide experimental data to improve the current knowledge of non-coding elements including cREs. However, it is important to keep in mind that these databases annotate data over an entire cell population which might be non-homogenous, and may result in a biased genetic or physiological representation (A user's guide to the encyclopedia, 2011). Furthermore, developing improved user-friendly analyses and visualization tools will facilitate the wider use of these comprehensive resources.

2.6. Functional validation of cREs

Application of the above discussed techniques and data resources has uncovered several putative cREs which raises the need for

functional validation to confirm which of these are truly functional in cells or tissues. Classical reporter assays such as fluorescence *in situ* hybridization (FISH) and luciferase assays have been used to study the enhancer activity. In luciferase assays, the enhancer is introduced into a reporter construct with a minimal promoter to activate the transcription of the reporter gene (luciferase), which can be quantified based on light signal capture. Additionally, transgenic reporter assays have been used to test cell-specific enhancers *in vivo* (Kwon, 2015). However, they can test only one individual candidate at a time making it arduous to validate multiple candidates, spurring the development of more high-throughput techniques such as massively parallel reporter assays (MPRA) for enhancer testing. Self-Transcribing Active Regulatory Region sequencing (STARR-seq) is one such technique where candidate enhancers are cloned downstream of a core promoter and within the 3' UTR of a reporter gene. These candidates will be transcribed as a part of the resulting reporter transcript when active and can be deep sequenced (Muerdter et al., 2015). This technique has been used to generate a *Drosophila* genome-wide quantified enhancer map (Arnold et al., 2013). However, STARR-seq might not be a preferred technique in mammalian genomes due to their large size. Functional Identification of Regulatory Elements Within Accessible Chromatin (FIREWACH) another high-throughput functional assay has been used to evaluate over 80,000 DNA fragments simultaneously in mammalian cells (Babbitt et al., 2015). This technique is able to handle large genomes by focusing the analyses to the most relevant 2% of the accessible chromatin within the cell. These MPRA techniques have yet to be implemented in cardiac cells to identify heart-relevant cREs on a genomic scale. Thus, despite all these significant milestones in annotating the non-coding part of the genome, key non-coding elements that are functionally proven to be implicated in CHD etiology still remain unknown.

3. Role of non-coding RNAs in cardiac gene regulation

3.1. Definition

Non-coding RNAs are functional RNAs that are transcribed yet not translated into protein, and are involved in the regulation of gene-expression at both transcriptional and post-transcriptional level. Non-coding RNAs (ncRNAs) can be broadly categorised into three groups based on their size, i.e. small ncRNAs (~16–31 nucleotides), medium size ncRNA (20–200 nucleotides) and long non-coding RNAs (lncRNA > 200 nucleotides) (Esteller, 2011).

3.2. miRNAs

microRNAs (miRNAs) are the best studied class of small ncRNAs that have been shown to be involved in various cardiovascular disorders (Papageorgiou et al., 2012), (Schulte and Zeller, 2015), (Wang et al., 2016a) and specifically in CHD (Smith et al., 2015), (Xing et al., 2013), (Xie et al., 2016). miRNAs generally downregulate their target mRNA by inhibiting translation or degrading it. A single miRNA commonly affects multiple target genes, therefore changes in expression of a miRNA can perturb an entire gene network resulting in complex pathological states. miRNAs such as *miR-1* and *miR-133a/b* have been identified for their roles in embryonic cardiogenesis (Bostjancic et al., 2015), (Kwon et al., 2005). These small RNAs have presumably multiple effects in development as well as in disease progression (Sayed and Abdellatif, 2011). For instance, high expression of *miR-126* has been observed in zebrafish heart and vasculature. Deletion of *miR-126* resulted in defective neoangiogenesis in mice models when subjected to myocardial infarction (Wang et al., 2008). *miR-1-2* deletion has been associated with the development of ventricular septum defects and increased cardiomyocyte proliferation (Zhao et al., 2007). Also, SNPs have been identified in the 3' untranslated region (3'UTR) of CHD-relevant genes such as *GATA4* which alter miRNA binding sites in CHD patients, indicating their role in the disease development (Sabina et al.,

2013). PIWI-interacting RNA (piRNA) and transfer RNA-derived small RNAs (tsRNA) are two recently discovered gene regulatory small ncRNA classes (Ross et al., 2014), (Li et al., 2018). piRNAs are involved in regulating expression of retrotransposons, have been shown to be differentially expressed in patients with myocardial infarction. However, any potential role in CHD for either piRNAs or tsRNAs still remains unexplored (Rajan et al., 2016).

3.3. lncRNAs

lncRNAs are a class of regulatory RNAs that individually have been implicated in varied roles such as chromatin modification (Gupta et al., 2010), (Rinn et al., 2007), transcriptional (Huarte et al., 2010) and post-transcriptional gene expression regulation and RNA transport (Geisler and Collier, 2013). lncRNAs function by interacting with DNA, RNA and regulatory proteins. Similar to protein-coding genes, lncRNAs could be contributing to tissue-specific response via their role in epigenetic regulation. They appear to be necessary for proper targeting of histone modifying complexes and play a role in DNA methylation and recruitment of regulatory proteins to the cREs of its target genes (Peschansky and Wahlestedt, 2014). Wang et al. referred to this class of epigenetic regulatory lncRNA as Epi-lncRNAs (Wang and Wang, 2015). In recent years, several studies have emerged highlighting the role of lncRNAs in cardiac development (Schonrock et al., 2012). Some examples are Myosin Heavy Chain Associated RNA Transcripts (MyHEART or *Mhrt*), Cardiac hypertrophy-associated transcript (*Chast*) (Viereck et al., 2016), Cardiac-hypertrophy-associated epigenetic regulator (*Chaer*) (Wang et al., 2016b) and Braveheart (*Bvht*) (Schmitz et al., 2016), (Klattenhoff et al., 2013). *Mhrts* have been identified as cardiac-specific lncRNA, with low level of transcripts in fetal hearts, and increased abundance as the heart developed with increased *Myh7* expression; and have been shown to protect heart from cardiac hypertrophy in mice (Han et al., 2014). *Bvht* is involved in activation of cardiac gene regulatory network by regulating the expression of *Mesp1* gene, and also interacts with protein SUZ12 involved in cardiomyocyte differentiation (Klattenhoff et al., 2013). Two other lncRNAs, *Chast* and *Chaer*, have recently been shown to be essential in the development of cardiac hypertrophy (Wang et al., 2016b), (Viereck et al., 2016). The latter directly interacts with chromatin remodeling factor PCR2 to interfere with normal gene expression regulation.

3.4. circRNAs

Another class of non-coding RNAs are circRNAs which are studied for their role in cardiovascular diseases (Fan et al., 2017). circRNAs are stable, covalently closed circular RNA molecules, formed initially by pairing of complementary motifs present at the two ends (which often have been observed to be Alu repeats and are expressed widely in human cells) after which the ends of the loop are covalently ligated through a “back-splicing” event. They regulate miRNA target gene expression by competitively binding to miRNAs, reducing their ability to down-regulate the expression of their target genes. Thus, they have been termed microRNA sponges (Fan et al., 2017). Recent studies in various cardiovascular diseases have demonstrated the regulatory role of circRNA. Wang et al. demonstrated the role of Heart-related circRNA (HRCR), which can sequester *miR-223*, (targeting the *ARC* gene), as a positive regulator of cardiac hypertrophy (Wang et al., 2016c). Similarly, other circRNAs have been identified in myocardial infarction and cardiac senescence (Geng et al., 2016), (Du et al., 2017). Since circRNAs represent another level of gene regulation and have been associated with various heart disorders, it would be worth investigating whether some circRNAs play a role in CHD.

3.5. Techniques to explore ncRNAs

ncRNAs are often profiled in whole-transcriptome studies by

applying high-throughput RNA sequencing technologies to total RNA preparations after targeted depletion of rRNA. Small RNAs in the range of 19–33 bp are isolated with specialized protocols and profiled using high-throughput sequencing, most often with the aim of detecting miRNAs (Hafner et al., 2008), (Lu et al., 2007). However, the results of RNA-seq experiments have to be interpreted carefully while identifying ncRNAs as they are short sequences (< 100 bp in length) and more often repetitive elements, which might map to multiple regions in the genome increasing the false positives (Uchida, 2017), (Weirick et al., 2016). Ultra deep sequencing will be required to detect lowly expressed ncRNAs in general.

lncRNA are harder to isolate since they vary both in length and expression level. The lncRNAs are found preferentially proximal (within ~10 kb) to a protein-coding locus and are often expressed in a tissue-specific manner (Cabili et al., 2011). Heart-specific localisation of lncRNA might also be indicative of its role in cardiac development and disease. Generally, to view the localisation of ncRNAs, *in situ* hybridization assay has been commonly used (Soares et al., 2018). For instance, myocardium-specific expression of lncRNA *Mhrt* was observed using *in situ* hybridization in mice indicating its role in heart development (Han et al., 2014). In order to identify the targets of lncRNAs, various hybridization techniques such as Capture Hybridization Analysis of RNA Targets (CHART) are applied to detect genome-wide binding locations of lncRNAs. The RNA bound chromatin across the genome is enriched using biotinylated complementary antisense oligonucleotides (22–28 nucleotides) that bind to the accessible target RNA. RNA-chromatin complexes are then immobilised on beads and eluted using RNase H over multiple washes. The targets of lncRNA in the genome are then determined by sequencing the associated DNA and mapping it genome-wide (Vance, 2017), (Simon, 2013). These hybridization-based methods are used to determine genome-wide associations of known lncRNA.

