International Journal of Cardiology 283 (2019) 144–150

Contents lists available at ScienceDirect

International Journal of Cardiology

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

Multivariate analysis of genome-wide data to identify potential pleiotropic genes for type 2 diabetes, obesity and coronary artery disease using MetaCCA

XiaoCan Jia [a], YongLi Yang [a], YuanCheng Chen [b], Zhenhua Xia [a], Weiping Zhang [a], Yu Feng [a], Yifan Li [a], Jiebing Tan [a], Chao Xu [c], Qiang Zhang [a], Hongwen Deng [c], XueZhong Shi [a,*]

[a] Department of Epidemiology and Biostatistics, College of Public Health, Zhengzhou University, Zhengzhou, Henan, China
[b] Department of Endocrinology and Metabolism, the Third Affiliated Hospital of Southern Medical University, Guangzhou, Guangdong, China
[c] Center for Bioinformatics and Genomics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA, USA

ARTICLE INFO

Article history:
Received 9 September 2018
Received in revised form 17 October 2018
Accepted 29 October 2018
Available online 31 October 2018

Keywords:
Obesity
Type 2 diabetes
Coronary artery disease
Multivariate statistical analysis
Genome-wide association study
Pleiotropic

ABSTRACT

Background: Although genome-wide association studies (GWAS) have been extensively applied in identifying SNP associated with metabolic diseases, the SNPs identified by this prevailing univariate approach only explain a small percentage of the genetic variance of traits. The extensive previous studies have repeatedly shown type2 diabetes (T2D), obesity and coronary artery disease (CAD) have common genetic mechanisms and the overlapping pathophysiological pathways.

Methods: The genetic pleiotropy-informed metaCCA method was applied on summary statistics data from three independent meta-GWAS summary statistics to identify shared variants and pleiotropic effect between T2D, obesity and CAD. Furthermore, to refine all genes, we performed gene-based association analyses for these three diseases respectively using VEGAS2. Gene enrichment analysis was applied to explore the potential functional significance of the identified genes.

Results: After metaCCA analysis, 833 SNPs reached the Bonferroni corrected threshold ($p < 7.99 \times 10^{-7}$) in the univariate SNP-multivariate phenotype analysis, and 327 genes with a significance threshold ($p < 3.73 \times 10^{-6}$) were identified as potentially pleiotropic genes in the multivariate SNP-multivariate phenotype analysis. By screening the results of gene-based p-values, we identified 22 putative pleiotropic genes which achieved significance threshold in metaCCA analyses and were also associated with at least one disease in the VEGAS2 analyses.

Conclusions: The metaCCA method identified novel variants associated with T2D, obesity and CAD by effectively incorporating information from different GWAS datasets. Our analyses may provide insights for some common therapeutic approaches of metabolic diseases based on the pleiotropic genes and common mechanisms identified.

1. Introduction

Coronary artery disease (CAD) is a multifactorial chronic disorder and one of leading causes of morbidity and mortality worldwide, which is characterized with coronary arteries progressing silently and usually has established to an advance stage by the time symptoms start appearing [1,2]. Epidemiological studies have estimated that CAD death accounts for approximately 25% of all cardiovascular deaths worldwide and the most significant risk factors associated with the development of CAD include genetic and environmental factors such as hypertension, high blood cholesterol levels, hyperglycemia, smoking, and obesity. Risk factors, particularly obesity and type 2 diabetes (T2D), have already had well-established associations with CAD [3]. Specifically, T2D is closely associated with several of the mechanisms that lead to CAD independently and T2D is associated with an increased risk of CAD by two to four-fold in observational studies [4–7]. Several previous epidemiological studies and meta-analysis have demonstrated that every 1 kg/m^2 increase in body mass index (BMI) leads to a 5–7% increase in the incidence of CAD, and obese participants had a significantly greater risk of CAD (relative risk – RR 1.81) after the adjustment for age, sex, physical activities, and smoking [8,9].

Pleiotropy describes the genetic effect of a single nucleotide polymorphisms (SNP) or gene on two or more phenotypic traits and its outcome is genetic correlation. Largely, this concept concerns across-trait architecture [10]. Previous studies have repeatedly indicated that T2D, obesity and CAD have common genetic mechanisms [1]. For example, *TCF7L2*, *IRS1*, and other 9 genes association with T2D have

* Corresponding author.
E-mail address: xzshi@zzu.edu.cn (X. Shi).

https://doi.org/10.1016/j.ijcard.2018.10.102
0167-5273/

been confirmed as the causal influence on CAD risk by the mendelian randomization methods, which means T2D could led to CAD [6]. *CNNM2*, *ATXN2* and *HIP1* were identified associated with BMI and CAD by the conditional false discovery rate method [2]. Furthermore, dozens of genetic loci associated with obesity and T2D have been demonstrated by Genome wide association studies (GWAS) [11]. It is therefore important to identify pleiotropic genes that acting through common biological mechanisms and assess overlapping pathophysiological relationships of these three disorders using effective analytical approaches.

GWAS is a powerful, systematic, and standard univariate approach to investigate and identify potentially causal or risk-conferring genetic variants for complex diseases in the individual level measurement [12,13]. GWAS, especially those with large sample size and meta-analysis of multiple studies, have identified thousands of genetic variants associated with complex traits or diseases. However, they only explain a small part of observed heritability for T2D, obesity and CAD [14,15]. Indeed, multivariate analysis may have higher statistical power to detect the unexplained heritability due to considering correlations not only among multiple SNPs but also among different traits or diseases [16,17]. Existing studies of genetic risk factors for complex traits have used bivariate analysis, and multivariate analysis based on multiple correlated phenotypes or diseases is rare [18]. Therefore, a multivariate analysis to identify pleiotropic genes, especially using the publicly available summary statistics of GWAS, is worth pursuing.

Cichonska et al. [19] recently performed a canonical correlation analysis (metaCCA) method allowing multivariate representation of both genotypic and phenotypic variables. This new approach aims to increase statistical power to identify novel genetic associations, and the core principle is to use the method of canonical correlation analysis (CCA) to identify linear relationships between two sets of variables: multiple SNPs against multiple traits based on the published univariate summary statistics from GWAS by meta-analysis [12,19]. Cichonska [19] has successfully applied this method to 9 lipid measures related from studies of three Finnish cohorts and the results showed metaCCA highly improved the statistical power by considering the correlations among multiple SNPS and multiple phenotypes.

