RSC Advances

ROYAL SOCIETY OF CHEMISTRY

www.rsc.org/advances

Discovery and Characterization of Functional Modules and Pathogenic Genes Associated with the Risk of Coronary Artery Disease

Wennan Nie[1#], Yana Lv[2, 3#], Tianbing Guan[1#], Leyu Yan[2#], Wenhu Liu[1], Qiang Li[1], Xiaojuan Guo[1], Mengpei Feng[1], Guang Xu[1], Xi Chen[2, 3, 4]* and Haitao Lv[1, 5]*

[1]Chongqing University Innovative Drug Research Centre, School of Chemistry and Chemical Engineering, Chongqing 401331, P.R. China

[2]Yunnan Branch, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Jinghong 666100, China

[3]Key Laboratory of Dai and Southern Medicine of Xishuangbanna Dai Autonomous Prefecture, Jinghong 666100, China

[4]Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences, Peking Union Medical College, Beijing 100193, China

[5]Tissue Repair and Regeneration Program, Institute of Health and Biomedical Innovation, Queensland University of Technology, Brisbane, QLD 4059, Australia

#these authors contributed equally to this manuscript.

Corresponding authors:

Haitao Lv, Ph.D.

E-mail: haitao.lu@cqu.edu.cn

Phone: 61-731386189/86-2365678464

Fax: 86-2365678450

Xi Chen, Ph.D.

E-mail: chenxi@implad.ac.cn

Abstract

Coronary artery disease (CAD) involves a complex interplay between multiple pathogenic genes that leads to a complex pathogenesis; its diagnosis and treatment remain significantly challenging. Here, we developed an integrated network biology approach to identify disease risk functional modules and pathogenic genes associated with CAD risk. First, we selected 72 known disease genes from the OMIM, GAD, and DO databases as an initial set of seed genes. We retrieved PPI data from HPRD to expand this gene set into a CAD-PPI gene network based on direct interactions and then performed topology analysis for this CAD-PPI gene network. Second, we utilized an MCL algorithm to identify 49 susceptible modules with high modularity. Third, we used functional consistency analysis to further identify 23 risk functional modules. Finally, according to existing cascades of known disease genes in KEGG pathways, we identified 82 pathogenic genes that are either directly or indirectly associated with CAD risk. Based on previous reports, 37 of our identified genes are involved in the development of CAD, whereas the other 45 genes remain to be associated with CAD by experimental evidence. Taken together, our results will provide a better understanding of CAD pathogenesis as well as new insights into its prognosis and treatment.

Key words: coronary artery disease, network biology approach, risk disease module, risk pathogenic gene

1 Introduction

Coronary artery disease (CAD), an important cardiovascular disease, involves a complex interplay between multiple pathogenic genes [1-3]. CAD development has a strong genetic component and has become a leading cause of morbidity and mortality worldwide [4, 5]. According to the 2013 Statistical Update published by the American Heart Association [6, 7], approximately one in every six people die of CAD at an estimated cost of US $180 billion each year. Although the CAD mortality rate has gradually decreased in many developed countries, it will remain quite a public health challenge in numerous industrial countries and the developing world in the near future [8, 9]. Therefore, identifying and characterizing CAD risk genes and their biological functions currently remains very vital and will aid the study of CAD pathogenesis. Previously, Li et al. [10] employed integrated network analysis to uncover comprehensive insights into cardiovascular disease (CVD) pathogenesis and to discover novel drug targets. Nair et al. [11] constructed a gene network based on integrated leukotrienes and inflammatory biomarkers to identify leukotriene-induced inflammatory molecules and the expression of important members of the leukotriene pathway in CAD patients and a cohort of healthy controls. In this study, they found that leukotrienes and inflammatory genes (LTA4H and IL-8) are closely related to cardiovascular disease. Moreno-Moral et al. [12] employed transcriptional network analysis to identify and characterize many co-expressed genes that play a key role in human heart disease. Huan et al. [13] built a coexpression network based on whole blood gene expression profiles to identifying a differential module (DM). In their

study, they further integrated tissue-specific Bayesian Networks and Protein-Protein Interaction Networks for their causal DM to identify key regulatory driver genes. Their results suggested that these key drivers (SPIB, TNFRSF13C and EBF1) are highly related to CAD development.

Biomolecular networks and Protein and Protein Interaction (PPI) networks have be used to screen candidate genes, predict gene function, and identify molecular markers of disease [14-16]. In PPI networks, direct interaction partners tend to share the same or similar functions, and causative genes for some complex diseases are likely to congregate in the same network communities, such as biological modules, protein complexes, pathways or sub-works [17-20]. For example, Oti et al. [21] and Chen et al. [22] successfully predicted novel genetic heterogeneity of disease causative genes by adopting direct interactive relationships with known disease genes in a protein-protein interaction network. In addition, some graph theoretical analyses have succeeded in mining network modules [23]. The MCL algorithm, a fast and scalable unsupervised cluster algorithm for graphs (also known as networks) based on the simulation of (stochastic) flow in graphs, is currently one of the best clustering algorithms. MCL only requires an inflation parameter and sets different values corresponding to different clustering results [24]. As an example, James et al. [25] studied protein complexes from binary interaction networks using the MCL algorithm and found that the MCL algorithm is superior to other cluster algorithms. Therefore, identification of disease modules based on PPI network is an important biological question that can be addressed by the MCL algorithm.

According to the degree of functional consistency and intensity of the interactive correlation, we proposed a joint strategy to identify functional modules and pathogenic genes associated with CAD risk. First, we selected 72 known disease genes as the initial set of seed genes and employed PPI knowledge to expand these genes into a CAD-PPI gene network based on genes with the same or similar functions as the known disease genes. Subsequently, our topology analysis demonstrated that the CAD-PPI gene network possesses the well-known small-world and scale-free properties. Second, we applied the MCL method to identify susceptible modules that include at least one known disease gene. Third, we screened a total of 23 risk functional modules and identified 82 risk pathogenic genes that were either directly or indirectly associated with CAD development based on this joint strategy. Finally, we found that 37 risk pathogenic genes are associated with CAD based on bioinformatics retrieval analyses.