3.6. Available resources

RNA-related databases such as RNACentral store datasets of different classes of RNAs, including non-coding RNA (Petrov et al., 2017). The RNACentral Consortium is formed by 40 expert databases, 23 of which have already been imported into RNACentral. It has a wide variety of lncRNA sources (LncBase (Paraskevopoulou et al., 2016), LNCipedia (Volders et al., 2013), LncRNAWiki (Ma et al., 2015), lncRNADB (Quek et al., 2015), GENCODE (Harrow et al., 2012)), miRNAs and their targets (including miRBase (Griffiths-Jones et al., 2006), miRTarBase (Chou et al., 2018)) and snoRNAs (NONCODE (Liu et al., 2005)) datasets. For circRNAs, circNet database contains tissue-specific information on circRNAs and their miRNA partners (Liu et al., 2016). PolymiRTS (Polymorphism in microRNAs and their Target Sites) is a functional analysis tool to explore miRNAs and their target sites in disease contexts, including coronary heart disease and heart failure (Bhattacharya et al., 2014). Finally the FANTOM consortium has built an atlas of various lncRNAs across more than 1800 human cell- and tissue-types including cardiac, based on Cap analysis gene expression (CAGE) data integrated with a collection of transcript models (Iyer et al., 2015).

Although there are increasingly large collections of ncRNA datasets becoming available for various disease and physiological states from human tissues, more specific heart or CHD examples are very few. Because ncRNAs operate at various regulatory levels (chromatin tertiary structure, splicing, transcription, RNA stability and translation) and mechanistic levels (interacting with DNA, protein and/or other RNAs), more integrative approach to find associations between genomic datasets from different experiment types are required to find new ncRNA players in heart and CHD-specific regulatory machinery.

4. Alterations to be expected in non-coding regulatory elements

4.1. Point mutations

4.1.1. Definition

A point mutation is a substitution of a single nucleotide in the DNA sequence with another nucleotide which may or may not have a pathogenic effect. Several point mutations have been observed in both syndromic and non-syndromic CHD. Here, we focus on SNPs that have been identified in non-coding regions.

Genome-wide association studies (GWAS) have been conducted in CHD patients to pinpoint CHD associated loci including point mutations (Cordell et al., 2013a), (Blue et al., 2017). GWAS aims to identify genetic variants that are associated with a phenotype (typically, disease) in large population cohorts, most commonly using single nucleotide polymorphism arrays (SNP arrays). These arrays allow to estimate the frequency of known variants or loci across a diseased genome, in comparison with their frequency of occurrence in healthy controls, in order to identify variants with a statistically significant association with the disease (Visscher et al., 2012). These studies have been widely used to identify SNPs associated with CHD risk (Hu et al., 2013), (Cordell et al., 2013b), (Zhao et al., 2014), (Lin et al., 2015).

Strikingly, ~88% of SNPs identified in GWAS lie in non-coding regions (Edwards et al., 2013), but they are often not pursued for further investigation. This could be on account of linkage disequilibrium observed in families which leads to indirect association of variants with the disease, making it difficult to narrow down the causal variant (Bush and Moore, 2012). Often, to reduce the number of variants under consideration, coding variants have been prioritized because it is easier to predict their immediate biological effect (e.g. producing a truncated protein). Furthermore, even statistically significant association with disease does not resolve the underlying biological mechanism of the variant's pathogenesis, which may be important in the long term for developing new potential therapies.

4.1.2. Relevant studies and techniques

Cordell and colleagues have identified the SNP rs870142 at a non-coding locus on chromosome 4p16, which might be involved in transcriptional regulation, to be associated with an *ostium secundum* atrial septal defect (ASD) in individuals of European ancestry (Cordell et al., 2013b). The same locus association was replicated in Han Chinese population with ASD by Zhao et al. (2014). In addition, Lin and colleagues identified SNPs upstream of genes including *TBX3* and *TBX5* (SNP rs2433752), *SMARCA2* (SNP rs7863990), *EDNRA* (SNP rs1400558) as CHD susceptibility loci in the Chinese population. These genes are known to have a role in various stages of heart development (Lin et al., 2015). These studies reinforce the need to understand the role of such loci in regulating the transcription of various genes to eventually predict CHD risks in the clinic.

The roles of point mutations in non-coding regions in syndromic disease are still unexplored. This can be attributed to the fact that a large set of alterations known in syndromic CHD constitute chromosomal aberrations where a large proportion of the chromosome is either translocated or inserted/deleted. In this scenario, it becomes very difficult to identify single point mutations (both in intronic and genic regions) that could underlie the pathology. However, genes involved in syndromic form of the disease such as *TBX5* (involved in Holt-Oram syndrome) have been observed to interact with various regulatory elements to control the expression of several other genes (Waldron et al., 2016). A non-coding pathogenic single nucleotide mutation may affect gene regulation by altering TF binding. A study identified a non-coding single nucleotide mutation present ~90 kb downstream of the *TBX5* gene in one of the enhancer elements in patients with non-syndromic ventricular septal defect. This variant was demonstrated to alter the capacity of the enhancer to drive cardiac-specific gene expression in transgenic mice and zebrafish during heart development (Smemo et al.,

2012).

Whole genome sequencing (WGS) unlike exome sequencing which captures only coding regions (around 2% of the entire genomic information), covers both coding and non-coding regions of the genome. Furthermore, WGS allows more reliable detection of copy number variants (CNVs) with a good read depth. The falling cost of WGS and improving turnaround time of data analysis will most likely drive it to overtake WES in the near future. However, the challenge to narrow down to the causal variants and avoid a bias in the filtering process based on what is already known in the genome remains a limitation in both approaches (Dorn et al., 2014). Indeed, similarly to cancer genomes (Gan et al., 2018), a large number of single nucleotide variations (SNVs) are likely to be observed in the non-coding regions of healthy versus CHD genomes. Features such as experimental observations of local epigenomic states, conservation of motifs, their vicinity to heart-relevant genes and integration of gene expression data can be used to narrow down the list of SNVs relevant for disease (May et al., 2011). However, prediction of these CHD-relevant SNPs still poses a challenge in terms of limited CHD-specific data sets, scarce conservation of these regulatory elements across animal models and humans, and associative nature of GWAS studies.

4.1.3. Available resources

Several GWAS databases are available which aim at integrating genotype-to-phenotype associations in various diseases including CHD (Brookes and Robinson, 2015). These include NHGRI-EBI GWAS catalogue and GWASdb (Li et al., 2012), (MacArthur et al., 2017), (Welter et al., 2014). GWAS Central (the former Human Genome Variation database of Genotype-to-Phenotype information) collates data from public portals of both small and large population studies and includes CHD datasets such as Tetralogy of Fallot (Beck et al., 2014). Moreover, the CHARGE consortium (Cohorts for Heart and Aging Research in Genomic Epidemiology) is one of the biggest large population GWAS studies in cardiovascular disease for which data is available at the database of Genotypes and Phenotypes (dbGaP), NIH. DNA banks of CHD patients like CONCOR (Netherlands) emphasise on genetic studies to identify markers to predict CHD risks and survival (van der Velde et al., 2005). Potential limitations of these datasets are biased participants' selection criteria and genotyping errors.

ClinVar integrates point mutations from dbSNP, archives variants including those in splice sites, UTRs, near gene regions and their clinical significance (Landrum et al., 2018). The Exome Aggregation Consortium (ExAC), though majorly known for exome datasets from large sequencing studies, covers various genes with CHD consequence and their variants, as well as intronic and promoter sequences [<http://exac.broadinstitute.org>]. Another widely used database is 1000 Genomes which catalogues the largest human variants and genotype data which is currently maintained by EMBL-EBI and provides access to the variant calling files [<http://www.internationalgenome.org/about>]. The Human Gene Mutation Database (HGMD) is another comprehensive repository of clinically identified variations including those in regulatory regions for inherited diseases including heart disorders [<http://www.hgmd.cf.ac.uk/ac/index.php>]. The Cardiovascular Research Grid (CVRG) Project is a collective effort by NIH and other institutions, aimed at integrating multi-scale data including genomics, proteomics, electrocardiographic, structural and functional cardiovascular clinical data, and is of great use for GWAS studies [<http://cvrgrid.org/>].

4.2. Structural abnormalities

4.2.1. Definition

Structural variation refers to a variety of microscopic and sub-microscopic alterations which encompass genomic DNA with size larger than 1 kb: quantitative (copy number variants (CNVs) comprising deletions, insertions and duplications) and/or positional (translocations) or orientational (inversions) [reviewed in (Scherer et al., 2007)]. These

variations initially were limited to cytogenetic detection, until genome sequencing technologies were introduced which narrowed this detection method-based distinction (Freeman et al., 2006), (Sharp et al., 2005). In this section we discuss the structural variants that are commonly observed in syndromic CHD, techniques used to detect them, and resources storing such data. In addition, we highlight the limited exploration of non-coding regions in these alterations, which warrant further investigation.

