

SCIENTIFIC REPORTS



OPEN

Tissue-specific Network Analysis of Genetic Variants Associated with Coronary Artery Disease

Xiao Miao¹, Xinlin Chen², Zhijun Xie^{3,4} & Honghuang Lin⁴

Coronary artery disease (CAD) is a leading cause of death worldwide. Recent genome-wide association studies have identified more than one hundred susceptibility loci associated with CAD. However, the underlying mechanism of these genetic loci to CAD susceptibility is still largely unknown. We performed a tissue-specific network analysis of CAD using the summary statistics from one of the largest genome-wide association studies. Variant-level associations were summarized into gene-level associations, and a CAD-related interaction network was built using experimentally validated gene interactions and gene coexpression in coronary artery. The network contained 102 genes, of which 53 were significantly associated with CAD. Pathway enrichment analysis revealed that many genes in the network were involved in the regulation of peripheral arteries. In summary, we performed a tissue-specific network analysis and found abnormalities in the peripheral arteries might be an important pathway underlying the pathogenesis of CAD. Future functional characterization might further validate our findings and identify potential therapeutic targets for CAD.

Coronary artery disease (CAD) is a leading cause of death worldwide^{1–3}. CAD is heritable, and a familial history of CAD is associated with a significantly increased CAD risk⁴. Recent genome-wide association studies (GWAS) have identified more than one hundred genetic loci associated with CAD^{5–11}. However, most of these loci are located outside of protein coding regions, and their implication on CAD is still largely unknown. Moreover, many studies have been focused on the association of single variants with CAD. Given that the majority of complex diseases including CAD are caused by the interplay of many genetic and environmental factors^{12,13}, it is important to jointly investigate the combinatory effects of multiple genetic variants on biological pathways and interaction networks^{14–16}.

The gene function is highly dependent on the tissue where the gene is expressed¹⁷, which is controlled by very distinct regulatory programs¹⁸. Genes with tissue-specific expression have shown important physiological processes for complex organisms¹⁹. However, functional studies of human genes have been traditionally carried out on specific cell lines, and the characterization of tissue-specific interactions is predominantly based on a small sample size. Recent advances in next generation sequencing provide an unprecedented opportunity to profile gene expression in a much larger scale of human samples^{20,21}. An excellent example is the Genotype-Tissue Expression (GTEx) project, which has characterized the association of genetic variants with gene expression in nearly 50 different types of human tissues that were collected from approximately 1000 individuals²².

The objective of our current study is to understand the function of CAD-related genetic loci through the gene interaction network. The associations of individual variants with CAD were summarized into gene-level associations, which were then combined with experimentally validated gene interactions as well as gene coexpression in coronary artery. A tissue-specific interaction network then built to examine the interactions between CAD-related genes and their potential functions in terms of biological pathways.

¹Innovation Research Institute of Traditional Chinese Medicine, Shanghai University of Traditional Chinese Medicine, Shanghai, China. ²Department of Cardiovascular Medicine, Longhua Hospital, Shanghai University of Traditional Chinese Medicine, Shanghai, China. ³College of Basic Medical Science, Zhejiang Chinese Medical University, Hangzhou, Zhejiang Province, China. ⁴Section of Computational Biomedicine, Department of Medicine, Boston University School of Medicine, Boston, MA, USA. Correspondence and requests for materials should be addressed to X.M. (email: lieast@263.net) or H.L. (email: hlin@bu.edu)

