Gene 531 (2013) 347–354

Contents lists available at ScienceDirect

Gene

journal homepage: www.elsevier.com/locate/gene

Identification of susceptibility modules for coronary artery disease using a genome wide integrated network analysis

Shiwei Duan, Xuhong Luo, Changzheng Dong *

Zhejiang Provincial Key Laboratory of Pathophysiology, School of Medicine, Ningbo University, 818 Fenghua Road, Ningbo, Zhejiang 315211, China

ARTICLE INFO

Article history:
Accepted 19 August 2013
Available online 29 August 2013

Keywords:
Coronary artery disease
Genome-wide association
Network analysis
Protein–protein interaction

ABSTRACT

Although recent genome-wide association studies (GWAS) have identified a handful of variants with best significance for coronary artery disease (CAD), it remains a challenge to summarize the underlying biological information from the abundant genotyping data. Here, we propose an integrated network analysis that effectively combines GWAS genotyping dataset, protein–protein interaction (PPI) database, literature and pathway annotation information. This three-step approach was illustrated for a comprehensive network analysis of CAD as the following. First, a network was constructed from PPI database and CAD seed genes mined from the available literatures. Then, susceptibility network modules were captured from the results of gene-based association tests. Finally, susceptibility modules were annotated with potential mechanisms for CAD via the KEGG pathway database. Our network analysis identified four susceptibility modules for CAD including a complex module that consisted of 15 functional inter-connected sub-modules, AGPAT3–AGPAT4–PPAP2B module, ITGA11–ITGB1 module and EMCN–SELL module. MAPK10 and COL4A2 among the top-scored focal adhesion pathway related module were the most significant genes (MAPK10: OR = 32.5, $P = 3.5 \times 10^{-11}$; COL4A2: OR = 2.7, $P = 2.8 \times 10^{-10}$). The significance of the two genes were further validated by other two gene-based association tests (MAPK10: P = 0.009 and 0.007; COL4A2: P = 0.001 and 0.023) and another independent GWAS dataset (MAPK10: P = 0.001; COL4A2: P = 0.0004). Furthermore, 34 out of 44 previously reported CAD susceptibility genes were captured by our CAD PPI network and 17 of them were also significant genes. The susceptibility modules identified in our study might provide novel clues for the clarification of CAD pathogenesis in the future.

1. Introduction

Coronary artery disease (CAD) is a narrowing or blockage of the coronary arteries that supply blood and oxygen to the heart. CAD is a serious disease and has become the leading cause of death in the world (Franco et al., 2011). CAD is a complex disease contributed by environmental factors and genes with minor effect. Candidate gene association tests (Huang et al., 2013; Peng et al., 2012; Zhou et al., 2012) as well as high-throughput genome wide association study (GWAS) (Lusis, 2012) have been performed to reveal the genetic aspects to the CAD pathogenesis, however, majority of the genes are estimated to remain undiscovered (Peden and Farrall, 2011).

GWAS using massive single nucleotide polymorphism (SNP) microarrays have provided an unbiased way to identify novel CAD susceptibility loci. Since the GWAS reported that CDKN2A–CDKN2B locus was associated with CAD (McPherson et al., 2007; Samani et al., 2007; The Wellcome Trust Case Control Consortium, 2007), tens of other GWAS have identified other susceptibility genes and loci for CAD and its risk factors (Lusis, 2012). Recently the CARDIoGRAMplusC4D Consortium published their GWAS results performed in 63,746 CAD cases and 130,681 controls and identified 15 new loci (Deloukas et al., 2013). Current GWAS focuses on single SNP association tests that only the SNPs with the best significance are presented as the final results. Although meta-analysis of GWAS may accumulate genome wide significance for some less significant SNPs in the involved GWAS, majority of SNPs with potential functional role in the disease development are buried under the genome wide significance threshold.

Integrated network analysis is able to detect new genes in which each single SNP conferred a minor disease risk and provided novel insights to clarify the biological mechanisms of complex diseases (Eleftherohorinou et al., 2009; Jia et al., 2011; Peng et al., 2009; Wang et al., 2007). In this study, we extend the current GWAS-based network analysis strategy for CAD and effectively integrate multi-source information, including GWAS, protein–protein interaction (PPI) database, publication libraries and pathway datasets. Our GWAS-based network analysis is efficient and flexible for the identification of biologically meaningful network modules that consist of both strong and moderate risk genes.

Abbreviations: CAD, coronary artery disease; GWAS, genome-wide association studies; HuGE, The Phenopedia component of the online Human Genome Epidemiology; OR, odds ratio; PPI, protein–protein interaction; QC, quality control; SNP, single nucleotide polymorphism; WTCCC, The Wellcome Trust Case Control Consortium.

* Corresponding author. Tel.: +86 574 87609603.
E-mail address: dongchangzheng@nbu.edu.cn (C. Dong).