In this study, we applied the genetic pleiotropy-informed metaCCA method on summary statistics data from three independent meta-GWAS summary statistics to identify shared variants and pleiotropic effect between T2D, obesity and CAD. By using this method, we could identify more common variants that are genetic risk factors for one or more common disorders and resulting potentially shared genetic influences should provide novel effective approaches for preventing and treating metabolic diseases ultimately.

2. Methods

2.1. GWAS datasets

The GWAS dataset for T2D contains association summary statistics of 2,473,441 imputed SNPs was downloaded from http://www.diagram-consortium.org/downloads.html. This meta-analysis compromising of 48,286 cases and 250,671controls of European Ancestry with body mass index adjustment performed by the Diabetes Genetics Replication and Meta-analysis (DIAGRAM) Consortium [20]. The dataset for BMI was based on a meta-analysis of 2,550,021 genotyped or imputed SNPs from 234,069 individuals of European ancestry performed by the Genetic Investigation of Anthropometric Traits (GIANT) Consortium, which was downloaded from http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files [21]. The dataset for CAD was a meta-analysis of 22 GWAS studies of European descent imputed to HapMap 2 involving 22,233 cases and 64,762 controls performed by the transatlantic Coronary Artery Disease Genome-wide Replication and Meta-analysis (CARDIoGRAM) Consortium, which was downloaded from http://www.cardiogramplusc4d.org/data-downloads [22]. All the samples in the GWAS datasets came from populations of European ancestry. The summary statistics have undertaken at least two times of genomic control separately at the individual study level and meta-analysis. Further detailed descriptions of the sample ascertainment and stringent quality control procedures can be found in the corresponding consortium publications [20–23]. We avoided reduplicating control individuals when selecting these datasets. The data contain summary statistics, only including p-values, regression coefficients and standard error after meta-analysis. Finally, 2,382,957 overlapping SNPs of T2D, BMI, and CAD were selected on which we performed the multivariate analysis.

2.2. Data preparation

The analytical workflow of our study was presented in Fig. 1. Before the implementation of the metaCCA method, several steps were undertaken. First, we combined the summary statistics for the 2,382,957 common SNPs included in the studies of T2D, BMI, and CAD and completed the gene annotation for the three GWAS according to the 1000 Genome datasets using PLINK1.9. The reference data, which contained 26,291 genes, were downloaded from the website: https://www.cog-genomics.org/static/bin/plink/glist-hg19. Second, a linkage disequilibrium (LD) based SNP pruning method was used to remove SNPs with large pairwise correlations. The SNP pruning method was proceeded by a window of 50 SNPs where LD was calculated between each pair of SNPs. The minor allele frequency (MAF) is also considered for the SNP pruning, and SNP with smaller MAF for pairs with $r^2 > 0.2$ were removed. Following this initial removal of SNPs in high LD, each sliding window of 5 SNPs forward and the process repeated until there were no pairs of SNPs with high LD [11]. All datasets were pruned using the HapMap 3 CEU genotypes as a reference panel. After gene annotation and SNP pruning, there remained 62,553 SNPs located in 13,420 gene regions on which we performed the metaCCA analysis. The regression coefficient beta was normalized before conducting the metaCCA analysis because the individual-level data set genotype and phenotype matrices were not standardized. Standardization was achieved afterwards by:

$$\beta_{gp}^{STANDR} = \frac{1}{\sqrt{n}SE_{gp}} \times \beta_{gp} \tag{1}$$

where SE_{gp} is the standard error of β_{gp}, as given by the original GWAS result, g is the number of genotypic, p is the number of phenotypic variables, and n *is* the sample number of each diseases.

2.3. MetaCCA analysis

MetaCCA is an extension of the method of CCA, which required a cross-covariance matrix between all genotypic and phenotypic variables ($\sum XY$), a genotypic correlation structure between SNPs ($\hat{\sum} XX$), and a phenotypic correlation structure between traits ($\hat{\sum} YY$) [19]. $\sum XY$ is constructed as the normalized regression coefficient βgp:

$$\sum XY = \frac{X^T Y}{N-1} = \begin{pmatrix} \beta_{11} & \beta_{12} & \cdots & \beta_{1P} \\ \beta_{21} & \beta_{22} & \cdots & \beta_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \beta_{G1} & \beta_{G2} & \cdots & \beta_{GP} \end{pmatrix} \tag{2}$$

where G and P are the number of genotypic and phenotypic variables, respectively.

In metaCCA, $\hat{\sum} XX$ is calculated using a reference database representing the study population, such as the 1000 Genomes database, or other genotypic data available on the target population. There will be better results if $\hat{\sum} XX$were estimated from the target population or the same ethnicity instead of interracial populations [19]. In our study, $\hat{\sum} XX$ was estimated using the reference SNP dataset of the HapMap 3 CEU.

The phenotypic correlation structure $\hat{\sum} YY$ is computed based on $\sum XY$. Each entry of $\hat{\sum} YY$corresponds to a Pearson correlation coefficient between the vector of β estimates from p phenotypic variables across g genetic variants. It has been demonstrated that the bigger the number of genotypic variables g, the more accurate the quality of the estimate. Thus, $\hat{\sum} YY$were calculated from summary statistics for all available genetic variants (2,382,957 overlapping SNPs), even if only a part were used for further analysis.

After calculation, the full covariance matrix ($\sum$), consisting of three covariance matrices, can be obtained:

$$\sum = \begin{pmatrix} \hat{\sum} XX & \sum XY \\ \sum_{XY}^{T} & \hat{\sum} YY \end{pmatrix} \tag{3}$$

We need determine whether the full covariance matrix is positive semidefinite (PSD). If it is not PSD, an iterative procedure is used to shrink the full covariance matrix until $\sum$ becomes PSD. In the next analysis, the PSD of the full covariance matrix is plugged into the CCA framework to get the final genotype-phenotype association result [19]. The correlation between genotype and phenotype is called the canonical correlation r [24].

