Integrative Bioinformatics in Post-GWAS cardiovascular genomics

Daiane Hemerich

ISBN 978-94-93019-83-6

Cover design by Elisa Calamita
Printed by ProefschriftMaken || Proefschriftmaken.nl

Financial support by the Dutch Heart Foundation for the publication of this thesis is gratefully acknowledged.

This thesis was also accomplished with financial support from Pfizer, the National Council for the Improvement of Higher Education (CAPES) and Science without Borders Project, process 13259/130.

Integrative Bioinformatics in Post-GWAS cardiovascular genomics

Geavanceerde bioinformatica in het post-GWAS tijdperk van cardiovasculair genoom-onderzoek
(met een samenvatting in het Nederlands)

Bioinformática integrativa em genômica cardiovascular pós-GWAS
(com um resumo em português)

Proefschrift

ter verkrijging van de graad van doctor aan de Universiteit Utrecht
op gezag van de rector magnificus, prof.dr. H.R.B.M. Kummeling,
ingevolge het besluit van het college voor promoties
in het openbaar te verdedigen op
donderdag 9 oktober 2018 des ochtends te 10.30 uur

door

Daiane Hemerich

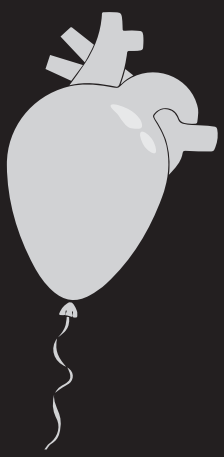
geboren op 13 maart 1989
te Marau, Brazilië

Promotor:	Prof. dr. F. W. Asselbergs
Copromotoren:	Dr. V. Tragante
	Dr. J. van Setten

To my mother Ana, who taught herself how to use a computer
and later a smartphone in order to keep updated
on the work of this thesis and my well-being on a daily basis

Este livro é dedicado à minha mãe Ana,
que ensinou a si mesma como usar um computador
e mais tarde um telefone celular, apenas para poder se manter atualizada
sobre o trabalho desta tese e meu bem-estar diariamente

Table of Contents



Chapter I

Introduction

Cardiovascular Diseases

Cardiovascular diseases (CVDs) are the leading cause of death in the world, taking the lives of 17.7 million people every year and accounting for 31% of all global deaths (Figure 1) [1, 2]. In the United States of America (USA), one third of all deaths is attributed to CVD, with a CVD-related death happening every 40 seconds [3]. The number of CVD-related deaths is predicted to increase to 24.2 million by 2030, with 40.5% of the USA population projected to be affected by some form of CVD [4]. As a result, real indirect costs (lost productivity from morbidity and premature mortality) for all CVD are estimated to increase from $172 billion in 2010 to $276 billion in 2030, an increase of 61% [5]. A similar scenario is predicted in Europe, where current overall CVD is estimated to cost €210 billion a year [6-9].

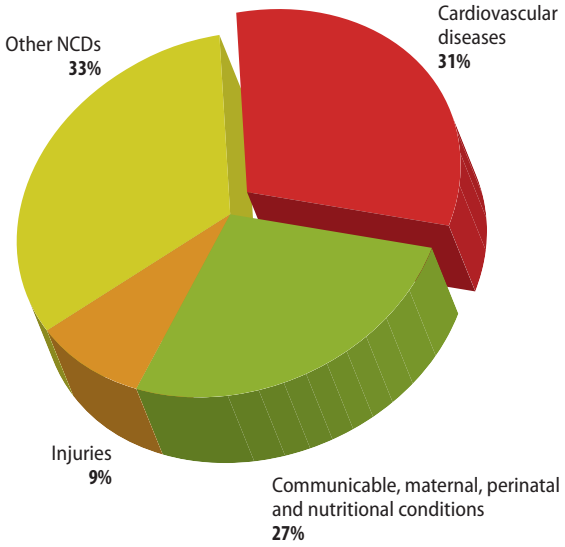


Figure 1. Distribution of major causes of death, including CVDs and other noncommunicable diseases (NCDs). Reproduced from the World Health Organization with permission.

CVDs are a group of disorders related to the heart and circulatory system, including coronary artery disease (CAD), myocardial infarction (MI), high blood pressure (BP), stroke, and heart failure. Key health factors and behaviors increase risks for CVDs, such as smoking, physical inactivity, unhealthy diet, high body weight, high cholesterol, hypertension, and high blood sugar. Indeed, a substantial number of CVD-related deaths can be attributed to tobacco smoking, which increases the risk of dying from coronary heart disease and cerebrovascular disease 2–3 fold [4]. In most cases, it is the

interplay between these modifiable habits with non-modifiable factors, such as age, sex and genetic predisposition which determines the chances of suffering from a CVD.

Genetic variation and human (epi)genetics

Genetic variation is the process of natural occurrence of variation in the DNA sequence among individuals of the same species. Although individuals of a species have similar characteristics, they are hardly ever identical. But it is also relatively evident that some physical characteristics, such as height and hair color, tend to be inherited from an individual's ancestors, and are often similar among members of the same family. Although less visible, the same characteristics dictate susceptibility to some diseases. This is the reason why some diseases are traditionally said to "run in the family".

The process behind it is called inheritance, by which genetic information is passed from the parents to their offspring. Each one of us has a genome, containing genetic information needed to build and maintain our organism. Half of this genome is inherited from each of our parents. This combined genome forms a set of instructions made of deoxyribonucleic acid (DNA), present in all cells of our body that have a nucleous, in the form of chromosomes. DNA is formed by a chemical code determined by the order of four nucleotide bases: adenine (A), cytosine (C), guanine (G) and thymine (T). The human genome consists of three billion pairs of such nucleotides. These encode genes, which dictate characteristics previously mentioned (height, hair color, susceptibility to many diseases) and many others. Subtle differences in our DNA can lead to phenotypic variation. The most common type of genetic variation is the single nucleotide polymorphism (SNP), which is a substitution at a single position in a DNA sequence among individuals. SNPs can influence genes and their expression, thus explaining different characteristics among individuals. SNPs are typically present across populations [10, 11], and inherited together in blocks as a haplotype.

CVDs constitute a wide spectrum of diseases. Severe cases such as cardiomyopathy at young age tend to run within families and cause sudden cardiac death. However, only a few percent of all morbidity and mortality caused by cardiovascular diseases include these severe cases. Most patients are sporadic cases; that is, they have no affected (close) relatives. In contrast to the familiar cases, whose disease is most likely caused by one or few genetic defects, cardiovascular diseases in the general population are multifactorial, complex traits, caused by combinations of environmental and genetic factors.

Unhealthy habits combined with genetic susceptibility increase the likelihood of developing diseases, including CVDs. Some of these unhealthy habits, such as smoking, can also influence the epigenome, thus affecting gene function. These epigenetic changes do not involve changes in the underlying DNA sequence, so the phenotype is affected without changes in the genotype. One form of epigenetic change occurs when molecular structures such as methyl groups attach to genes (a process called DNA

methylation), altering their expression (Figure 2). Chemical elements can also attach to histones, the molecules around which DNA is wrapped. This is the case of acetylation, for instance, when acetyl groups bind to one of the four core histones (Figure 2). In eukaryotes, DNA typically is found as a repeating array of nucleosomes [12], in which 146 base pairs (bp) of DNA are wound around a histone octamer (consisting of two each of histone proteins H2A, H2B, H3, and H4). Histones can be affected by several post-translational modifications, mainly targeting amino acid residues of the N-terminal tails that protrude from the chromatin fiber [13, 14]. Acetylation of histone H3 on lysine residue number 27 (H3K27ac), for instance, is one of several histone modifications known to enhance transcription of genes by opening the chromatin structure and making the DNA more accessible for binding of transcription factors [15]. This histone modification is a marker of regulatory elements known as enhancers and promoters. Other histone modifications, such as trimethylation of histone H3 on lysine 27 (H3K27me3), exert the opposite effect, and thus highlight transcriptionally repressed regions and silent genes. This complex mechanism of chromatin structure is critical to cell function, allowing different cell-types to express (or repress) different genes. Epigenetic modifications of the genome, such as DNA methylation and histone modifications, have also been reported to play a role in diseases such as CVDs, including atherosclerosis, inflammation, hypertension and diabetes [14, 16]. Cardiac hypertrophy, for instance, has been linked to histone acetylation [17] and histone methylation, in particular with H3K9[18-20] and H3K4 methylation[21, 22].

Technological progress made possible the discovery of not only several SNPs associated to CVDs, but also some of their underlying epigenetic mechanisms [14, 16, 23]. This new knowledge can provide a better insight into the biological mechanisms leading to this devastating spectrum of diseases, and ultimately, also provide better prevention and treatment.

Relating genetic variation to disease: Genetic association studies

Since Gregor Mendel's identification of inheritance patterns in 1865 [24], progressing to the discovery of DNA structure by James Watson, Francis Crick, Rosalind Franklin, and Maurice Wilkins in 1944 [25-28], important advances have been made in the genetics field. One of the most recent such groundbreaking moments was the completion of the Human Genome Project (HGP) [29-31], set up in 1990 with the aim of determining a full reference human genome. Thirteen years and approximately $2.7 billion later, the HGP completed the sequence of the 3 billion chemical base pairs that make up human DNA [31]. The International HapMap Consortium [32] was an important follow-up, aiming at determining the population diversity, frequency, and genomic location of common genetic variation and haplotypes in individuals of different ethnical origins.

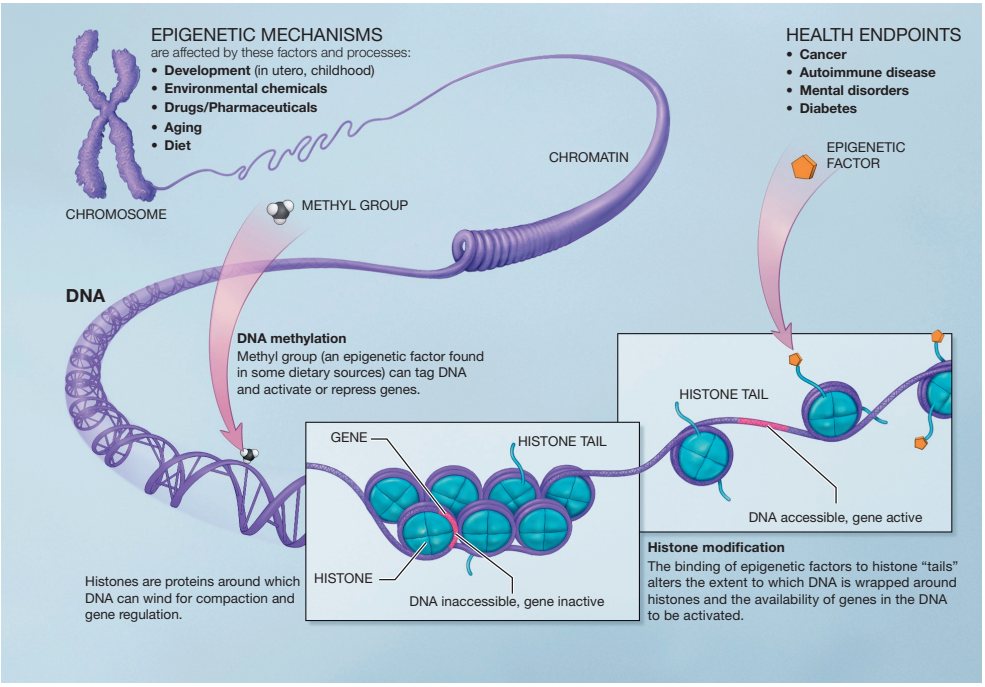


Figure 2. Epigenetic mechanisms, including DNA methylation and histone modification. Credit: National Institutes of Health via Wikimedia.

After the HapMap, the 1000 Genomes Project [33] sequenced 2,504 individuals from 26 populations. Its final phase catalogued over 88 million variants: almost 85 million SNPs (defined as having an allele frequency > 1%) or single nucleotide variants (SNVs, a one-base pair change without any limit on allele frequency), approximately 3.6 million insertions and deletions (indels), and 60,000 structural variants [33]. At the same time, the "SNP chip" technology was developed, based on the concept of microarrays. Companies such as Affymetrix (www.affymetrix.com) and Illumina (www.illumina.com) developed DNA microarrays [34, 35]. All these technological advances made it possible to test genetic variation at the genome-wide level for association with heritable complex traits and diseases [36], and so started the era of genome-wide association studies (GWAS).

A GWAS compares the frequencies of genetic variation between individuals with a given disease against control individuals from the same population. A strong association between a DNA position and a particular disease or trait marks the general location of the associated genetic locus, even if the associated SNP itself is not directly responsible for the disease. When the *p*-value for association is below the significance threshold, it is considered statistically significant. This threshold is adjusted by Bonferroni correction[37], avoiding type I error or false-positive result with a division of the significance threshold

(usually p=0.05) by the number of tests being performed, assuming that all tests are independent. Since it has been determined that the human genome contains approximately 1 million [38, 39] independent regions, the standard genome-wide significance threshold became $p<5 \times 10^{-8}$ (0.05/1,000,000).
The GWAS effort has identified to date 62,652 unique SNP-trait associations in 3,395 studies (as of June 25th, 2018) [40] (Figure 3).

The use of larger sample sizes has helped to identify an ever-increasing number of significantly associated variants, as demonstrated by the most recent study including over 1 million people that has identified 535 novel loci associated to BP traits [41], compared to just above 300 variants identified in all previous efforts. The next challenge is the interpretation of associations in order to get a better insight into the biology of disease, and possible translation of these statistical results to clinical practice.

Interpreting GWAS results: genetic risk scores and data integration

GWAS tend to identify genetic changes that have a small effect on disease risk but are common in the general population. Each single gene variant increases disease risk only slightly, underscoring the polygenic nature of complex diseases, but the accumulation of risk variants in several different genes may account for a greater risk. The combination of these small effects can therefore lead to significant risk assessment. This is the reasoning behind the concept of genetic risk scores (GRS).
Risk factors have a long history of use in medicine for decision-making in cardiovascular disease [42]. The Framingham Study developed a risk score as a framework for CVD risk assessment including major risk factors, such as age, BP, tobacco use, total cholesterol, high density lipoprotein cholesterol, and diabetes [43]. Metrics like the C statistic were further shown to improve prediction of CVDs beyond these traditional risk factors [44-46].
Genetic polymorphisms contributing to CVDs became integrated in these models, first focused on *a priori* selected candidate genes [47-49]. But because the reported contribution of any single genetic variation to CVD risk is small, a robust risk prediction from genetic markers was still lacking. As larger cohorts were used in GWAS, more robust and replicable associations were identified. Aggregating the risk effect of multiple loci into a single GRS was shown to improve risk prediction in CVD, beyond that afforded by traditional risk factors [49, 50]. The calculation of a GRS can thus add interpretation and application of GWAS data for clinical assessment of outcomes of interest, converting genetic data to a predictive measure of disease susceptibility[51].
A main challenge in GWAS interpretation is identifying the effect of associated SNPs. It is generally assumed that a SNP within a coding region affects that particular gene; however, the majority of SNPs fall in non-coding regions and many of them are intergenic

Figure 3. All catalogued SNP to trait association (p<5x10^{-8}) per chromosome [40], updated until June 25th, 2018.

[38, 52]. This makes it difficult to identify the causal variant inside each associated locus, as well as its target gene [53]. Any variant inside the same haplotype as the reported genome-wide significant SNP, said to be in high linkage disequilibrium (LD), can be causal [54]. Assessment of the causality of coding variants has become relatively straightforward in recent years with development of bioinformatics tools that predict the effect of amino acid changes, such as SIFT [55], PROVEAN [56], PolyPhen [57], CONDEL [58] and CADD [59]. The evaluation of non-coding variants mechanisms and target genes is much more challenging, due to our lack of fundamental knowledge in the non-coding regions of the genome. Based on data from the Encyclopedia of DNA Elements (ENCODE), only 1.2% of human DNA is estimated to encode protein-coding exons, while the majority of the genome is transcribed at some point in at least one cell type [60], influenced by regulatory elements located in non-coding regions of the genome. There is now vast evidence that non-coding variants associated with disease may interrupt the action of these regulatory elements, crucial in relevant tissues for that particular disease [61]. Investigating the influence of associated variants in regulatory landscape of critical cell-types is essential for understanding disease risk and biology.

Next-generation sequencing (NGS) technologies have made sequencing of DNA and RNA quicker and cheaper than the previous method of Sanger sequencing. This has resulted in the development of new methods that help greatly in the field of post-GWAS analyses. Chromatin Immunoprecipitation Sequencing (ChIP-seq), for example, combines chromatin immunoprecipitation (ChIP) with massively parallel DNA sequencing to identify the binding sites of DNA-associated proteins. It can be used to map global binding sites and regions of histone modifications highlighting regulatory elements [62, 63]. RNA sequencing (RNA-Seq), another method based on NGS, reveals the presence and quantity of RNA in a biological sample at a given moment [64, 65]. Maps of these regulatory annotations and expression of genes in several tissues have been made publically available by projects such as ENCODE [60], Roadmap Epigenomics [66] and GTEx [67, 68], enabling their integration with GWAS results.

Yet another challenge in post-GWAS analyses is the identification of genes influenced by non-coding associated variants, due to our yet limited knowledge on genome folding dynamics. SNPs may affect distant genes [69], at times up to 2 Megabase pairs (Mbp) away from the associated SNP, and these genes are not necessarily the closest genes to the SNP [70]. Effects on distant genes have been demonstrated when associated variants act on enhancers that may interact with far away genes [69, 71], sometimes even in other chromosomes [72, 73]. This occurs due to the 3D organization of the genome in a robust yet dynamic looping architecture, which facilitates fine-tuning of gene expression by mediating the contacts between distantly located *cis*-regulatory elements [74-76]. Chromosome conformation capture techniques (for instance 3C[77], 4C [78, 79] and Hi-C [80, 81]), capture chromosome interactions [82], resulting in networks of interacting genetic loci [78, 79]. Maps of chromosomal conformation have been made publically available and

incorporated into online tools [83, 84], allowing the scientific community to map GWAS variation to target genes [85].
Integration of different layers of information is crucial in interpretation of GWAS results and allows the prioritization of genes from known GWAS loci for functional follow-up, detecting novel gene-trait associations, inferring the directions of associations, and potential druggability[86-89].

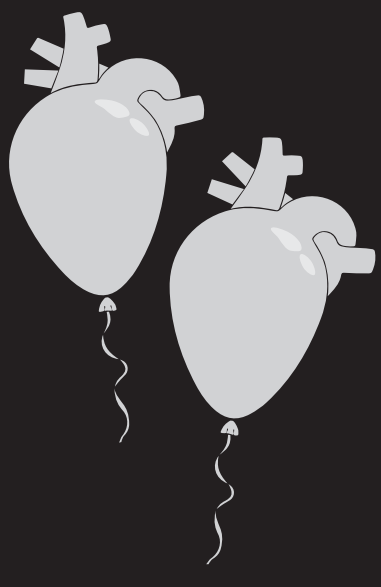
Applying GWAS results: translation into CVD prevention and treatment

In the scope of CVD therapeutics, the predictors identified over the last 30 years and described in the previous section are commonly used today to detect individuals at risk. These include hypertension [90], obesity, and diabetes, whose prevalence has increased together with lifestyle changes in low-income and middle-income countries – therefore resulting in higher rates of CVD [91]. Current recommendations for the treatment of CVD focus on reducing modifiable risk factors [92], including lifestyle changes (diet, exercise, smoking cessation), lipid lowering (e.g. statins) and blood pressure lowering medication (e.g. ACE inhibitors). Much research is still required to get better insight in the biological mechanisms leading to specific phenotypes, which can in turn lead to more efficacious and cost-effective preventive and therapeutic approaches. Currently, it is estimated that only one in every 5000 new drug compounds makes it to market [93]. Novel approaches are needed to support drug development, an expensive and difficult process, that may take 10-15 years and costs billions of dollars [94, 95]. These novel approaches have emerged in recent years based on genetic strategies. Bioinformatics analyses and follow-up of GWAS have unraveled the involvement of genes affected by GWAS SNPs in different pathophysiological pathways involved in CVDs, such as LDL cholesterol and triglyceride metabolism, blood pressure, vascular remodeling and inflammation [96]. These pathways contain possible candidate drug targets. A key challenge is to prioritize GWAS hits and their corresponding gene products for pharmacological intervention. In this process, bioinformatics methods may yield novel insights into the potential "druggability" [97] of each of these loci in order to translate genetic knowledge into clinical care.
Bioinformatics tools applied to druggability analyses on genetic data can also identify approved and already marketed drugs for treating a new phenotype other than the one the drug was originally developed for. This strategy is referred to as drug repositioning or repurposing, an approach proposed and improved in the past 15 years [98, 99], based on new discoveries including, more recently, genetic information [100, 101]. In such cases, where an existing drug targets a gene product or pathway of a disease different from the original indication, fewer clinical trials may need to be conducted to alter the label and indicate a treatment for another disease as safety has already been demonstrated.

An example of repurposing is sildenafil, initially produced with the expectation of reducing angina, and later found to be effective to treat erectile dysfunction [102] and pulmonary hypertension [103], leading to the subsequent releases of Viagra® in 1998 and Revatio® in 2005 [104]. Repurposing efforts based on GWAS results are described further in this thesis (Chapter 7).
Overall, it is expected that multiple GWAS loci, pointing to currently unexplored mechanisms, can be utilized in new drug development. This approach is exemplified in the development of antibodies that were successfully applied in the case of *PCSK9* [105-107]. Further research and clinical trials are still needed, however, for proper translation of statistical signals into therapeutic targets.

Outline of this thesis

The work in this thesis aims both at integrating genome, epigenome and transcriptome information in order to gain more insight into the mechanisms causing CVDs and inform future fine-mapping efforts that aim to understand the role of genetic variation in humans. In **Chapter 2** we focus on data integration for interpretation and potential translation of blood pressure-associated loci into drug targets. In **Chapter 3**, we perform functional annotation of 52 loci influencing myocardial mass. In **Chapter 4**, we make use of layers of information provided by NGS technology on the study of the regulation and expression profile of dilated cardiomyopathy. In **Chapter 5** we apply the concept of GRS to study the potential relevance of cardiovascular susceptibility loci in other vascular beds. In **Chapter 6** we investigate cardiovascular disease progression profiles, aided by GRSs. In **Chapter 7**, we describe a pipeline for identification of new drug targets, and another for potential repurposing of currently marketed ones. We conclude in **Chapter 8** with a discussion on how these studies provide a better insight into the biology of CVDs, the current challenges in translation of risk variants into clinical practice, and the opportunities that advances in the field will bring in the future.



Chapter 2

Integrative bioinformatics approaches for identification of drug targets in hypertension

Hemerich D, van Setten J, Tragante V, Asselbergs FW

2018. Front. Cardiovasc. Med.
DOI: 10.3389/fcvm.2018.00025

2.1 Abstract

High blood pressure or hypertension is an established risk factor for a myriad of cardiovascular diseases. Genome-wide association studies have successfully found over nine hundred loci that contribute to blood pressure. However, the mechanisms through which these loci contribute to disease are still relatively undetermined as less than 10% of hypertension-associated variants are located in coding regions. Phenotypic cell-type specificity analyses and expression quantitative trait loci show predominant vascular and cardiac tissue involvement for blood pressure-associated variants. Maps of chromosomal conformation and expression quantitative trait loci (eQTL) in critical tissues identified 2424 genes interacting with blood pressure-associated loci, of which 517 are druggable. Integrating genome, regulome and transcriptome information in relevant cell-types could help to functionally annotate blood pressure associated loci and identify drug targets.

2.2 Introduction

Elevated blood pressure (BP) or hypertension is a heritable chronic disorder [108-110], considered the single largest contributing risk factor in disease burden and premature mortality [111]. High systolic and/or diastolic BP reflects a higher risk of cardiovascular diseases [111]. Genome-wide association studies (GWAS) have found association of 905 loci to BP traits (systolic - SBP, diastolic - DBP and pulse pressure -PP) to date (Supplemental Table 1) [41, 112-139]. The use of larger sample sizes has helped to identify additional variants, as demonstrated by the most recent study including over 1 million people that has identified 535 novel BP loci [41]. Still, this collective effort thus far has not entirely elucidated the complete genetic contribution to BP, estimated to be approximately 50%–60% [140].

To add to this complexity, 90.7% of the 905 BP-associated index variants are located in intronic or intergenic regions (Supplemental Table 1). Causal variants are also difficult to pinpoint because of linkage disequilibrium (LD) [54]. There is now vast evidence that non-coding variants associated with disease interrupt the action of regulatory elements crucial in relevant tissues for that particular disease [61]. BP loci are not only linked to cardiovascular disease but also to other diseases (Figure 1), suggesting that BP-associated variants can result in a wide range of phenotypes. Tissue specificity of genetic loci may be relevant for organ specific disease progression. For example, variants altering expression in heart may more likely affect disease progression through heart-mediated processes rather than kidney-mediated processes, and some patients may suffer of left ventricular hypertrophy while others may develop nephropathy. Thus, investigating the influence of BP variants in critical cell-types is essential in understanding disease risk and biology, and assessing the possible translation of an associated locus into a drug target. The public availability of regulatory annotations in several tissues by projects such as ENCODE [60], Roadmap [66] and GTEx [67, 68] has enabled integration of epigenetic modifications, expression quantitative trait loci (eQTLs) and –omics information with GWAS data. Integrative approaches are useful for prioritizing genes from known GWAS loci for functional follow-up, detecting novel gene-trait associations, inferring the directions of associations, and potential druggability [86-89].

Here we summarize the advances made in recent years towards unraveling the mechanisms of non-coding BP variants in disease progression with the resources mentioned above. We focus on integrative approaches that aim to prioritize BP-associated SNPs located in regulatory regions of the genome for follow-up studies (Figure 2). Genetic and molecular aspects of hypertension have been reviewed previously by others [142, 143].

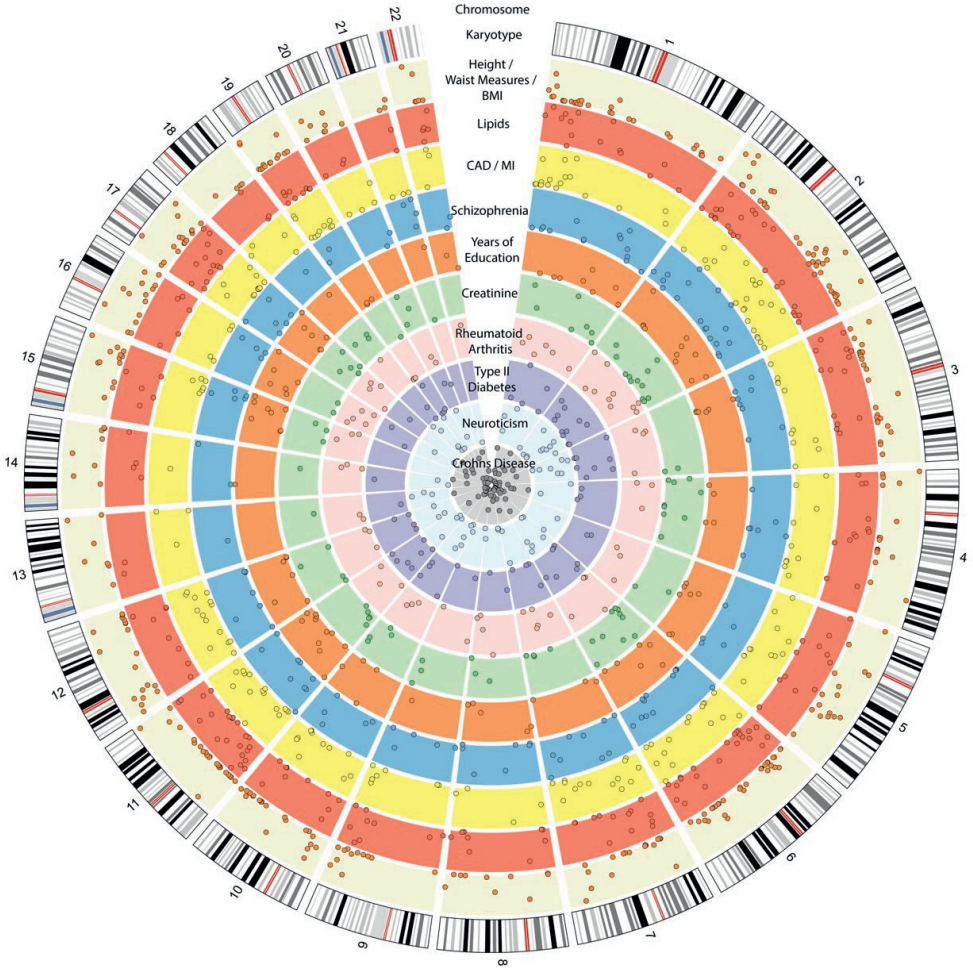


Figure 1. Circos plot showing the 10 traits from the GWAS catalog [52] with the largest number of loci also associated to BP, as identified by PhenoScanner [141] at $p<0.05$ (Supplemental Methods). The outer ring represents the genomic/chromosomal location (hg19). The following inner rings show the associations to different traits. Beige: body measurements (height, body mass index (BMI), weight, waist/hip ratio, hip circumference, waist circumference. N=358). Red: lipids (high-density lipoprotein (HDL), low-density lipoprotein (LDL), triglycerides, total cholesterol. N=226). Yellow: coronary artery disease (CAD)/myocardial infarction (MI) (N=206). Light blue: schizophrenia (N=135). Light orange: years of education attendance (N=101). Light green: creatinine (N=88). Light pink: rheumatoid arthritis (N=78). Light purple: type II diabetes (N=73). Light turquoise: neuroticism (N=69). Light grey: Crohn's disease (N=67).

2.3 Integrative approaches using –omics data

Remarkable advances have been made recently towards a better comprehension of BP genetics, the biology of disease and translation towards new therapeutics, boosted by the widespread application of high-throughput genotyping technologies. At the same

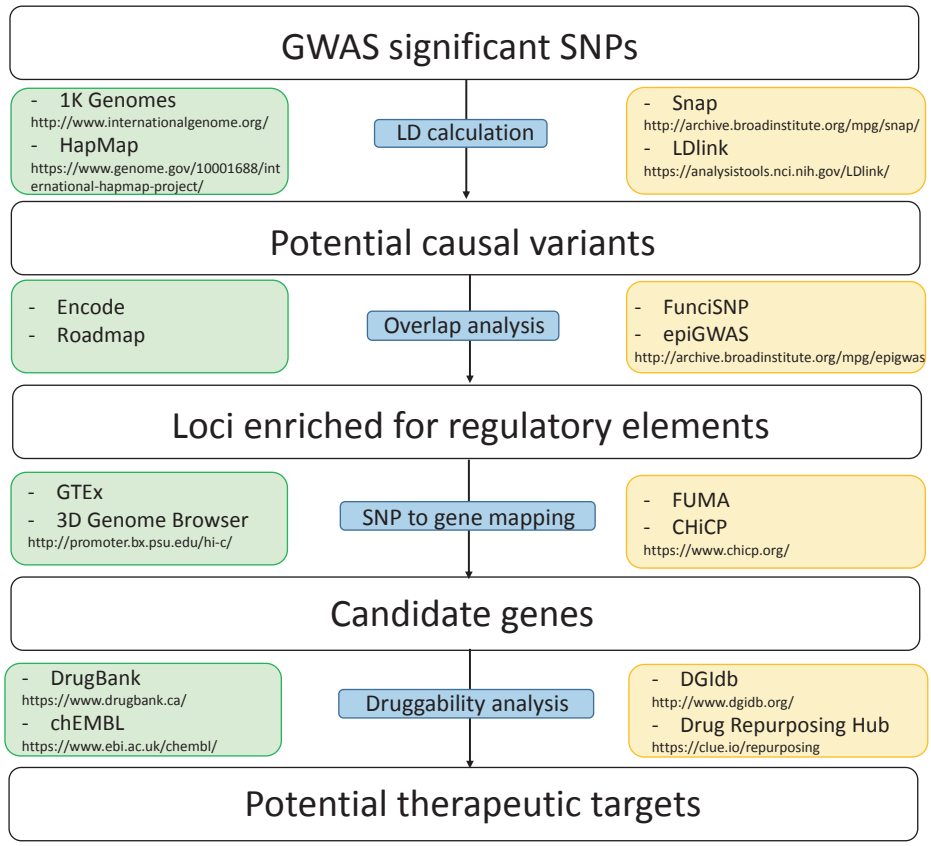


Figure 2. Diagram of analytical steps that can be followed for variant prioritization and translation of association to a potential drug target. Each step is accompanied by examples of publicly available data (green boxes on the left) and tools (yellow boxes on the right) that can be used.

time, most BP-associated variants are non-coding, making the conversion of statistical associations into target genes a great challenge. SIFT [55], PROVEAN [56], PolyPhen [57], CONDEL [58] and more recently CADD [59] are scoring algorithms developed for predicting the effect of amino acid changes. Only 98 out of the 905 lead BP-associated SNPs reflect a CADD score above 12.37 (Supplemental Table 5), a threshold suggested by Kicher *et al.* as deleterious [59]. However, the causal variant inside the locus might reflect a different CADD score than the lead SNP, and pinpointing the mechanisms disturbed by the variation remains a challenge. New strategies that make use of regulatory annotations in disease-relevant tissues have greatly expanded our ability to investigate the processes involved in BP. In particular, annotation of histone modifications and regions of open chromatin allow the identification of active transcription in specific-cell types. Similarly, maps of DNA variants affecting expression in a cell-type specific manner will be integral

in BP loci interpretation. A list of cardiovascular-related cell-types researched by the ENCODE Project is presented by Munroe *et al.* [144]. Such data can be integrated with GWAS results using bioinformatics tools [83, 145, 146]. For instance, FUMA provides extensive functional annotation for all SNPs in associated loci and annotates the identified genes in biological context [83]. FunciSNP investigates functional SNPs in regulatory regions of interest [146]. Ensemble's Variant Effect Predictor (VEP) determines the effect of variants on genes, transcripts, and protein sequence, as well as regulatory regions, also outputting SIFT, Polyphen and CADD scores for each variant, among other information [147]. Although such integrative tools are useful for variant prioritization and interpretation, not all take into consideration tissue specificity aspects. RegulomeDB, for example, is a database that annotates SNPs with known and predicted regulatory elements in the intergenic regions of the human genome, calculating a score that reflects its evidence for regulatory potential [148]. However, the scoring procedure can only be performed across all available tissue types. In addition, several databases containing a broad range of tissues were made publicly available since the last update of RegulomeDB, that could be included in the tool. Together, these resources have been useful in prioritizing genes and variants in associated loci for functional follow-up experiments in many post-GWAS analyses, and can be implemented in interpretation of BP-associated loci.

2.3.1 Transcription regulation: histone modifications and open chromatin

As genomic coordinates of active regulatory elements may be mapped using unique functions of chromatin, the characterization of chromatin changes in the genome in specific cell-types can be used to identify DNA variants disturbing active regulatory elements. The four core chromatin histones, H2A, H2B, H3, and H4, can suffer posttranslational modifications, such as acetylation or methylation [149]. These histone modifications indicate active (euchromatin) or repressed (heterochromatin) chromatin structure, defining regulation and gene transcription [150, 151]. Acetylation of histones H3 and H4, and H3 methylation at Lys4 (H3K4me3), for instance, correlate with gene transcription, whereas methylation at Lys9 correlates with gene silencing [150, 152]. These modifications provide a robust readout of active regulatory positions in the genome, and have been employed for annotation in several studies [130]. Histone modifications influencing arterial pressure have been observed in many tissues, including vascular smooth muscle [153]. An updated phenotypic cell-type specificity analysis of the 905 BP loci using H3K4me3 mark in 125 tissues is shown in Figure 3. The most significant cell-types are cardiovascular-related (Supplemental Methods, Supplemental Table 2). Other tissues with high rank in specificity are smooth muscle, fetal adrenal gland, embryonic kidney cells, CD34 and stem-cell derived CD56+ mesoderm cultured cells.

These results are consistent with analyses using DNase I hypersensitivity sites (DHSs), which indicate likely binding sites of transcription factors. Moreover, these results add

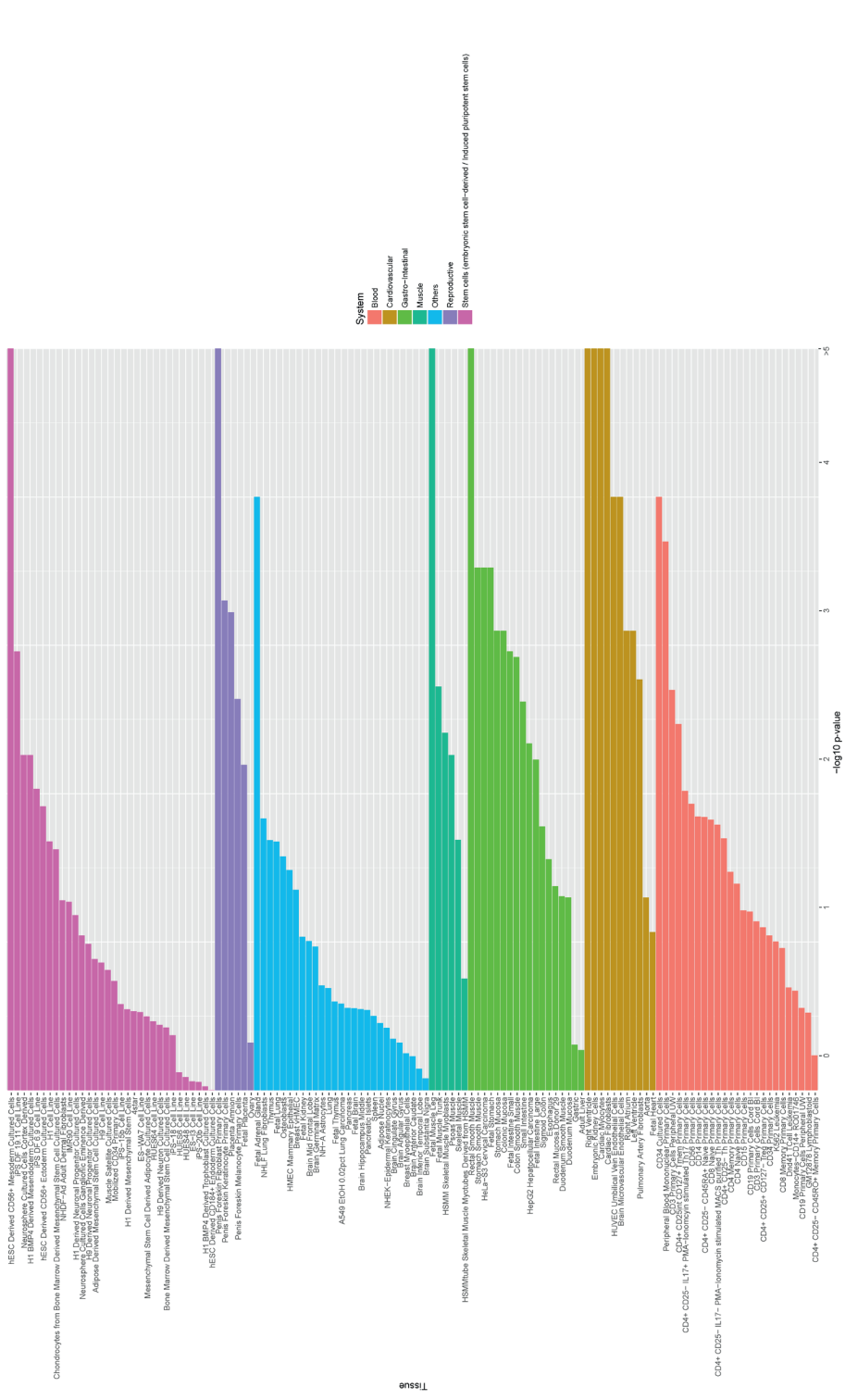


Figure 3. Ranked tissues after phenotypic cell-type specificity analysis of 905 BP SNPs using 125 H3K4me3 datasets on human tissue (Supplemental Methods, Supplemental Table 2).

more evidence that BP loci are enriched on regions of open chromatin [41, 126, 127, 130] (Supplemental Figure 1), regulating transcription in a broad range of tissues.

2.3.2 Methylation

In addition to histone modifications that promote transcription, BP loci have also been studied for their enrichment on DNA methylation, known to have the opposite regulatory effect. The methylation of CpG sites, presented by CpG islands in promoters, affects binding of transcription factors, resulting in gene silencing [154, 155]. Abnormal CpG methylation is found in hypertension [156, 157, 158], and in many other complex diseases [159-160]. Recently, Kato *et al.* identified a ~2-fold enrichment associating BP variants and local DNA methylation [126]. The study also demonstrates that DNA methylation in blood correlates with methylation in several other tissues. These observations add to previous indications on the function of DNA methylation in regulating BP.

2.3.3 Measuring the impact of BP risk alleles on gene expression: eQTLs

Expression quantitative trait loci (eQTL) are regions harbouring nucleotides correlating with alterations in gene expression [161]. Linking transcription levels to complex traits has been a follow-up step adopted by many studies [86, 162-164], driven by the increase in available data of expression patterns across tissues and populations [41, 89, 165-169]. Warren *et al.* found that 55.1% of their identified BP-associated loci have SNPs with eQTLs in at least one tissue from GTex repository [67], with arterial tissue most frequently observed (29.9% of loci had eQTL in aorta and/or tibial artery) [128]. A great enrichment of eQTLs in artery was also observed by Evangelou *et al.*, who identified 92 novel loci with eQTL enrichment in arterial tissue and 48 in adrenal tissue [41]. In summary, these studies also suggest that BP loci exert a regulatory effect mostly in vascular and cardiac tissues.

2.4 Finding the targets: chromosome confirmation capture techniques

Mapping variation to target genes is one of the greatest challenges in the post-GWAS era, and different strategies have been developed to this end [170]. One approach is the use of chromosome confirmation capture (3C [77], 4C [78, 79], Hi-C [80, 81]). These techniques capture chromosome interactions [82], resulting in networks of interacting genetic loci [78,79].
Warren *et al.* made use of this resource to investigate the target genes of non-coding SNPs, using Hi-C data from endothelial cells (HUVECs). Distal potential genes were found on 21 loci, and these genes were enriched for regulators of cardiac hypertrophy in pathway analysis [127]. Kraja *et al.* also explored long-range chromatin interactions using endothelial precursor cell Hi-C data [84, 171], finding the link between an associated loci and a gene known to affect cell growth and death [172]. More recently, Evangelou *et al.* used chromatin interaction Hi-C data from HUVECs [173], neural progenitor cells

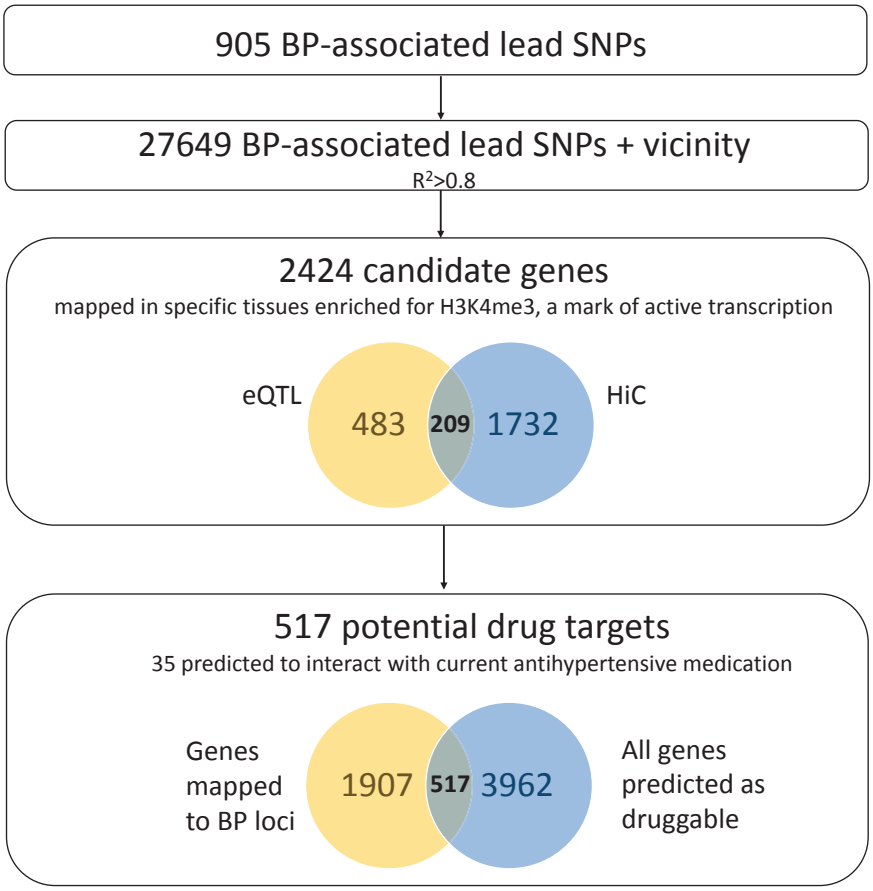


Figure 4. Diagram illustrating the results of our integrative approach.