4.2.2. Relevant studies and techniques

A variety of structural aberrations have been associated with CHD. However, the studies reporting these aberrations have been focusing on the characterisation of the coding regions within the structural variant. A large population-based study (4430) on infants with CHD, demonstrated the occurrence of chromosomal abnormalities in 12.3% of the population, with the common ones being trisomy 21, trisomy 18, 22q11.2 deletion, and trisomy 13 (Hartman et al., 2011). These structural variations have been frequently associated with syndromic CHD. Williams syndrome is another disorder caused by a microdeletion of chromosome 7q11.23. The deletion affects 26–28 genes spanning 1.5–1.8 Mb and results in multiple defects including supravalvular aortic stenosis (Adams and Schmaier, 2012). Another study reported *de novo* and germline aberrations to have a role in CHD. The aberrations are of median size varying from 2.74 Mb to 6.21 Mb with around 17–20 genes involved. Some of these aberrations span regions genes with well-known cardiac development role, such as *NKX2-5*, while others affect novel regions, whose significance in CHD is yet to be explored (Thienpont et al., 2007). These structural variants, including CNVs, apart from altering gene dosage, can also modify the topological interaction of regulatory elements lying within chromatin loops (Ibn-Salem et al., 2014). 3C and HiC techniques have captured topological associating domains (TADs) in chromatin structure where regulatory elements, such as promoters, enhancers and insulators interact. The boundaries of these domains have been found to play a role in inhibiting these interactions and are called TAD boundaries. CNVs especially large deletions leading to disruption of TAD boundaries have been associated with congenital diseases (Ibn-Salem et al., 2014), (Kyrchanova and Georgiev, 2014). However, such studies need to be applied to CHD.

Cytogenetic analyses including fluorescence *in situ* hybridization (FISH) and metaphase karyotype analysis, have a resolution of 5 Mb and higher, and are commonly used in the clinic to detect chromosomal aberrations in CHD syndromes such as Williams-Beuren, DiGeorge, and Alagille (Chaix et al., 2016). Probes in FISH assay can be designed to detect alterations in a known DNA sequence or even RNA including lncRNAs and miRNAs (Zhang et al., 2012), (Lee et al., 2011), (Soares et al., 2018). Multiplex ligation-dependent probe amplification (MLPA) has been used in detecting chromosomal deletions and duplications in CHD patients including Tetralogy of Fallot and chromosomal 8p23 deletion syndrome (Zhu et al., 2016), (Guida et al., 2010) as well as the detection of presence or absence of an anomaly in a known genic region and/or at intron-exon boundaries (Yoshida et al., 2016). MLPA is most useful in cases of CHDs that are genetically well-characterised and hence provide an opportunity for better clinical management of the disease. However, the drawback of these techniques lies in the fact that they are targeted probing, where aberrations in any novel region cannot be detected (Stuppia et al., 2012).

A more commonly used high-throughput detection method is array-based comparative genomic hybridization (aCGH), also known as chromosomal microarray analysis (CMA) (Haeri et al., 2016). This method detects CNVs genome-wide albeit with a resolution lower than that of FISH. However, aCGH has been used to detect other anomalies including deletions, duplications and translocations. Thienpont et al. evaluated the diagnostic value of this technique in detecting such aberrations (Thienpont et al., 2007). Defects including atrial septal defect, ventricular septal defect, pulmonary valve stenosis, aortic

regurgitation, sinus venosus atrial septal defect, have been successfully detected with submicroscopic aberrations using aCHG in fetuses with abnormal cardiac ultrasound results (Yan et al., 2014). The ability to identify structural variants by high-throughput sequencing using methodologies such as CNV-seq (Zhu et al., 2016) has led to development of bioinformatics tools to call these variants with high specificity and sensitivity (Tattini et al., 2015). DELLY, is one of the earliest tools developed to call various structural variants including deletions, tandem duplications and balanced rearrangements like inversions or reciprocal translocations (Rausch et al., 2012). Other tools such as Lumpy and Soft search are now available for calling structural variants, while others are also integrating these to design pipelines with a combinatorial parameter approach (Tattini et al., 2015), (Layner et al., 2014), (Hart et al., 2013). For instance, genome STRiP (Genome STRucture In Populations) is a collection of tools for identifying and genotyping structural variants (Handsaker et al., 2015).

4.2.3. Available resources

Curated databases have collated structural aberrations in numerous diseases including CHD. The European Bioinformatics Institute (EBI) and the National Centre for Biotechnology Information (NCBI) maintain public archives dbVAR and DGVA for structural variants in the genome (Lappalainen et al., 2013). dbVAR (also available under ClinVar database) contains variants majoritarily greater than 50 bp in length and offers freely available raw data (Harrison et al., 2016). DECIPHER (Database of Chromosomal Imbalance and Phenotype in Humans Using Ensembl Resources) allows access to analysis and identification of rare pathogenic aberrations in patient whole-genome and exome data and uses Ensembl resources. There are datasets of more than 1700 patients with cardiovascular diseases including phenotypical details of fetal patients (Firth et al., 2009).

Given the large number of resources and technologies available to detect CNVs, clinical interpretation and identification of syndromic and non-syndromic CHD still need clear genetic marker boundaries. Breckpot and colleagues proposed at step-by-step interpretation of these CNVs based on various features including size of the aberration and insertions. They also reported that the frequency of causal CNVs in non-syndromic CHD is lower than in syndromic CNVs (Breckpot et al., 2011).

5. Conclusion

In this article, we have discussed the role of non-coding regions in the etiology of CHD and available resources to help predict and interpret their function. We reviewed examples where alterations in various non-coding elements including enhancers, promoters and ncRNAs are associated to CHD development. However, unlike other diseases such as cancer, the exploration of non-coding regions in CHD still remains limited.

Considering the number of putative enhancers identified at various stages of heart development, there is a need to collate and annotate them to understand their role in pathogenesis of CHD. In this review, we have discussed the techniques involved in the identification of these regions in the genome including high-throughput sequencing, and we have described resources with non-coding variant data sets. These resources collate multi-scale genetic information of different stages of heart development in humans as well as other model organisms, and across various CHD populations. However, due to the weakly conserved nature of these sequences across various animal models, the identification of functional enhancers still poses a challenge. Having a comprehensive set of coding and non-coding genome markers will assist to improve the efficiency of CHD diagnosis, while further investigations into the mechanisms of pathogenesis of these variants will boost our understanding of the underlying causes of CHD.

Funding

This work was supported by Rotary Global Grant, Rotary Foundation International [GG1637399] and Australian Genome Research Facility (AGRF) to GC; NHMRC/National Heart Foundation Career Development Fellowship [1049980] to MR. The Australian Regenerative Medicine Institute is supported by grants from the State Government of Victoria and the Australian Government.

Acknowledgements

We would like to thank Mr. Markus Tondl (Centre for Cancer Biology, UniSA), Dr. Stuart Archer (Monash Bioinformatics Platform), and Dr. Ekaterina Salimova (ARMI) for their critical input on the manuscript.