2 Materials and Methods

In this study, we implemented a joint strategy based on interaction correlations and functional consistency to identify functional modules and pathogenic genes associated with CAD risk. The detailed protocol of this strategy is shown in **Figure 1**.

Global identification of known CAD-associated genes

OMIM (Online Mendelian Inheritance in Man) (http://www.ncbi.nlm.nih.gov/omim) is an online and open-source database that incorporates a comprehensive, authoritative compendium of human genes and genetic phenotypes [26]. In addition,

GAD (Genetic Association Database) (http://geneticassociationdb.nih.gov/) is a collective archive of human genetic association studies of complex diseases and disorders[27], and DO (disease ontology) (http://www.disease-ontology.org/) is another comprehensive knowledge database of inherited, developmental and acquired human diseases [28]. Through high-throughput data mining, 72 known disease genes were captured from these three databases and defined as the initial set of seed genes for our analysis.

3 Systems Construction of CAD-PPI gene network

HPRD (Human Protein Reference Database) (http://www.hprd.org/) is a rich resource of experiments that confirm human protein functions. This resource provides information on human protein functions including protein–protein interactions, post-translational modifications, enzyme-substrate relationships and disease associations [29]. We directly collected protein–protein interaction (PPI) data from HPRD (HPRD_Release9). This dataset consisted of 9453 proteins and 36,867 pairs of confirmed interactions excluding intra-molecular protein interactions and self oligomerization.

The analysis steps were as follows. First, the initial seed genes map was subjected to the PPI network in an effort to discover directly interacting genes. Second, we constructed a CAD-PPI gene network by actually mapping the known disease genes and their direct interaction partners. Finally, to assess the availability of the CAD-PPI gene network, we manipulated the topological analysis of the CAD-PPI

gene network.

4 Identification of risk functional modules

Based on the functional consistency of identifying risk functional modules, we first utilized the MCL clustering algorithm to mine disease modules. MCL is a fast and expandable unsupervised clustering algorithm. This algorithm simulates random walks within a graph through repeat implementation of expansion and inflation [30]. In this study, susceptible modules are defined as those with ≥3 genes ≥1 known disease gene. Then, we carried out pathway enrichment analysis targeting susceptible modules using the hypergeometric distribution test. Furthermore, we screened the risk functional modules with significant functional consistency ($p<0.01$) between the susceptible modules and the known disease genes.

5 Identification of risk pathogenic genes

Using the CAD-PPI gene network, we employed our joint strategy based on the degree of functional consistency and intensity of interaction correlations to screen for and identify risk pathogenic genes.

Functional consistency: We performed significant pathway enrichment analysis for susceptible modules using the hypergeometric distribution test to screen the candidate gene set, note $S_{K|p_k}$.

$$S_{K|p_k} = \bigcap_{i=1}^{n} \bigcap_{j=1}^{m} S_{K|p_{ij}}$$

$$p_{ij}(k|N, M, n) = \frac{\binom{M}{k}\binom{N-M}{n-k}}{\binom{N}{n}}$$

where i denotes a module, n denotes the number of modules, m denotes the number of pathways, j denotes a pathway, N denotes the whole human genome, n is the number of genes in a module, M is the number of genes in a pathway, and k is the number of overlapping genes in M and n. Pathways with $p<0.01$ were regarded as significantly enriched.

Interaction correlation: We screened for the risk pathogenic genes from among candidate genes with direct interactions with known disease genes in our CAD-PPI gene network and with relationships that exist within KEGG pathways. Let S be a set that contains all risk pathogenic genes:

$$S = \begin{cases} \{g_{c_i} | i = 1,2,3 \cdots n\}, & \langle g_{c_i}, g_{b_j} \rangle \exists PPI \ and \ \langle g_{c_i}, g_{b_j} \rangle \exists KEGG \\ \emptyset, & \langle g_{c_i}, g_{b_j} \rangle \exists PPI \ and \ \langle g_{c_i}, g_{b_j} \rangle \nexists KEGG \end{cases}$$

where g_c is the candidate gene, g_b is a known disease gene, $\langle g_{c_i}, g_{b_j} \rangle$ denotes the direct interaction correlation, and $j = 1,2,3 \cdots n$.

Results and Discussion

Construction of gene networks for CAD

By utilizing the PPI knowledge regarding seed genes and their direct relationships in

HPRD, we mapped 72 initial seed genes into a CAD-PPI gene network with 811 gene nodes and 1079 interactions **(Figure 2).** To estimate the availability of the CAD-PPI gene network, we performed topological analysis and illustrated that the CAD-PPI gene network has well-known small-world and scale-free features. As shown in **Figure 3**, the topological features of the CAD-PPI gene network were visually demonstrated.

Identification of functional modules associated with CAD risk. The MCL algorithm was selectively utilized to mine the risk functional modules; we identified a total of 49 susceptible modules that relied on the known disease genes. We further performed pathway analysis using a hypergeometric distribution test to target these susceptible modules; 23 risk functional modules were discovered and identified ($p<0.01$) **(Figure 4).** These modules correlated remarkably with many biological processes such as the Wnt signalling pathway, complement and coagulation cascades, the Jak-STAT signalling pathway, cytokine-cytokine receptor interactions, the renin-angiotensin system, Type II diabetes mellitus, the MAPK signalling pathway, etc. Basically, we characterized the pathways which were either directly or indirectly associated with CAD development. For example, M14 was enriched in factors involved in Wnt signalling. The Wnt signalling pathway encompasses a series of proteins that are required for developmental processes such as cell fate specification [31-33]. In addition, defects in Wnt signalling were found in a family with autosomal dominant early coronary artery disease, and the developmentally important Wnt

pathway has also been associated with the early stages of coronary artery disease [34, 35]. M11, M30, M31, M36 and M41 were enriched in the complement and coagulation cascades. Complement activation affects many biochemical processes including the development and progression of atherosclerosis, and it also influences thrombosis development [36]. TF produced by macrophages and smooth muscle cells is capable of activating the coagulation cascade, and thrombosis and coagulation can be stimulated by disrupted plaque [37, 38]. Similarly, M4, M34, M37, and M43 were enriched in cytokine-cytokine receptor interaction pathways as well, which may be involved in the inflammatory processes of CAD progression. Similarly, Rull et al. [39] reported that arteriosclerosis lesions and plaque composition are related to enrichment in the cytokine–cytokine receptor interaction pathways. In addition, the data collected by Liu et al. [40] and Zhang et al. [41] also suggested that cytokine–cytokine receptor interaction pathways play important roles in CAD development.