In this study, two types of multivariate analysis were considered. First, univariate SNP-multivariate phenotype association analysis was tested at the SNP level. Manhattan plots presented all SNPs within an LD block in relation to their chromosomal locations. To identify any potential pleiotropic gene, we did multivariate SNP-multivariate phenotype association analysis at the gene level. The result was the canonical correlation of a gene with all three diseases. We checked the summary statistics of GWAS for a set of 62,553 pre-selected SNPs and found that the mean standardized β's is close to zero and the median p-value for those β's close to 0.5 in all the three GWAS datasets, which means the sample of SNPs behaves like a random sample of the whole genome, then

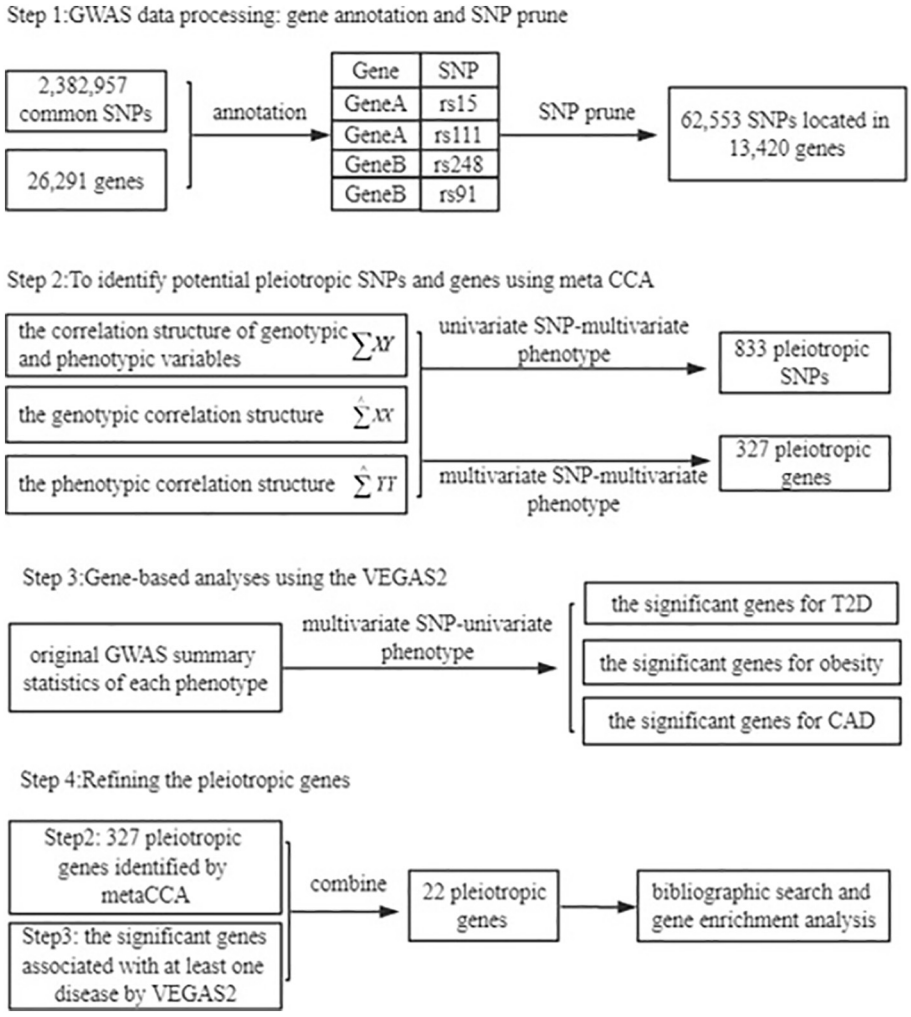

Fig. 1. The analytical workflow of the present study.

Bonferroni corrected p-value < 0.05 was used as the threshold for nominal significance. If the p-value of the canonical correlation r of any SNP was smaller than 7.99×10^{-7} (=0.05/62,553), it was deemed significantly associated with the three diseases. Because the β's of genes could not be obtained and computed from the summary statistics of GWAS, a conservative corrected method-Bonferroni corrected threshold is used in the gene level. Similarly, genes with a canonical correlation p-value smaller than 3.73×10^{-6} (=0.05/13,420) were significantly associated with all the three diseases.

2.4. Gene-based analyses

To refine the identified genes by MetaCCA, we performed gene-based association analyses, using the VEGAS2 (Versatile Gene-based Association Study–2) method (performed at: https://vegas2.qimrberghofer.edu.au/) [25]. This method calculates the correlation analysis of multiple SNPs in a gene region with one phenotype using original GWAS summary statistics, which has previously shown higher sensitivity and lower false positive rates compared to other gene-based approaches [26]. All SNPs in each gene was analysed using the 1000 Genomes European reference genotypes. We obtained the gene-based p-value of each gene for one disease and selected the pleiotropic genes that were associated with at least one disorder using 1E-06 as the threshold.

2.5. Functional annotation and gene enrichment analysis

An useful way to understand polygenic associations is to determine whether the implicated genetic variants occur in genes that comprise a biological pathway or not [2]. To evaluate the potential biological function of all putative pleiotropic genes, we conducted the GO enrichment analysis using Enrichr (http://amp.pharm.mssm.edu/Enrichr/). All significant genes re-identified by VEGAS2 in our study were annotated and enriched based on three main categories: biological processes, cellular component and molecular functions. An adjusted p-value < 0.05 in the enrichment analysis indicates the nominal significance [27].

3. Results

To identify common variants shared among T2D, obesity, and CAD, we undertook a two-step analysis strategy. First, by using the metaCCA method, we inspected the potential pleiotropic SNPs and genes association with these three phenotypes. Next, by adopting the VEGAS2 method, we checked those potential common variants for their specific associations with individual disorders.