References

- A user's guide to the encyclopedia of DNA elements (ENCODE). *PLoS Biol.* 9, e1001046. <https://doi.org/10.1371/journal.pbio.1001046>.
- Adams, G.N., Schmaier, A.H., 2012. The Williams-Beuren Syndrome—a window into genetic variants leading to the development of cardiovascular disease. *PLoS Genet.* 8, e1002479. <https://doi.org/10.1371/journal.pgen.1002479>.
- Akavia, U.D., Benayahu, D., 2008. Meta-analysis and profiling of cardiac expression modules. *Physiol. Genom.* 35, 305–315. <https://doi.org/10.1152/physiolgenomics.90248.2008>.
- Albrecht, F., List, M., Bock, C., Lengauer, T., 2016. DeepBlue epigenomic data server: programmatic data retrieval and analysis of epigenome region sets. *Nucleic Acids Res.* 44, W581–W586. <https://doi.org/10.1093/nar/gkw211>.
- Arnold, C.D., et al., 2013. Genome-wide quantitative enhancer activity maps identified by STARR-seq. *Science (New York, N.Y.)* 339, 1074–1077. <https://doi.org/10.1126/science.1232542>.
- Babbitt, C.C., Markstein, M., Gray, J.M., 2015. Recent advances in functional assays of transcriptional enhancers. *Genomics* 106, 137–139. <https://doi.org/10.1016/j.ygeno.2015.06.002>.
- Bae, J.B., 2013. Perspectives of international human epigenome consortium. *Genom. Inform.* 11, 7–14. <https://doi.org/10.5808/gi.2013.11.1.7>.
- Barth, J.L., et al., 2010. Jarid2 is Among a Set of Genes Differentially Regulated by Nkx2.5 during Outflow Tract Morphogenesis, vol. 239. *Developmental dynamics: an official publication of the American Association of Anatomists*, pp. 2024–2033. <https://doi.org/10.1002/dvdy.22341>.
- Baumgartner, H., et al., 2010. ESC Guidelines for the management of grown-up congenital heart disease (new version 2010). *Eur. Heart J.* 31, 2915–2957. <https://doi.org/10.1093/eurheartj/ehq249>.
- Beck, T., Hastings, R.K., Gollapudi, S., Free, R.C., Brookes, A.J., 2014. GWAS Central: a comprehensive resource for the comparison and interrogation of genome-wide association studies. *Eur. J. Hum. Genet. EJHG* 22, 949–952. <https://doi.org/10.1038/ejhg.2013.274>.
- Ben-Gal, I., et al., 2005. Identification of transcription factor binding sites with variable-order Bayesian networks. *Bioinformatics (Oxf. Engl.)* 21, 2657–2666. <https://doi.org/10.1093/bioinformatics/bti410>.
- Bhattacharya, A., Ziebarth, J.D., Cui, Y., 2014. PolymiRTS Database 3.0: linking polymorphisms in microRNAs and their target sites with human diseases and biological pathways. *Nucleic Acids Res.* 42, D86–D91. <https://doi.org/10.1093/nar/gkt1028>.
- Blakeslee, W.W., et al., 2017. Histone deacetylase adaptation in single ventricle heart disease and a young animal model of right ventricular hypertrophy. *Pediatr. Res.* 82, 642–649. <https://doi.org/10.1038/pr.2017.126>.
- Blow, M.J., et al., 2010. ChIP-Seq identification of weakly conserved heart enhancers. *Nat. Genet.* 42, 806–810. <https://doi.org/10.1038/ng.650>.
- Blue, G.M., Kirk, E.P., Sholler, G.F., Harvey, R.P., Winlaw, D.S., 2012. Congenital heart disease: current knowledge about causes and inheritance. *Med. J. Aust.* 197, 155–159.
- Blue, G.M., et al., 2014. Targeted next-generation sequencing identifies pathogenic variants in familial congenital heart disease. *J. Am. Coll. Cardiol.* 64, 2498–2506. <https://doi.org/10.1016/j.jacc.2014.09.048>.
- Blue, G.M., et al., 2017. Advances in the genetics of congenital heart disease: a clinician's guide. *J. Am. Coll. Cardiol.* 69, 859–870. <https://doi.org/10.1016/j.jacc.2016.11.060>.
- Bostjancic, E., Jerse, M., Glavac, D., Zidar, N., 2015. miR-1, miR-133a/b, and miR-208a in human fetal hearts correlate to the apoptotic and proliferation markers. *Exp. Biol. Med. (Maywood, N.J.)* 240, 211–219. <https://doi.org/10.1177/1535370214546268>.
- Bouveret, R., et al., 2015. NKX2-5 mutations causative for congenital heart disease retain functionality and are directed to hundreds of targets. *eLife* 4. <https://doi.org/10.7554/eLife.06942>.
- Breckpot, J., et al., 2011. Challenges of interpreting copy number variation in syndromic and non-syndromic congenital heart defects. *Cytogenet. Genome Res.* 135, 251–259. <https://doi.org/10.1159/000331272>.
- Brivanlou, A.H., Damell Jr., J.E., 2002. Signal transduction and the control of gene expression. *Science (New York, N.Y.)* 295, 813–818. <https://doi.org/10.1126/science.1066355>.
- Brookes, A.J., Robinson, P.N., 2015. Human genotype-phenotype databases: aims, challenges and opportunities. *Nat. Rev. Genet.* 16, 702–715. <https://doi.org/10.1038/nrg3932>.
- Bruneau, B.G., et al., 2001. A murine model of Holt-Oram syndrome defines roles of the T-box transcription factor Tbx5 in cardiogenesis and disease. *Cell* 106, 709–721.
- Buenrostro, J.D., Wu, B., Chang, H.Y., Greenleaf, W.J., 2015. ATAC-seq: a method for assaying chromatin accessibility genome-wide. *Curr. Protoc. Mol. Biol.* 109. <https://doi.org/10.1002/0471142727.mb2129s109>. 21.29.21–29.
- Bush, W.S., Moore, J.H., 2012. Chapter 11: genome-wide association studies. *PLoS Comput. Biol.* 8, e1002822. <https://doi.org/10.1371/journal.pcbi.1002822>.
- Cabili, M.N., et al., 2011. Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. *Genes Dev.* 25, 1915–1927. <https://doi.org/10.1101/gad.17446611>.
- Chaix, M.A., Andelfinger, G., Khairy, P., 2016. Genetic testing in congenital heart disease: a clinical approach. *World J. Cardiol.* 8, 180–191. <https://doi.org/10.4330/wjc.v8.i2.180>.
- Chou, C.H., et al., 2018. miRTarBase update 2018: a resource for experimentally validated microRNA-target interactions. *Nucleic Acids Res.* 46, D296–d302. <https://doi.org/10.1093/nar/gkx1067>.
- Consortium, E.P., 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74. <https://doi.org/10.1038/nature11247>.
- Cordell, H.J., et al., 2013a. Genome-wide association study identifies loci on 12q24 and 13q32 associated with tetralogy of Fallot. *Hum. Mol. Genet.* 22, 1473–1481. <https://doi.org/10.1093/hmg/dd552>.
- Cordell, H.J., et al., 2013b. 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. <https://doi.org/10.1038/ng.2637>.
- Cowan, J.R., Ware, S.M., 2015. Genetics and genetic testing in congenital heart disease. *Clin. Perinatol.* 42, 373–393. <https://doi.org/10.1016/j.clp.2015.02.009>. ix.
- Crawford, G.E., et al., 2004. Identifying gene regulatory elements by genome-wide recovery of DNase hypersensitive sites. *Proc. Natl. Acad. Sci. U. S. A.* 101, 992–997. <https://doi.org/10.1073/pnas.0307540100>.
- Crawford, G.E., et al., 2006. DNase-chip: a high-resolution method to identify DNase I hypersensitive sites using tiled microarrays. *Nat. Methods* 3, 503–509. <https://doi.org/10.1038/nmeth888>.
- Dang, L.T., et al., 2018. TrawlerWeb: an online de novo motif discovery tool for next-generation sequencing datasets. *BMC Genomics* 19, 238. <https://doi.org/10.1186/s12864-018-4630-0>.
- de Wit, E., de Laat, W., 2012. A decade of 3C technologies: insights into nuclear organization. *Genes Dev.* 26, 11–24. <https://doi.org/10.1101/gad.179804.111>.
- Denker, A., de Laat, W., 2016. The second decade of 3C technologies: detailed insights into nuclear organization. *Genes Dev.* 30, 1357–1382. <https://doi.org/10.1101/gad.281964.116>.
- Dickel, D.E., et al., 2016. Genome-wide compendium and functional assessment of in vivo heart enhancers. *Nat. Commun.* 7, 12923. <https://doi.org/10.1038/ncomms12923>.
- Dorn, C., Grunert, M., Sperling, S.R., 2014. Application of high-throughput sequencing for studying genomic variations in congenital heart disease. *Brief. Funct. Genom.* 13, 51–65. <https://doi.org/10.1093/bfgp/elt040>.
- Du, W.W., et al., 2017. Foxo3 circular RNA promotes cardiac senescence by modulating multiple factors associated with stress and senescence responses. *Eur. Heart J.* 38, 1402–1412. <https://doi.org/10.1093/eurheartj/ehw001>.
- Edwards, S.L., Beesley, J., French, J.D., Dunning, A.M., 2013. Beyond GWAS: illuminating the dark road from association to function. *Am. J. Hum. Genet.* 93, 779–797. <https://doi.org/10.1016/j.ajhg.2013.10.012>.
- Esteller, M., 2011. Non-coding RNAs in human disease. *Nat. Rev. Genet.* 12, 861–874. <https://doi.org/10.1038/nrg3074>.
- Fan, X., et al., 2017. Circular RNAs in cardiovascular disease: an overview. *BioMed Res. Int.* 2017, 5135781. <https://doi.org/10.1155/2017/5135781>.
- Fickett, J.W., Hatziargiou, A.G., 1997. Eukaryotic promoter recognition. *Genome Res.* 7, 861–878.
- Firth, H.V., et al., 2009. DECIPHER: database of chromosomal imbalance and phenotype in humans using Ensembl resources. *Am. J. Hum. Genet.* 84, 524–533. <https://doi.org/10.1016/j.ajhg.2009.03.010>.
- Foster, S.R., et al., 2015. Cardiac gene expression data and in silico analysis provide novel insights into human and mouse taste receptor gene regulation. *N. Schmied. Arch. Pharmacol.* 388, 1009–1027. <https://doi.org/10.1007/s00210-015-1118-1>.
- Freeman, J.L., et al., 2006. Copy number variation: new insights in genome diversity. *Genome Res.* 16, 949–961. <https://doi.org/10.1101/gr.3677206>.
- Freire-Pritchett, P., et al., 2017. Global reorganisation of cis-regulatory units upon lineage commitment of human embryonic stem cells. *eLife* 6. <https://doi.org/10.7554/eLife.21926>.
- Gan, K.A., Carrasco Pro, S., Sewell, J.A., Fuxman Bass, J.I., 2018. Identification of single nucleotide non-coding driver mutations in cancer. *Front. Genet.* 9, 16. <https://doi.org/10.3389/fgene.2018.00016>.
- Garg, V., et al., 2003. GATA4 mutations cause human congenital heart defects and reveal an interaction with TBX5. *Nature* 424, 443–447. <https://doi.org/10.1038/nature01827>.
- Geisler, S., Collier, J., 2013. RNA in unexpected places: long non-coding RNA functions in diverse cellular contexts. *Nat. Rev. Mol. Cell Biol.* 14, 699–712. <https://doi.org/10.1038/nrm3679>.
- Geng, H.H., et al., 2016. The circular RNA Cdr1a promotes myocardial infarction by mediating the regulation of miR-7a on its target genes expression. *PLoS One* 11, e0151753. <https://doi.org/10.1371/journal.pone.0151753>.
- Giltsch, R., et al., 2018. Distinct epigenetic programs regulate cardiac myocyte development and disease in the human heart in vivo. *Nat. Commun.* 9, 391. <https://doi.org/10.1038/s41467-017-02762-z>.

