Gene 539 (2014) 30–36

Contents lists available at ScienceDirect

Gene

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

Landscape of the relationship between type 2 diabetes and coronary heart disease through an integrated gene network analysis

Changzheng Dong [a,*,1], Linlin Tang [a,1], Zhifang Liu [a,1], Shizhong Bu [a,b], Qiong Liu [a,b], Qinwen Wang [a,b], Yifeng Mai [c], Dao Wen Wang [d,*], Shiwei Duan [a,b,**]

[a] Zhejiang Provincial Key Laboratory of Pathophysiology, School of Medicine, Ningbo University, Ningbo, Zhejiang, China
[b] Diabetes Center, School of Medicine, Ningbo University, Ningbo, Zhejiang, China
[c] The Affiliated Hospital, Ningbo University, Ningbo, Zhejiang, China
[d] Institute of Hypertension and Department of Internal Medicine, Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology, Wuhan, China

ARTICLE INFO

Article history:
Received 17 December 2013
Received in revised form 23 January 2014
Accepted 2 February 2014
Available online 5 February 2014

Keywords:
Type 2 diabetes
Coronary artery disease
Genome-wide association
Network analysis
Protein–protein interaction
Pathway
Hub

ABSTRACT

Type 2 diabetes (T2D) and coronary artery disease (CAD) are closely related chronic diseases with high prevalence and morbidity. However, a comprehensive comparison of the two diseases is lacking. Recent genome-wide association studies (GWAS) have identified a handful of single nucleotide polymorphisms (SNPs) that are significantly associated with the risk of T2D and CAD. These most significant findings may help interpret the pathogenesis of T2D and CAD. However, tremendous results from these GWAS are ignored. Here we revisited the raw datasets of these GWAS and performed an integrated gene network analysis to unveil the relationship between T2D and CAD by combining multiple datasets including protein–protein interaction (PPI) database, publication libraries, and pathway datasets. Our results showed that majority of genes were involved in the first module (1122 genes in T2D and 895 in CAD). Four pathways were found to be common in both T2D and CAD, including regulation of actin cytoskeleton, calcium signaling pathway, MAPK signaling pathway and focal adhesion (all $P < 0.00001$). *MAX* which was involved in small cell lung cancer pathway was a hub gene unique to T2D (OR = 1.2, P = 0.006) but not in CAD. In contrast, three hub genes including *PLEKHG5* (T2D: OR = 1, P = 1; CAD: OR = 1.12, P = 0.006), *TIAM1* (T2D: OR = 1, P = 1; CAD: OR = 1.48, P = 0.004) and *AKAP13* (T2D: OR = 1, P = 1; CAD: OR = 1.38, P = 0.001) were hub genes unique to CAD. Moreover, for some hub genes (such as *SMAD3*) that were susceptible to both T2D and CAD, their associated polymorphisms were unique to each of the two diseases. Our findings might provide a landscape of the relationship between T2D and CAD.

1. Introduction

Type 2 diabetes (T2D) and coronary artery disease (CAD) are chronic diseases with high morbidity and prevalence in the world (Chiha et al., 2012). T2D is a metabolic disease with high glucose in the blood, and CAD is caused by atherosclerosis plaques which narrow the arteries in the heart. T2D and CAD are both complex diseases influenced by genetic and environmental factors (Baccarelli and Ghosh, 2012; Murea et al., 2012). T2D is a risk factor for the development of CAD as the blood glucose is associated with the carotid artery intima-media thickness (Zhang et al., 2013).

Genome-wide association studies (GWAS) have identified a handful of single nucleotide polymorphisms (SNPs) with significant association of T2D or CAD (Dastani et al., 2012; van Setten et al., 2013). Other genetic studies have also revealed similar results (Kaur et al., 2013; Rahimi et al., 2012). For example, −2518A>G polymorphism of *MCP-1* reduces the risk of T2D but increases the risk of CAD in Indians (Kaur et al., 2013). The interaction between T allele of *NOS3* G894T and B1 allele of *CETP* TaqIB has a combined effect on increasing the risk of T2D and CAD in Western Iran (Rahimi et al., 2012). In addition, there are other risk gene polymorphisms acting as important roles to contributing to augment the risk of T2D and CAD, such as G8790A of *ACE2* (Chaoxin et al., 2013), *APOE* E3/E4 (Chaudhary et al., 2012) and *PPARG* rs1801282 (Ho et al., 2012).