3.1. Potential pleiotropic SNPs and genes by metaCCA analysis

After gene annotation and SNP pruning, there were 62,553 SNPs located in 13,420 gene regions available for the metaCCA analysis. The size of SNP representation of the genes ranged from 1 to 244 SNPs; the median number of SNPs in each gene was 4.66. For the univariate SNP-multivariate phenotype analysis, 833 SNPs reached the Bonferroni corrected threshold ($p < 7.99 \times 10^{-7}$), and the canonical correlation r between each SNP and phenotype ranged from 0.0118 to 0.0483. The results are presented by the Manhattan plot in Fig. 2. If the $-\log_{10}$ (metaCCA) value of a certain SNP was >6.10, this SNP was flagged as a potential pleiotropic SNP for T2D, obesity, and CAD. For the multivariate SNP-multivariate phenotype analysis, 327 genes with a significance threshold of p-value < 3.73×10^{-6} were identified as the potential pleiotropic genes. The canonical correlation r between genotype and phenotype ranged from 0.0234 to 0.6292.

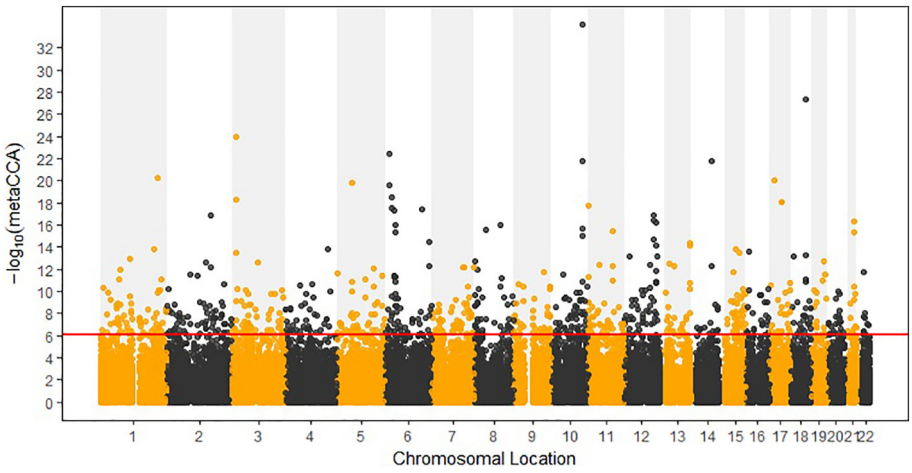

Fig. 2. Manhattan plot of $-\log_{10}$(metaCCA) values for univariate SNP-T2D, obesity, and CAD analysis. The red line marks the $-\log_{10}$(metaCCA) value of 6.10 corresponding to $p < 7.99 \times 10^{-7}$. If the $-\log_{10}$(metaCCA) value of a certain SNP was >6.10, this SNP was identified as a pleiotropic SNP for these three correlated diseases.

3.2. Refining the pleiotropic genes by gene-based analyses

After the metaCCA analysis, we refined the list of 327 pleotropic genes associated with more than one disorder to identify their association with specific traits using the gene-based p-value calculation using the VEGAS-2 algorithm. 9 genes were identified for T2D, 75 associated genes were identified for obesity, and 5 significant genes were identified for CAD with the adjusted p-value < 0.05.

By screening the results of gene-based analysis p-values, we identified 22 putative pleiotropic yielding significance in the metaCCA analyses and were associated with at least one disease in the VEGAS2 analyses. 6 of the possible 22 pleiotropic genes were associated with T2D,11 genes were identified as the associated genes with obesity, and 8 genes were identified as being associated with CAD. In particular, 3 genes (*CADM2*, *MASP1*, *TCF7L2*) were identified as pleiotropic genes in the original GWAS. The findings of the metaCCA and VEGAS2 analyses are summarized in Table 1.

Specifically, 10 of these 22 putative pleiotropic genes (*ANKS1A*, *CADM2*, *DPP4*, *KCNQ1*, *KIF11*, *MASP1*, *NEDD4L*, *NEGR1*, *PHACTR1*, *TCF7L2*) had been previously reported to be associated with more than one of these three disorders. Of these 10 confirmed pleiotropic genes, 6 genes (*CADM2*, *DPP4*, *KCNQ1*, *KIF11*, *TCF7L2*) were repeatedly reported to be associated with T2D, obesity and CAD in published studies, 3 genes (*ANKS1A*, *MASP1*, *PHACTR1*) were associated with T2D and CAD, *NEDD4L* was associated with obesity and CAD, and *NEGR1*were associated with T2D and obesity.

Of the 12 detected novel putative pleiotropic genes, *NCOA1* was previously reported to be associated with T2D by the contributing GWAS, 3 genes (*MIR4721*, *NR5A2*, *PTPMT1*) were associated with obesity, and *AGBL1*was associated with CAD. Other remaining significant genes (*CCDC158*, *COL4A3BP*, *FAT3*, *RFWD2*, *SLC24A1*, *SLC39A13*, *TMEM219*), might represent candidate novel pleiotropic genes for these three diseases. The detailed features of 22 significant pleiotropic genes are shown in Table 2.

Table 1
The 22 pleiotropic genes identified by the metaCCA and VEGAS2 analysis.

Locus	Gene	MetaCCA adjusted p-value	VEGAS adjusted p-value		
			T2D	obesity	CAD
1	*AGBL1*	7.49E-10	0.14	1.00E-06	0.53
2	*ANKS1A*	3.62E-11	1.03E-04	1.00E-06	9.09 E-04
3	*CADM2*	1.71E-08	1.00E-06	1.00E-06	0.13
4	*CCDC158*	1.75E-10	0.06	1.00E-06	0.91
5	*COL4A3BP*	4.58E-07	0.52	1.00E-06	0.07
6	*DPP4*	1.56E-225	0.45	1.00E-06	0.17
7	*FAT3*	7.68E-28	1.00E-06	1.75 E-06	0.75
8	*KCNQ1*	3.49E-98	1.00E-06	0.17	0.15
9	*KIF11*	6.81E-08	1.00E-06	0.02	0.34
10	*MASP1*	1.26E-16	1.00E-06	0.18	1.00E-06
11	*MIR4721*	1.26E-08	0.70	1.00E-06	0.23
12	*NCOA1*	1.15E-06	0.79	1.00E-06	0.36
13	*NEDD4L*	9.43E-21	1.08E-04	1.00E-06	0.24
14	*NEGR1*	1.33E-35	0.13	1.00E-06	0.44
15	*NR5A2*	2.32E-10	0.35	1.00E-06	1.00E-03
16	*PHACTR1*	5.70E-46	0.61	0.37	1.00E-06
17	*PTPMT1*	8.94E-31	0.45	0.13	1.00E-06
18	*RFWD2*	4.22E-07	0.59	0.03	1.00E-06
19	*SLC24A1*	3.56E-10	0.65	0.77	1.00E-06
20	*SLC39A13*	4.53E-10	0.03	0.75	1.00E-06
21	*TCF7L2*	1.60E-113	1.00E-06	0.10	1.00E-06
22	*TMEM219*	4.42E-12	0.42	0.10	1.00E-06

Table 2
The features of 22significant pleiotropic genes.