(NPC), mesenchymal stem cells (MSC) and tissue from the aorta and adrenal gland [174] to identify distal affected genes. They found 498 novel loci that contained a potential regulatory SNP, and in 484 loci long-range interactions were found in at least one cell-type [41].

A list of human HiC data available on BP relevant tissues is presented in Supplemental Table 3. An updated version of variant to gene mapping making use of this chromatin conformation data is shown in Supplemental Table 4. Promoter regions of 1941 genes were found to interact with the 27649 candidate SNPs (905 BP associated SNPs and vicinity) (Supplemental Methods, Figure 4). Integration with eQTL data on relevant tissues confirmed 209 of the genes mapped, and added additional 483 genes. One main goal of understanding biological mechanisms of GWAS associations and affected genes is to be able to therapeutically target them. Assessment of the druggability of a BP-associated locus depends on several factors, but overlap of these results with a

recent effort on druggability suggests that 517 of these 2424 genes are druggable [175], and 35 mapped genes are also predicted to interact with common drugs for treatment of hypertension (Supplemental Table 5, Figure 4, Supplementary Methods). Interestingly, 1774 of the genes mapped are physically located outside BP-associated loci. These results support the hypothesis that BP GWAS loci act on tissue specific regulatory gene networks. Importantly, they also show that the use of long range chromatin interaction maps can reliably identify target genes even outside the risk locus.

2.5 Discussion

GWAS have pinpointed over 900 loci associated with BP, and increasing sample size has shown to be crucial to identify more signals [41]. However, efforts are needed to translate these results into biological inferences on causal mechanisms and understanding of disease biology. The integration of data beyond the DNA sequence is crucial to identify genes involved in BP regulated by epigenetic mechanisms.

BP variants show eQTL, histone modification and open chromatin enrichment in a broad range of tissues, mostly vascular and cardiac-related. As the interplay of regulatory elements is highly cell-type specific, the study of changes that influence chromatin structure and accessibility needs to be extended to a broad range of tissues and conditions, including disease and its stages. Rosa-Garrido *et al.* observed chromatin structural abnormalities when comparing healthy and diseased cardiac myocytes, concluding that heart failure involves altered enhancer-gene interactions [176]. Thus, alterations in chromatin structure underlying heart disease perturbs significant interactions that contribute to gene expression. This finding suggests that high resolution chromatin conformation and epigenetic data in disease state can help in understanding how regulatory variants confer risk to disease. The availability of data in different populations will also allow fine-mapping and functional annotation across ethnic groups.

By mapping of BP-associated variants to genes using maps of chromosomal conformation in specific cell-types, we identified 1941 genes, of which 209 show supported by eQTL mapping. Of all genes mapped (n=2424), 517 are predicted as druggable and 35 are predicted to interact with common antihypertensive drugs. These include successful cases such as *APOB* gene, predicted to be targeted by Ibersartan, an angiotensin II receptor antagonist used mainly for the treatment of hypertension [177]. Interestingly, in this analysis we were also able to identify *ABCC9* gene on both eQTL and HiC mapping, a gene that interacts with Minoxidil. Although originally developed as an antihypertensive vasodilator, side effects provided limitations and currently its main application occurs topically for treatment of hair loss [178, 179]. This highlights the several factors involved in druggability of a target and need for extensive validation and trials. With *in-silico* experimental evidence supporting a plausible mechanism for association, definitive

assignment of functions to putative cis-regulatory elements requires perturbation of these elements. Although the majority of associated variants add only modest effects on risk, more studies suggest combinations of SNPs are frequently necessary in order to explain these effects [180-182]. CRISPR–Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats) editing technology [183] permits targeted manipulation of epigenetic mechanisms linked to risk alleles [184]. Finally, genes that show consequent differential expression can be further validated *in vivo* with the use of animal models.

In summary, the integrative approaches presented in this review help understanding the underlying biology of GWAS loci by mapping SNPs to genes and determine cell and tissue-specificity. The increase in availability of regulatory data in a broad range of tissues and disease states will expand the possibilities for integration and interpretation of association results. Studies validating the genes prioritized may identify new drug targets, enabling more effective prevention and treatment of hypertension and its consequences.

2.6 Supplemental Material

2.6.1 Supplemental Methods

2.6.1.1 Association lookups with other diseases

We used PhenoScanner [141] database to evaluate cross-trait effects for the 905 blood pressure (BP) -associated sentinel SNPs including proxies in linkage disequilibrium (LD, $r^2 \geq 0.8$) with variants of the GWAS catalog [52] at $p < 0.05$.

2.6.1.2 Prediction of coding consequences

We used Variant Effect Predictor (VEP) from Ensemble [3] to predict the coding consequences of the canonical transcript of each lead BP-associated SNP, as well as its CADD score [4]. Predicted coding consequences other than "missense_variant" were renamed as "synonymous_variant".

2.6.1.3 Phenotypic cell-type specificity

We obtained publicly available data on H3K4me3 histone modification on 126 tissues from ENCODE [5] and NIH Roadmap Epigenomics Mapping Consortium [6] (Supplemental Table 1). We downloaded alignment files in hg-19 coordinates of both treatment and control. Duplicated, unmapped reads and reads with mapping quality less than 5 (-b -F 4 -q 5) were removed using Samtools v1.3 [7]. Peak calling was performed using MACS v2.1.0 [8] using the respective input samples, -gsize=hg -nomodel parameters, and estimated fragment sizes (-extsize) predicted by PhantomPeakQualTools v1.1 [9]. We used the method described by Trynka et al. [10] for phenotypic cell-type specificity analysis, in which a locus is defined by identifying SNPs in tight LD with each variant (r2 ≥ 0.8), using data from the 1000 Genomes Project [47]. Each variant is scored based on the distance and height of the nearest peak. If the physical distance to the nearest peak is more than 2.5 kb, then the score is set to 0 to obviate any confounding distal effects. The statistical significance of cell-type specificity is assessed by deriving a 95th percentile threshold based on the permutation of 10,000 matched sets of SNPs not associated with the phenotype.

We used FORGE [11] to investigate cell-type-specific enrichment (FDR, $p < 0.05$) within DNase I–hypersensitive sites in 125 cell samples from ENCODE project [5], using default parameters. Briefly, FORGE compares the frequency of query variants in different cell types with a reference set of 1204 control variants from the GWAS catalog [2] with discovery $p < 5 \times 10–8$ in European ancestry populations. For each cell-type and p-value threshold, the enrichment of query variants mapping to footprints is expressed as a p-value derived from a logistic mixed effect model.

2.6.1.4 Variant to gene mapping through chromosomal conformation

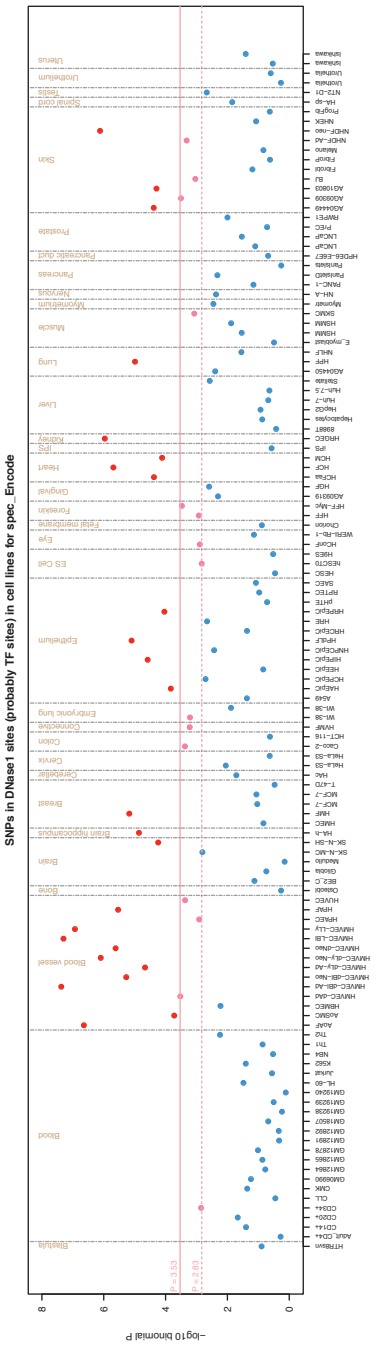
We used FUMA v1.3.0 [12] to map the 905 BP-associated sentinel SNPs and proxies in LD (r2 ≥ 0.8) to genes in relevant cell-types. We used maps of chromosomal conformation (HiC) in HUVEC [13], adrenal gland, aorta, left ventricle, right ventricle [14] and eQTL from GTEx v7 in adrenal gland, aorta, artery coronary, artery tibial, atrial appendage and left ventricle [15], keeping default parameters.

2.6.1.5 Interaction of mapped genes and antihypertensive drug targets

In order to assess the overlap between drug targets and BP loci, we obtained a list of medications used to treat hypertension at WebMD [16] and checked for their predicted interactions with gene products using DGIdb [17, 18]. We used the genes identified by our integrative mapping as representatives of each significant SNP to find matches with the output of DGIdb. A list of genes predicted as druggable was retrieved from the study by Finan et al. (2017) [19].

2.6.2 Supplemental Figures and Tables

2.6.2.1 Supplemental Figures



Supplemental Figure 1. Phenotypic cell-type specificity analysis results using DNase 1, showing enrichment predominantly on cardiovascular tissues.

2.6.2.2 Supplemental Tables

Supplemental Table 1.
Summary information on the 905 BP-associated variants identified to date (available on Hemerich *et al.* [85]).

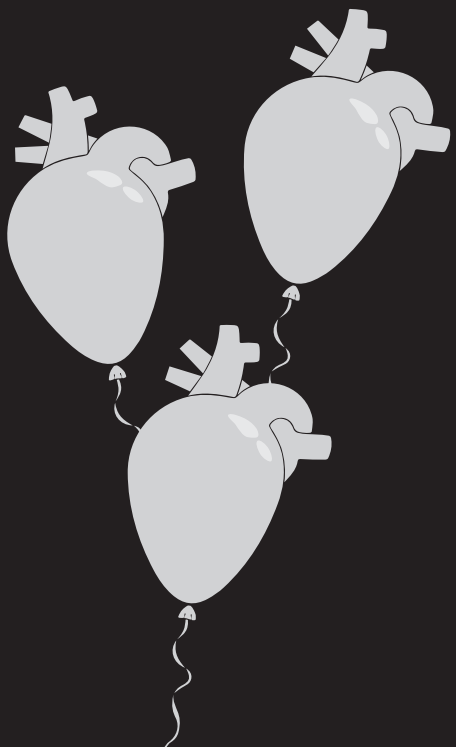
Supplemental Table 2.
Results of the integrative approach. CADD score and coding consequences predicted for the canonical transcript of each lead BP-associated SNP. "Gene HiC" and "Gene eQTL" columns show genes mapped by HiC and/or eQTL in relevant tissues. Genes identified by both resources are highlighted in red. Column "Gene predicted as druggable" shows which mapped gene is predicted to be druggable according to Finan *et al.* (2017). Last column shows genes predicted to interact with current antihypertensive medicines (available on Hemerich *et al.* [85]).

Supplemental Table 3.
Phenotypic cell-type specificity analysis results using H3K4me3, showing enrichment predominantly on cardiovascular tissues (available on Hemerich *et al.* [85]).

Supplemental Table 4.
Tools for integration and visualization of publically available datasets of chromosomal conformation experiments performed on tissues relevant for BP.

Tool	URL	Adrenal Gland	Aorta	Endothelial Precursor Cell	HUVEC	LV	RV
3D Genome Browser	http://promoter.bx.psu.edu/hi-c/				x	x	x
Capture HiC Plotter (CHiCP)	https://www.chicp.org/			x			
FUMA	http://fuma.ctglab.nl/	x	x			x	x

Supplemental Table 5.
Protein-coding genes mapped to the 905 BP-associated loci using maps of chromosomal conformation (HiC) in relevant cell-types (HUVEC, adrenal gland, aorta, endothelial precursor cells, LV=left ventricle, RV=right ventricle) and eQTL (in adrenal gland, aorta, artery coronary, artery tibial, atrial appendage, LV). Positions given in GRCh37 (hg19) genomic coordinates (available on Hemerich *et al.* [85]).



Chapter 3

Integrative functional annotation of 52 genetic loci influencing myocardial mass identifies candidate causal regulatory variants and target genes

Hemerich D, Pei J, Harakalova M, van Setten J, Boymans S, Boukens BJ, Efimov IR, Michels M, van der Velden J, Vink A, Cheng C, van der Harst P, Moore JH, Mokry M, Tragante V, Asselbergs FW

Submitted for publication

3.1 Abstract

Regulatory elements may be involved in the mechanisms by which 52 loci influence myocardial mass, reflected by abnormal amplitude and duration of the QRS complex on the electrocardiogram (ECG). Functional annotation thus far did not take into account how these elements are affected in disease context. We hypothesize that differentially active regulatory elements between hypertrophic cardiomyopathy (HCM) patients and healthy controls can offer more insights into the mechanisms of QRS-associated loci than elements not affected by disease. We generated maps of regulatory elements on HCM and non-diseased hearts. We tested enrichment of QRS-associated loci on elements differentially acetylated and/or directly regulating differentially expressed genes between HCM patients and controls. We further performed functional annotation on QRS-associated loci using these maps of differentially active regulatory elements. Regions differentially affected in disease showed a stronger enrichment (p=8.6x10^{-5}) for QRS-associated variants than those not showing differential activity (p=0.01). Promoters of genes differentially regulated between HCM patients and controls showed more enrichment (p=0.001) than differentially acetylated enhancers (p=0.8) and super-enhancers (p=0.025). We also identified 74 potential causal variants overlapping these differential regulatory elements. Twenty-seven of the genes mapped confirmed previous findings, now also pinpointing the potentially affected regulatory elements and candidate causal variants. Fourteen new genes were also mapped. These results provide insights into how regulatory QRS-associated variants account risk for increased myocardial mass.

3.2 Introduction

The QRS complex on the electrocardiogram (ECG) represents cardiac depolarization and conduction of the electrical signal through the ventricular muscle. Duration and amplitude of the QRS complex is used as a proxy for left ventricular mass [185, 186]. Abnormalities of the QRS complex are associated with an increased risk of cardiovascular (CV) mortality and morbidity [187-190]. A recent large-scale genome-wide association study (GWAS) meta-analysis of four correlated and clinically used QRS traits (Sokolow-Lyon, Cornell, 12-lead-voltage duration products (12-leadsum), and QRS duration) identified 52 independent loci at $p<1x10^{-8}$ [191]. However, the identification of causal variants, their target genes and disturbed mechanisms remain an important challenge. In a given GWAS locus, the SNP with the most significant association with the disease (lowest *p*-value) is usually reported as the 'lead' SNP. This lead SNP is not necessarily the causal variant, and a SNP in high linkage disequilibrium (LD) with the lead SNP may be the causal one [192]. Only three of the 52 SNPs reported in the original study are located in coding regions of the genome [191]. A non-synonymous variant in high LD with a high deleteriousness score, such as CADD score > 12.37 [59], can also be considered candidate for causality, and that is the case of 14 variants in nine QRS-associated loci (Supplemental Table 1). Potential mechanisms for causality in the remaining loci include disturbance of regulatory elements, genomic regions that play a crucial role in transcriptional regulation. Indeed, increasing evidence shows multiple GWAS variants regulate transcription [66, 193, 194]. Non-coding variants often affect gene expression in a cell-type specific manner by altering the function of enhancer and promoter elements [145, 195-197]. Results from *in silico* analyses have suggested strong enrichment of QRS-associated variants in specific chromatin states associated with active enhancers, promoters, and transcription in the human heart [66]. By contrast, no enrichment was observed for transcriptionally repressive histone marks [139, 191]. Although using relevant cell-types, these previous studies did not include regulatory information obtained from tissue in disease state. Highlighting possible differences in transcription from healthy to diseased myocardial tissue may help understand the underlying mechanisms of changes in left ventricular mass. To gain insight into these mechanisms, we integrate genome, regulome and transcriptome information to test whether regions differentially affected in disease can be more informative than the regulatory landscape of non-diseased tissue. We use regulatory information obtained from non-diseased myocardium and diseased tissue from patients with hypertrophic cardiomyopathy (HCM). We also co-localize candidate causative QRS-associated variants and regulatory regions that show differential activity (proxied by H3K27ac levels or by gene expression) in HCM tissues compared to controls, in order to identify regulatory elements potentially altered by the variation. Taken together, our results may enhance our understanding on the regulatory mechanisms underlying increased myocardial mass.

3.3 Methods

3.3.1 Human Material

The procedures for obtaining human samples were approved by the scientific advisory board of the biobank of the University Medical Center Utrecht (protocol number 12/387), the Washington University School of Medicine Ethics Committee (Institutional Review Board) and the local ethics committee of the Erasmus MC, and written consent was obtained. Biopsies on HCM patients are septal myomectomy specimens. Control samples were obtained from donor hearts not used for transplantation.

3.3.2 RNA sequencing

We performed RNA-seq on twenty-two biological replicates of human myocardial tissue, eleven of which on diseased (HCM) and eleven in non-diseased state (Supplemental Table 2). RNA was isolated using Qiagen AllPrep Micro Kit according to the manufacturer's instructions. After the selection of mRNA, libraries were prepared using the NEXTflex™ Rapid RNA-seq Kit (Bio Scientific). Libraries were sequenced on the Nextseq500 platform (Illumina), producing single end reads of 75bp. Reads were aligned to the human reference genome GRCh37 using STAR v2.4.2a [198]. Picard's AddOrReplaceReadGroups v1.98 (http://broadinstitute.github.io/picard/) was used to add read groups to the BAM files, which were sorted with Sambamba v0.4.5 [199] and transcript abundances were quantified with HTSeq-count v0.6.1p1 [200] using the union mode. Subsequently, reads per kilobase million reads sequenced (RPKMs) were calculated with edgeR's RPKM function [201]. Genes with low counts (whose sum of all counts across samples included in the analysis was < 10) were removed. In order to obtain a list of genes expressed in either HCM or control group, we categorized transcripts that are expressed as having >0.5 RPKM averaged across samples of each group (expected reads per kilobase of transcripts per million fragments sequenced). In order to obtain a list of differentially expressed genes between HCM and controls at FDR<0.05, we employed Deseq2 v1.10.1 package [202]. We calculated *p*-values using Wald statistics and corrected for multiple testing using the Benjamini-Hochberg method. We retrieved transcription start sites (TSS) of the obtained genes from Ensembl Genes 89 using Biomart (Human genes GRCh37.13) [203]. We defined promoter regions as ranging from -2500 to +2500 base-pairs (bp) from the TSS. Differentially regulated genes can be found in Supplemental Table 3.

3.3.3 Chromatin Immunoprecipitation and Sequencing

Chromatin immunoprecipitation and sequencing (ChIP-seq) using H3K27ac mark was performed on human myocardial samples from 8 HCM patients and 8 healthy controls were included in the study, partially matching the samples included in the RNA-seq experiment described above (Supplemental Table 2). Chromatin was isolated

from each sample using the MAGnify™ Chromatin Immunoprecipitation System kit (Life Technologies) according to the manufacturer's instructions. Immunoprecipitations were performed with antibody H3K27ac (ab4729, Abcam) for ChIP-seq as described previously [204]. ChIP DNA Clean & Concentrator kit (Zymo Research) was used to purify captured DNA fragments. Libraries were prepared using the NEXTflex™ Rapid DNA Sequencing Kit (Bioo Scientific) and sequenced on Illumina NextSeq500 sequencer. Alignment to the human reference genome (hg19) was performed using BWA v0.7.5a [205] . Duplicated, unmapped reads and reads with mapping quality less than 5 were removed using Samtools v1.3 [206]. Peak calling was performed using MACS v 2.1.0 [207] using the respective input samples, -gsize=hg -nomodel parameters, and estimated fragment sizes (-extsize) predicted by PhantomPeakQualTools v1.1 [208]. Epigenomic profiles were compared using DiffBind v2.2.12 [209] to identify enhancer peaks that are quantitatively different (FDR<0.05, Deseq2 algorithm) between healthy and HCM populations. Consensus peaksets were formed by peaks that overlapped in at least two samples. Differentially acetylated regions can be found in Supplemental Table 4.

3.3.4 Super-enhancer identification

We identified super-enhancer regions on the 8 HCM H3K27ac datasets described previously by employing ROSE algorithm [210] using a stitching distance of 12.5kb. Peaks fully contained in the region spanning 2500bp upstream and downstream of an annotated TSS were excluded. HCM-specific super-enhancers were defined as super-enhancer in HCM not overlapping super-enhancers in the control septum samples. These can be found in Supplemental Table 5. BEDtools v2.17.0 [211] and BEDOPS v2.4.35 [212] were used for manipulation of bed-files. Super-enhancer regions from non-diseased human left ventricle (LV) were retrieved from the Super-Enhancer Archive (SEA) [213].

3.3.5 Regions of Open Chromatin

All datasets of regulatory features generated in this study were further narrowed down to regions overlapping maps of open chromatin in heart, in order to retain only the sequences accessible to transcription factors (TF). We retrieved peaks from the "open chromatin" tracks from ENCODE portal in heart tissue. These datasets were generated using DNaseI hypersensitivity or Formaldehyde-Assisted Isolation of Regulatory Elements (FAIRE) assays. For the DNaseI assays, 3 biological replicates from 3 non-diseased adult hearts were included. For the FAIRE assays, 2 biological replicates from 2 non-diseased adult hearts were included. A peaks file with GEO accession GSM1008559, identified as Heart_OC, was downloaded from http://genome.ucsc.edu/cgi-bin/hgFileUi?db=hg19&g=wgEncodeOpenChromDnase .

3.3.6 Integrative analysis and statistical enrichment of QRS-associated variants on specific regulatory regions

We used FUMA [83] to expand the set of 52 lead SNPs to candidate SNPs in LD, using the following parameters: *p*-value of lead SNPs <1x10^{-8}, reference panel population 1000G Phase3 EUR, minor allele frequency ≥ 0.01, maximum distance between LD blocks to merge into a locus < 1000kb. We generated sets of LD variants in the following r^2 windows: 0.05 =< r^2 > 0.1, 0.1 =< r^2 > 0.2, 0.2 =< r^2 > 0.3, 0.3 =< r^2 > 0.4, 0.4 =< r^2 > 0.5, 0.5 =< r^2 > 0.6, 0.6 <= r^2 > 0.7, 0.7 <= r^2 > 0.8, 0.8 <= r^2 > 0.9 and 0.9 <= r^2 >= 1. We used bedtools v2.19.1 [211] to identify candidate variants from the generated sets that overlap with differential regulatory regions of interest, including HCM-specific super-enhancers, LV-specific super-enhancers, regions differentially acetylated between HCM patients versus controls, and promoter regions of genes differentially expressed between HCM patients and controls. A SNP falling within coordinates of regulatory regions of interest was considered as overlapping SNP. To assess the significance and enrichment of the amount of overlaps found, we generated background distribution for the number of overlaps by sampling SNPs from the total set of HapMap phase 2 CEU SNPs [214] with similar minor allele frequency and distance to gene. We carried out permutation tests using overlaps with all promoters/enhancer elements identified in HCM and control hearts, and with regions resulting from differential analysis between HCM patients and controls.

3.3.6 Variant to gene mapping

We used FUMA [83] for eQTL mapping, linking SNPs to genes whose expression is likely affected up to 1 Mb (*cis*-eQTL). Since eQTLs are highly tissue specific, LV was selected from GTEx portal. We defined significant eQTLs as FDR ≤ 0.05. The gene FDR is pre-calculated by GTEx and every gene-tissue pair has a defined p-value threshold for eQTLs based on permutation [83]. eQTL maximum *p*-value was defined as ≤ 1x10^{-3}. In addition, we identified the three closest differentially expressed genes to QRS-candidate causative variants overlapping differential regulatory elements.

3.4 Results

In this study, we functionally annotated and fine-mapped 52 QRS-associated loci. We used regulatory information (promoters from RNA-seq experiments, H3K27ac regions highlighting active promoters and enhancers, and super-enhancers from ChIP-seq experiments) obtained from non-diseased myocardium and diseased tissue from HCM patients. Differential expression analysis between 11 HCM patients and 11 controls resulted in 1557 up-regulated and 1202 down-regulated genes (Supplemental Table 3). Differential acetylation analysis between 14 HCM patients and 4 controls resulted

in 4068 up-acetylated and 2983 down-acetylated regions (Supplemental Table 4). From the total set of super-enhancers identified, 1048 were unique to the HCM group, not overlapping with the control group (Supplemental Table 5). Promoter, enhancer and super-enhancer regions were further narrowed down to those overlapping regions of open chromatin, in order to retain only sequences accessible to TF binding. We used these regulatory features to test the enrichment of QRS-associated variants and vicinity in LD. In order to identify an optimal LD cutoff, we identified variants in LD with the 52 QRS-associated SNPs, and divided them into bins, as described in the Methods section. We overlapped each bin with regulatory elements differentially regulated/acetylated in HCM, and performed 10k permutation tests to calculate enrichment (Supplemental Figure 1). Although LD thresholds showed fluctuations on enrichment, we observed a continuous decrease in enrichment as the LD threshold becomes more lenient. We defined LD r^2>0.5 as cutoff to expand the set of candidate causal QRS-associated variants, from 52 lead SNPs to a total of 4620. We co-localized these candidate causal QRS-associated variants and regulatory regions that show differential activity in disease, in order to identify regulatory elements potentially altered by the variation. We also investigated which set of regulatory features showed more enrichment for QRS-associated variants, and thus has the potential to be more informative for fine-mapping efforts.

3.4.1 Disease-affected regulatory regions show more enrichment for QRS-associated variants than all tissue-specific regulatory regions

We performed enrichment tests aiming to identify which set of regulatory regions is more informative for fine-mapping efforts. Regulatory features differentially regulated/acetylated in HCM were more enriched (p=8.6x10^{-5}) than those not showing differential activity (p=0.01), suggesting these can offer more insight into mechanisms altered by genetic variation than general tissue-specific regulatory elements (Figure 1A).

3.4.2 Differentially expressed promoters show more enrichment than regions highlighted by differential acetylation

Given the strong enrichment of QRS-associated variants in differentially regulated promoters and differentially acetylated regulatory elements in HCM (Figure 1A), we overlapped candidate causative SNPs and enriched regulatory features to identify variants that might be altering the function of these elements. Of the 4620 candidate QRS-associated SNPs, 74 co-localized with differential regulatory features (Figures 1 and 2, Supplemental Table 1), more than expected by chance (p=8.6e-5). These variants show more enrichment in differentially expressed promoters from RNA-seq experiments (p=0.001) than differentially acetylated regulatory elements highlighted by H3K27 from ChIP-seq experiments (enhancers p=0.8, super-enhancers p=0.02) (Figure 1B). These results highlight the potential of integrating regulatory elements that show differential behavior in disease, especially promoter regions.

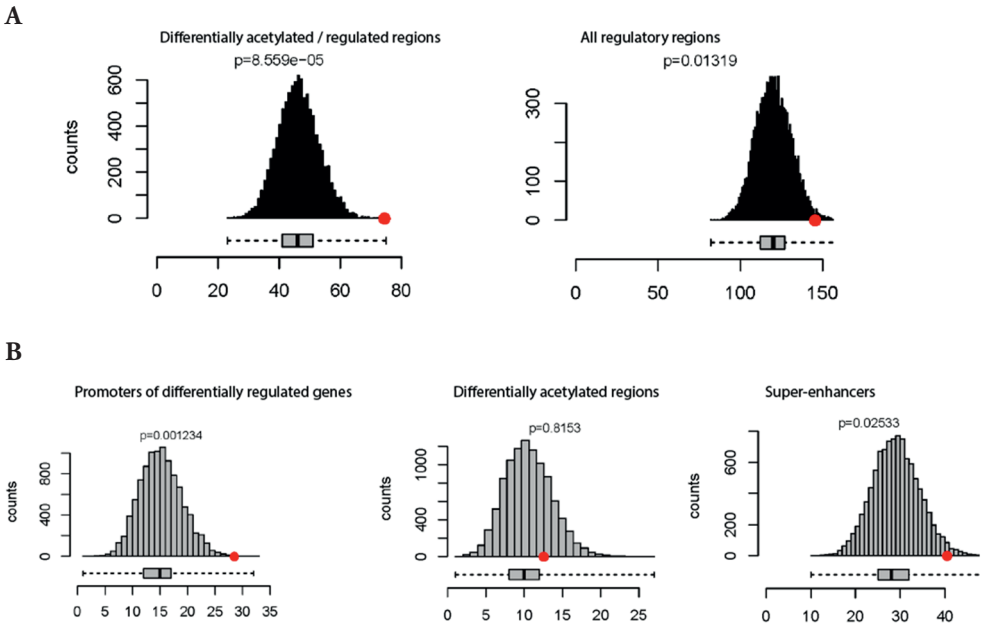


Figure 1. A) Mean number of regulatory elements overlapping with QRS-associated loci (red circle) compared with 10,000 matched control sets (gray bars). Differentially acetylated / regulated elements in HCM show more enrichment (p=8.6x10^{-5}) for QRS-associated candidate causal variants than those not showing differential activity (p=0.01). **B)** Promoters of genes differentially regulated between HCM patients and controls showed more enrichment (p=0.001) for QRS-associated variants than differentially acetylated enhancers (p=0.8) and super-enhancers (p=0.025).

3.4.3 Fine-mapping pinpoints potential causal variants and candidate genes

The 74 variants overlapping regulatory elements affected in HCM are spread through 20 QRS-associated loci. We investigated which genes are the potential targets of the 74 QRS-associated variants. We retrieved the three genes nearest to each of the 74 variants. We found 13 of the nearest genes are down-regulated and 12 are up-regulated in HCM versus controls (Supplemental Table 1). eQTL mapping with LV tissue from GTEx portal [89] confirmed the involvement of these differentially regulated genes with QRS-associated candidate regulatory variants in three loci (*ACP2/MADD, FADS1/FADS2, PROCR/EDEM2*). Eighteen of the 68 candidate genes that were identified by the original study were confirmed with this new approach [191]. The functional annotation performed in this study pinpoints the likely regulatory elements affected by the variation that in turn affect expression of these genes, as well as candidate causative SNPs (Supplemental Tables 1 and 7). The function of the new fourteen candidate genes potentially involved in myocardial mass is described in Supplemental Table 6. Visualization and further description of each of the 20 QRS-associated loci overlapping regulatory regions of interest can be found on the Supplemental Material.

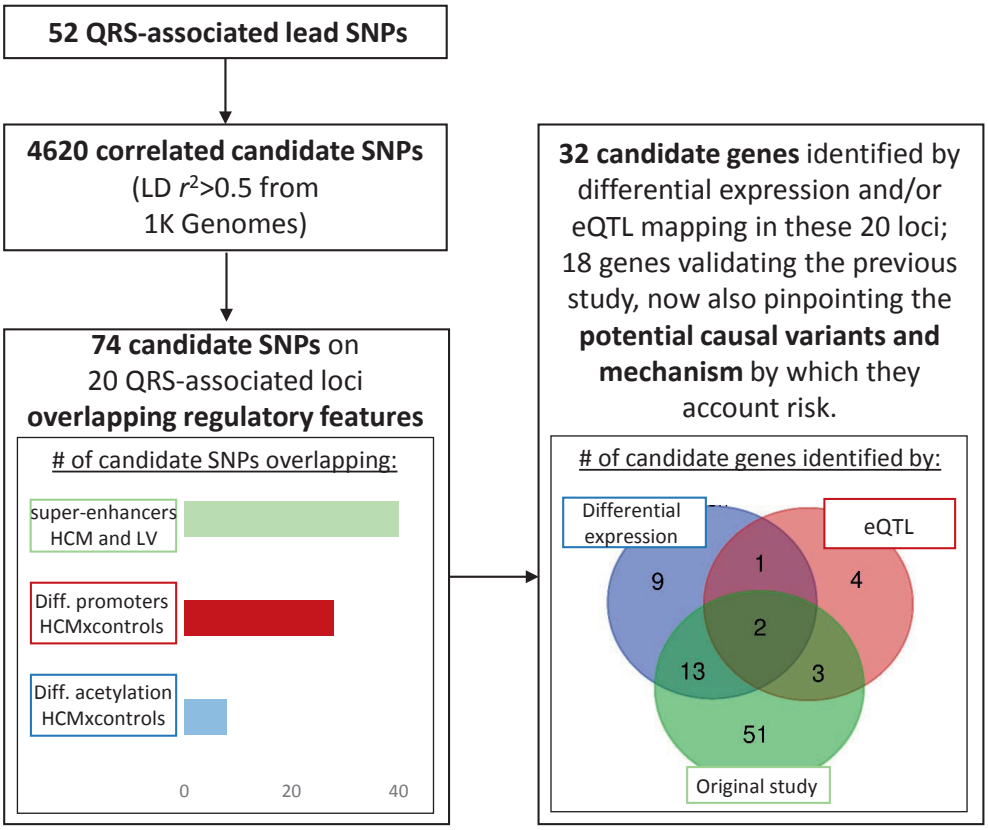


Figure 2. Overview of functional annotation pipeline. We expanded the search from 52 lead SNPs to include variants in LD (r^2>0.5), leading to 4620 candidate variants. Of these, 74 overlapped one or more regulatory feature of interest: differentially acetylated (H3K27ac) regions between HCM and control patients, differentially expressed promoter regions between HCM and control patients, and super-enhancers specific to HCM or LV. In the 20 loci where these overlaps were observed, differential expression analysis on HCM patients and controls and or/eQTL mapping in healthy LV identified 32 candidate genes. Twenty-seven of these validate the findings of the previous study, now also pinpointing causal variants and the potential mechanisms by which they account risk. In addition, 14 new candidate genes were identified, some with cardiac function and previously implicated in other studies on ECG traits (Supplemental Tables 1, 6 and 7).

3.5 Discussion

We performed functional annotation of 52 loci influencing myocardial mass, considering regulatory elements with differential activity between diseased (HCM) heart samples and healthy controls may act as mediators between genetic variation and gene expression. The 52 QRS-associated variants showed enrichment on the set of differential regulatory elements in comparison to non-differential regulatory elements. In addition, differentially expressed promoters identified by RNA-seq in HCM and controls showed

more enrichment of QRS-associated variants, than the other differential regulatory elements analyzed. These results can aid the study design of future fine-mapping studies. Our functional annotation showed twenty loci showed overlapping with the enriched differential regulatory regions. We identified the variants in LD with the lead QRS-associated SNP overlapping regulatory features, thus pinpointing the potential causative variants and mechanisms through which these variants account risk for disease. We observed that some loci showed enrichment of more than one candidate SNP in the same regulatory feature. Studies indeed show that several variants can be causal inside a disease-associated locus [180, 181]. In addition, more than one gene can be causal inside the same loci [215], and we indeed identified more than one candidate gene per locus in some cases. Further assays, such as chromatin conformation capture [77-80, 171], can add evidence on the interaction between QRS-candidate regulatory SNPs and each candidate gene, preferably also taking into account the disease context [176]. These experiments can also reveal whether regions identified as promoters also function as enhancers [216-218]. Moreover, genes potentially influenced by the regulatory elements affected by QRS-associated variants that are not in the immediate vicinity can also be detected [219].

We observed that 18 genes mapped by the previous study were confirmed by differential expression and/or eQTL mapping, adding evidence for the role of these genes in natural variance in myocardial mass. We also identified 14 genes not previously mapped to the QRS-associated variants, some with known cardiac function and previously implicated in studies on ECG parameters and/or heart phenotypes. Definitive assignment of function(s) to putative cis-regulatory elements requires perturbation of these elements [220]. A promising approach is deletion and manipulation of nucleases through CRISPR/Cas9, which showed to be successful in recent studies [195, 221].

Future efforts can unravel the mechanisms of the remaining QRS-associated loci by also taking into account other regulatory elements apart from enhancers and promoters, such as silencers, highlighted by different histone modifications. In addition, other histone modifications can also help identify additional enhancer elements. A subset of active enhancers has been identified in mouse embryonic stem cells that lack H3K27ac but are marked by H3K122ac and/or H3K64ac, H3K4me1 and enhancer RNAs (eRNAs) [222]. This evidence shows that although H3K27ac highlights enhancer activity, enhancers that lack H3K27ac are not necessarily inactive, and other histone marks can carry out the activating role of H3K27ac [184]. Moreover, as chromatin states are spatially and temporally dynamic along disease progression, regulatory information is needed in more different conditions including different stages of disease progression. This will allow a more precise and complete fine-mapping and functional annotation. It can also help answer whether QRS-associated SNPs promote aberrant function of regulatory elements, or the SNPs alter the activity of regulatory elements following disease establishment.

Taken together, our results show the importance of using differential regulatory elements in disease tissue on fine-mapping studies, since influence from cell conditions are crucial for epigenetic modifications, chromatin accessibility, TF binding and consequently gene regulation [223]. We also expand the view of the epigenetic and regulatory aspects involved in myocardial mass by pinpointing potential causal variants, mechanisms potentially affected by genetic variation, and candidate genes. Validation of the identified candidate causal variants through methods such as genome-editing technologies will broaden our knowledge on how QRS-associated variants contribute to disease.

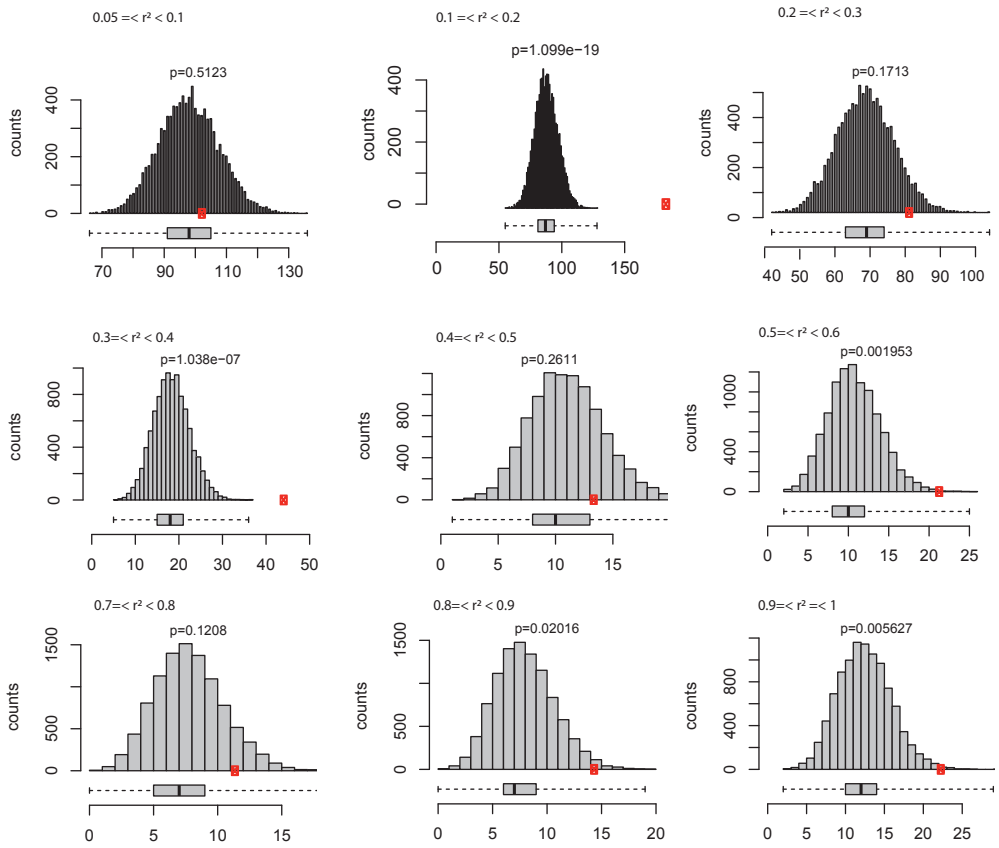
3.6 Supplemental Material

3.6.1 Supplemental Methods

3.6.1.1 Visualization of overlaps

All 4620 QRS-associated candidate SNPs (52 lead SNPs and variants in LD r^2>0.5) were plotted on UCSC Genome Browser on Human GRCh37/hg19 assembly [224]. The following tracks were included in their full form (not filtered by open chromatin regions): down-regulated promoters in HCM versus controls, up-regulated promoters in HCM versus controls, down-acetylated regions in HCM versus controls, up-acetylated regions in HCM versus controls, super-enhancers specific to HCM (not overlapping super-enhancers found in the control group), LV super-enhancers from the Super-Enhancer Archive [213]. Finally, the heart DNase track from ENCODE [60], showing the regions to which differential regulatory elements were further narrowed down as regions of transcription factor binding. The 74 variants overlapping regulatory elements affected in HCM overlapping open chromatin in heart and spread through 20 QRS-associated loci can be seen on Supplemental Figures 2 to 21. Resolution is set to 10kb or 20kb (information present in each figure). UCSC genes, a layered mark of H3K27ac on 7 cell-lines and a transcription factor track from ENCODE are also shown.

3.6.2 Supplemental Figures



Supplemental Figure 1.
Mean number of differential regulatory elements overlapping with QRS-associated loci (red circle) belonging to 9 windows of LD r^2 values, compared with 10,000 matched control sets (gray bars).

Supplemental Figures 2 to 21. Loci of lead QRS-associated SNPs overlapping with regulatory elements of differential activity between HCM patients and control donors (available upon request).

3.6.3 Supplemental Tables

Supplemental Table 1.

