Pathway-Based Genome-Wide Association Studies for Plasma Triglycerides in Obese Females and Normal-Weight Controls

Hongxiao Jiao[1☯], Kai Wang[2☯], Fuhua Yang[1☯], Struan F. A. Grant[3,4], Hakon Hakonarson[3,4], R. Arlen Price[5]*, Wei-Dong Li[1,5]*

1 Research Center of Basic Medical Sciences, Tianjin Medical University, Tianjin, 300070, China, 2 Zilkha Neurogenetic Institute and Norris Comprehensive Cancer Center, University of Southern California, Los Angeles, CA, 90089, United States of America, 3 Center for Applied Genomics, Children's Hospital of Philadelphia, Philadelphia, PA, 19104, United States of America, 4 Department of Pediatrics, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, United States of America, 5 Center for Neurobiology and Behavior, Department of Psychiatry, University of Pennsylvania Perelman School of Medicine, Philadelphia, PA, 19104, United States of America

☯ These authors contributed equally to this work.
* liweidong98@tijmu.edu.cn (WDL); arlen@exchange.upenn.edu (RAP)

CrossMark
click for updates

OPEN ACCESS

Abstract

Pathway-based analysis as an alternative approach can provide complementary information to single-marker genome-wide association studies (GWASs), which always ignore the epistasis and does not have sufficient power to find rare variants. In this study, using genotypes from a genome-wide association study (GWAS), pathway-based association studies were carried out by a modified Gene Set Enrichment Algorithm (GSEA) method (GenGen) for triglyceride in 1028 unrelated European-American extremely obese females (BMI$\geq$35kg/m^2) and normal-weight controls (BMI<25kg/m^2), and another pathway association analysis (ICSNPathway) was also used to verify the GenGen result in the same data. The GO0009110 pathway (vitamin anabolism) was among the strongest associations with triglyceride (empirical P<0.001); the result remained significant after FDR correction ($P = 0.022$). MMAB, an obesity-related locus, included in this pathway. The ABCG1 and BCL6 gene was found in several triglyceride-related pathways (empirical P<0.05), which were also replicated by ICSNPathway (empirical P<0.05, FDR<0.05). We also performed single-marked GWAS using PLINK for TG levels (log-transformed). Significant associations were found between ASTN2 gene SNPs and plasma triglyceride levels (rs7035794, $P = 2.24 \times 10^{-10}$). Our study suggested that vitamin anabolism pathway, BCL6 gene pathways and ASTN2 gene may contribute to the genetic variation of plasma triglyceride concentrations.

Introduction

Genome-wide association studies (GWASs) have rapidly become a powerful method for genetic studies in complex diseases. Many disease-related genes or loci have been identified with hundreds of thousands of common variants. Nevertheless, the variants identified by GWASs capture only a minor fraction of disease heritability [1, 2]. Many variants with modest associations are often ignored after multiple testing correction in GWASs[3]. Researchers can expand sample size to identify more associated loci, however, this approach will have huge cost and diminishing returns. The limitations of GWAS make it difficult to use this approach to find rare variants and epistasis [2, 4]. Pathway-based analysis is an alternative approach that detects trait-associated loci in GWAS data, which can provide complementary information to single-marker analysis, such as providing additional biological insights and highlighting new candidate genes[5, 6].

Pathway-based approaches are based on the principle that the genes always collected from the same biological or functional pathway interact with each other and constitute a network[7, 8]. Note that the context "pathway" means a gene set, rather than an interconnected biological process. Gene Set Enrichment Analysis(GSEA)[9] is a pathway-based approach, which can used to measure how much association signals are enriched in a gene set defined by known biological knowledge of genes and pathways. Wang et al. developed the GSEA-based pathway analysis, the modified GSEA method (GenGen) [10]. This approach has been successfully applied to uncover many disease-related pathways, such as the IL12/IL23 pathway associated with Crohn's disease[11], the Vasoactive Intestinal Peptide (VIP) pathway important for Obesity[12], as well as the WNT-signaling pathway in Type2 diabetes(T2D)[13]. Unlike the above disease studies, our study focus on a quantitative trait, plasma triglyceride (TG) levels, with previous American obesity GWAS data, to examine whether TG-related gene set can be enriched.