| Chromosome | Gene | #Variants within the gene | Gene association with CAD (<i>P</i> value) | Best variant with the gene | Best variant association with CAD (<i>P</i> value) |
|------------|-------------------|---------------------------|---|----------------------------|---|
| 9p21.3 | <i>CDKN2B-AS1</i> | 283 | 9.45E-69 | rs2891168 | 2.29E-98 |
| 9p21.3 | <i>CDKN2B</i> | 96 | 2.04E-38 | rs7028268 | 4.98E-48 |
| 9p21.3 | <i>CDKN2A</i> | 119 | 3.42E-32 | rs3217992 | 1.03E-42 |
| 9p21.3 | <i>CDKN2A-AS1</i> | 104 | 2.26E-26 | rs3217992 | 1.03E-42 |
| 6q25.3 | <i>LPA</i> | 180 | 8.34E-25 | rs55730499 | 5.39E-39 |
| 19p13.2 | <i>SMARCA4</i> | 219 | 5.88E-21 | rs56289821 | 4.44E-15 |
| 15q25.1 | <i>ADAMTS7</i> | 216 | 9.37E-20 | rs4468572 | 4.44E-16 |
| 19p13.2 | <i>LDLR</i> | 256 | 1.71E-16 | rs56289821 | 4.44E-15 |
| 1p13.3 | <i>SORT1</i> | 114 | 2.18E-16 | rs7528419 | 1.97E-23 |
| 15q25.1 | <i>MORF4L1</i> | 134 | 6.37E-16 | rs4468572 | 4.44E-16 |
| 19p13.2 | <i>MIR6886</i> | 182 | 2.13E-15 | rs56289821 | 4.44E-15 |
| 1p13.3 | <i>MYBPHL</i> | 113 | 3.96E-15 | rs7528419 | 1.97E-23 |
| 10q11.21 | <i>C10orf142</i> | 135 | 1.31E-14 | rs1746050 | 6.28E-13 |
| 6p24.1 | <i>PHACTR1</i> | 656 | 1.59E-14 | rs9349379 | 1.81E-42 |
| 1p13.3 | <i>PSRC1</i> | 128 | 1.88E-14 | rs7528419 | 1.97E-23 |
| 6q23.2 | <i>LINC01312</i> | 113 | 9.18E-14 | rs12202017 | 1.98E-11 |
| 2q33.2 | <i>WDR12</i> | 51 | 2.08E-13 | rs115396314 | 5.11E-18 |
| 6q26 | <i>PLG</i> | 184 | 2.28E-13 | rs2315065 | 2.88E-34 |
| 1q41 | <i>MIA3</i> | 99 | 7.04E-13 | rs67180937 | 1.01E-12 |
| 1q41 | <i>TAF1A-AS1</i> | 111 | 9.66E-13 | rs35700460 | 1.38E-12 |
| 1p13.3 | <i>CELSR2</i> | 159 | 9.70E-13 | rs7528419 | 1.97E-23 |
| 2q33.2 | <i>CARF</i> | 67 | 1.28E-12 | rs115654617 | 3.12E-18 |
| 10q11.21 | <i>LINC00841</i> | 222 | 1.57E-12 | rs1870634 | 5.55E-15 |
| 1q41 | <i>TAF1A</i> | 132 | 1.96E-12 | rs35700460 | 1.38E-12 |
| 1p32.2 | <i>PLPP3</i> | 211 | 2.40E-12 | rs9970807 | 5.00E-14 |

Table 1. Top 25 genes associated with CAD.

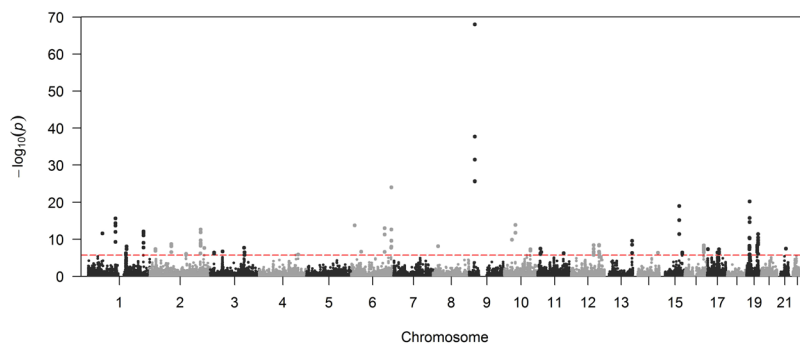


Figure 1. Manhattan plot of gene-level association with CAD. Each dot represents one gene. The x-axis represents chromosome positions, whereas the y-axis represents the $\log_{10}(P)$. The red dash indicates $P < 0.05/26,228 = 1.91 \times 10^{-6}$. Only autosomal variants were included in the analysis.