Overview of the 20 QRS-associated loci overlapping regulatory regions of interest. rsID, chromosome, position (hg19), NEF = Non effect allele, EF = effect allele, MAF = maximum allele frequency, CADD score, r^2, lead SNP, overlap with promoters of down-regulated genes in HCM, overlap with promoters of up-regulated genes in HCM, overlap with H3K27ac regions down-acetylated in HCM, overlap with H3K27ac regions up-acetylated in HCM, overlap with HCM super-enhancers, overlap LV super-enhancers, first, second and third NG = nearest gene, nearest gene (from the nearest three) down-regulated in HCM, nearest gene (from the nearest three) up regulated in HCM, gene identified by eQTL mapping in LV.

rsID	chr	pos	NEF	EF	MAF	CADD	r^2	Lead SNP	PROM down	PROM up	H3K27 ac down	H3K27 ac up
rs680386	chr1	23695275	G	T	0.19	5.7	0.6	rs2849028	X			
rs72694106	chr1	51762346	G	A	0.02	4.7	0.6	rs17391905	X			
rs2274316	chr1	156446242	A	C	0.33	12.3	1.0	rs2274317				
rs12131289	chr1	156446450	T	C	0.33	11.4	1.0	rs2274317				
rs2274317	chr1	156446903	C	T	0.33		1.0	rs2274317				
rs4450010	chr1	156448209	G	T	0.33		1.0	rs2274317				
rs1185700	chr1	156450719	G	A	0.24	14.7	0.6	rs2274317				
rs1925950	chr1	156450740	A	G	0.33	11.7	1.0	rs2274317				
rs2274319	chr1	156450873	C	T	0.33	2.2	1.0	rs2274317				
rs3790457	chr1	156458840	T	C	0.33	12.0	1.0	rs2274317				
rs12038396	chr1	156459535	T	C	0.33	7.3	1.0	rs2274317				
rs3790458	chr1	156459699	A	T	0.24	9.6	0.6	rs2274317				
rs746527	chr1	156460511	C	T	0.24	15.7	0.6	rs2274317				
rs3790461	chr1	156469364	G	A	0.24	7.9	0.6	rs2274317				
rs12136856	chr1	156473114	G	C	0.33	12.5	1.0	rs2274317				
rs538050729	chr1	156474929	GC	G	0.35	12.6	0.9	rs2274317				
rs1104859	chr1	201331554	G	T	0.26	16.3	0.6	rs10920184				
rs2365652	chr1	201331664	A	C	0.35	5.1	1.0	rs10920184				
rs1892026	chr1	201332886	T	G	0.26	2.1	0.6	rs10920184				
rs3729547	chr1	201334382	A	G	0.26	14.2	0.6	rs10920184				
rs10800776	chr1	201338553	C	T	0.26	0.6	0.6	rs10920184				
rs10920183	chr1	201338586	G	A	0.35	1.8	1.0	rs10920184				
rs4915232	chr1	201347946	T	C	0.43	2.1	0.7	rs10920184				
rs2252860	chr2	26987481	A	G	0.41	7.6	0.7	rs6710065	X			
rs2304337	chr2	26987488	G	C	0.41	5.9	0.7	rs6710065	X			
rs34902032	chr2	27084942	TC	T	0.38	1.6	0.9	rs6710065				X
rs11677841	chr2	27239949	T	G	0.38	1.3	0.5	rs6710065				X
rs968702	chr3	53276485	G	C	0.18	0.5	0.7	rs4687718		X		X
rs62256006	chr3	53289634	G	A	0.16	12.3	0.6	rs4687718		X		X
rs17523471	chr3	185304018	T	C	0.32	4.4	1.0	rs10937226				X
rs146170154	chr6	36646768	C	CTA	0.17	8.4	0.7	rs1321311		X		
rs3176326	chr6	36647289	G	A	0.17	7.9	0.7	rs1321311		X		
rs6913012	chr6	118973137	A	G	0.46	8.4	0.7	rs11153730		X		
rs1003549	chr7	35295491	T	C	0.13	0.8	0.7	rs1419856		X		
rs2270188	chr7	116140524	T	G	0.48	11.8	0.5	rs11773845				
rs2270189	chr7	116140616	G	A	0.48	7.8	0.5	rs11773845				
rs4367519	chr8	124666429	C	T	0.05	8.9	1.0	rs4367519				
rs143950919	chr8	124666582	T	TA	0.05	2.5	1.0	rs4367519				
rs11000728	chr10	75404300	C	G	0.13	5.6	0.8	rs7099599				
rs3812629	chr10	75407290	G	A	0.13		0.8	rs7099599				

SE HCM	SE LV	NG 1	NG 2	NG 3	NG Down	NG Up	eQTL gene
		ZNF436	C1orf213	TCEA3	ZNF436		TCEA3
		TTC39A	RNF11	EPS15	TTC39A		
	X	MEF2D	C1orf61	IQGAP3			MEF2D
	X	MEF2D	C1orf61	IQGAP3			MEF2D
	X	MEF2D	C1orf61	IQGAP3			MEF2D
	X	MEF2D	IQGAP3	C1orf61			MEF2D
	X	MEF2D	IQGAP3	C1orf61			
	X	MEF2D	IQGAP3	C1orf61			MEF2D
	X	MEF2D	IQGAP3	C1orf61			
	X	MEF2D	IQGAP3	C1orf61			MEF2D
	X	MEF2D	IQGAP3	C1orf61			MEF2D
	X	MEF2D	IQGAP3	C1orf61			
	X	MEF2D	IQGAP3	C1orf61			
	X	MEF2D	IQGAP3	C1orf61			
	X	MEF2D	IQGAP3	C1orf61			
	X	MEF2D	IQGAP3	C1orf61			
	X	TNNT2	LAD1	PKP1			
	X	TNNT2	LAD1	PKP1			
	X	TNNT2	LAD1	PKP1			
	X	TNNT2	LAD1	PKP1			
	X	TNNT2	LAD1	TNNI1			
	X	TNNT2	LAD1	TNNI1			
	X	LAD1	TNNT2	TNNI1			
		SLC35F6	CENPA	KCNK3	CENPA	KCNK3	KHK
		SLC35F6	CENPA	KCNK3	CENPA	KCNK3	KHK
		DPYSL5	CENPA	SLC35F6	CENPA		KHK
		MAPRE3	TMEM214	AGBL5	AGBL5		KHK
		TKT	DCP1A	PRKCD		PRKCD,TKT	
		TKT	DCP1A	PRKCD		PRKCD,TKT	
		SENP2	LIPH	IGF2BP2	LIPH		
		CDKN1A	RAB44	CPNE5		CPNE5,CDKN1A	
		CDKN1A	RAB44	CPNE5		CPNE5,CDKN1A	
		CEP85L	PLN	MCM9		CEP85L	
		TBX20	DPY19L1	HERPUD2		TBX20,DPY19L1	
	X	CAV2	CAV1	MET			
	X	CAV2	CAV1	MET			
	X	KLHL38	ANXA13	FBXO32	FBXO32		
	X	KLHL38	ANXA13	FBXO32	FBXO32		
	X	SYNPO2L	MYOZ1	USP54	SYNPO2L,MYOZ1		FUT11
	X	SYNPO2L	MYOZ1	USP54	SYNPO2L,MYOZ1		FUT11

rs4746139	chr10	75407649	A	C	0.13	0.0	0.8	rs7099599				
rs60632610	chr10	75415677	C	T	0.14	32.0	0.9	rs7099599	X			
rs2167079	chr11	47270255	C	T	0.30	22.6	0.8	rs2269434		X		
rs3758673	chr11	47278917	C	T	0.30	6.7	0.8	rs2269434		X		
rs1449627	chr11	47290984	T	G	0.31	6.5	0.8	rs2269434		X		
rs142960070	chr11	47291817	T	TG	0.31	6.6	0.8	rs2269434		X		
rs3781622	chr11	47348702	T	C	0.31	1.4	0.8	rs2269434				
rs753993	chr11	47349969	C	A	0.20	13.1	0.5	rs2269434				
rs10838693	chr11	47350553	G	C	0.31		0.9	rs2269434				
rs4752825	chr11	47352409	G	A	0.31	4.0	0.8	rs2269434				
rs11570050	chr11	47371484	A	AG	0.29	1.0	0.6	rs2269434				
rs7948705	chr11	47447955	C	G	0.29	8.9	0.6	rs2269434				X
rs11307826	chr11	47448334	CA	C	0.29	10.4	0.6	rs2269434				X
rs174538	chr11	61560081	G	A	0.31	9.2	0.8	rs174577	X			
rs174561	chr11	61582708	T	C	0.30	7.9	0.8	rs174577	X			
rs3834458	chr11	61594920	CT	C	0.35	10.9	0.9	rs174577	X			
rs5792235	chr11	61596322	CA	C	0.35	9.4	0.9	rs174577	X			
rs99780	chr11	61596633	C	T	0.36	8.6	1.0	rs174577	X			
rs2958153	chr12	57081517	G	A	0.27	5.9	0.9	rs2926743				X
rs3214051	chr12	57119236	A	G	0.37	8.2	0.6	rs2926743	X			
rs8039472	chr15	85361644	G	A	0.49	5.7	0.8	rs7183401				
rs4633690	chr15	85361960	C	T	0.45	18.3	0.9	rs7183401				
rs2879828	chr15	85361977	G	T	0.45	15.8	1.0	rs7183401				
rs871014	chr17	27945339	C	T	0.49	7.2	0.5	rs7211246	X			
rs3760456	chr17	27948844	C	T	0.45	7.4	0.6	rs7211246	X			
rs2617865	chr17	28049804	T	G	0.47	15.5	0.8	rs7211246				
rs2628179	chr17	28071796	C	G	0.46	5.3	0.7	rs7211246				
rs4795529	chr17	28443549	A	G	0.47	5.5	0.9	rs7211246	X			
rs6505162	chr17	28444183	A	C	0.44		0.8	rs7211246	X			
rs8067576	chr17	28444254	T	A	0.44	15.6	1.0	rs7211246	X			
rs530081	chr18	32621231	G	A	0.33		0.6	rs617759	X			
rs6060266	chr20	33733078	T	C	0.23	9.1	0.6	rs2025096		X		
rs111275419	chr20	33739804	CCCAT	C	0.23	15.8	0.6	rs2025096				
rs6058194	chr20	33739831	G	A	0.23	20.1	0.6	rs2025096				

	X	SYNPO2L	MYOZ1	USP54	SYNPO2L,MYOZ1		
		SYNPO2L	MYOZ1	AGAP5	SYNPO2L,MYOZ1		FUT11
		ACP2	NR1H3	DDB2		NR1H3	ACP2
		NR1H3	ACP2	MADD		NR1H3	ACP2
		MADD	NR1H3	ACP2		NR1H3	ACP2
		MADD	NR1H3	ACP2		NR1H3	ACP2
	X	MADD	MYBPC3	SPI1		MYBPC3	ACP2
	X	MADD	MYBPC3	SPI1		MYBPC3	ACP2,MADD
	X	MADD	MYBPC3	SPI1		MYBPC3	ACP2
	X	MYBPC3	MADD	SPI1			ACP2
	X	MYBPC3	SPI1	MADD			
		PSMC3	SLC39A13	RAPSN			
		PSMC3	SLC39A13	RAPSN			
		TMEM258	FEN1	FADS2	FADS2		FADS1, FADS2
		FADS2	FADS1	FEN1	FADS2		FADS1, FADS2
		FADS2	FADS1	FEN1	FADS2		FADS1, FADS2
		FADS2	FADS1	FEN1	FADS2		FADS1, FADS2
		FADS2	FADS1	FEN1	FADS2		FADS1, FADS2
		PTGES3	NACA	ATP5B	NACA		
		NACA	PRIM1	HSD17B6	NACA		
	X	ALPK3	ZNF592	SLC28A1			
	X	ALPK3	ZNF592	SLC28A1			
	X	ALPK3	ZNF592	SLC28A1			
		CORO6	ANKRD13B	SSH2	CORO6		
		CORO6	SSH2	ANKRD13B	CORO6		
	X	SSH2	CORO6	ANKRD13B	CORO6		
	X	SSH2	CORO6	ANKRD13B	CORO6		
		NSRP1	EFCAB5	SLC6A4	NSRP1		
		NSRP1	EFCAB5	SLC6A4	NSRP1		
		NSRP1	EFCAB5	SLC6A4	NSRP1		
		MAPRE2	DTNA	ZNF397	MAPRE2		
		EDEM2	PROCR	TRPC4AP		PROCR,EDEM2	PROCR,EDEM2
X		EDEM2	PROCR	TRPC4AP		PROCR,EDEM2	PROCR,EDEM2
X		EDEM2	PROCR	TRPC4AP		PROCR,EDEM2	PROCR,EDEM2

Supplemental Table 2. Human myocardial samples included in this study.

Sample name	ChIP-seq	RNA-seq	SEX
HCM_1	x	x	M
HCM_2	x		M
HCM_3	x	x	M
HCM_4	x		M
HCM_5	x	x	M
HCM_6	x	x	M
HCM_7	x	x	F
HCM_8	x	x	M
HCM_9	x	x	M
HCM_10	x	x	F
HCM_11	x	x	F
HCM_12	x	x	M
HCM_13	x		M
HCM_14	x	x	M
CONTROL_ChIP_1	x		F
CONTROL_ChIP_2	x		M
CONTROL_ChIP_3	x		M
CONTROL_ChIP_4	x		F
CONTROL_RNA_1		x	M
CONTROL_RNA_2		x	M
CONTROL_RNA_3		x	M
CONTROL_RNA_4		x	F
CONTROL_RNA_5		x	F
CONTROL_RNA_6		x	F
CONTROL_RNA_7		x	M
CONTROL_RNA_8		x	M
CONTROL_RNA_9		x	M
CONTROL_RNA_10		x	F
CONTROL_RNA_11		x	F

Supplemental Table 3. Ensembl IDs of differentially regulated genes identified by RNA-seq experiment in HCM patients versus controls (available upon request).

Supplemental Table 4. Differentially acetylated regions identified by ChIP-seq experiment in HCM patients versus controls (available upon request).

Supplemental Table 5. HCM-specific super-enhancer regions (available upon request).

Supplemental Table 6. Description of new candidate genes identified mapped to QRS-associated loci overlapping regulatory regions with abnormal expression / acetylation between HCM and control patients.

LIPH - Lipase H	Lipase H is a phosphatidic acid-selective phospholipase A1 (PLA1) that produces 2-acyl lysophosphatidic acid (LPA). LPA is a lipid mediator with diverse biologic properties, including induction of platelet aggregation, smooth muscle contraction, and stimulation of cell proliferation [225].
CPNE5 - Copine 5	This gene is one of several genes that encode a calcium-dependent protein containing two N-terminal type II C2 domains and an integrin A domain-like sequence in the C-terminus. Has been implicated in previous GWAS in heart rate [226].
DPY19L1 - Dpy-19 Like C-Mannosyltransferase 1	Function unknown. Automated annotations supported by experiments on knockout mouse models associate this gene to abnormal heart morphology and enlarged heart [227].
TTC39A - Tetratricopeptide repeat domain 39A	Function unknown. Automated annotations supported by experiments on knockout mouse models associate this gene to increased heart weight and increased circulating phosphate level [228].
PROCR - Endothelial protein C receptor	The protein encoded by this gene is a receptor for activated protein C, a serine protease activated by and involved in the blood coagulation pathway. The encoded protein is an N-glycosylated type I membrane protein that enhances the activation of protein C. Mutations in this gene have been associated with venous thromboembolism, myocardial infarction and coronary artery disease [229, 230].
PRKCD - Protein kinase C delta	Protein kinase C (PKC) is a family of serine- and threonine-specific protein kinases that can be activated by calcium and the second messenger diacylglycerol. Studies both in human and mice demonstrate that PRKCD kinase is involved in B cell signaling and in the regulation of growth, apoptosis, and differentiation of a variety of cell types [231].
CORO6 - Coronin 6	Involved in actin filament binding. Overexpressed in heart and skeletal muscle [232]. Has been implicated in previous GWAS in coronary artery disease [229].
TCEA3 - Transcription Elongation Factor A3	This gene is associated to other measure of amplitude of the electrocardiogram, ST-T wave [233] and QT-interval [234]. Overexpressed in heart [232].
FUT11 - Fucosyltransferase 11	Function unknown.
CENPA - Centromere protein A	Histone H3-like variant which exclusively replaces conventional H3 in the nucleosome core of centromeric chromatin at the inner plate of the kinetochore [235]. Required for recruitment and assembly of kinetochore proteins, mitotic progression and chromosome segregation. May serve as an epigenetic mark that propagates centromere identity through replication and cell division [151, 235-237].
AGBL5 - ATP/GTP binding protein-like 5	Function unknown.
KCNK3 - Potassium channel, subfamily K, member 3	pH-dependent, voltage-insensitive, background potassium channel protein. Rectification direction results from potassium ion concentration on either side of the membrane. Acts as an outward rectifier when external potassium concentration is low [238].

KHK - ketohexokinase	Catalyzes the first step of metabolism of dietary fructose, conversion of fructose to fructose-1-phosphate. It has been shown that myocardial hypoxia influences fructose metabolism in human and mouse models of pathologic cardiac hypertrophy through hypoxia-inducible factor 1-alpha activation of *SF3B1* and *SF3B1*-mediated splice switching of *KHK-A* to *KHK-C* [239, 240]. In mice, heart-specific depletion of *SF3B1* or genetic ablation of *KHK* suppressed pathologic stress-induced fructose metabolism, growth, and contractile dysfunction [240].
FADS1 - Fatty Acid Desaturase 1	A member of the fatty acid desaturase (FADS) gene family. Component of a lipid metabolic pathway that catalyzes the biosynthesis of highly unsaturated fatty acids from precursor essential polyunsaturated fatty acids, linoleic acid, and alpha-linolenic acid. SNPs of the *FADS* gene cluster have been associated with polyunsaturated fatty acids in a cohort of patients with cardiovascular disease [241].

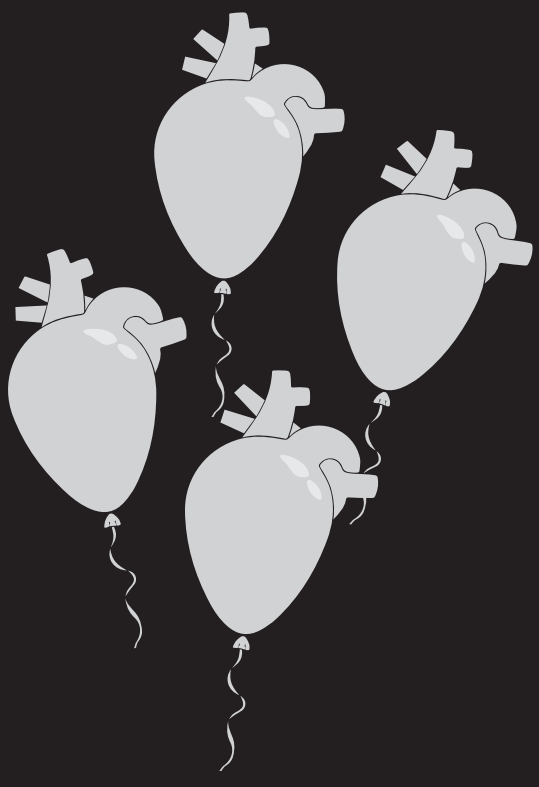
Supplemental Table 7.
Summary of the functional annotation using differential regulatory elements in HCM patients versus controls.

Locus	Lead QRS-associated SNP	Genes from original study	# LD SNPs overlapping exons with CADD > 12.37
1p36.12	rs2849028	ZNF436; C1orf213	
1p32.3	rs17391905	CDKN2C	
1p31.3	rs2207790	NFIA	
1p13.1	rs12039739	CASQ2	
1q22	rs2274317	MEF2D	1
1q23.3	rs12036340	OLFML2B	
1q32.1	rs10920184	TNNT2	1
1q32.1	rs4288653	PLEKHA6	
2p23.3	rs6710065	DPYSL5	1
2p22.2	rs3770770	STRN	
2q31.2	rs3816849	TTN	
3p22.2	rs6801957	SCN10A	
3p21.1	rs4687718	TKT	
3p14.1	rs2242285	LRIG1	
3p14.1	rs13314892	MITF	
3q27.2	rs10937226	SENP2	
4p15.31	rs1344852	SLIT2	
5q33.2	rs13185595	HAND1	
6p21.31	rs1321311	CDKN1A	
6p21.1	rs1015150	TFEB	
6q22.31	rs11153730	PLN;SLC35F1;CEP85L	
7p14.3	rs1419856	TBX20	
7p12.3	rs6968945	TNS3	
7q31.2	rs11773845	CAV1	
8q24.13	rs4367519	FBXO32;KLHL38	1
8q24.13	rs10105974	MTSS1	
10q21.1	rs1733724	DKK1	
10q21.3	rs12414364;rs10509289	CTNNA3	
10q22.2	rs7099599	SYNPO2L;SEC24C;CAMK2G	2
10q25.2	rs7918405	VTI1A	
11p11.2	rs2269434	ACP2;MADD;MYBPC3;NR1H3	1
11q12.2	rs174577	FADS2;TMEM258	

# LD SNPs overlapping differentially regulated promoters in HCM x controls	# LD SNPs overlapping differentially acetylated regions in HCM x controls	# LD SNPs overlapping HCM or LV super-enhancers	Genes from this study
1			ZNF436; TCEA3
1			TTC39A
		14	MEF2D
		7	TNNT2
2	2		CENPA; AGBL5; KCNK3; KHK
2	2		PRKCD;TKT
	1		LIPH
2			CPNE5; CDKN1A
1			CEP85L
1			TBX20; DPY19L1
		2	
1		3	SYNPO2L; MYOZ1; FUT11
4	2	5	ACP2; MADD; MYBPC3; NR1H3
5			FADS1; FADS2

12q13.13	rs736825	HOXC5;HOXC4;HOXC6	
12q13.3	rs2926743	NACA	3
12q24.21	rs7132327	TBX3	
13q14.13	rs1408224	LRCH1	
13q22.1	rs728926	KLF12	
14q24.2	rs12880291	SIPA1L1	
15q25.3	rs7183401	ALPK3	
15q26.3	rs8038015	IGF1R	
16q23.3	rs6565060	CDH13	
17q11.2	rs7211246	NSRP1;EFCAB5	
17q21.31	rs242562	MAPT	
17q24.2	rs9912468	PRKCA	
18q12.1	rs617759	MAPRE2	
18q12.2	rs879568	FHOD3	3
18q12.3	rs10853525	SETBP1	
20p12.3	rs3929778	BMP2	
20q11.22	rs2025096	MYH7B;GSS;EDEM2	1
21q21.1	rs7283707	USP25	
21q21.3	rs13047360	ADAMTS5	

1	1			
			3	
5			2	NSRP1; CORO6
1				MAPRE2
			2	PROCR; EDEM2



Chapter 4

Regulatory and transcriptional profile in dilated cardiomyopathy

Hemerich D, Pei J, Harakalova M, van Setten J, Boukens BJ,
Efimov IR, Vink A, Cheng C, Mokry M, Tragante V, Asselbergs FW

Manuscript in preparation

4.1 Abstract

Dilated cardiomyopathy (DCM) is a major cause of heart failure and mortality caused by a long list of etiologies, ranging from myocardial infarction, DNA mutations in cardiac genes, to toxics. Despite the identification of over 100 genes involved in DCM, much of the heritability and the regulatory processes of DCM are not fully understood. Integrating transcriptomic and regulatory information can help to identify pathways involved in this heterogeneous disease. In this study, we compared the transcriptome and regulome of six DCM patients with eight control donors. We identified 454 differentially expressed genes between the two groups, 84 of which also show differential activity in their promoters. We confirmed several pathways implicated in the etiology of DCM, such as mitochondrial myopathy, ion channel binding and SMAD binding. Furthermore, we predicted 148 motifs enriched in the sequences of candidate distal enhancer regions with differential activity between DCM and control groups, six of which are also differentially expressed between the two conditions (*BACH2*, *JUN*, *TEAD4*, *PRDM1*, *SOX4*, *FOXO1*). Our results add evidence of the role of genes and pathways previously associated to the process of cardiomyopathy and shed light on the complexity and heterogeneity of DCM.

4.2 Introduction

Dilated cardiomyopathy (DCM) is the most common type of cardiomyopathies, occurring mostly in adults aged 20 to 60, and is a major cause of heart failure and mortality [242, 243]. It has been linked to over 100 genes to date [244-250]. The gene *TTN*, for instance, is responsible for approximately 25% of familial cases of idiopathic dilated cardiomyopathy and 18% of sporadic cases [251]. However, the remaining heritability and the regulatory processes that mediate biological mechanisms of DCM remain incompletely understood. Systems analysis integrating next generation sequencing-based "omics" approaches, such as the genome-wide sequencing of DNA, RNA, and chromatin (genomics, transcriptomics, and regulomics, respectively [60]), may provide valuable insights into our understanding of DCM [252].

Distal regulatory sequences, such as enhancers, have been shown to contribute to gene activation in a cell-type specific manner [15, 60, 253-260]. Disruption of such sequences is believed to contribute significantly to the etiology of several traits, including congenital heart disease and some forms of adult heart diseases [261, 262]. Thus, identification of specific regulatory elements influencing the biological mechanisms of DCM may enhance our understanding of the human heart and ultimately lead to better prevention of CVDs and death.

In this study, we examined gene expression and the activity of putative enhancers and promoters in the human heart. We characterized differences between control donors and individuals with genetic forms of DCM (caused by mutations on *TTN*, *MYH7* and *MYBPC3* genes) that can shed a light on the mechanisms of this complex heterogeneous disease, and the role of regulatory elements.

4.3 Methods

4.3.1 ChIP-seq data generation and differential analysis

Myocardial samples from six DCM patients and eight controls (obtained from autopsy) were included in the study (Supplemental Table 1). All samples were collected from the left ventricular region. Chromatin immunoprecipitation and sequencing (Chip-seq) using the H3K27ac mark was performed to study the difference of chromatin regulation between control and disease groups on a genome-wide scale (Supplemental Figure 1). Chromatin was isolated from each sample using the MAGnify™ Chromatin Immunoprecipitation System kit (Life Technologies) according to the manufacturer's instructions. Immunoprecipitations were performed with the antibody against H3K27ac (ab4729, Abcam) for ChIP-seq as described previously [204]. ChIP DNA Clean & Concentrator kit (Zymo Research) was used to purify captured DNA fragments. Libraries were prepared using the NEXTflex™ Rapid DNA Sequencing Kit (Bioo

Scientific) and sequenced on the Illumina NextSeq500 sequencer. Alignment to the human reference genome (hg19) was performed using BWA v0.7.5a [205]. Duplicated, unmapped reads and reads with mapping quality less than 5 (-b -F 4 -q 5) were removed using Samtools v1.3 [206]. Peak calling was performed using MACS v 2.1.0 [207] using the respective input samples, -gsize=hg -nomodel parameters, and estimated fragment sizes (-extsize) predicted by PhantomPeakQualTools v1.1 [208].
Epigenomic profiles were compared using DiffBind v2.6.6 [209] to identify candidate enhancer peaks that are quantitatively different (FDR<0.05, EdgeR algorithm) between the healthy and DCM populations. Consensus peaksets were formed by peaks that overlapped in at least two samples. We defined regions between +2500 and -2500 base pairs (bps) from the transcription start site (TSS) of genes as promoter regions. Coordinates of all human genes (GRCh37.13, Ensembl Genes 92) were retrieved using Biomart [203]. The remaining differentially acetylated regions were classified as candidate differential distal enhancers.

4.3.2 RNA-seq data generation and differential analysis

We performed RNA-seq on the same fourteen biological replicates of human myocardial tissue, six from DCM patients and eight from control donors (Supplemental Table 1, Supplemental Figure 2). RNA was isolated using Qiagen AllPrep Micro Kit according to the manufacturer's instructions. After the selection of mRNA, libraries were prepared using the NEXTflex™ Rapid RNA-seq Kit (Bio Scientific). Libraries were sequenced on the Nextseq500 platform (Illumina), producing single-end reads of 75bp. Reads were aligned to the human reference genome GRCh37 using STAR v2.4.2a [198]. The tool AddOrReplaceReadGroups v1.98 from the Picard package (http://broadinstitute.github.io/picard/) was used to add read groups to the BAM files, which were sorted with Sambamba v0.4.5 [199] and transcript abundances were quantified with HTSeq-count v0.6.1p1 [200] using the union mode. Subsequently, reads per kilobase million reads sequenced (RPKMs) were calculated with edgeR's RPKM function [201]. In order to obtain a list of differentially expressed genes between DCM and controls, we employed Deseq2 v1.10.1 package [202]. We defined differentially expressed genes as those having adjusted *p*-value<0.05 and fold change less than 0.8 or greater than 1.2. We calculated *p*-values using Wald statistics and corrected for multiple testing using the Benjamini-Hochberg method. The transcription start sites (TSS) of the obtained genes were retrieved from Ensembl Genes 89 using Biomart (Human genes GRCh37.13) [203]. Promoter regions were defined as ranging from -2500 to +2500 base-pairs (bp) from the TSS.

4.3.3 Pathway and Gene Ontology (GO) enrichment

We employed ToppFun from ToppGene Suite [263] for pathway and GO enrichment on differentially expressed genes. We used the same tools for enrichment analysis on genes that showed differentially acetylation in promoters between DCM patients and

control donors. We used an FDR cutoff of $p<0.05$, grouping results within a category by source. We used GREAT v3.0.0 [264] to identify enriched GO terms and pathways on differentially acetylated regions classified as potential distal enhancers, with an FDR cutoff of $p<0.05$.

4.3.4 Motifs predicted as disturbed in DCM

We employed Analysis of Motif Enrichment (AME) v5.0.1 for transcription factor binding detection and motif enrichment [265] from HOCOMOCO Human v11 database [266]. We used as input unique DNase I hypersensitive sites (DHSs) on left ventricle (LV) overlapping the identified putative distal enhancers. DHSs not overlapping with candidate distal enhancers were used as background control regions.

4.4 Results

4.4.1 Differential gene expression in DCM

We identified 430 protein coding and 24 noncoding genes that were differentially expressed between DCM and control groups using RNA-seq (Figure 1, Supplemental Tables 2 and 3), ten of which are well-known genes associated to DCM (odds ratio (OR)=2.27, p=0.02, Fisher's exact test (FET)) (Supplemental Table 5) [249, 267]. The top 20 most differentially expressed genes are shown in Table 1. Interestingly, the differentially expressed genes with the highest fold changes are mitochondrially encoded tRNAs. The presence of these tRNAs in the set of noncoding differentially regulated genes resulted in an enrichment to GO terms such as mitochondrial myopathy (p=2.5E-9) and muscle abnormality related to mitochondrial dysfunction (p=3.2E-9).

Differentially expressed coding genes were enriched for pathways involved in extracellular matrix (p=2.40E-9), collagen formation (p=1.35E-5) and tubulin folding (p=2.80E-5) (Figure 2).

4.4.2 Differential acetylation in DCM

In total, 20,979 regions were differentially acetylated between DCM and the control group (FDR <0.05); 8,315 regions of those considered hyper-acetylated, and the remaining 12,664 regions considered hypo-acetylated (Supplemental Figure 1). We then identified differentially acetylated peaks overlapping promoter regions, defined as -2500 and +2500 bp from TSSs of all genes. In total, we obtained 2,651 hypo-acetylated peaks and 729 hyper-acetylated peaks overlapping promoters. These 3,380 promoter regions in total belong to 2,784 genes (Supplemental Table 5). Pathways enriched for this set of genes include adrenergic signaling in cardiomyocytes (p=1.24E-6), *Rap1* signaling pathway (p=3.19E-6) and signaling by Rho GTPases (p=7.91E-9) (Figure 2).

Table 1. Differentially expressed genes with greatest absolute log fold changes.

Gene symbol	Description	pvalue	padj	Log fold change	DCM candidate	Comments
MT-TQ	mitochondrially encoded tRNA glutamine	1.24E-17	1.71E-13	-2.84	X	Previously associated to DCM [249]
MT-TP	mitochondrially encoded tRNA proline	8.50E-10	1.67E-06	-2.23	X	Previously associated to DCM [249, 268]
TUBA3D	tubulin alpha 3d	1.07E-10	4.90E-07	-2.01		Member of the alpha-tubulin family. Tubulin is a major component of microtubules that maintain cellular structure, function, and intracellular transport [269]
MT-TA	mitochondrially encoded tRNA alanine	1.87E-07	0.000168	-1.99	X	Previously associated to DCM [249, 270, 271]
FCN3	ficolin 3	1.81E-09	3.11E-06	-1.90		Low ficolin-3 concentration in serum has already been associated with heart failure [272], cardiac syndrome X [273] and ischemic cardiomyopathy[274]
TUBA3E	tubulin alpha 3e	4.39E-10	1.51E-06	-1.88		See *TUBA3D*.
C1orf105	chromosome 1 open reading frame 105	8.18E-07	0.000555	-1.86		Previously associated to inter-adventitial common carotid artery diameter [275]
MOG	myelin oligodendrocyte glycoprotein	9.24E-07	0.000555	-1.81		Member of the immunoglobulin superfamily, found on the surface of myelinating oligodendrocytes and external lamellae of myelin sheaths in the central nervous system[276]
PDCD1	programmed cell death 1	2.88E-06	0.001135	-1.79		Previously associated to DCM [277-279].
HOPX	HOP homeobox	5.33E-06	0.001506	-1.74	X	Previously associated to DCM [267]
ANKRD2	ankyrin repeat domain 2	3.63E-06	0.001275	-1.72		Member of the muscle ankyrin repeat protein (MARP) family, whose expression is induced in response to hypertrophy [280-282]. Differential regulation associated to DCM [283], congenital myopathies [284] and muscular dystrophy [282, 284]

LINC01405	long intergenic non-protein coding RNA 1405	2.55E-06	0.001094	-1.67	Shown to be involved in mitochondrial myopathy pathogenesis [285]
LCN6	lipocalin 6	1.78E-05	0.003211	-1.63	Differential regulation associated to ischemic DCM [286]. Identified as a myocardial biomarkers of RV fate and function in end-stage human heart failure [287]
LYVE1	lymphatic vessel endothelial hyaluronan receptor 1	8.97E-07	0.000555	-1.60	Differential regulation associated to DCM [288], plays a role in early muscle degeneration[289]
FCGBP	Fc fragment of IgG binding protein	2.59E-05	0.004277	-1.59	Differential regulation associated to acute ischemic stroke [290]
CHGB	Chromogranin B	3.08E-06	0.001	0.46	A putative cardiac biomarker in heart failure [291, 292]
EGR1	early growth response 1	2.36E-05	0.00405	1.61	Transcription factor involved in the promotion of cardiac hypertrophy and damage [293]. Previously associated to cardiac calcium dynamics [294, 295] and cardiac fibroblast apoptosis[296]
CBLN1	cerebellin 1 precursor	1.93E-06	0.000916	1.62	Encodes a cerebellum-specific precursor protein, precerebellin, with similarity to the globular (non-collagen-like) domain of complement component C1qB. Shown to be upregulated during embryonic stem cell differentiation in a study with isolated multipotent cardiovascular progenitors [297]
COL24A1	collagen type XXIV alpha 1 chain	2.12E-06	0.000952	1.64	Member of the collagen gene family, may participate in regulating type I collagen fibrillogenesis at specific anatomical locations during fetal development [298]
IGHM	immunoglobulin heavy constant mu	3.71E-06	0.001275	1.75	Differential regulation associated to ischemic cardiomyopathy [299]

Among these 2,784 genes with differentially acetylated promoters are 28 well-known DCM-associated genes (Figure 1, Supplemental Tables 4 and 5) [249, 267]. In addition, 85 of these 2,784 genes also showed differential expression in the previous RNA-seq analysis between the same DCM and control samples (Figure 1, Supplemental Table 2). In addition to differentially acetylated promoters, 17,559 distal differentially acetylated regions were identified and classified as putative distal enhancers with differential activity (Supplemental Table 6). The enrichment of GO terms add evidence of the involvement of ion channel binding (p=4E-29) [300, 301], actinin binding (p=1.3E-23) [302] and SMAD binding (p=3.2E-18) [303-305] in the process of DCM (Figure 2).

4.4.3 Motifs potentially disturbed in DCM

We also investigated whether the differential regulatory regions classified as candidate distal enhancers are enriched for known DNA-binding motifs. All differentially acetylated peaks were intersected with DHSs of LV, in order to retain only possible transcription factor binding sites. We identified 148 transcription factors (TFs) predicted to bind to the regions differentially acetylated in DCM patients compared to control donors (Supplemental Table 7). We observed that six of the TFs identified - *BACH2*, *JUN*, *TEAD4*, *PRDM1*, *SOX4*, *FOXO1* - also show differential regulation (OR=3.47, p=0.01, FET) (Figure 1). Of these, *BACH2*, *PRDM1* and *FOXO1* also show differential acetylation on their promoters.

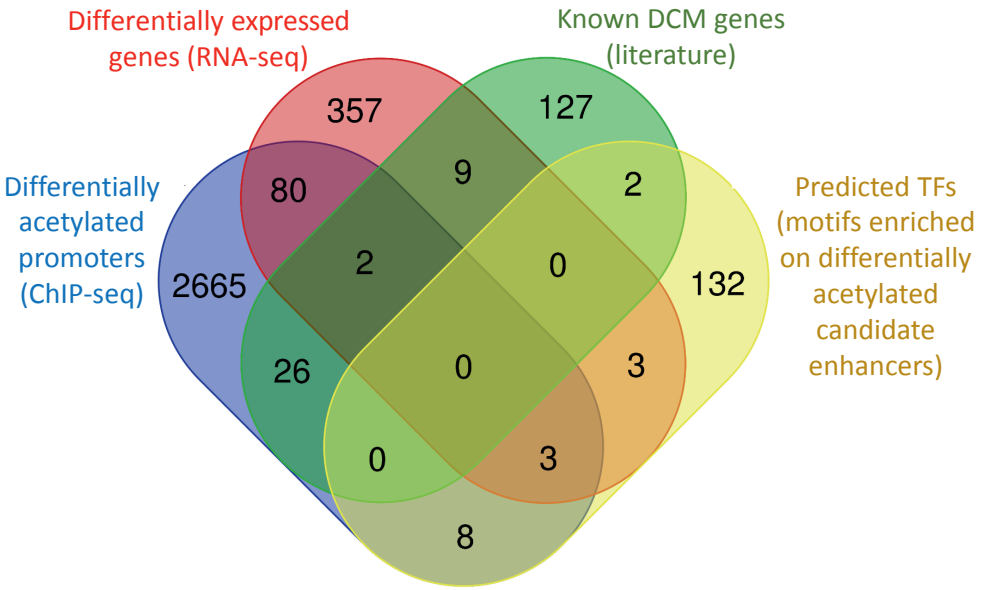


Figure 1. Overlap of genes with differential expression and/or regulatory activity between DCM patients and control donors identified in this study, also overlapping with well-known DCM genes [249, 267].

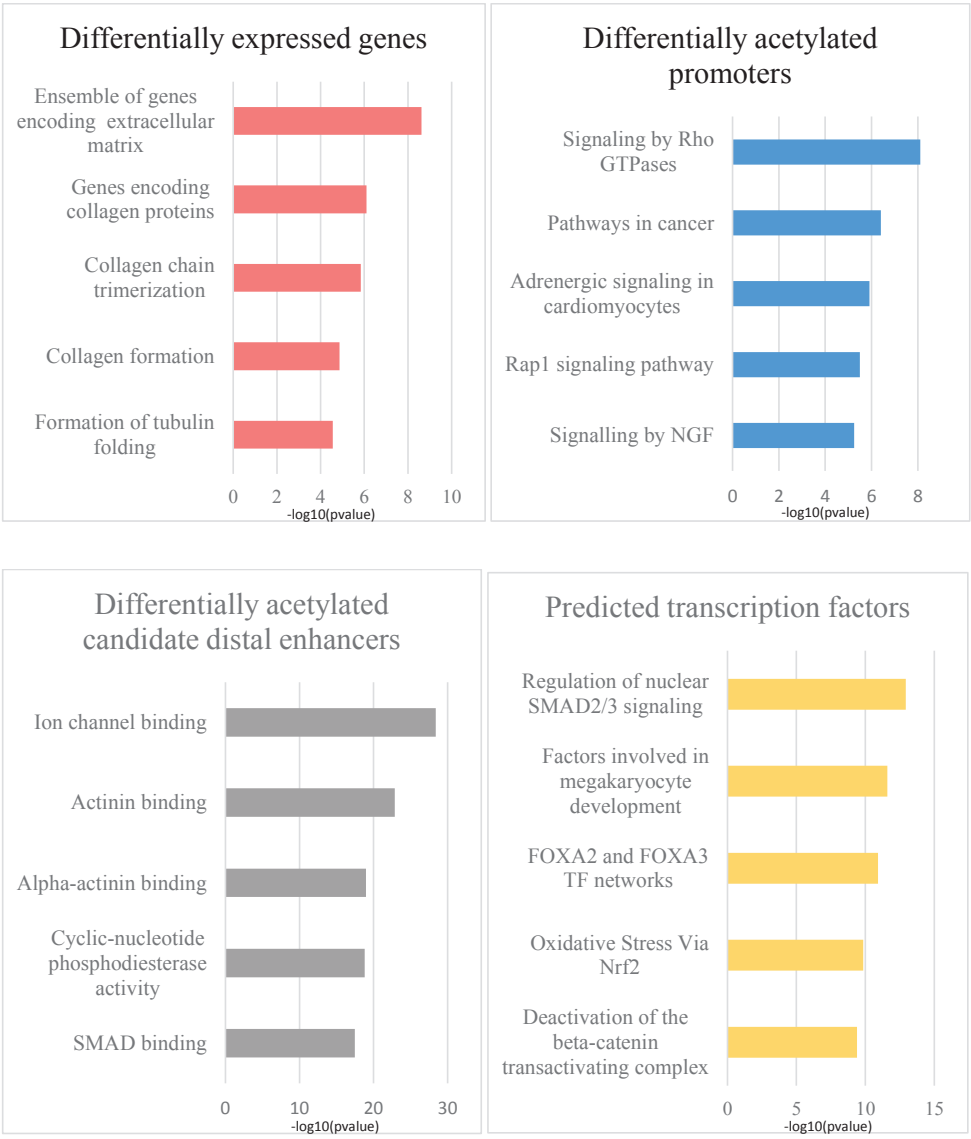


Figure 2. Top 5 pathways enriched in each set of genes or regulatory elements with differential activity or expression between DCM patients and control donors, as well as predicted TFs.

4.5 Discussion

In this study, we integrated regulome and transcriptome information on the investigation of genes and mechanisms involved in differential expression and regulation of DCM. We identified 454 genes that show differential expression between DCM patients and control donors. These are enriched for GO terms that implicate collagen formation and

extracellular matrix organization. Of the 20 genes with higher fold change between DCM and control groups, 15 have been previously identified in studies on DCM and/or other cardiovascular traits (Table 1). Interestingly, differentially expressed non-coding genes are enriched for pathways such as mitochondrial myopathy and muscle abnormality related to mitochondrial dysfunction. Genetic variation disturbing mitochondria dynamics are responsible for neurodegenerative conditions and cardiac failure [306, 307], and its role on DCM has been under investigation [308]. From the twenty-four mitochondrial genes previously implicated as associated to DCM [249], seven showed to be differentially regulated in our study, supporting the association of mitochondrial (mt)DNA defects with cardiomyopathies (Supplemental Tables 4 and 5) [270].

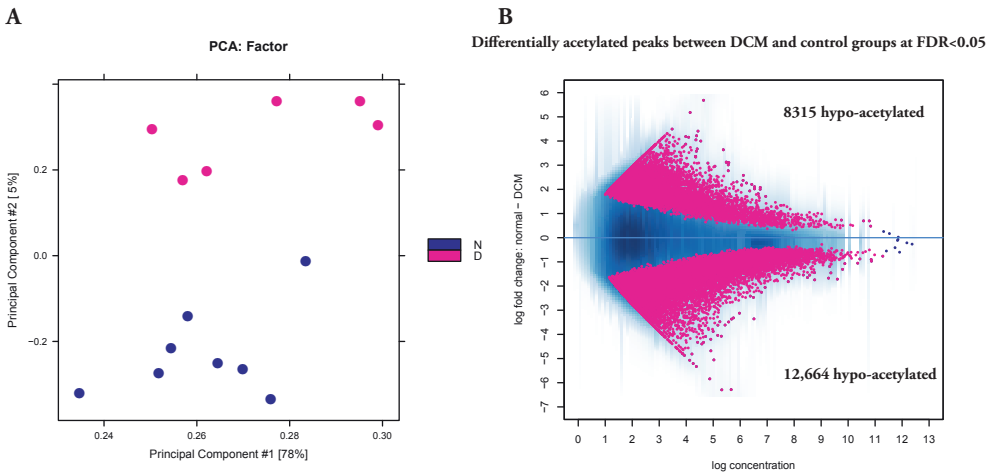
In order to investigate the regulome landscape with differential activity in DCM, the same samples used for the differential expression analysis with RNA-seq were also subject to ChIP-seq experiments on H3K27ac, a mark of active enhancers and promoters. With the ChIP-seq data generated on DCM and control samples, we analysed the differential acetylation activity between the two groups. We identified differentially acetylated promoter regions of 2,784 genes, 85 of which also showed differential expression in the previous RNA-seq analysis between the same samples (Figure 1). Among these 2,784 genes with differential promoter activity are 28 well-known DCM genes [249, 267] (Figure 1, Supplemental Tables 4 and 5). Enrichment analysis on this set of 2,784 genes identified pathways known to be involved in the process of cardiomyopathy, such as adrenergic signaling in cardiomyocytes [309] and signaling by Rho GTPases [310, 311] (Figure 2). We also observed the enrichment of *Rap1* signaling pathway, which has been implicated in vascular morphogenesis and dysfunction [312]. It has also been linked to the regulation of extracellular matrix, cellular proliferation and migration in certain fibroblast cells, suggesting a role for *Rap1a* in cardiac fibroblasts with potential implications in cardiac fibrosis [313].

Differentially acetylated regions not overlapping with promoter coordinates were classified as putative distal enhancers with differential activity (Supplemental Table 6). On these distal enhancers, we also observed enrichment of pathways known to be involved in cardiomyopathy processes, such as ion channel binding [300, 301], actinin binding [302] and SMAD binding [303-305] (Figure 2). Furthermore, we predicted 148 motifs enriched in these sequences. These might be TFs whose activity is disrupted along with the differential activity of the enhancers they bind to. Pathway enrichment on these motifs again highlights the role of SMAD binding in the process of DCM (Figure 2). Six of the 148 TFs identified also showed different expression levels between DCM patients and controls (*BACH2*, *JUN*, *TEAD4*, *PRDM1*, *SOX4*, *FOXO1*) (Figure 1). Of these six, *BACH2*, *PRDM1* and *FOXO1* also show differential acetylation on their promoters. These TFs and regulatory regions can be prioritized on future discovery studies and follow-up experiments.

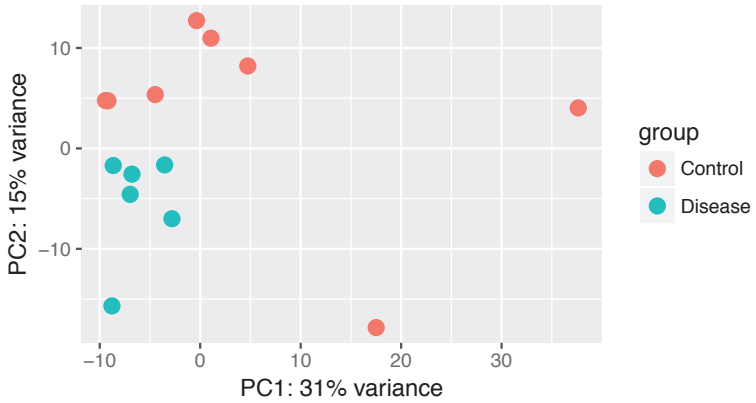
Overall, our results shed light into the complexity of DCM, and add evidence of the role of pathways and genes previously associated to the process of cardiomyopathy. The differential regulatory elements, genes and TFs identified here may serve as a guide for future discovery and follow-up studies. Increasing sample sizes for experiments on a homogeneous selection of DCM patients may help describe the epigenetic landscape of DCM [252, 314]. Identifying the underlying molecular mechanisms of DCM will improve candidate gene prioritization and ultimately the discovery of novel therapeutic approaches, also improving the assessment of this complex disease.

