



# Integrative functional analysis of super enhancer SNPs for coronary artery disease

Juexiao Gong<sup>1,2</sup> · Chuan Qiu<sup>3</sup> · Dan Huang<sup>1,2</sup> · Yiyan Zhang<sup>1,2</sup> · Shengyong Yu<sup>1,2</sup> · Chunping Zeng<sup>4</sup>

Received: 7 July 2017 / Revised: 13 January 2018 / Accepted: 29 January 2018 / Published online: 28 February 2018  
© The Author(s) under exclusive licence to The Japan Society of Human Genetics 2018

## Abstract

Clinical research in coronary artery disease (CAD) primarily focused on genetic variants located in protein-coding regions. Recently, mutations fall within non-coding regions have been suggested to be essential to the pathogenesis of human complex disease. Super enhancer is a densely spaced cluster of transcriptional enhancers located in non-coding regions, which is critical for regulating cell-type specific gene expression. However, the underlying mechanism of the super enhancer single-nucleotide polymorphisms (SNPs) affecting the risk of CAD remains unclear. By integrating genome-wide association study (GWAS) meta-analysis of CAD and cell/tissue-specific histone modification data set, we identified 366 potential CAD-associated super enhancer SNPs in 67 loci, including 94 SNPs that are involved in regulating chromatin interactive and/or affecting the transcription factors binding affinity. Interestingly, we found 7 novel functional loci (*CBFA2T3*, *ZMIZ1*, *DIP2B*, *SCNN1D/ACAP3*, *TMEM105*, *CAMK2G*, and *MAPK1*) that CAD-associated super enhancer SNPs were clustered into the same or neighboring super enhancers. Pathway analysis showed a significant enrichment in several well-known signaling and regulatory processes, e.g., cAMP signaling pathway and ErbB signaling pathway, which play a key role in CAD metabolism. Our results highlight the potential functional importance of CAD-associated super enhancer SNPs and provide the targets for further insights on the pathogenesis of CAD.

---

These authors contributed equally: Juexiao Gong, Chuan Qiu.

**Electronic supplementary material** The online version of this article (<https://doi.org/10.1038/s10038-018-0422-2>) contains supplementary material, which is available to authorized users.

✉ Shengyong Yu  
yshy66@163.com

✉ Chunping Zeng  
zcp193@163.com

<sup>1</sup> Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing, China

<sup>2</sup> Jiangsu Province Academy of Traditional Chinese Medicine, Nanjing, China

<sup>3</sup> Center for Bioinformatics and Genomics, Department of Global Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA

<sup>4</sup> Department of Endocrinology, The Fifth Affiliated Hospital of Guangzhou Medical University, Guangzhou, China

## Introduction

Coronary artery disease (CAD), also named ischemic heart disease, clinically manifest as a group of disorder including stable or unstable angina pectoris, acute myocardial infarction, and sudden cardiac death [1]. Over the past two decades, the increasing prevalence of CAD has become a serious public health burden which affect over 110 million individual and caused ~9 million deaths worldwide in 2015 [2], and it is estimated that CAD will be account for over 11.1 million deaths globally in 2020 [3]. CAD has become one of the leading causes of death, disability, and rapidly rising costs in medical care. There exists some compelling and consistent evidence showing that the risk of CAD is strongly influenced by a wide range of genetic factors [4]. Thus far, over 50 robust CAD-associated single nucleotide polymorphisms (SNPs) have been successfully inferred by genome-wide association studies (GWASs), but in aggregate, these SNPs together explain around 10% of the CAD heritability [5].

It has been found that SNPs in the protein-coding region are associated with hundreds of human complex disease by altering gene expression [6]. However, over

90% of common SNPs lie within non-coding regions [7] and the mechanism of these SNPs underlie disease susceptibility remains unclear. Super enhancer, overlapped with noncoding regions (e.g., DNA methylation valleys, etc.), is a genomic region comprising multiple enhancers that is collectively bound by transcription factor proteins to drive gene transcription [8]. Super enhancer involves in cell identity and plays a critical role in cell-type specific gene expression [9]. Recently Denes et al. point out that GWAS-identified SNPs are specifically enriched in the super enhancer regions of disease-relevant cell or tissue types [10]. For example, the super enhancer identified in T and B cells are enriched for rheumatoid arthritis-associated SNPs, and the SNPs that are significantly associated with the risk for Alzheimer's disease are enriched in the super enhancers of human brain [10]. However, the precise mechanism of cell/tissue-type specific super enhancer SNPs affecting risk of CAD remains unclear.

In this study, we aimed to identify CAD-associated super enhancer SNPs by integrating CAD cell/tissue-specific histone modification ChIP-seq data with CAD GWAS meta-analysis summary statistics, followed by comprehensive bioinformatics analyses to further illustrate the functional importance of super enhancer SNPs. Our results shed light on super enhancer SNPs may a novel paradigm for the pathogenesis of CAD.

## Materials and methods

### CAD GWAS Data sets

We retrieved 2236 CAD-associated SNPs at GWAS significance threshold ( $p$ -value  $< 5.0 \times 10^{-8}$ ) with a false discovery rate (FDR)  $q$ -value  $< 2.1 \times 10^{-4}$  from Coronary Artery Disease Genome-wide Replication and Meta-analysis (CARDIoGRAM) plus The Coronary Artery Disease (C4D) Genetics Consortium (<http://www.cardiogramplusc4d.org/data-downloads/>). To the best of our knowledge, this study is the largest and the standard GWAS meta-analysis of CAD which incorporated a total of 184,305 subjects including 60,801 CAD cases from 48 studies. This study interrogated 6.7 million common variants and 2.7 million rare variants. Next, we applied proxy SNP [11] to identify SNPs in linkage disequilibrium (LD) with retrieved CAD-associated SNPs. The searching process depends on genotyping results from the 1000 Genomes CEU (Utah residents with European ancestry) population panel. The inclusion criteria for LD SNPs were set as 500 kb region around the query SNP with  $r^2 > 0.8$ . A total of 6568 candidates of CAD-associated SNPs were identified.

### Histone modification data sets

We selected CAD-relevant cell/tissues based on their crucial role in pathogenesis of CAD through comprehensive, systematic literature searches in PubMed. In total, we identified 7 CAD-relevant cell/tissue types, including cardiac muscle, vascular endothelial cells and smooth muscle cells [12–15], adipose nuclei/tissue [16, 17], brain [18, 19], CD14 [20–22], skeletal muscle [23–25], and spleen [26]. We downloaded the histone modification ChIP-seq data sets (H3K27ac) with WIG/BIGWIG format in 7 CAD-relevant cell/tissue types from Gene Expression Omnibus (GEO) database. A detailed list of CAD-relevant cell/tissue types can be found in Supplemental Table 1.

### Identification of CAD-relevant super enhancer & CAD-associated super enhancer SNPs

We used Bowtie v1.1.2 (<http://bowtie-bio.sourceforge.net/index.shtml>) to align ChIP-seq WIG/BIGWIG files to the human reference genome (hg19). A significant threshold of  $p$ -value  $< 1.0 \times 10^{-9}$  was used to distinguish enhancer regions by MACS v1.4.1 [27]. Super-enhancers were separated from enhancers using the ROSE algorithm [8]. A detail description of the super enhancer identification process was illustrated in the previously published pipeline [10]. A total of 11,925 super enhancers were identified in 7 CAD-relevant cell/tissues. We mapped 6568 candidates of CAD-associated SNPs to these super enhancers. Finally, we achieved 366 CAD-associated super enhancer SNPs (Supplemental Table 2).