Current GWAS apply single-locus approach to indentify the most significant SNPs (Manolio et al., 2009). In these large-scale GWAS, only a few SNPs were highlighted and many other SNPs with less significance were ignored. These omitted SNPs may be of importance to explain the risk of diseases such as T2D and CAD. With the accumulative contribution of free online databases, it is feasible to explore the role of these omitted SNPs in the risk of diseases by the integration of

Abbreviations: T2D, type 2 diabetes; CAD, coronary artery disease; GWAS, genome-wide association studies; SNP, single nucleotide polymorphism; PPI, protein–protein interaction.

* Corresponding authors.
** Correspondence to: S. Duan, Zhejiang Provincial Key Laboratory of Pathophysiology, School of Medicine, Ningbo University, Ningbo, Zhejiang, China.
E-mail addresses: dongchangzheng@nbu.edu.cn (C. Dong), dwwang@tjh.tjmu.edu.cn (D.W. Wang), duanshiwei@nbu.edu.cn (S. Duan).

[1] CD, LT and ZL are co-first authors of this work.

0378-1119/$ – see front matter © 2014 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.gene.2014.02.001

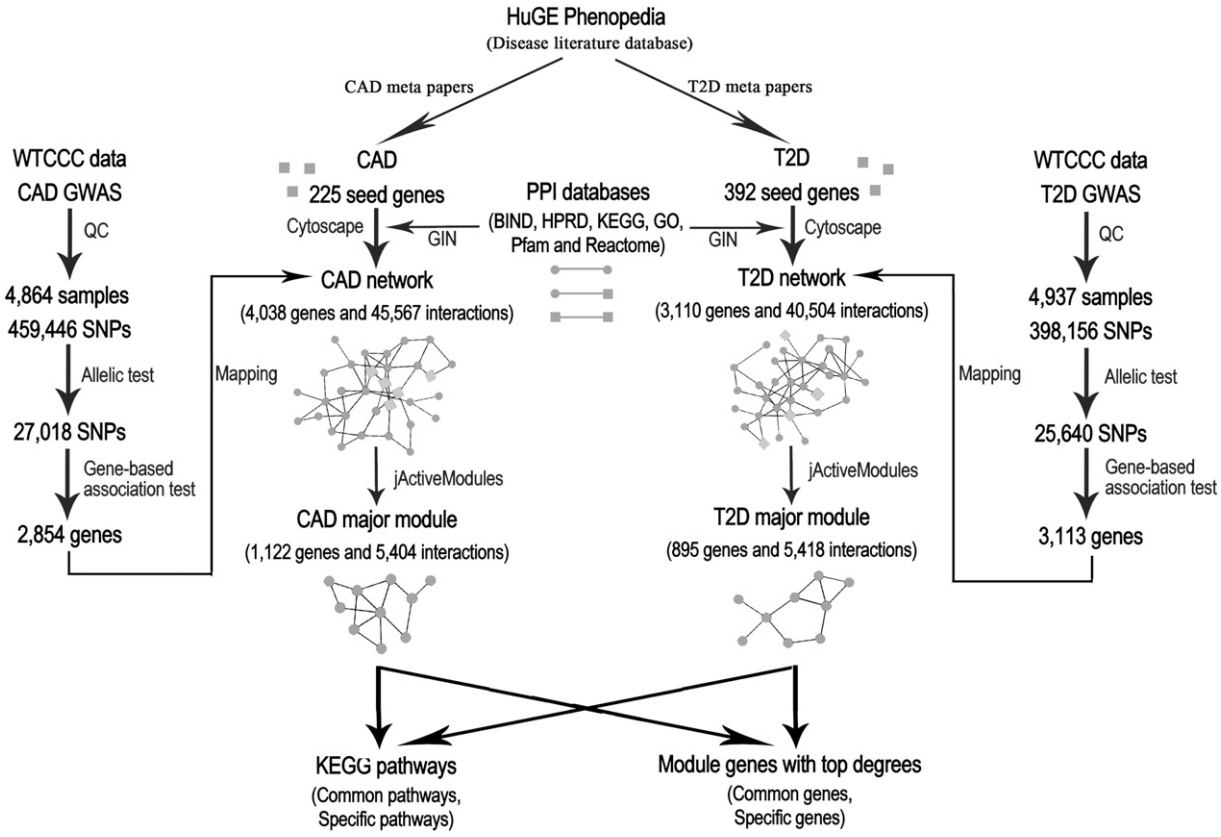

Fig. 1. Flow chart of the integrated network analysis.

resources such as protein–protein interaction (PPI) database, publication libraries and pathway datasets (Fig. 1) (Jia et al., 2011; Wang et al., 2007).

Recently, we have performed an integrated network analysis to identify the susceptible modules for CAD (Duan et al., 2013). The goal of our study is to find out the genetic landscapes of the relationship between CAD and T2D.

2. Methods and materials

2.1. GWAS and gene-based association test

GWAS dataset was granted and retrieved from WTCCC (Wellcome Trust Case Control Consortium, 2007). Data quality control (QC) was performed according to the previous report (Wellcome Trust Case Control Consortium, 2007). In addition, we excluded the SNPs with a maximum individual missing rate of 3%, a maximum SNP missing rate of 5%, a minor allele frequency (MAF) of 1% and a Hardy–Weinberg equilibrium $P < 5.7 \times 10^{-7}$ for controls. Altogether there were 459,446 SNPs among 4864 samples (1926 CAD cases and 2938 controls) for CAD dataset and 398,156 SNPs among 