4.6 Supplemental Material

4.6.1 Supplemental Figures



Supplemental Figure 1. a) PCA plot showing the separation between normal (N) samples and DCM (D) patients in ChIP-seq data. **b)** MA plot depicting the 20,979 differentially acetylated peaks found between the two groups.



Supplemental Figure 2. PCA plot showing the separation between control and disease (DCM) patients in RNA-seq data.

Supplemental Table 1. Characteristics of the samples included in the study.

Group	ID	Sex	Mutation	Diagnosis	Age at diagnosis (years)	Age at explant (years)	Heart weight (g)	ICD	CRT-D	LVAD	Second HTx	Significant CAD	Last known EF before explant
DCM	DCM1	F	*TTN*	DCM	33	35	257	Yes	No	No	No	No	0.19
	DCM2	F	*TTN*	DCM with marks of hypertrophy	24	39	403	Yes	No	Yes	Yes	No	NA
	DCM3	M	*MYBPC3*	HCM in a dilated phase	17	37	486	Yes	Yes	No	No	No	0.21
	DCM4	F	*MYBPC3*	HCM	12	34	266	No	No	No	No	No	0.25
	DCM5	M	*MYH7*	HCM in a dilated phase	6	36	505	Yes	No	No	No	No	0.13
	DCM6	F	*MYH7*	DCM with marks of hypertrophy	26	26	174*	No	No	Yes	No	NA	NA
control	CONTROL 1	M											
	CONTROL 2	M											
	CONTROL 3	M											
	CONTROL 4	F											
	CONTROL 5	F											
	CONTROL 6	M											
	CONTROL 7	M											
	CONTROL 8	F											

*after removal of heart slice. F = female (Sex); M = male (Sex); NA = not available.
No clinical information available for control donors.

Supplemental Table 2. Protein-coding genes differentially expressed between DCM patients and controls (available upon request).

Supplemental Table 3. Noncoding genes differentially expressed between DCM patients and controls.

Supplemental Table 4. DCM associated genes from Roberts *et al.* (2015) [267] and Harakalova *et al.* (2015) [249].

DCM candidate gene	Reference	DCM candidate gene	Reference
ACTC1	Roberts *et al.* (2015)	*TMPO*	Roberts *et al.* (2015)
EMD	Roberts *et al.* (2015)	*DES*	Roberts *et al.* (2015)
LDB3	Roberts *et al.* (2015)	*HFE*	Roberts *et al.* (2015)
SGCA	Roberts *et al.* (2015)	*PRDM16*	Roberts *et al.* (2015)
ACTN2	Roberts *et al.* (2015)	*TNNC1*	Roberts *et al.* (2015)
EYA4	Roberts *et al.* (2015)	*DMD*	Roberts *et al.* (2015)
LMNA	Roberts *et al.* (2015)	*HOPX*	Roberts *et al.* (2015)
SGCB	Roberts *et al.* (2015)	*PSEN1*	Roberts *et al.* (2015)
ALMS1	Roberts *et al.* (2015)	*TNNI3*	Roberts *et al.* (2015)
FHL2	Roberts *et al.* (2015)	*DNAJC19*	Roberts *et al.* (2015)
MYBPC3	Roberts *et al.* (2015)	*ILK*	Roberts *et al.* (2015)
SGCD	Roberts *et al.* (2015)	*PSEN2*	Roberts *et al.* (2015)
ANKRD1	Roberts *et al.* (2015)	*TNNT2*	Roberts *et al.* (2015)
FKRP	Roberts *et al.* (2015)	*DSG2*	Roberts *et al.* (2015)
MYH6	Roberts *et al.* (2015)	*LAMA2*	Roberts *et al.* (2015)
SGCG	Roberts *et al.* (2015)	*RBM20*	Roberts *et al.* (2015)
APOA1	Roberts *et al.* (2015)	*TPM1*	Roberts *et al.* (2015)
FKTN	Roberts *et al.* (2015)	*DSP*	Roberts *et al.* (2015)
MYH7	Roberts *et al.* (2015)	*LAMA4*	Roberts *et al.* (2015)
SOD2	Roberts *et al.* (2015)	*SCN5A*	Roberts *et al.* (2015)
BAG3	Roberts *et al.* (2015)	*TTN*	Roberts *et al.* (2015)
FLT1	Roberts *et al.* (2015)	*DTNA*	Roberts *et al.* (2015)
NEXN	Roberts *et al.* (2015)	*LAMP2*	Roberts *et al.* (2015)
SYNE1	Roberts *et al.* (2015)	*SDHA*	Roberts *et al.* (2015)
CAV3	Roberts *et al.* (2015)	*VCL*	Roberts *et al.* (2015)
FOXD4	Roberts *et al.* (2015)	*ACAD8*	Harakalova *et al* (2015)
NFKB1	Roberts *et al.* (2015)	*ACTA1*	Harakalova *et al* (2015)
TAZ	Roberts *et al.* (2015)	*ADCY5*	Harakalova *et al* (2015)
CRYAB	Roberts *et al.* (2015)	*ALG6*	Harakalova *et al* (2015)
FXN	Roberts *et al.* (2015)	*ANO5*	Harakalova *et al* (2015)
PDLIM3	Roberts *et al.* (2015)	*BOLA3*	Harakalova *et al* (2015)
TBX20	Roberts *et al.* (2015)	*BRCC3/MTCP1a*	Harakalova *et al* (2015)
CSRP3	Roberts *et al.* (2015)	*CHKB*	Harakalova *et al* (2015)
GATAD1	Roberts *et al.* (2015)	*CHRM2*	Harakalova *et al* (2015)
PLEC	Roberts *et al.* (2015)	*CPT2*	Harakalova *et al* (2015)
TCAP	Roberts *et al.* (2015)	*DMPK*	Harakalova *et al* (2015)
CTF1	Roberts *et al.* (2015)	*DPM3*	Harakalova *et al* (2015)
HADHA	Roberts *et al.* (2015)	*DYSF*	Harakalova *et al* (2015)
PLN	Roberts *et al.* (2015)	*EMD/LMNAb*	Harakalova *et al* (2015)

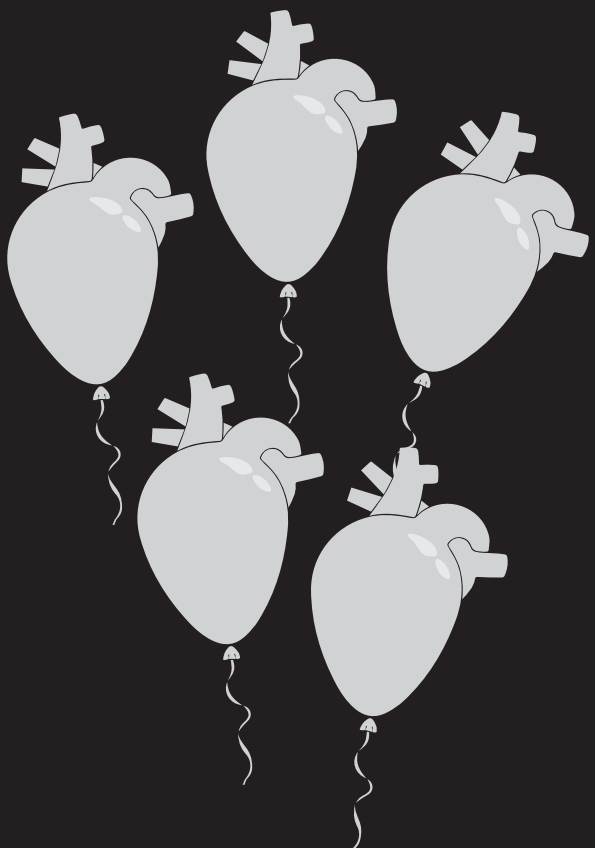
DCM candidate gene	Reference
EPG5	Harakalova *et al* (2015)
ERBB3	Harakalova *et al* (2015)
FBN1	Harakalova *et al* (2015)
FHL1	Harakalova *et al* (2015)
GATA4	Harakalova *et al* (2015)
GBE1	Harakalova *et al* (2015)
GLB1	Harakalova *et al* (2015)
GNPTAB	Harakalova *et al* (2015)
HADHB	Harakalova *et al* (2015)
HFE2/HFEc	Harakalova *et al* (2015)
HFE2	Harakalova *et al* (2015)
HFE2d	Harakalova *et al* (2015)
HSPB6	Harakalova *et al* (2015)
ISL1	Harakalova *et al* (2015)
ITGB1BP2	Harakalova *et al* (2015)
JUP	Harakalova *et al* (2015)
KCNH2	Harakalova *et al* (2015)
LAMP2/MYH7e	Harakalova *et al* (2015)
MGME1	Harakalova *et al* (2015)
MURC	Harakalova *et al* (2015)
MYL2	Harakalova *et al* (2015)
NEBL	Harakalova *et al* (2015)
NKX2-5	Harakalova *et al* (2015)
PGM1	Harakalova *et al* (2015)
PKP2/DSC2f	Harakalova *et al* (2015)
PKP2/DSPg	Harakalova *et al* (2015)
PNPLA2	Harakalova *et al* (2015)
PNPLA2h	Harakalova *et al* (2015)
POLG	Harakalova *et al* (2015)
POMT1	Harakalova *et al* (2015)
POU1F1	Harakalova *et al* (2015)
RET	Harakalova *et al* (2015)
RYR2	Harakalova *et al* (2015)
SCARB2	Harakalova *et al* (2015)
SKI	Harakalova *et al* (2015)
SYNE1/SYNE2i	Harakalova *et al* (2015)
TERT	Harakalova *et al* (2015)
TMEM43	Harakalova *et al* (2015)
TMEM70	Harakalova *et al* (2015)
TXNRD2	Harakalova *et al* (2015)
UBR1	Harakalova *et al* (2015)
XPNPEP3	Harakalova *et al* (2015)
ABCC9	Harakalova *et al* (2015)
BRCC3	Harakalova *et al* (2015)

DCM candidate gene	Reference
DOLK	Harakalova *et al* (2015)
DSC2	Harakalova *et al* (2015)
MTCP1	Harakalova *et al* (2015)
MYPN	Harakalova *et al* (2015)
PKP2	Harakalova *et al* (2015)
SLC22A5	Harakalova *et al* (2015)
SYNE2	Harakalova *et al* (2015)
MT-RNR1	Harakalova *et al* (2015)
MT-RNR2	Harakalova *et al* (2015)
MT-ATP6	Harakalova *et al* (2015)
MT-ATP8	Harakalova *et al* (2015)
MT-CYB	Harakalova *et al* (2015)
MT-CO1	Harakalova *et al* (2015)
MT-CO2	Harakalova *et al* (2015)
MT-CO3	Harakalova *et al* (2015)
MT-ND1	Harakalova *et al* (2015)
MT-ND2	Harakalova *et al* (2015)
MT-ND3	Harakalova *et al* (2015)
MT-ND4	Harakalova *et al* (2015)
MT-ND4L	Harakalova *et al* (2015)
MT-ND5	Harakalova *et al* (2015)
MT-ND6	Harakalova *et al* (2015)
MT-TA	Harakalova *et al* (2015)
MT-TR	Harakalova *et al* (2015)
MT-TN	Harakalova *et al* (2015)
MT-TD	Harakalova *et al* (2015)
MT-TC	Harakalova *et al* (2015)
MT-TE	Harakalova *et al* (2015)
MT-TQ	Harakalova *et al* (2015)
MT-TG	Harakalova *et al* (2015)
MT-TH	Harakalova *et al* (2015)
MT-TI	Harakalova *et al* (2015)
MT-TL1	Harakalova *et al* (2015)
MT-TL2	Harakalova *et al* (2015)
MT-TK	Harakalova *et al* (2015)
MT-TM	Harakalova *et al* (2015)
MT-TF	Harakalova *et al* (2015)
MT-TP	Harakalova *et al* (2015)
MT-TS1	Harakalova *et al* (2015)
MT-TS2	Harakalova *et al* (2015)
MT-TT	Harakalova *et al* (2015)
MT-TW	Harakalova *et al* (2015)
MT-TY	Harakalova *et al* (2015)
MT-TV	Harakalova *et al* (2015)

Supplemental Table 5. Genes whose promoter regions show differential acetylation (H3K27ac concentration) between DCM patients and control donors (available upon request).

Supplemental Table 6. Differentially acetylated regions between DCM patients and control donors classified as distal enhancers (available upon request).

Supplemental Table 7. Motifs predicted to bind on candidate distal enhancers differenially acetylated between DCM patients and control donors.



Chapter 5

Impact of carotid atherosclerosis loci on cardiovascular events

Hemerich D, van der Laan SW, Tragante V, den Ruijter HM,
de Borst GJ, Pasterkamp G, de Bakker PI, Asselbergs FW

2015. Atherosclerosis
DOI: 10.1016/j.atherosclerosis.2015.10.017

5.1 Abstract

Genome-wide association studies (GWAS) have identified six single-nucleotide polymorphisms (SNPs) related to carotid intima media thickness (cIMT) or plaque. However, whether these loci relate to other vascular diseases and subsequent vascular events is unclear. We tested six SNPs (rs4888378, rs11781551, rs445925, rs6601530, rs17398575 and rs1878406) for association with subclinical atherosclerotic measures (cIMT, plaque presence and ankle-brachial index), as well as ischemic stroke, abdominal aortic aneurysm, peripheral or coronary artery disease (CAD) in the Second Manifestations of ARTerial disease (SMART) cohort. Four SNPs were associated with cIMT and two with plaque ($p < 0.05$). One SNP was also significantly associated to CAD (rs1878406, OR = 1.24, 95% CI = 1.08-1.42, $p=2\times10^{-3}$). A genetic risk score (GRS) based on the cIMT-related SNPs was associated to increased risk of cIMT itself ($p=1\times10^{-3}$), but not to other secondary outcomes or vascular events during follow-up ($p = 0.86$). In addition to replicating previously published associations for cIMT, we confirmed a nominally significant effect between the GRS and cIMT.

5.2 Introduction

Coronary artery disease (CAD) and stroke are leading causes of death worldwide [315], influenced by common genetic factors. Subclinical atherosclerosis, a thickening of the artery wall caused by the deposition of cholesterol material, often precedes these events [316, 317]. Carotid intima-media thickness (cIMT) and plaque, measures of subclinical atherosclerosis, have been shown to predict incident atherosclerosis-related cardiovascular disease [318-320]. In this sense, genetic association studies may identify susceptibility genes and pathways involved in the initiation and early phases of these diseases.
Recently, genome-wide association studies (GWAS) identified 6 single-nucleotide polymorphisms (SNPs) [321, 322] associated to cIMT or plaque. The extent to which these loci are related to other subclinical phenotypes, clinically manifested vascular diseases and subsequent cardiovascular events is unclear. Therefore, we aimed to demonstrate the external validity of these findings by testing whether these SNPs relate to cIMT, plaque, ankle-brachial index (ABI), ischaemic stroke (IS), abdominal aortic aneurysm (AAA), peripheral artery disease (PAD), and CAD in the SMART (Second Manifestations of ARTerial disease) cohort. We also modelled the 3 cIMT-related SNPs as a multilocus genetic risk score (GRS) and tested for association with CAD, IS, AAA, PAD, ABI, and cIMT itself, as well as with recurrent vascular events.

5.3 Methods

5.3.1 Study Populations and Phenotyping

We used data from the SMART cohort, consisting of patients from the University Medical Center Utrecht (UMCU), the Netherlands, included on the basis of manifest atherosclerotic vascular disease or the treatment of atherosclerotic risk factors [323, 324] (Supplemental Table S1). From the 8,210 patients included in the study, 3,743 had CAD, 640 had AAA, 1,726 had PAD, and 1,764 had IS, with overlap among traits. Patients free of cardiovascular disease who had one or more risk factors for cardiovascular disease and were included in the SMART follow-up phase, served as a control group (n=1,981). All patients provided informed consent, and the Medical Ethics Committee of the UMCU approved the study.

5.3.2 SNP selection, genotyping and quality control

Based on a GWAS meta-analysis for cIMT and plaque [321], we selected 3 cIMT-associated SNPs (rs11781551, rs445925, and rs6601530), and 2 plaque-associated SNPs (rs17398575 and rs1878406). A second study associated rs4888378 with cIMT [322] and was added to this analysis, resulting in a total of 4 cIMT-associated and 2 plaque-associated SNPs. Community standard quality control (QC) [325] was applied in all 6

SNPs. We excluded rs445925 from further analysis as it was out of Hardy-Weinberg Equilibrium (p=4.14x10^{-9}).
Wet-lab genotyping was carried out at KBiosciences (Hertfordshire, UK, www.kbioscience.co.uk) whose personnel were blinded to patient status, using the proprietary KASPar PCR technique and TaqMan. Genotype calling was done using an automated system, with results checked manually using SNPviewer software.

5.3.3 Statistical Analysis

Single SNP and GRS analyses were performed using linear and logistic regression models where appropriate, adjusting for sex and age. We considered the presence of plaque when cIMT>1.1, following the same criteria as in Bis *et al.* [321].
For each individual in our cohort we constructed an unweighted GRS using PASW Statistics 21 (SPSS, Inc., 2012, Chicago, IL, USA, www.spss.com). The GRS was calculated as the sum of the number of risk alleles at the 3 cIMT-related SNPs (rs11781551, rs6601530, and rs4888378).
We used Cox proportional hazards model to analyze the association between GRS and clinical events during follow-up, considering age, sex and inclusion criteria as covariates. To estimate quantitative effect sizes of the GRS on subsequent disease risk, we divided individuals into quartiles according to the GRS distribution and computed hazard ratios (HR) between the quartiles, with the first quartile as reference.
Since all 5 SNPs that passed QC were previously associated with cIMT or plaque at genome-wide significance ($p<5x10^{-8}$), we considered a $p<0.05$ threshold (for the same risk allele in the same direction previously reported) as significant for these two phenotypes. Correcting for multiple testing, the Bonferroni threshold defined for the other five tested vascular beds (IS, AAA, CAD, PAD, and ABI) was $p<2x10^{-3}$.

5.4 Results and Discussion

In this study SMART comprised a total of 8,210 individuals (6,229 cases and 1,981 controls), aged 17 to 83 years, whose majority (67.5%) is male.
The single-SNP analysis for association with cIMT and plaque in SMART (Table 1) resulted in 4 SNPs associated with cIMT and 2 with plaque ($p<0.05$), all showing a concordant direction of effect with the originally reported associations (binomial p=0.03). This independent replication supports the role of these loci as genetic determinants of cIMT and plaque.
In order to test the relation between these loci and other vascular beds, we tested their association with IS, AAA, CAD, PAD, and ABI (Table 2). One variant (rs1878406) was significantly associated with CAD (odds ratio [OR]=1.24, 95% confidence interval [C.I.]=1.08-1.42, p=2x10^{-3}). This SNP is located 8.5 kb from *EDNRA*, a gene related

Table 1. cIMT/plaque associated SNPs reported by literature and the association results for cIMT/plaque in SMART.

Reported by literature								**This study**			
Plaque associated variants											
Locus	*SNP*	*Chr*	*Alleles*	*EAF*	*HWE*	*OR (95%CI)*	*P*	*EAF*	*OR (95%CI)*	*P (cIMT)*	*P (plaque)*
EDNRA	rs1878406	4	T/C	0.13	0.5535	1.22 (1.15–1.29)	6.90×10^{-12}	0.16	1.11 (0.98-1.26)	**0.01**	**0.05**
PIK3CG	rs17398575	7	A/G	0.25	0.01141	1.18 (1.12–1.23)	2.30×10^{-12}	0.24	1.17(1.04-1.30)	**0.009**	**0.004**
cIMT associated variants											
Locus	*SNP*	*Chr*	*Alleles*	*EAF*		*ß (95%CI)*	*P*	*EAF*	*ß (95%CI)*	*P (cIMT)*	*P (plaque)*
PINX1	rs6601530	8	G/A	0.45	0.4141	0.0078	1.70×10^{-8}	0.46	0.0009	0.404	0.176
ZHX2	rs11781551	8	A/G	0.48	0.5388	-0.0078	2.40×10^{-11}	0.45	-0.0071	**0.033**	0.133
BCAR1-CFDP1-TMEM170A	rs4888378	16	A/G	0.43	0.003173	-0.0045	7.24×10^{-6}	0.39	-0.006873	**0.039**	0.496

SNP: single-nucleotide polymorphism. Alleles: effect and non-effect alleles. EAF: effect allele frequency. Results listed in bold are nominally significant (p=0.05). HWE: Hardy-Weinberg Equilibrium.

to endothelial dysfunction. The endothelin receptor is a target to reduce blood pressure, given the vasoconstrictor role of endothelins in blood pressure elevation and vascular hypertrophy [326]. Interestingly, genetic variations in this gene have also been associated with atherosclerosis in hypertensive patients [327] and with ambulatory blood pressure [328]. Convincingly, previous reports also linked the same locus to CAD [321, 329]. This supports the hypothesis that *EDNRA* might affect atherosclerosis causing changes in blood pressure and thereby increasing the risk on CAD. SNP rs6601530 showed a nominally significant association to CAD (p=0.04) and ABI (p=0.006). This SNP locates in an intron of PINX1, encoding Pin2-interacting protein 1, a telomerase inhibitor [330] that relates to chromosomal segregation in mitosis [331] and has been implicated in cancer [332, 333]. A recent study [334] found that SNP rs7840785, also located in the intron of PINX1, was significantly associated with right carotid IMT (p=0.0003) in a non-European population. The same study also conducted a gene-based analysis in which PINX1 was significantly associated with right carotid IMT. Even after removing the significant SNP rs7840785, PINX1 was still significantly associated in the overall sample ($p=1\times10^{-7}$). However, after correcting for multiple testing, we did not find significant associations with IS, AAA, PAD, or ABI.

We further tested the combined effects of the cIMT-related SNPs in a multilocus GRS for association with disease risk (Supplemental Table S2). We confirmed the association between the GRS and cIMT (p=0.04) [321]. We did not find a significant association between the risk score and IS, AAA, CAD, PAD, or ABI. This can be due to the small effect the combined SNPs confer on the susceptibility of disease, or cardiovascular traits. We also analyzed the association between the GRS and vascular events during follow-up using Cox proportional hazard models (HR=0.977, 95% C.I. 0.93-1.07, survival curves in Supplemental Figure S1). We found no increased risk of vascular events during follow-up when comparing individuals in the highest GRS quartile with those in the lowest quartile (p=0.86). We may have had limited power due to the modest number of secondary events to find such an association, if it exists.

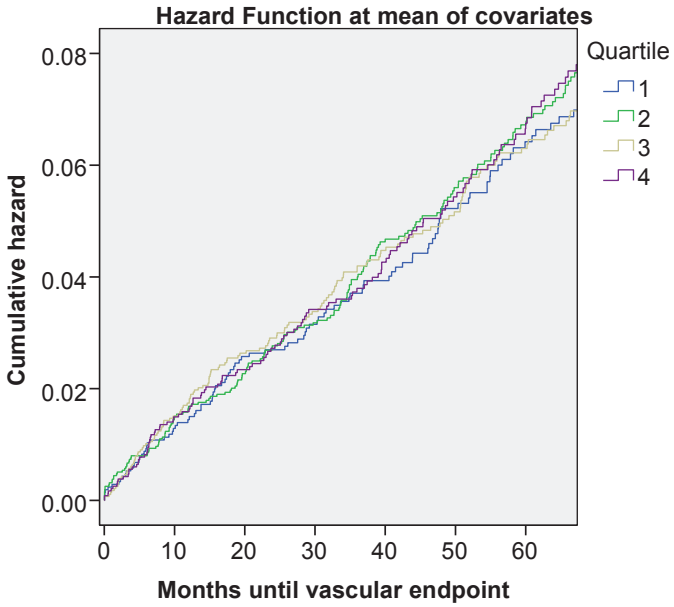
The results showed in this study confirm the original findings [321, 322], adding to the evidence of the effect these loci have on cIMT or plaque presence. The GRS based on cIMT-related SNPs did not show a significant effect on other vascular beds, or secondary vascular events, suggesting against pleiotropy. An explanation for this finding could be that risk factors and their underlying genetic background may impact differently on each vascular bed in the atherogenic process. In conclusion, our findings provide support for previously claimed SNP associations for cIMT and plaque, specifically highlighting the role of rs1878406 in both atherosclerosis and CAD. This should motivate further research focused on the underlying mechanisms involved.

Table 2. Single SNP association results for IS, AAA, CAD, PAD and ABI in SMART.

		IS N=1,764		**AAA N=640**		**CAD N=3,743**		**PAD N=1,726**		**ABI N=7,953**	
SNP	**Alleles**	*OR (95%CI)*	*P*	*OR (95%CI)*	*P*	*OR (95%CI)*	*P*	*OR (95%CI)*	*P*	*ß (95%CI)*	*P*
rs1878406	T/C	1.02 (0.88-1.19)	0.791	1.10 (0.88-1.39)	0.407	1.24 (1.08-1.42)	**0.002**	1.14 (0.98-1.34)	0.094	-0.001 (-0.01-0.01)	0.886
rs17398575	A/G	1.07 (0.95-1.21)	0.260	1.03 (0.85-1.25)	0.743	1.01 (0.90-1.13)	0.905	1.15 (1.01-1.30)	0.035	-0.003 (-0.01-0.00)	0.314
rs6601530	G/A	1.04 (0.93-1.16)	0.523	1.06 (0.89-1.25)	0.514	1.11 (1.00-1.22)	0.040	1.09 (0.97-1.22)	0.132	-0.008 (-0.01-0.00)	0.006
rs11781551	A/G	1.00 (0.90-1.11)	0.975	1.12 (0.95-1.33)	0.170	1.03 (0.94-1.14)	0.515	1.06 (0.95-1.19)	0.307	0.000 (-0.01-0.01)	0.923
rs4888378	A/G	1.00 (0.90-1.11)	0.966	1.13 (0.95-1.33)	0.160	0.97 (0.88-1.07)	0.569	0.97 (0.86-1.08)	0.546	0.003 (0.00-0.01)	0.213

5.5 Supplemental Material

5.5.1 Supplemental Figures



Supplemental Figure S1. Survival plot stratified per quartile of the genetic risk score for cIMT in association with composite clinical endpoints. The GRS is not associated with an increased risk of developing recurrent vascular events when comparing individuals in the highest GRS quartile with those in the lowest quartile (p=0.86).

5.5.2 Supplemental Tables

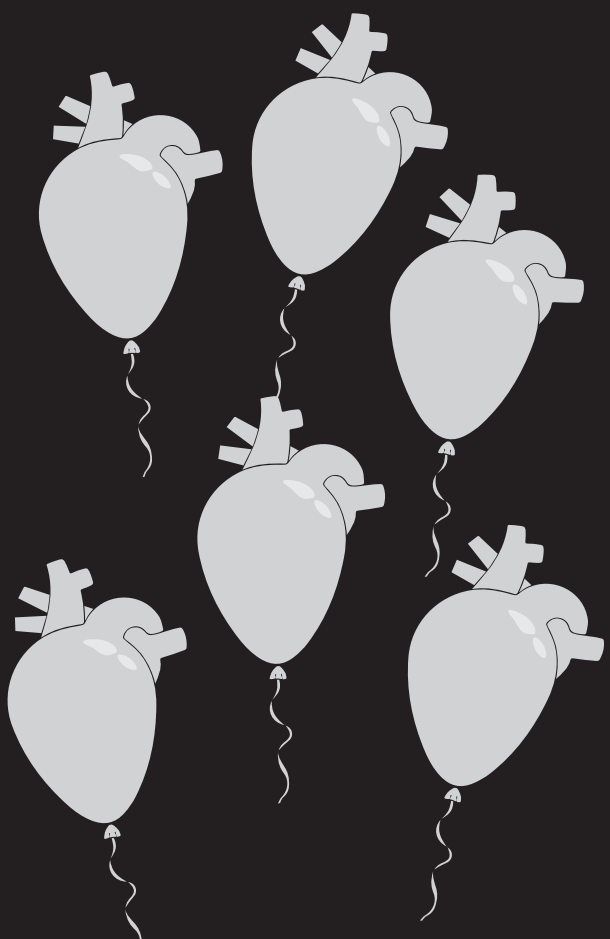
Supplemental Table S1. Baseline characteristics of SMART. *Cerebrovascular disease* history includes ischemic stroke and transient ischemic attack (TIA). *Cardiovascular disease* history includes coronary artery disease, myocardial infarction, percutaneous coronary intervention, and coronary artery bypass graft. CAD is defined as having one or more events in the past (such as fatal or non-fatal MI) or revascularization procedure (such as Coronary Artery Bypass Graft or Percutaneous Coronary Intervention), or one or more events during the period of follow-up. Peripheral disease is defined as resting ABPI ≤ 0.90 or post exercise ABPI decreasing 20% or more in at least one leg, with signs of intermittent claudication, rest pain or gangrene/ulcers. *Type 2 diabetes* and *hypertension* include all individuals with diagnosed type 2 diabetes or hypertension, respectively, and/or those on appropriate medication. *Current smokers* include all individuals smoking up to 6 months until the surgery date. *BMI*: body mass index, kg/m2. *eGFR*: estimated glomerular filtration rate was based on the Modification of Diet in Renal Disease formula, mL/min/1.73 m^2. *TC*: total cholesterol, *LDL*: low-density lipoprotein, *HDL*: high-density lipoprotein, *TG*: triglycerides; all lipids are in mmol/L. *Antihypertensives* include all anti-hypertension medication. *LLDs*: lipid-lowering drugs including statins, and other lipid-lowering drugs. *Anticoagulants* include clopidogrel, dipyridamole, acenocoumarin, ascal, and anti-platelet drugs. *N* is the total sample size.

Characteristic	**SMART N=8,210**
Male, N (%)	5,541 (67.5)
Age, years (sd)	57 (12)
History, N (%)	
Cerebrovascular	1,779 (21.7)
Cardiovascular	3,788 (46.1)
Peripheral	1,751 (21.3)
Risk factors	
Type 2 diabetes	1,655 (20.2)
Hypertension	2.223 (27.1)
Current smokers	2,610 (31.8)
BMI	26.8 [22.49-31.17]
TC	5.2 [3.8-6.6]
HDL	1.26 [0.87-1.65]
TG	1.67 [0.71-2.63]
LDL	3.17 [1.98-4.36]
Medication, N (%)	
Antihypertensives	3,709 (45.2)
LLDs	3,574 (43.5)
Anticoagulants	4,555 (55.5)

Supplemental Table S2. Association between the genetic risk score and CAD, IS, PAD, AAA, ABI and cIMT in SMART.

	Overall[a] (*n* = 8,210) p-value (OR or Exp(ß))
CAD	0.308 (1.19)
IS	0.647 (1.09)
PAD	0.340 (1.21)
AAA	0.058 (1.74)
ABI	0.494 (0.99)
cIMT	**0.041** (0.97)

CAD: coronary artery disease; IS: ischaemic stroke; AAA: abdominal aortic aneurysm; PAD: peripheral artery disease; HR: hazard ratio. Covariates: age and sex. Results listed in bold are nominally significant (p=0.05). [a]Considering genetic risk score as a continuous variable.



Chapter 6

Use of tissue-specific genetic risk scores on the investigation of disease progression profiles of patients with oucomes induced by hypertension

Hemerich D, van Setten J, van der Laan SW, Kofink D,
Munroe PB, Tragante V, Asselbergs FW

Manuscript in preparation.

6.1 Abstract

High blood pressure is an established risk factor for a multitude of diseases. Genome-wide association studies have successfully identified over nine hundred loci that contribute to natural variation in blood pressure, including hypertension. However, the mechanisms and extent through which these loci contribute to blood pressure leading to hypertension and all its subsequent outcomes remain elusive. We hypothesize that each individual's genetic profile provides a specific combination of tissue-specific regulatory variants, identifying subphenotypes of hypertension, leading to different hypertension related outcomes. We searched for associations of DNA variation with tissue-specific regulation highlighted by the presence of H3K27ac, a mark of enhancers and promoters, in tissues implicated by hypertension-induced outcomes. We calculated genetic risk scores for each subset of blood pressure-associated variants acting on specific tissues and correlated these with ten outcomes using participants from the UK Biobank. In eight of the nine hypertension-induced outcomes analyzed, tissue-specific variant sets showed significant risk (p<0.05) on tissue-specific outcomes. In three of these cases, tissue-specific sets result in increased risk on outcomes related to that tissue, compared to other tissue-specific sets of variants (kidney-specific regulatory variants implicated in high risk of kidney failure (pulse pressure (PP) p=0.001, systolic blood pressure (SBP p=0.001), diastolic blood pressure (DBP) p=0.005), pancreas-specific variants implicated on outcomes including diabetes (PP p=9.89x10-10), and LV-specific regulatory variants associated with dilated cardiomyopathy (DBP p=0.005, PP p=0.018, SBP p=0.032). Increasing sample sizes and expanding this analysis to a broader range of histone modifications and tissues might help unravel more cell-type specific sets of variants accounting risk for tissue-specific outcomes. This may lead to a better understanding of disease progression and more personalized intervention for better care.

6.2 Introduction

High blood pressure (BP) is a highly prevalent chronic disorder, considered the single highest contributing risk factor to disease burden and premature mortality [335]. Elevated systolic (SBP) and/or diastolic BP (DBP) increases the risk of several disorders, including heart failure (HF), coronary artery disease (CAD), peripheral arterial disease (PAD), abdominal aortic aneurysms, fatty liver disease and kidney failure [335-340]. Genome-wide association studies (GWAS) have identified over 900 genetic regions associated with SBP, DBP, pulse pressure (PP) and/or hypertension to date [112-124, 134, 135, 137, 144, 341-345]. Over 90% of these variants are located in intronic or intergenic regions [85], and might interrupt the action of regulatory elements crucial in relevant cell-types. The identification of genes affected by regulatory elements is a challenge, given distal regulatory elements such as enhancers, insulators and silencing elements are often located far away from the genes they control [69, 346]. Moreover, genes controlled by regulatory elements affected by DNA variation act in different pathways, and disturbance in gene expression is often reflected in a variety of outcomes of which hypertension is a risk factor. BP-associated variants altering the expression of genes in the heart have a higher likelihood of affecting disease progression through heart-mediated processes rather than kidney-mediated processes. According to this example, some patients may develop coronary artery disease while others may suffer kidney failure, although initially they are considered only hypertensive patients (Figure 1).

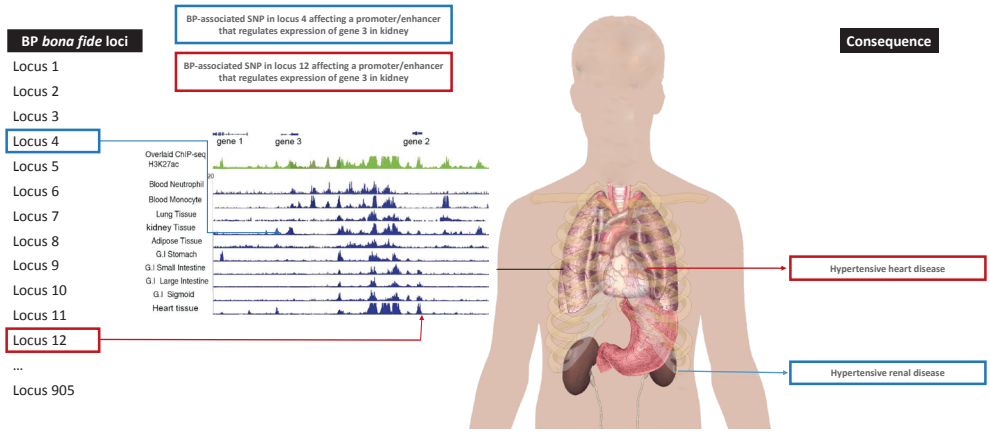


Figure 1. Illustration of our hypothesis. A number of *bona fide* SNPs has been associated with BP levels. A subset of these SNPs overlaps with H3k27ac peaks (a mark of active enhancers and promoters) in different tissues. The variant might affect the regulatory element it overlaps, and consequently the expression of the gene regulated by this element. Carriers of particular sets of tissue-specific regulatory SNPs might manifest consequences of hypertension in a different way, with different progression and outcome profiles.

We hypothesize that a combination of tissue-specific BP-associated regulatory variants might confer greater risk for a high BP-induced outcome affecting that tissue, than BP-associated variants specifically acting on other tissues not primarily implicated in the outcome. In this study, we investigated the impact of genetic risk profiles of disease progression of BP-associated variants in specific tissues. The results provided by applying this strategy on a broader range of tissues and histone modifications may contribute to the practice of personalized medicine for complex diseases.

6.3 Methods

6.3.1 SNP selection

We obtained 905 loci identified by GWAS of 4 BP traits (DBP, SBP, PP and hypertension). A list summarizing all BP-associated variants can be found on Hemerich *et al* (2018) [85].

6.3.2 Maps of regulatory elements: enhancers and promoters

Active regulatory regions (promoters or enhancers, for instance), are marked by specific histone modifications, such as monomethylation of the fourth residue of histone H3 (H3K4me1) and acetylation of the 27th residue of H3 (H3K27ac) [15, 347-350]. Chromatin immunoprecipitation followed by sequencing (ChIP-seq) captures these maps of histone modification. The co-localization of GWAS variants with tissue-specific maps of histone modifications can be used to identify which DNA variants potentially act by disturbing regulatory elements [194, 351]. Thus, we obtained publicly available ChIP-seq data on H3K27ac histone modification on tissues from ENCODE [60] and Roadmap Epigenomics Consortium [66] on tissues mostly affected by hypertension-induced outcomes. These tissues are the left ventricle (LV), right ventricle (RV), right atrium (RA), aorta, adrenal gland, kidney and pancreas. We downloaded alignment files in hg-19 coordinates of both treatment and control, from https://egg2.wustl.edu/roadmap/data/byFileType/alignments/consolidated/. Duplicated, unmapped reads and reads with mapping quality less than 5 (-b -F 4 -q 5) were removed using Samtools v1.3 [352]. Peak calling was performed using MACS v2.1.0 [207] using the respective input samples, -gsize=hg -nomodel parameters, and estimated fragment sizes (-extsize) predicted by PhantomPeakQualTools v1.1 [9].

6.3.3 Candidate tissue-specific regulatory variants

We used the method described by Trynka *et al* (2013) [207] to identify subsets of BP-associated variants acting on regulatory elements in critical cell-types. According to this method, a locus is defined by identifying SNPs in tight linkage disequilibrium (LD) with each associated variant (r^2>0.8), using data from the 1000 Genomes Project [353]. Each variant of this expanded set is scored based on the distance and height of the

nearest ChIP-seq peak, which in this study highlights an active enhancer or promoter. If the physical distance to the nearest peak is more than 2.5 kilobases (kb), then the score is set to 0 to remove any confounding distal effects. The statistical significance of cell-type specificity is assessed by deriving a 95th percentile threshold based on the permutation of 10,000 matched sets of SNPs not associated with the phenotype.

6.3.4 Study design and participants

The UK Biobank (UKB) is a large and detailed prospective study with over 500,000 British individuals aged 40–69 years at recruitment, identified through primary care lists, who accepted an invitation to attend one of the 22 assessment centers that were serially established across the UK between 2006 and 2010 [354]. At recruitment, detailed information was collected via a standardized questionnaire on socio-demographic characteristics, health status and physician-diagnosed medical conditions, family history and lifestyle factors. Selected physical and functional measurements were obtained including height, weight, waist-hip ratio, SBP and DBP. The UKB data were subsequently linked to Hospital Episode Statistics (HES) data, as well as national death and cancer registries. The HES data available for the current analysis covers all hospital admissions to NHS hospitals in England and Scotland from April 1997 to March 2015, with the Scottish data dating back as early as 1981. HES uses International Classification of Diseases ICD 9 and 10 to record diagnosis information, and OPCS-4 (Office of Population, Censuses and Surveys: Classification of Interventions and Procedures, version 4) to code operative procedures. Death registries include all deaths in the UK up to January 2016, with both primary and contributory causes of death coded in ICD-10. Genotyping of UK Biobank participants was undertaken using a custom-built genome-wide array (the UK Biobank Axiom array: http://www.ukbiobank.ac.uk/wp-content/uploads/2014/04/UKBiobank-Axiom-Array-Datasheet-2014.pdf) of ~826,000 markers. Genotyping was done in two phases. 50,000 subjects were initially typed as part of the UK BiLEVE project . The rest of the participants were genotyped using a slightly modified array. Imputation to ~92 million markers was subsequently carried out using the Haplotype Reference Consortium (HRC) [355] and UK10K/1000Genomes haplotype resource panels.

The determination of phenotypes used for analyses in this study was based on ICD codes, extracted directly from the phenotype file (data fields #41202 and #41204). We extracted case/control status for CAD (defined as percutaneous transluminal coronary angioplasty (PTCA), or coronary artery bypass graft (CABG), data field #41200, hospital admission OPCS-4 K40–K46, K49, K50.1, or K75; 4,412 cases), MI (ICD codes I21-I24 and I25.2; 11,316 cases), hypertensive heart disease (I11.0, I11.9; 161 cases), heart failure (I50.0, I50.1 and I50.9; 5,094 cases), dilated cardiomyopathy (I42.0; 617 cases), aortic aneurysm and dissection (I71; 1,286 cases), renal failure and others (N00-N29; 13,357 cases), hypertensive renal disease (I12.0, I12.9; 1,443 cases) and diseases affecting

pancreas, including diabetes (E10-E1, K85 and K86; 19,689 cases). We used bipolar disorder as a control, grouping participants from UKB with ICD code F31 (965 cases).

6.3.5 Genetic association and construction of genetic risk scores (GRS)

We calculated a GRS for each set of tissue-specific variants, implicating regulation on organs that are known to be affected by hypertension-induced outcomes (Supplemental Table 1). According to these criteria, the sets of variants with a high regulatory score used for further analysis are specific to LV, RV, right atrium, aorta, kidney and the adrenal gland. In order to retain only independent variants, we used the *clump* function from PLINK v1.07, keeping for each locus (defined by a 500kb window) only the best SNPs using an r^2 of 0.2 and 1000 Genomes CEU v3 as reference panel. We then assessed associations between these SNPs and BP components (DBP, SBP and PP) in UKB samples of European descent, including year of birth, age squared, sex and 40 principal components (PCs) as covariates. We next computed individual risk scores (GRS) for each tissue and BP component, using beta estimates from UKB for the tissue-specific independent BP-associated variants (either the lead associated SNP or variants in high LD, r^2>0.8) (Supplemental Table 1). For each tissue, we identified a set of SNPs significantly associated with DBP, SBP and PP, using a Bonferroni-corrected significance threshold of 0.05 / (810 SNPs * 3 BP components) » $2.1*10^{-5}$. Based on these SNPs, tissue-specific GRS were generated by calculating the sum of their effect allele dosages weighted by the corresponding regression coefficients from the association analyses, using the *crossprod* function in R (v 3.2.2). The GRS was associated with case/control status of each of the outcomes described above (nine hypertension-induced and one control, bipolar disorder) using a logistic regression model with age, sex, BMI, smoking status (current, past, never), and 10 PCs as covariates.