### Functional annotation of CAD-associated super enhancer SNPs

Genomic mapping and annotation of CAD-associated super enhancer SNPs was carried out through SNPnexus [28] which allows submitting either known or novel variants for analysis in a single query according to genomic region of interest. We characterized the regulatory features of CAD-associated super enhancer SNPs by rVarBase v2.0 [29]. rVarBase utilizes experimentally supported regulatory elements from ENCODE and other data resources to annotate regulatory feature of variants from chromatin state of region surrounding the queried SNPs, regulatory elements overlapping with the queried SNPs, and functional candidate genes of queried SNPs. To further validate the potential functional consequence of CAD-associated super enhancer SNPs, we applied HaploReg v4.1 [30], a comprehensive resource which utilizes data from Roadmap Epigenomics, ENCODE and other data resources, to explore the effect of CAD-associated super enhancer SNPs on regulatory motif alterations within sets of genetically linked CAD-associated

super enhancer SNPs. In addition, the effects of CAD-associated super enhancer SNPs on the expression of target genes (eQTLs) in different cell/tissue types were investigated by using the data from the Genotype-Tissue Expression (GTEx) v7.0. Furthermore, transcription factor enrichment analysis was performed through SNP2TFBS v1.0 [31] tool which selects and visualizes user defined variants that affect single or multiple transcription factors.

### Long-range interaction analysis of CAD-associated super enhancer SNPs

GWAS3D was used to systematically estimate the probability of genomic variants which affect disease-associated pathways by assembling sequence motif, chromatin status, functional genomics, and conservation knowledge [32]. And also, GWAS3D provides a detailed annotation and expanded visualization to illustrate the long-range interaction between genetic variants. In the current study, GWAS3D v1.0 was applied to identify the most significant CAD-associated super enhancer SNPs which have a long-range interaction signal with their distal regulatory elements by applying the threshold value of Fisher's combined  $p$ -value  $< 1.0 \times 10^{-5}$ ,  $r^2 = 0.8$ , HapMap CEU population panel (phase I + II + III), size of 30 variants, three interaction size, and binding site  $p$ -value = 0.01.

### Pathway enrichment analysis

To identify biological enrichment pathways that are moderately associated with CAD, we applied an adapted Gene Set Enrichment Analysis (GSEA) framework to CAD-associated super enhancer SNPs through WebGestalt [33]. WebGestalt is a web-based comprehensive functional annotation tools which has been used for gene functional enrichment and classification analysis by several novel clustering algorithms. The GSEA algorithm in WebGestalt tested for over-representation of genes in a given gene set above a predetermined gene score rank cutoff. Here gene set was defined using the molecular signature database, the Kyoto Encyclopedia of Genes and Genomes (KEGG) [34]. To minimize biases caused by LD on pathway analysis, the CAD-associated super enhancer SNPs were pruned by plink (<https://www.cog-genomics.org/plink2>) using the “indep-pairwise” option with the parameters for window size, step and  $r^2$  set to 5, 1, and 0.3, respectively. 59 SNPs remained after pruning. We defined a nominal uncorrected significance level of  $p$ -value  $< 0.05$ . To correct for multiple hypothesis testing, the FDR was used for KEGG database (186 pathways), and the significant threshold value was set as FDR  $< 0.05$ .

### Protein–protein interaction (PPI) network

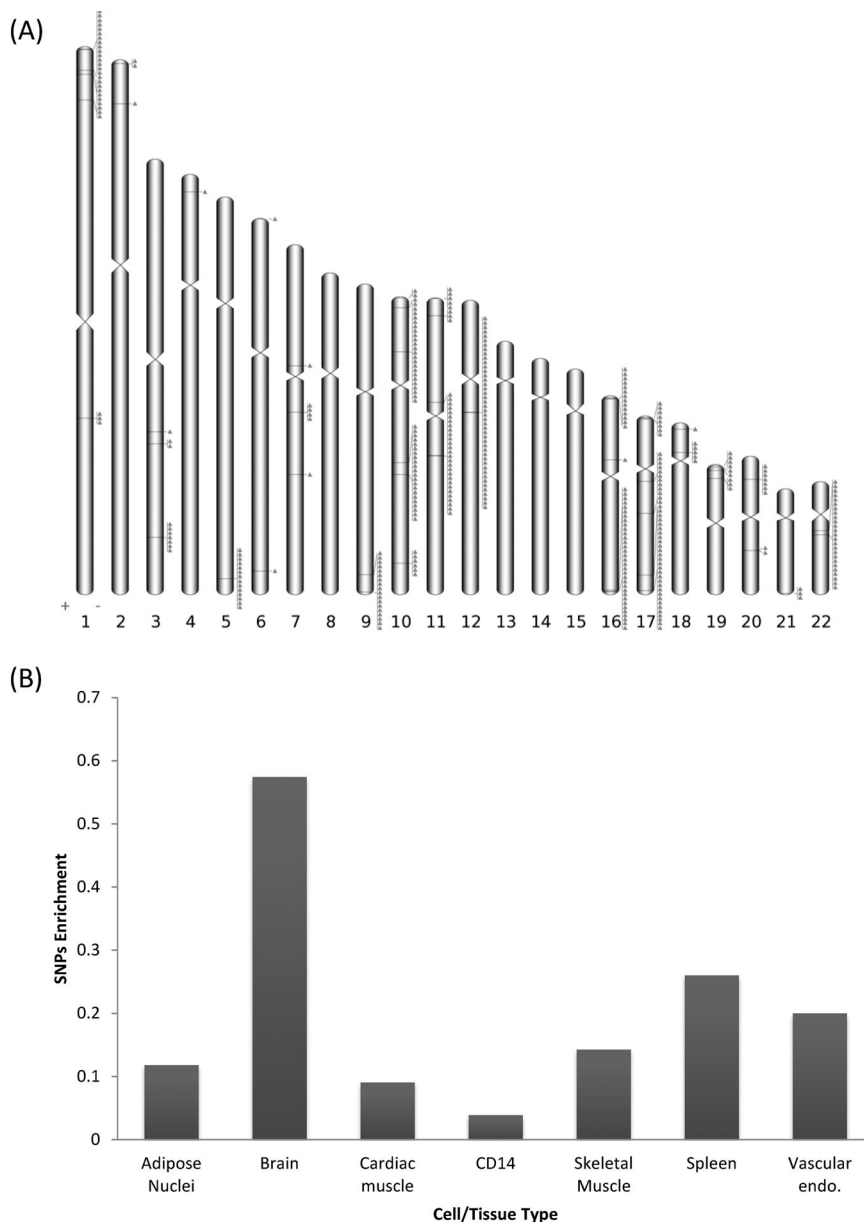
STRING [35] is used to construct the interaction network of the known or the predicted proteins. It quantify the genetic interaction based on physical and function association which was derived from multiple data sources, e.g., high-throughput experiments and previous knowledge. At present, the database contains 9.6 million proteins from 2031 organism constructed  $>184$  million interactions. In order to explore the functional partnership and interaction of the identified CAD-associated super enhancer SNPs target genes, we construct gene (protein) association networks via the STRING v10.0 database with default settings (observed interaction, 70; expected interaction, 8.83; Benjamini adjust  $p$ -value  $< 1 \times 10^{-10}$ ; proteins, 76) [36].

