Journal of Molecular and Cellular Cardiology 112 (2017) 1–7

Contents lists available at ScienceDirect

Journal of Molecular and Cellular Cardiology

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

Original article

Novel common variants associated with body mass index and coronary artery disease detected using a pleiotropic cFDR method

Wan-Qiang Lv[a], Xue Zhang[b], Qiang Zhang[a], Jing-Yang He[a], Hui-Min Liu[a], Xin Xia[a], Kun Fan[a], Qi Zhao[c,d], Xue-Zhong Shi[a], Wei-Dong Zhang[a], Chang-Qing Sun[a], Hong-Wen Deng[a,d,*]

[a] College of Public Health, Zhengzhou University, No.100 Kexue Road, High-Tech Development Zone of States, Zhengzhou, People's Republic of China
[b] Department of Geriatrics, Renmin Hospital of Wuhan University, Hubei Zhang Road (Formerly Ziyang Road), Wuchang District No. 99 Jiefang Road 238, Wuhan 430060, People's Republic of China
[c] Department of Epidemiology, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA
[d] Center for Genomics and Bioinformatics, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA

ARTICLE INFO

Keywords:
Body mass index
Conditional FDR
Coronary artery disease
Obesity
Pleiotropy

ABSTRACT

Genome-wide association studies (GWAS) have been successfully applied in identifying single nucleotide polymorphisms (SNPs) associated with body mass index (BMI) and coronary heart disease (CAD). However, the SNPs to date can only explain a small percentage of the genetic variances of traits. Here, we applied a genetic pleiotropic conditional false discovery rate (cFDR) method that combines summary statistic *p* values from different multi-center GWAS datasets, to detect common genetic variants associated with these two traits. The enrichment of SNPs associated with BMI and CAD was assessed by conditional Q-Q plots and the common variants were identified by the cFDR method. By applying the cFDR level of 0.05, 7 variants were identified to be associated with CAD (2 variants being novel), 34 variants associated with BMI (11 variants being novel), and 3 variants associated with both BMI and CAD (2 variants being novel). The SNP rs653178 (*ATXN2*) is noteworthy as this variant was replicated in an independent analysis. SNP rs12411886 (*CNNM2*) and rs794356 (*HIP1*) were of note as the annotated genes may be associated with processes that are functionally important in lipid metabolism. In conclusion, the cFDR method identified novel variants associated with BMI and/or CAD by effectively incorporating different GWAS datasets.

1. Introduction

Epidemiological studies have estimated that the prevalence of overweight/obesity increased by approximately 41% between 1980 and 2013, making it a major contributor to the rise in coronary heart disease (CAD) [1]. CAD is one of the leading causes of morbidity and mortality worldwide [1]. Risk factors, particularly obesity, have already had well-established associations with CAD [2].

Epidemiological studies supported that obesity is an independent predictor of clinical CAD [3]. A previous study has demonstrated that every 1 kg/m^2 increase in body mass index (BMI) leads to a 5–7% increase in the incidence of CAD across all BMI categories [4], supporting an inverse relationship between obesity (measured, as conventional, by high BMI) and risk of CAD. Similarly, a large meta-analysis showed that obese participants had a significantly greater risk of CAD (relative risk – RR 1.81, 95% confidence interval – CI 1.56–2.10) after the adjustment for age, sex, physical activities, and smoking [5]. Additional adjustments based on blood pressure and cholesterol levels reduced the RR of obesity to 1.49 (1.32–1.67), but the obesity impact still remained statistically significant. Despite the association between high BMI with an increased occurrence of CAD, studies have also reported the obesity paradox phenomenon that obese patients with established CAD have better clinical outcomes as compared with normal weight patients [6]. Accumulative molecular evidence suggested that obesity might directly be involved in the pathogenesis of CAD [7]. For example, accumulation of body fat could lead to classic metabolic abnormalities [8], including insulin resistance, hyperinsulinemia, hypercholesterolemia, and impaired glucose tolerance, all of which combined would further increase the likelihood of development into CAD [9].