- Giresi, P.G., Kim, J., McDaniel, R.M., Iyer, V.R., Lieb, J.D., 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res.* 17, 877–885. <https://doi.org/10.1101/gr.5533506>.
- Granados-Riveron, J.T., et al., 2012. Combined mutation screening of NKX2-5, GATA4, and TBX5 in congenital heart disease: multiple heterozygosity and novel mutations. *Congenit. Heart Dis.* 7, 151–159. <https://doi.org/10.1111/j.1747-0803.2011.00573.x>.
- Griffiths-Jones, S., Grocock, R.J., van Dongen, S., Bateman, A., Enright, A.J., 2006. miRBase: microRNA sequences, targets and gene nomenclature. *Nucleic Acids Res.* 34, D140–D144. <https://doi.org/10.1093/nar/gkj112>.
- Guida, V., et al., 2010. Multiplex ligation-dependent probe amplification analysis of GATA4 gene copy number variations in patients with isolated congenital heart disease. *Dis. Markers* 28, 287–292. <https://doi.org/10.3233/dma-2010-0703>.
- Gupta, R.A., et al., 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* 464, 1071–1076. <https://doi.org/10.1038/nature08975>.
- Haeri, M., Gelovani, V., Beaudet, A.L., 2016. Chromosomal microarray analysis, or comparative genomic hybridization: a high throughput approach. *MethodsX* 3, 8–18. <https://doi.org/10.1016/j.mex.2015.11.005>.
- Hafner, M., et al., 2008. Identification of microRNAs and other small regulatory RNAs using cDNA library sequencing. *Methods (San Diego, Calif.)* 44, 3–12. <https://doi.org/10.1016/j.ymeth.2007.09.009>.
- Han, P., et al., 2014. A long noncoding RNA protects the heart from pathological hypertrophy. *Nature* 514, 102–106. <https://doi.org/10.1038/nature13596>.
- Handsaker, R.E., et al., 2015. Large multiallelic copy number variations in humans. *Nat. Genet.* 47, 296–303. <https://doi.org/10.1038/ng.3200>.
- Harrison, S.M., et al., 2016. Using ClinVar as a resource to Support variant interpretation. *Curr. Protoc. Human Genet.* 89. <https://doi.org/10.1002/0471142905.hg0816s89>. 8. 16.11–18.16.23.
- Harrow, J., et al., 2012. GENCODE: the reference human genome annotation for the ENCODE Project. *Genome Res.* 22, 1760–1774. <https://doi.org/10.1101/gr.135350.111>.
- Hart, S.N., et al., 2013. SoftSearch: integration of multiple sequence features to identify breakpoints of structural variations. *PLoS One* 8, e83356. <https://doi.org/10.1371/journal.pone.0083356>.
- Hartman, R.J., et al., 2011. The contribution of chromosomal abnormalities to congenital heart defects: a population-based study. *Pediatr. Cardiol.* 32, 1147–1157. <https://doi.org/10.1007/s00246-011-0034-5>.
- He, A., Kong, S.W., Ma, Q., Pu, W.T., 2011. Co-occupancy by multiple cardiac transcription factors identifies transcriptional enhancers active in heart. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5632–5637. <https://doi.org/10.1073/pnas.1016959108>.
- He, A., et al., 2014. Dynamic GATA4 enhancers shape the chromatin landscape central to heart development and disease. *Nat. Commun.* 5, 4907. <https://doi.org/10.1038/ncomms5907>.
- Heintzman, N.D., et al., 2007. Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. *Nat. Genet.* 39, 311–318. <https://doi.org/10.1038/ng1966>.
- Heintzman, N.D., et al., 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459, 108–112. <https://doi.org/10.1038/nature07829>.
- Herrmann, F., Gross, A., Zhou, D., Kestler, H.A., Kuhl, M., 2012. A boolean model of the cardiac gene regulatory network determining first and second heart field identity. *PLoS One* 7, e46798. <https://doi.org/10.1371/journal.pone.0046798>.
- Hestand, M.S., et al., 2008. CORE.TF: a user-friendly interface to identify evolutionary conserved transcription factor binding sites in sets of co-regulated genes. *BMC Bioinf.* 9, 495. <https://doi.org/10.1186/1471-2105-9-495>.
- Hoffman, J., 2013. The global burden of congenital heart disease. *Cardiovasc. J. Afr.* 24, 141–145. <https://doi.org/10.5830/cvja-2013-028>.
- Hornshoj, H., et al., 2018. Pan-cancer screen for mutations in non-coding elements with conservation and cancer specificity reveals correlations with expression and survival. *NPJ Genom. Med.* 3, 1. <https://doi.org/10.1038/s41525-017-0040-5>.
- Hu, Z., et al., 2013. A genome-wide association study identifies two risk loci for congenital heart malformations in Han Chinese populations. *Nat. Genet.* 45, 818–821. <https://doi.org/10.1038/ng.2636>.
- Huarte, M., et al., 2010. A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. *Cell* 142, 409–419. <https://doi.org/10.1016/j.cell.2010.06.040>.
- Ibn-Salem, J., et al., 2014. Deletions of chromosomal regulatory boundaries are associated with congenital disease. *Genome Biol.* 15, 423. <https://doi.org/10.1186/s13059-014-0423-1>.
- Iyer, M.K., et al., 2015. The landscape of long noncoding RNAs in the human transcriptome. *Nat. Genet.* 47, 199–208. <https://doi.org/10.1038/ng.3192>.
- Jenkins, K.J., et al., 2007. Noninherited risk factors and congenital cardiovascular defects: current knowledge: a scientific statement from the American Heart Association Council on Cardiovascular Disease in the Young: endorsed by the American Academy of Pediatrics. *Circulation* 115, 2995–3014. <https://doi.org/10.1161/circulationaha.106.183216>.
- Jia, Y., et al., 2015. The diagnostic value of next generation sequencing in familial nonsyndromic congenital heart defects. *Am. J. Med. Genet. Part A* 167a, 1822–1829. <https://doi.org/10.1002/ajmg.a.37108>.
- Jia, G., et al., 2018. Single cell RNA-seq and ATAC-seq analysis of cardiac progenitor cell transition states and lineage settlement. *Nat. Commun.* 9, 4877. <https://doi.org/10.1038/s41467-018-07307-6>.
- Jin, B., Li, Y., Robertson, K.D., 2011. DNA methylation: superior or subordinate in the epigenetic hierarchy? *Genes Cancer* 2, 607–617. <https://doi.org/10.1177/1947601910393957>.
- Khan, A., et al., 2018. JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. *Nucleic Acids Res.* 46, D260–D266. <https://doi.org/10.1093/nar/gkx1126>.
- Khoshnood, B., et al., 2005. Trends in prenatal diagnosis, pregnancy termination, and perinatal mortality of newborns with congenital heart disease in France, 1983–2000: a population-based evaluation. *Pediatrics* 115, 95–101. <https://doi.org/10.1542/peds.2004-0516>.
- Klattenhoff, C.A., et al., 2013. Braveheart, a long noncoding RNA required for cardiovascular lineage commitment. *Cell* 152, 570–583. <https://doi.org/10.1016/j.cell.2013.01.003>.
- Kolesnikov, N., et al., 2015. ArrayExpress update—simplifying data submissions. *Nucleic Acids Res.* 43, D1113–D1116. <https://doi.org/10.1093/nar/gku1057>.
- Kulakovskiy, I.V., et al., 2013. HOCOMOCO: a comprehensive collection of human transcription factor binding sites models. *Nucleic Acids Res.* 41, D195–D202. <https://doi.org/10.1093/nar/gks1089>.
- Kulakovskiy, I.V., et al., 2018. HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Res.* 46, D252–d259. <https://doi.org/10.1093/nar/gkx1106>.
- Kurdyukov, S., Bullock, M., 2016. DNA methylation analysis: choosing the right method. *Biology* 5. <https://doi.org/10.3390/biology5010003>.
- Kvon, E.Z., 2015. Using transgenic reporter assays to functionally characterize enhancers in animals. *Genomics* 106, 185–192. <https://doi.org/10.1016/j.ygeno.2015.06.007>.
- Kwon, C., Han, Z., Olson, E.N., Srivastava, D., 2005. MicroRNA1 influences cardiac differentiation in Drosophila and regulates Notch signaling. *Proc. Natl. Acad. Sci. U. S. A.* 102, 18986–18991. <https://doi.org/10.1073/pnas.0509535102>.
- Kyrchanova, O., Georgiev, P., 2014. Chromatin insulators and long-distance interactions in Drosophila. *FEBS Lett.* 588, 8–14. <https://doi.org/10.1016/j.febslet.2013.10.039>.
- Landrum, M.J., et al., 2018. ClinVar: improving access to variant interpretations and supporting evidence. *Nucleic Acids Res.* 46, D1062–d1067. <https://doi.org/10.1093/nar/gkx1153>.
- Lappalainen, L., et al., 2013. DbVar and DGVA: public archives for genomic structural variation. *Nucleic Acids Res.* 41, D936–D941. <https://doi.org/10.1093/nar/gks1213>.
- Layer, R.M., Chiang, C., Quinlan, A.R., Hall, I.M., 2014. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol.* 15, R84. <https://doi.org/10.1186/gb-2014-15-6-r84>.
- Lee, K., et al., 2011. Precursor miR-886, a novel noncoding RNA repressed in cancer, associates with PKR and modulates its activity. *RNA (New York, N.Y.)* 17, 1076–1089. <https://doi.org/10.1261/rna.2701111>.
- Lenhard, B., et al., 2003. Identification of conserved regulatory elements by comparative genome analysis. *J. Biol.* 2, 13. <https://doi.org/10.1186/1475-4924-2-13>.
- Li, M.J., et al., 2012. GWASdb: a database for human genetic variants identified by genome-wide association studies. *Nucleic Acids Res.* 40, D1047–D1054. <https://doi.org/10.1093/nar/gkr1182>.
- Li, S., Xu, Z., Sheng, 2018. J. tRNA-derived small RNA: a novel regulatory small non-coding RNA. *Genes* 9. <https://doi.org/10.3390/genes9050246>.
- Liberatore, C.M., Searcy-Schrick, R.D., Vincent, E.B., Yutzey, K.E., 2002. Nkx-2.5 gene induction in mice is mediated by a Smad consensus regulatory region. *Dev. Biol.* 244, 243–256. <https://doi.org/10.1006/dbio.2002.0604>.
- Lin, Y., et al., 2015. Association analysis identifies new risk loci for congenital heart disease in Chinese populations. *Nat. Commun.* 6, 8082. <https://doi.org/10.1038/ncomms9082>.
- Liu, C., et al., 2005. NONCODE: an integrated knowledge database of non-coding RNAs. *Nucleic Acids Res.* 33, D112–D115. <https://doi.org/10.1093/nar/gki041>.
- Liu, X., Yu, X., Zack, D.J., Zhu, H., Qian, J., 2008. TIGER: a database for tissue-specific gene expression and regulation. *BMC Bioinf.* 9, 271. <https://doi.org/10.1186/1471-2105-9-271>.
- Liu, Y.C., et al., 2016. CircNet: a database of circular RNAs derived from transcriptome sequencing data. *Nucleic Acids Res.* 44, D209–D215. <https://doi.org/10.1093/nar/gkv940>.
- Lu, C., Meyers, B.C., Green, P.J., 2007. Construction of small RNA cDNA libraries for deep sequencing. *Methods (San Diego, Calif.)* 43, 110–117. <https://doi.org/10.1016/j.ymeth.2007.05.002>.
- Ma, L., et al., 2015. LncRNAWiki: harnessing community knowledge in collaborative curation of human long non-coding RNAs. *Nucleic Acids Res.* 43, D187–D192. <https://doi.org/10.1093/nar/gku1167>.
- MacArthur, J., et al., 2017. The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog). *Nucleic Acids Res.* 45, D896–d901. <https://doi.org/10.1093/nar/gkw1133>.
- Martinez, S.R., Gay, M.S., Zhang, L., 2015. Epigenetic mechanisms in heart development and disease. *Drug Discov. Today* 20, 799–811. <https://doi.org/10.1016/j.drudis.2014.12.018>.
- Maston, G.A., Evans, S.K., Green, M.R., 2006. Transcriptional regulatory elements in the human genome. *Annu. Rev. Genom. Hum. Genet.* 7, 29–59. <https://doi.org/10.1146/annurev.genom.7.080505.115623>.
- Mathelier, A., Shi, W., Wasserman, W.W., 2015. Identification of altered cis-regulatory elements in human disease. *Trends Genet.* TIG 31, 67–76. <https://doi.org/10.1016/j.tig.2014.12.003>.
- May, D., et al., 2011. Large-scale discovery of enhancers from human heart tissue. *Nat. Genet.* 44, 89–93. <https://doi.org/10.1038/ng.1006>.
- McCulley, D.J., Black, B.L., 2012. Transcription factor pathways and congenital heart disease. *Curr. Top. Dev. Biol.* 100, 253–277. <https://doi.org/10.1016/b978-0-12-387786-4.00008-7>.
- McDermott, D.A., et al., 2005. TBX5 genetic testing validates strict clinical criteria for Holt-Oram syndrome. *Pediatr. Res.* 58, 981–986. <https://doi.org/10.1203/01.Pdr.0000182593.95441.64>.
- Miskimen, K.L., Chan, E.R., Haines, J.L., 2017. Assay for transposase-accessible chromatin