As one of three major lipid phenotypes in the human serum, TG is a heritable trait that is a risk factor for cardiovascular disease [14, 15], insulin resistance, obesity are also characterized by increased plasma concentration of TG-rich lipoproteins[13, 14, 16]. TG level largely controlled by genetic factors, many genetic variants have been identified by GWASs. The Global Lipids Genetics Consortium(GLGC) has conducted two large-scale GWASs to identify genetic loci associated with lipid traits, and provided ~40 TG-associated genes[17, 18]. However, genetic variation at these loci explains only about 11.7% of overall TG variation within the population, corresponding to approximately 40–60% of the heritability for plasma TG levels [18–21]. Therefore, it is likely that some of genes regulating TG levels remain to be undiscovered.

In this study, we used the modified GSEA (GenGen) with extremely obese individuals and normal-weight controls, performed pathway-based association analyses to find additional susceptibility loci and pathways. In addition to GenGen, ICSNPathway (Identify candidate Causal SNPs and Pathways) [22] provides a feasible solution to bridge the gap between GWAS and disease mechanism study by generating hypothesis of SNP → gene → pathway(s). Our findings reveal several gene sets are associated with the plasma TG level, these genes may contribute to the heritability of TG.

Results

Single-marker association analysis for plasma TG levels

We examined an obesity cohort for TG levels, which previously genotyped on Illumina HumanHap550 SNP arrays for a GWAS for body weight traits[23]. TG levels were transformed by logarithms [log(TG)], and than linear regressions were carried out between log(TG) and

Table 1. Distribution of triglyceride levels in European-American subjects.

	N	Minimum	Maximum	Mean	SD*	Skewness	Kurtosis
Log(TG) **	1022	1.53	2.84b	2.08	0.25	0.28	-0.31
BMI-log(TG) ***	1022	-2.5	3.82	0.00	1.00	0.45	0.26

* Outliers (>±3 SD) were deleted in this study.

**Log(TG), log-transformed triglyceride levels.

***BMI-log(TG), BMI-adjusted log-transformed triglyceride levels.

doi:10.1371/journal.pone.0134923.t001

BMI, written BMI-adjusted log(TG), we performed single-marked association analysis for the two quantitative traits, log(TG) and BMI-adjusted log(TG), in 1022 samples after extreme values (>3 SD) were deleted. Distributions of log(TG) and BMI-adjusted log(TG) levels in all samples were shown in Table 1, their ranges were 1.53–2.84 and -2.5–3.82, respectively. Q-Q plots showed distributions of log(TG) and BMI-adjusted log(TG) (Fig 1), denoted that the two phenotypes in line with normal distribution.

Three loci on 9q33 reached genome-wide significance levels of $P < 5 \times 10^{-8}$, with the most significant SNP rs7035794 (log(TG), $P = 2.24 \times 10^{-10}$; BMI-adjusted log(TG), $P = 6.24 \times 10^{-8}$) within the ASTN2 (astrotactin 2) gene. Other three SNPs in the ASTN2 gene, rs1929010, rs4091697 and rs4333683, reached $P < 1 \times 10^{-6}$ (Table 2, Fig 2). rs2420511 ($P = 7.38 \times 10^{-8}$) and rs3817859 ($P = 1.67 \times 10^{-7}$) in C1orf112 were also associated with BMI-adjusted log(TG). Manhattan plots for GWASs of logTG and BMI-adjusted logTG were shown as (Fig 2). ASTN2 is a neuronal adhesion-related gene [24], associated with Schizophrenia[24, 25], the ASTN2 association with TG was not reported before this study, although ASTN2 gene SNPs were nominally associated with total cholesterol and LDL in a GWAS for lipid phenotypes in the Framingham Heart Study[26].