Identification of pathogenic genes associated with CAD risk

We screened a total of 82 risk pathogenic genes based on our joint strategy, and 37 risk genes were identified that are associated with the pathogenesis of CAD based on bioinformatics analyses. Another 45 risk pathogenic genes were discovered that might be related to CAD development but require more experimental evidence to confirm this relationship (Table 1).

Table 1. Pathogenic genes and enriched pathways of the identified CAD-

associated modules.

Module	pathway	Pathogenic genes
M3	Type II diabetes mellitus Neurotrophin signalling pathway Insulin signalling pathway	INSR, PIK3R2, PIK3R3, IKBKB, MAPK8, GRB2, MAPK9, PTPN, SOCS1, PTPR, MTOR, PRKCD, RPTOR
M4	Chemokine signalling pathway Cytokine-cytokine receptor interaction Toxoplasmosis	CCL11, CCL13, CCL14, CCL16, CCL3, CCL3L1, CCL3, CCL4, CCL5, CCL7, CCL8, GNAI1 CXCR4, GNAI3, JAK1, JAK2
M9	African trypanosomiasis	HPR, APOL1
M10	PPAR signalling pathway	RXRB, RXRG
M11	Complement and coagulation cascades	PLAT, PLAU,
M12	Renin-angiotensin system	CMA1, CTSG, ENPEP, MAS1, ACE2
M14	Wnt signalling pathway	FZD5, FZD8, WNT1, WNT3A, DKK1, DKK2
M18	Cell adhesion molecules	SELL, GLG1
M22	Apoptosis Osteoclast differentiation	IL1A, MYD88, IRAK2, IL1RAP
M25	Leukocyte transendothelial migration Staphylococcus aureus infection	EZR, MSN, SPN
M26	Cell adhesion molecules	SELPLG
M28	Regulation of actin cytoskeleton Calcium signalling pathway	GNA12, GNA13, GAN15
M30	Complement and coagulation cascades	PROC, PROS1
M31	Complement and coagulation cascades Staphylococcus aureus infection	MASP1 MASP2
M32	Endocytosis	LDLRAP1
M34	NOD-like receptor signalling pathway Amoebiasis Cytokine-cytokine receptor interaction MAPK signalling pathway	CASP1, IL1R2
M35	Leukocyte transendothelial migration Osteoclast differentiation Phagosome Leishmaniasis	RAC1, CYBB, NCF1, NCF2, NCF4, NOX1
M36	Complement and coagulation cascades	KLKB1
M37	Chemokine signalling pathway Cytokine-cytokine receptor interaction	CCR1 CCR10
M41	Complement and coagulation cascades	F2
M43	Jak-STAT signalling pathway Cytokine-cytokine receptor interaction	IL6R IL6ST
M46	Renin-angiotensin system Chagas disease	AGTR2 BDKRB2
M47	Salivary secretion Calcium signalling pathway	ADRB2 GNAS

Note: Red genes have reported evidence of a CAD association

As shown in **Figure 5**, the genes CCL3, CCL4, CCL5, CCL7, CCL8, CCL11, CCL13, CCL14, CCL16, CCL3L1, CCL3L3, CCR2, CCR5, CXCR4, GNAI1, GNAI3, JAK1, and JAK2 were contained in M4; of these genes, CCR2 and CCR5 are

known disease genes for CAD. We identified CCL3, CCL4, CCL5, CCL7, CCL8, CCL11, CCL13, CCL14, CCL16, CCL3L1, and CCL3L3, which not only belong to the CC chemokine family, but also participated in the chemokine signalling pathway and cytokine-cytokine receptor interaction pathways **(Figure 3).** Basically, cytokines play remarkable roles as regulators of inflammatory activity, in the development of atherosclerotic plaques and in the process of plaque destabilization [42]. The chemokine signalling pathway also plays a vital role in lesion development in atherosclerosis [43, 44].

The chemokines CCL3, CCL4, and CCL5 have been reported to play a positive role in the formation of atherosclerotic plaques [45-47]. CCL7, also known as MCP-3, has been shown by Maddaluno et al. [48] and Schenk et al. [49] to play a key role in the development of atherosclerosis and restenosis as well as other vascular pathologies. CCL14, CCL, CCL3L1, and CCL3L3 are likely to be linked to CAD; however, these relationships need further validation by biochemical experiments.

The CCL2, CCR1, and CCR10 genes were incorporated into M37, and CCL2 is the known disease gene. CCR1 and CCR10 are members of the chemokine receptor family and are also part of the chemokine signalling pathway and cytokine-cytokine receptor interaction pathways **(Figure 3).** Chemokine receptors play an important role in the inflammatory response that is associated with CAD development, suggesting their involvement in inflammation during CAD development. The chemokine receptor CCR1 is bound to chemokines present in arterial plaques [50]. Cha et al. [51] also verified that CCR1 is associated with the pathogenesis of CAD. Additionally, CCR10

may be a novel risk gene associated with CAD, but this relationship requires verification by future experiments.

The identified genes shown in **Figure 6**, CYBB, RAC1, NCF1, NCF2, NCF4, and NOX1, were mapped in M35, and CYBB was confirmed to be the known disease gene. NCF1, NCF2, NCF4, and NOX1 are members of the NADPH (nicotinamide adenine dinucleotide phosphate) family, which are enriched in the phagosome pathway **(Figure 3).** NADPH oxidases are important sources of superoxide in the vasculature, and different NADPH oxidase isoforms have potential contributions to vascular diseases [52]. Jiang et al. [53] found that Nox1 inhibitors have clinical significance in the treatment of cardiovascular disease. The remaining genes, NCF1, NCF2, and NCF4, were discovered to correlate with CAD in this study, but this association requires further confirmation by experimental evidence.