Results

Gene-level Association with CAD. The association of genetic variants with CAD was previously investigated⁷. The study included 9,455,778 common variants, among which 2,213 variants passed the genome-wide significance ($P < 5 \times 10^{-8}$). Variant-level associations were then summarized into gene-level associations using fastBAT²³. Variants within 50 kb of the gene region were collapsed and jointly modelled. Figure 1 is the Manhattan plot showing the association of each of the 26,228 genes with CAD. The Q-Q plot of the associations is shown in Supplementary Figure 1. A total of 143 genes were significantly associated with CAD after Bonferroni correction of multiple testing ($P < 0.05/26,228 = 1.91 \times 10^{-6}$). Table 1 shows the top 25 genes associated with CAD, and the full list of CAD-related genes is shown in Supplementary Table 1. The most significant gene was *CDKN2B-AS1* ($P = 9.45 \times 10^{-69}$), which is the antisense of *CDKN2B* that is located at the 9p21 locus. The locus has long been recognized to be associated with various cardiovascular diseases^{24–28}. We also performed a sensitivity analysis by expanding the flanking sequence to 100 kb to include more regulatory variants. As shown in Supplementary Figure 2, the association was highly correlated ($R^2 = 0.88$).

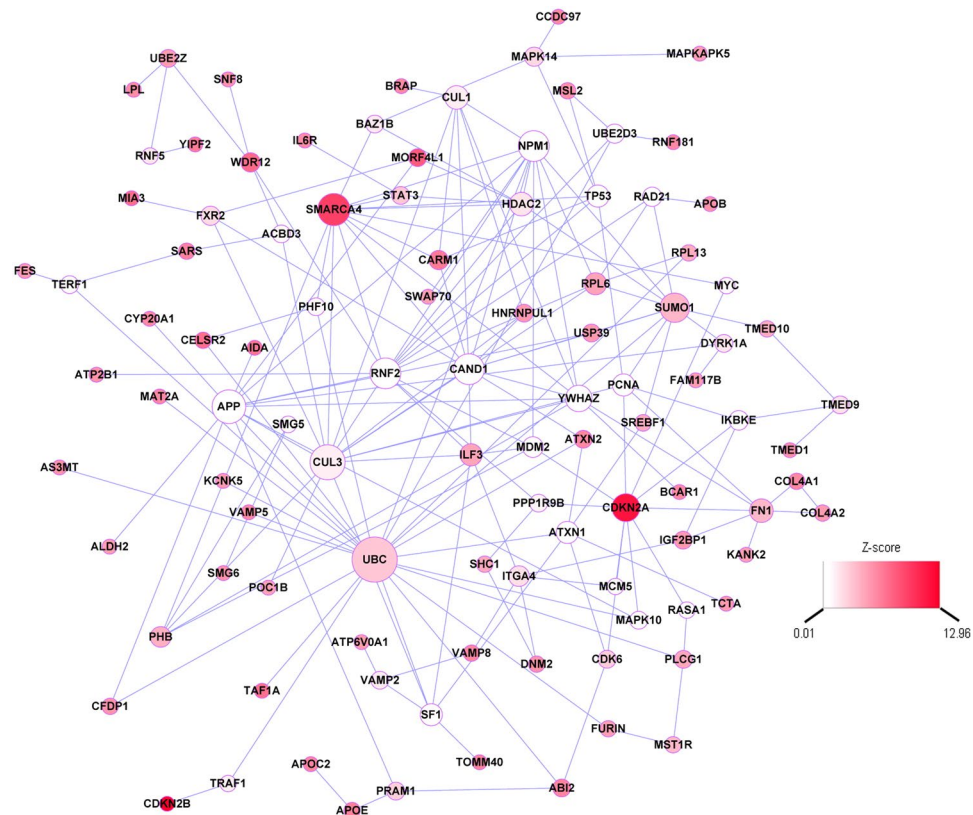


Figure 2. CAD-related network derived from protein-protein interaction. Each node represents one gene, whereas each edge represents the interaction between two genes. The nodes were colored to represent their association with CAD: red color represents genes that were associated with CAD, white color represents genes that were not associated with CAD. The node size is proportional to the number of edges that the node is connected to.

| Disease pathway | #Genes in the pathway | Ratio of enrichment | P value | FDR | Overlapping Genes |
|--|-----------------------|---------------------|----------|----------|---|
| Abnormalities of the peripheral arteries | 97 | 8.98 | 1.63E-06 | 3.52E-03 | COL4A1; COL4A2; APOB; APOE; APP; LPL; TP53; BAZ1B |
| Xanthomatosis | 19 | 22.93 | 2.06E-05 | 2.96E-02 | APOB; APOC2; APOE; LPL |
| Cerebral ischemia | 46 | 11.84 | 4.94E-05 | 5.33E-02 | COL4A1; COL4A2; APP; TP53; BAZ1B |
| Coronary artery disease | 51 | 10.68 | 8.20E-05 | 6.80E-02 | APOB; APOE; LPL; TP53; BAZ1B |
| Abnormality of the coronary arteries | 53 | 10.28 | 9.89E-05 | 6.80E-02 | APOB; APOE; LPL; TP53; BAZ1B |
| Neoplasm of the adrenal cortex | 11 | 29.71 | 0.00011 | 6.80E-02 | CDKN2B; MDM2; TP53 |

Table 2. Most significant disease pathways enriched with CAD-related genes (FDR < 0.1).