No genome-wide association ($P < 5 \times 10^{-8}$) was reached for binary TG (S1 Table), although "top" associations (including CDKN2A/B, PABPC4L, and ABCG1) were associated with lipid and/or body weight related phenotypes[27], or had direct biology connections with cholesterol transport.
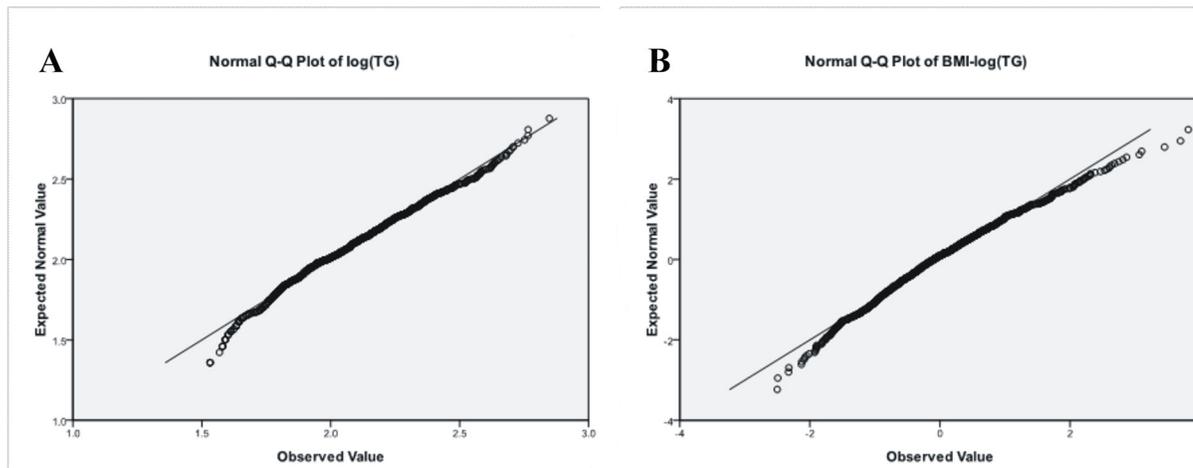

Fig 1. Q-Q plots of log(TG) (A) and BMI-adjusted log(TG)(B) levels in all subjects.

doi:10.1371/journal.pone.0134923.g001

Table 2. Quantitative association studies (PLINK) for log(TG) ($P<10^{-6}$).

CHR	SNP	BP	MAF	GENE	P[log(TG)]	P[BMI-log(TG)]
9	rs7035794	118692318	0.058	ASTN2	2.24×10^{-10}	6.24×10^{-8}
9	rs10118539	118695771	0.077	ASTN2	2.38×10^{-9}	1.00×10^{-7}
9	rs4836907	118706675	0.072	ASTN2	3.79×10^{-9}	1.51×10^{-7}
9	rs1929010	118728015	0.068	ASTN2	7.45×10^{-7}	1.80×10^{-6}
9	rs4091697	118694342	0.093	ASTN2	1.52×10^{-7}	8.04×10^{-6}
9	rs4333683	118677830	0.074	ASTN2	3.68×10^{-7}	3.33×10^{-5}
1	rs2420511	168079207	0.087	C1orf112	1.20×10^{-6}	7.38×10^{-8}
1	rs3817859	168085689	0.089	C1orf112	1.99×10^{-6}	1.67×10^{-7}

doi:10.1371/journal.pone.0134923.t002

Pathway-based analyses for plasma TG level

To further examine novel biological pathways or collections of functionally related genes that associated with TG levels, we carried out two binary pathway-based analyses: the modified GSEA (GenGen) [6] and ICSNPathway on this GWAS data, with the same subjects and SNPs used in single-marker association analysis. In 2001, the National Cholesterol Education Program (NCEP) released recommendations on triglyceride levels that Normal triglyceride concentration is less than 150 mg/dL, borderline is 150 to 199 mg/dL, high is 200 to 499 mg/dL, and concentrations of 500 mg/dL or higher are considered very high. In this study, we defined TG levels > 200 mg/dl as "cases", N = 203, and TG levels < 150 mg/dl as "controls", N = 667 (Fig 3). Deleted the borderline value, a total of 870 samples were used for the next binary pathway association analyses. Average ages of cases and controls were 42.6±9.6 years and 42.3±8.9 years, respectively. Distributions of TG in cases and controls are shown in Table 3.
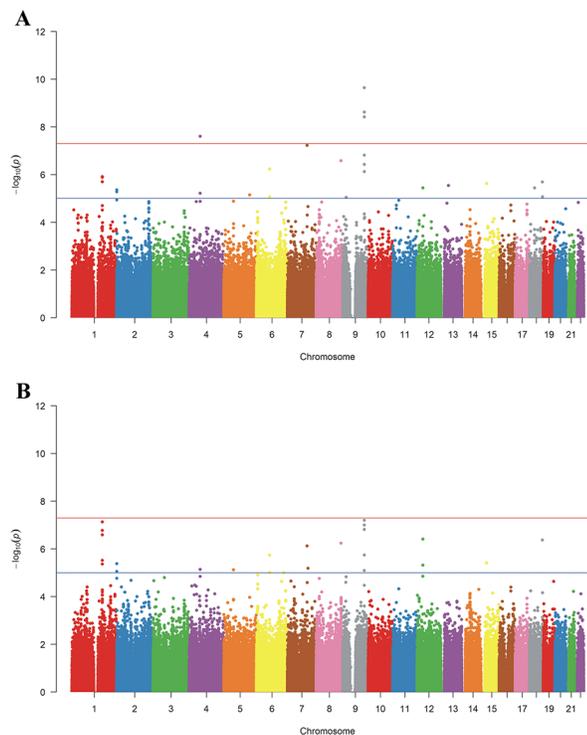