Conclusions

Coronary artery disease is a complex disease that is often triggered by the combinational effects of multiple genes [54]. In this study, we explored and exploited a joint strategy based on interaction correlations and functional consistency to identify function modules and pathogenic genes associated with the risk of CAD development.

First, based on genes that share the same or similar functions with known disease genes, we built a CAD-PPI gene network using PPI data and data on known disease genes. Because the CAD-PPI gene network has a reliable biological foundation, we performed topology analysis on our CAD-PPI gene network, which rendered well-

known small-world and scale-free features.

Then, in the context of this integrated CAD-PPI gene network, we employed the MCL algorithm to mine susceptible modules. Finally, we identified a total of 23 risk function modules and 82 risk pathogenic genes. The modules were remarkably enriched in the Wnt signalling pathway, complement and coagulation cascades, the Jak-STAT signalling pathway, cytokine-cytokine receptor interactions, and the renin-angiotensin system. These pathways provide novel insights into the pathogenesis of CAD. To implement further bioinformatics analyses, we identified 35 risk pathogenic genes that are associated with the pathogenesis of CAD and 45 risk pathogenic genes that may be involved in the development of CAD, but which need further confirmation through experimental evidence **(Figure 7).**

In this study, our strategy had two advantages. We built the CAD-PPI gene network based on genes that share the same or similar functions with known disease genes. Simultaneously, the relationships between genes were based on credible data within PPI databases that are documented by expert biologists. However, the proposed strategy was also deficient: first, there is a time lag for updating the data in these databases (HPRD, OMIM, GAD, and DO), and second, our PPI network was not large enough to be thoroughly optimized for the prediction and understanding of CAD.

Altogether, despite the few shortcomings of this study, we are still confident in our findings, as we successfully identified pathogenic genes and modules associated with the risk of CAD development that may be redefined as pathogenic genes for the

diagnosis and treatment of CAD. Meanwhile, the joint strategy we developed can be further explored and exploited to gain a better understanding of the pathogenesis of different cardiovascular diseases and other health issues.

Acknowledgements

This study was funded by a National Natural Science Foundation of China Grant (81274175), the Startup Funding for the "Hundred Young-Talent Scheme" Professorship provided by the Chongqing University in China (0236011104401), the Queensland University of Technology Vice Chancellor's Research Fellowship Grant (150410-0070/08), the Open Grant for Key Instrumental Platform Usage Supported by the Chongqing University in China (2013121564, 201406150008 and 201412150114), a Natural Science Foundation of Chongqing Grant (CSTC2014JCYIA10109), the National Natural Science Foundation of China (81173645), Significant New Drugs Creation Special Science and Technology Major of China (2014ZX09201-022-06) and the Program for Innovative Research Team in IMPLAD (PIRTI), as well as Director's Funds provided by Yunnan Branch, Institute of Medicinal Plant Development, Chinese Academy of Medical Sciences and Peking Union Medical College.

The authors declare no conflicts of interest.

References

1. Zhou Y, Liu T, Jia C: **Joint effects of serum ferritin and body mass index on the risk of coronary artery disease: a case-control study**. *BMJ open* 2013, **3**(11):e003695.
2. Adi D, Xie X, Ma YT, Fu ZY, Yang YN, Li XM, Xiang Y, Liu F, Chen BD: **Association of COL4A1 genetic polymorphisms with coronary artery disease in Uygur population in Xinjiang, China**. *Lipids Health Dis* 2013, **12**:153.
3. Zheng YY, Xie X, Ma YT, Yang YN, Fu ZY, Li XM, Ma X, Chen BD, Liu F: **A novel polymorphism (901G > a) of C5L2 gene is associated with coronary artery disease in Chinese Han and Uyghur population**. *Lipids in health and disease* 2013, **12**:139.
4. Asifa GZ, Liaquat A, Murtaza I, Kazmi SA, Javed Q: **Tumor necrosis factor-alpha gene promoter region polymorphism and the risk of coronary heart disease**. *TheScientificWorldJournal* 2013, **2013**:203492.
5. Zhou W, Wang Y: **A network-based analysis of the types of coronary artery disease from traditional Chinese medicine perspective: potential for therapeutics and drug discovery**. *Journal of ethnopharmacology* 2014, **151**(1):66-77.
6. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, Bravata DM, SHIFAN D, Ford ES, Fox CS: **Executive Summary: Heart Disease and Stroke Statistics: 2013 Update: A Report From the American Heart Association**. *Circulation* 2013, **127**(1):143-146.
7. Go AS, Mozaffarian D, Roger VL, Benjamin EJ, Berry JD, Borden WB, Bravata DM, Dai S, Ford ES, Fox CS: **Heart disease and stroke statistics--2013 update: a report from the American Heart Association**. *Circulation* 2013, **127**(1):e6.
8. Heidenreich PA, Trogdon JG, Khavjou OA, Butler J, Dracup K, Ezekowitz MD, Finkelstein EA, Hong Y, Johnston SC, Khera 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):933-944.
9. Rambiharilal Shrivastava S, Saurabh Shrivastava P, Ramasamy J: **Coronary heart disease: pandemic in a true sense**. *Journal of cardiovascular and thoracic research* 2013, **5**(3):125-126.
10. Li P, Fu Y, Ru J, Huang C, Du J, Zheng C, Chen X, Li P, Lu A, Yang L *et al*: **Insights from systems pharmacology into cardiovascular drug discovery and therapy**. *BMC systems biology* 2014, **8**(1):3.
11. Nair J, Shanker J, Jambunathan S, Arvind P, Kakkar VV: **Expression analysis of leukotriene-inflammatory gene interaction network in patients with coronary artery disease**. *Journal of atherosclerosis and thrombosis* 2014, **21**(4):329-345.
12. Moreno-Moral A, Mancini M, D'Amati G, Camici P, Petretto E: **Transcriptional network analysis for the regulation of left ventricular hypertrophy and microvascular remodeling**. *Journal of cardiovascular translational research* 2013, **6**(6):931-944.
13. Huan T, Zhang B, Wang Z, Joehanes R, Zhu J, Johnson AD, Ying S, Munson PJ, Raghavachari N, Wang R *et al*: **A systems biology framework identifies molecular underpinnings of coronary heart disease**. *Arteriosclerosis, thrombosis, and vascular biology* 2013, **33**(6):1427-1434.
14. Li BQ, You J, Chen L, Zhang J, Zhang N, Li HP, Huang T, Kong XY, Cai YD: **Identification of lung-cancer-related genes with the shortest path approach in a protein-protein interaction network**. *BioMed research international* 2013, **2013**:267375.
15. Jaromerska S, Praus P, Cho YR: **Distance-wise pathway discovery from protein-protein interaction networks weighted by semantic similarity**. *Journal of bioinformatics and*

computational biology 2014, **12**(1):1450004.