## Results

### Summary statistics for CAD-associated super enhancer SNPs

We first retrieved 2236 potential CAD-associated SNPs at genome-wide significant threshold ( $p$ -value  $< 5.0 \times 10^{-8}$ ) with FDR  $q$ -value  $< 2.1 \times 10^{-4}$  from the largest CAD GWAS meta-analysis performed by CARDIoGRAMplusC4D Genetics Consortium. By assembling LD information from the 1000 Genomes Project based on CEU population panel, we inferred 6568 potential CAD-associated SNPs. Then we took advantage of publicly available human histone modification ChIP-seq data sets from GEO database to build an extensive list of super enhancer among CAD-relevant cell/tissue types. In total, we identified and confirmed 11,925 super enhancer regions in 7 human cell/tissue types (e.g., cardiac muscle, vascular endothelial cells and smooth muscle cells, adipose nuclei/tissue, brain, CD14, skeletal muscle and spleen etc.). Then 6568 candidates of CAD-associated SNPs were mapped to the identified CAD-relevant cell/tissue-specific super enhancers. Finally, we successfully obtained a total of 366 potential CAD-associated super enhancer SNPs in 67 loci, including 127 SNPs identified by original GWAS meta-analysis (Supplemental Table 2). We observed that the proportion of 366 functional CAD-associated super enhancer SNPs within 6568 SNPs is significantly larger than that of genome average (OR = 4.76). Interestingly, we identified 9 potential functional loci 1p36.33 (*ACAP3/PUSL1*), 9q34.3 (*TPRN/SSNA1*), 10p12.1, 10q22.3 (*ZMIZ1*), 11q13.4 (*IL18BP/NUMA1*), 12q13.12 (*DIP2B*), 16p13.3, 16q24.3 (*CBFA2T3*), 17q25.3 (*TMEM105*) that CAD-associated super enhancer SNPs were clustered into the same or neighboring super enhancers (Fig. 1a).

**Fig. 1** Genome-wide distribution of CAD-associated super enhancer SNPs. **a** CAD-associated super enhancer SNPs were annotated in the latest build of the human genome (hg19). The enrichment of SNPs was observed on chromosome 1p36.33, 9q34.3, 10p12.1, 10q22.3, 11q13.4, 12q13.12, 16p13.3, 16q24.3, 17q25.3, etc. **b** The relative enrichment of CAD-associated super enhancer SNPs in seven different cell/tissue types. The SNP enrichment was calculated as a ratio of number of CAD-associated super enhancer in each different cell/tissue to 366 CAD-associated super enhancer SNPs (Color figure online)



To identify whether CAD-associated super enhancer SNPs enriched in specific cell/tissue types, we characterized the distribution of CAD-associated super enhancer SNPs in 7 different cell/tissue types and identified a significant SNP enrichment in brain (210/366), spleen (95/366) and vascular endothelial cells and smooth muscle cells (73/366) (Fig. 1b and Supplemental Table 2). Notably, 58 SNPs were mapped to super enhancer in a wide variety of cell/tissue types (Supplemental Table 2). In addition, we checked the LD status among each of the SNPs pairs that mapped to the same super enhancers. To our surprise, we identified 854 SNP pairs with low LD ( $r^2 < 0.2$ ) which were mapped to the

same super enhancers. These SNP pairs account for 80 CAD-associated super enhancer SNPs in 15 super enhancers from 7 CAD-relevant cell/tissue (Supplemental Table 3). For example, the SNP rs586965 in gene *SCNN1D* has a low LD ( $r^2 = 0.009$ ) with rs3737719 in gene *ACAP3*, this SNP pair was mapped to the same super enhancer (1p36.33) from human brain. Interestingly, previous genetic studies have linked *SCNN1D* deficiency with rare genetic diseases with developmental and functional disorders in the brain, heart, and respiratory systems [37]. The full list of SNPs that map to a specific super enhancer and LD ( $r^2 < 0.2$ ) between each SNP is included in Supplemental Table 3.

**Table 1** CAD-associated super enhancer SNPs in *CBFA2T3* affect the gene expression itself in artery

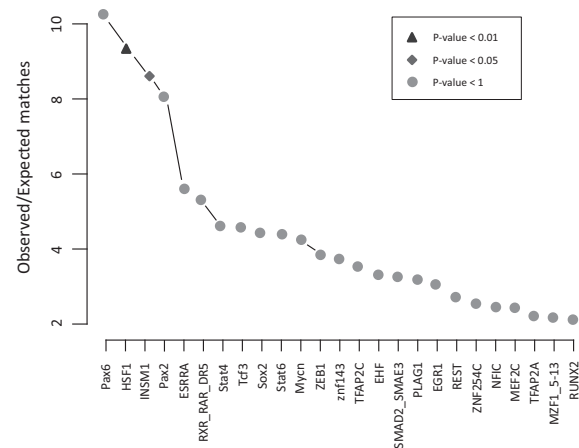
SNP	Chr	Position	Gene	P-Value	Effect Size	Tissue
rs502258	chr16	88968547	CBFA2T3	3.30E-07	-0.24	Artery-Aorta
rs844396	chr16	88966667	CBFA2T3	3.50E-07	-0.24	Artery-Aorta
rs2968474	chr16	88970566	CBFA2T3	4.20E-07	-0.24	Artery-Aorta
rs562812	chr16	88970944	CBFA2T3	4.20E-07	-0.24	Artery-Aorta
rs488251	chr16	88972554	CBFA2T3	5.70E-07	-0.24	Artery-Aorta
rs871756	chr16	88969076	CBFA2T3	1.20E-06	-0.23	Artery-Aorta
rs579740	chr16	88973937	CBFA2T3	3.20E-06	-0.21	Artery-Aorta
rs475796	chr16	88970776	CBFA2T3	5.70E-06	-0.21	Artery-Aorta
rs1559446	chr16	88975332	CBFA2T3	6.40E-06	-0.2	Artery-Aorta
rs2968475	chr16	88976663	CBFA2T3	6.50E-06	-0.2	Artery-Aorta
rs533581	chr16	88967437	CBFA2T3	7.80E-06	-0.2	Artery-Aorta
rs561795	chr16	88970818	CBFA2T3	7.80E-06	-0.21	Artery-Aorta
rs865102	chr16	88969969	CBFA2T3	1.30E-05	-0.2	Artery-Aorta
rs555191	chr16	88978850	CBFA2T3	1.70E-05	-0.2	Artery-Aorta
rs548591	chr16	88974516	CBFA2T3	2.80E-05	-0.18	Artery-Aorta

### Functional annotation of CAD-associated super enhancer SNPs

To investigate the potential functional impact of the identified CAD-associated super enhancer SNPs, we annotated 366 CAD-associated super enhancer SNPs to various regulatory elements through rVarBase v2.0 [29]. rVarBase is a database which utilizes experimentally supported regulatory elements from multiple data resources to make relevant functional prediction, it provides reliable regulatory feature for human variants. We identified that 94 SNPs in 39 loci (e.g., *CBFA2T3*, *ZMIZ1*, *DIP2B* etc.) showed evidence in gene regulatory process, including 80 SNPs involved in chromatin interactive regulations, 6 SNPs have effect on CpG islands and 2 SNPs involved in regulation of lncRNA expression (Supplemental Table 4).