6.4 Results

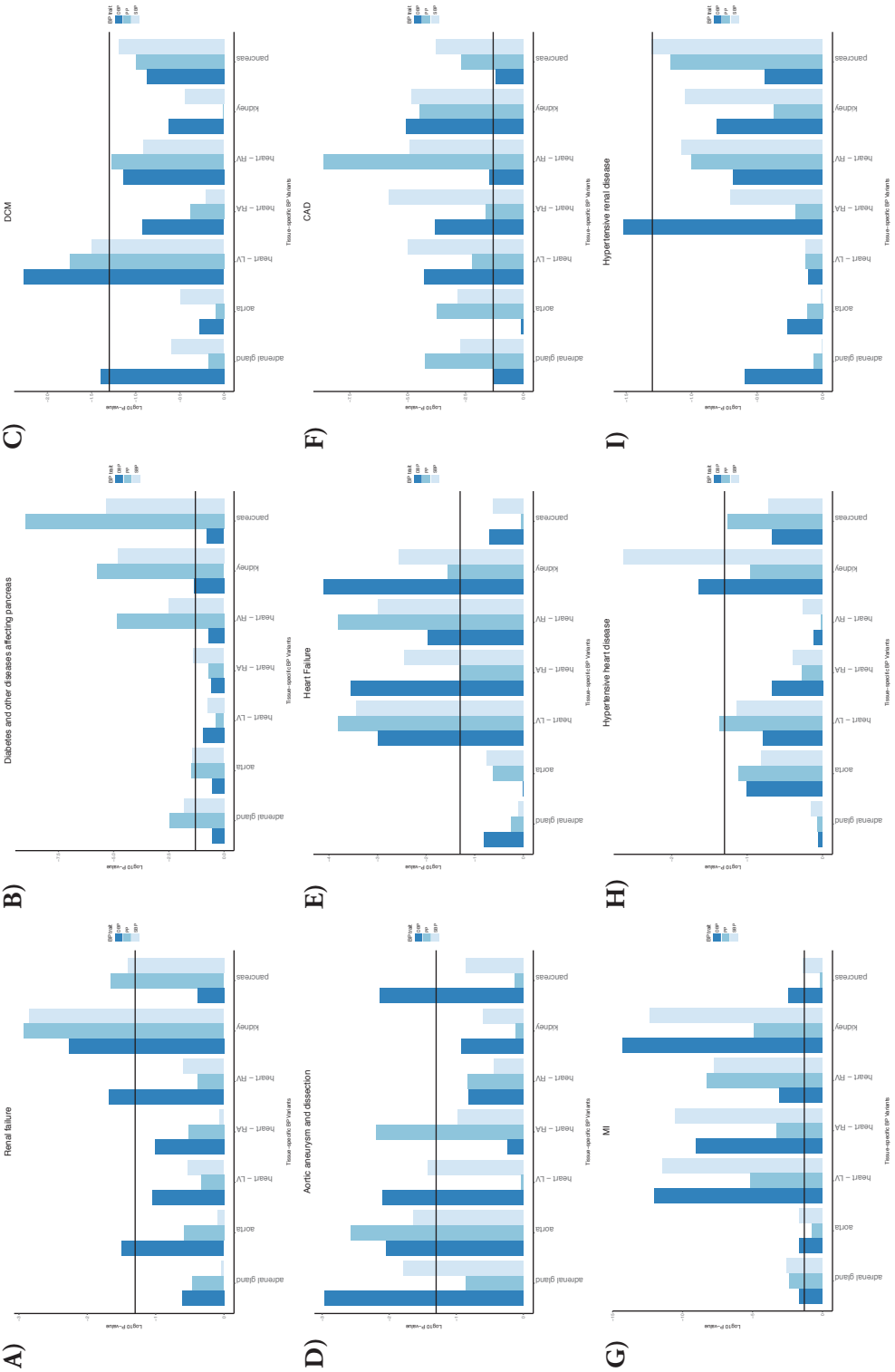
6.4.1 Genetic risk scores of cell-type specific BP-associated SNPs related to hypertension-induced outcomes

We investigated the impact of GRSs on cell-type specific regulatory SNPs and different hypertension-induced outcomes affecting specific tissues, focusing on heart (LV, RV and right atrium), aorta, kidney and adrenal gland (Supplemental Table 1). We observed that in eight out of the nine hypertension-induced outcomes analyzed, tissue-specific sets have significant risk (p<0.05) on tissue-specific outcomes (Figure 2A, B, C, D, E, F, G and H). In three of these cases, tissue-specific sets have increased risk on outcomes related to that tissue, compared to other tissue-specific sets of variants (Figure 2A, B and C). These cases are kidney-specific regulatory variants implicating high risk of kidney failure (PP p=0.001, SBP p=0.001, DBP p=0.005) (Figure 2A),

pancreas-specific variants on outcomes including diabetes (PP p=9.89x10^{-10}) (Figure 2B), and LV-specific regulatory variants on DCM (DBP p=0.005, PP p=0.018, SBP p=0.032) (Figure 2C). We also observed that only one of the BP cell-type specific sets of variants (in pancreas, SBP p=0.01) showed significance (p<0.05) of association to our control outcome, bipolar disorder (Figure 2J). Overall, the majority of cases tested (9/10) did not associate to the control outcome analyzed, whereas in most cases (8/9) tissue-specific variants showed risk for tissue-specific outcomes. In three of such cases (LV, kidney and pancreas), tissue-specific variants showed increased risk for tissue-specific outcomes, compared to regulatory SNPs affecting other BP-specific tissues - which are not the organ where the outcome in question primarily manifests (Figure 2A, B and C).

6.5 Discussion

In this study, we investigated the impact of tissue-specific regulatory variants in hypertension-induced outcomes, utilizing GRSs to aid in the identification of disease progression profiles. Tissue-specific variants were associated to tissue-specific outcomes in the majority of cases tested (8/9). A tenth test was performed against a control outcome, bipolar disorder. In this last analysis, tissue-specific variants implicating regulation in pancreas associated to one of the BP traits (PP) showed an association to bipolar disorder. This was the only association to the control outcome, out of the 21 tests performed for each outcome (7 tissues and 3 BP traits). From the nine hypertension-induced outcomes tested, three showed higher association of tissue-specific variants that implicate regulation in the tissue affected by the outcome. These are LV-specific regulatory variants in DCM, kidney-specific regulatory variants in kidney failure, and pancreas-specific variants in outcomes implicating pancreas tissue including diabetes (Figures 1A, B and C). These studies suggest a role of tissue-specific regulatory variants in accounting risk for outcomes that manifest in the same tissue. Some outcomes were shown to be more heterogeneous, such as aortic aneurysm, heart failure, CAD and MI. This suggests that BP-associated regulatory variants that account risk for these outcomes act in a wide range of tissues. Moreover, patients included in the UKB often have a combination of diagnoses and can be concurrently cases for heart-related outcomes as well as kidney-related diseases. Hypertensive renal disease and hypertensive heart disease did not show enrichment for tissue-specific sets of variants, which can be due to the limitations of this study. The fact that we did not manage to capture difference across specific tissues in some cases might be due to lack of statistical power, given the number of cases observed in the UKB cohort manifesting some outcomes is limited (Methods). In addition, we only focused on one of the several histone modifications (acetylation of lysine 27 on histone 3) that are so thus far known to highlight regions of activity of regulatory elements. Histones are subject to a vast array of posttranslational modifications including acetylation, methylation, phosphorylation, and ubiquitylation.


A)
Renal failure
B)
Diabetes and other diseases affecting pancreas
C)
DCM
D)
Aortic aneurysm and dissection
E)
Heart Failure
F)
CAD
G)
MI
H)
Hypertensive heart disease
I)
Hypertensive renal disease
BP Mat
DBP
PP
SBP
adrenal gland
aorta
heart – LV
heart – RA
heart – RV
kidney
pancreas
Tissue-specific BP Variants
Log10 P-value

<<< **Figure 2.** Association of the GRSs measures in each set of tissue-specific variants with nine outcomes influenced by hypertension affecting BP-specific tissues (and bipolar disorder as control outcome). The black horizontal line indicates the nominal significance threshold (p=0.05). CAD = coronary artery disease. MI = myocardial infarction. DCM = dilated cardiomyopathy.

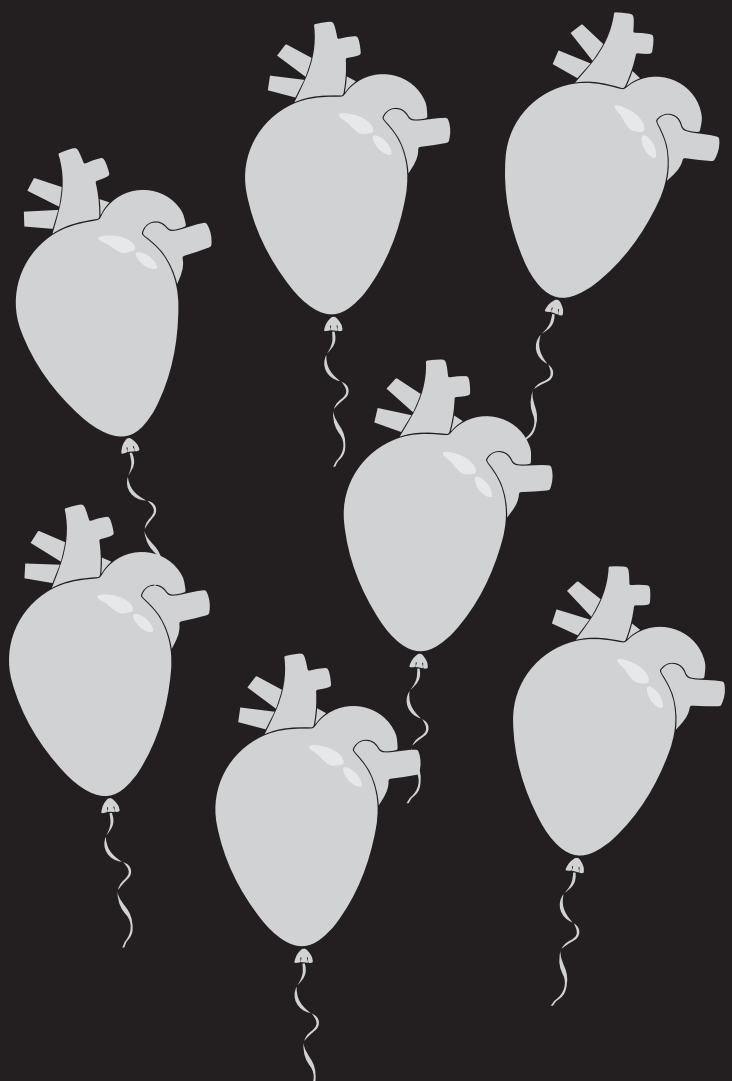
There is a complex interplay between the different covalent modifications occurring on the histone tails, and the hierarchical order of events and dependencies leading to altered gene expression is not yet fully understood [356, 357]. One modification can affect subsequent modifications on histone tails and this crosstalk between active and repressive modifications regulate diverse chromatin functions, such as gene expression [358-360]. Thus, focusing on only one histone modification limits the identification of the full set of potentially regulatory variants, and their risk on disease.

Other reasons for not detecting association of tissue-specific BP-associated variants in outcome-specific tissues in some cases can be due to the fact that simply there is none. According to this reasoning, while there is a genetic component for high BP that acts on a cell-type specific manner, the impact of high BP on subsequent outcomes affecting specific tissues may not be dependent on genetics, but dependent on physiological mechanisms due to the excessive pressure on artery walls. A thorough investigation on a more complete set of tissue-specific BP-associated regulatory variants is therefore needed. Increasing sample sizes is a strategy that can overcome the limitations faced in this study, by increasing both in the number of participants as well as the number of regulatory variants identified in each specific tissue. As more experiments on the investigation of the regulatory landscapes in a variety of tissues are performed, more data will be available for such integration, and knowledge will increase on how regulatory variants act on specific tissues and the interplay of regulatory elements. The further investigation on tissue-specific genetic risk profiles can not only help us understand the disease mechanisms, but also build a basis for tissue-specific, genetic profile-driven treatment.

6.6 Supplemental Material

6.6.1 Supplemental tables

Supplemental Table S1. Tissue-specific variants used in the calculation of the GRS (available upon request).



Chapter 7

Druggability of coronary artery disease risk loci

Tragante V*. Hemerich D*, Alshabeeb M, Brænne I, Lempiainen H, Patel R, den Ruijter HM, Barnes MR, Moore JH, Schunkert H, Erdmann J, Asselbergs FW
* *These authors contributed equally*

Accepted for publication.

7.1 Abstract

Genome-wide association studies (GWAS) have identified multiple loci associated with coronary artery disease (CAD) and myocardial infarction (MI), but only a few of these loci are current targets for on-market medications. To identify drugs suitable for repurposing and their targets, we created two unique pipelines integrating public data on 49 CAD/MI-GWAS loci, drug-gene interactions, side effects and chemical interactions. We first used publicly available GWAS results on all phenotypes to predict relevant side effects, identified drug-gene interactions, and prioritized candidates for repurposing among existing drugs. Secondly, we prioritized gene product targets by calculating a druggability score to estimate how accessible pockets of CAD/MI associated gene products are, then used again the GWAS results to predict side effects, excluded loci with widespread cross-tissue expression to avoid housekeeping and genes involved in vital processes and accordingly ranked the remaining gene products. These pipelines ultimately led to three suggestions for drug repurposing: pentolinium, adenosine triphosphate and riociguat (to target *CHRNB4*, *ACSS2* and *GUCY1A3*, respectively); and three proteins for drug development: *LMOD1*, *HIP1* and *PPP2R3A*.

7.2 Introduction

Coronary artery disease (CAD) is a major cause of death worldwide, leading to a yearly estimated 8.5 million cases of myocardial infarction (MI) [361] and loss of an expected ~150 million disability-adjusted life years globally in 2020 [362]. Current therapeutics for prevention of CAD mainly comprise the control of risk factors, e.g. the prescription of HMG-CoA reductase inhibitors, known as statins, or *PCSK9* inhibitors, to reduce low-density cholesterol (LDL-C) [363-365]. More recently, the CANTOS study has shown that also non-lipid pathways, such as inflammatory processes, also influence atherothrombotic development [366, 367]. In addition, platelet inhibition may be used for prevention of coronary events in certain, high-risk patient groups.

The discrepancy between the overwhelming clinical need and the small number of agents used in the preventive treatment of CAD and MI is largely explained by a high attrition in drug development, which is mostly attributable to unacceptable side effects and/or lack of efficacy [368]. Currently, it is estimated that only one in every 5000 new drug compounds makes it to market [93]. Furthermore, this process may take 10-15 years and costs billions of dollars for conducting clinical trials to clear the stringent requirements set by health agencies around the world [94, 95]. Drug development has therefore become an expensive and difficult process, hindering the clinical implementation of potentially beneficial new drugs. Novel approaches to support drug development have emerged in recent years based on genetic strategies. For example, one may now conduct in-silico druggability analyses on genetic data, using bioinformatics tools, in order to identify approved and already marketed drugs for treating a new phenotype other than the one the drug was originally developed for. This strategy is referred to as drug repositioning or repurposing, an approach proposed and improved in the past 15 years [98, 99], based on new discoveries including, more recently, genetic information [100, 101]. In such case, where an existing drug targets a gene product or pathway of a disease different from the original indication, fewer clinical trials may need to be conducted to alter the label and indicate a treatment for another disease as safety has already been demonstrated. An example of repurposing is sildenafil, initially produced with the expectation of reducing angina, and later found to be effective to treat erectile dysfunction [102] and pulmonary hypertension [103], leading to the subsequent releases of Viagra® in 1998 and Revatio® in 2005 [104]. Other successful examples of repurposing include gemfibrozil, duloxetine, dapoxetine and thalidomide [369] (original indications, repurposed indications and evidence for repurposing available in Suppl. Table 1).

Genome wide association studies (GWAS) have identified multiple independent loci that contribute to the genetic susceptibility of CAD/MI [370-373]. Many of these loci include genes involved in diverse and currently unexplored biological mechanisms. Thereby these loci represent novel drug targets for treatment and prevention of CAD/MI. A key challenge is to prioritize GWAS hits and their products for pharmacological

intervention. In this process, bioinformatics methods may yield novel insights into the potential "druggability" [97] of each of these loci in order to translate genetic knowledge into clinical care.

When assessing the druggability of a GWAS hit, several factors need consideration. A particular gene may not be "druggable", which means that developing a molecule to target this gene product is not feasible, due to the lack of a defined drug binding pocket (known as a pharmacophore), or the druggability cannot be assessed due to unavailable relevant protein structural information. Although a target may be druggable, it still may not be suitable for clinical exploration, as immediate toxicity issues, buffering effects, redundancy, robustness and possible undesired pleiotropic effects in downstream biological pathways need to be clarified. For example, inhibition of the cardiac expressed HERG gene causes severe QT-interval prolongation, which is now screened as a liability in all drug discovery programs [374]. Other adverse events may be more subtle, for example genetic variability in HMGCR has recently been identified as a risk factor for type 2 diabetes (T2D), which partially explains the relationship between statin use and risk of developing T2D [375,376, 377]. Today, GWAS have found associations between thousands of loci and hundreds of phenotypes, thereby enabling a robust exploration of possible pleiotropic effects for any given locus or SNP [378]. Here, we present two unique pipelines integrating currently available public data on GWAS, drug-gene interactions, side effects and chemical interactions. The first pipeline aims to identify approved drugs that may be suitable for repurposing for treatment of CAD/MI, while the second pipeline ranks non-targeted genes for their suitability to be a target for development of new drugs. The pipelines make use of numerous sources of information made available publicly in the past few years.

7.3 Methods

The detailed pipeline, scripts used and example data are available on https://github.com/drugab/drugab. Figures 1 and 2 present a flowchart of the pipelines developed in this project. Supplementary Figure 1 details the pipelines by also presenting data and tools used. We break down the steps in the following sessions.

7.3.1 Loci selection

We obtained data from the CARDIoGRAMplusC4D consortium [370], which reported the largest GWAS to date on CAD/MI, aggregating 63,746 CAD cases and 130,681 controls. This study identified 49 SNPs as showing significant associations for CAD/MI ($p<5 \times 10^{-8}$), and further 104 SNPs considered most important to explain variance (Suppl. Table 9 of the original study [370]).

7.3.2 Pipeline 1: drug-repurposing pipeline (Figure 1)

The drug repurposing pipeline uses publicly available GWAS results from PhenoScanner [379] to predict relevant side/discordant effects of targeting CAD/MI loci, identifies drug-gene interactions using DGIdb [380] and filters most suitable candidates for repurposing among drugs existing in the market.

7.3.2.1 Phenome-wide ranking

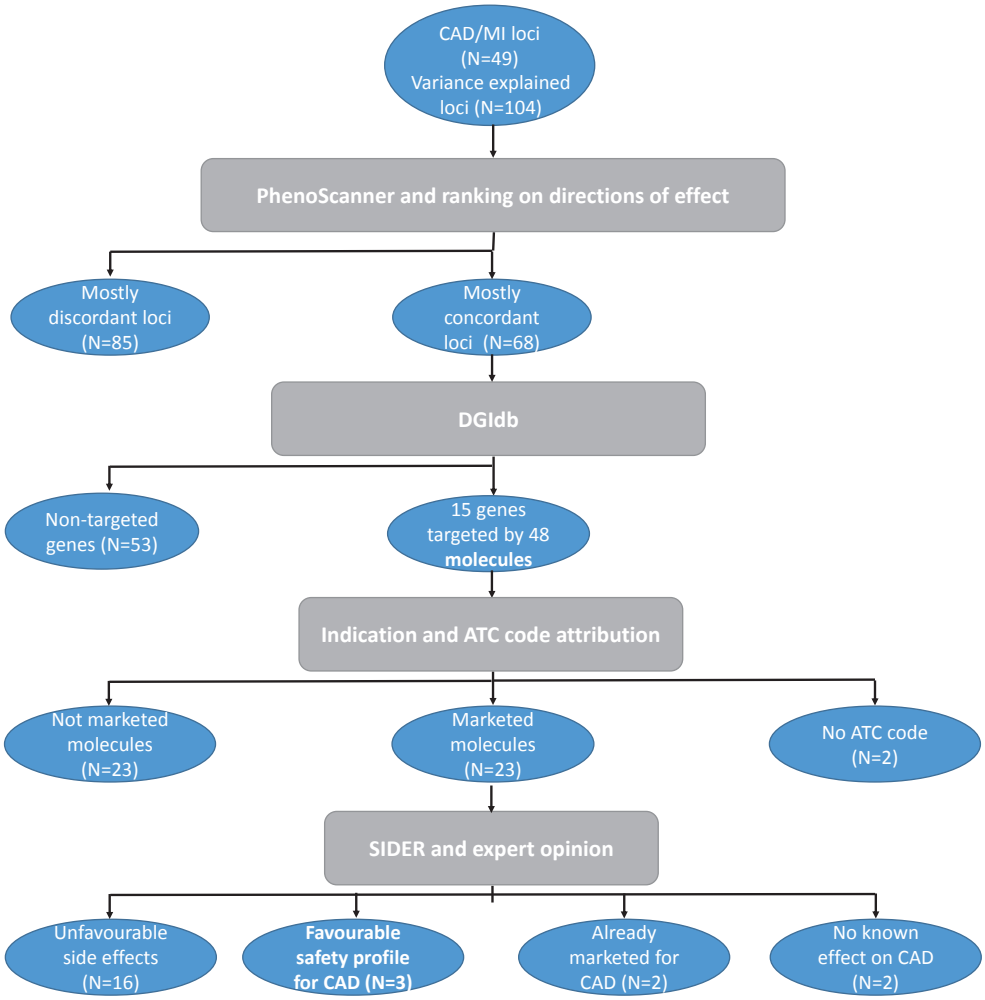


Figure 1. Pipeline 1 to evaluate druggability of CAD/MI loci for repurposing of currently available drugs. It starts with a set of 153 SNPs, 49 of those GWAS-significant (p<5 x 10^{-8}) and 104 considered most important to explain heritable variance. Downstream analyses filter some of these signals based on their likelihood to affect other biological processes, being targeted by drugs and side effect profile of these drugs. For CAD/MI, the pipeline provided 3 potential candidates for repurposing, represented in bold in the last part of the pipeline.

To make optimal use on all the available GWAS data on multiple co-morbidities and phenotypes, we ranked CAD/MI loci based on their directions of effect (scores +1 in the concordant count for same direction of effect or +1 in the discordant count for different direction of effect), and ranked the SNPs according to the difference of these counts.
We hypothesize that GWAS on co-morbidities can provide a landscape of processes affected by increasing/decreasing the concentration of gene products (in case of drugs affecting these gene products). Therefore, targeting genes associated with other CAD-related phenotypes would be more helpful to manage CAD/MI than focusing on CAD specific risk loci. Furthermore, we can estimate possible side effects from an intervention with this approach, and avoid therapies predicted to have a severe side effect profile. Ideally, a good candidate alters no biological processes other than the phenotype of interest. To this intent, we performed a prioritization step among the 153 (49 + 104 as explained above) CAD SNPs using PhenoScanner [379], a comprehensive catalog of published GWAS results, in order to obtain a landscape of the associations with multiple phenotypes (Figure 1 for the flowchart). This has helped us to identify the impact of each SNP on other traits that are known to increase the risk of CAD/MI episodes, so we know that targeting these loci may have extended benefits by tackling multiple risk factors. We investigated also SNPs in high linkage disequilibrium (LD) (r^2>=0.8) with the reported SNP using 1000 Genomes Phase 3 [379, 381] using European population within 500kb upstream and 500kb downstream the variant, for completeness, and assessed in total 1303 SNPs. Nominally affected phenotypes per SNP region (p<0.05) were summed and the directions of effect on different phenotypes (effects for all phenotypes are aligned on the effect allele) were compared with those presented by CARDIoGRAMplusC4D, in two different setups. The first involved all phenotypes available on PhenoScanner, and the second was limited to traits affecting CAD/MI, namely LDL cholesterol, triglycerides, total cholesterol, hypertension (both diastolic [DBP] and systolic blood pressure [SBP] were considered), type 2 diabetes and obesity/BMI. We then ranked the variants based on a score formed by the sum of phenotypes of same direction (+1 for concordant count) minus the signals of opposite direction (+1 for discordant count) for the risk factor set, and in case of ties, the broad score (all phenotypes). Phenotype association results without betas available were discarded from the calculations.

7.3.2.2 Drug-gene mapping

Next, we mapped each SNP to genes based on the original annotation from CARDIoGRAMplusC4D (based on nearest gene) and on GENCODE (based on regulatory data from Ensemble and functional annotation from ENCODE) [382]. We hypothesize that each SNP considered maps to the same gene driving the CAD/MI development in all different phenotypes involved. We then sought publicly available drug-gene interaction databases to identify small molecules or marketed drugs that can target their gene products. We used an integrated database of drug-gene interactions

called DGIdb [380], which merged the most known drug-gene interaction databases, such as DrugBank [383] and PharmGKB [384]. We have not filtered results from DGIdb.

7.3.2.3 Side effect profiling and expert opinion

Full information on indications and side effects of drugs recommended for repurposing was obtained to exclude those drugs with severe and medically significant adverse reactions, such as reactions from chemotherapy drugs and severe cardiovascular complications such as long QT syndrome. Hence, we used the SIDER database version 3 [385], containing information on indications and side effects extracted from drug labels. Approximately 1,000 drugs are available in this dataset, with over 60,000 side effects described. Most of the drugs in the dataset are represented by their International Nonproprietary Names (INN), some are referred to by their United States Adopted Names (USAN). We used two data sources to identify all possible names: STITCH [386] and WHO's INN [387], available at https://mednet-communities.net/inn/. Vigibase was used [388] to aid in filtering drug candidates for repurposing based on side effect profiles. Vigibase is the largest repository of adverse event reports in the world, with over 14 million reports (as of 2017). This database was created by the World Health Organization (WHO) and maintained by the Uppsala Monitoring Centre [388]. Finally, a clinical pharmacist/pharmacogeneticist and a cardiologist validated the findings and made a final decision of which drugs to prioritize. We used the Drug Repurposing Hub [389] to compare our results with available data provided by this portal.

7.3.3 Pipeline 2: prioritization of drug target candidates (Figure 2)

For the prioritization of drug target candidates we calculated a druggability score to estimate how likely a compound would physicochemically interact with the drug target, i.e., how dock-able the pockets of CAD/MI associated gene products are. Afterwards we used PhenoScanner [379] to predict side effects (similar to what we did in the first pipeline), excluded loci with widespread cross-tissue expression to avoid housekeeping and essential genes and ranked the remaining gene products, determining the most suitable genes to be targeted for drug development.

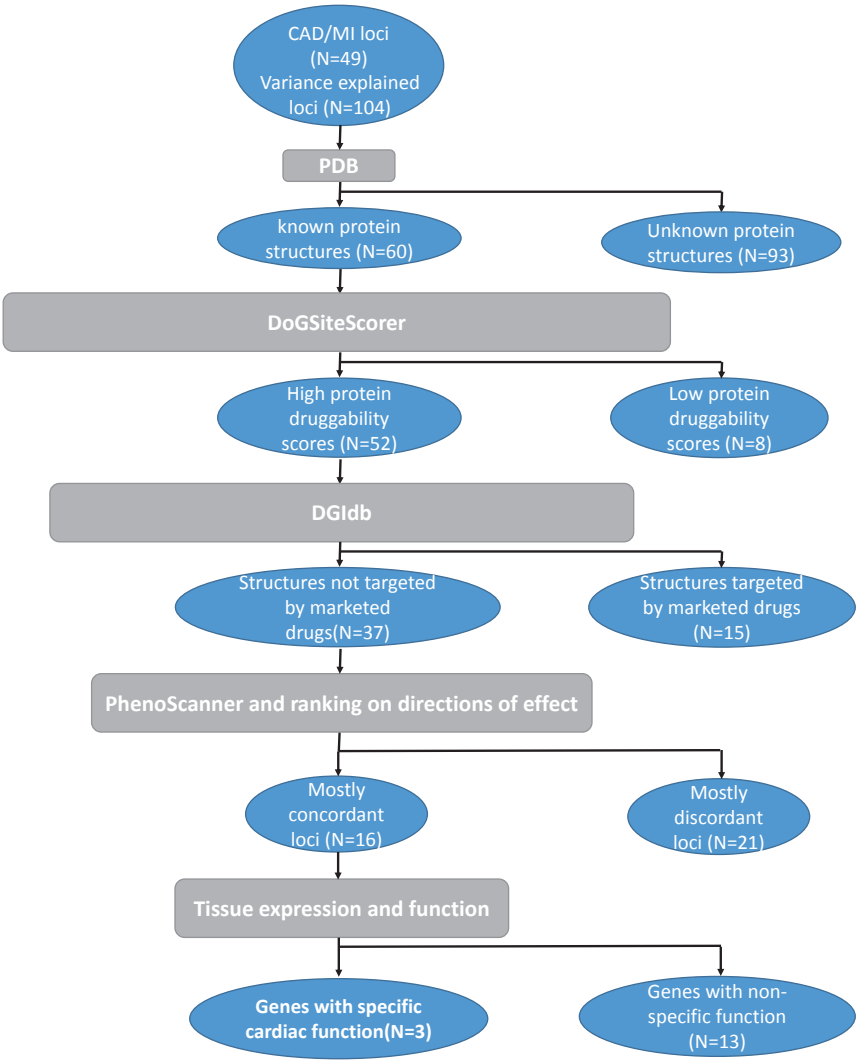


Figure 2. Pipeline 2 to evaluate CAD/MI loci as potential drug targets. Starting from the original publication SNP-to-gene mapping, it identifies chemical structures of the proteins produced by these genes, evaluates the likelihood of pockets of these proteins being connected to small molecules and observes the possible range of side effects based on GWAS results. Finally, it analyses cross-tissue gene expression and function, in order to exclude vital gene products involved in vital processes. It outputs three genes as most likely suitable candidates for drug development, represented in bold in the last part of the pipeline.

7.3.3.1 Druggability of docking pockets

We calculated the druggability score of each gene product using DoGSiteScorer [390]. This resource uses Protein Data Bank (PDB) [391] structures to calculate physicochemical interactions and identify pockets that can be targeted by molecules, and scores these pockets with values between 0 and 1. As mentioned in the original publication [390],

pockets with a DoGSiteScorer value above 0.5 would be in theory druggable [390]. Protein structures are determined most often using crystallography or magnetic resonance, as a picture of a specific conformation of a protein. With the druggability scores, the number of concordant/discordant directions of effect and information on tissue-specific expression obtained from Protein Atlas [392], we can hypothesize on most suitable candidate genes for drug development.

7.4 Results

7.4.1 Candidates for repurposing

All 153 SNP regions, selected for their association with CAD/MI, queried on PhenoScanner had a nominally significant association with another phenotype, 68 of which with a positive score of associations in the same direction as for CAD/MI, i.e., the risk of developing these phenotypes can be decreased together with CAD/MI in case of targeting given gene products. We filtered the list further by using DGIdb [380] to identify existing medications targeting the gene products of these regions, and found drugs targeting 15 of these gene products (*ABO, ACSS2, ARVCF, CDKN1A, CHRNB4, CKM, GUCY1A3, HDAC9, IL6R, LPL, MAP4, MTAP, PCSK9, SCARB1 and SLC22A4*). Due to gene products affected by multiple drugs, the targeted genes showed interactions with 48 drugs. Further analysis of the results showed that 22 of them are not clinically available drugs, either due to halted development (N=15) or still under evaluation in clinical trials (N=7). After excluding these, 26 medicinal products mapping to 15 genes of interest remain.

Out of the 26 products on our list, three are not classified as drugs (L-carnitine, phosphatidylserine and adenine), leaving 23 marketed drugs that have 26 assigned Anatomical Therapeutic Chemical (ATC) codes in total. The most common ATC groups involved were cardiovascular (group C - 8 drugs), cancer (group L - 7 drugs) and nervous system (group N - 6 drugs). Table 1 presents drug-gene interactions and ATC codes for these drugs.

We determined side effects of the 23 drugs using different databases, in particular SIDER; accordingly, we suggested possible candidates for repurposing. We excluded chemotherapy compounds (N=6), based on their serious side effect profile and unsuitability for continuous use in cardiovascular indications. Further exclusions involved medications which are already marketed for CAD prevention (N=2), expected to cause tachycardia (N=4), MI (N=1), liver injury (N=3), kidney damage (N=1) and stroke (N=1). Full counts of indications and side effects are available in Suppl. Table 2 and 3, respectively. Out of the remaining medications (N=5), we suggest two possible candidates for repurposing based on their positive impact on cardiovascular system seen in previous clinical trials. These include pentolinium (predicted as antagonist of the gene

Table 1. Predicted drug-gene interactions by DGIdb, ATC codes and nominally significant loci in the same LD block of the SNP presented

Reported SNP	Gene	Interacting drug	ATC code 1	ATC code 2	ATC code 3	GWAS predicted concordant phenotypes	GWAS predicted discordant phenotypes
rs11206510	*PCSK9*	EVOLOCUMAB	C10AX13			LDL cholesterol,Total cholesterol,Triglycerides,Obesity class 2	-
rs11206510	*PCSK9*	ALIROCUMAB	C10AX14			LDL cholesterol,Total cholesterol,Triglycerides,Obesity class 2	-
rs6088638	*ACSS2*	ADENOSINE TRIPHOSPHATE	C01EB10			Triglycerides,Obesity class 2,Obesity class 1,BMI	-
rs264	*LPL*	ORLISTAT	A08AB01			Triglycerides,Type II diabetes,Obesity class 2,BMI	-
rs264	*LPL*	CLOFIBRATE	C10AB01			Triglycerides,Type II diabetes,Obesity class 2,BMI	-
rs264	*LPL*	GEMFIBROZIL	C10AB04			Triglycerides,Type II diabetes,Obesity class 2,BMI	-
rs4845625	*IL6R*	TOCILIZUMAB	L04AC07			LDL	-
rs1034565	*ARVCF*	RISPERIDONE	N05AX08			Type II diabetes,BMI	-
rs1034565	*ARVCF*	BUPROPION	N06AX12			Type II diabetes,BMI	-
rs11072794	*CHRNB4*	VARENICLINE	N07BA03			Type II diabetes	-
rs11072794	*CHRNB4*	PENTOLINIUM	C02xxxx			Type II diabetes	-
rs11072794	*CHRNB4*	DEXTROMETHORPHAN	N07XX59	R05DA09		Type II diabetes	-
rs11072794	*CHRNB4*	ETHANOL	V03AB16	V03AZ01	D08AX08	Type II diabetes	-
rs11072794	*CHRNB4*	NICOTINE	N07BA01			Type II diabetes	-
rs8111989	*CKM*	CREATINE	C01EB06			BMI	-
rs2023938	*HDAC9*	VORINOSTAT	L01XX38			-	-
rs2023938	*HDAC9*	BELINOSTAT	L01XX49			-	-
rs2023938	*HDAC9*	VALPROIC ACID	N03AG01			-	-
rs2023938	*HDAC9*	PANOBINOSTAT	L01XX42			-	-
rs2023938	*HDAC9*	ROMIDEPSIN	L01XX39			-	-
rs7692387	*GUCY1A3*	RIOCIGUAT	C02KX05			-	-
rs7642590	*MAP4*	PACLITAXEL	L01CD01			Triglycerides	BMI
rs7642590	*MAP4*	DOCETAXEL	L01CD02			Triglycerides	BMI
rs273909	*SLC22A4*	L-CARNITINE	A16AA01			LDL,Total cholesterol,Triglycerides	

product of *CHRNB4*), adenosine triphosphate (targeting the gene product of *ACSS2*) and riociguat (antianginal agent works as a stimulator for the product of *GUCY1A3*).

7.4.2 Druggability of docking pockets

We also investigated druggability of CAD/MI loci, by analyzing the chemical structures of the respective gene products in search of pockets suitable for docking with novel molecules. We obtained PDB structures for 60 out of the 153 proteins produced by the estimated genes these CAD/MI loci belong to. Six of those were not human, but from a homologous animal model, and we decided to keep them for the analyses. Fifty-two structures had good druggability scores (>=0.5), and the eight remaining structures with low druggability scores were excluded from further analyses. Thirty-seven structures of the 52 remaining are not targeted by drugs available in the market, according to DGIdb. The statement of the original DoGSiteScorer manuscript about threshold for druggability were confirmed (scores above 0.5 would be in theory druggable [390]) by the 15 structures currently targeted by drugs, with scores ranging from 0.5-0.89. We also used the PhenoScanner ranking explained above for ranking the most promising candidates for drug development, and excluded a further 21 loci due to its predicted negative effect on related phenotypes. We then evaluated the remaining 16 loci for their function and tissue expression using Protein Atlas [392] to observe tissue expression of the gene products in different tissues, and excluded those with high levels of expression (more than 1SD from the average expression cross-tissue) across multiple tissues (e.g., brain, kidney, pancreas, muscle), since those are most likely housekeeping genes or necessary for cell cycle, and therefore not suitable for intervention [393]. We concluded with this approach that the most suitable targets to be considered are leiomodin 1 (*LMOD1*), huntingtin-interacting protein 1 (HIP1) and protein phosphatase 2, regulatory subunit b-double prime, alpha (PPP2R3A) (druggability scores were 0.73, 0.79 and 0.85, respectively). Score of effect directions were 4, 3, and 7 for CAD/MI related phenotypes (max. possible score 8) and 24, 24, and 14 for all other phenotypes, respectively. Full description of all molecules that passed the filters are shown in Table 2. Druggability scores for all proteins with a PDB entry are presented on Suppl. Table 4.

7.5 Discussion

GWAS have identified multiple loci and genes that appear to play a causal role in CAD/MI. While new efforts may unveil other associated loci (and indeed have already, with the current loci count at 164 [229, 394]), it is essential to maximize the value of the current data to translate this knowledge into clinical care, and improve management of CAD/MI. One way to utilize genetic information is by identifying suitable targets for drugs and possible repurposing of already existing drugs. Here, we cross-referenced multiple

Table 2. Most suitable drug targets according to predicted pocket interactions and nominally significant loci in the same LD block of the SNP presented

Reported SNP	Genes	DoGSiteScorer pocket score	PDB code	GWAS predicted concordant phenotypes	GWAS predicted discordant phenotypes
rs1393786	*PPP2R3A*	0.85	4i5j	LDL cholesterol,Total cholesterol,Triglycerides,Obesity class 2,Obesity class 1,Obesity class 3,BMI	
rs2820315	*LMOD1*	0.73	4z79	Obesity class 2,Obesity class 1,Obesity class 3,BMI	
rs1167800	*HIP1*	0.79	3i00	Triglycerides,Obesity class 1,BMI	
rs15563	*UBE2Z*	0.73	5a4p	LDL cholesterol,Total cholesterol,Type II diabetes	
rs6544713	*ABCG8*	0.89	5d07	LDL cholesterol,Total cholesterol,Triglycerides	BMI
rs9326246	*BUD13, ZNF259, APO5A, APOA1*	0.62	4uqt	LDL cholesterol,Total cholesterol,Triglycerides	LDL cholesterol
rs972158	*SNX10*	0.73	4pzg	Triglycerides,Type II diabetes	
rs10797416	*SKI*	0.73	1sbx	LDL cholesterol,Total cholesterol	
rs12205331	*ANKS1A*	0.81	2lmr	Triglycerides,BMI	
rs7139492	*COL4A1*	0.85	1li1	Total cholesterol	
rs816889	*RND3*	0.83	1m7b	Type II diabetes	
rs7173743	*MORF4L1*	0.79	2f5j	Obesity class 2	
rs10495907	*7SK*		2kx8	LDL,Total cholesterol,Triglycerides,Type II diabetes,Obesity class 2,Obesity class 1,Obesity class 3,BMI	
rs2281727	*SMG6*	0.89	4um2	BMI	Total cholesterol
rs2294461	*LY86*	0.82	3b2d	Type II diabetes	LDL
rs6984210	*BMP1*	0.69	3edg		

bioinformatics databases to identify potentially druggable genes and related compounds that may be suitable for repurposing in order to treat CAD/MI.

Until very recently, drug development has not been guided by genetic profiles and risks, therefore expecting that GWAS hits perfectly correspond to current treatments is unrealistic. That said, we compiled a list of medications used in the treatment of CAD (N=79), obtained from different sources, searched for drug-gene interactions among those with any level of confidence (N=608), checked how many unique genes are represented in these interactions (N=251) and compared to the genes we obtained from the CAD loci, either original mapping or GENCODE annotation (N=144). Expectedly, our data indicate that the drugs from the ATC group "cardiovascular system" (C) are overrepresented among our results. Indeed, our pipeline was able to identify three of the main medication groups used for treatment of CAD: statins (e.g. simvastatin), *PCSK9* inhibitors (e.g. evolocumab) and angiotensin II receptor blockers (e.g. irbesartan), which may serve as a validation of the method.

We found an overlap of 9 genes, namely *APOA1, APOB APOC1, APOE, EDNRA, GUCY1A3, LIPA, LPL* and *PCSK9*. Among those, our top results, in order, include *PCSK9* (target of newest medications in the field, evolocumab and alirocumab), *LPL* (indirectly a target of gemfibrozil), *APOC1* and *APOE* (indirectly representing statins, i.e. first in line drugs against CAD, and ritonavir) and *GUCY1A3* (target of suggested for repurposing drug riociguat) (Suppl. Table 5). Starting with a set of 153 loci identified through GWAS experiments for association with CAD/MI, through a series of filtering steps, we add evidence of the value of our druggability approach and suggest specifically three hits to be targeted by three drug compounds that show promise for repurposing including adenosine triphosphate (*ATP*), pentolinium, and riociguat.

ATP is a promising novel candidate with a ranking score (+4) similar to those obtained for the main rediscovered agents. Here the route of administration (intravenous) is an obstacle that needs to be addressed in future studies. *ATP*, which canalizes the reactions involving *ACCS2* gene product, is one of the top ranked repurposing candidates; this makes it the best promising agent suggested for CAD patients. It is has a role in regulating various biological cascades such as cardiac function, muscle contractility and blood circulation [395]. Through the period of 80s and 90s, *ATP* was useful in managing several clinical conditions such as haemorrhagic shock, pulmonary hypertension and paroxysmal supraventricular tachycardias [396]. ATP is not currently marketed in the United States but is available certain European countries; it is indicated as an adjunct therapy for low back pain in France [397] and was tested as therapeutic agent for patients with Alzheimer Disease in a recent clinical trial (https://clinicaltrials.gov/ct2/show/NCT02279511). According to DGIdb, *ACSS2* gene is a suggested target for *ATP*; it synthesizes acetyl *CoA* from *ATP* and *CoA* through an acetyl-adenosine monophosphate (*AMP*). The Drug repurposing hub confirms the drug-gene interaction. The proposed repurposing candidates also included riociguat (targeting *GUCY1A3*) and pentolinium

(targeting *CHRNB4*), albeit with lower rankings (scores of 0 - +1 for all phenotypes), which may suggest additive roles for these medications in CAD. The findings of several GWAS suggest the gene *GUCY1A3*, coding for the alpha-3 subunit of soluble guanylate cyclase in chromosome 4 as a drug target to manage individuals with CAD/MI. The variant rs7692387 was strongly associated with CAD [373] and later found to modulate *GUCY1A3* promoter activity [398] and rs13139571 was identified as a risk factor for hypertension [116]. This gene codes for a protein that acts as a major receptor for nitric oxide and other nitro-derivative products (e.g. nitroglycerin) to induce vasodilatation and platelet inhibition [399]. *GUCY1A3* is highly expressed in the vascular smooth-muscle cells and has the potential to modulate vascular tone and to induce venous and arterial relaxation. Interaction of nitric oxide with different isoforms of soluble guanylate cyclase 1 (*GUCY1*) plays an important role in regulating platelet aggregation [400] as well as accelerating thrombus formation [401]. Deletion of *GUCY1A3* is known to cause asymptomatic moyamoya (intracranial stenosis) angiopathy [400]. Its deletion is also known to cause myocardial infarction [401], and detailed molecular analyses identified how this variant modulates expression of the gene, soluble guanylyl protein levels, activity of the enzyme and platelet function [398]. In addition, an exonic variant rs201558687 was reported as protective marker against pulmonary hypertension [402]. This finding suggests a role for *GUCY1A3* gene in pulmonary hypertension in addition to its association with CAD/MI.

The newly marketed drug riociguat was approved by the U.S. Food and Drug Administration (FDA) in 2013 [403] and later by European Medicines Agency in 2014 to manage patients with pulmonary arterial hypertension (PAH) and chronic thromboembolic pulmonary hypertension (CTEPH) [404]. Riociguat is a positive allosteric modulator for *GUCY1A3*. It is a unique drug that acts as guanylate cyclase stimulator [405]; this mechanism makes a hope to repurpose the usage of riociguat as an antianginal agent [406]. Hypotension, headache and dizziness are the common side effects of riociguat [407]. A clinical trial was prepared to study the effects of riociguat for CAD (clinicaltrials.gov ID: NCT01165931) but the study was cancelled before recruitment for unknown reasons. Further evidence is provided by the Drug Repurposing Hub [389], which mentions that molsidomine, a drug not available in DGIdb and predicted to interact with *GUCY1A3*, is prescribed against CAD.

Pentolinium is an old antihypertensive agent indicated to control malignant hypertension and hypertensive crises, in particular, throughout surgery [408]. Pentolinium is predicted to antagonize *CHRNB4*. It was marketed by WYETH AYERST under the trade name of Ansolysen then it was decided to stop its manufacturing in January 1982. The manufacturer did not indicate the reasons of discontinuation but possibly because of induction of severe postural and exertional hypotension [409]. Moreover, the drug is rather non-specific (it targets different subunits of nicotinic acetylcholine receptors (*nAChRs*) at a time [410], in particular α*3,* α*10,* α*2* and α*4*), and had to compete with newer

medications with higher efficacy in lowering blood pressure. It has a potent peripheral ganglionic blocking action and acts as an antagonist to nicotinic receptors which inhibit the release of both adrenaline and noradrenaline [411]. Although these receptors are abundant on somatic and central nervous system they are also expressed on aortic valves and atrial appendages [412]. Nicotinic receptors are considered part of a superfamily of ligand-gated ion channels which mediate fast signal transmission at synapses [413]. Blockage of the receptor results in relaxation as well as vasodilatation of smooth muscles. A single nucleotide polymorphism (SNP, rs11072794) in the gene cholinergic receptor, neuronal nicotinic, beta polypeptide 4 (*CHRNB4*) - in chromosome 15- which codes for nicotinic receptors was found to be a risk factor for CAD/MI [370] and nominally significantly associates to type 2 diabetes (T2D) [414] in a GWAS study.