16. Li BQ, Niu B, Chen L, Wei ZJ, Huang T, Jiang M, Lu J, Zheng MY, Kong XY, Cai YD: **Identifying chemicals with potential therapy of HIV based on protein-protein and protein-chemical interaction network**. *PloS one* 2013, **8**(6):e65207.
17. Spirin V, Mirny LA: **Protein complexes and functional modules in molecular networks**. *Proceedings of the National Academy of Sciences of the United States of America* 2003, **100**(21):12123-12128.
18. Barabasi A-L, Oltvai ZN: **Network biology: understanding the cell's functional organization**. *Nature Reviews Genetics* 2004, **5**(2):101-113.
19. Narayanan T, Gersten M, Subramaniam S, Grama A: **Modularity detection in protein-protein interaction networks**. *BMC research notes* 2011, **4**:569.
20. Lee SA, Chan CH, Chen TC, Yang CY, Huang KC, Tsai CH, Lai JM, Wang FS, Kao CY, Huang CY: **POINeT: protein interactome with sub-network analysis and hub prioritization**. *BMC bioinformatics* 2009, **10**:114.
21. Oti M, Snel B, Huynen MA, Brunner HG: **Predicting disease genes using protein-protein interactions**. *Journal of medical genetics* 2006, **43**(8):691-698.
22. Chen JY, Shen C, Sivachenko AY: **Mining Alzheimer disease relevant proteins from integrated protein interactome data**. *Pacific Symposium on Biocomputing Pacific Symposium on Biocomputing* 2006:367-378.
23. Zou J, Zheng MW, Li G, Su ZG: **Advanced systems biology methods in drug discovery and translational biomedicine**. *BioMed research international* 2013, **2013**:742835.
24. Vlasblom J, Wodak SJ: **Markov clustering versus affinity propagation for the partitioning of protein interaction graphs**. *BMC bioinformatics* 2009, **10**:99.
25. Vlasblom J, Wodak SJ: **Markov clustering versus affinity propagation for the partitioning of protein interaction graphs**. *BMC bioinformatics* 2009, **10**(1):99.
26. Hamosh A, Scott AF, Amberger JS, Bocchini CA, McKusick VA: **Online Mendelian Inheritance in Man (OMIM), a knowledgebase of human genes and genetic disorders**. *Nucleic acids research* 2005, **33**(Database issue):D514-517.
27. Becker KG, Barnes KC, Bright TJ, Wang SA: **The genetic association database**. *Nature genetics* 2004, **36**(5):431-432.
28. Schriml LM, Arze C, Nadendla S, Chang YW, Mazaitis M, Felix V, Feng G, Kibbe WA: **Disease Ontology: a backbone for disease semantic integration**. *Nucleic acids research* 2012, **40**(Database issue):D940-946.
29. Keshava Prasad TS, Goel R, Kandasamy K, Keerthikumar S, Kumar S, Mathivanan S, Telikicherla D, Raju R, Shafreen B, Venugopal A *et al*: **Human Protein Reference Database--2009 update**. *Nucleic acids research* 2009, **37**(Database issue):D767-772.
30. Srihari S, Ning K, Leong HW: **Refining Markov Clustering for protein complex prediction by incorporating core-attachment structure**. *Genome informatics International Conference on Genome Informatics* 2009, **23**(1):159-168.
31. Logan CY, Nusse R: **The Wnt signaling pathway in development and disease**. *Annual review of cell and developmental biology* 2004, **20**:781-810.
32. Katoh M, Katoh M: **WNT signaling pathway and stem cell signaling network**. *Clinical cancer research : an official journal of the American Association for Cancer Research* 2007, **13**(14):4042-4045.