To further validate the potential functional consequence of the CAD-associated super enhancer SNPs, we test the effect of CAD-associated super enhancer SNPs on regulatory motifs by using data from ENCODE and Roadmap Epigenomics projects through HaploReg v4.1. We identified that 315 CAD-associated super enhancer SNPs change transcription factor binding motif (Supplemental Table 5). Interestingly, among the 94 potential functional SNPs identified by rVarBase, 81 SNPs including 6 SNPs have effect on CpG islands and 2 SNPs involved in regulation of lncRNA expression were replicated by HaploReg analysis. These results highlight the strong and reliable regulatory potential of CAD-associated super enhancer SNPs.

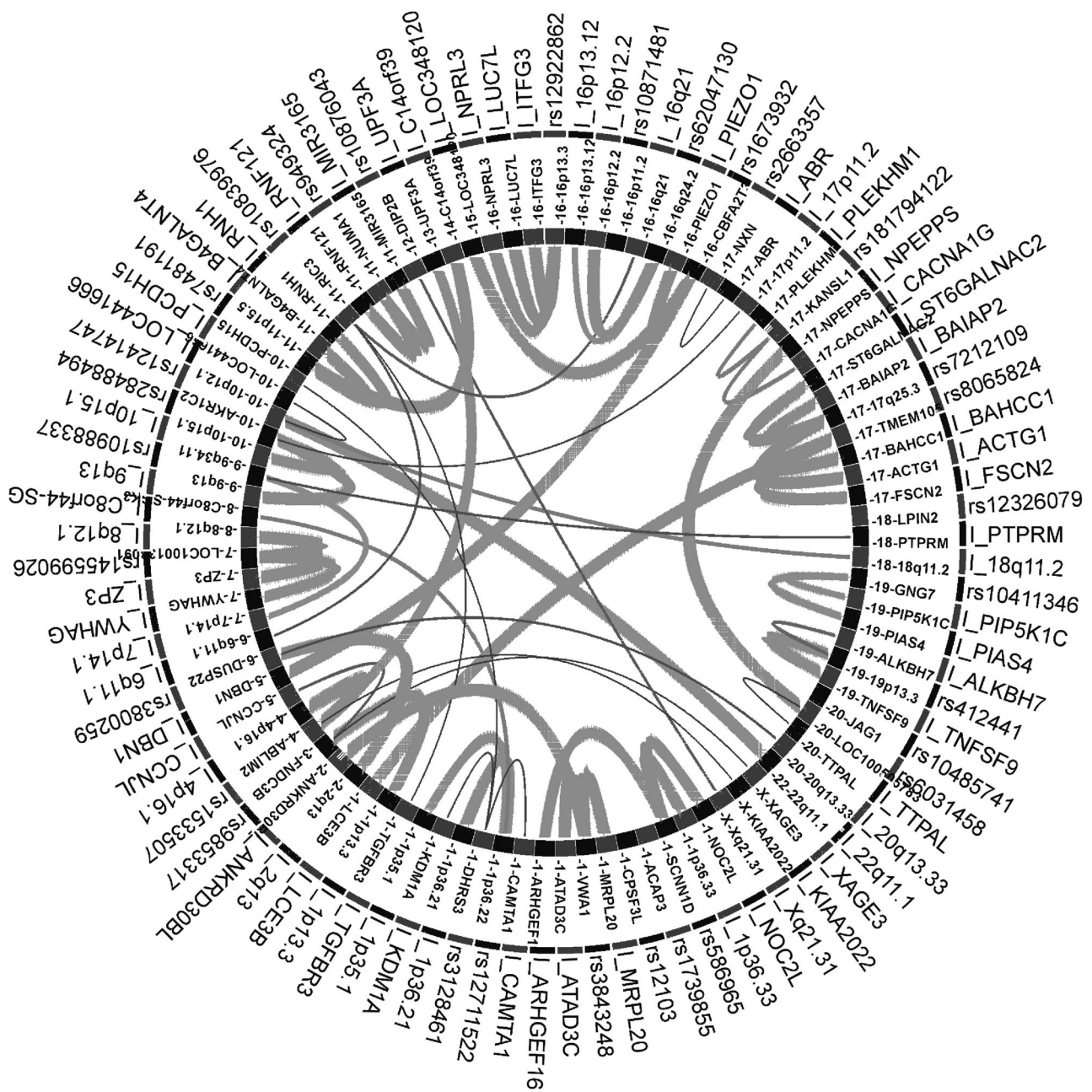
In addition, we investigated the effect of CAD-associated super enhancer SNPs on gene expression by using the data from GTEx v7.0. Notably, 205 CAD-associated super



**Fig. 2** Magnified transcription factors enrichment plot for CAD-associated super enhancer SNPs. Transcription factors are sorted based on their enrichment (Color figure online)

enhancer SNPs in 38 regions were reported eQTL evidences, including 181 SNPs have numerous reported eQTL evidences in a wide variety of tissues and cell types (Supplemental Table 6). This result showed that 56.7% (38/67) of the identified CAD-associated loci linked with eQTLs. We identified that 15 CAD-associated super enhancer SNPs in *CBFA2T3* affect the gene expression itself in artery with  $p$ -value  $< 2.8 \times 10^{-5}$  and effect size  $-0.24$  to  $-0.18$  (Table 1). The full list of eQTLs of CAD-associated super enhancer SNPs is included in Supplemental Table 6.

The effect of CAD-associated super enhancer SNPs on the transcription factor binding affinity was further analyzed by SNP2TFBS. We identified 2 significant enriched transcription factors, HSF1 and INSM1 (Fig. 2). Transcription factors HSF1 plays a critical role in normal lifespan



**Fig. 3** The circle plot of GWAS3D for CAD-associated super enhancer SNPs. Analysis was based on all cell line and CEU population. The red line indicated long-range interaction signals, and the

intensity of interaction was represented by the width of the line. Interactive elements with significant SNP will start with 'L' (Color figure online)

regulation. It induces rapidly after temperature stress and binds to heat shock promoter elements to activate transcription process. Recently, several studies suggested that HSF1 may regulate some of the differences during the development of physiologic and pathologic cardiac hypertrophy [38]. Moreover, in the context of cardiac ischemia/reperfusion, HSF1 expression can up-regulate heat shock protein expression to protect against subsequent ischemia/reperfusion injury [39]. Transcription factors INSM1

plays an important role in embryonic neurogenesis and differentiation of neuroendocrine cell in fetal development [40]. Although there is no direct evidence of association between CAD with transcription factors INSM1, a recent study showed that INSM1 has an effect on regulation of beta-cell development [41]. Interestingly, reduced beta-cell function can result in CAD with normal glucose tolerance [42].