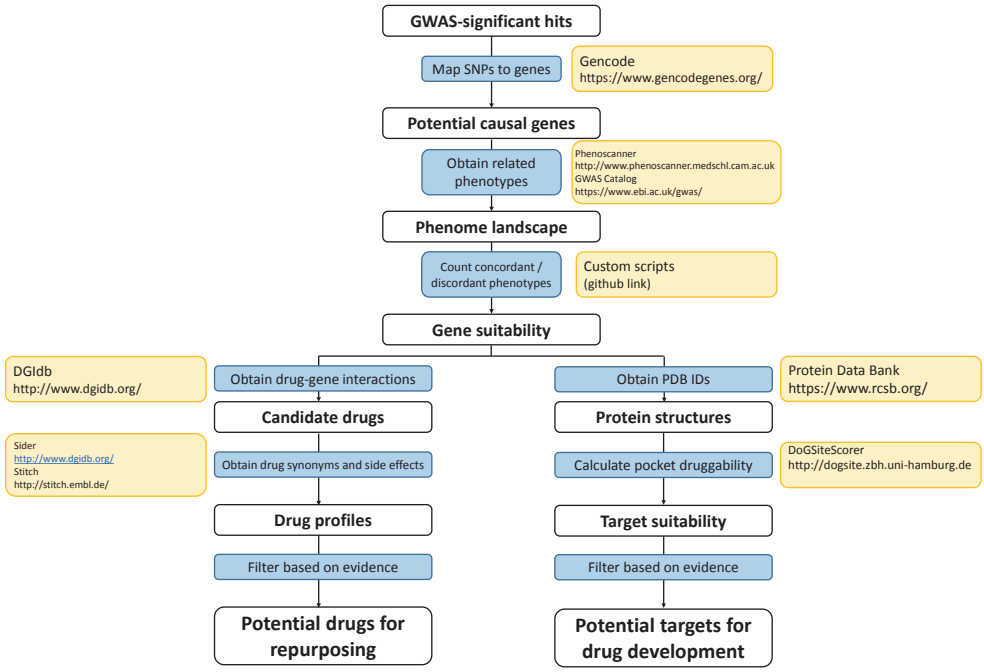
The SNP rs11072794, located in an intronic region, is in complete LD (r^2=1) with other two variants (rs899997 and rs12899940), that are located in regulatory regions that possibly affect gene function; rs12899940 is located in promoter flanking region, while rs899997 is located in a transcription factor binding site. The SNP rs899997 was recently identified as a risk factor to develop coronary artery disease and ischemic stroke according to recent analysis of three different genome-wide association studies; the METASTROKE, CARDIoGRAM, and C4D consortia [415]. Further, another marker (rs8023822) in *CHRNB4* gene was also detected as a susceptibility loci for CAD in T2D in a meta-analysis that involve several GWAS studies among Scottish population [416]. These significant associations between different loci in CHRNB4 and CAD suggest the gene product as a good drug target; therefore, pentolinium can possibly repurposed in the management of CAD/MI conditions. In this case, however, we found no further evidence from the Drug Repurposing Hub.Regarding novel targets, we were able to elaborate a ranking of the most suitable candidates. The most promising candidates for targeting are LMOD1, HIP1 and PPP2R3A. Leiomodin 1 (*LMOD1*) is related to smooth muscle contraction and cardiac conduction [417] and has been identified to have smooth muscle cell-specific eQTLs in SNP rs34091558 [418]. SNP rs2820315 at *LMOD1* has reached genome-wide significance in the latest GWAS in CAD/MI [419]. Huntingtin-interacting protein 1 (HIP1) codes for a protein significantly expressed in coronary artery endothelium cell and play a major role in cell endocytosis [420], having one of the top significant cis-eQTL expression patterns among CAD/MI loci [421]. Finally, protein phosphatase 2, regulatory subunit b-double prime, alpha (*PPP2R3A*) is abundantly expressed in heart and skeletal muscles and responsible for intracellular signal regulation [422], being associated to several regulatory networks of CAD [423].

The pipelines are modular and can be easily generalized for other diseases. Once determined an appropriate set of *bona fide* associated SNPs for a given trait, the pipelines can provide candidates for repurposing and most suitable targets for drug development. However, a number of limitations need to be considered in translating our results to clinical studies. First, our approach relies on the well-informed yet unproven relationship

between CAD/MI loci and a nearby gene and the druggability of its gene product. It is worth mentioning that the genes investigated in both pipelines were mapped to variants by positional mapping in their original study. We extended this analysis using GENCODE annotations to identify additional genes, a comprehensive resource that also integrates regulatory data in the process of annotation. However, variants can act on elements regulating genes that are not physically located in their immediate vicinity, but often even outside the locus [83]. Experiments on chromosomal conformation can capture these dynamics by generating a map of tissue-specific genomic regions that physically interact [77-79, 82]. The ongoing generation and integration of such maps on disease-relevant cell-types will enable identification of target genes that might not have been unraveled by using current approaches, which in turn might improve results of our pipelines. Second, it needs to be investigated as to whether the drugs discussed modulate the gene product in a beneficial way. Third, we illustrate in our examples that some of the drugs are pleiotropic or not suitable for chronic application, and that they may be only a starting point for further developments. Fourth, the databases used were complete versions at the moment of usage, but those are ongoing efforts that will improve coverage and reliability, so iterative development may yield more reliable results. Finally, we focused this investigation on repurposing of established drugs. We expect that multiple GWAS loci, pointing to currently unexplored mechanisms, can be addressed by new drug development on antibodies, like what has been successfully achieved in case of *PCSK9* [105-107]. In conclusion, we have found evidence for repurposing of drugs and candidates for drug development in the context of CAD/MI, suggesting that *in-silico* analysis using existing databases and genetic findings may be useful to accelerate translation into clinical practice. Clinical trials are now needed to explore the potential value of these agents.

7.6 Supplemental Material

7.6.1 Supplemental Figures



Supplemental Figure S1. Detailed flowchart presenting the pipelines developed in this study together with data and tools used.

7.6.1 Supplemental Tables

Supplementary Table 1. Examples of repurposed drugs.

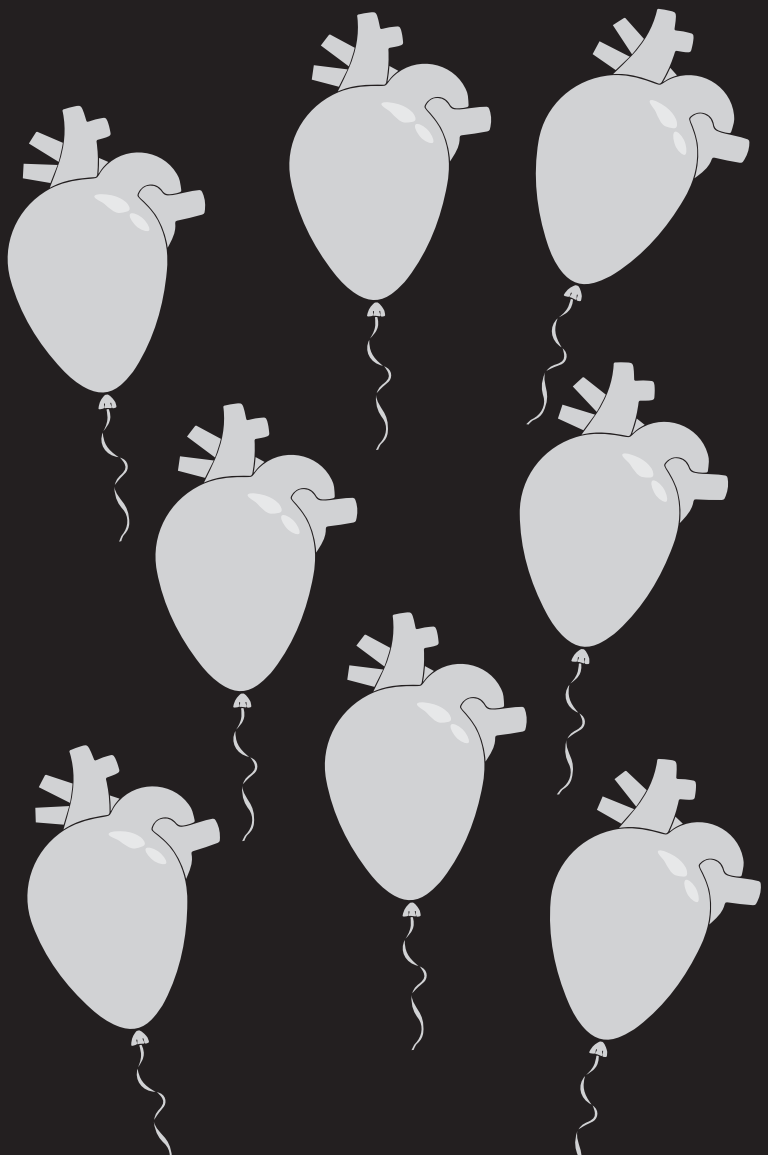
Medication	Initial indication	Repurposed indication	Evidence for repurposing
sildenafil	Angina	erectile dysfunction, pulmonary hypertension	PDE5 mechanisms [PMID: 10422958]
duloxetine	depression	stress urinary incontinence	Animal models [PMID: 14501737]
dapoxetine	depression	premature ejaculation	Side effect of depression treatment [PMID: 16962882]
thalidomide	morning sickness	multiple myeloma	Teratogenicty in fetuses [PMID: 12046682]
gemfibrozil	dyslipidemia	coronary artery disease	relation between cholesterol and CAD [PMID: 10438259]

Supplementary Table 2. Number of drugs prescribed per indication (available upon request).

Supplementary Table 3. Number of side effects known for the drugs available on SIDER (available upon request).

Supplementary Table 4. Number of concordant and discordant phenotypes per locus associated with CAD/MI (available upon request).

Supplementary Table 5. CAD drug-gene interactions, sorted by CAD GWAS hits (available upon request).



Chapter 8

Summary and Discussion

8.1 Summary

CVDs are the leading cause of death globally [2]. Risk factors were identified over the past decades, both modifiable and non-modifiable, which increase the risk of developing CVD. Modifiable risk factors include smoking, hypertension, obesity, excessive alcohol consumption and stress. Non-modifiable risk factors include age, sex, and genetics. GWAS efforts have been successful in identifying loci that contribute to CVD risk. However, the mechanisms through which these loci contribute to disease are still relatively undetermined. In this thesis, I present a set of strategies for interpretation of what has been discovered in recent years associating genetic loci and CVDs, aiming at translating these findings into better disease understanding and better health care.

In **Chapter 2** we use publically available data and bioinformatics tools for interpretation and potential translation of BP associated loci into drug targets. By integrating genome, regulome and transcriptome information in relevant cell-types, we performed functional annotation of 905 BP loci, showing that these loci implicate primarily vascular and cardiac tissues. Maps of chromosomal conformation and eQTL in these critical tissues identified 2424 genes interacting with BP-associated loci, of which 517 are predicted to be druggable.

In **Chapter 3,** we integrated several layers of data for fine-mapping and functional annotation of 52 loci influencing myocardial mass, measured by the QRS parameter of the ECG. We demonstrated that differentially active regulatory elements between HCM and controls show more enrichment for QRS-associated variants than the combination of all regulatory elements identified in HCM and controls. In addition, we pinpointed candidate genes and potential mechanisms by which candidate causative QRS-associated variants account risk for increased left ventricular mass.

In **Chapter 4**, we highlighted the contribution of data integration for investigation of regulation and expression profiles of dilated cardiomyopathy (DCM). We identified sets of differentially regulated and/or differentially acetylated genes previously implicated in DCM, adding evidence of their role in the process of this heterogeneous disease.

In **Chapter 5** we calculated GRSs to study the potential relevance of cardiovascular susceptibility loci to other vascular beds. We tested 6 loci for association with cIMT (a biomarker for CVD) and plaque presence. We replicated 4 of these associations with cIMT and 2 with plaque, and associated a variant near *EDNRA* with CAD. We thus provide support for previously claimed SNP associations for cIMT and plaque, specifically highlighting the role of rs1878406 in both atherosclerosis and CAD. In addition, we show the potential of GRS application in extending GWAS findings for interpretation of associations and investigation of pleiotropy.

In **Chapter 6**, we calculated GRSs from BP-associated variants with high tissue-specific regulatory potential, and measured these GRSs against outcomes influenced by high BP that manifest in different organs. We showed that phenotypically cell-type specific sets

of BP-associated variants correlate better to BP-induced outcomes (such as renal failure, diabetes and CVDs) than non BP-induced outcomes (bipolar disorder). In addition, from the nine hypertension-induced outcomes tested, three showed higher association of tissue-specific variants that implicate regulation in the tissue affected by the outcome. These are LV-specific regulatory variants in DCM, kidney-specific regulatory variants in kidney failure, and pancreas-specific variants in outcomes implicating pancreas tissue, including diabetes. These studies suggest a role of tissue-specific regulatory variants in accounting risk for outcomes that manifest in the same tissue. Increasing sample sizes and extending this analysis to a broader range of tissues and histone modifications can help capture the high risk sets of variants for tissue-specific BP-induced outcomes.
In **Chapter 7**, we described a pipeline for identification of new drug targets, with the calculation of a druggability score to estimate the dockability of pockets of CAD/MI associated gene products and then utilized GWAS results again to predict side effects. We were able to identify three proteins with high potential for drug development: *LMOD1*, *HIP1* and *PPP2R3A*. We also presented a pipeline for potential repurposing of currently marketed drugs integrating publicly available information on current marketed medicines with GWAS results. This pipeline led to three suggestions for drug repurposing: pentolinium, adenosine and riociguat, to target *CHRNB4*, *ACSS2* and *GUCY1A3*, respectively.

8.2 GWAS limitations and perspectives

Since its initiation, the GWAS approach has identified thousands of disease-associated loci in hundreds of traits and diseases [424], despite initial skepticism. Findings from GWAS have made, and will continue to make, a very important contribution to the investigation of genetic architecture of complex diseases. However, critics continue to currently call attention to the problem of "missing heritability". Twin studies have helped estimate the amount of phenotypic variation in traits or diseases that can be attributed to genetic variation, thus having a heritable component. Most complex traits and common diseases are estimated to have 30-60% heritability [425]. In some cases, this estimate increases to 80% [426-428]. As a substantial proportion of individual differences in disease susceptibility is known to be due to genetic factors, understanding this genetic variation can contribute to better prevention, diagnosis and treatment of disease [425]. Most variants identified by GWAS thus far confer relatively small increments in risk, and explain only a small proportion of familial clustering. Skeptics thus question as to whether GWAS efforts could ever elucidate all the heritability of complex disease. In the example of CAD, heritability is estimated to be 20-40%, of which the largest GWAS meta-analysis explains only 11% [329, 429]. However, with the increase in sample sizes over the years, more variants have been continually revealed and have had an impact on

explained heritability estimates. This is the case of most recent and largest GWAS effort to date on BP including over 1 million people, identifying 535 novel BP loci [41]. Other recent efforts with increased sample sizes have shown a similar trend in explaining heritability [430-432]. This suggests much larger numbers of variants of smaller effect are yet to be found in order to explain the missing heritability of some traits and diseases. Other factors that might help explain heritability in the future are the discovery of rarer variants (possibly with larger effects), detection of structural variants that are currently poorly detected by available genotyping arrays, an increase of power to detect gene–gene interactions, and adequate accounting for shared environment among relatives [425, 433, 434]. The greatest promise of GWAS, however, was to agnostically identify genetic factors that modify disease, by using a purely statistical approach. In this regard, GWAS have succeeded [41, 428, 435-438]. The other main goal was to identify causal genes that would help elucidate disease biology and potentially improve diagnostics and treatment [36], and this brings the question as to whether GWAS results thus far show potential to be biologically or clinically relevant. Recently, studies based on large-scale collaboration, meta-analysis and national projects such as UK Biobank [354, 439] are beginning to demonstrate that genetic factors provide robust and powerful risk estimation across diseases that is additive to traditional risk factors [440-442]. In addition, identification of causal genes affected by statistically associated variants (that are located mostly on non-coding regions of the genome) is a challenge currently being addressed by taking into account other data dimensions beyond the DNA sequence. Data integration is a key step in the post-GWAS era towards the interpretation of non-coding genetic variation. It can help to both fine-map disease risk loci as well as to improve understanding of human physiology and disease etiology, so that more effective means of diagnosis, treatment and prevention can be developed.

8.3 Incorporating genetic risk scores into clinical practice for better risk assessment and personalized medicine

In addition to helping us to understand the role of genetics in disease mechanisms, GWAS results can be used to improve our ability to risk-stratify individuals [50, 443]. This thesis showed an example in **Chapter 5** of the use of GRSs to study the potential relevance of CVD susceptibility loci to other vascular beds. Indeed, recent studies have shown that GRSs may serve not only as prognostic, but also as predictive markers. The assessment of an individual's risk for CVDs has the potential to enable development of individualized treatment strategies, contributing to the progress of precision medicine [444].

Most recent GRSs have demonstrated significant improvements in performance for risk prediction in CVD [50]. Moreover, costs of genome-wide genotyping are now under

US$100 per person, and because genotyping chips survey common variants across the genome, reflecting risk for hundreds of conditions besides CVD, with a single test it would be possible to predict risk of multiple diseases. Thus, the low cost, strong prognostic and predictive power of increasly accurate risk prediction are reasons to consider incorporating genetic information and GRSs into clinical practice [42].

We showed in **Chapter 6** that hypertension-associated variants are significantly associated to vascular and cardiovascular hypertension induced outcomes, whereas this was not observed in the outcome used as control. The aforementioned study did not manage to capture subsets of tissue-specific regulatory variants that account for greater risk for tissue-specific outcomes in all outcomes tested. These results could be due to lack of power in detecting such association given the current study design, or the fact that only one class of regulatory elements was used, focused on one epigenetic modification. As data becomes available that maps the regulatory landscape of more tissue types and focuses on different regulatory elements (highlighted by different histone modifications, for example), we might be able to identify disease-associated loci that enables us to generate genetic disease progression profiles.

As the genetic risk of an individual can be detected for a CVD, measures can be taken to lower this risk by working on modifiable risk factors. Genetic risk for hypertension, for example, has been shown to be mitigated by a healthy lifestyle [445]. Furthermore, informing at-risk subjects of their genetic risk has the potential to motivate not only lifestyle changes, but also adherence to prescribed medication to a greater extent than the classic risk factor information alone [446].

8.4 Improving CVD biological understanding with data integration

GWAS were an early statistical method to relate genetic variation to diseases and human traits. The evident follow-up is to focus on the specific regions of the genome that show significant statistical association, and perform further experiments that can help understand biological mechanisms or explain disease etiology. Thus, biological interpretation of genetic variation is a key step of the post-GWAS era, and can only be enabled by the integration of other data dimensions, such as gene expression, epigenetic data and genome folding. The evolution of genome sequencing has allowed the development of protocols and experiments for interrogating features of DNA that influence its function, at constantly lower costs. This has enabled collaborative projects such as ENCODE [60], Roadmap Epigenomics [66] and GTEx [67, 89] to make publicly available a wealth of data on epigenetic modifications, open chromatin, methylation, expression quantitative trait loci (eQTLs) and "omics" information. In addition, maps of chromosomal conformation capture that analyze the spatial organization of chromatin in a cell can inform on the complex interactions between genes and their

regulatory elements. Integration of all these layers of information with GWAS results is a critical step towards explaining the contribution of non-coding variation to disease, identifying potentially disrupted regulatory mechanisms and genes with consequent altered expression (**Chapters 2** and **3** of this thesis).

The complex regulatory machinery of the genome is highly tissue-specific. Although the regulatory landscapes of a wide variety of cell-types are available on public sources, the generation of more data on different cell-types will help us understand further the mechanisms of complex diseases. In addition, the regulatory maps currently available are not yet complete, given that not all epigenetic marks and their combinations have yet been identified. Furthermore, how these features work together in concert to regulate epigenetic mechanisms and gene expression is still unknown [23]. In fact, even our catalog of discovered genes is still under construction, evolving as sample sizes increase [447]. As our gene and regulatory elements lists evolve, so will the explanation on the genetic causes of many complex diseases.

Another important aspect to take into account is the changing regulatory landscape of cells during disease progression. This process also occurs with the chromosomal conformation inside the nucleus, while conformation capture techniques reflect only a snapshot in a single time-point [176]. The influence from cell conditions is crucial for epigenetic modifications, chromatin accessibility, TF binding and consequently gene regulation [448]. Thus, a more complete view of regulatory changes in specific cell-types in different conditions will enable a better understanding of regulatory elements affected in disease, and the contribution of genetic variation on increased risk.

Epigenetic studies in cardiovascular medicine thus far revealed a significant number of modifications affecting the development and progression of CVD [14]. These epigenetic mechanisms can be employed not only for understanding the pathophysiology of CVD [449-453](**Chapter 4** of this thesis), but also its diagnosis and treatment [16, 454]. Although there is currently no practical application of epigenetics in treatment in CVDs, histone methylation inhibitors and histone deacetylase inhibitors have been successful so far in cancer therapeutics, and there are many more examples of diseases for which epigenetic treatment holds great promise [16, 455-458]. Targeting these enzymatic regulators of chromatin modification is also promising for modulating the transcriptional regulation of genes that are involved in atherogenesis, inflammation, smooth muscle cell proliferation, and matrix formation [23].

Overall, the main contribution of epigenetics in the field of CVDs thus far has been in elucidating the role that functional variants play in accounting for risk of disease. Understanding the epigenetic mechanisms underlying disease biology will make a significant contribution to our understanding of CVDs [14]. The resulting knowledge from data integration on post-GWAS analyses may lead to better prevention and treatment, which in turn may cause a decline in healthcare costs and an improved quality of life.

8.5 Further experimental steps towards new drug targets

Over a quarter of drugs that enter clinical development fail because they are ineffective[96]. Human genetics and large-scale genomics hold great promise on how therapeutic target validation is approached [459], either by discovery of new drug targets or repurposing of existing ones, as proposed on **Chapter 7** of this thesis.

Bioinformatics analyses and experimental studies have revealed the involvement of genes affected by GWAS risk alleles in new and different pathophysiological pathways [96]. The investigation of the genes and regulatory elements involved in these pathways can enable translation of risk-related genetic variants to new preventive and therapeutic strategies. Validation of GWAS findings beyond *in silico* approach is a much-needed step, and experimental animal and *in vitro* models will be required. The most promising follow-up assay appears to be the use of CRISPR-Cas9 gene editing technology (clustered regularly interspaced short palindromic repeat (CRISPR) -associated protein 9) [460]. This technology has the advantage to be able to insert any form of mutation (single nucleotide changes, insertion or deletions) with precision and at any given location [460]. A repurposing of the CRISPR-Cas9 system has been used to edit the methylation status of genomic sequences, inducing altered DNA methylation of specific loci altering gene expression [461]. Techniques for remodeling chromatin loops with designer CRISPR tools have also emerged [462]. Genome editing experiments can be designed to reverse epigenetic changes by targeting intermediate chromatin features, such as accessibility and structure [449]. Targeted manipulation of epigenetic mechanisms via CRISPR–Cas9 can thus be a powerful tool for validation of findings from fine-mapping and integrative studies, helping elucidate the causal role, if any, of epigenetic marks in a locus-specific manner [184].

From discovery of statistical associations to fine-mapping and validation, closer partnerships between academic researchers and pharmaceutical companies will be required to seize the opportunities that human genetics and genomics offer on drug target research. Most of this need relies on the fact that the vast majority of relevant existing data is found in open databases but still often is not interoperable. Integrating these databases is a critical step to maximize the benefits of such wealth of data, which is much better done by partnerships involving domain experts than by being tackled by individual companies [459]. Collaborations on data integration and analysis can thus expand our ability to translate GWAS discoveries into effective disease prevention or treatment, at low cost and with minimal side effects.

References

1. Yang, Q., et al., *Trends in cardiovascular health metrics and associations with all-cause and CVD mortality among US adults.* JAMA, 2012. **307**(12): p. 1273-83.

2. Benjamin, E.J., et al., *Heart Disease and Stroke Statistics-2018 Update: A Report From the American Heart Association.* Circulation, 2018. **137**(12): p. e67-e492.

3. Writing Group, M., et al., *Heart Disease and Stroke Statistics-2016 Update: A Report From the American Heart Association.* Circulation, 2016. **133**(4): p. e38-360.

4. McKay, J., *The atlas of heart disease and stroke*. 2004: Geneva : World Health Organization, [2004] ©2004.

5. Heidenreich, P.A., et al., *Forecasting the future of cardiovascular disease in the United States: a policy statement from the American Heart Association.* Circulation, 2011. **123**(8): p. 933-44.

6. Oldridge, N.B., *Economic burden of physical inactivity: healthcare costs associated with cardiovascular disease.* Eur J Cardiovasc Prev Rehabil, 2008. **15**(2): p. 130-9.

7. Luengo-Fernandez, R., et al., *Cost of cardiovascular diseases in the United Kingdom.* Heart, 2006. **92**(10): p. 1384-9.

8. Leal, J., et al., *Economic burden of cardiovascular diseases in the enlarged European Union.* Eur Heart J, 2006. **27**(13): p. 1610-9.

9. Wilkins E, W.L., Wickramasinghe K, Bhatnagar P, Leal J, Luengo-Fernandez R, Burns R, Rayner M, Townsend N, *European Cardiovascular Disease Statistics 2017*, E.H. Network, Editor. 2017: Brussels.

10. Gabriel, S.B., et al., *The structure of haplotype blocks in the human genome.* Science, 2002. **296**(5576): p. 2225-9.

11. Rosenberg, N.A., et al., *Genetic structure of human populations.* Science, 2002. **298**(5602): p. 2381-5.

12. P., W.A., *Chromatin: structure and function*. 1992, Academic Press: London, England.

13. Natsume-Kitatani, Y., M. Shiga, and H. Mamitsuka, *Genome-wide integration on transcription factors, histone acetylation and gene expression reveals genes co-regulated by histone modification patterns.* PLoS One, 2011. **6**(7): p. e22281.

14. Abi Khalil, C., *The emerging role of epigenetics in cardiovascular disease.* Ther Adv Chronic Dis, 2014. **5**(4): p. 178-87.

15. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.

16. Muka, T., et al., *The role of epigenetic modifications in cardiovascular disease: A systematic review.* Int J Cardiol, 2016. **212**: p. 174-83.

17. Gusterson, R.J., et al., *The transcriptional co-activators CREB-binding protein (CBP) and p300 play a critical role in cardiac hypertrophy that is dependent on their histone acetyltransferase activity.* J Biol Chem, 2003. **278**(9): p. 6838-47.

18. Haddad, F., et al., *Role of antisense RNA in coordinating cardiac myosin heavy chain gene switching.* J Biol Chem, 2003. **278**(39): p. 37132-8.

19. Majumdar, G., et al., *Epigenetic regulation of cardiac muscle-specific genes in H9c2 cells by Interleukin-18 and histone deacetylase inhibitor m-carboxycinnamic acid bis-hydroxamide.* Mol Cell Biochem, 2008. **312**(1-2): p. 47-60.

20. Zhang, Q.J., et al., *The histone trimethyllysine demethylase JMJD2A promotes cardiac hypertrophy in response to hypertrophic stimuli in mice.* J Clin Invest, 2011. **121**(6): p. 2447-56.

21. Bingham, A.J., et al., *The repressor element 1-silencing transcription factor regulates heart-specific gene expression using multiple chromatin-modifying complexes.* Mol Cell Biol, 2007. **27**(11): p. 4082-92.

22. Stein, A.B., et al., *Loss of H3K4 methylation destabilizes gene expression patterns and physiological functions in adult murine cardiomyocytes.* J Clin Invest, 2011. **121**(7): p. 2641-50.

23. Ordovas, J.M. and C.E. Smith, *Epigenetics and cardiovascular disease.* Nat Rev Cardiol, 2010. **7**(9): p. 510-9.

24. Sturtevant, A.H., *A History of Genetics - With a New Introduction by Edward B. Lewis*. 2001: California Institute of Technology

25. Dahm, R., *Friedrich Miescher and the discovery of DNA.* Dev Biol, 2005. **278**(2): p. 274-88.

26. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii.* J Exp Med, 1944. **79**(2): p. 137-58.

27. Franklin, R.E. and R.G. Gosling, *Molecular configuration in sodium thymonucleate.* Nature, 1953. **171**(4356): p. 740-1.

28. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.

29. Lander, E.S., et al., *Initial sequencing and analysis of the human genome.* Nature, 2001. **409**(6822): p. 860-921.

30. Venter, J.C., et al., *The sequence of the human genome.* Science, 2001. **291**(5507): p. 1304-51.

31. International Human Genome Sequencing, C., *Finishing the euchromatic sequence of the human genome.* Nature, 2004. **431**(7011): p. 931-45.

32. International HapMap, C., *The International HapMap Project.* Nature, 2003. **426**(6968): p. 789-96.

33. Genomes Project, C., et al., *A global reference for human genetic variation.* Nature, 2015. **526**(7571): p. 68-74.

34. Meaburn, E., et al., *Genotyping pooled DNA using 100K SNP microarrays: a step towards genomewide association scans.* Nucleic Acids Res, 2006. **34**(4): p. e27.

35. Oliphant, A., et al., *BeadArray technology: enabling an accurate, cost-effective approach to high-throughput genotyping.* Biotechniques, 2002. **Suppl**: p. 56-8, 60-1.

36. Hirschhorn, J.N. and M.J. Daly, *Genome-wide association studies for common diseases and complex traits.* Nat Rev Genet, 2005. **6**(2): p. 95-108.

37. Bonferroni, C.E., *Teoria statistica delle classi e calcolo delle probabilita*. 1936: Libreria Internazionale Seeber.

38. Altshuler, D., M.J. Daly, and E.S. Lander, *Genetic mapping in human disease.* Science, 2008. **322**(5903): p. 881-8.

39. International HapMap, C., *A haplotype map of the human genome.* Nature, 2005. **437**(7063): p. 1299-320.

40. MacArthur, J., et al., *The new NHGRI-EBI Catalog of published genome-wide association studies (GWAS Catalog).* Nucleic Acids Res, 2017. **45**(D1): p. D896-D901.

41. Evangelou, E., et al., *Genetic analysis of over one million people identifies 535 novel loci for blood pressure.* bioRxiv, 2017.

42. Knowles, J.W. and E.A. Ashley, *Cardiovascular disease: The rise of the genetic risk score.* PLoS Med, 2018. **15**(3): p. e1002546.

43. Wilson, P.W., et al., *Prediction of coronary heart disease using risk factor categories.* Circulation, 1998. **97**(18): p. 1837-47.

44. Chambless, L.E., et al., *Coronary heart disease risk prediction in the Atherosclerosis Risk in Communities (ARIC) study.* J Clin Epidemiol, 2003. **56**(9): p. 880-90.

45. Folsom, A.R., et al., *An assessment of incremental coronary risk prediction using C-reactive protein and other novel risk markers: the atherosclerosis risk in communities study.* Arch Intern Med, 2006. **166**(13): p. 1368-73.

46. Koenig, W., et al., *C-reactive protein modulates risk prediction based on the Framingham Score: implications for future risk assessment: results from a large cohort study in southern Germany.* Circulation, 2004. **109**(11): p. 1349-53.

47. Stephens, J.W. and S.E. Humphries, *The molecular genetics of cardiovascular disease: clinical implications.* J Intern Med, 2003. **253**(2): p. 120-7.

48. Morrison, A.C., et al., *ADD1 460W allele associated with cardiovascular disease in hypertensive individuals.* Hypertension, 2002. **39**(6): p. 1053-7.

49. Morrison, A.C., et al., *Prediction of coronary heart disease risk using a genetic risk score: the Atherosclerosis Risk in Communities Study.* Am J Epidemiol, 2007. **166**(1): p. 28-35.

50. Assimes, T.L. and R. Roberts, *Genetics: Implications for Prevention and Management of Coronary Artery Disease.* J Am Coll Cardiol, 2016. **68**(25): p. 2797-2818.

51. N., C.B.J. and I.R. P., *Genetic Risk Scores.* Current Protocols in Human Genetics, 2016. **91**(1): p. 1.29.1-1.29.9.

52. Hindorff, L.A., et al., *Potential etiologic and functional implications of genome-wide association loci for human diseases and traits.* Proc Natl Acad Sci U S A, 2009. **106**(23): p. 9362-7.

53. Witte, J.S., *Genome-wide association studies and beyond.* Annu Rev Public Health, 2010. **31**: p. 9-20 4 p following 20.

54. Cowper-Sal lari, R., et al., *Breast cancer risk-associated SNPs modulate the affinity of chromatin for FOXA1 and alter gene expression.* Nat Genet, 2012. **44**(11): p. 1191-8.

55. Kumar, P., S. Henikoff, and P.C. Ng, *Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm.* Nat Protoc, 2009. **4**(7): p. 1073-81.

56. Choi, Y., et al., *Predicting the functional effect of amino acid substitutions and indels.* PLoS One, 2012. **7**(10): p. e46688.

57. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

58. Gonzalez-Perez, A. and N. Lopez-Bigas, *Improving the assessment of the outcome of nonsynonymous SNVs with a consensus deleteriousness score, Condel.* Am J Hum Genet, 2011. **88**(4): p. 440-9.

59. Kircher, M., et al., *A general framework for estimating the relative pathogenicity of human genetic variants.* Nat Genet, 2014. **46**(3): p. 310-5.

60. Consortium, E.P., *An integrated encyclopedia of DNA elements in the human genome.* Nature, 2012. **489**(7414): p. 57-74.

61. Trynka, G., et al., *Chromatin marks identify critical cell types for fine mapping complex trait variants.* Nat Genet, 2013. **45**(2): p. 124-30.

62. Barski, A., et al., *High-resolution profiling of histone methylations in the human genome.* Cell, 2007. **129**(4): p. 823-37.

63. Johnson, D.S., et al., *Genome-wide mapping of in vivo protein-DNA interactions.* Science, 2007. **316**(5830): p. 1497-502.

64. Chu, Y. and D.R. Corey, *RNA sequencing: platform selection, experimental design, and data interpretation.* Nucleic Acid Ther, 2012. **22**(4): p. 271-4.

65. Wang, Z., M. Gerstein, and M. Snyder, *RNA-Seq: a revolutionary tool for transcriptomics.* Nat Rev Genet, 2009. **10**(1): p. 57-63.

66. Roadmap Epigenomics, C., et al., *Integrative analysis of 111 reference human epigenomes.* Nature, 2015. **518**(7539): p. 317-30.

67. Consortium, G.T., *The Genotype-Tissue Expression (GTEx) project.* Nat Genet, 2013. **45**(6): p. 580-5.

68. Carithers, L.J. and H.M. Moore, *The Genotype-Tissue Expression (GTEx) Project.* Biopreserv Biobank, 2015. **13**(5): p. 307-8.

69. Visel, A., E.M. Rubin, and L.A. Pennacchio, *Genomic views of distant-acting enhancers.* Nature, 2009. **461**(7261): p. 199-205.

70. Brodie, A., J.R. Azaria, and Y. Ofran, *How far from the SNP may the causative genes be?* Nucleic Acids Res, 2016. **44**(13): p. 6046-54.

71. Blackwood, E.M. and J.T. Kadonaga, *Going the distance: a current view of enhancer action.* Science, 1998. **281**(5373): p. 60-3.

72. Williams, A., C.G. Spilianakis, and R.A. Flavell, *Interchromosomal association and gene regulation in trans.* Trends Genet, 2010. **26**(4): p. 188-97.

73. Lieberman-Aiden, E., et al., *Comprehensive mapping of long-range interactions reveals folding principles of the human genome.* Science, 2009. **326**(5950): p. 289-93.

74. Ong, C.T. and V.G. Corces, *Enhancer function: new insights into the regulation of tissue-specific gene expression.* Nat Rev Genet, 2011. **12**(4): p. 283-93.

75. Petrascheck, M., et al., *DNA looping induced by a transcriptional enhancer in vivo.* Nucleic Acids Res, 2005. **33**(12): p. 3743-50.

76. Mishra, A. and R.D. Hawkins, *Three-dimensional genome architecture and emerging technologies: looping in disease.* Genome Med, 2017. **9**(1): p. 87.

77. Dekker, J., et al., *Capturing chromosome conformation.* Science, 2002. **295**(5558): p. 1306-11.

78. Simonis, M., et al., *Nuclear organization of active and inactive chromatin domains uncovered by chromosome conformation capture-on-chip (4C).* Nat Genet, 2006. **38**(11): p. 1348-54.

79. Zhao, Z., et al., *Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions.* Nat Genet, 2006. **38**(11): p. 1341-7.

80. Belton, J.M., et al., *Hi-C: a comprehensive technique to capture the conformation of genomes.* Methods, 2012. **58**(3): p. 268-76.

81. van Berkum, N.L., et al., *Hi-C: a method to study the three-dimensional architecture of genomes.* J Vis Exp, 2010(39).

82. Rivera, C.M. and B. Ren, *Mapping human epigenomes.* Cell, 2013. **155**(1): p. 39-55.

83. Watanabe, K., et al., *Functional mapping and annotation of genetic associations with FUMA.* Nat Commun, 2017. **8**(1): p. 1826.

84. Schofield, E.C., et al., *CHiCP: a web-based tool for the integrative and interactive visualization of promoter capture Hi-C datasets.* Bioinformatics, 2016. **32**(16): p. 2511-3.

85. Hemerich, D., et al., *Integrative Bioinformatics Approaches for Identification of Drug Targets in Hypertension.* Front Cardiovasc Med, 2018. **5**: p. 25.

86. Zhu, Z., et al., *Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets.* Nat Genet, 2016. **48**(5): p. 481-7.

87. Pickrell, J.K., et al., *Detection and interpretation of shared genetic influences on 42 human traits.* Nat Genet, 2016. **48**(7): p. 709-17.

88. Visscher, P.M. and J. Yang, *A plethora of pleiotropy across complex traits.* Nat Genet, 2016. **48**(7): p. 707-8.

89. Consortium, G.T., et al., *Genetic effects on gene expression across human tissues.* Nature, 2017. **550**(7675): p. 204-213.

90. Kannel, W.B., et al., *Factors of risk in the development of coronary heart disease--six year follow-up experience. The Framingham Study.* Ann Intern Med, 1961. **55**: p. 33-50.

91. Gersh, B.J., et al., *Novel therapeutic concepts: the epidemic of cardiovascular disease in the developing world: global implications.* Eur Heart J, 2010. **31**(6): p. 642-8.

92. Yusuf, S., et al., *Effect of potentially modifiable risk factors associated with myocardial infarction in 52 countries (the INTERHEART study): case-control study.* Lancet, 2004. **364**(9438): p. 937-52.

93. Walmsley, R.M. and N. Billinton, *How accurate is in vitro prediction of carcinogenicity?* British journal of pharmacology, 2011. **162**(6): p. 1250-1258.

94. Lipsky, M.S. and L.K. Sharp, *From idea to market: the drug approval process.* The Journal of the American Board of Family Practice, 2001. **14**(5): p. 362-367.

95. Mullard, A., *Industry R&D returns slip.* Nature Reviews Drug Discovery, 2016. **15**(1): p. 7-7.

96. Kessler, T., B. Vilne, and H. Schunkert, *The impact of genome-wide association studies on the pathophysiology and therapy of cardiovascular disease.* EMBO Mol Med, 2016. **8**(7): p. 688-701.

97. Hajduk, P.J., J.R. Huth, and C. Tse, *Predicting protein druggability.* Drug Discovery Today, 2005. **10**(23–24): p. 1675-1682.

98. Rockey, W.M. and A.H. Elcock, *Progress toward virtual screening for drug side effects.* Proteins, 2002. **48**(4): p. 664-71.

99. Li, Y.Y., J. An, and S.J.M. Jones, *A Large-Scale Computational Approach to Drug Repositioning.* Genome Informatics, 2006. **17**(2): p. 239-247.

100. Sanseau, P., et al., *Use of genome-wide association studies for drug repositioning.* Nat Biotechnol, 2012. **30**(4): p. 317-20.

101. Oprea, T. and J. Mestres, *Drug repurposing: far beyond new targets for old drugs.* The AAPS journal, 2012. **14**(4): p. 759-763.

102. Arrowsmith, J. and R. Harrison, *Drug Repositioning: The Business Case and Current Strategies to Repurpose Shelved Candidates and Marketed Drugs*. Drug Repositioning: Bringing New Life to Shelved Assets and Existing Drugs. 2012. 9.

103. Melnikova, I., *Rare diseases and orphan drugs.* Nature Reviews Drug Discovery, 2012. **11**(4): p. 267-268.

104. Galiè, N., et al., *Sildenafil citrate therapy for pulmonary arterial hypertension.* New England Journal of Medicine, 2005. **353**(20): p. 2148-2157.

105. Everett, B.M., R.J. Smith, and W.R. Hiatt, *Reducing LDL with PCSK9 inhibitors—the clinical benefit of lipid drugs.* New England Journal of Medicine, 2015. **373**(17): p. 1588-1591.

106. Schwartz, G.G., et al., *Effect of alirocumab, a monoclonal antibody to PCSK9, on long-term cardiovascular outcomes following acute coronary syndromes: rationale and design of the ODYSSEY outcomes trial.* American heart journal, 2014. **168**(5): p. 682-689. e1.

107. Raal, F.J., et al., *PCSK9 inhibition with evolocumab (AMG 145) in heterozygous familial hypercholesterolaemia (RUTHERFORD-2): a randomised, double-blind, placebo-controlled trial.* The Lancet, 2015. **385**(9965): p. 331-340.

108. Munoz, M., et al., *Evaluating the contribution of genetics and familial shared environment to common disease using the UK Biobank.* Nat Genet, 2016. **48**(9): p. 980-3.

109. Feinleib, M., et al., *The NHLBI twin study of cardiovascular disease risk factors: methodology and summary of results.* Am J Epidemiol, 1977. **106**(4): p. 284-5.

110. Mongeau, J.G., P. Biron, and C.F. Sing, *The influence of genetics and household environment upon the variability of normal blood pressure: the Montreal Adoption Survey.* Clin Exp Hypertens A, 1986. **8**(4-5): p. 653-60.

111. Bromfield, S. and P. Muntner, *High blood pressure: the leading global burden of disease risk factor and the need for worldwide prevention programs.* Curr Hypertens Rep, 2013. **15**(3): p. 134-6.

112. Levy, D., et al., *Genome-wide association study of blood pressure and hypertension.* Nat Genet, 2009. **41**(6): p. 677-87.

113. Newton-Cheh, C., et al., *Association of common variants in NPPA and NPPB with circulating natriuretic peptides and blood pressure.* Nat Genet, 2009. **41**(3): p. 348-53.

114. Talmud, P.J., et al., *Gene-centric association signals for lipids and apolipoproteins identified via the HumanCVD BeadChip.* Am J Hum Genet, 2009. **85**(5): p. 628-42.

115. Padmanabhan, S., et al., *Genome-wide association study of blood pressure extremes identifies variant near UMOD associated with hypertension.* PLoS Genet, 2010. **6**(10): p. e1001177.

116. International Consortium for Blood Pressure Genome-Wide Association, S., et al., *Genetic variants in novel pathways influence blood pressure and cardiovascular disease risk.* Nature, 2011. **478**(7367): p. 103-9.

117. Franceschini, N., A.P. Reiner, and G. Heiss, *Recent findings in the genetics of blood pressure and hypertension traits.* Am J Hypertens, 2011. **24**(4): p. 392-400.

118. Consortium, I.K.C., *Large-scale gene-centric analysis identifies novel variants for coronary artery disease.* PLoS Genet, 2011. **7**(9): p. e1002260.

119. Johnson, A.D., et al., *Association of hypertension drug target genes with blood pressure and hypertension in 86,588 individuals.* Hypertension, 2011. **57**(5): p. 903-10.

120. Johnson, T., et al., *Blood pressure loci identified with a gene-centric array.* Am J Hum Genet, 2011. **89**(6): p. 688-700.

121. Padmanabhan, S., C. Newton-Cheh, and A.F. Dominiczak, *Genetic basis of blood pressure and hypertension.* Trends Genet, 2012. **28**(8): p. 397-408.

122. Franceschini, N., et al., *Genome-wide association analysis of blood-pressure traits in African-ancestry individuals reveals common associated genes in African and non-African populations.* Am J Hum Genet, 2013. **93**(3): p. 545-54.

123. Ganesh, S.K., et al., *Loci influencing blood pressure identified using a cardiovascular gene-centric array.* Hum Mol Genet, 2013. **22**(8): p. 1663-78.

124. Tragante, V., et al., *Gene-centric meta-analysis in 87,736 individuals of European ancestry identifies multiple blood-pressure-related loci.* Am J Hum Genet, 2014. **94**(3): p. 349-60.

125. Surendran, P., et al., *Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension.* Nat Genet, 2016. **48**(10): p. 1151-1161.

126. Kato, N., et al., *Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation.* Nat Genet, 2015. **47**(11): p. 1282-1293.

127. Warren, H.R., et al., *Genome-wide association analysis identifies novel blood pressure loci and offers biological insights into cardiovascular risk.* Nat Genet, 2017. **49**(3): p. 403-415.

128. Hoffmann, T.J., et al., *Genome-wide association analyses using electronic health records identify new loci influencing blood pressure variation.* Nat Genet, 2017. **49**(1): p. 54-64.