33. Staal FJ, Sen JM: **The canonical Wnt signaling pathway plays an important role in lymphopoiesis and hematopoiesis**. *European journal of immunology* 2008, **38**(7):1788-1794.
34. Goliasch G, Wiesbauer F, Kastl SP, Katsaros KM, Blessberger H, Maurer G, Schillinger M, Huber K, Wojta J, Speidl WS: **Premature myocardial infarction is associated with low serum levels of Wnt-1**. *Atherosclerosis* 2012, **222**(1):251-256.
35. Mani A, Radhakrishnan J, Wang H, Mani A, Mani M-A, Nelson-Williams C, Carew KS, Mane S, Najmabadi H, Wu D: **LRP6 mutation in a family with early coronary disease and metabolic risk factors**. *Science* 2007, **315**(5816):1278-1282.
36. Carter AM: **Complement Activation: An Emerging Player in the Pathogenesis of Cardiovascular Disease**. *Scientifica* 2012, **2012**.
37. Libby P, Theroux P: **Pathophysiology of coronary artery disease**. *Circulation* 2005, **111**(25):3481-3488.
38. Toschi V, Gallo R, Lettino M, Fallon JT, Gertz SD, Ferna A, Chesebro JH, Badimon L, Nemerson Y, Fuster V: **Tissue factor modulates the thrombogenicity of human atherosclerotic plaques**. *Circulation* 1997, **95**(3):594-599.
39. Rull A, Beltrán-Debón R, Aragonès G, Rodríguez-Sanabria F, Alonso-Villaverde C, Camps J, Joven J: **Expression of cytokine genes in the aorta is altered by the deficiency in MCP-1: effect of a high-fat, high-cholesterol diet**. *Cytokine* 2010, **50**(2):121-128.
40. Liu H, Liu W, Liao Y, Cheng L, Liu Q, Ren X, Shi L, Tu X, Wang QK, Guo AY: **CADgene: a comprehensive database for coronary artery disease genes**. *Nucleic acids research* 2011, **39**(Database issue):D991-996.
41. Zhang L, Li X, Tai J, Li W, Chen L: **Predicting candidate genes based on combined network topological features: a case study in coronary artery disease**. *PloS one* 2012, **7**(6):e39542.
42. Konenkov V, Shevchenko A, Prokof'ev V, Maksimov V: **[Complex of genotypes of cytokines as a genetic factor of risk of development of myocardial infarction of in Europien population of Russia men]**. *Kardiologiia* 2011, **52**(7):22-29.
43. Moatti D, Faure S, Fumeron F, Amara MEW, Seknadji P, McDermott DH, Debré P, Aumont MC, Murphy PM, de Prost D: **Polymorphism in the fractalkine receptor CX3CR1 as a genetic risk factor for coronary artery disease**. *Blood* 2001, **97**(7):1925-1928.
44. Combadière C, Potteaux S, Gao J-L, Esposito B, Casanova S, Lee EJ, Debré P, Tedgui A, Murphy PM, Mallat Z: **Decreased atherosclerotic lesion formation in CX3CR1/apolipoprotein E double knockout mice**. *Circulation* 2003, **107**(7):1009-1016.
45. Charo IF, Ransohoff RM: **The many roles of chemokines and chemokine receptors in inflammation**. *The New England journal of medicine* 2006, **354**(6):610-621.
46. Tereshchenko IP, Petrkova J, Voevoda MI, Taborsky M, Navratilova Z, Romaschenko AG, Maksimov VN, Mrazek F, Petrek M: **CCL5/RANTES gene polymorphisms in Slavonic patients with myocardial infarction**. *Mediators of inflammation* 2011, **2011**:525691.
47. Montecucco F, Braunersreuther V, Lenglet S, Delattre BM, Pelli G, Buatois V, Guilhot F, Galan K, Vuilleumier N, Ferlin W *et al*: **CC chemokine CCL5 plays a central role impacting infarct size and post-infarction heart failure in mice**. *European heart journal* 2012, **33**(15):1964-1974.
48. Maddaluno M, Di Lauro M, Di Pascale A, Santamaria R, Guglielmotti A, Grassia G, Ialenti A: **Monocyte chemotactic protein-3 induces human coronary smooth muscle cell proliferation**. *Atherosclerosis* 2011, **217**(1):113-119.

49. Schenk S, Mal N, Finan A, Zhang M, Kiedrowski M, Popovic Z, McCarthy PM, Penn MS: **Monocyte chemotactic protein-3 is a myocardial mesenchymal stem cell homing factor**. *Stem Cells* 2007, **25**(1):245-251.

50. Braunersreuther V, Zernecke A, Arnaud C, Liehn EA, Steffens S, Shagdarsuren E, Bidzhekov K, Burger F, Pelli G, Luckow B *et al*: **Ccr5 but not Ccr1 deficiency reduces development of diet-induced atherosclerosis in mice**. *Arteriosclerosis, thrombosis, and vascular biology* 2007, **27**(2):373-379.

51. Cha SH, Lee JK, Lee JY, Kim HT, Ryu HJ, Han BG, Kim JW, Oh B, Kimm K, Shin HD *et al*: **Association of CCR2 polymorphisms with the number of closed coronary artery vessels in coronary artery disease**. *Clinica chimica acta; international journal of clinical chemistry* 2007, **382**(1-2):129-133.

52. Takac I, Schroder K, Brandes RP: **The Nox family of NADPH oxidases: friend or foe of the vascular system?** *Current hypertension reports* 2012, **14**(1):70-78.

53. Jiang S, Streeter J, Schickling BM, Zimmerman K, Weiss RM, Miller FJ, Jr.: **Nox1 NADPH oxidase is necessary for late but not early myocardial ischaemic preconditioning**. *Cardiovascular research* 2014, **102**(1):79-87.

54. Scantamburlo G, Scheen AJ: **[Role of psychosocial stress in complex diseases]**. *Revue medicale de Liege* 2012, **67**(5-6):234-242.

Figure Legends

Figure 1. Workflow for identifying the functional modules and pathogenic genes

associated with CAD risk.

Figure 2. The CAD-PPI gene network. Red denotes known genes, and green denotes gene that directly interact with the known genes.

Figure 3. The properties of the CAD-PPI gene network. (a) Degree distribution. The connectivity distribution of the proteins obeys the power-law distribution. (b) Clustering coefficient. (c) Shortest path. (d) Between centrality.

Figure 4. The functional relationships between the identified disease risk function modules. Green triangles denote KEGG pathways, and orange ovals denote modules.

Figure 5. Chemokine signalling pathway. Yellow rectangles denote disease risk disease, and red rectangles denote known disease genes.

Figure 6. Phagosome pathway. Yellow rectangles denote disease risk, and red rectangles denote known disease genes.

Figure 7. Disease gene network for CAD risk. Red denotes known disease genes, green denotes confirmed risk disease genes, and grey denotes unconfirmed risk disease genes.