Locus	Gene	Chr	Gene type[a]	r-Value	Adjusted p-value	Number of SNPs
1	*AGBL1*	11	Novel*	0.05	7.49E-10	52
2	*ANKS1A*	6	Confirmed	0.04	3.62E-11	4
3	*CADM2*	3	Confirmed	0.03	1.71E-08	18
4	*CCDC158*	12	Novel	0.03	1.75E-10	5
5	*COL4A3BP*	5	Novel	0.03	4.58E-07	2
6	*DPP4*	15	Confirmed	0.13	1.56E-225	3
7	*FAT3*	11	Novel	0.05	7.68E-28	27
8	*KCNQ1*	11	Confirmed	0.10	3.49E-98	41
9	*KIF11*	10	Confirmed	0.03	6.81E-08	3
10	*MASP1*	3	Confirmed	0.04	1.26E-16	6
11	*MIR4721*	16	Novel*	0.04	1.26E-08	30
12	*NCOA1*	2	Novel*	0.04	1.15E-06	5
13	*NEDD4L*	18	Confirmed	0.04	9.43E-21	15
14	*NEGR1*	1	Confirmed	0.06	1.33E-35	24
15	*NR5A2*	1	Novel*	0.04	2.32E-10	11
16	*PHACTR1*	6	Confirmed	0.07	5.70E-46	23
17	*PTPMT1*	11	Novel*	0.06	8.94E-31	9
18	*RFWD2*	1	Novel	0.03	4.22E-07	14
19	*SLC24A1*	15	Novel	0.04	3.56E-10	3
20	*SLC39A13*	11	Novel	0.04	4.53E-10	1
21	*TCF7L2*	10	Confirmed	0.10	1.60E-113	15
22	*TMEM219*	16	Novel	0.04	4.42E-12	2

Note:
Confirmed: This gene was previously reported to be associated with more than one disease.
Novel*: This gene had been reported to be associated with only one disease of T2D, obesity and CAD.
Novel: This gene had never been reported to be associated with T2D, obesity or CAD.
[a] The references that support this gene was a confirmed or novel* gene had been listed in the file of e-component.

3.3. Functional term enrichment analysis

GO enrichment analyses revealed that the biological functions of these pleiotropic genes were mainly involved in the metabolism of lipids. When 22 pleiotropic genes associated with T2D, BMI and CAD were used as the gene sets for the GO term enrichment analysis (conforming to the up-to-date 2017 database), several functional terms were identified as being enriched. For the GO biological process, the top five significant GO terms were regulation of membrane repolarization (GO:0060306), ventricular cardiac muscle cell action potential (GO:0086005), cardiac muscle cell action potential involved in contraction (GO:0086002), regulation of ion transmembrane transport (GO:0034765), and regulation of potassium ion transmembrane transport (GO:1901379). As a result, *KCNQ1* and *NEDD4L* were the overlapping genes of the five related terms. For the GO molecular function, the top significant GO term was hormone receptor binding (GO:0051427), which involved two genes of *NCOA1* and *TCF7L2*. This GO term enrichment analysis furnished supporting evidence for our results from a functional aspect and may contribute to the illumination of etiology of T2D, BMI and CAD. Detailed information is shown in Table 3.

4. Discussion

In the present study, a novel analytical approach – metaCCA was used to explore the common genetic variants for T2D, obesity and CAD by combining three independent GWAS meta-analyses with available summary statistics. After verification using gene-based analyses, we successfully identified a total of 22 putative pleiotropic genes and performed the functional term enrichment analysis based on these results. In particularly, 10 confirmed genes were identified as pleiotropic in previous different types of studies and were validated in the present study, 5 novel pleiotropic genes were reported to be associated with one disease in previous study, and 7 candidate novel pleiotropic genes were never reported to be associated with T2D, obesity or CAD. The improved detection not only yielded the potential shared genetic components but also provide better understanding for further exploring potential common biological pathogenesis of these metabolic diseases.

Among the 10 confirmed pleiotropic genes, some was shown to play an important role on the pathomechanism of metabolic diseases. For example, common genetic variants in *TCF7L2* are associated with the risk for T2D, obesity and CAD [28]. As a transcription factor, *TCF7L2* is not only the main susceptibility gene for T2D, primarily through impairing the insulin secretion by pancreatic beta cells, but also has an influence on vascular GLP-1 receptor expression in obesity population [29,30]. Most of all, it has been thought that incretin signaling prevents arteriosclerosis, and very recently anti-arteriosclerotic effects through GLP-1 receptor were finally demonstrated in clinical human experiment [29]. Srivastava et al. [31] also shown that altered function of *TCF7L2* axis can induce vascular smooth muscle cells plasticity and initiate vascular wall remodeling, which may be the potential targets for the pharmacotherapy of CAD. GO terms enrichment analysis results also suggest hormone receptor binding could be a key biological process for T2D, obesity and CAD. Another important and confirmed pleiotropic gene for these three diseases is *KCNQ1*, which alters expression, function and sensitivity of potassium voltage-gated channel. It has been confirmed *KCNQ1* is hypermethylated in the obese subjects [32]. It causes smooth muscle dysfunction and probably endothelial dysfunction which makes people particularly prone to premature cardiovascular disease [33]. Several recent GWAS and pathway/GO terms enrichment analyses have found variants in *KCNQ1* were associated with risk of T2D and CAD [34,35].