**Table 2** The significant pathways for CAD super enhancer SNPs target genes

Term ID	Term Description	P-Value	FDR
hsa05214	Glioma - Homo sapiens (human)	3.13E-04	4.41E-02
hsa04024	cAMP signaling pathway - Homo sapiens (human)	6.28E-04	4.41E-02
hsa04012	ErbB signaling pathway - Homo sapiens (human)	7.29E-04	4.41E-02
hsa04540	Gap junction - Homo sapiens (human)	7.29E-04	4.41E-02
hsa04912	GnRH signaling pathway - Homo sapiens (human)	8.30E-04	4.41E-02
hsa04713	Circadian entrainment - Homo sapiens (human)	9.40E-04	4.41E-02
hsa01522	Endocrine resistance - Homo sapiens (human)	9.98E-04	4.41E-02
hsa04066	HIF-1 signaling pathway - Homo sapiens (human)	1.15E-03	4.45E-02
hsa04725	Cholinergic synapse - Homo sapiens (human)	1.43E-03	4.62E-02
hsa04320	Dorso-ventral axis formation - Homo sapiens (human)	1.49E-03	4.62E-02
hsa04144	Endocytosis - Homo sapiens (human)	1.68E-03	4.71E-02

### Long-range interaction of CAD-associated super enhancer SNPs

GWAS3D was used to visualize the CAD-associated super enhancer SNPs that have a long-range genomic interaction with other loci. Finally, 86 SNPs in 27 loci were detected as significant long-range interaction SNPs based on the HapMap CEU population (Fig. 3, Supplemental Table 7). The interesting SNP rs8065824 (Fisher's combined  $p$ -value =  $2.72 \times 10^{-14}$ ) located within the intron of *TMEM105* on chromosome 17 has a long-range genetic interaction with two loci *FSCN2* and *LCE3B*. Another interesting SNP rs12922862 (Fisher's combined  $p$ -value =  $1.35 \times 10^{-10}$ ) located in the high enrichment region of CAD-associated super enhancer SNPs (16p13.3) has a long-range genetic interaction with locus *NPRL3*. Furthermore, GWAS3D also revealed 103 SNPs in 34 loci which may affect promoter activity by changing transcription factor binding site affinity (Supplemental Table 7), including 93 SNPs have direct effect by GWAS leading SNPs and 10 variants have indirect effect by high LD of GWAS leading SNPs. Interestingly, among the 103 functional SNPs identified by GWAS3D, 72 SNPs were also confirmed in HaploReg regulatory prediction results (Supplemental Table 5).

### Pathway enrichment analysis for CAD-associated super enhancer SNPs

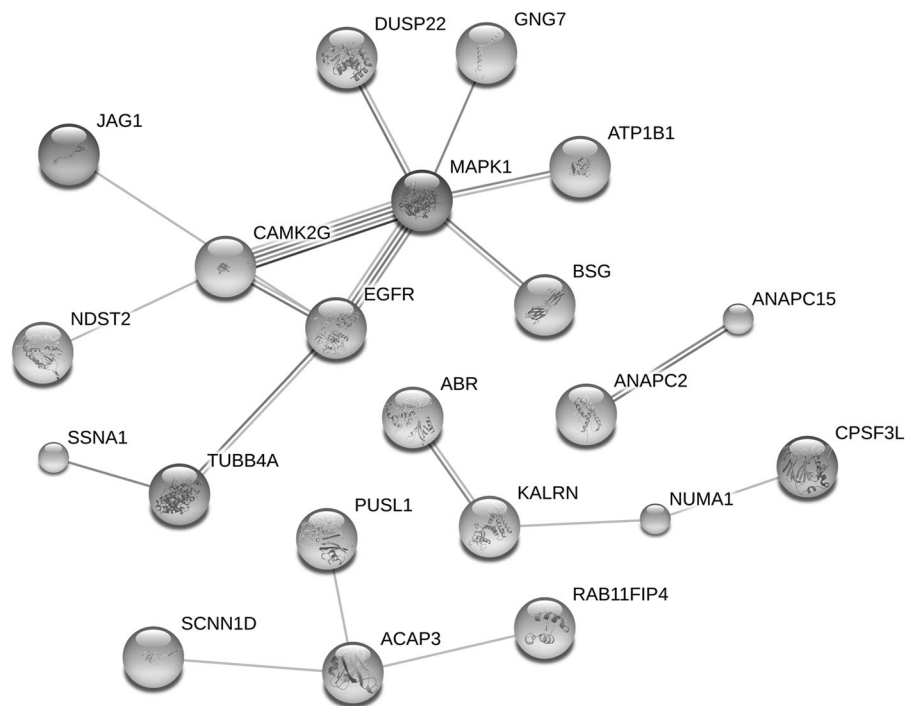
To systematically investigate whether the identified CAD-associated super enhancer SNPs were enriched in the specific biological process, we conduct pathway enrichment analysis through WebGestalt. We identified 11 significant pathways with FDR < 0.05. Interestingly, most of them were enriched in signaling pathways and regulation process (Table 2), e.g., cAMP signaling pathway and HIF-1

signaling pathway, which play a key role in pathogenesis of CAD [43–45]. Another interesting pathway is ErbB signaling pathway. Patients with risk for cardiotoxicity treated by trastuzumab therapy have demonstrated the critical role of ErbB signaling in anti-apoptotic process, maintenance of cardiomyocyte survival and cardiac hypertrophy, which may have a potential function in the CAD mechanism [46].

### Protein–protein Interaction (PPI) Network Associated with CAD

In order to partially explore and characterize the functional partnership among the identified CAD-associated super enhancer SNPs and the interaction network involved in the biological process of CAD, the 67 CAD-associated super enhancer SNPs target genes were uploaded into the STRING v10.0 database with default threshold value (Benjamini and Hochberg adjust  $p < 1.0 \times 10^{-10}$ ). Figure 4 revealed a strong association between the topological property and biological function of CAD-associated super enhancer SNPs target genes. Hub genes with the strong connections are *CAMK2G* and *MAPK1*. Gene *CAMK* encodes a number of serine-threonine protein kinases families which catalyzes the formation of many second messenger of  $Ca^{2+}$ . Interestingly, a previous study showed that the activity of *CAMK2* leads to cardiac arrhythmogenesis in transgenic *CAMK2C* mouse having heart failure [47]. *MAPK1* is a protein coding gene, *MAPK1/ERK2* cascade participates various biological processes, e.g., cell growth, migration, and differentiation, regulation of transcription, translation, and rearrangement of cytoskeletal. A recent study identified that the SNPs rs11913721, rs9340, and rs6928 within *MAPK1* can alter the susceptibility to CAD in Chinese Han perimenopausal women [45].

**Fig. 4** A functional protein association network analysis for CAD SNPs target genes. Connections are based on co-expression and experimental evidence with a STRING v10.0 summary score above 0.4. Each filled node denotes a gene; edges between nodes indicate protein-protein interactions between protein products of the corresponding genes. Different edge colors represent the types of evidence for the association (Color figure online)



## Discussion

There are much more genetic variations located in the non-coding regions compared with protein-coding regions in the human genome. Some important functional regulation elements, such as the super enhancer, have a great impact on cell-type specific gene expression. SNPs located in super enhancer may play an essential role in disease metabolism. In this study, we integrated GWAS meta-analysis result from CARDIoGRAMplusC4D and CAD cell/tissue-specific histone modification ChIP-seq data from GEO to identify CAD-associated SNPs in super enhancer. To systematically investigate the function and mechanism of these SNPs and their target genes, we further characterized their regulatory feature, performed transcription factor enrichment analysis, and conduct pathway and network analysis *etc.*