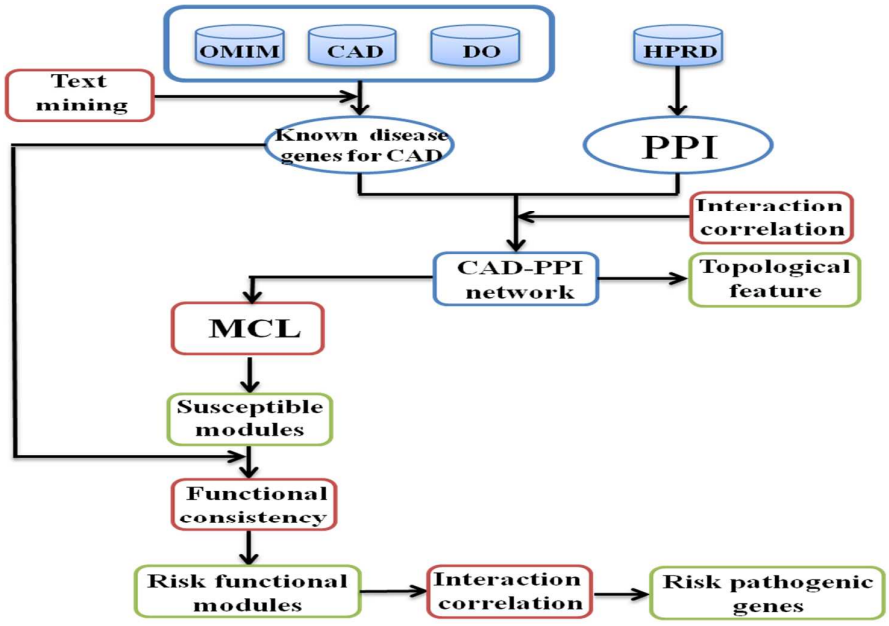

Figure 1

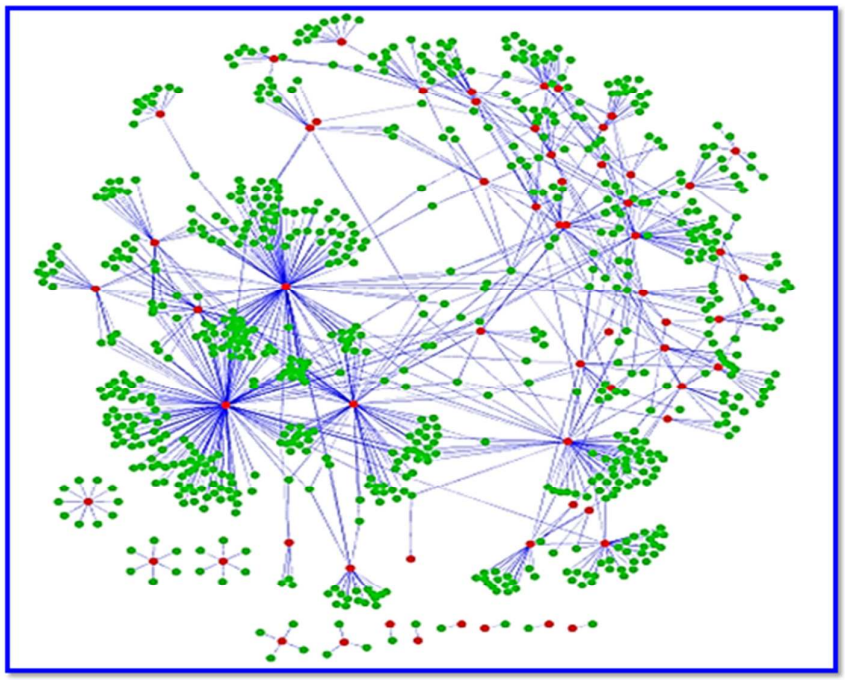

Figure 2

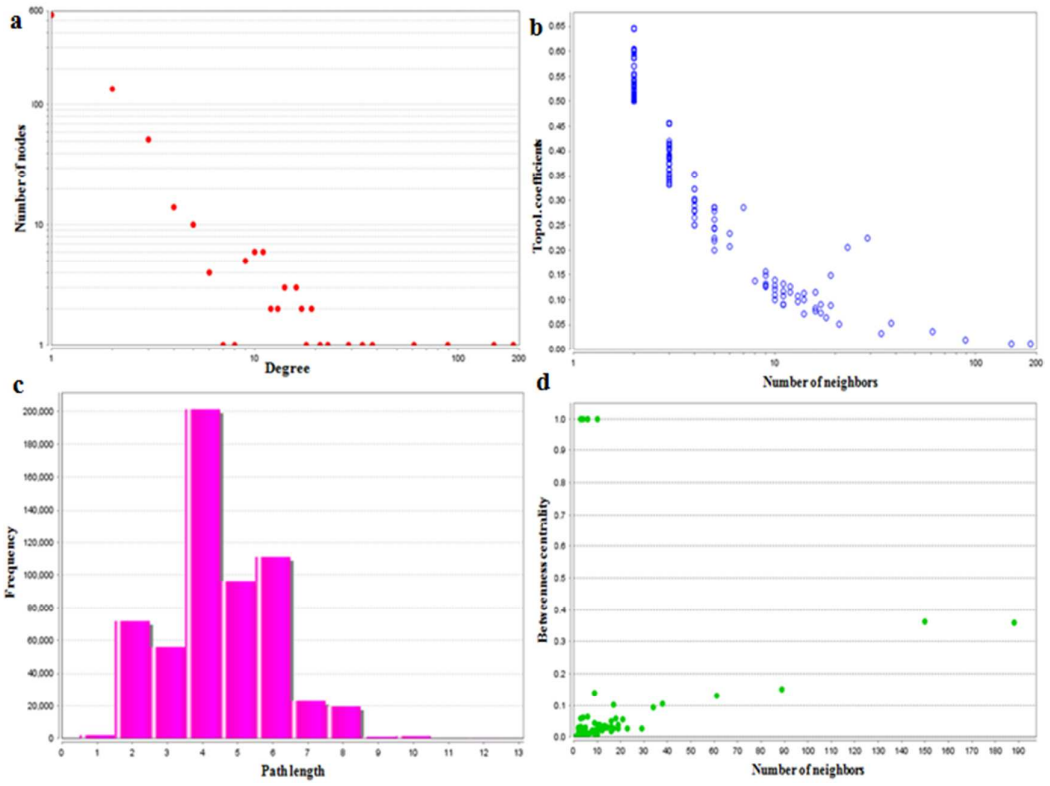

Figure 3

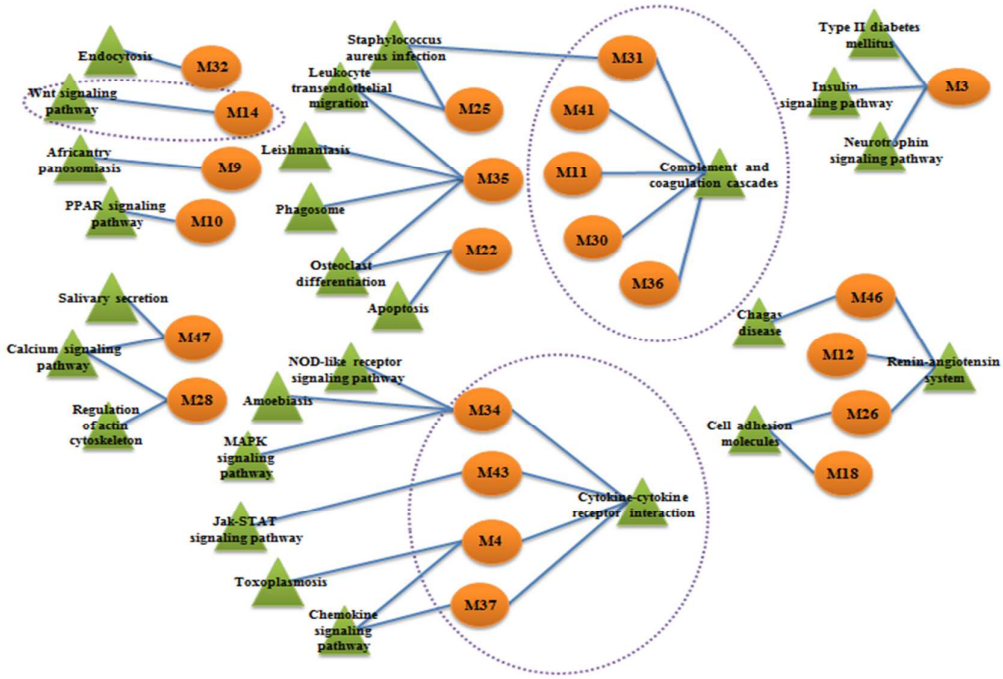