Interestingly, 5 (*AGBL1*, *NCOA1*, *MIR4721*, *NR5A2*, *PTPMT1*) of the 12 detected novel putative pleiotropic genes had been validated associated with one kind of T2D, obesity and CAD. *AGBL1* is an associated gene with activated partial thromboplastin time, which has been identified though assessment of existing gene expression and CAD databases from the Atherosclerosis Risk in Communities (ARIC) study [36]. Meerson et al. [37] suggest that miR-4443 acts in a tumor-suppressive manner by down-regulating *NCOA1* downstream of MEK-C/EBP-mediated leptin and insulin signaling, and eventually insulin and/or leptin resistance may suppress this pathway and increase the risk of metabolic disease such as obesity and T2D [38]. Among the 5 novel pleiotropic genes, for genes, *MIR4721*, *NR5A2* and *PTPMT1*, were suggested to be pleiotropic genes for BMI based on the results of previous studies [39,40]. The *MIR4721* gene, is located in gene promoters, which is associated with proximal gene regulation including DNA methylation and other chromatin marks [39]. However, the detailed molecular pathway associated with human diseases and traits is unclear. *NR5A2* is a nuclear receptor which regulates the expression of genes involved in cholesterol metabolism, pluripotency maintenance and cell differentiation. It has been recently shown that a ligand of *NR5A2* prevents liver steatosis and improves insulin sensitivity in mouse models of insulin resistance, which may have an effect on the formation of pancreatic cancer and other metabolic diseases [41]. Similar to *MIR4721*, *PTPMT1* is also associated with DNA methylation. Fortunately, a new experimental study has

Table 3
Top five significant GO Term enrichment of the 22 pleiotropic genes.

	p-Value	Adjusted p-value	Genes
Term (GO_Biological_Process)			
Regulation of membrane repolarization (GO:0060306)	1.56E-05	0.02	*KCNQ1; NEDD4L*
Ventricular cardiac muscle cell action potential (GO:0086005)	1.37E-05	0.02	*KCNQ1; NEDD4L*
Cardiac muscle cell action potential involved in contraction (GO:0086002)	4.63E-04	0.04	*KCNQ1; NEDD4L*
Regulation of ion transmembrane transport (GO:0034765)	7.11E-04	0.04	*KCNQ1; NEDD4L*
Regulation of potassium ion transmembrane transport (GO:1901379)	5.97E-04	0.04	*KCNQ1; NEDD4L*
Term (GO_Cellular_Component)			
beta-Catenin-TCF complex (GO:1990907)	1.31E-02	0.16	*TCF7L2*
Invadopodium (GO:0071437)	1.42E-02	0.16	*DPP4*
Cul4A-RING E3 ubiquitin ligase complex (GO:0031464)	1.53E-02	0.16	*RFWD2*
Nuclear chromatin (GO:0000790)	3.14E-02	0.19	*NCOA1; TCF7L2*
Cul4-RING E3 ubiquitin ligase complex (GO:0080008)	3.99E-02	0.19	*RFWD2*
Term (GO_Molecular_Function)			
Hormone receptor binding (GO:0051427)	9.23E-04	0.04	*NCOA1; TCF7L2*
Nuclear hormone receptor binding (GO:0035257)	1.78E-03	0.07	*NCOA1; TCF7L2*
Protein homodimerization activity (GO:0042803)	5.50E-03	0.07	*DPP4; CADM2; SLC39A13; MASP1*
Sodium channel inhibitor activity (GO:0019871)	7.98E-03	0.07	*NEDD4L*
Aryl hydrocarbon receptor binding (GO:0017162)	9.86E-03	0.07	*NCOA1*

reported that knockdown of *PTPMT1* expression in the pancreatic insulinoma cell line alters the mitochondrial phosphoprotein profile and markedly enhances insulin secretion, which indicates *PTPMT1* is a potential drug target for the treatment of T2D [42].

Compared with genetic findings in previous GWAS of T2D, obesity and CAD, there are 7 candidate novel genes (*CCDC158*, *COL4A3BP*, *FAT3*, *RFWD2*, *SLC24A1*, *SLC39A13*, *TMEM219*) not previously reported. However, *RFWD2* (also called the ubiquitin ligase COP1) in beta cells is critical for insulin secretion [43]. *TMEM219* has been designated as receptors for insulin-like growth factor binding protein-3 (IGFBP-3), and IGFBP-3 act as an important contributor to the regulation of somatic growth by acting as the major circulating transport protein [44]. Therefore, *RFWD 2* and *TMEM219* may play some important role in the development and therapeutics of T2D. In addition, Lee, Yet al [45] statistical analysis demonstrated that the *CCDC158* gene is strongly associated with body weight and cold carcass weight in Hanwoo, which become a new research direction of risk gene related to obesity in population. Here, we don't have a detailed description of each candidate novel gene because pathomechanisms are unclear apparently, and further experimental studies will need to be conducted to confirm our novel findings.

Many genes and pathways may have pleiotropic effects on more than one disease, which is a common phenomenon in chronic disease and metabolic disease [46]. Systematically and comprehensively searching for the pleiotropic genes and their effects is essential and necessary. Compared to the animal experiments or cross-sectional population-based studies, the advantages of this study are as follows. First, the statistical power of present study is increased through the metaCCA method by integrating three large GWAS summary statistics, which provided an increase in effective sample size. Second, jointly analyzing multiple related traits including T2D, obesity and CAD lead to richer findings compared univariate disease analyses. Not only a few reported pleiotropic genes existing in T2D, obesity and CAD were verified, but also novel pleiotropic genes were detected in this study. In addition, it is a cost-effective study based on the data of GWAS summary statistics compared to the conventional standard GWAS sequencing technology. However, this study could not relate to the information about the direction of effects of pleiotropic genes on risk to these diseases because of a lack of detailed original individual measures. Alternative approaches and experimental studies may be applied to check whether novel genes could still be identified/substantiated with these methods in order to confirm novel findings in the further study.