Fig 2. Manhattan plots for GWAS of logTG (A) and BMI-adjusted log TG (B).

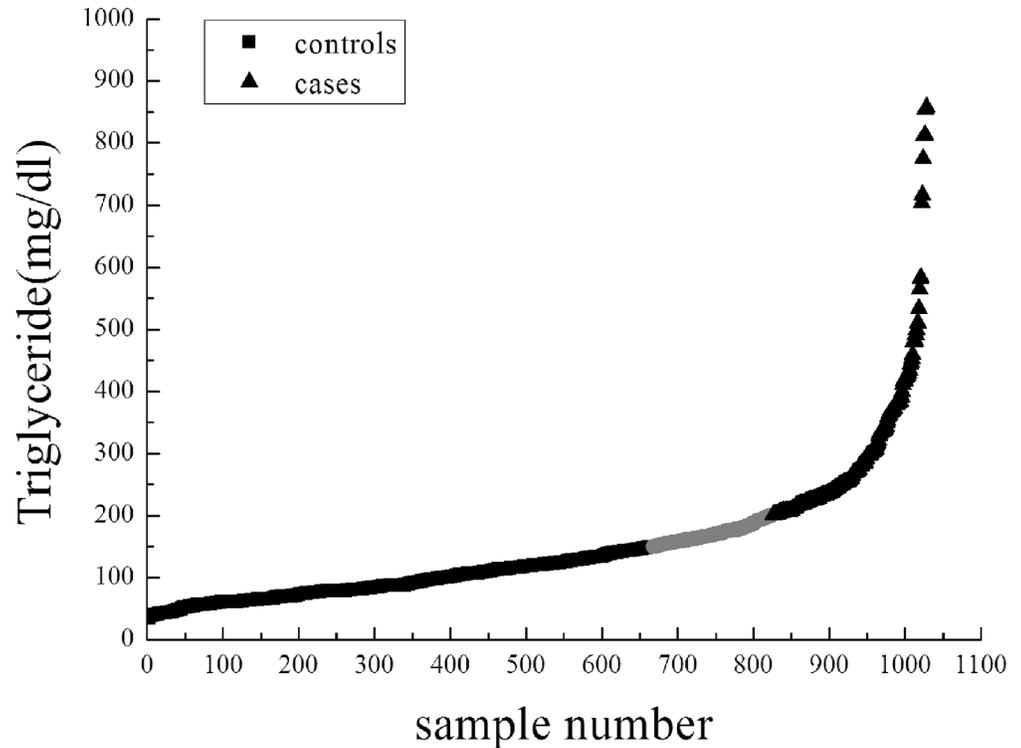doi:10.1371/journal.pone.0134923.g002

Fig 3. Scatter diagram of triglyceride in 1028 samples. Arranged in accordance with triglyceride levels from low to high, TG>200mg/dl as cases (N = 203) marked by black boxes, TG<150mg/dl as controls (N = 667) marked by black triangle, 150mg/dl<TG<200mg/dl as gray area were ignored in our pathway-based analysis by modified GSEA (GenGen).