Our results indicate a total of 366 potential functional CAD-associated super enhancer SNPs in 67 loci. Interestingly, some of these SNPs showed a strong enrichment in CAD-related genes or novel candidate genes, e.g., *CBFA2T3*, *ZMIZ1* and *DIP2B* *etc.* We also performed a similar analysis based on 127 CAD-associated super enhancer SNPs identified by meta-analysis. We observed a similar gene pattern and functional annotation results. These 127 CAD-associated super enhancer SNPs were mapped to 47 loci which include the novel candidates *CBFA2T3*, *ZMIZ1*, and *DIP2B* *etc.* (Supplementary Table 2). Gene *CBFA2T3* (*CBFA2/RUNX1* Translocation Partner 3) includes 24 CAD-associated super enhancer SNPs. This gene encodes a family of the myeloid translocation gene

that interacts with transcription factors of DNA-binding sites and recruits multiple co-repressors to facilitate repression of transcription initiation. In a previous study, Gattenlöhner S. et al. [48] showed that the constant over-expression of *RUNX1* is unique molecular characteristics of cardiomyocytes within old post infarct scars in patients with ischemic cardiomyopathy. Recently, Voora et al. [49] performed a system pharmacogenomics study and identified that *RUNX1*, acts as an aspirin-responsive transcription factor, is associated with increased risk of myocardial infarction and death in patients living with cardiovascular disease. Another interesting gene is *ZMIZ1*. There are 18 CAD-associated super enhancer SNPs enriched in *ZMIZ1*-gene. This gene encodes a family of PIAS proteins which regulate the transcriptional activity of multiple transcription factors, such as Smad3/4, p53 and androgen receptor. In a previous study, androgen receptor gene CAG polymorphism has been reported to be associated with the severity of CAD both in postmenopausal women and in men [50, 51]. Importantly, we identified that CAD-associated super enhancer SNPs were highly enriched (44/366) in *DIP2B* gene which encodes a member of DIP2b protein and may participates in DNA methylation. Wong et al. [52] recently found a significant polygenic sharing between major depressive disorder and cardiometabolic traits, including positive associations with CAD, fat percentage, low-density lipoprotein *etc.*, and a negative association with high-density lipoprotein. Interestingly, one SNP rs10876041 in *DIP2B* is shared by major depressive disorder with CAD, low-density lipoprotein and total cholesterol.



Functional annotation of CAD-associated super enhancer SNPs through rVarBase showed that 94 SNPs in 39 loci have evidence of regulatory function, including 80 SNPs involved in chromatin interactive regulation, 6 SNPs overlapped with CpG islands, and 2 SNPs involved in regulation of lncRNA expression. We validated this result in haploReg v4.1 and found that 81 SNPs including 6 SNPs have effect on CpG islands and 2 SNPs involved in regulation of lncRNA expression were replicated by HaploReg analysis. These results highlighted the strong and reliable regulatory potential of these CAD-associated super enhancer SNPs. There is another interesting gene *ANAPC15* including 5 CAD-associated super enhancer SNPs, it encodes a subunit of the anaphase promoting complex/cyclosome and has a critical role in regulation of mitotic cell cycle spindle assembly checkpoint. A recent study which systematically characterized the role of lncRNA in control of heart development showed a positive correlation between lncRNA and *ANAPC15* that shares the same promoter in both fetal and adult hearts [53]. We also examined the effect of CAD-associated super enhancer SNPs on expression of their target genes by using data from GTEx project, 205 CAD-associated super enhancer SNPs in 38 loci showed eQTL evidences, including 181 CAD-associated super enhancer SNPs have numerous reported eQTL evidences in a wide variety of tissues and cell types. Furthermore, we indicated that CAD-associated super enhancer SNPs may alter the transcription factor binding affinity of HSF1 and INSM1. For example, HSF1 protects cardiomyocytes from death via activation of Akt kinase, as well as inactivation of caspase 3, and c-Jun N-terminal kinase [54]. In the context of cardiac ischemia/reperfusion, HSF1 expression up-regulate heat shock protein expression to protect against subsequent ischemia/reperfusion injury [39].

GWAS3D revealed 86 CAD-associated super enhancer SNPs which have long-range interaction signals. The interesting SNP rs8065824 located within the intron of *TMEM105* on chromosome 17 has a long-range genetic interaction with the locus *LCE3B*. *LCE3B* is a protein coding gene considered as precursors of the cornified envelope of the stratum corneum, recent study showed that psoriasis is associated with systemic inflammation and it is an increased risk factor for developing cardiovascular disease [55]. Interestingly, psoriasis and CAD share many underlying etiologic mechanisms, including increased T-helper type 1-mediated inflammation and dysregulation of angiogenesis [56]. This has led to the concept that systemic inflammation from psoriasis may predispose to CAD initiation and progression. Another interesting SNP rs12922862 located in the high enrichment region of CAD-associated super enhancer SNPs (16p13.3) has a long-range genetic interaction with locus *NPRL3*. *NPRL3* is a

component of the GATOR1 complex and acts as an inhibitor of the amino acid-sensing branch of the TORC1 pathway. Recent study demonstrated that *NPRL3* leads to a substantial increase in expression of many ribosomal protein genes and reduction of cell cycle-associated genes, it is required for normal development of the cardiovascular system in mouse [57]. Further investigation of mammalian *NPRL3* may elucidate how it affects TOR signaling and how it controls protein synthesis and affects the pathogenesis of CAD.

Pathway enrichment analysis of CAD-associated SNPs target genes not only confirmed well known cAMP signaling pathway and HIF-1 signaling pathway, but also identified some interesting candidates which may also contribute to CAD pathogenesis, such as ErbB signaling pathway. The cAMP signaling pathway regulates a multitude of diverse cellular events. In myocardial cells, cAMP formed by catecholamine-modulated,  $\beta$ -adrenergic receptors mediated excitation contraction coupling through the activation of downstream targets, e.g. cAMP dependent PKA, and the phosphorylation of the ryanodine receptor and the L-type  $\text{Ca}^{2+}$  channel, thus result in the increased amount of intracellular  $\text{Ca}^{2+}$  available for cardiac myocyte contraction [43]. HIF-1 is expressed in all metazoan species and has an essential role in regulating oxygen delivery and utilization to maintain oxygen homeostasis. Analysis of animal models suggests that HIF-1 has a crucial role in protection of pressure overload-induced heart failure and the pathophysiology of ischemic heart disease [44]. ErbB signaling pathway is an interesting candidate. Patients with risk for cardiotoxicity treated by trastuzumab demonstrated the critical role of ErbB signaling in anti-apoptotic process, maintenance of cardiomyocyte survival and cardiac hypertrophy, which suggest a potential function in the CAD pathogenesis [46].

PPI network analysis construct an interaction network and identified two hub genes *CAMK2G* and *MAPK1*. Interestingly, 4 regulatory SNPs targeted to gene *CAMK2G*, *CAMK2G* is the gamma isoform of *CAMK2* which encodes a serine-threonine protein kinases family that catalyzes the formation of many second messenger of  $\text{Ca}^{2+}$ . In the past decade, *CAMK2* has been shown to act as a critical regulator of cardiac function, e.g. involvement of transcriptional activation in cardiac hypertrophy [58], and the heart failure due to aberrant  $\text{Ca}^{2+}$  handling and apoptosis [59]. It has also been linked to electrical remodeling following myocardial infarction, as well as atrial and ventricular arrhythmias [47, 60]. Interestingly, a recent study reported a significant interaction of cAMP signaling pathway with *CAMK2* under normal condition and upon  $\beta$ -adrenergic stimulation [61] via regulation of phosphodiesterase 4D. There are 2 super enhancer SNPs in protein coding gene *MAPK1*. *MAPK1/ERK2* cascade participate cell growth, migration, and differentiation through regulation of transcription, translation,

and rearrangement of cytoskeletal. Recent study showed that variation in *MAPK1* can also alter the susceptibility to CAD in Chinese Han perimenopausal women [45].