Tissue-specific Interaction Network Related to CAD. We then built a CAD-related interaction network by integrating GWAS and gene coexpression in coronary artery. As shown in Fig. 2, the network is comprised of 102 nodes and 182 edges. Each node represents one gene, and each edge represents the interaction between two genes. These genes were interconnected together, so not all of the 143 CAD-related gene were included. Among the 102 genes in the network, 53 were CAD-related genes ($P < 1.91 \times 10^{-6}$).

We then examined the potential functional implication of the network using WebGestalt²⁹. As shown in Table 2, six disease pathways were significantly enriched with genes in the network (FDR < 0.1). The most significant pathway is abnormalities of the peripheral arteries ($P = 1.63 \times 10^{-6}$, FDR = 3.52×10^{-3}), which includes eight CAD-related genes. In addition, many CAD-related genes were involved in xanthomatosis ($P = 2.06 \times 10^{-5}$, FDR = 2.96×10^{-2}) and cerebral ischemia ($P = 4.94 \times 10^{-5}$, FDR = 5.33×10^{-2}).

A recent study³⁰ of CAD prioritized 184 candidate genes as the most likely causal genes for CAD based on functional evidence. These genes were involved in 286 modules. Interestingly, 26 of those candidate genes were also included in our network (enrichment $P < 2.2e-16$). We also examined if genes in the network are potential drug targets. DrugBank³¹ was queried, and 13 genes in the CAD-related network were reported as therapeutic targets for at least one known or developing drug, suggesting the potential of network analysis to identify drug targets.

Identification of Key Drivers in the Network. We also examined the structure of the CAD-related network, and tested if they were any key drivers of the network. Each gene in the network was evaluated by the association of its neighbors with CAD, and the distribution was then compared with random distribution. A single gene, *UBC*, turned out to be one of the most important genes in the network. The gene was nominally associated with CAD ($P = 1.25 \times 10^{-3}$) but did not reach the genome-wide significance cutoff ($P < 1.91 \times 10^{-6}$). However, it interacted with 22 other genes, including 10 CAD-related genes, suggesting that it might be an important regulator of the CAD-related pathways.

We also calculated the weighted centroid of the network, which took into consideration of both the weight of neighboring genes as well as the strength of interactions (represented by the absolute value of the correlation coefficient between two genes). As shown in Supplementary Table 2, *UBC*, together with *CAND1* and *SMARCA4*, were among the top centroid genes.

Discussion

One challenge in the post-GWAS era is to understand molecular mechanisms underlying the association of GWAS loci with diseases. Given that many genes are only expressed in certain tissues, it is important to study gene functions in a tissue-specific manner. Here we built a CAD-related interaction network in coronary artery, and found that many of genes in the network were involved in abnormalities of the peripheral arteries. Our study suggests that the disturbing of peripheral artery functions might be an important pathway leading to CAD.

UBC was found to be a key driver of the network. The gene encodes ubiquitin C, a highly conserved gene³² that acts as a potential target of N-Formylmethionine³³. In combination with proteasome, the ubiquitin-proteasome system (UPS) is responsible for the degradation of up to 80–90% of proteins in mammalian cells³⁴, which is essential to the removal of non-functional or damaged polypeptides³⁵. The ubiquitination also regulates multiple cardiac signal transduction pathways^{36–38} as well as the promotion of pathologic hypertrophic growth of cardiomyocytes^{39,40}.

It is interesting to see that *APOB* and *APOE* were both involved in multiple pathways enriched with CAD-related genes. *APOB* encodes apolipoprotein B, a lipoprotein that is known to be involved in the development of CAD^{41,42}. The inclusion of apolipoprotein B level into risk models would significantly improve the prediction performance of future risk of coronary heart disease⁴³. Genetic variations in *APOB* were also associated with Mendelian diseases such as familial hypobetalipoproteinemia⁴⁴. *APOE* encodes a ligand that binds to both low density lipoprotein receptor and APOE-specific receptor. It is involved in the regulation of cholesterol and the metabolism of lipoproteins, and the polymorphisms of *APOE* are associated with atherosclerosis⁴⁵ and coronary heart disease^{46,47}.