- using sequencing (ATAC-seq) data analysis. *Curr. Protoc. Human Genet.* 92 <https://doi.org/10.1002/cphg.32>. 20.24.21–20.24.13.
- Mitchell, S.C., Korones, S.B., Berendes, H.W., 1971. Congenital heart disease in 56,109 births. Incidence and natural history. *Circulation* 43, 323–332.
- Moon, A., 2008. Mouse models of congenital cardiovascular disease. *Curr. Top. Dev. Biol.* 84, 171–248. [https://doi.org/10.1016/s0070-2153\(08\)00604-2](https://doi.org/10.1016/s0070-2153(08)00604-2).
- Montefiori, L.E., et al., 2018. A promoter interaction map for cardiovascular disease genetics. *Elife* 7, 826. <https://doi.org/10.7554/eLife.35788>.
- Morin, S., Charron, F., Robitaille, L., Nemer, M., 2000. GATA-dependent recruitment of MEF2 proteins to target promoters. *EMBO J.* 19, 2046–2055. <https://doi.org/10.1093/emboj/19.9.2046>.
- Muerdter, F., Boryn, L.M., Arnold, C.D., 2015. STARR-seq - principles and applications. *Genomics* 106, 145–150. <https://doi.org/10.1016/j.ygeno.2015.06.001>.
- Narlikar, L., Ovcharenko, I., 2009. Identifying regulatory elements in eukaryotic genomes. *Briefings Funct. Genomics Proteomics* 8, 215–230. <https://doi.org/10.1093/bfpg/elp014>.
- Narlikar, L., et al., 2010. Genome-wide discovery of human heart enhancers. *Genome Res.* 20, 381–392. <https://doi.org/10.1101/gr.098657.109>.
- Newburger, D.E., Bulyk, M.L., 2009. UniPROBE: an online database of protein binding microarray data on protein-DNA interactions. *Nucleic Acids Res.* 37, D77–D82. <https://doi.org/10.1093/nar/gkn660>.
- OGeen, H., Echipare, L., Farnham, P.J., 2011. Using ChIP-seq technology to generate high-resolution profiles of histone modifications. *Methods Mol. Biol. (Clifton, N.J.)* 791, 265–286. https://doi.org/10.1007/978-1-61779-316-5_20.
- Papageorgiou, N., et al., 2012. The role of microRNAs in cardiovascular disease. *Curr. Med. Chem.* 19, 2605–2610.
- Papatheodorou, I., et al., 2018. Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res.* 46, D246–d251. <https://doi.org/10.1093/nar/gkx1158>.
- Paraskevopoulou, M.D., et al., 2016. DIANA-LncBase v2: indexing microRNA targets on non-coding transcripts. *Nucleic Acids Res.* 44, D231–D238. <https://doi.org/10.1093/nar/gkv1270>.
- Patel, S.S., Burns, T.L., 2013. Nongenetic risk factors and congenital heart defects. *Pediatr. Cardiol.* 34, 1535–1555. <https://doi.org/10.1007/s00246-013-0775-4>.
- Pertea, M., Salzberg, S.L., 2010. Between a chicken and a grape: estimating the number of human genes. *Genome Biol.* 11, 206. <https://doi.org/10.1186/gb-2010-11-5-206>.
- Peschansky, V.J., Wahlestedt, C., 2014. Non-coding RNAs as direct and indirect modulators of epigenetic regulation. *Epigenetics* 9, 3–12. <https://doi.org/10.4161/epi.27473>.
- Petrov, A.I., et al., 2017. RNACentral: a comprehensive database of non-coding RNA sequences. *Nucleic Acids Res.* 45, D128–d134. <https://doi.org/10.1093/nar/gkw1008>.
- Prall, O.W., et al., 2007. An Nkx2-5/Bmp2/Smad1 negative feedback loop controls heart progenitor specification and proliferation. *Cell* 128, 947–959. <https://doi.org/10.1016/j.cell.2007.01.042>.
- Quek, X.C., et al., 2015. lncRNADB v2.0: expanding the reference database for functional long noncoding RNAs. *Nucleic Acids Res.* 43, D168–D173. <https://doi.org/10.1093/nar/gku988>.
- Raha, D., Hong, M., Snyder, M., 2010. ChIP-Seq: a method for global identification of regulatory elements in the genome. *Curr. Protoc. Mol. Biol.* <https://doi.org/10.1002/0471142727.mb2119s91>. (Chapter 21), Unit 21.19.21–14.
- Rajan, K.S., et al., 2016. Abundant and altered expression of PIWI-interacting RNAs during cardiac hypertrophy. *Heart Lung Circ.* 25, 1013–1020. <https://doi.org/10.1016/j.hlc.2016.02.015>.
- Ramialison, M., et al., 2017. Analysis of steric effects in DamID profiling of transcription factor target genes. *Genomics* 109, 75–82. <https://doi.org/10.1016/j.ygeno.2017.01.006>.
- Rausch, T., et al., 2012. DELLY: structural variant discovery by integrated paired-end and split-read analysis. *Bioinformatics (Oxf. Engl.)* 28, i333–i339. <https://doi.org/10.1093/bioinformatics/bts378>.
- Reecy, J.M., et al., 1999. Identification of Upstream Regulatory Regions in the Heart-Expressed Homeobox Gene Nkx2-5, vol. 126. Development, Cambridge, England, pp. 839–849.
- Richards, Garg, 2010. Genetics of Congenital Heart disease.pdf.
- Rinn, J.L., et al., 2007. Functional demarcation of active and silent chromatin domains in human HOX loci by noncoding RNAs. *Cell* 129, 1311–1323. <https://doi.org/10.1016/j.cell.2007.05.022>.
- Rizzo, J.M., Bard, J.E., Buck, M.J., 2012. Standardized collection of MNase-seq experiments enables unbiased dataset comparisons. *BMC Mol. Biol.* 13, 15. <https://doi.org/10.1186/1471-2199-13-15>.
- Romanoski, C.E., Glass, C.K., Stunnenberg, H.G., Wilson, L., Almouzni, G., 2015. Epigenetics: Roadmap for regulation. *Nature* 518, 314–316. <https://doi.org/10.1038/518314a>.
- Rosa-Garrido, M., et al., 2017. High-resolution mapping of chromatin conformation in cardiac myocytes reveals structural remodeling of the epigenome in heart failure. *Circulation* 136, 1613–1625. <https://doi.org/10.1161/circulationaha.117.029430>.
- Ross, R.J., Weiner, M.M., Lin, H., 2014. PIWI proteins and PIWI-interacting RNAs in the soma. *Nature* 505, 353–359. <https://doi.org/10.1038/nature12987>.
- Sabina, S., et al., 2013. Germline hereditary, somatic mutations and microRNAs targeting-SNPs in congenital heart defects. *J. Mol. Cell. Cardiol.* 60, 84–89. <https://doi.org/10.1016/j.yjmcc.2013.04.002>.
- Sandve, G.K., Drablos, F., 2006. A survey of motif discovery methods in an integrated framework. *Biol. Direct* 1, 11. <https://doi.org/10.1186/1745-6150-1-11>.
- Sayed, D., Abdellatif, M., 2011. MicroRNAs in development and disease. *Physiol. Rev.* 91, 827–887. <https://doi.org/10.1152/physrev.00006.2010>.
- Scacheri, C.A., Scacheri, P.C., 2015. Mutations in the noncoding genome. *Curr. Opin. Pediatr.* 27, 659–664. <https://doi.org/10.1097/MOP.0000000000000283>.