4937 samples (1999 T2D cases and 2938 controls) for T2D dataset after the data QC procedures. An allelic test was used for the association between SNPs and diseases (CAD and T2D). Altogether there were 27,018 positive SNPs for CAD and 25,640 for T2D (uncorrected $P < 0.05$). These positive SNPs were further mapped to 2854 for CAD and 3113 positive genes for T2D. Gene regions are defined as the range from 2 kb upstream to 2 kb downstream of transcribed regions according to the hg19 coordinates in the UCSC website. Haplotype blocks were estimated using Gabriel's confidence intervals' algorithm to split genes into independent blocks, and the most significant SNP for each block (risk SNPs) was selected to represent the block. Fisher's combination test was calculated for these independent risk SNPs in each gene to perform gene-based association test. The odds ratios (ORs) for each gene were the product of the ORs of all independent SNPs (the reciprocal was used if the OR was smaller than 1), which represented the maximum relative risk of the gene. All GWAS analysis was performed by PLINK (Purcell et al., 2007).

2.2. Construction and analysis of the PPI network

Non-redundant PPI database was built up by applying MiMI software to collect gene interaction information from numerous protein interaction databases (Tarcea et al., 2009), including BIND (Bader et al., 2001), HPRD (Keshava Prasad et al., 2009), GO (Ashburner et al., 2000), KEGG (Kanehisa et al., 2010), Pfam (Famula et al., 2000) and Reactome (Matthews et al., 2009). Using GIN software, the PPI database was organized into molecule specific networks (Ozgur et al., 2008) that were initialized from several seed genes. Phenopedia component of the online Human Genome Epidemiology (HuGE) encyclopedia provided a disease-centered view of genetic association studies (Yu et al., 2010).

Table 1
Modules and network of T2D and CAD.

Organizations	T2D		CAD	
	Genes	Interactions	Genes	Interactions
Network	4038	45,567	3110	40,504
Modules				
1	1122	5404	895	5418
2	2	1	2	1
3	2	1	2	1
4	3	2	2	1
5	2	1	2	1
6	2	1	NA	NA
7	2	1	NA	NA

Genes: genes involved in network and modules; interactions: the interactions between the genes involved in the study; NA: not applicable.