doi:10.1371/journal.pone.0134923.g003

In the process of pathway analysis, the case/control labels were permuted 1000 times in our study, and for each permutation, the association test statistics for all SNP markers were recalculated, and then re-evaluation the signifcance of the pre-defined gene sets by comparing the observed test statistics with the null distribution generated by the permutations. Total 1347 gene sets were assessed, which passed the size threshold (5–200 genes), and retrieved from Bio-Carta, KEGG(Kyoto Encyclopedia of Genes and Genomes Pathway), and GO (Gene Ontology) databases. We found that twenty-three pathways were associated with TG-levels at an empirical *P*-value less than 0.01 in the pathway-based analysis by GenGen (**S2 Table**). Among these pathways, GO0009110 (vitamin anabolism) showed the most significant association with TG level (empirical P<0.001) and have false discovery rate (FDR) less than 0.05 passed multiple testing corrections (FDR = 0.022), and another eight pathways have FDR less than 0.2 (**Fig 4**). There were 19 genes in the vitamin anabolism pathway, including MMAB, PNPO, PDXK, and ME1 (**S3 Table**). The SNP rs7953794 locate in the MMAB gene yielded moderate single locus association with TG (*P* = 0.016), MMAB is a HDL-cholesterol-related gene [28]. We need to

Table 3. Distribution of triglyceride levels in "cases" (TG>200mg/dL) and "controls" (TG<150mg/dL) for binary GWAS and pathway association analyses.

	N	Age	Triglycerides (mg/dL)	Maximum	Minimum	Mean	SD
Cases	203	42.6±9.6	>200	858	201	307.4	126.3
Controls	667	42.3±8.9	<150	149	34	92.4	30.0

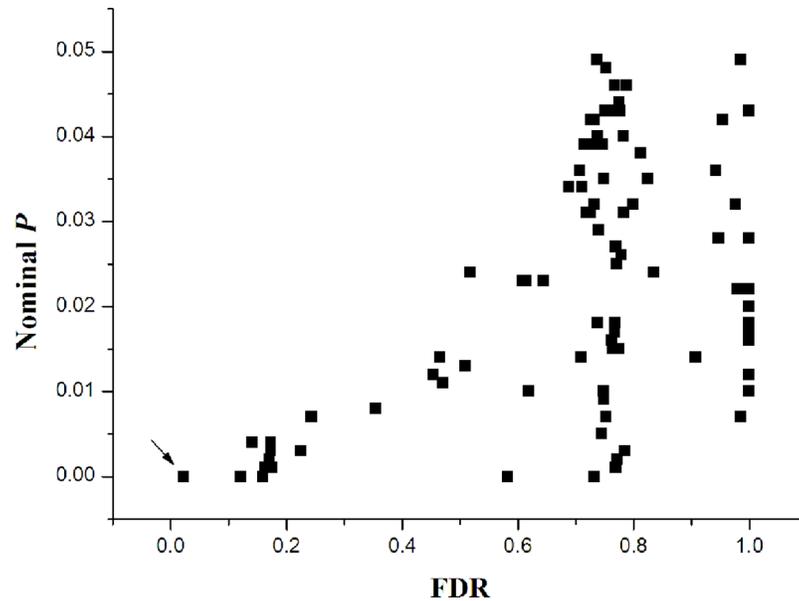doi:10.1371/journal.pone.0134923.t003

Fig 4. The distribution of empirical *P*-FDR for triglyceride. Empirical *P*-FDR for triglyceride related pathways (empirical *P*<0.05, denoted as "nominal" *P* values in the GenGen program) obtained by modified GSEA (GenGen), GO0009110 pathway is indicated by the arrow.

point out that the "nominal" *P* values in the GenGen program[10] were actually empirical *P* values based on permutations, therefore we used "empirical *P*" for our GenGen results.