Heritability (h^2) studies demonstrate a substantial genetic contribution to obesity risk ($h^2 \sim 40$–70%) [10] and CAD (male twins: $h^2 \sim 45$–69% and female twins: $h^2 \sim 26$–50%) [11]. However, the identified genes to date only can explain a small percentage of the variances of BMI and CAD [10,11]. Considering the high degree of

* Corresponding author at: Tulane Center of Bioinformatics and Genomics, Department of Biostatistics and Data Science, School of Public Health and Tropical Medicine, Tulane University, New Orleans, LA 70112, USA.
E-mail address: hdeng2@tulane.edu (H.-W. Deng).

http://dx.doi.org/10.1016/j.yjmcc.2017.08.011
Received 24 July 2017; Received in revised form 18 August 2017; Accepted 22 August 2017
Available online 24 August 2017
0022-2828/

heritability in obesity and CAD, more efforts are needed to improve the detection of additional variants that may underlie the "missing heritability". GWAS have the potential to explain a larger proportion of the heritability, mainly by using enlarged larger sample sizes [12]. However, the additional statistical power gained per subject by increasing recruitments of additional study subjects is limited [13], cost-effective analytical methods are therefore needed to fully utilize the existing GWAS datasets. Such methods have recently been developed and successfully applied, including meta-CCA method [13], Genetic analysis incorporating Pleiotropy and Annotation (GPA) method [14], the method by Zhu X and coworkers [15].

The pleiotropic effect is defined as a single gene or variant being associated with more than two distinct phenotypes [16]. Evidence indicated that pleiotropic effect exists in BMI and CAD, which suggests that they may share common genetic variants [17]. By combining the independent GWAS from associated trait and disease of BMI and CAD, the samples sizes are effectively enlarged for detection of the pleiotropic genes [18,19]. Based on pleiotropic effect, statistical power and detection of shared variants will be highly improved by leveraging multi-center GWAS datasets of BMI and CAD.

Recently, a pleiotropy-informed cFDR method is developed with the aim to identify some of the missing heritability [20] with GWAS on individual traits/diseases. So far this method has been successfully applied, *e.g.*, in schizophrenia and cardiovascular disease risk factors [20], and blood pressure and other phenotypes [21] by other groups, and by our own group for height and femoral neck bone mineral density [22], type 2 diabetes and birth weight [23], and for CAD and bone mineral density [24]. This method thus theoretically (21) and practically have been shown to have improved the statistical power and improved variants discovery in the studied associated traits or diseases. Here, we applied the genetic pleiotropy-informed cFDR method on summary statistics of two independent meta-GWAS to identify shared variants and pleiotropic effect between BMI and CAD. By using this method, we hypothesized that we could identify more common variants for BMI and/or CAD, and discover some novel etiologic relationship between BMI and CAD.

2. Materials and methods

2.1. GWAS datasets

The dataset for BMI was downloaded from http://portals.broadinstitute.org/collaboration/giant/index.php/GIANT_consortium_data_files. This GWAS meta-analysis compromising of 249,796 individuals of white European Ancestry was performed by the Genetic Investigation of Anthropometric Traits (GIANT) Consortium [19]. Two datasets for CAD were downloaded from http://www.cardiogramplusc4d.org/data-downloads. The C4D dataset performed by the Coronary Artery Disease (C4D) Genetics Consortium was derived from a meta-analysis of four large GWAS of European and South Asian descent involving 15,420 CAD cases and 15,062 controls [25]. The second dataset conducted by the transatlantic Coronary ARtery Disease Genome-wide Replication and Meta-analysis (CARDIoGRAM) Consortium was derived from a meta-analysis of 22 GWAS of European descent comprising of 22,233 cases and 64,762 controls [18]. All the datasets consist of summary statistics for each SNP based on each meta-analysis publication, providing *p* values after using genomic control (GC) both at the individual study level and after meta-analysis [26]. Further details of the GWAS samples and methods employed within each group were presented in the original references [18,19,25]. Additionally, the CARDIoGRAM dataset for CAD in our analysis was used as the replication dataset. This study used only summary statistics from publically available datasets for previous GWAS. It does not involve human subjects directly. Informed consent was obtained from all participants of contributing studies in the published respective GWAS. Contributing studies received ethical approval from their respective institutional review boards. This study was approved by the Ethical Committee of the Life Sciences of Zhengzhou University.