In the current study, we performed comprehensive bioinformatics analyses to demonstrate the functional importance of the CAD-associated super enhancer SNPs. However, the databases used in these bioinformatics analyses are specialized, and we only focus on SNPs from the 1000 genomes project with  $MAF > 0.001$ . We admit that a number of known variants satisfying the MAF criterion are missing in our analysis. In addition, the results of functional annotation exclusively depend on computationally predicted regulation features, model and algorithm selection wherefore is critical for these kind of analysis. For example, the reliability of the transcription factor binding site predicted by SNP2TFBS is a function of the accuracy of the PWM model which is less accurate than recently proposed approaches [31]. Further, genomic sequence with higher affinity to a transcription factor is often insufficient for in vivo binding. A number of other factors may impact on binding such as chromatin accessibility [62]. Therefore, the functional importance of these candidate CAD-associated super enhancer SNPs may be overshadowed, further experiment validation should be conduct to confirm the functional mechanism of these candidates on CAD.

In conclusion, by integrating CAD GWAS and histone modification ChIP-seq data, we identified several CAD-associated super enhancer SNPs. Comprehensive bioinformatics analyses shed light on the SNPs located in super enhancer may a novel paradigm for the pathogenesis of CAD.

**Acknowledgements** We thank the DIAGRAM Consortium and GEO project for sharing the data set used in this research. This study was partially supported or benefited by grants from Technological Innovation Project of Foshan City (2017AG100102), and Medical Research Fund of Guangdong Province, China (A2017575).

## Compliance with ethical standards

**Conflict of interest** The authors declare that they have no conflict of interest.

## References

1. Wong ND. Epidemiological studies of CHD and the evolution of preventive cardiology. *Nat Rev Cardiol*. 2014;11:276–89.
2. Disease, GBD., Injury, I & Prevalence, C. Global, regional, and national incidence, prevalence, and years lived with disability for 310 diseases and injuries, 1990–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet*. 2016;388:1545–602.
3. Mathers CD, Loncar D. Projections of global mortality and burden of disease from 2002 to 2030. *PLoS Med*. 2006;3:e442.
4. Lloyd-Jones DM, Nam BH, D'Agostino RB Sr, Levy D, Murabito JM, Wang TJ, et al. Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA*. 2004;291:2204–11.
5. Consortium CAD, Deloukas P, Kanoni S, Willenborg C, Farrall M, Assimes TL, et al. Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet*. 2013;45:25–33.
6. Hindorff LA, Sethupathy P, Junkins HA, Ramos EM, Mehta JP, Collins FS, et al. Potential etiologic and functional implications of genome-wide association loci for human diseases and traits. *Proc Natl Acad Sci USA*. 2009;106:9362–7.
7. Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337:1190–5.
8. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. *Cell*. 2013;153:307–19.
9. Bulger M, Groudine M. Functional and mechanistic diversity of distal transcription enhancers. *Cell*. 2011;144:327–39.
10. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. *Cell*. 2013;155:934–47.
11. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. *Bioinformatics*. 2008;24:2938–9.
12. Tousoulis D, Charakida M, Stefanadis C. Endothelial function and inflammation in coronary artery disease. *Postgrad Med J*. 2008;84:368–71.
13. Kinlay S, Libby P, Ganz P. Endothelial function and coronary artery disease. *Curr Opin Lipidol*. 2001;12:383–9.
14. Motterle A, Pu X, Wood H, Xiao Q, Gor S, Ng FL, et al. Functional analyses of coronary artery disease associated variation on chromosome 9p21 in vascular smooth muscle cells. *Hum Mol Genet*. 2012;21:4021–9.
15. Rivard A, Andres V. Vascular smooth muscle cell proliferation in the pathogenesis of atherosclerotic cardiovascular diseases. *Histol Histopathol*. 2000;15:557–71.
16. Larsson B, Svardsudd K, Welin L, Wilhelmsen L, Bjorntorp P, Tibblin G. Abdominal adipose tissue distribution, obesity, and risk of cardiovascular disease and death: 13 year follow up of participants in the study of men born in 1913. *Br Med J*. 1984;288:1401–4.
17. Lapidus L, Bengtsson C, Larsson B, Pennert K, Rybo E, Sjostrom L. Distribution of adipose tissue and risk of cardiovascular disease and death: a 12 year follow up of participants in the population study of women in Gothenburg, Sweden. *Br Med J*. 1984;289:1257–61.
18. Szczepanska-Sadowska E, Cudnoch-Jedrzejewska A, Ufnal M, Zera T. Brain and cardiovascular diseases: common neurogenic background of cardiovascular, metabolic and inflammatory diseases. *J Physiol Pharmacol*. 2010;61:509–21.
19. Hess DC, D'Cruz IA, Adams RJ, Nichols FT 3rd. Coronary artery disease, myocardial infarction, and brain embolism. *Neurol Clin*. 1993;11:399–417.
20. Shantsila E, Tapp LD, Wrigley BJ, Pamukcu B, Apostolakis S, Montoro-Garcia S, et al. Monocyte subsets in coronary artery disease and their associations with markers of inflammation and fibrinolysis. *Atherosclerosis*. 2014;234:4–10.
21. Schlitt A, Heine GH, Blankenberg S, Espinola-Klein C, Doppeide JF, Bickel C, et al. CD14+CD16+ monocytes in coronary artery disease and their relationship to serum TNF-alpha levels. *Thromb Haemost*. 2004;92:419–24.
22. Ghattas A, Griffiths HR, Devitt A, Lip GY, Shantsila E. Monocytes in coronary artery disease and atherosclerosis: where are we now? *J Am Coll Cardiol*. 2013;62:1541–51.
23. Gayda M, Merzouk A, Choquet D, Ahmaidi S. Assessment of skeletal muscle fatigue in men with coronary artery disease using