In summary, we have performed a systematic multivariate analysis of the open genome-wide data using metaCCA, verified 10 confirmed pleiotropic genes in the previous studies and highlighted 12 significant genes that may be the novel pleiotropic candidate genes for at least two of the three diseases. Furthermore, we also illustrated potential biological functions of this pleiotropic genes and our results may provide with novel insights into the shared genetic factors in development of T2D, obesity and CAD.

Conflict of interest

All authors declared no financial/personal interest.

Funding

This work was supported by the National Institutes of Health (https://www.nih.gov/) to HWD (AR069055, U19 AG055373, R01 MH104680, R01AR059781 and P20GM109036), Key Technologies R & D Program of Henan Province in China (http://www.hnkjt.gov.cn/) to YLY (152102310263, 152102410007), and Edward G. Schlieder Endowment to HWD.

References

[1] S.L.A. Yeung, C.M. Schooling, Adiponectin and coronary artery disease risk: a bi-directional Mendelian randomization study, Int. J. Cardiol. 268 (2018) 222–226.
[2] W.-Q. Lv, X. Zhang, Q. Zhang, J.-Y. He, H.-M. Liu, X. Xia, et al., Novel common variants associated with body mass index and coronary artery disease detected using a pleiotropic cFDR method, J. Mol. Cell. Cardiol. 112 (2017) 1–7.
[3] G. Chen, D. Levy, Contributions of the Framingham heart study to the epidemiology of coronary heart disease, JAMA Cardiol. 1 (2016) 825–830.
[4] H.L. Hillage, The Emerging Risk Factors Collaboration. Diabetes mellitus, fasting blood glucose concentration, and risk of vascular disease: a collaborative meta-analysis of 102 prospective studies (vol 375, pg 2215, 2010), Lancet 376 (2010) 958.
[5] M. Laakso, J. Kuusisto, Insulin resistance and hyperglycaemia in cardiovascular disease development, Nat. Rev. Endocrinol. 10 (2014) 293–302.
[6] O.S. Ahmad, J.A. Morris, M. Mujammami, V. Forgetta, A. Leong, R. Li, et al., A Mendelian randomization study of the effect of type-2 diabetes on coronary heart disease, Nat. Commun. 6 (2015).
[7] M.R. Movahed, M. Hashemzadeh, M.M. Jamal, Diabetes mellitus is a strong, independent risk for atrial fibrillation and flutter in addition to other cardiovascular disease, Int. J. Cardiol. 105 (2005) 315–318.
[8] E.Y. Cho, J.E. Manson, M.J. Stampfer, C.G. Solomon, G.A. Colditz, F.E. Speizer, et al., A prospective study of obesity and risk of coronary heart disease among diabetic women, Diabetes Care 25 (2002) 1142–1148.
[9] R.P. Bogers, W.J.E. Bemelmans, R.T. Hoogenveen, H.C. Boshuizen, M. Woodward, P. Knekt, et al., Association of overweight with increased risk of coronary heart disease partly independent of blood pressure and cholesterol levels - a meta-analysis of 21 cohort studies including more than 300,000 persons, Arch. Intern. Med. 167 (2007) 1720–1728.
[10] F.W. Stearns, One hundred years of pleiotropy: a retrospective (vol 186, pg 767, 2010), Genetics 187 (2011) 355.
[11] Q. Zhang, K.-H. Wu, J.-Y. He, Y. Zeng, J. Greenbaum, X. Xia, et al., Novel common variants associated with obesity and type 2 diabetes detected using a cFDR method, Sci. Rep. 7 (2017).
[12] C.S. Tang, M.A.R. Ferreira, A gene-based test of association using canonical correlation analysis, Bioinformatics 28 (2012) 845–850.
[13] J. Kettunen, T. Tukiainen, A.-P. Sarin, A. Ortega-Alonso, E. Tikkanen, L.-P. Lyytikainen, et al., Genome-wide association study identifies multiple loci influencing human serum metabolite levels, Nat. Genet. 44 (2012) 269–U65.
[14] P.M. Visscher, M.A. Brown, M.I. McCarthy, J. Yang, Five years of GWAS discovery, Am. J. Hum. Genet. 90 (2012) 7–24.
[15] D. Welter, J. MacArthur, J. Morales, T. Burdett, P. Hall, H. Junkins, et al., The NHGRI GWAS catalog, a curated resource of SNP-trait associations, Nucleic Acids Res. 42 (2014) D1001–D1006.
[16] M. Inouye, S. Ripatti, J. Kettunen, L.-P. Lyytikainen, N. Oksala, P.-P. Laurila, et al., Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis, PLoS Genet. 8 (2012).
[17] P. Marttinen, J. Gillberg, A. Havulinna, J. Corander, S. Kaski, Genome-wide association studies with high-dimensional phenotypes, Stat. Appl. Genet. Mol. Biol. 12 (2013) 413–431.
[18] E. Evangelou, J.P.A. Ioannidis, Meta-analysis methods for genome-wide association studies and beyond, Nat. Rev. Genet. 14 (2013) 379–389.
[19] A. Cichonska, J. Rousu, P. Marttinen, A.J. Kangas, P. Soininen, T. Lehtimaki, et al., metaCCA: summary statistics-based multivariate meta-analysis of genome-wide association studies using canonical correlation analysis, Bioinformatics 32 (2016) 1981–1989.
[20] A. Mahajan, J. Wessel, S.M. Willems, W. Zhao, N.R. Robertson, A.Y. Chu, et al., Refining the accuracy of validated target identification through coding variant fine-mapping in type 2 diabetes, Nat. Genet. 50 (2018) 559.
[21] A.E. Locke, B. Kahali, S.I. Berndt, A.E. Justice, T.H. Pers, F.R. Day, et al., Genetic studies of body mass index yield new insights for obesity biology, Nature 518 (2015) 197–U401.
[22] N.N. Mehta, Large-scale association analysis identifies 13 new susceptibility loci for coronary artery disease, Circ. Cardiovasc. Genet. 4 (2011) 327–329.
[23] T.W. Winkler, F.R. Day, D.C. Croteau-Chonka, A.R. Wood, A.E. Locke, R. Maegi, et al., Quality control and conduct of genome-wide association meta-analyses, Nat. Protoc. 9 (2014) 1192–1212.
[24] J.A. Seoane, C. Campbell, I.N.M. Day, J.P. Casas, T.R. Gaunt, Canonical correlation analysis for gene-based pleiotropy discovery, PLoS Comput. Biol. 10 (2014).
[25] A. Mishra, S. Macgregor, VEGAS2: software for more flexible gene-based testing, Twin Res. Hum. Genet. 18 (2015) 86–91.
[26] G.L. Wojcik, W.H.L. Kao, P. Duggal, Relative performance of gene- and pathway-level methods as secondary analyses for genome-wide association studies, BMC Genet. 16 (2015).
[27] E.Y. Chen, C.M. Tan, Y. Kou, Q. Duan, Z. Wang, G.V. Meirelles, et al., Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool, BMC Bioinf. 14 (2013).
[28] R. Gong, M. Chen, C. Zhang, M. Chen, H. Li, A comparison of gene expression profiles in patients with coronary artery disease, type 2 diabetes, and their coexisting conditions, Diagn. Pathol. 12 (2017).
[29] T. Kimura, A. Obata, M. Shimoda, I. Shimizu, GdS Xavier, S. Okauchi, et al., Down-regulation of vascular GLP-1 receptor expression in human subjects with obesity, Sci. Rep. 8 (2018).
[30] S. Jainandunsing, H.R. Koole, J.N.I. van Miert, T. Rietveld, J.L. DarcosWattimena, E.J.G. Sijbrands, et al., Transcription factor 7-like 2 gene links increased in vivo insulin synthesis to type 2 diabetes, EBioMedicine 30 (2018) 295–302.