Tissue-specific expression has been recognized as an important pattern for many diseases. Traditional expression quantitative trait loci (eQTLs) studies were mostly focused on gene expression in the blood due to convenience and accessibility. However, these studies have limited power to detect tissue-specific expression profiles. Recent efforts are being devoted to identify eQTLs across a variety of tissue types^{22,48,49}, which would empower an in-depth study of gene function among different types of tissues^{50,51}.

GWAS provided an unbiased approach to screen a huge amount of genetic variants with complex diseases. In order to correct for multiple testing and reduce false positives, a stringent significance cutoff is applied (typically $P < 5 \times 10^{-8}$), which on the other hand, could result in the loss of many true associations. By integrating gene interactions, we were able to reprioritize gene signals, and identify some additional associations that might be otherwise missed because of moderate significance. One example is *TGFBI*, which encodes a transforming growth factor, which plays an important role in the regulation vascular smooth muscle⁵². The gene was significantly associated with CAD after summarizing genetic variants within the gene region ($P = 6.26 \times 10^{-9}$). However, the most significant SNP within the gene region was rs15052 ($P < 2.21 \times 10^{-7}$), which did not reach the genome-wide significance cutoff ($P < 5 \times 10^{-8}$). A recent study¹⁰ that combined UK Biobank together with CARDIoGRAMplusC4D 1000 Genomes-based GWAS and the Myocardial Infarction Genetics and CARDIoGRAM Exome found that rs8108632 within the *TGFBI* region was significantly associated with CAD ($P = 4.04 \times 10^{-8}$). Our results suggest that a combination of joint testing of multiple variants together with gene interactions could offer additional power to identify novel susceptibility loci for complex diseases.

Given that many top variants identified by GWAS are just tagging variants, it is challenging to pinpoint the causal variants/genes. For example, the most significant variant associated with body mass index is rs9930506, which is located within the intron of *FTO*. However, it was found that the causative gene at the locus is actually *IRX3*, which is located ~480 kb away from the top SNP rs9930506⁵³. Appropriately 6.22% of *cis*-eQTLs in coronary artery are even more than 500 kb away from the corresponding genes²², suggesting that causal genetic variants could be far away from the genes. On the other hand, the inclusion of long-range regulation could also result in an increase of irrelevant genes. Therefore it is useful to include additional annotations such as those derived from the ENCODE Project⁵⁴ and the Roadmap Epigenomics Project⁵⁵.

Our study has several limitations. The interaction network was built from previously reported protein interactions. However, it was estimated that only a small proportion of interactions have been characterized^{56–58}. Therefore, many important interactions might still be missing and thus were not included in the current analysis. In addition, tissue-specific interactions were estimated from the correlation of gene expression, which had limited power to identify interactions between genes with low expression. It is also worth to note that the key driver analysis is a hypothesis generating approach solely based on the strength of associations of individual genes and the network structure. It is by no means to indicate causality between genes. Further experimental validation is needed to understand the mechanism of associations and the biological processes involved in the etiology of CAD.

In conclusion, we performed a tissue-specific network analysis of genetic variants associated with CAD. Our study underscored the role of abnormalities in the peripheral arteries in the pathogenesis of CAD. Future functional characterization of CAD-related gene might identify potential therapeutic targets for CAD.

Methods

Association of Genetic Variants with CAD. The summary statistics of genetic variants associated with CAD was obtained from the CARDIoGRAMplusC4D 1000 Genomes-based genome-wide association study (<http://www.cardiogramplusc4d.org/data-downloads/>), which included 48 studies with a total of 60,801 CAD cases and 123,504 controls⁷. The raw genotypes were imputed to 1000 Genomes phase 1 v3 that included more than 38 million variants, but only common variants (minor allele frequency $\geq 1\%$) were included in the current analysis.

Derivation of Gene-level Association with CAD. Variant-level associations were summarized into gene-level associations using fastBAT²³. For each gene, we took into account of all the variants within 50 kb of the gene region, and joint tested their association with CAD. Given that tens of thousands of genes were tested, a nominal significance cutoff ($P < 0.05$) could result in a number of false positives. Therefore Bonferroni correction was used, which is a conservative adjustment that assumes all tests are independent. Genes with p-value less than $0.05/N$ were considered as significant, where N was number of tested genes.