129. Liu, C., et al., *Meta-analysis identifies common and rare variants influencing blood pressure and overlapping with metabolic trait loci.* Nat Genet, 2016. **48**(10): p. 1162-70.

130. Ehret, G.B., et al., *The genetics of blood pressure regulation and its target organs from association studies in 342,415 individuals.* Nat Genet, 2016. **48**(10): p. 1171-1184.

131. Ganesh, S.K., et al., *Effects of long-term averaging of quantitative blood pressure traits on the detection of genetic associations.* Am J Hum Genet, 2014. **95**(1): p. 49-65.

132. Kato, N., et al., *Meta-analysis of genome-wide association studies identifies common variants associated with blood pressure variation in east Asians.* Nat Genet, 2011. **43**(6): p. 531-8.

133. Wang, Y., et al., *From the Cover: Whole-genome association study identifies STK39 as a hypertension susceptibility gene.* Proc Natl Acad Sci U S A, 2009. **106**(1): p. 226-31.

134. Wain, L.V., et al., *Genome-wide association study identifies six new loci influencing pulse pressure and mean arterial pressure.* Nat Genet, 2011. **43**(10): p. 1005-11.

135. Simino, J., et al., *Gene-age interactions in blood pressure regulation: a large-scale investigation with the CHARGE, Global BPgen, and ICBP Consortia.* Am J Hum Genet, 2014. **95**(1): p. 24-38.

136. Fox, E.R., et al., *Association of genetic variation with systolic and diastolic blood pressure among African Americans: the Candidate Gene Association Resource study.* Hum Mol Genet, 2011. **20**(11): p. 2273-84.

137. Zhu, X., et al., *Meta-analysis of correlated traits via summary statistics from GWASs with an application in hypertension.* Am J Hum Genet, 2015. **96**(1): p. 21-36.

138. Ho, J.E., et al., *Discovery and replication of novel blood pressure genetic loci in the Women's Genome Health Study.* J Hypertens, 2011. **29**(1): p. 62-9.

139. Parmar, P.G., et al., *International Genome-Wide Association Study Consortium Identifies Novel Loci Associated With Blood Pressure in Children and Adolescents.* Circ Cardiovasc Genet, 2016. **9**(3): p. 266-278.

140. Kupper, N., et al., *Heritability of daytime ambulatory blood pressure in an extended twin design.* Hypertension, 2005. **45**(1): p. 80-5.

141. Staley, J.R., et al., *PhenoScanner: a database of human genotype-phenotype associations.* Bioinformatics, 2016. **32**(20): p. 3207-3209.

142. Padmanabhan, S., M. Caulfield, and A.F. Dominiczak, *Genetic and molecular aspects of hypertension.* Circ Res, 2015. **116**(6): p. 937-59.

143. Wang, X., et al., *Beyond genome-wide association studies: new strategies for identifying genetic determinants of hypertension.* Curr Hypertens Rep, 2011. **13**(6): p. 442-51.

144. Munroe, P.B., M.R. Barnes, and M.J. Caulfield, *Advances in blood pressure genomics.* Circ Res, 2013. **112**(10): p. 1365-79.

145. Tak, Y.G. and P.J. Farnham, *Making sense of GWAS: using epigenomics and genome engineering to understand the functional relevance of SNPs in non-coding regions of the human genome.* Epigenetics Chromatin, 2015. **8**: p. 57.

146. Coetzee, S.G., et al., *FunciSNP: an R/bioconductor tool integrating functional non-coding data sets with genetic association studies to identify candidate regulatory SNPs.* Nucleic Acids Res, 2012. **40**(18): p. e139.

147. McLaren, W., et al., *The Ensembl Variant Effect Predictor.* Genome Biol, 2016. **17**(1): p. 122.

148. Boyle, A.P., et al., *Annotation of functional variation in personal genomes using RegulomeDB.* Genome Res, 2012. **22**(9): p. 1790-7.

149. Marmorstein, R. and R.C. Trievel, *Histone modifying enzymes: structures, mechanisms, and specificities.* Biochim Biophys Acta, 2009. **1789**(1): p. 58-68.

150. Jenuwein, T. and C.D. Allis, *Translating the histone code.* Science, 2001. **293**(5532): p. 1074-80.

151. Sullivan, B.A. and G.H. Karpen, *Centromeric chromatin exhibits a histone modification pattern that is distinct from both euchromatin and heterochromatin.* Nat Struct Mol Biol, 2004. **11**(11): p. 1076-83.

152. Fischle, W., Y. Wang, and C.D. Allis, *Binary switches and modification cassettes in histone biology and beyond.* Nature, 2003. **425**(6957): p. 475-9.

153. Wise, I.A. and F.J. Charchar, *Epigenetic Modifications in Essential Hypertension.* Int J Mol Sci, 2016. **17**(4): p. 451.

154. Holliday, R. and J.E. Pugh, *DNA modification mechanisms and gene activity during development.* Science, 1975. **187**(4173): p. 226-32.

155. Riggs, A.D., *X inactivation, differentiation, and DNA methylation.* Cytogenet Cell Genet, 1975. **14**(1): p. 9-25.

156. Frey, F.J., *Methylation of CpG islands: potential relevance for hypertension and kidney diseases.* Nephrol Dial Transplant, 2005. **20**(5): p. 868-9.

157. Smolarek, I., et al., *Global DNA methylation changes in blood of patients with essential hypertension.* Med Sci Monit, 2010. **16**(3): p. CR149-155.

158. Raftopoulos, L., et al., *Epigenetics, the missing link in hypertension.* Life Sci, 2015. **129**: p. 22-6.

159. Muhonen, P. and H. Holthofer, *Epigenetic and microRNA-mediated regulation in diabetes.* Nephrol Dial Transplant, 2009. **24**(4): p. 1088-96.

160. Mill, J., et al., *Epigenomic profiling reveals DNA-methylation changes associated with major psychosis.* Am J Hum Genet, 2008. **82**(3): p. 696-711.

161. Albert, F.W. and L. Kruglyak, *The role of regulatory variation in complex traits and disease.* Nat Rev Genet, 2015. **16**(4): p. 197-212.

162. Gusev, A., et al., *Integrative approaches for large-scale transcriptome-wide association studies.* Nat Genet, 2016. **48**(3): p. 245-52.

163. Musunuru, K., et al., *From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus.* Nature, 2010. **466**(7307): p. 714-9.

164. Gamazon, E.R., et al., *A gene-based association method for mapping traits using reference transcriptome data.* Nat Genet, 2015. **47**(9): p. 1091-8.

165. Xia, K., et al., *seeQTL: a searchable database for human eQTLs.* Bioinformatics, 2012. **28**(3): p. 451-2.

166. Consortium, G.T., *Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans.* Science, 2015. **348**(6235): p. 648-60.

167. Yang, T.P., et al., *Genevar: a database and Java application for the analysis and visualization of SNP-gene associations in eQTL studies.* Bioinformatics, 2010. **26**(19): p. 2474-6.

168. Westra, H.J., et al., *Systematic identification of trans eQTLs as putative drivers of known disease associations.* Nat Genet, 2013. **45**(10): p. 1238-1243.

169. Lappalainen, T., et al., *Transcriptome and genome sequencing uncovers functional variation in humans.* Nature, 2013. **501**(7468): p. 506-11.

170. Spain, S.L. and J.C. Barrett, *Strategies for fine-mapping complex traits.* Hum Mol Genet, 2015. **24**(R1): p. R111-9.

171. Javierre, B.M., et al., *Lineage-Specific Genome Architecture Links Enhancers and Non-coding Disease Variants to Target Gene Promoters.* Cell, 2016. **167**(5): p. 1369-1384 e19.

172. Kraja, A.T., et al., *New Blood Pressure-Associated Loci Identified in Meta-Analyses of 475 000 Individuals.* Circ Cardiovasc Genet, 2017. **10**(5).

173. Rao, S.S., et al., *A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping.* Cell, 2014. **159**(7): p. 1665-80.

174. Schmitt, A.D., et al., *A Compendium of Chromatin Contact Maps Reveals Spatially Active Regions in the Human Genome.* Cell Rep, 2016. **17**(8): p. 2042-2059.

175. Finan, C., et al., *The druggable genome and support for target identification and validation in drug development.* Sci Transl Med, 2017. **9**(383).

176. Rosa-Garrido, M., et al., *High-Resolution Mapping of Chromatin Conformation in Cardiac Myocytes Reveals Structural Remodeling of the Epigenome in Heart Failure.* Circulation, 2017. **136**(17): p. 1613-1625.

177. Liljedahl, U., et al., *Single nucleotide polymorphisms in the apolipoprotein B and low density lipoprotein receptor genes affect response to antihypertensive treatment.* BMC Cardiovasc Disord, 2004. **4**(1): p. 16.

178. Chandrashekar, B.S., et al., *Topical minoxidil fortified with finasteride: An account of maintenance of hair density after replacing oral finasteride.* Indian Dermatol Online J, 2015. **6**(1): p. 17-20.

179. Harakalova, M., et al., *Dominant missense mutations in ABCC9 cause Cantu syndrome.* Nat Genet, 2012. **44**(7): p. 793-6.

180. Trynka, G., et al., *Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease.* Nat Genet, 2011. **43**(12): p. 1193-201.

181. Stahl, E.A., et al., *Bayesian inference analyses of the polygenic architecture of rheumatoid arthritis.* Nat Genet, 2012. **44**(5): p. 483-9.

182. Edwards, S.L., et al., *Beyond GWASs: illuminating the dark road from association to function.* Am J Hum Genet, 2013. **93**(5): p. 779-97.

183. Cong, L., et al., *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

184. Atlasi, Y. and H.G. Stunnenberg, *The interplay of epigenetic marks during stem cell differentiation and development.* Nat Rev Genet, 2017. **18**(11): p. 643-658.

185. Levy, D., et al., *Determinants of sensitivity and specificity of electrocardiographic criteria for left ventricular hypertrophy.* Circulation, 1990. **81**(3): p. 815-20.

186. Okin, P.M., et al., *Time-voltage QRS area of the 12-lead electrocardiogram: detection of left ventricular hypertrophy.* Hypertension, 1998. **31**(4): p. 937-42.

187. Kannel, W.B., T. Gordon, and D. Offutt, *Left ventricular hypertrophy by electrocardiogram. Prevalence, incidence, and mortality in the Framingham study.* Ann Intern Med, 1969. **71**(1): p. 89-105.

188. Verdecchia, P., et al., *Prognostic value of a new electrocardiographic method for diagnosis of left ventricular hypertrophy in essential hypertension.* J Am Coll Cardiol, 1998. **31**(2): p. 383-90.

189. Usoro, A.O., et al., *Risk of mortality in individuals with low QRS voltage and free of cardiovascular disease.* Am J Cardiol, 2014. **113**(9): p. 1514-7.

190. Kamath, S.A., et al., *Low voltage on the electrocardiogram is a marker of disease severity and a risk factor for adverse outcomes in patients with heart failure due to systolic dysfunction.* Am Heart J, 2006. **152**(2): p. 355-61.

191. van der Harst, P., et al., *52 Genetic Loci Influencing Myocardial Mass.* J Am Coll Cardiol, 2016. **68**(13): p. 1435-1448.

192. Corradin, O. and P.C. Scacheri, *Enhancer variants: evaluating functions in common disease.* Genome Med, 2014. **6**(10): p. 85.

193. Maurano, M.T., et al., *Systematic localization of common disease-associated variation in regulatory DNA.* Science, 2012. **337**(6099): p. 1190-5.

194. Schaub, M.A., et al., *Linking disease associations with regulatory information in the human genome.* Genome Res, 2012. **22**(9): p. 1748-59.

195. Gupta, R.M., et al., *A Genetic Variant Associated with Five Vascular Diseases Is a Distal Regulator of Endothelin-1 Gene Expression.* Cell, 2017. **170**(3): p. 522-533 e15.

196. Smemo, S., et al., *Obesity-associated variants within FTO form long-range functional connections with IRX3.* Nature, 2014. **507**(7492): p. 371-5.

197. Yao, L., et al., *Functional annotation of colon cancer risk SNPs.* Nat Commun, 2014. **5**: p. 5114.

198. Dobin, A., et al., *STAR: ultrafast universal RNA-seq aligner.* Bioinformatics, 2013. **29**(1): p. 15-21.

199. Tarasov, A., et al., *Sambamba: fast processing of NGS alignment formats.* Bioinformatics, 2015. **31**(12): p. 2032-4.

200. Anders, S., P.T. Pyl, and W. Huber, *HTSeq--a Python framework to work with high-throughput sequencing data.* Bioinformatics, 2015. **31**(2): p. 166-9.

201. Robinson, M.D., D.J. McCarthy, and G.K. Smyth, *edgeR: a Bioconductor package for differential expression analysis of digital gene expression data.* Bioinformatics, 2010. **26**(1): p. 139-40.

202. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.

203. Kinsella, R.J., et al., *Ensembl BioMarts: a hub for data retrieval across taxonomic space.* Database (Oxford), 2011. **2011**: p. bar030.

204. Eijkelenboom, A., et al., *Genome-wide analysis of FOXO3 mediated transcription regulation through RNA polymerase II profiling.* Mol Syst Biol, 2013. **9**: p. 638.

205. Li, H. and R. Durbin, *Fast and accurate short read alignment with Burrows-Wheeler transform.* Bioinformatics, 2009. **25**(14): p. 1754-60.

206. Li, H., et al., *The Sequence Alignment/Map format and SAMtools.* Bioinformatics, 2009. **25**(16): p. 2078-9.

207. Zhang, Y., et al., *Model-based analysis of ChIP-Seq (MACS).* Genome Biol, 2008. **9**(9): p. R137.

208. Kharchenko, P.V., M.Y. Tolstorukov, and P.J. Park, *Design and analysis of ChIP-seq experiments for DNA-binding proteins.* Nat Biotechnol, 2008. **26**(12): p. 1351-9.

209. Ross-Innes, C.S., et al., *Differential oestrogen receptor binding is associated with clinical outcome in breast cancer.* Nature, 2012. **481**(7381): p. 389-93.

210. Whyte, W.A., et al., *Master transcription factors and mediator establish super-enhancers at key cell identity genes.* Cell, 2013. **153**(2): p. 307-19.

211. Quinlan, A.R. and I.M. Hall, *BEDTools: a flexible suite of utilities for comparing genomic features.* Bioinformatics, 2010. **26**(6): p. 841-2.

212. Neph, S., et al., *BEDOPS: high-performance genomic feature operations.* Bioinformatics, 2012. **28**(14): p. 1919-20.

213. Wei, Y., et al., *SEA: a super-enhancer archive.* Nucleic Acids Res, 2016. **44**(D1): p. D172-9.

214. International HapMap, C., et al., *A second generation human haplotype map of over 3.1 million SNPs.* Nature, 2007. **449**(7164): p. 851-61.

215. Flister, M.J., et al., *Identifying multiple causative genes at a single GWAS locus.* Genome Res, 2013. **23**(12): p. 1996-2002.

216. Catarino, R.R., C. Neumayr, and A. Stark, *Promoting transcription over long distances.* Nat Genet, 2017. **49**(7): p. 972-973.

217. Dao, L.T.M., et al., *Genome-wide characterization of mammalian promoters with distal enhancer functions.* Nat Genet, 2017. **49**(7): p. 1073-1081.

218. Diao, Y., et al., *A tiling-deletion-based genetic screen for cis-regulatory element identification in mammalian cells.* Nat Methods, 2017. **14**(6): p. 629-635.

219. Sandhu, K.S., et al., *Large-scale functional organization of long-range chromatin interaction networks.* Cell Rep, 2012. **2**(5): p. 1207-19.

220. Elkon, R. and R. Agami, *Characterization of noncoding regulatory DNA in the human genome.* Nat Biotechnol, 2017. **35**(8): p. 732-746.

221. Xia, Q., et al., *The type 2 diabetes presumed causal variant within TCF7L2 resides in an element that controls the expression of ACSL5.* Diabetologia, 2016. **59**(11): p. 2360-2368.

222. Pradeepa, M.M., et al., *Histone H3 globular domain acetylation identifies a new class of enhancers.* Nat Genet, 2016. **48**(6): p. 681-6.

223. Liu, L., G. Jin, and X. Zhou, *Modeling the relationship of epigenetic modifications to transcription factor binding.* Nucleic Acids Res, 2015. **43**(8): p. 3873-85.

224. Kent, W.J., et al., *The human genome browser at UCSC.* Genome Res, 2002. **12**(6): p. 996-1006.

225. Sonoda, H., et al., *A novel phosphatidic acid-selective phospholipase A1 that produces lysophosphatidic acid.* J Biol Chem, 2002. **277**(37): p. 34254-63.

226. den Hoed, M., et al., *Identification of heart rate-associated loci and their effects on cardiac conduction and rhythm disorders.* Nat Genet, 2013. **45**(6): p. 621-31.

227. *Gene: Dpy19l1*. June 11, 2018]; Available from: http://www.mousephenotype.org/data/genes/MGI:1915685.

228. *Gene: Ttc39a*. June 11, 2018]; Available from: http://www.mousephenotype.org/data/genes/MGI:2444350.

229. van der Harst, P. and N. Verweij, *Identification of 64 Novel Genetic Loci Provides an Expanded View on the Genetic Architecture of Coronary Artery Disease.* Circ Res, 2018. **122**(3): p. 433-443.

230. Dennis, J., et al., *The endothelial protein C receptor (PROCR) Ser219Gly variant and risk of common thrombotic disorders: a HuGE review and meta-analysis of evidence from observational studies.* Blood, 2012. **119**(10): p. 2392-400.

231. O'Leary, N.A., et al., *Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.* Nucleic Acids Res, 2016. **44**(D1): p. D733-45.

232. Fagerberg, L., et al., *Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics.* Mol Cell Proteomics, 2014. **13**(2): p. 397-406.

233. Verweij, N., et al., *Twenty-eight genetic loci associated with ST-T-wave amplitudes of the electrocardiogram.* Hum Mol Genet, 2016. **25**(10): p. 2093-2103.

234. Arking, D.E., et al., *Genetic association study of QT interval highlights role for calcium signaling pathways in myocardial repolarization.* Nat Genet, 2014. **46**(8): p. 826-36.

235. Orthaus, S., et al., *Assembly of the inner kinetochore proteins CENP-A and CENP-B in living human cells.* Chembiochem, 2008. **9**(1): p. 77-92.

236. Shelby, R.D., O. Vafa, and K.F. Sullivan, *Assembly of CENP-A into centromeric chromatin requires a cooperative array of nucleosomal DNA contact sites.* J Cell Biol, 1997. **136**(3): p. 501-13.

237. Black, B.E., et al., *Structural determinants for generating centromeric chromatin.* Nature, 2004. **430**(6999): p. 578-82.

238. Plant, L.D., et al., *SUMOylation silences heterodimeric TASK potassium channels containing K2P1 subunits in cerebellar granule neurons.* Sci Signal, 2012. **5**(251): p. ra84.

239. Mirtschink, P. and W. Krek, *Hypoxia-driven glycolytic and fructolytic metabolic programs: Pivotal to hypertrophic heart disease.* Biochim Biophys Acta, 2016. **1863**(7 Pt B): p. 1822-8.

240. Mirtschink, P., et al., *HIF-driven SF3B1 induces KHK-C to enforce fructolysis and heart disease.* Nature, 2015. **522**(7557): p. 444-449.

241. Malerba, G., et al., *SNPs of the FADS gene cluster are associated with polyunsaturated fatty acids in a cohort of patients with cardiovascular disease.* Lipids, 2008. **43**(4): p. 289-99.

242. Mestroni, L., et al., *Guidelines for the study of familial dilated cardiomyopathies. Collaborative Research Group of the European Human and Capital Mobility Project on Familial Dilated Cardiomyopathy.* Eur Heart J, 1999. **20**(2): p. 93-102.

243. Hershberger, R.E., et al., *Progress with genetic cardiomyopathies: screening, counseling, and testing in dilated, hypertrophic, and arrhythmogenic right ventricular dysplasia/cardiomyopathy.* Circ Heart Fail, 2009. **2**(3): p. 253-61.

244. Kayvanpour, E., et al., *Genotype-phenotype associations in dilated cardiomyopathy: meta-analysis on more than 8000 individuals.* Clin Res Cardiol, 2017. **106**(2): p. 127-139.

245. Kayvanpour, E., H.A. Katus, and B. Meder, *Determined to Fail--the Role of Genetic Mechanisms in Heart Failure.* Curr Heart Fail Rep, 2015. **12**(5): p. 333-8.

246. Watkins, H., H. Ashrafian, and C. Redwood, *Inherited cardiomyopathies.* N Engl J Med, 2011. **364**(17): p. 1643-56.

247. MacRae, C.A., *Mendelian forms of structural cardiovascular disease.* Curr Cardiol Rep, 2013. **15**(10): p. 399.

248. Kathiresan, S. and D. Srivastava, *Genetics of human cardiovascular disease.* Cell, 2012. **148**(6): p. 1242-57.

249. Harakalova, M., et al., *A systematic analysis of genetic dilated cardiomyopathy reveals numerous ubiquitously expressed and muscle-specific genes.* Eur J Heart Fail, 2015. **17**(5): p. 484-93.

250. Hershberger, R.E., D.J. Hedges, and A. Morales, *Dilated cardiomyopathy: the complexity of a diverse genetic architecture.* Nat Rev Cardiol, 2013. **10**(9): p. 531-47.

251. Herman, D.S., et al., *Truncations of titin causing dilated cardiomyopathy.* N Engl J Med, 2012. **366**(7): p. 619-28.

252. Harakalova, M. and F.W. Asselbergs, *Systems analysis of dilated cardiomyopathy in the next generation sequencing era.* Wiley Interdiscip Rev Syst Biol Med, 2018. **10**(4): p. e1419.

253. Atchison, M.L., *Enhancers: mechanisms of action and cell specificity.* Annu Rev Cell Biol, 1988. **4**: p. 127-53.

254. Maston, G.A., S.K. Evans, and M.R. Green, *Transcriptional regulatory elements in the human genome.* Annu Rev Genomics Hum Genet, 2006. **7**: p. 29-59.

255. Bulger, M. and M. Groudine, *Functional and mechanistic diversity of distal transcription enhancers.* Cell, 2011. **144**(3): p. 327-39.

256. Ernst, J., et al., *Mapping and analysis of chromatin state dynamics in nine human cell types.* Nature, 2011. **473**(7345): p. 43-9.

257. Visel, A., J. Bristow, and L.A. Pennacchio, *Enhancer identification through comparative genomics.* Semin Cell Dev Biol, 2007. **18**(1): p. 140-52.

258. Visel, A., et al., *ChIP-seq accurately predicts tissue-specific activity of enhancers.* Nature, 2009. **457**(7231): p. 854-8.

259. Consortium, E.P., et al., *Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project.* Nature, 2007. **447**(7146): p. 799-816.

260. Thurman, R.E., et al., *The accessible chromatin landscape of the human genome.* Nature, 2012. **489**(7414): p. 75-82.

261. Bruneau, B.G., *The developmental genetics of congenital heart disease.* Nature, 2008. **451**(7181): p. 943-8.

262. Zaidi, S., et al., *De novo mutations in histone-modifying genes in congenital heart disease.* Nature, 2013. **498**(7453): p. 220-3.

263. Chen, J., et al., *ToppGene Suite for gene list enrichment analysis and candidate gene prioritization.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W305-11.

264. McLean, C.Y., et al., *GREAT improves functional interpretation of cis-regulatory regions.* Nat Biotechnol, 2010. **28**(5): p. 495-501.

265. McLeay, R.C. and T.L. Bailey, *Motif Enrichment Analysis: a unified framework and an evaluation on ChIP data.* BMC Bioinformatics, 2010. **11**: p. 165.

266. Kulakovskiy, I.V., et al., *HOCOMOCO: towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis.* Nucleic Acids Res, 2018. **46**(D1): p. D252-D259.

267. Roberts, A.M., et al., *Integrated allelic, transcriptional, and phenomic dissection of the cardiac effects of titin truncations in health and disease.* Sci Transl Med, 2015. **7**(270): p. 270ra6.

268. Cardena, M.M., et al., *A new duplication in the mitochondrially encoded tRNA proline gene in a patient with dilated cardiomyopathy.* Mitochondrial DNA, 2013. **24**(1): p. 46-9.

269. Hao, X.D., et al., *De novo mutations of TUBA3D are associated with keratoconus.* Sci Rep, 2017. **7**(1): p. 13570.

270. Arbustini, E., et al., *Mitochondrial DNA mutations and mitochondrial abnormalities in dilated cardiomyopathy.* Am J Pathol, 1998. **153**(5): p. 1501-10.

271. Li, Y.Y., et al., *Point mutations in mitochondrial DNA of patients with dilated cardiomyopathy.* J Mol Cell Cardiol, 1997. **29**(10): p. 2699-709.

272. Prohaszka, Z., et al., *Association of ficolin-3 with severity and outcome of chronic heart failure.* PLoS One, 2013. **8**(4): p. e60976.

273. Horvath, Z., et al., *Association of Low Ficolin-Lectin Pathway Parameters with Cardiac Syndrome X.* Scand J Immunol, 2016. **84**(3): p. 174-81.

274. Li, Y., et al., *Identification of a Common Different Gene Expression Signature in Ischemic Cardiomyopathy.* Genes (Basel), 2018. **9**(1).

275. Harrison, S.C., et al., *A gene-centric study of common carotid artery remodelling.* Atherosclerosis, 2013. **226**(2): p. 440-6.

276. Pham-Dinh, D., et al., *Myelin/oligodendrocyte glycoprotein is a member of a subset of the immunoglobulin superfamily encoded within the major histocompatibility complex.* Proc Natl Acad Sci U S A, 1993. **90**(17): p. 7990-4.

277. Nishimura, H., et al., *Autoimmune dilated cardiomyopathy in PD-1 receptor-deficient mice.* Science, 2001. **291**(5502): p. 319-22.

278. Okazaki, T., et al., *Autoantibodies against cardiac troponin I are responsible for dilated cardiomyopathy in PD-1-deficient mice.* Nat Med, 2003. **9**(12): p. 1477-83.

279. Okazaki, T., et al., *Hydronephrosis associated with antiurothelial and antinuclear autoantibodies in BALB/c-Fcgr2b-/-Pdcd1-/- mice.* J Exp Med, 2005. **202**(12): p. 1643-8.

280. Miller, M.K., et al., *The muscle ankyrin repeat proteins: CARP, ankrd2/Arpp and DARP as a family of titin filament-based stress response molecules.* J Mol Biol, 2003. **333**(5): p. 951-64.

281. Kemp, T.J., et al., *Identification of Ankrd2, a novel skeletal muscle gene coding for a stretch-responsive ankyrin-repeat protein.* Genomics, 2000. **66**(3): p. 229-41.

282. Pallavicini, A., et al., *Characterization of human skeletal muscle Ankrd2.* Biochem Biophys Res Commun, 2001. **285**(2): p. 378-86.

283. Nagueh, S.F., et al., *Altered titin expression, myocardial stiffness, and left ventricular function in patients with dilated cardiomyopathy.* Circulation, 2004. **110**(2): p. 155-62.

284. Nakada, C., et al., *Altered expression of ARPP protein in skeletal muscles of patients with muscular dystrophy, congenital myopathy and spinal muscular atrophy.* Pathobiology, 2004. **71**(1): p. 43-51.

285. Wang, W., et al., *Identification of miRNA, lncRNA and mRNA-associated ceRNA networks and potential biomarker for MELAS with mitochondrial DNA A3243G mutation.* Sci Rep, 2017. **7**: p. 41639.

286. Zhang, B., et al., *Kcnj11 Ablation Is Associated With Increased Nitro-Oxidative Stress During Ischemia-Reperfusion Injury: Implications for Human Ischemic Cardiomyopathy.* Circ Heart Fail, 2017. **10**(2).

287. di Salvo, T.G., et al., *Right ventricular myocardial biomarkers in human heart failure.* J Card Fail, 2015. **21**(5): p. 398-411.

288. Dashkevich, A., et al., *Immunohistochemical study of remodeling of myocardial lymphatic and blood microvascular structures in terminal heart failure: differences between ischemic and dilated cardiomyopathy.* Lymphology, 2010. **43**(3): p. 110-7.

289. Wardrop, K.E. and J.A. Dominov, *Proinflammatory signals and the loss of lymphatic vessel hyaluronan receptor-1 (LYVE-1) in the early pathogenesis of laminin alpha2-deficient skeletal muscle.* J Histochem Cytochem, 2011. **59**(2): p. 167-79.

290. Sharma, R., et al., *Proteomic Signature of Endothelial Dysfunction Identified in the Serum of Acute Ischemic Stroke Patients by the iTRAQ-Based LC-MS Approach.* J Proteome Res, 2015. **14**(6): p. 2466-79.

291. Heidrich, F.M., et al., *Regulation of circulating chromogranin B levels in heart failure.* Biomarkers, 2018. **23**(1): p. 78-87.

292. Rosjo, H., et al., *Chromogranin B in heart failure: a putative cardiac biomarker expressed in the failing myocardium.* Circ Heart Fail, 2010. **3**(4): p. 503-11.

293. Buitrago, M., et al., *The transcriptional repressor Nab1 is a specific regulator of pathological cardiac hypertrophy.* Nat Med, 2005. **11**(8): p. 837-44.

294. Kasneci, A., et al., *Egr-1 negatively regulates calsequestrin expression and calcium dynamics in ventricular cells.* Cardiovasc Res, 2009. **81**(4): p. 695-702.

295. Pacini, L., et al., *Altered calcium regulation in isolated cardiomyocytes from Egr-1 knock-out mice.* Can J Physiol Pharmacol, 2013. **91**(12): p. 1135-42.

296. Zins, K., et al., *Egr-1 upregulates Siva-1 expression and induces cardiac fibroblast apoptosis.* Int J Mol Sci, 2014. **15**(1): p. 1538-53.

297. Bondue, A., et al., *Defining the earliest step of cardiovascular progenitor specification during embryonic stem cell differentiation.* J Cell Biol, 2011. **192**(5): p. 751-65.

298. Koch, M., et al., *Collagen XXIV, a vertebrate fibrillar collagen with structural features of invertebrate collagens: selective expression in developing cornea and bone.* J Biol Chem, 2003. **278**(44): p. 43236-44.

299. Kittleson, M.M., et al., *Identification of a gene expression profile that differentiates between ischemic and nonischemic cardiomyopathy.* Circulation, 2004. **110**(22): p. 3444-51.

300. Koenig, X., et al., *Voltage-gated ion channel dysfunction precedes cardiomyopathy development in the dystrophic heart.* PLoS One, 2011. **6**(5): p. e20300.

301. Molina-Navarro, M.M., et al., *Differential gene expression of cardiac ion channels in human dilated cardiomyopathy.* PLoS One, 2013. **8**(12): p. e79792.

302. Hein, S., et al., *Deposition of nonsarcomeric alpha-actinin in cardiomyocytes from patients with dilated cardiomyopathy or chronic pressure overload.* Exp Clin Cardiol, 2009. **14**(3): p. e68-75.

303. Mahmoudabady, M., et al., *Activin-A, transforming growth factor-beta, and myostatin signaling pathway in experimental dilated cardiomyopathy.* J Card Fail, 2008. **14**(8): p. 703-9.

304. Dobaczewski, M., W. Chen, and N.G. Frangogiannis, *Transforming growth factor (TGF)-beta signaling in cardiac remodeling.* J Mol Cell Cardiol, 2011. **51**(4): p. 600-6.

305. Burke, M.A., et al., *Molecular profiling of dilated cardiomyopathy that progresses to heart failure.* JCI Insight, 2016. **1**(6).

306. Wai, T., et al., *Imbalanced OPA1 processing and mitochondrial fragmentation cause heart failure in mice.* Science, 2015. **350**(6265): p. aad0116.

307. Franco, A., et al., *Correcting mitochondrial fusion by manipulating mitofusin conformations.* Nature, 2016. **540**(7631): p. 74-79.

308. Bondue, A., et al., *Complex Roads from Genotype to Phenotype in Dilated Cardiomyopathy: Scientific update from the Working Group of Myocardial Function of the European Society of Cardiology.* Cardiovasc Res, 2018.

309. Freeman, K., et al., *Alterations in cardiac adrenergic signaling and calcium cycling differentially affect the progression of cardiomyopathy.* J Clin Invest, 2001. **107**(8): p. 967-74.

310. Sussman, M.A., et al., *Altered focal adhesion regulation correlates with cardiomyopathy in mice expressing constitutively active rac1.* J Clin Invest, 2000. **105**(7): p. 875-86.

311. Wei, L., et al., *Inhibition of Rho family GTPases by Rho GDP dissociation inhibitor disrupts cardiac morphogenesis and inhibits cardiomyocyte proliferation.* Development, 2002. **129**(7): p. 1705-14.

312. Chrzanowska-Wodnicka, M., *Distinct functions for Rap1 signaling in vascular morphogenesis and dysfunction.* Exp Cell Res, 2013. **319**(15): p. 2350-9.

313. Jeyaraj, S.C., N.T. Unger, and M.A. Chotani, *Rap1 GTPases: an emerging role in the cardiovasculature.* Life Sci, 2011. **88**(15-16): p. 645-52.

314. Ching, T., S. Huang, and L.X. Garmire, *Power analysis and sample size estimation for RNA-Seq differential expression.* RNA, 2014. **20**(11): p. 1684-96.

315. Lloyd-Jones, D., et al., *Executive summary: heart disease and stroke statistics--2010 update: a report from the American Heart Association.* Circulation, 2010. **121**(7): p. 948-54.

316. Falk, E., *Pathogenesis of atherosclerosis.* J Am Coll Cardiol, 2006. **47**(8 Suppl): p. C7-12.

317. Santos-Gallego, C.G., B. Picatoste, and J.J. Badimón, *Pathophysiology of acute coronary syndrome.* Current atherosclerosis reports, 2014. **16**(4): p. 1-9.

318. Amato, M., et al., *Carotid intima-media thickness by B-mode ultrasound as surrogate of coronary atherosclerosis: correlation with quantitative coronary angiography and coronary intravascular ultrasound findings.* Eur Heart J, 2007. **28**(17): p. 2094-101.

319. Lorenz, M.W., et al., *Prediction of clinical cardiovascular events with carotid intima-media thickness: a systematic review and meta-analysis.* Circulation, 2007. **115**(4): p. 459-67.

320. van der Meer, I.M., et al., *Predictive value of noninvasive measures of atherosclerosis for incident myocardial infarction: the Rotterdam Study.* Circulation, 2004. **109**(9): p. 1089-94.

321. Bis, J.C., et al., *Meta-analysis of genome-wide association studies from the CHARGE consortium identifies common variants associated with carotid intima media thickness and plaque.* Nat Genet, 2011. **43**(10): p. 940-7.

322. Gertow, K., et al., *Identification of the BCAR1-CFDP1-TMEM170A locus as a determinant of carotid intima-media thickness and coronary artery disease risk.* Circ Cardiovasc Genet, 2012. **5**(6): p. 656-65.

323. Simons, P.C., et al., *Second manifestations of ARTerial disease (SMART) study: rationale and design.* Eur J Epidemiol, 1999. **15**(9): p. 773-81.

324. Achterberg, S., et al., *Patients with coronary, cerebrovascular or peripheral arterial obstructive disease differ in risk for new vascular events and mortality: the SMART study.* Eur J Cardiovasc Prev Rehabil, 2010. **17**(4): p. 424-30.

325. Laurie, C.C., et al., *Quality control and quality assurance in genotypic data for genome-wide association studies.* Genet Epidemiol, 2010. **34**(6): p. 591-602.

326. Nakov, R., et al., *Darusentan: an effective endothelinA receptor antagonist for treatment of hypertension.* Am J Hypertens, 2002. **15**(7 Pt 1): p. 583-9.

327. Yasuda, H., et al., *Association of single nucleotide polymorphisms in endothelin family genes with the progression of atherosclerosis in patients with essential hypertension.* J Hum Hypertens, 2007. **21**(11): p. 883-92.

328. Rahman, T., et al., *Common genetic variation in the type A endothelin-1 receptor is associated with ambulatory blood pressure: a family study.* J Hum Hypertens, 2008. **22**(4): p. 282-8.

329. Consortium, C.A.D., et al., *Large-scale association analysis identifies new risk loci for coronary artery disease.* Nat Genet, 2013. **45**(1): p. 25-33.

330. Zhou, X.Z. and K.P. Lu, *The Pin2/TRF1-interacting protein PinX1 is a potent telomerase inhibitor.* Cell, 2001. **107**(3): p. 347-59.

331. Yuan, K., et al., *PinX1 is a novel microtubule-binding protein essential for accurate chromosome segregation.* J Biol Chem, 2009. **284**(34): p. 23072-82.

332. Zhang, B., et al., *Silencing PinX1 compromises telomere length maintenance as well as tumorigenicity in telomerase-positive human cancer cells.* Cancer research, 2009. **69**(1): p. 75-83.

333. Cai, M.Y., et al., *Decreased expression of PinX1 protein is correlated with tumor development and is a new independent poor prognostic factor in ovarian carcinoma.* Cancer science, 2010. **101**(6): p. 1543-1549.

334. Li, C., et al., *Genetic association and gene-smoking interaction study of carotid intima-media thickness at five GWAS-indicated genes: the Bogalusa Heart Study.* Gene, 2015. **562**(2): p. 226-31.

335. Lim, S.S., et al., *A comparative risk assessment of burden of disease and injury attributable to 67 risk factors and risk factor clusters in 21 regions, 1990-2010: a systematic analysis for the Global Burden of Disease Study 2010.* Lancet, 2012. **380**(9859): p. 2224-60.

336. Organization, W.H., *Global atlas on cardiovascular disease prevention and control.* 2011.

337. Brookes, M.J. and B.T. Cooper, *Hypertension and fatty liver: guilty by association?* J Hum Hypertens, 2007. **21**(4): p. 264-70.

338. MacMahon, S., et al., *Blood pressure, stroke, and coronary heart disease. Part 1, Prolonged differences in blood pressure: prospective observational studies corrected for the regression dilution bias.* Lancet, 1990. **335**(8692): p. 765-74.

339. Leonardi-Bee, J., et al., *Blood pressure and clinical outcomes in the International Stroke Trial.* Stroke, 2002. **33**(5): p. 1315-20.

340. Rodgers, A., et al., *Blood pressure and risk of stroke in patients with cerebrovascular disease. The United Kingdom Transient Ischaemic Attack Collaborative Group.* BMJ, 1996. **313**(7050): p. 147.

341. Tomaszewski, M., et al., *Genetic architecture of ambulatory blood pressure in the general population: insights from cardiovascular gene-centric array.* Hypertension, 2010. **56**(6): p. 1069-76.

342. Salfati, E., et al., *Direct Estimates of the Genomic Contributions to Blood Pressure Heritability within a Population-Based Cohort (ARIC).* PLoS One, 2015. **10**(7): p. e0133031.

343. Schork, N.J., et al., *Common vs. rare allele hypotheses for complex diseases.* Curr Opin Genet Dev, 2009. **19**(3): p. 212-9.

344. Surendran, P., et al., *Trans-ancestry meta-analyses identify rare and common variants associated with blood pressure and hypertension.* Nat Genet, 2016. **48**(10): p. 1151-61.

345. Kato, N., et al., *Trans-ancestry genome-wide association study identifies 12 genetic loci influencing blood pressure and implicates a role for DNA methylation.* Nat Genet, 2015. **47**(11): p. 1282-93.

346. Heintzman, N.D. and B. Ren, *Finding distal regulatory elements in the human genome.* Curr Opin Genet Dev, 2009. **19**(6): p. 541-9.

347. Heintzman, N.D., et al., *Histone modifications at human enhancers reflect global cell-type-specific gene expression.* Nature, 2009. **459**(7243): p. 108-12.

348. Rada-Iglesias, A., et al., *A unique chromatin signature uncovers early developmental enhancers in humans.* Nature, 2011. **470**(7333): p. 279-83.

349. Zentner, G.E., P.J. Tesar, and P.C. Scacheri, *Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions.* Genome Res, 2011. **21**(8): p. 1273-83.

350. Shlyueva, D., G. Stampfel, and A. Stark, *Transcriptional enhancers: from properties to genome-wide predictions.* Nat Rev Genet, 2014. **15**(4): p. 272-86.

351. Hardison, R.C., *Genome-wide epigenetic data facilitate understanding of disease susceptibility association studies.* J Biol Chem, 2012. **287**(37): p. 30932-40.

352. Li, H., *A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data.* Bioinformatics, 2011. **27**(21): p. 2987-93.

353. Genomes Project, C., et al., *A map of human genome variation from population-scale sequencing.* Nature, 2010. **467**(7319): p. 1061-73.

354. Sudlow, C., et al., *UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age.* PLoS Med, 2015. **12**(3): p. e1001779.

355. McCarthy, S., et al., *A reference panel of 64,976 haplotypes for genotype imputation.* Nat Genet, 2016. **48**(10): p. 1279-83.

356. Zhang, T., S. Cooper, and N. Brockdorff, *The interplay of histone modifications - writers that read.* EMBO Rep, 2015. **16**(11): p. 1467-81.

357. Swygert, S.G. and C.L. Peterson, *Chromatin dynamics: interplay between remodeling enzymes and histone modifications.* Biochim Biophys Acta, 2014. **1839**(8): p. 728-36.

358. Vaissiere, T., C. Sawan, and Z. Herceg, *Epigenetic interplay between histone modifications and DNA methylation in gene silencing.* Mutat Res, 2008. **659**(1-2): p. 40-8.

359. Zhang, Y. and D. Reinberg, *Transcription regulation by histone methylation: interplay between different covalent modifications of the core histone tails.* Genes Dev, 2001. **15**(18): p. 2343-60.

360. Strahl, B.D. and C.D. Allis, *The language of covalent histone modifications.* Nature, 2000. **403**(6765): p. 41-5.

361. Vos, T., et al., *Global, regional, and national incidence, prevalence, and years lived with disability for 301 acute and chronic diseases and injuries in 188 countries, 1990–2013: a systematic analysis for the Global Burden of Disease Study 2013.* The Lancet, 2015. **386**(9995): p. 743-800.

362. Perk, J., et al., *European Guidelines on cardiovascular disease prevention in clinical practice (version 2012).* European heart journal, 2012. **33**(13): p. 1635-1701.

363. Pursnani, A., et al., *Guideline-Based Statin Eligibility, Coronary Artery Calcification, and Cardiovascular Events.* JAMA, 2015. **314**(2): p. 134-41.

364. Schwartz, G.G., et al., *Effect of alirocumab, a monoclonal antibody to PCSK9, on long-term cardiovascular outcomes following acute coronary syndromes: rationale and design of the ODYSSEY outcomes trial.* Am Heart J, 2014. **168**(5): p. 682-9.

365. Raal, F.J., et al., *Inhibition of PCSK9 with evolocumab in homozygous familial hypercholesterolaemia (TESLA Part B): a randomised, double-blind, placebo-controlled trial.* Lancet, 2015. **385**(9965): p. 341-50.

366. Ridker, P.M., et al., *Interleukin-1beta inhibition and the prevention of recurrent cardiovascular events: rationale and design of the Canakinumab Anti-inflammatory Thrombosis Outcomes Study (CANTOS).* Am Heart J, 2011. **162**(4): p. 597-605.

367. Weber, C. and P. von Hundelshausen, *CANTOS Trial Validates the Inflammatory Pathogenesis of Atherosclerosis: Setting the Stage for a New Chapter in Therapeutic Targeting.* Circ Res, 2017. **121**(10): p. 1119-1121.

368. Arrowsmith, J. and P. Miller, *Trial watch: phase II and phase III attrition rates 2011-2012.* Nature Reviews Drug Discovery, 2013. **12**(8): p. 569-569.

369. Ashburn, T.T. and K.B. Thor, *Drug repositioning: identifying and developing new uses for existing drugs.* Nature reviews Drug discovery, 2004. **3**(8): p. 673-683.

370. Deloukas, P., et al., *Large-scale association analysis identifies new risk loci for coronary artery disease.* Nature genetics, 2013. **45**(1): p. 25-33.

371. Schunkert, H., et al., *Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease.* Nature Genetics, 2011. **43**(4): p. 333-338.

372. Lu, X., et al., *Genome-wide association study in Han Chinese identifies four new susceptibility loci for coronary artery disease.* Nature genetics, 2012. **44**(8): p. 890-894.

373. CARDIoGRAMplusC4D Consortium, *A comprehensive 1000 Genomes-based genome-wide association meta-analysis of coronary artery disease.* Nature Genetics, 2015.

374. Guth, B.D. and G. Rast, *Dealing with hERG liabilities early: diverse approaches to an important goal in drug development.* Br J Pharmacol, 2010. **159**(1): p. 22-4.

375. Preiss, D., et al., *Risk of incident diabetes with intensive-dose compared with moderate-dose statin therapy: a meta-analysis.* Jama, 2011. **305**(24): p. 2556-2564.

376. Medina, M.W., et al., *Alternative splicing of 3-hydroxy-3-methylglutaryl coenzyme A reductase is associated with plasma low-density lipoprotein cholesterol response to simvastatin.* Circulation, 2008. **118**(4): p. 355-362.