We further used ICSNPathway to analyze pathway associations in the same data (S4 Table). In this method, a SNP *P*-value file (from GWAS) and a gene set file were needed. In order to reduce the external differences between the ICSNPathway and the modified GSEA (GenGen), SNP *P*-value file was obtained from binary association analyses by PLINK, the same gene set file was used as GenGen. The results of ICSNPathway were shown as (Fig 3). Nine pathways, which were among the most significant TG-associated pathways in GenGen, also reached the *P*-value<0.01 and FDR less than 0.2. Vitamin anabolism pathway had the empirical *P* = 0.002. GO0022618 (Protein-RNA complex assembly), GO0051320 (S-phase), GO0016447 (somatic recombination of immunoglobulin gene segments), and GO0045930 (negative regulation of mitotic cell cycle) were further verified to be associated with TG by ICSNPathway analyses after multiple testing corrections (nominal *P*<0.01, FDR<0.05). The BCL6 gene (B-cell CLL/ lymphoma 6) was found in 6 of the top 9 TG-associated pathways, GO0051320, GO0016447, GO0045930, GO0002449 (lymphocyte mediated immunity), GO0007346 (regulation of mitotic cell cycle), and GO0002460 (adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin super-family domains).

Discussion

In the single-marker GWAS and pathway-based association studies, we identified the gene ASTN2 and the "vitamin anabolism" pathway were significantly associated with TG levels, respectively.

The neuronal adhesion-related gene ASTN2 (astrotactin 2) is expressed in the brain and has a domain structure similar to that of ASTN1; the ASTN2 + ASTN1 protein complex is important for proper cell-surface expression of ASTN1 [29]. ASTN2 is involved in neuronal adhesion and performs a key role in neural development [30, 31]. Neuronal pathways can produce an

effect on insulin sensitivity [32], and contribute to insulin resistance syndrome components in humans, especially for type 2 diabetes and obesity [33]. In our study, neuronal adhesion-related gene ASTN2 was associated with plasma TG levels. A previous linkage analysis for lipid-related latent gene-expression quantitative traits with metabolic syndrome found that ASTN2 was associated with lipid levels [34].

Traditional GWASs (also called single-marker GWASs) have many limitations, including the inability to identify many "minor" genes. In our single-marker association analysis, only one gene, ASTN2, reached genome-wide significance levels of $P<5\times10^{-8}$. With the purpose of collecting more TG-related genes, we carried out pathway-based analysis on our GWAS data (Table 4), with the same subjects and SNPs used in previous single-marker GWASs. Many gene sets were associated with TG levels through the GenGen, especially the vitamin anabolism gene set.

GO0009110 is a vitamin anabolism pathway, the interactions among the genes exist in this pathway are related to triglyceride. MMAB [methylmalonic aciduria (cobalamin deficiency) cblB type] encodes a protein that catalyzes the conversion of vitamin B_{12} into adenosylcobalamin, an active coenzyme form of B_{12}. HDL-cholesterol-associated SNP (rs7298565) is associated with higher MMAB mRNA and protein levels [28], MMAB_3U3527G→C variant also contribute to variation of HDL-cholesterol concentrations [35]. In our whole genome linkage analysis, the chromosome region 12q23-24 yielded significant linkage (LOD score = 4.08) for percentage fat [36]. The MMAB locus was only 50 kb away from the linkage peak D12S1339.

MMAB was associated with TG levels, which can be interpreted by examining its function. On the one hand, methylmalonyl-CoA mutase, as an adenosylcobalamin-dependent enzyme, can catalyze the 1,2-rearrangement of methylmalonyl-CoA to succinyl-CoA [37]. Succinyl-CoA joins the tricarboxylic acid cycle, which is the common metabolic pathway of carbohydrates, lipids, and amino acids. MMAB can affect TG levels through adenosylcobalamin and methylmalonyl-CoA mutase. Moreover, plasma B_{12} correlates inversely with homocysteine, which is an intermediate product of methionine metabolism [38]. B_{12} deficiency is a common cause of hyperhomocysteinemia [39]. Homocysteine is associated with plasma TG levels. Geoff et al. found that homocysteine-induced endoplasmic reticulum stress causes dysregulation of the endogenous sterol response pathway and leads to uptake of TG [40].

BCL6 was present in many TG-associated pathways, especially in the three pathways of the ICSNPathway analysis that passed multiple testing correction (nominal $P<0.001$, FDR<0.05). BCL6 is a transcriptional repressor which frequently disrupted by translocations in B-cell

Table 4. Top pathways that associated with binary TG stratification (case-control).