Figure 4

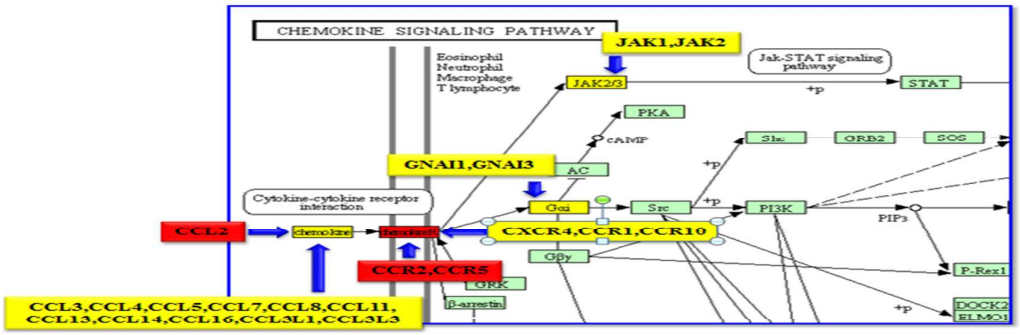

Figure 5

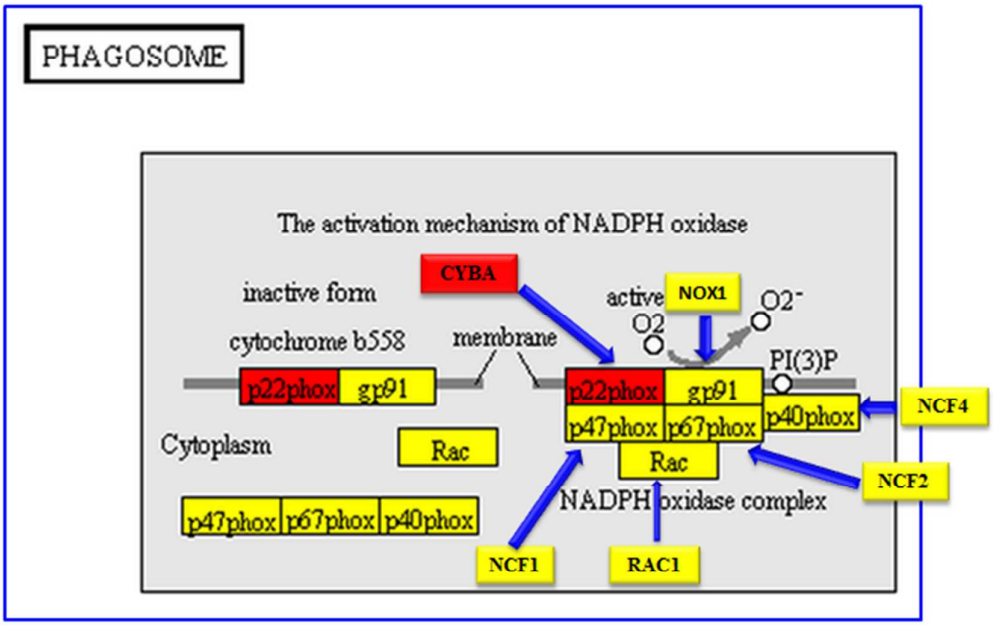

Figure 6

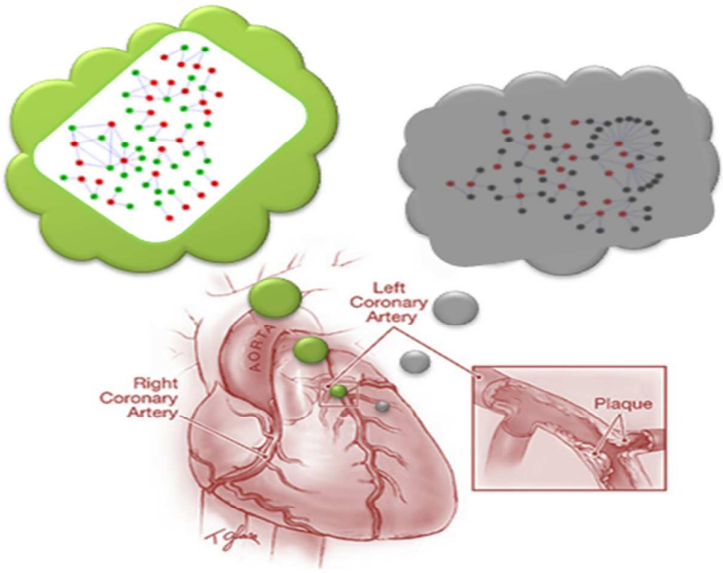

Figure 7