- Scherer, S.W., et al., 2007. Challenges and standards in integrating surveys of structural variation. *Nat. Genet.* 39, S7–S15. <https://doi.org/10.1038/ng2093>.
- Schlesinger, J., et al., 2011. The cardiac transcription network modulated by Gata4, Mef2a, Nkx2.5, Srf, histone modifications, and microRNAs. *PLoS Genet.* 7, e1001313. <https://doi.org/10.1371/journal.pgen.1001313>.
- Schmitt, A.D., et al., 2016. A Compendium of chromatin contact maps reveals spatially active regions in the human genome. *Cell Rep* 17 (8), 2042–2059. <https://doi.org/10.1016/j.celrep.2016.10.061>.
- Schmitz, S.U., Grote, P., Herrmann, B.G., 2016. Mechanisms of long noncoding RNA function in development and disease. *Cell. Mol. Life Sci. CMLS* 73, 2491–2509. <https://doi.org/10.1007/s00018-016-2174-5>.
- Schonrock, N., Harvey, R.P., Mattick, J.S., 2012. Long noncoding RNAs in cardiac development and pathophysiology. *Circ. Res.* 111, 1349–1362. <https://doi.org/10.1161/circresaha.112.268953>.
- Schott, J.J., et al., 1998. Congenital heart disease caused by mutations in the transcription factor NKX2-5. *Science (New York, N.Y.)* 281, 108–111.
- Schulte, C., Zeller, T., 2015. microRNA-based diagnostics and therapy in cardiovascular disease—Summing up the facts. *Cardiovasc. Diagn. Ther.* 5, 17–36. <https://doi.org/10.3978/j.issn.2223-3652.2014.12.03>.
- Schwartz, R.J., Olson, E.N., 1999. Building the Heart Piece by Piece: Modularity of cis-Elements Regulating Nkx2-5 Transcription, vol. 126. Development, Cambridge, England, pp. 4187–4192.
- Sharp, A.J., et al., 2005. Segmental duplications and copy-number variation in the human genome. *Am. J. Hum. Genet.* 77, 78–88. <https://doi.org/10.1086/431652>.
- Shi, H., et al., 2016. Gestational stress induces the unfolded protein response, resulting in heart defects. *Dev. (Camb. Engl.)* 143, 2561–2572. <https://doi.org/10.1242/dev.136820>.
- Simon, M.D., 2013. Capture hybridization analysis of RNA targets (CHART). *Curr. Protoc. Mol. Biol.* <https://doi.org/10.1002/0471142727.mb2125s101>. (Chapter 21), Unit 21.25.
- Smedley, D., et al., 2015. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res.* 43, W589–W598. <https://doi.org/10.1093/nar/gkv350>.
- Smemo, S., et al., 2012. Regulatory variation in a TBX5 enhancer leads to isolated congenital heart disease. *Hum. Mol. Genet.* 21, 3255–3263. <https://doi.org/10.1093/hmg/dds165>.
- Smith, T., Rajakaruna, C., Caputo, M., Emanueli, C., 2015. MicroRNAs in congenital heart disease. *Ann. Transl. Med.* 3, 333. <https://doi.org/10.3978/j.issn.2305-5839.2015.12.25>.
- Soares, R.J., et al., 2018. Evaluation of fluorescence in situ hybridization techniques to study long non-coding RNA expression in cultured cells. *Nucleic Acids Res.* 46, e4. <https://doi.org/10.1093/nar/gkx946>.
- Song, L., Crawford, G.E., 2010. DNase-seq: a high-resolution technique for mapping active gene regulatory elements across the genome from mammalian cells. *Cold Spring Harbor Protoc.* 2010 <https://doi.org/10.1101/pdb.prot5384>. <https://doi.org/10.1101/pdb.prot5384>.
- Spitz, F., Furlong, E.E., 2012. Transcription factors: from enhancer binding to developmental control. *Nat. Rev. Genet.* 13, 613–626. <https://doi.org/10.1038/nrg3207>.
- Stuppia, L., Antonucci, I., Palka, G., Gatta, V., 2012. Use of the MLPA assay in the molecular diagnosis of gene copy number alterations in human genetic diseases. *Int. J. Mol. Sci.* 13, 3245–3276. <https://doi.org/10.3390/ijms13033245>.
- Tattini, L., D'Aurizio, R., Magi, A., 2015. Detection of genomic structural variants from next-generation sequencing data. *Front. Bioeng. Biotechnol.* 3, 92. <https://doi.org/10.3389/fbioe.2015.00092>.
- Tennant, P.W., Pearce, M.S., Bythell, M., Rankin, J., 2010. 20-year survival of children born with congenital anomalies: a population-based study. *Lancet (Lond. Engl.)* 375, 649–656. [https://doi.org/10.1016/s0140-6736\(09\)61922-x](https://doi.org/10.1016/s0140-6736(09)61922-x).
- Thienpont, B., et al., 2007. Submicroscopic chromosomal imbalances detected by array-CGH are a frequent cause of congenital heart defects in selected patients. *Eur. Heart J.* 28, 2778–2784. <https://doi.org/10.1093/eurheartj/ehl560>.
- Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., de Laat, W., 2002. Looping and interaction between hypersensitive sites in the active beta-globin locus. *Mol. Cell* 10, 1453–1465.
- Uchida, S., 2017. High-throughput methods to detect long non-coding RNAs. High-throughput 6. <https://doi.org/10.3390/ht6030012>.
- Vallaster, M., Vallaster, C.D., Wu, S.M., 2012. Epigenetic mechanisms in cardiac development and disease. *Acta Biochim. Biophys. Sin.* 44, 92–102. <https://doi.org/10.1093/abbs/gmr090>.
- van den Boogaard, M., et al., 2014. A common genetic variant within SCN10A modulates cardiac SCN5A expression. *J. Clin. Investig.* 124, 1844–1852. <https://doi.org/10.1172/jci73140>.
- van der Linde, D., et al., 2011. Birth prevalence of congenital heart disease worldwide: a systematic review and meta-analysis. *J. Am. Coll. Cardiol.* 58, 2241–2247. <https://doi.org/10.1016/j.jacc.2011.08.025>.
- van der Velde, E.T., et al., 2005. CONCOR, an initiative towards a national registry and DNA-bank of patients with congenital heart disease in The Netherlands: rationale, design, and first results. *Eur. J. Epidemiol.* 20, 549–557.
- van Duijvenboden, K., Ruijter, J.M., Christoffels, V.M., 2014. Gene regulatory elements of the cardiac conduction system. *Brief. Funct. Genom.* 13, 28–38. <https://doi.org/10.1093/bfpg/elt031>.
- van Steensel, B., Henikoff, S., 2000. Identification of in vivo DNA targets of chromatin proteins using tethered dam methyltransferase. *Nat. Biotechnol.* 18, 424–428. <https://doi.org/10.1038/74487>.
- Vance, K.W., 2017. Mapping long noncoding RNA chromatin occupancy using capture hybridization analysis of RNA targets (CHART). *Methods Mol. Biol. (Clifton, N.J.)* 1468, 39–50. https://doi.org/10.1007/978-1-4939-4035-6_5.
- Vecoli, C., Pulignani, S., Foffa, I., Andreassi, M.G., 2014. Congenital heart disease: the