Construction of a Tissue-specific Interaction Network. Gene interactions were obtained from the iRefIndex database⁵⁹. Self-interactions or interactions involved in non-human genes were removed from downstream analyses. Gene coexpression was based on 173 coronary artery samples from the GTEx project²². Pearson correlation was calculated for each interaction gene pair, and pairs with normalized correlation higher than 0.25 were considered as coexpression in the tissue.

Construction of Gene Interaction Network Related to CAD. A dense module searching strategy⁶⁰ was implemented to identify modules enriched with CAD-related genes. Each gene was assigned a score equivalent to the absolute value of the z-score of the association with CAD. Seed genes were defined as those significantly associated with CAD after Bonferroni correction. The module searching started with a single seed gene. Neighboring genes were then added sequentially to the module if the addition would increase the overall module score⁶¹, which was defined as $Z_m = \frac{\sum g_i}{\sqrt{k}}$, where k was the number of genes in the module, and g_i was the score of gene i. The process iterated until no more genes could be added. Modules derived from different seed genes were highly overlapped and thus were merged to create a combined interaction network.

Identification of Key Drivers in the Network. Key drivers were defined as genes that interact with more CAD-related genes than what would be expected from a randomly selected gene set with an equal number of genes. These genes are pivotal to the structure of the network and might be potential targets for further functional characterization. The score of each gene was defined as the z-score of the association with CAD. The Kolmogorov–Smirnov test was then used to assess the deviation of scores of neighboring genes from the random expectation. To correct for multiple testing, Bonferroni adjustment was used and genes with $P < 0.05/N$ were defined as key drivers, where N was the total number of genes in the network. In addition, we also calculated the weighted centroid of the network, which was defined as $C_i = g_i + \sum_k |w_{i,k}| * g_k$, where C_i was the weighted centroid of gene i, and $w_{i,k}$ was the correlation coefficient between gene i and gene k, and g_i is the score of gene k.

References

1. Benjamin, E. J. *et al.* Heart Disease and Stroke Statistics-2017 Update: A Report From the American Heart Association. *Circulation* **135**, e146–e603 (2017).
2. Townsend, N. *et al.* Cardiovascular disease in Europe: epidemiological update 2016. *Eur Heart J* **37**, 3232–3245 (2016).
3. Mortality, G. B. D. & Causes of Death, C. Global, regional, and national life expectancy, all-cause mortality, and cause-specific mortality for 249 causes of death, 1980–2015: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet* **388**, 1459–1544 (2016).
4. Lloyd-Jones, D. M. *et al.* Parental cardiovascular disease as a risk factor for cardiovascular disease in middle-aged adults: a prospective study of parents and offspring. *JAMA* **291**, 2204–2211 (2004).
5. Consortium, C. A. D. *et al.* Large-scale association analysis identifies new risk loci for coronary artery disease. *Nat Genet* **45**, 25–33 (2013).
6. Coronary Artery Disease Genetics, C. A genome-wide association study in Europeans and South Asians identifies five new loci for coronary artery disease. *Nat Genet* **43**, 339–344 (2011).
7. Nikpay, M. *et al.* A comprehensive 1,000 Genomes-based genome-wide association meta-analysis of coronary artery disease. *Nat Genet* **47**, 1121–1130 (2015).
8. Schunkert, H. *et al.* Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease. *Nat Genet* **43**, 333–338 (2011).
9. Webb, T. R. *et al.* Systematic Evaluation of Pleiotropy Identifies 6 Further Loci Associated With Coronary Artery Disease. *J Am Coll Cardiol* **69**, 823–836 (2017).
10. Nelson, C. P. *et al.* Association analyses based on false discovery rate implicate new loci for coronary artery disease. *Nat Genet* **49**, 1385–1391 (2017).
11. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet* **44**(369–375), S361–363 (2012).
12. O'Donnell, C. J. & Nabel, E. G. Genomics of cardiovascular disease. *N Engl J Med* **365**, 2098–2109 (2011).
13. Dai, X., Wiernek, S., Evans, J. P. & Runge, M. S. Genetics of coronary artery disease and myocardial infarction. *World journal of cardiology* **8**, 1–23 (2016).
14. Zhang, Y. *et al.* Network analysis reveals functional cross-links between disease and inflammation genes. *Scientific reports* **3**, 3426 (2013).
15. Kontou, P. I., Pavlopoulou, A., Dimou, N. L., Pavlopoulos, G. A. & Bagos, P. G. Network analysis of genes and their association with diseases. *Gene* **590**, 68–78 (2016).
16. Stevens, A. *et al.* Network analysis: a new approach to study endocrine disorders. *Journal of molecular endocrinology* **52**, R79–93 (2014).
17. Huttenhower, C. *et al.* Exploring the human genome with functional maps. *Genome Res* **19**, 1093–1106 (2009).
18. Sonawane, A. R. *et al.* Understanding Tissue-Specific GeneRegulation. *Cell reports* **21**, 1077–1088 (2017).