2.2. Data preparation

Before the analysis, we checked whether there were overlapped samples included in these datasets of the cohorts. We found no individuals were overlapped between C4D and GIANT datasets, and no overlapped individuals between CARDIoGRAM and C4D datasets. However, the datasets used for BMI and CAD had different ancestors, *i.e.*, European Ancestry for BMI and Europeans and South Asians for CAD (45% of the CAD case are Asians).

When dealing with the various datasets, we combined the two phenotypes' summary *p* values for the common SNPs studied in both datasets. After annotating the common SNPs, we applied a LD-based pruning method to remove the large correlations between pairs of variants. The minor allele frequency (MAF) was used as a criterion in the SNP pruning method, which removed the SNP with smaller MAF for pairs with $R^2 > 0.2$. The datasets were pruned using the HapMap3 genotypes of the corresponding matched ethnicity references. First, this method proceeds by using 50 SNPs as a group where LD is calculated between each pair of SNPs. SNPs with smaller MAF were removed from our analysis if their measured LD had an $R^2 > 0.2$. Then, move forward by 5 SNPs and the process is repeated until there are no pairs of SNPs that are > 0.2. This pruning method ensures that SNPs are not in LD with each other in the follow-up analysis. It is suggested that complex correlations among the test-statistics may bias the estimate of the conditional FDR, including LD score regression [27] and the effect of correlation in FDR estimation [28]. There are two regions with complex LD structures [29], including the extended major histocompatibility complex (MHC) (chr6:25652429–33368333) and chromosomal region 8p23.1 (chr8:7242715–12483982). Thus, we constructed conditional Q-Q plots and the cFDR analysis after excluding SNPs within these regions to remove potential bias introduced by them.

2.3. Statistical analysis

2.3.1. Genomic control (GC)

Population stratification can be a problem for association studies, where the association could be found due to the underlying structure of the population and not a disease associated locus. GC is one of the most widely used approaches to correct for this problem, which controls the inflation of test statistics due to population structures [30]. GC works by using markers that are not linked with the trait in question to correct for any inflation of the test statistic caused by population stratification [26]. GC has been applied in the original GWAS at the individual study level and for the meta-analysis, there is no need for us to repeat it here in our analyses.

2.3.2. Conditional Q-Q plots for accessing pleiotropic effect enrichment

Q-Q plots are a descriptive tool for visualizing the difference between observed distribution and theoretical distribution. In the analysis of GWAS, quantiles of the observed *p*-values (nominal), denoted by 'p', are plotted on the y-axis, and quantiles of the theoretical null distribution (the uniform distribution estimated by the empirical cumulative distribution function), here denoted by 'q', are plotted on the x-axis. Commonly, the negative log transformation was used, we denoted the y-axis as nominal $-\log10$ (p), and the x-axis as empirical $-\log10$ (q). The enrichment of pleiotropic effect is graphically accessed by conditional Q-Q plots [20]. Under the null hypothesis, enrichment of pleiotropic effect can be reflected by leftward deflections of the observed distribution from the null line. If the Q-Q plot falls on the line $x = y$, with no deviation between lines, this would indicate no enrichment of genetic pleiotropic effect. If pleiotropic enrichment does exist, an earlier leftward shift from the null line will be present. Larger spacing in the Q-Q plots is interpreted as a greater extent of pleiotropic