- crossroads of genetics, epigenetics and environment. *Curr. Genom.* 15, 390–399. <https://doi.org/10.2174/1389202915666140716175634>.
- Viereck, J., et al., 2016. Long noncoding RNA Chast promotes cardiac remodeling. *Sci. Transl. Med.* 8, 326ra322. <https://doi.org/10.1126/scitranslmed.aaf1475>.
- Vincentz, J.W., Barnes, R.M., Firulli, B.A., Conway, S.J., Firulli, A.B., 2008. Cooperative Interaction of Nkx2.5 and Mef2c Transcription Factors during Heart Development, vol. 237. Developmental dynamics : an official publication of the American Association of Anatomists, pp. 3809–3819. <https://doi.org/10.1002/dvdy.21803>.
- Visel, A., Minovitsky, S., Dubchak, I., Pennacchio, L.A., 2007. VISTA Enhancer Browser—a database of tissue-specific human enhancers. *Nucleic Acids Res.* 35, D88–D92. <https://doi.org/10.1093/nar/gkl822>.
- Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five years of GWAS discovery. *Am. J. Hum. Genet.* 90, 7–24. <https://doi.org/10.1016/j.ajhg.2011.11.029>.
- Vlieghe, D., et al., 2006. A new generation of JASPAR, the open-access repository for transcription factor binding site profiles. *Nucleic Acids Res.* 34, D95–D97. <https://doi.org/10.1093/nar/gkj115>.
- Volders, P.J., et al., 2013. LNCipedia: a database for annotated human lncRNA transcript sequences and structures. *Nucleic Acids Res.* 41, D246–D251. <https://doi.org/10.1093/nar/gks915>.
- Waldron, L., et al., 2016. The cardiac TBX5 interactome reveals a chromatin remodeling network essential for cardiac septation. *Dev. Cell* 36, 262–275. <https://doi.org/10.1016/j.devcel.2016.01.009>.
- Wamstad, J.A., et al., 2012. Dynamic and coordinated epigenetic regulation of developmental transitions in the cardiac lineage. *Cell* 151, 206–220. <https://doi.org/10.1016/j.cell.2012.07.035>.
- Wamstad, J.A., Wang, X., Demuren, O.O., Boyer, L.A., 2014. Distal enhancers: new insights into heart development and disease. *Trends Cell Biol.* 24, 294–302. <https://doi.org/10.1016/j.tcb.2013.10.008>.
- Wang, Z., Wang, Y., 2015. Dawn of the Epi-LncRNAs: new path from Myheart. *Circ. Res.* 116, 235–236. <https://doi.org/10.1161/circresaha.114.305490>.
- Wang, S., et al., 2008. The endothelial-specific microRNA miR-126 governs vascular integrity and angiogenesis. *Dev. Cell* 15, 261–271. <https://doi.org/10.1016/j.devcel.2008.07.002>.
- Wang, J., Liew, O.W., Richards, A.M., Chen, Y.T., 2016a. Overview of MicroRNAs in cardiac hypertrophy, fibrosis, and apoptosis. *Int. J. Mol. Sci.* 17. <https://doi.org/10.3390/ijms17050749>.
- Wang, Z., et al., 2016b. The long noncoding RNA Chaer defines an epigenetic checkpoint in cardiac hypertrophy. *Nat. Med.* 22, 1131–1139. <https://doi.org/10.1038/nm.4179>.
- Wang, K., et al., 2016c. A circular RNA protects the heart from pathological hypertrophy and heart failure by targeting miR-223. *Eur. Heart J.* 37, 2602–2611. <https://doi.org/10.1093/eurheartj/ehv713>.
- Wang, X., Lin, P., Ho, J.W.K., 2018. Discovery of cell-type specific DNA motif grammar in cis-regulatory elements using random Forest. *BMC Genomics* 19, 929. <https://doi.org/10.1186/s12864-017-4340-z>.
- Wasserman, W.W., Sandelin, A., 2004. Applied bioinformatics for the identification of regulatory elements. *Nat. Rev. Genet.* 5, 276–287. <https://doi.org/10.1038/nrg1315>.
- Wei, W., Yu, X.D., 2007. Comparative analysis of regulatory motif discovery tools for transcription factor binding sites. *Genom. Proteom. Bioinform.* 5, 131–142. [https://doi.org/10.1016/s1672-0229\(07\)60023-0](https://doi.org/10.1016/s1672-0229(07)60023-0).
- Weingarten-Gabbay, S., Segal, E., 2014. The grammar of transcriptional regulation. *Hum. Genet.* 133, 701–711. <https://doi.org/10.1007/s00439-013-1413-1>.
- Weirick, T., et al., 2016. The identification and characterization of novel transcripts from RNA-seq data. *Briefings Bioinform.* 17, 678–685. <https://doi.org/10.1093/bib/bbv067>.
- Welter, D., et al., 2014. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. *Nucleic Acids Res.* 42, D1001–D1006. <https://doi.org/10.1093/nar/gkt1229>.
- Wong, K.C., 2017. MotifHyades: expectation maximization for de novo DNA motif pair discovery on paired sequences. *Bioinformatics (Oxf. Engl.)* 33, 3028–3035. <https://doi.org/10.1093/bioinformatics/btx381>.
- Wu, C., et al., 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome Biol.* 10, R130. <https://doi.org/10.1186/gb-2009-10-11-r130>.
- Xie, W.Q., Zhou, L., Chen, Y., Ni, B., 2016. Circulating microRNAs as potential biomarkers for diagnosis of congenital heart defects. *World J. Emerg. Med.* 7, 85–89. <https://doi.org/10.5847/wjem.j.1920-8642.2016.02.001>.
- Xing, H.J., et al., 2013. Identification of microRNAs present in congenital heart disease associated copy number variants. *Eur. Rev. Med. Pharmacol. Sci.* 17, 2114–2120.
- Yan, Y., et al., 2014. Detection of submicroscopic chromosomal aberrations by array-based comparative genomic hybridization in fetuses with congenital heart disease. *Ultrasound Obstet. Gynecol. Offic. J. Int. Soc. Ultrasound Obstet. Gynecol.* 43, 404–412. <https://doi.org/10.1002/uog.13236>.
- Yoshida, A., et al., 2016. Genetic mutation analysis in Japanese patients with non-syndromic congenital heart disease. *J. Hum. Genet.* 61, 157–162. <https://doi.org/10.1038/jhg.2015.126>.
- Yuan, X., et al., 2018. Heart enhancers with deeply conserved regulatory activity are established early in zebrafish development. *Nat. Commun.* 9, 4977. <https://doi.org/10.1038/s41467-018-07451-z>.
- Zahavich, L., Bowdin, S., Mital, S., 2017. Use of clinical exome sequencing in isolated congenital heart disease. *Circ. Cardiovasc. Genet.* 10. <https://doi.org/10.1161/circgenetics.116.001581>.
- Zaidi, S., et al., 2013. De novo mutations in histone-modifying genes in congenital heart disease. *Nature* 498, 220–223. <https://doi.org/10.1038/nature12141>.
- Zerbino, D.R., et al., 2016. Ensembl regulation resources. *Database J. Biol. Databases Curation* 2016. <https://doi.org/10.1093/database/bav119>.
- Zhang, Q.J., Liu, Z.P., 2015. Histone methylations in heart development, congenital and adult heart diseases. *Epigenomics* 7, 321–330. <https://doi.org/10.2217/epi.14.60>.
- Zhang, B., et al., 2012. The lncRNA Malat1 is dispensable for mouse development but its transcription plays a cis-regulatory role in the adult. *Cell Rep.* 2, 111–123. <https://doi.org/10.1016/j.celrep.2012.06.003>.
- Zhang, W., Spector, T.D., Deloukas, P., Bell, J.T., Engelhardt, B.E., 2015. Predicting genome-wide DNA methylation using methylation marks, genomic position, and DNA regulatory elements. *Genome Biol.* 16, 14. <https://doi.org/10.1186/s13059-015-0581-9>.
- Zhao, Y., et al., 2007. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* 129, 303–317. <https://doi.org/10.1016/j.cell.2007.03.030>.
- Zhao, B., et al., 2014. Replication of the 4p16 susceptibility locus in congenital heart disease in Han Chinese populations. *PLoS One* 9, e107411. <https://doi.org/10.1371/journal.pone.0107411>.
- Zhou, J., Troyanskaya, O.G., 2015. Predicting effects of noncoding variants with deep learning-based sequence model. *Nat. Methods* 12, 931–934. <https://doi.org/10.1038/nmeth.3547>.
- Zhou, Y., Kim, J., Yuan, X., Braun, T., 2011a. Epigenetic modifications of stem cells: a paradigm for the control of cardiac progenitor cells. *Circ. Res.* 109, 1067–1081. <https://doi.org/10.1161/circresaha.111.243709>.
- Zhou, V.W., Goren, A., Bernstein, B.E., 2011b. Charting histone modifications and the functional organization of mammalian genomes. *Nat. Rev. Genet.* 12, 7–18. <https://doi.org/10.1038/nrg2905>.
- Zhu, X., et al., 2016. Identification of copy number variations associated with congenital heart disease by chromosomal microarray analysis and next-generation sequencing. *Prenat. Diagn.* 36, 321–327. <https://doi.org/10.1002/pd.4782>.
- Zinzen, R.P., Girardot, C., Gagneur, J., Braun, M., Furlong, E.E., 2009. Combinatorial binding predicts spatio-temporal cis-regulatory activity. *Nature* 462, 65–70. <https://doi.org/10.1038/nature08531>.