377. Cai, R., et al., *Statins worsen glycemic control of T2DM in target LDL-c level and LDL-c reduction dependent manners: a meta-analysis.* Expert Opin Pharmacother, 2016. **17**(14): p. 1839-49.

378. Tragante, V., et al., *Harnessing publicly available genetic data to prioritize lipid modifying therapeutic targets for prevention of coronary heart disease based on dysglycemic risk.* Human genetics, 2016: p. 1-15.

379. Staley, J.R., et al., *PhenoScanner: a database of human genotype-phenotype associations.* Bioinformatics, 2016.

380. Griffith, M., et al., *DGIdb: mining the druggable genome.* Nature methods, 2013.

381. Consortium, G.P., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.

382. Harrow, J., et al., *GENCODE: the reference human genome annotation for The ENCODE Project.* Genome research, 2012. **22**(9): p. 1760-1774.

383. Knox, C., et al., *DrugBank 3.0: a comprehensive resource for 'omics' research on drugs.* Nucleic acids research, 2011. **39**(suppl 1): p. D1035-D1041.

384. Altman, R.B., *PharmGKB: a logical home for knowledge relating genotype to drug response phenotype.* Nature genetics, 2007. **39**(4): p. 426.

385. Kuhn, M., et al., *A side effect resource to capture phenotypic effects of drugs.* Molecular systems biology, 2010. **6**(1).

386. Kuhn, M., et al., *STITCH 3: zooming in on protein–chemical interactions.* Nucleic acids research, 2012. **40**(D1): p. D876-D880.

387. Van Bever, E., et al., *Operational rules for the implementation of INN prescribing.* International journal of medical informatics, 2014. **83**(1): p. 47-56.

388. Lindquist, M., *VigiBase, the WHO global ICSR database system: basic facts.* Drug Information Journal, 2008. **42**(5): p. 409-419.

389. Corsello, S.M., et al., *The Drug Repurposing Hub: a next-generation drug library and information resource.* Nat Med, 2017. **23**(4): p. 405-408.

390. Volkamer, A., et al., *DoGSiteScorer: a web server for automatic binding site prediction, analysis and druggability assessment.* Bioinformatics, 2012. **28**(15): p. 2074-2075.

391. Berman, H., et al., *The worldwide Protein Data Bank (wwPDB): ensuring a single, uniform archive of PDB data.* Nucleic acids research, 2007. **35**(suppl 1): p. D301-D303.

392. Uhlen, M., et al., *Towards a knowledge-based human protein atlas.* Nature biotechnology, 2010. **28**(12): p. 1248-1250.

393. Uhlén, M., et al., *Tissue-based map of the human proteome.* Science, 2015. **347**(6220): p. 1260419.

394. Nelson, C.P., et al., *Association analyses based on false discovery rate implicate new loci for coronary artery disease.* Nat Genet, 2017. **49**(9): p. 1385-1391.

395. Agteresch, H.J., et al., *Adenosine triphosphate: established and potential clinical applications.* Drugs, 1999. **58**(2): p. 211-32.

396. Belhassen, B. and A. Pelleg, *Adenosine triphosphate and adenosine: perspectives in the acute management of paroxysmal supraventricular tachycardia.* Clin Cardiol, 1985. **8**(9): p. 460-4.

397. Jager, R., et al., *Oral adenosine-5'-triphosphate (ATP) administration increases blood flow following exercise in animals and humans.* J Int Soc Sports Nutr, 2014. **11**: p. 28.

398. Kessler, T., et al., *Functional Characterization of the GUCY1A3 Coronary Artery Disease Risk Locus.* Circulation, 2017: p. CIRCULATIONAHA. 116.024152.

399. Zabel, U., et al., *Human soluble guanylate cyclase: functional expression and revised isoenzyme family.* Biochem J, 1998. **335 (Pt 1)**: p. 51-7.

400. Herve, D., et al., *Loss of alpha1beta1 soluble guanylate cyclase, the major nitric oxide receptor, leads to moyamoya and achalasia.* Am J Hum Genet, 2014. **94**(3): p. 385-94.

401. Erdmann, J., et al., *Dysfunctional nitric oxide signalling increases risk of myocardial infarction.* Nature, 2013. **504**(7480): p. 432.

402. Wilkins, M.R., et al., *alpha1-A680T variant in GUCY1A3 as a candidate conferring protection from pulmonary hypertension among Kyrgyz highlanders.* Circ Cardiovasc Genet, 2014. **7**(6): p. 920-9.

403. Traynor, K., *Riociguat approved for pulmonary hypertension.* Am J Health Syst Pharm, 2013. **70**(22): p. 1960.

404. Bishop, B.M., *Riociguat for pulmonary arterial hypertension and chronic thromboembolic pulmonary hypertension.* Am J Health Syst Pharm, 2014. **71**(21): p. 1839-44.

405. Guha, M., *First-in-class guanylate cyclase stimulator approved for PAH.* Nat Biotechnol, 2013. **31**(12): p. 1064.

406. Rask-Andersen, M., S. Masuram, and H.B. Schioth, *The druggable genome: Evaluation of drug targets in clinical trials suggests major shifts in molecular class and indication.* Annu Rev Pharmacol Toxicol, 2014. **54**: p. 9-26.

407. Galie, N., et al., *PATENT PLUS: a blinded, randomised and extension study of riociguat plus sildenafil in pulmonary arterial hypertension.* Eur Respir J, 2015. **45**(5): p. 1314-22.

408. Haas, E. and H. Goldblatt, *Effects of an antihypertensive drug, pentolinium.* Am J Physiol, 1959. **196**(4): p. 763-8.

409. Freis, E.D. and I.M. Wilson, *Results of Prolonged Treatment with Pentolinium Tartrate.* Circulation, , 1956. **13**.

410. Cachelin, A.B. and G. Rust, *Beta-subunits co-determine the sensitivity of rat neuronal nicotinic receptors to antagonists.* Pflugers Arch, 1995. **429**(3): p. 449-51.

411. Sethi, O.P. and O.D. Gulati, *Analysis of mode of action of some nicotinic blocking drugs.* Jpn J Pharmacol, 1973. **23**(4): p. 437-51.

412. Nebion, A.G., *Top 10 tissues expressing genes*. 2016, Genevisible: Zurich, Switzerland.

413. Khakh, B.S., et al., *State-dependent cross-inhibition between transmitter-gated cation channels.* Nature, 2000. **406**(6794): p. 405-10.

414. Mahajan, A., et al., *Genome-wide trans-ancestry meta-analysis provides insight into the genetic architecture of type 2 diabetes susceptibility.* Nat Genet, 2014. **46**(3): p. 234-44.

415. Dichgans, M., et al., *Shared genetic susceptibility to ischemic stroke and coronary artery disease: a genome-wide analysis of common variants.* Stroke, 2014. **45**(1): p. 24-36.

416. Natalie, V.Z., *The genetic determinants of cardiovascular disease in patients with type 2 diabetes*. 2013, University of Dundee. p. 309.

417. Matic, L.P., et al., *Phenotypic Modulation of Smooth Muscle Cells in Atherosclerosis Is Associated With Downregulation of LMOD1, SYNPO2, PDLIM7, PLN, and SYNMHighlights.* Arteriosclerosis, thrombosis, and vascular biology, 2016. **36**(9): p. 1947-1961.

418. Miller, C.L., et al., *Integrative functional genomics identifies regulatory mechanisms at coronary artery disease loci.* Nature communications, 2016. **7**.

419. Howson, J.M., et al., *Fifteen new risk loci for coronary artery disease highlight arterial-wall-specific mechanisms.* Nature genetics, 2017. **49**(7): p. 1113-1119.

420. Metzler, M., et al., *Disruption of the endocytic protein HIP1 results in neurological deficits and decreased AMPA receptor trafficking.* Embo j, 2003. **22**(13): p. 3254-66.

421. Joehanes, R., et al., *Integrated genome-wide analysis of expression quantitative trait loci aids interpretation of genomic association studies.* Genome biology, 2017. **18**(1): p. 16.

422. Hendrix, P., et al., *Structure and expression of a 72-kDa regulatory subunit of protein phosphatase 2A. Evidence for different size forms produced by alternative splicing.* J Biol Chem, 1993. **268**(20): p. 15267-76.

423. Zhao, Y., et al., *Network-based identification and prioritization of key regulators of coronary artery disease loci.* Arteriosclerosis, thrombosis, and vascular biology, 2016: p. ATVBAHA. 115.306725.

424. Welter, D., et al., *The NHGRI GWAS Catalog, a curated resource of SNP-trait associations.* Nucleic Acids Res, 2014. **42**(Database issue): p. D1001-6.

425. Manolio, T.A., et al., *Finding the missing heritability of complex diseases.* Nature, 2009. **461**(7265): p. 747-53.

426. Wood, A.R., et al., *Defining the role of common variation in the genomic and biological architecture of adult human height.* Nat Genet, 2014. **46**(11): p. 1173-86.

427. Neale, B.M., et al., *Patterns and rates of exonic de novo mutations in autism spectrum disorders.* Nature, 2012. **485**(7397): p. 242-5.

428. Schizophrenia Working Group of the Psychiatric Genomics, C., *Biological insights from 108 schizophrenia-associated genetic loci.* Nature, 2014. **511**(7510): p. 421-7.

429. Nikpay, M., et al., *A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease.* Nat Genet, 2015. **47**(10): p. 1121-1130.

430. Locke, A.E., et al., *Genetic studies of body mass index yield new insights for obesity biology.* Nature, 2015. **518**(7538): p. 197-206.

431. Willer, C.J., et al., *Discovery and refinement of loci associated with lipid levels.* Nat Genet, 2013. **45**(11): p. 1274-1283.

432. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119-24.

433. Maher, B., *Personal genomes: The case of the missing heritability.* Nature, 2008. **456**(7218): p. 18-21.

434. Zuk, O., et al., *Searching for missing heritability: designing rare variant association studies.* Proc Natl Acad Sci U S A, 2014. **111**(4): p. E455-64.

435. Klarin, D., et al., *Genetic analysis in UK Biobank links insulin resistance and transendothelial migration pathways to coronary artery disease.* Nat Genet, 2017. **49**(9): p. 1392-1397.

436. Visscher, P.M., et al., *10 Years of GWAS Discovery: Biology, Function, and Translation.* Am J Hum Genet, 2017. **101**(1): p. 5-22.

437. de Lange, K.M., et al., *Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease.* Nat Genet, 2017. **49**(2): p. 256-261.

438. van Leeuwen, E.M., et al., *Meta-analysis of 49 549 individuals imputed with the 1000 Genomes Project reveals an exonic damaging variant in ANGPTL4 determining fasting TG levels.* J Med Genet, 2016. **53**(7): p. 441-9.

439. Collins, R., *What makes UK Biobank special?* Lancet, 2012. **379**(9822): p. 1173-4.

440. Khera, A.V., et al., *Genetic Risk, Adherence to a Healthy Lifestyle, and Coronary Disease.* N Engl J Med, 2016. **375**(24): p. 2349-2358.

441. Khera, A.V., et al., *Genome-wide polygenic score to identify a monogenic risk-equivalent for coronary disease.* bioRxiv, 2017.

442. Inouye, M., et al., *Genomic risk prediction of coronary artery disease in nearly 500,000 adults: implications for early screening and primary prevention.* bioRxiv, 2018.

443. Manolio, T.A., *Genomewide association studies and assessment of the risk of disease.* N Engl J Med, 2010. **363**(2): p. 166-76.

444. Jameson, J.L. and D.L. Longo, *Precision medicine--personalized, problematic, and promising.* N Engl J Med, 2015. **372**(23): p. 2229-34.

445. Pazoki, R., et al., *Genetic Predisposition to High Blood Pressure and Lifestyle Factors: Associations With Midlife Blood Pressure Levels and Cardiovascular Events.* Circulation, 2018. **137**(7): p. 653-661.

446. Humphries, S.E., *Common Variants for Cardiovascular Disease: Clinical Utility Confirmed.* Circulation, 2017. **135**(22): p. 2102-2105.

447. Pertea, M., et al., *Thousands of large-scale RNA sequencing experiments yield a comprehensive new human gene list and reveal extensive transcriptional noise.* bioRxiv, 2018.

448. Liu, L., W. Zhao, and X. Zhou, *Modeling co-occupancy of transcription factors using chromatin features.* Nucleic Acids Res, 2016. **44**(5): p. e49.

449. Rosa-Garrido, M., D.J. Chapski, and T.M. Vondriska, *Epigenomes in Cardiovascular Disease.* Circ Res, 2018. **122**(11): p. 1586-1607.

450. Dickel, D.E., et al., *Genome-wide compendium and functional assessment of in vivo heart enhancers.* Nat Commun, 2016. **7**: p. 12923.

451. Ounzain, S., et al., *Functional importance of cardiac enhancer-associated noncoding RNAs in heart development and disease.* J Mol Cell Cardiol, 2014. **76**: p. 55-70.

452. Smith, J.G., et al., *Discovery of Genetic Variation on Chromosome 5q22 Associated with Mortality in Heart Failure.* PLoS Genet, 2016. **12**(5): p. e1006034.

453. Meder, B., et al., *Epigenome-Wide Association Study Identifies Cardiac Gene Patterning and a Novel Class of Biomarkers for Heart Failure.* Circulation, 2017. **136**(16): p. 1528-1544.

454. Schleithoff, C., et al., *On the epigenetics of vascular regulation and disease.* Clin Epigenetics, 2012. **4**(1): p. 7.

455. Gal-Yam, E.N., et al., *Cancer epigenetics: modifications, screening, and therapy.* Annu Rev Med, 2008. **59**: p. 267-80.

456. Sharma, S., T.K. Kelly, and P.A. Jones, *Epigenetics in cancer.* Carcinogenesis, 2010. **31**(1): p. 27-36.

457. Heerboth, S., et al., *Use of epigenetic drugs in disease: an overview.* Genet Epigenet, 2014. **6**: p. 9-19.

458. Hatzimichael, E., et al., *Epigenetics in diagnosis, prognostic assessment and treatment of cancer: an update.* EXCLI J, 2014. **13**: p. 954-76.

459. Barrett, J.C., I. Dunham, and E. Birney, *Using human genetics to make new medicines.* Nature Reviews Genetics, 2015. **16**(10): p. 561-562.

460. Sander, J.D. and J.K. Joung, *CRISPR-Cas systems for editing, regulating and targeting genomes.* Nat Biotechnol, 2014. **32**(4): p. 347-55.

461. Liu, X.S., et al., *Editing DNA Methylation in the Mammalian Genome.* Cell, 2016. **167**(1): p. 233-247 e17.

462. Morgan, S.L., et al., *Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping.* Nat Commun, 2017. **8**: p. 15993.

Summary
(in English)

Samenvatting
(in het Nederlands)

Resumo
(em Português)

Summary for the general public
(General Summary in English)

Cardiovascular diseases (CVDs) are the leading cause of death globally. Risk factors were identified over the past decades, both modifiable and non-modifiable, which increase the risk of developing CVD. Modifiable risk factors include smoking, hypertension, obesity, excessive alcohol consumption and stress. Non-modifiable risk factors include age, sex, and genetics. Genomic research made remarkable progress in the last decades, identifying thousands of regions in the DNA that increase risk for CVDs and other diseases. However, the mechanisms through which these genomic regions contribute to disease are still relatively undetermined. Most of this genetic variation associated with disease identified so far is located in what are called noncoding parts of the genome. Noncoding DNA sequences are components of an organism's DNA that do not encode proteins. In recent years, parts of the noncoding genome was found to have regulatory function, affecting expression of the coding genome. A better understanding of this relationship is important to identify mechanisms of organism development and disease. Advances in technology made possible to identify the regulatory landscape of different tissues and cell-types, and this information was made publicly available by projects such as ENCODE and Roadmap Epigenomics. Integration of this functional information with genetic associations identified by genome-wide association studies (GWAS) allows better understanding of disease mechanisms and helps translating this knowledge into better disease prevention and treatment. In this thesis, I present a set of strategies for integrating several layers of information on the genome and gene regulation, aiming at translating GWAS discoveries into better disease understanding and better health care.

In **Chapter 2** we used publically available data and bioinformatics tools for interpretation of DNA variants associated to increased risk of high blood pressure. We identified an enrichment of these DNA variants in regulatory elements active in cardiovascular tissues and other tissues of organs affected by high blood pressure outcomes. We also assessed which genes potentially affected by these risk variants can be considered candidates for development of new drugs.

In **Chapter 3,** we integrated several layers of data to identify the possible function of 52 genetic regions influencing myocardial mass, measured by the QRS parameter of the electrocardiogram (ECG). We demonstrated that regulatory elements that show differential activity between patients with hypertrophic cardiomyopathy (HCM) and control donors show more enrichment for QRS-associated variants than regulatory elements that do not show differential activity in HCM. In addition, we pinpointed candidate genes and potential mechanisms by which QRS-associated variants account risk for increased left ventricular mass.

In **Chapter 4**, we highlighted the contribution of data integration for investigation of regulation and gene expression in dilated cardiomyopathy (DCM). We identified

regulatory elements and genes that show differential activity or expression in DCM. Some of these genes were previously identified in other studies, and our findings reinforce the evidence of their role in the process of DCM.

In **Chapter 5** we investigated whether 6 variants associated to carotid intima media thickness (cIMT, a biomarker for CVD) and/or plaque presence are also associated to other CVDs. We found that indeed one of these variants, near *EDNRA* gene, is also associated with coronary artery disease (CAD).

In **Chapter 6**, we investigated whether variants associated to high blood pressure that implicate specific gene regulation in a given tissue account more risk for an outcome that affects the tissue tested, compared to other tissues. In the nine hypertension-induced outcomes tested, three showed higher association of tissue-specific variants that implicate regulation in the tissue affected by the outcome. These are LV-specific regulatory variants in DCM, kidney-specific regulatory variants in kidney failure, and pancreas-specific variants in diabetes.

In **Chapter 7**, we described a pipeline for identification of new drug targets. We were able to identify three genes with high potential for drug development: *LMOD1*, *HIP1* and *PPP2R3A*. We also presented a pipeline for potential repurposing of currently marketed drugs, by integration of genetic variants that increase risk for coronary artery disease (CAD) and pharmacological data. This pipeline led to three suggestions of drugs that were originally developed to treat other diseases, but can be repurposed for treatment of CAD, saving both time and money in the drug development process.

Geavanceerde bioinformatica in het post-GWAS tijdperk van cardiovasculair genoom-onderzoek

(Nederlandse samenvatting)

Hart- en vaatziekten zijn de nummer één doodsoorzaak wereldwijd. De afgelopen tientallen jaren zijn veel risicofactoren gevonden, welke zijn in the delen in twee groepen: risicofactoren waar de patiënt iets aan kan doen, en risicofactoren die niet door de patiënt te veranderen zijn. Voorbeelden van de eerste groep zijn roken, overgewicht, overmatige alcoholconsumptie en stress. Voorbeelden van risicofactoren waar de patiënt geen invloed op heeft zijn leeftijd, geslacht en genetische factoren. Over de laatste jaren is enorme vooruitgang geboekt in het genetisch onderzoek, wat heeft geleidt tot identificatie van een groot aantal regio's in het DNA als risicofactor voor hart- en vaatziekten en voor vele andere ziektes en aandoeningen. Het is echter nog nauwelijks bekend waarom deze regio's in het DNA een invloed hebben op deze ziektes. Het grootste deel van deze regio's zijn zogenaamde "niet coderende" delen van het DNA; dit is DNA dat niet codeert voor een eiwit. Sinds enkele jaren is bekend dat sommige van deze niet coderende gebieden een regulerende rol hebben op genen, waarmee ze de hoeveelheid afgeschreven eiwit bepalen. Onderzoek naar de relatie tussen het niet coderende deel en coderende deel van het DNA is belangrijk om meer te leren over biologische mechanismes die voor ziektes zorgen. Recente technologische ontwikkelingen hebben het mogelijk gemaakt om regulerende gebieden in het DNA aan te wijzen in cellen in verschillende weefsels van het lichaam. Veel van deze informatie is vrijelijk beschikbaar, zoals de projecten ENCODE en Roadmap Epigenomics. Door deze informatie samen te voegen met informatie over specifieke variaties in het DNA uit zogenaamde genoomwijde associatie studies (GWAS), kunnen we beter begrijpen hoe ziektes ontstaan. Dit helpt ons vervolgens in het ontwikkelen van betere preventiestrategieën en behandelingen. In dit proefschrift presenteer ik een set strategieën om verschillende soorten informatie over variaties in het DNA en regulatie van genen te combineren. Hiermee wil ik de vertaalslag maken van genetische varianten die zijn gevonden in GWAS naar het beter begrijpen van mechanismes die hart- en vaatziekten veroorzaken.

In **Hoofdstuk 2** hebben we gebruik gemaakt van publiekelijk beschikbare gegevens en software voor het interpreteren van varianten in het DNA die van invloed zijn op bloeddruk. We laten zien dat deze varianten vaker dan verwacht een regulerende functie hebben op genen in weefsels die te maken hebben met het hart en vaatstelsel. We hebben vervolgens onderzocht welke genen mogelijk betrokken zijn. Deze kunnen mogelijk dienen als aanhechtingspunten voor nieuwe medicatie.

In **Hoofdstuk 3** hebben we verschillende soorten informatie gecombineerd om de functie van 52 regio's in het DNA te bepalen. Deze regio's zijn geselecteerd omdat ze invloed hebben op de dikte van de hartspier, gemeten als QRS-complex in een hartfilmpje. We laten zien dat deze regio's vaker dan verwacht in gebieden in het genoom liggen

die een verschil in genregulering tussen patiënten met hypertrofische cardiomyopathie (HCM) en gezonde controles. Daarnaast hebben we genen en biologische mechanismes gevonden die mogelijk een causale rol spelen in de dikte van de hartspier.
Ook in **Hoofdstuk** 4 hebben we verschillende lagen aan informatie gecombineerd om de invloed van genregulering op gedilateerde cardiomyopathie (DCM) te onderzoeken. We hebben regulerende gebieden en genen in het DNA gevonden die een verschil in activiteit lieten zien tussen DCM patiënten en gezonde controles. Een aantal van deze genen zijn eerder al in andere studies gevonden, wat hun invloed op het krijgen van DCM benadrukt.
In **Hoofdstuk 5** hebben we onderzocht of varianten in het DNA die van invloed zijn op de dikte van de wand van de halsslagader (cIMT) en op aderverkalking, ook een invloed hebben op andere vormen van hart- en vaatziekten. We laten zien de al bekende varianten voor cIMT en aderverkalking ook in ons cohort een relatie met deze aandoeningen laten zien. Daarnaast laat één van de varianten, dicht bij het gen *EDNRA* een correlatie zien met coronair vaatlijden.
In **Hoofdstuk 6** hebben we onderzoek gedaan naar varianten in het DNA die invloed hebben op bloeddruk. Een deel van deze varianten ligt in gebieden in het DNA die in bepaalde weefsels en organen een regulerend effect op genen hebben. Wij hebben onderzocht of varianten die genen reguleren in een bepaald weefsel, ook van invloed zijn op ziektes die te maken hebben met dat weefsel. We hebben negen ziektes getest die gerelateerd zijn aan een te hoge bloeddruk. Drie daarvan hadden een relatief hoog aantal varianten in het bijbehorende weefsel, in vergelijking tot andere weefsels en organen. Dit waren 1) regulerende varianten in het linkerkamer van het hart voor de ziekte DCM, 2) regulerende varianten in de nieren voor nierfalen en 3) genregulerende varianten in de alvleesklier voor diabetes (suikerziekte) en andere alvleesklier-gerelateerde ziektes. Dit hoofdstuk laat zien dat er een potentiele rol is van weefselspecifieke genregulerende varianten in het DNA die risico geven op bepaalde ziektes in mensen met een hoge bloeddruk.
In **Hoofdstuk 7** hebben we een methode beschreven waarmee we nieuwe aanhechtingspunten voor medicijnen kunnen bestuderen. Hiermee hebben we drie genen gevonden, *LMOD1*, *HIP1* en *PPP2R3A*, waarvoor medicatie het risico op bepaalde hart- en vaatziekten zou kunnen verminderen. Daarnaast beschrijven we een methode waarmee we onderzoeken welke bestaande medicatie potentieel voor andere ziektes gebruikt zou kunnen worden. Dit doen we door het combineren van farmacologische informatie (de exacte werking van alle medicijnen op de markt) en varianten in het DNA die een verhoogd risico geven op coronair vaatlijden. Op deze manier hebben we drie medicijnen gevonden die mogelijk ook gebruikt kunnen worden voor het behandelen van coronair vaatlijden.

Bioinformática integrativa em genômica cardiovascular pós-GWAS
(Resumo em Português)

Doenças cardiovasculares (DCVs) são a principal causa de morte no mundo. Fatores que aumentam o risco de desenvolvimento de DCVs foram identificados nas últimas décadas e classificados como modificáveis e não-modificáveis. Fatores de risco modificáveis incluem tabagismo, pressão alta, obesidade, consumo excessivo de álcool e estresse. Fatores de risco não-modificáveis incluem idade, sexo e genética. Houve muito avanço em pesquisa genética nas últimas décadas, o que inclui a identificação de milhares de regiões do DNA que representam maior risco para desenvolvimento de doenças, incluindo cardiovasculares. No entanto, os mecanismos através dos quais essas regiões contribuem para a doença ainda são relativamente desconhecidos. A maior parte dessa variação genética associada a doenças identificada até agora está localizada naquilo que é chamado de partes não codificantes do genoma. Sequências de DNA não-codificantes não codificam sequências de proteínas, mas nos últimos anos, foi descoberto que partes desse genoma não-codificante têm função reguladora, afetando a expressão do genoma codificante. Uma melhor compreensão dessa relação é importante para identificar mecanismos de desenvolvimento do organismo e de doenças. Avanços tecnológicos possibilitaram identificar o mapa regulatório de diferentes tecidos e células, e essa informação foi disponibilizada publicamente por projetos como ENCODE e o Roadmap Epigenomics. A integração desta informação funcional com associações genéticas identificadas por estudos de associação genômica ampla (Genome-Wide Association Studies, GWAS) permite uma melhor compreensão dos mecanismos da doença e ajuda a traduzir este conhecimento em melhor prevenção e tratamento. Nesta tese, eu apresento um conjunto de estratégias para integrar várias camadas de informação sobre o genoma e seus elementos regulatórios, visando traduzir as descobertas de GWAS em melhor compreensão de doenças e desenvolvimento de melhores intervenções terapêuticas.

No **Capítulo 2**, usamos dados publicamente disponíveis e ferramentas de bioinformática para interpretação de variantes de DNA associadas ao aumento do risco de pressão alta. Observamos que essas variantes tendem a se concentrar em regiões de elementos regulatórios que estão ativos em tecidos cardiovasculares e outros tecidos de órgãos afetados por pressão alta. Também investigamos quais genes potencialmente afetados por essas variantes de risco têm potencial para serem considerados candidatos ao desenvolvimento de novos medicamentos.

No **Capítulo 3**, integramos várias camadas de dados para identificar a possível função de 52 regiões genéticas que influenciam aumento de massa miocárdica, medida pelo parâmetro QRS do eletrocardiograma. Identificamos elementos regulatórios cuja atividade é diferente entre pacientes com cardiomiopatia hipertrófica (CMH) e controles. Demonstramos que variantes genéticas associadas ao QRS estão altamente concentradas

nessas regiões diferenciais, em comparação a todas as regiões regulatórias identificadas. Com esses resultados, avançamos em direção a um maior entendimento sobre como a variação genética associada a QRS perturba a atividade de elementos regulatórios e os genes que eles afetam, aumentando o risco de hipertrofia cardíaca.

No **Capítulo 4**, destacamos a contribuição de integração de dados para investigação de regulação e expressão gênica em cardiomiopatia dilatada (CMD). Identificamos elementos regulatórios e genes que mostram atividade ou expressão diferencial em CMD, comparado a tecido cardíaco sem a doença. Alguns desses genes foram previamente identificados em outros estudos, e nossos resultados reforçam a evidência de seu papel no processo de CMD.

No **Capítulo 5**, investigamos se 6 variantes genéticas associadas à espessura da camada íntima-média de artérias carótidas (EIMC) e presença de placa nas artérias também estão associadas a outras DCVs. Descobrimos que, de fato, uma dessas variantes também está associada à doença arterial coronariana, sugerindo que estas DCVs compartilham alguns dos mesmos mecanismos.

No **Capítulo 6**, investigamos se as variantes associadas a pressão alta que envolvem elementos regulatórios específicos em um determinado tecido representam um risco maior para uma doença que afeta o tecido em questão, em comparação com outros tecidos. Das nove doenças induzidas por pressão alta testadas, três apresentaram maior associação de variantes regulatórias específicas ao tecido afetado pela doença. Esse foi o caso no teste em que as variantes regulatórias com atividade específica no ventrículo esquerdo aumentam o risco de desenvolver CMD, variantes regulatórias específicas em rim representam mais risco de desenvolvimento de insuficiência renal e variantes específicas em pâncreas implicam mais risco de desenvolvimento de diabetes.

No **Capítulo 7**, descrevemos um método para identificação de genes que podem ser considerados para desenvolvimento de novos medicamentos para tratamento de doença arterial coronariana (DAC) e infarto do miocárdio. Identificamos três genes com alto potencial para o desenvolvimento de drogas. Também apresentamos uma série de métodos para o potencial redirecionamento de medicamentos atualmente comercializados, que foram desenvolvidos para o tratamento de outras doenças, mas que podem potencialmente serem usados para tratamento de DAC. Integramos variantes genéticas que representam maior risco para o desenvolvimento de DAC com dados farmacológicos públicos. Com este método, sugerimos três drogas que podem potencialmente ser reaproveitadas para o tratamento de DAC. Com isso, muito dinheiro e tempo investido em pesquisa e ensaios clínicos são economizados no desenvolvimento de melhores tratamentos para DAC e outras doenças.

Acknowledgements

Writing my thesis acknowledgements section was a bit like a dream to me, and I am so extremely happy to be actually doing it right now. It means my thesis is ready and my PhD mission is almost accomplished. This however would not have been possible without the contribution and support of a lot of people, to whom I am ever so grateful.

First of all, my profound gratitude to my supervisor, **Folkert Asselbergs** and my co-supervisors, **Vinicius Tragante** and **Jessica van Setten**.

Folkert, thank you for giving this Brazilian computer scientist the opportunity to work on cardiovascular (epi)genetics within you group, and connecting me to so many amazing people in the field around the globe. Your enthusiasm, sense of humor and care about your students, patients and projects play an essential role in keeping your group motivated and confident. I don't know a lot of supervisors who ask their students if they are happy almost on a weekly basis ☺ Thank you for being such an inspiring mentor and helping me so much. **Paul de Bakker**, thank you for sharing the mentorship task with Folkert during my first and most challenging PhD year.

Vinicius, thank you for your help and support from writing my project proposal to finding a place to live in The Netherlands, with everything in between and after that. It was also great to have someone to speak Portuguese (and laugh about jokes that only Brazilians understand), to help adjust to the Dutch culture and to the life and work in the (slighlty new to me) field of bioinformatics.

Jessica, just like with Vinicius, I could go on and on and write pages of everything you helped me with during these PhD years, both inside and outside the office (seriously, so many things), for which I'll be forever grateful. Since you became my co-supervisor you were there every single day for every single issue, sharing the joy, frustration, ups and downs, and the fun of conducting science. I wish you would have shared my daily supervision with Vinicius from the very start, because it has been absolutely fantastic. Thank you for everything.

Connie R. Bezzina, Tanja Zeller, Eva Van Rooij, Gerard Pasterkamp, Pieter Doevendans, Jan Veldink: thank you for taking the time to read and comment on this work as my reading committee.

Jonne, Ineke, Ingrid, Monique and **Marie-José**: thank you for helping me solve all bureaucratic issues and making my life much easier.

The Asselbergs and **Hester den Ruijter**'s groups, specially **Sander, Robin, Daniel, Ema, Magdalena, Michal, Jiayi, Mark, Arjan, Laurens, Gideon, Katrien, Mimout, Karim, Janine, Rob, Marijke, Alicia, Ayoub, Zhyiong, Aisha, Saskia**: thank you all for the fruitful discussions, suggestions and feedbacks.

Thank you **Jason H. Moore** and everyone from the Institute for Biomedical Informatics at the UPenn, for teaching me so much during my period there.

Thank you to the amazing people I could always run to when desperate for help on statistics / computational issues: **Sara**, thank you also for helping me in issues regarding expat life in NL and for sharing your passion for science and for the feminist cause. **Joep**, always ready and up for brainstorming and troubleshooting! Thanks for the amazing vegetarian sushi, too. **Ies**, thank you for all your help during my very first project. **Hinri**, thank you for your patience in helping me in my very start on analyzing OMICS data. **Roel**, thank you for all your help, either computational, domestic – helping carry my luggage around or fix things at home – and fun-related – sharing good food and helping me to learn how to play the guitar.

To the current / former members of the Medical Genetics Department, with whom I shared an office in the two first PhD years and a lot of fun moments during / after that: **Mircea**, it was always fun to have the most random conversations in coffee breaks / spontaneous alcoholic adventures outside the office, or just saving each other every time one of us lost the key from home; **Sakshi**, I'll never forget my trip to India thanks to you showing me around, among many other special memories from our friendship; **Laurent**, thank you for helping me with the first bioinformatic tools I had contact with; **Camilla**, thank you, bonita, for all the fun and help and balloon gifts and evenings of dancing; **Ingrid**, thank you for introducing me to all your amazing friends and taking me to parties as fun as I haven't seen before in Utrecht. Thank you **Flip, Ruben, José, Chris, Christina, Wout, Balder, Kirsten, Glenn, Maarten, Marijn, Alessio, Rozemarijn, Mirjam, Edith, Karen, Helen** and **Iris** for all the good moments in and outside the office, including coffee breaks, lunchs and borrels.

To my paranymphs: **Glen** and **Sander**. I did not know anyone when I moved to The Netherlands and was obviously very apprehensive with the new life in a country where I would have to build new friendships from scratch. Little I knew that my new colleagues would include you, my two new best friends. During these years, I could always count on you for every issue possible related to adaptation as an expat, scientific / computational questions, uncertainties about projets and the future, the frustrations and celebrations of the PhD life. Outside of the office, I think we've covered pretty much everything: roadtrips, festivals, concerts, dinners, cosplaying, awful movies nights, better movie nights, countless much longer-than-expected evenings, consequently countless hangovers, hiking, camping, to give a few examples. I'm indeed very lucky to have met you, and forever grateful for everything.

To the other members of "The Team", **Nayia**, **Sanne** and **Ellen**: together with **Glen** and **Sander** you made my 4 Dutch years the most fun possible and were everything I could ever have asked for in terms of friendship. I'm very lucky for having met you at

the office and hope our very special connection will remain after my PhD. Thank you for being always there no matter what, for helping me in every challenge an expat PhD student faces, and for always lifting me up. I intend to come back to NL in the future specially to have Ellen's speculaas ice-cream and hold Sanne's cute little baby. Also thank you Team for introducing me to **Marlous** and **Stef**, with whom we had memorable – and sometimes not very much memorable given the amount of whisky involved – super gezellig times together.

I was also very lucky to have made friends in The Netherlands outside of the office. Some of you contributed to this thesis by reviewing parts of it, by encouraging me, supporting me and lifting me up, understanding my absence due to deadlines, or in general just by making my life lighter and better. I believe all of these friendships made outside the office started due to our passion for good music and jazz dance, so thank you for all good long nights of conversation listening to vinyls (**Damion** and **Andrea**, I'm specially looking at you). In the same line, thank you all awesome **lindy hop and swing jazz lovers** from Utrecht, Amsterdam, Rotterdam, and everyone I met on dance classes, socials, tea gatherings, practices and parties. Great part of this thesis was written while I looked forward to meeting you in the evening and letting go of all the worries of science and academia for a moment on the dancefloor. A very special thanks to **Ilja**, my favorite dance partner in crime, one of my "bestest" Dutch friends ever who lent me his family for Xmas and gave me his favorite childhood teddy bear, to name just a couple of awesome things.

Thank you **Willemieke**, whose advices pretty much changed my life.

To the Brazilian friends who were always there for me even from the other side of the ocean: **Cristiano, Suelen, Renata, Verônica, Sabrina**, I love you no matter the distance. Thank you my beloved **Anna**, for your love and letters straight from Italy.

Todo o meu amor e agradecimento, acima de tudo, à minha família. Obrigada a todos os meus **tio(a)s**, **primo(a)s** e **vó** querida pelo suporte e carinho, pelo pensamento positivo e boas vibrações, e por sempre me receberem tão bom quando volta pra casa de férias com tanta saudade. Muito obrigada mano **Jardel** por sempre me apoiar e por manter as coisas em ordem em casa. E finalmente, mãe **Ana** e pai **Jairo**, muito obrigada por todo o esforço de vocês em criarem essa filha que sempre teve sonhos e objetivos tão ambiciosos pra nossa origem como agricultores numa cidade super pequena do sul do Brasil. Obrigada por me apoiarem e me ajudarem nesses sonhos, e serem portanto também responsáveis pela realização deles. Amo muito vocês.

List of Publications

Tragante V*, **Hemerich D***, Alshabeeb M, Brænne I, Lempiainen H, Patel R, Ruijter HM, Barnes MR, Moore JH, Schunkert H, Erdmann J, Asselbergs FW, Druggability of Coronary Artery Disease Risk Loci, accepted for publication on Circulation: Genomic and Precision Medicine, 2018, Aug 13. doi: 10.1161/CIRCGEN.117.001977.

Hemerich D*, van Setten, J**,** Tragante V, Asselbergs FW. Integrative bioinformatics approaches for identification of drug targets in hypertension. Frontiers in Cardiovascular Medicine, 2018, Apr 4;5:25. doi: 10.3389/fcvm.2018.00025.

van der Laan SW*, Harshfield EL*, **Hemerich D**, Stacey D, Wood AM, Asselbergs FW. From lipid locus to drug target through human genomics. Cardiovascular Research, 2018, Jul 15;114(9):1258-1270. doi: 10.1093/cvr/cvy120.

Nolte IM*, Munoz ML, Tragante V, Amare AT, Jansen R, Vaez A, von der Heyde B, Avery CL, Bis JC, Dierckx B, van Dongen J, Gogarten SM, Goyette P, Hernesniemi J, Huikari V, Hwang SJ, Jaju D, Kerr KF, Kluttig A, Krijthe BP, Kumar J, van der Laan SW, Lyytikäinen LP, Maihofer AX, Minassian A, van der Most PJ, Müller-Nurasyid M, Nivard M, Salvi E, Stewart JD, Thayer JF, Verweij N, Wong A, Zabaneh D, Zafarmand MH, Abdellaoui A, Albarwani S, Albert C, Alonso A, Ashar F, Auvinen J, Axelsson T, Baker DG, de Bakker PIW, Barcella M, Bayoumi R, Bieringa RJ, Boomsma D, Boucher G, Britton AR, Christophersen I, Dietrich A, Ehret GB, Ellinor PT, Eskola M, Felix JF, Floras JS, Franco OH, Friberg P, Gademan MGJ, Geyer MA, Giedraitis V, Hartman CA, **Hemerich D**, Hofman A, Hottenga JJ, Huikuri H, Hutri-Kähönen N, Jouven X, Junttila J, Juonala M, Kiviniemi AM, Kors JA, Kumari M, Kuznetsova T, Laurie CC, Lefrandt JD, Li Y, Li Y, Liao D, Limacher MC, Lin HJ, Lindgren CM, Lubitz SA, Mahajan A, McKnight B, Zu Schwabedissen HM, Milaneschi Y, Mononen N, Morris AP, Nalls MA, Navis G, Neijts M, Nikus K, North KE, O'Connor DT, Ormel J, Perz S, Peters A, Psaty BM, Raitakari OT, Risbrough VB, Sinner MF, Siscovick D, Smit JH, Smith NL, Soliman EZ, Sotoodehnia N, Staessen JA, Stein PK, Stilp AM, Stolarz-Skrzypek K, Strauch K, Sundström J, Swenne CA, Syvänen AC, Tardif JC, Taylor KD, Teumer A, Thornton TA, Tinker LE, Uitterlinden AG, van Setten J, Voss A, Waldenberger M, Wilhelmsen KC, Willemsen G, Wong Q, Zhang ZM, Zonderman AB, Cusi D, Evans MK, Greiser HK, van der Harst P, Hassan M, Ingelsson E, Järvelin MR, Kääb S, Kähönen M, Kivimaki M, Kooperberg C, Kuh D, Lehtimäki T, Lind L, Nievergelt CM, O'Donnell CJ, Oldehinkel AJ, Penninx B, Reiner AP, Riese H, van Roon AM, Rioux JD, Rotter JI, Sofer T, Stricker BH, Tiemeier H, Vrijkotte TGM, Asselbergs FW, Brundel BJJM, Heckbert SR, Whitsel EA, den Hoed M, Snieder H, de Geus EJC. Genetic loci associated with heart rate variability and their effects on cardiac disease risk. Nature Communications 2017 Jun 14;8:15805. doi: 10.1038/ncomms15805.

Hemerich D*, van der Laan SW*, Tragante V, Asselbergs FW. Impact of carotid atherosclerosis loci on cardiovascular events. Atherosclerosis, 2015. 243(2): 466-8. doi: 10.1016/j.atherosclerosis.2015.10.017.

Blomberg L*; **Hemerich D***; Ruiz DDA. Evaluating the performance of regression algorithms on datasets with missing data. International Journal of Business Intelligence and Data Mining, 2013. 8: p.105.

Manuscripts submitted

Ferraz MAMM*, Rho HS, **Hemerich D**, Henning H, van Tol L, Hoelker M, Besenfelder U, Mokry M, Vos P, Stout T, Le Gac S, Gadella B. An oviduct-on-a-chip provides an enhanced in vitro environment for zygote genome reprogramming. Nature Communications.

Hemerich D*, Pei J, Harakalova M , van Setten J, Boymans S, Boukens BJ, Efimov IR, Michels M, van der Velden J, Vink A, Cheng C, van der Harst P, Moore JH, Mokry M, Tragante V, Asselbergs FW. Integrative functional annotation of 52 genetic loci influencing myocardial mass identifies candidate causal regulatory variants and target genes. JACC: Basic to Translational Science.

van Setten J*, Verweij N, Mbarek H, Niemeijer MN, Trompet S, Arking DE, Brody JA, Gandin I, Grarup N, Hall LM, **Hemerich D**, Lyytikäinen L, Mei H, Müller-Nurasyid M, Prins BP, Robino A, Smith AV, Warren HR, Asselbergs FW, Boomsma DI, Caulfield MJ, Eijgelsheim M, Ford I, Hansen T, Harris TB, Heckbert SR, Hottenga J, Iorio A, Kors JA, Linneberg A, MacFarlane P, Meitinger T, Nelson CP, Raitakari OT, Aldana CTS, Sinagra G, Sinner M, Soliman EZ, Stoll M, Uitterlinden A, van Duijn CM, Waldenberger M, Alonso A, Gasparini P, Gudnason V, Jamshidi Y, Kääb S, Kanters JK, Lehtimäki T, Munroe PB, Peters A, Samani NJ, Sotoodehnia N, Ulivi S, Wilson JG, de Geus EJC, Jukema JW, Stricker B, van der Harst P, de Bakker PIW, Isaacs A. GWAS meta-analysis of 30,000 samples identifies seven novel loci for quantitative ECG traits. European Journal of Human Genetics.

Articles in preparation

Hemerich D*, Pei J, Harakalova M , van Setten J, Boukens BJ, Efimov IR, Vink A, Cheng C, Mokry M, Tragante V, Asselbergs FW. Regulatory and transcriptional profile in dilated cardiomyopathy. In preparation.

Hemerich D*, Pei J, Harakalova M , van Setten J, Boymans S, Boukens BJ, Efimov IR, Michels M, van der Velden J, Vink A, Cheng C, Moore JH, Mokry M, Tragante V, Asselbergs FW. A bioinformatics method to improve detection of genetic associations considering prior biological knowledge. In preparation.

Hemerich D*, van Setten J, van der Laan SW, Kofink D, Munroe PB, Tragante V, Asselbergs FW. Use of tissue-specific genetic risk scores on the investigation of disease progression profiles of patients with oucomes induced by hypertension. In preparation.

Curriculum Vitae

Daiane Hemerich was born on 13 March 1989 in Marau, Brazil. After completing secondary education in Marau, she moved to Passo Fundo and completed a bachelor in Computer Science at the University of Passo Fundo. She then moved to Porto Alegre and started her master in Computer Science at the Pontifical Catholic University of Porto Alegre. During this period, she participated in projects on Data Mining and Machine Learning, with a period of collaboration at the University of Trento in Italy. After having received her masters, she moved to The Netherlands to perform PhD studies at the University of Utrecht, whose results are described in this thesis. In the future, Daiane wants to apply her knowledge of bioinformatics, genetics and epigenetics to improve disease understanding, prevention and treatment in her post-doctoral research.