0378-1119/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2013.08.059

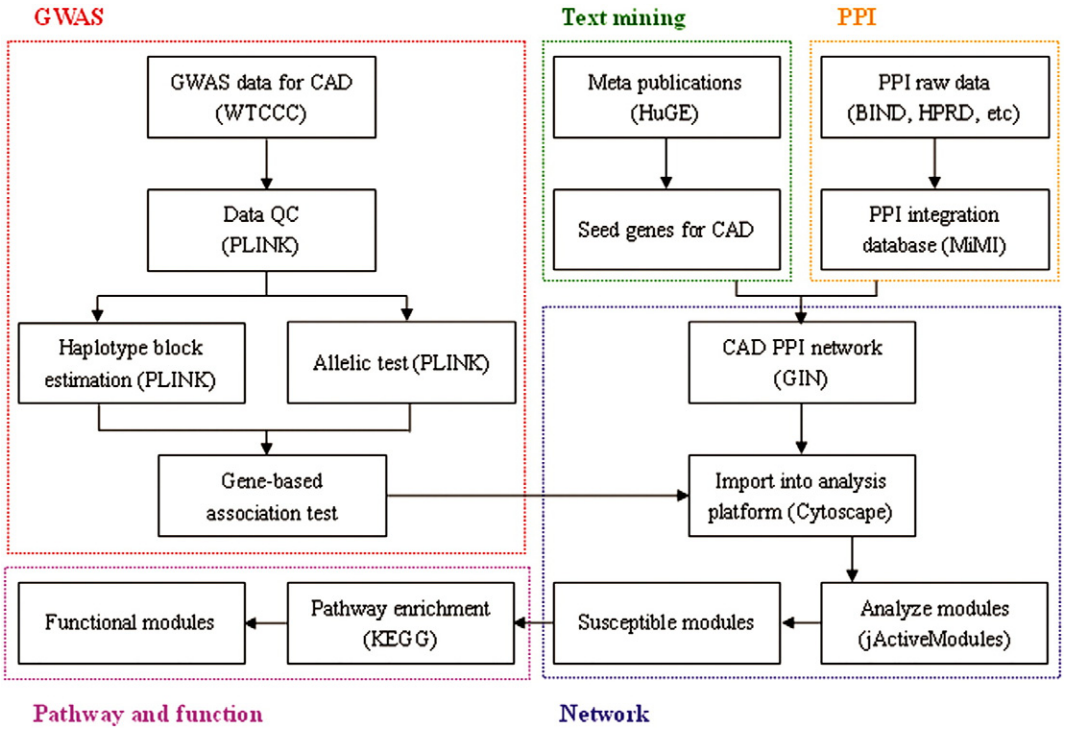

Fig. 1. Workflow for constructing and analyzing the CAD PPI network.

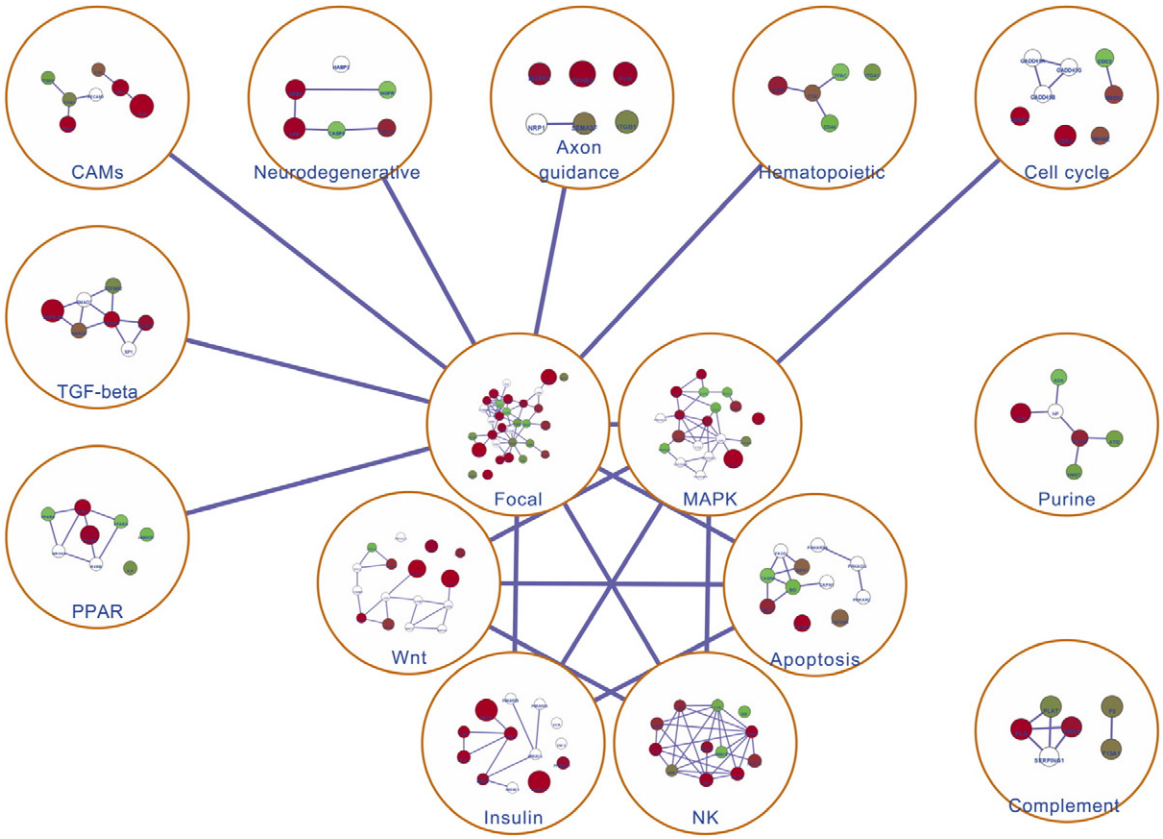

Fig. 2. Fifteen functional modules for the CAD PPI network, Module top. The nested network shows that Module top is organized into fifteen functional modules. Each module is part of a certain KEGG pathway, and the modules are connected by genes that overlap between the modules. Six modules (focal, MAPK, Wnt, insulin, NK and apoptosis) form the core module of Module top, and seven modules (PPAR, TGF-beta, CAMs, neurodegenerative, axon guidance, hematopoietic and cell cycle) connect to the core module via the focal module and the MAPK module. Two other modules (purine and complement) act independently. The node size mapping and node color mapping are detailed in Fig. 3.