- surface electromyography during isometric contraction of quadriceps muscles. *Arch Phys Med Rehabil.* 2005;86:210–5.
24. Paternostro G, Camici PG, Lammerstma AA, Marinho N, Baliga RR, Kooner JS, et al. Cardiac and skeletal muscle insulin resistance in patients with coronary heart disease. A study with positron emission tomography. *J Clin Invest.* 1996;98:2094–9.
  25. Zoll J, Steiner R, Meyer K, Vogt M, Hoppeler H, Fluck M. Gene expression in skeletal muscle of coronary artery disease patients after concentric and eccentric endurance training. *Eur J Appl Physiol.* 2006;96:413–22.
  26. Emami H, Singh P, MacNabb M, Vucic E, Lavender Z, Rudd JH, et al. Splenic metabolic activity predicts risk of future cardiovascular events: demonstration of a cardiopleenic axis in humans. *JACC Cardiovasc Imaging.* 2015;8:121–30.
  27. Zhang Y, Liu T, Meyer CA, Eeckhoutte J, Johnson DS, Bernstein BE, et al. Model-based analysis of ChIP-Seq (MACS). *Genome Biol.* 2008;9:R137.
  28. Dayem Ullah AZ, Lemoine NR, Chelala C. A practical guide for the functional annotation of genetic variations using SNPnexus. *Brief Bioinform.* 2013;14:437–47.
  29. Guo L, Du Y, Qu S, Wang J. rVarBase: an updated database for regulatory features of human variants. *Nucleic Acids Res.* 2016;44:D888–893.
  30. Ward LD, Kellis M. HaploRegv4: systematic mining of putative causal variants, cell types, regulators and target genes for human complex traits and disease. *Nucleic Acids Res.* 2016;44:D877–881.
  31. Kumar S, Ambrosini G, Bucher P. SNP2TFBS—a database of regulatory SNPs affecting predicted transcription factor binding site affinity. *Nucleic Acids Res.* 2017;45:D139–D144.
  32. Li MJ, Wang LY, Xia Z, Sham PC, Wang J. GWAS3D: detecting human regulatory variants by integrative analysis of genome-wide associations, chromosome interactions and histone modifications. *Nucleic Acids Res.* 2013;41:W150–158.
  33. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 2017;45:W130–W137.
  34. Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 2000;28:27–30.
  35. Szklarczyk D, Franceschini A, Wyder S, Forslund K, Heller D, Huerta-Cepas J, et al. STRINGv10: protein–protein interaction networks, integrated over the tree of life. *Nucleic Acids Res.* 2015;43:D447–452.
  36. Benjamini Y, Controlling HY. The false discovery rate: a practical and powerful approach to multiple testing. *J Roy Stat Soc B.* 1995;57:289–300.
  37. Ji HL, Zhao RZ, Chen ZX, Shetty S, Idell S, Matalon S. delta ENaC: a novel divergent amiloride-inhibitable sodium channel. *Am J Physiol Lung Cell Mol Physiol.* 2012;303:L1013–1026.
  38. Toko H, Minamino T, Komuro I. Role of heat shock transcription factor 1 and heat shock proteins in cardiac hypertrophy. *Trends Cardiovasc Med.* 2008;18:88–93.
  39. Willis MS, Patterson C. Hold me tight: role of the heat shock protein family of chaperones in cardiac disease. *Circulation.* 2010;122:1740–51.
  40. Mendelson CR. Role of transcription factors in fetal lung development and surfactant protein gene expression. *Annu Rev Physiol.* 2000;62:875–915.
  41. Zhu Y, Liao HL, Wang N, Yuan Y, Ma KS, Verna L, et al. Lipoprotein promotes caveolin-1 and Ras translocation to caveolae: role of cholesterol in endothelial signaling. *Arterioscler Thromb Vasc Biol.* 2000;20:2465–70.
  42. Wang H, Hu B, Feng B. Decreased beta cell function and insulin sensitivity contributed to coronary artery disease in patients with normal glucose tolerance. *J Atheroscler Thromb.* 2012;19:806–13.
  43. Zaccolo M. cAMP signal transduction in the heart: understanding spatial control for the development of novel therapeutic strategies. *Br J Pharmacol.* 2009;158:50–60.
  44. Semenza GL. Hypoxia-inducible factor 1 and cardiovascular disease. *Annu Rev Physiol.* 2014;76:39–56.
  45. Guo N, Zhang N, Yan L, Cao X, Wang J, Wang Y. Correlation between genetic polymorphisms within the MAPK1/HIF-1/HO-1 signaling pathway and risk or prognosis of perimenopausal coronary artery disease. *Clin Cardiol.* 2017;40:597–604.
  46. Sanchez-Soria P, Camenisch TD. ErbB signaling in cardiac development and disease. *Semin Cell Dev Biol.* 2010;21:929–35.
  47. Sag CM, Wadsack DP, Khabbazzadeh S, Abesser M, Grefe C, Neumann K, et al. Calcium/calmodulin-dependent protein kinase II contributes to cardiac arrhythmogenesis in heart failure. *Circ Heart Fail.* 2009;2:664–75.
  48. Gattenlohner S, Waller C, Ertl G, Bultmann BD, Muller-Hermelink HK, Marx A. NCAM(CD56) and RUNX1(AML1) are up-regulated in human ischemic cardiomyopathy and a rat model of chronic cardiac ischemia. *Am J Pathol.* 2003;163:1081–90.
  49. Voora D, Rao AK, Jalagadugula GS, Myers R, Harris E, Ortel TL, et al. Systems pharmacogenomics finds RUNX1 is an aspirin-responsive transcription factor linked to cardiovascular disease and coloncancer. *EBioMedicine.* 2016;11:157–64.
  50. Alevizaki M, Cimponeriu AT, Garofallaki M, Sarika HL, Alevizaki CC, Papamichael C, et al. The androgen receptor gene CAG polymorphism is associated with the severity of coronary artery disease in men. *Clin Endocrinol.* 2003;59:749–55.
  51. Saltiki K, Cimponeriu A, Garofalaki M, Sarika L, Papatoma A, Stamatelopoulos K, et al. Severity of coronary artery disease in postmenopausal women: association with the androgen receptor gene (CAG)n repeat polymorphism. *Menopause.* 2011;18:1225–31.
  52. Brian Chi-Fung Wong, CK-LC, Hon-Cheong So. Genetic overlap of depression with cardiometabolic diseases and implications for drug repurposing for comorbidities. *bioRxiv.* (2017).
  53. He C, Hu H, Wilson KD, Wu H, Feng J, Xia S, et al. Systematic characterization of long noncoding RNAs reveals the contrasting coordination of cis- and trans-molecular regulation in human fetal and adult hearts. *Circ Cardiovasc Genet.* 2016;9:110–8.
  54. Zou Y, Zhu W, Sakamoto M, Qin Y, Akazawa H, Toko H, et al. Heat shock transcription factor 1 protects cardiomyocytes from ischemia/reperfusion injury. *Circulation.* 2003;108:3024–30.
  55. Prodanovich S, Kirsner RS, Kravetz JD, Ma F, Martinez L, Federman DG. Association of psoriasis with coronary artery, cerebrovascular, and peripheral vascular diseases and mortality. *Arch Dermatol.* 2009;145:700–3.
  56. Armstrong AW, Voyles SV, Armstrong EJ, Fuller EN, Rutledge JC. Angiogenesis and oxidative stress: common mechanisms linking psoriasis with atherosclerosis. *J Dermatol Sci.* 2011;63:1–9.
  57. Kowalczyk MS, Hughes JR, Babbs C, Sanchez-Pulido L, Szumska D, Sharpe JA, et al. Npr13 is required for normal development of the cardiovascular system. *Mamm Genome.* 2012;23:404–15.
  58. Zhang T, Brown JH. Role of Ca<sup>2+</sup>/calmodulin-dependent protein kinase II in cardiac hypertrophy and heart failure. *Cardiovasc Res.* 2004;63:476–86.
  59. Hoch B, Meyer R, Hetzer R, Krause EG, Karczewski P. Identification and expression of delta-isoforms of the multifunctional Ca<sup>2+</sup>/calmodulin-dependent protein kinase in failing and non-failing human myocardium. *Circ Res.* 1999;84:713–21.
  60. Mohler PJ, Hund TJ. Role of CaMKII in cardiovascular health, disease, and arrhythmia. *Heart Rhythm.* 2011;8:142–4.

61. Mika D, Richter W, Conti M. A CaMKII/PDE4D negative feedback regulates cAMP signaling. *Proc Natl Acad Sci USA*. 2015;112:2023–8.
62. Slattery M, Zhou T, Yang L, Dantas Machado AC, Gordan R, Rohs R. Absence of a simple code: how transcription factors read the genome. *Trends Biochem Sci*. 2014;39:381–99.