19. Winter, E. E., Goodstadt, L. & Ponting, C. P. Elevated rates of protein secretion, evolution, and disease among tissue-specific genes. *Genome Res* **14**, 54–61 (2004).
20. Morozova, O. & Marra, M. A. Applications of next-generation sequencing technologies in functional genomics. *Genomics* **92**, 255–264 (2008).
21. Metzker, M. L. Sequencing technologies - the next generation. *Nature reviews* **11**, 31–46 (2010).
22. Consortium, G. T. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* **348**, 648–660 (2015).
23. Bakshi, A. *et al.* Fast set-based association analysis using summary data from GWAS identifies novel gene loci for human complex traits. *Scientific reports* **6**, 32894 (2016).
24. Doria, A. *et al.* Interaction between poor glycemic control and 9p21 locus on risk of coronary artery disease in type 2 diabetes. *JAMA* **300**, 2389–2397 (2008).
25. Helgadottir, A. *et al.* A common variant on chromosome 9p21 affects the risk of myocardial infarction. *Science* **316**, 1491–1493 (2007).
26. McPherson, R. *et al.* A common allele on chromosome 9 associated with coronary heart disease. *Science* **316**, 1488–1491 (2007).
27. Wellcome Trust Case Control, C. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* **447**, 661–678 (2007).
28. Samani, N. J. *et al.* Genomewide association analysis of coronary artery disease. *N Engl J Med* **357**, 443–453 (2007).
29. Wang, J., Duncan, D., Shi, Z. & Zhang, B. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res* **41**, W77–83 (2013).
30. Lempiainen, H. *et al.* Network analysis of coronary artery disease risk genes elucidates disease mechanisms and druggable targets. *Scientific reports* **8**, 3434 (2018).
31. Wishart, D. S. *et al.* DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res* **46**, D1074–D1082 (2018).
32. Wiborg, O. *et al.* The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. *EMBO J* **4**, 755–759 (1985).
33. Law, V. *et al.* DrugBank 4.0: shedding new light on drug metabolism. *Nucleic Acids Res* **42**, D1091–1097 (2014).
34. Zolk, O., Schenke, C. & Sarikas, A. The ubiquitin-proteasome system: focus on the heart. *Cardiovasc Res* **70**, 410–421 (2006).
35. Goldberg, A. L. Protein degradation and protection against misfolded or damaged proteins. *Nature* **426**, 895–899 (2003).
36. Laine, A. & Ronai, Z. Ubiquitin chains in the ladder of MAPK signaling. *Science's STKE: signal transduction knowledge environment* **2005**, re5 (2005).
37. Abriel, H., Kamynina, E., Horisberger, J. D. & Staub, O. Regulation of the cardiac voltage-gated Na⁺ channel (H1) by the ubiquitin-protein ligase Nedd4. *FEBS Lett* **466**, 377–380 (2000).
38. van Bemmelen, M. X. *et al.* Cardiac voltage-gated sodium channel Nav1.5 is regulated by Nedd4-2 mediated ubiquitination. *Circ Res* **95**, 284–291 (2004).
39. Wilkins, B. J. *et al.* Calcineurin/NFAT coupling participates in pathological, but not physiological, cardiac hypertrophy. *Circ Res* **94**, 110–118 (2004).
40. Willis, M. S. & Patterson, C. Into the heart: the emerging role of the ubiquitin-proteasome system. *J Mol Cell Cardiol* **41**, 567–579 (2006).
41. Saleheen, D. *et al.* Association of HDL cholesterol efflux capacity with incident coronary heart disease events: a prospective case-control study. *The lancet. Diabetes & endocrinology* **3**, 507–513 (2015).
42. Sniderman, A. *et al.* Association of coronary atherosclerosis with hyperapobetalipoproteinemia increased protein but normal cholesterol levels in human plasma low density (beta) lipoproteins. *Proc Natl Acad Sci USA* **77**, 604–608 (1980).
43. Pencina, M. J. *et al.* Apolipoprotein B improves risk assessment of future coronary heart disease in the Framingham Heart Study beyond LDL-C and non-HDL-C. *European journal of preventive cardiology* **22**, 1321–1327 (2015).