Pathway ID	Description	Set Size	GenGen		ICSNPathway		Database
			empirical P	FDR	Nominal P	FDR	
GO0009110	Vitamin anabolism	19	**<0.001**	**0.022**	0.002	0.157	GO
GO0022618	Protein-RNA complex assembly	43	<0.001	0.121	**<0.001**	**0.02**	GO
GO0051320	S-phase	13	<0.001	0.159	**<0.001**	**0.005**	GO
GO0002449	Lymphocyte mediated immunity	46	0.001	0.163	0.006	0.181	GO
GO0007346	Regulation of mitotic cell cycle	56	0.001	0.176	0.007	0.184	GO
GO0016447	Somatic recombination of immunoglobulin gene segments	24	0.002	0.170	**<0.001**	**0.012**	GO
GO0045930	Negative regulation of mitotic cell cycle	14	0.003	0.173	**<0.001**	**0.013**	GO
GO0005178	Integrin binding	47	0.004	0.173	0.001	0.156	GO
GO0002460	Adaptive immune response based on somatic recombination of immune receptors built from immunoglobulin superfamily domains	57	0.004	0.141	0.053	0.359	GO

doi:10.1371/journal.pone.0134923.t004

lymphomas. BCL6 can repress the transcription of BCL6 target genes, mainly involved in cell activation and differentiation, cell cycle regulation, and inflammation [41, 42]. Previous studies have demonstrated that the FOXO/BCL6/cyclin D2 pathway linked to β-cell proliferation and may therefore be considered to be associated with diabetes [43, 44]. In addition, the BCL6-SMRT and BCL6-NcoR cistromes were reported to be able to repress NF-κB-driven inflammatory immune responses [45]. Free BCL6 can attenuating inflammatory gene expression through suppressing MCP-1 in the kidney, and related to anti-inflammatory effects. For patients with severe hypertriglyceridemia, anti-inflammatory drug therapy significantly reduces TG levels [46]. Studies also have shown that elevation of TG-rich lipoprotein can induce endothelial cell inflammation [47, 48]. TG levels are usually increased in obese and diabetic individuals, and BCL6 is related to diabetes and obesity through the above description. BCL6 is associated with TG levels in our results.

Many genes were enriched in our pathway-based analysis that probably relate to TG. Several diabetes-related or inflammation-related genes are included in the TG-associated pathways, such as SMAD3 [49], TGFB1 [50], CDKN2B [13], and IL10 [51]. Needless to say, further analyses are needed to decipher the interaction among those genes.

In this study, we performed pathway-based studies using two different methods to better verify our findings. Vitamin anabolism related pathways were associated with plasma TG levels, which therefore might account for TG-related cardiovascular events. Larger sample sizes are needed in future pathway association studies in different populations to verify this result.

Materials and Methods

Subjects

These samples were originally collected to study obesity, and they were used to analyze associations for TG in this study. One thousand and twenty eight (1028) unrelated European Americans were chosen from an ongoing study, comprising 490 extremely obese females (BMI>35 kg/m^2) and 538 normal-weight controls (BMI<25 kg/m^2). All cases were obese probands, selected from obese families and trios, and unrelated normal weight controls were selected who had a current and lifetime BMI<25 kg/m^2[23]. Clinical characteristics have been described previously [52, 53].

Note that these samples were collected originally in order to investigate obesity genes in female subjects, so we have a small fraction of males during the recruitment.

All participants gave written informed consent, and the research protocol was approved by the Institutional Review Board (IRB) on Studies Involving Human Beings at the University of Pennsylvania.

Phenotypes

Blood samples were obtained after subjects had fasted overnight (>6 h). Plasma TG levels were measured by Quest Diagnostics (Philadelphia, PA). The standard formula, Weight (kg) divided by Height (m^2), was used calculated Body mass index. Height was measured by a standing position using a stadiometer. Weight was measured by a scale with a maximum weight of 600 pounds (270 kg). All measurements were taken when subjects dressed in light clothing. Log-transformed TG [log(TG)] and BMI-adjusted log(TG) were the phenotypes of association analyses. For BMI-adjusted log(TG), linear regressions were carried out between log(TG) and BMI (SPSS, version 17.0), with BMI as independent variable, log(TG) as the dependent variable, and standardized residuals saved to make mean = 0 and standard deviation = 1. Threshold selected binary triglycerides were used for discrete GWAS and GenGen analyses: individuals with TG>200mg/dL were set as "cases" and those with TG<150mg/dL

were used as "controls", affection statuses for those with 150mg/dL≤TG≤200mg/dL were considered as "unknown".