[31] R. Srivastava, J. Zhang, G.-w. Go, A. Narayanan, T.P. Nottoli, A. Mani, Impaired LRP6-TCF7L2 activity enhances smooth muscle cell plasticity and causes coronary artery disease, Cell Rep. 13 (2015) 746–759.
[32] A.M. Gomez-Uriz, F.I. Milagro, M.L. Mansego, P. Cordero, I. Abete, A. De Arce, et al., Obesity and ischemic stroke modulate the methylation levels of KCNQ1 in white blood cells, Hum. Mol. Genet. 24 (2015) 1432–1440.
[33] K.-C. Huang, T.-M. Li, X. Liu, J.-H. Chen, W.-K. Chien, Y.-T. Shiao, et al., KCNQ1 variants associate with hypertension in type 2 diabetes and affect smooth muscle contractility in vitro, J. Cell. Physiol. 232 (2017) 3309–3316.
[34] Y. Sakamoto, K. Oniki, A. Kajiwara, A. Yoshida, K. Taharasako, H. Ijima, et al., The clinical impact of the KCNQ1 rs2237892 polymorphism on the treatment of type 2 diabetes, Diabetes 66 (2017) (A666-A).
[35] S. Yoshino, R. Cilluffo, M. Prasad, P.J.M. Best, E.J. Atkinson, T. Aoki, et al., Sex-specific genetic variants are associated with coronary endothelial dysfunction, J. Am. Heart Assoc. 5 (2016).
[36] W. Tang, C. Schwienbacher, L.M. Lopez, Y. Ben-Shlomo, T. Oudot-Mellakh, A.D. Johnson, et al., Genetic associations for activated partial thromboplastin time and prothrombin time, their gene expression profiles, and risk of coronary artery disease, Am. J. Hum. Genet. 91 (2012) 152–162.
[37] A. Meerson, H. Yehuda, Leptin and insulin up-regulate miR-4443 to suppress NCOA1 and TRAF4, and decrease the invasiveness of human colon cancer cells, BMC Cancer 16 (2016).
[38] J.P. Bradfield, H.-Q. Qu, K. Wang, H. Zhang, P.M. Sleiman, C.E. Kim, et al., A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci, PLoS Genet. 7 (2011).
[39] S. Voisin, M.S. Almen, G.Y. Zheleznyakova, L. Lundberg, S. Zarei, S. Castillo, et al., Many obesity-associated SNPs strongly associate with DNA methylation changes at proximal promoters and enhancers, Genome Med. 7 (2015).
[40] H. Tang, X. Dong, M. Hassan, J.L. Abbruzzese, D. Li, Body mass index and obesity- and diabetes-associated genotypes and risk for pancreatic cancer, Cancer Epidemiol. Biomark. Prev. 20 (2011) 779–792.
[41] A. Bolado-Carrancio, J.A. Riancho, J. Sainz, J.C. Rodriguez-Rey, Activation of nuclear receptor NR5A2 increases Glut4 expression and glucose metabolism in muscle cells, Biochem. Biophys. Res. Commun. 446 (2014) 614–619.
[42] D.J. Pagliarini, S.E. Wiley, M.E. Kimple, J.R. Dixon, P. Kelly, C.A. Worby, et al., Involvement of a mitochondrial phosphatase in the regulation of ATP production and insulin secretion in pancreatic beta cells, Mol. Cell 19 (2005) 197–207.
[43] R. Suriben, K.A. Kaihara, M. Paolino, M. Reichelt, S.K. Kummerfeld, Z. Modrusan, et al., beta-Cell insulin secretion requires the ubiquitin ligase COP1, Cell 163 (2015).
[44] R.C. Baxter, Insulin-like growth factor binding protein-3 (IGFBP-3): novel ligands mediate unexpected functions, J. Cell Commun. Signal. 7 (2013) 179–189.
[45] Y.S. Lee, D.Y. Oh, J.J. Kim, J.H. Lee, H.S. Park, J.S. Yeo, A single nucleotide polymorphism in LOC534614 as an unknown gene associated with body weight and cold carcass weight in Hanwoo (Korean cattle), Asian Australas. J. Anim. Sci. 23 (2010) 1543–1551.
[46] O.A. Makeeva, A.A. Sleptsov, E.V. Kulish, O.L. Barbarash, A.M. Mazur, E.B. Prokhorchuk, et al., Genomic study of cardiovascular continuum comorbidity, Acta Nat. 7 (2015) 89–99.