44. Linton, M. F., Farese, R. V. Jr. & Young, S. G. Familial hypobetalipoproteinemia. *J Lipid Res* **34**, 521–541 (1993).
45. Davignon, J., Gregg, R. E. & Sing, C. F. Apolipoprotein E polymorphism and atherosclerosis. *Arteriosclerosis* **8**, 1–21 (1988).
46. Tiret, L. *et al.* ApoE polymorphism and predisposition to coronary heart disease in youths of different European populations. The EARS Study. European Atherosclerosis Research Study. *Arteriosclerosis and thrombosis: a journal of vascular biology* **14**, 1617–1624 (1994).
47. Humphries, S. E. *et al.* Apolipoprotein E4 and coronary heart disease in middle-aged men who smoke: a prospective study. *Lancet* **358**, 115–119 (2001).
48. Lin, H. *et al.* Gene expression and genetic variation in human atria. *Heart Rhythm* **11**, 266–271 (2014).
49. Nica, A. C. & Dermitzakis, E. T. Expression quantitative trait loci: present and future. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences* **368**, 20120362 (2013).
50. Greene, C. S. *et al.* Understanding multicellular function and disease with human tissue-specific networks. *Nat Genet* **47**, 569–576 (2015).
51. Song, A. *et al.* Network-based analysis of genetic variants associated with hippocampal volume in Alzheimer's disease: a study of ADNI cohorts. *BioData mining* **9**, 3 (2016).
52. Goumans, M. J., Liu, Z. & ten Dijke, P. TGF-beta signaling in vascular biology and dysfunction. *Cell research* **19**, 116–127 (2009).
53. Claussnitzer, M. *et al.* FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* **373**, 895–907 (2015).
54. ENCODE Project, C. *et al.* An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74 (2012).
55. Chadwick, L. H. The NIH Roadmap Epigenomics Program data resource. *Epigenomics* **4**, 317–324 (2012).
56. Hermjakob, H. *et al.* IntAct: an open source molecular interaction database. *Nucleic Acids Res* **32**, D452–455 (2004).
57. Zanzoni, A. *et al.* MINT: a Molecular INTeraction database. *FEBS Lett* **513**, 135–140 (2002).
58. Cowley, M. J. *et al.* PINA v2.0: mining interactome modules. *Nucleic Acids Res* **40**, D862–865 (2012).
59. Razick, S., Magkharas, G. & Donaldson, I. M. iRefIndex: a consolidated protein interaction database with provenance. *BMC Bioinformatics* **9**, 405 (2008).
60. Jia, P., Zheng, S., Long, J., Zheng, W. & Zhao, Z. dmGWAS: dense module searching for genome-wide association studies in protein-protein interaction networks. *Bioinformatics* **27**, 95–102 (2011).
61. Ideker, T., Ozier, O., Schwikowski, B. & Siegel, A. F. Discovering regulatory and signalling circuits in molecular interaction networks. *Bioinformatics* **18**(Suppl 1), S233–240 (2002).

Acknowledgements

This project was supported by the National Natural Science Foundation of China (Grant No. 81673977). It was also supported by the Boston University Digital Health Initiative, and the National Center for Advancing Translational Sciences, National Institutes of Health, through BU-CTSI Grant Number 1UL1TR001430. The views expressed in this manuscript are those of the authors and do not necessarily represent the views of the National Institutes of Health or the U.S. Department of Health and Human Services. Data on coronary artery disease/myocardial infarction have been contributed by CARDIoGRAMplusC4D investigators and have been downloaded from www.cardiogramplusc4d.org.

Author Contributions

X.M. and H.L. initiated the study and drafted the manuscript. Z.X. and H.L. performed the analyses. X.C. and Z.X. critically reviewed the manuscript. All authors approve the final version of the manuscript.

Additional Information

Supplementary information accompanies this paper at <https://doi.org/10.1038/s41598-018-29904-7>.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018