Genotyping

Genomic DNA was extracted from peripheral blood using a high-salt method [54] and diluted to 10 ng/μl. Genotyping was performed for our previous GWAS for body weight traits[23]. In brief, Illumina HumanHap550 SNP arrays (Illumina, San Diego, CA), with about 550,000 SNP markers, were used to genotype DNA samples at the Center for Applied Genomics, Children's Hospital of Philadelphia. Standard Illumina data normalization procedures and canonical genotype clustering files were used to process the genotyping signals. Hardy-Weinberg equilibrium (HWE) was tested for all SNPs in the array, SNPs with genotype frequencies that depart from HWE were deleted.

Statistical analyses

Basic statistical descriptions were performed using SPSS 17.0. TG (log-transformed) outliers (>3 SD) were excluded from the data set, 6/1028 samples were removed, and then linear regression was carried out between log(TG) and BMI. Genome-wide quantitative association analyses were performed by PLINK 1.07 [55] for log(TG) and BMI-adjusted log(TG). SNPs with minor allele frequencies (MAF) <1% were excluded from quantitative association studies. We also performed a discrete GWAS for triglyceride using threshold-selected binary TG (TG>200mg/dL vs. TG<150mg/dL) (**Table 3**).

The pathway-based genome-wide studies were divided into two steps, firstly, we used a modified Gene Set Enrichment Analysis (GSEA) method (GenGen) developed by Wang et al. [10] to carry out binary association analysis; Secondly, we used ICSNPathway [22] to verify our pathway association results obtained by GenGen.

Gene set enrichment analysis

This method was used performed pathway-based test for genome-wide association data. The main process was as follows. For each gene, the SNP with the highest test statistic (chi-square detection/F-test) among all SNPs mapped to the gene was selected to represent the gene [10]. All genes were ranked by sorting their statistic values from the largest to smallest, denoted by r(1),..., $r(N)$, where N represents the total number of genes. For any given gene set S composed of N_H genes, an enrichment score (ES) was calculated, which was a weighted Kolmogorov-Smirnov-like running sum statistic that reflects the overrepresentation of genes within S at the top of the entire ranking list of genes in the genome:

$$\text{ES}(S) = \max_{1 \leq j \leq N} \left\{ \sum_{G_{j^*} \in S, j^* \leq j} \frac{|r_{(j^*)}|^p}{N_R} - \sum_{G_{j^*} \notin S, j^* \leq j} \frac{1}{N - N_H} \right\},$$

Where $N_R = \sum_{G_{j^*} \in S} |r_{(j^*)}|^p$ and p is a parameter that gives higher weight to genes with extremes statistic values (default $P = 1$). The phenotype label were permutated 1000 times to adjust the size of different genes. For each permutation, enrichment scores were calculated. We then calculated the normalized enrichment score (NES):

$$\text{NES} = \frac{\text{ES}(S) - \text{mean}[\text{ES}(S, \pi)]}{SD[\text{ES}(S, \pi)]}.$$

Finally, a false-discovery rate (FDR) procedure was conducted to control the fraction of expected false-positive findings:

$$\text{FDR} = \frac{\%\text{of all } (S, \pi) \text{ with } \text{NES}(S, \pi) \geq NES*}{\%\text{of observed } S \text{ with } \text{NES}(S) \geq NES*}.$$

where NES^* denotes the normalized ES in the observed data. Approximately 520K SNPs passed the initial quality-control threshold in the analysis of GenGen(defined as minor-allele frequency > 0.01 and Hardy-Weinberg equilibrium P-value > 0.001), which covered 17,438 genes; 20k bp upstream and downstream of each gene was considered to be a part of the gene. We retrieved 301 annotated pathways from the BioCarta database and 212 annotated pathways from Kyoto Encyclopedia of Genes and Genomes Pathway database and constructed 2,058 gene sets on the basis of Gene Ontology (GO) annotation files, which were downloaded from the GO website. We also limited testing to those pathways that contained between 5 and 200 genes represented by markers in our GWAS database. Thus, a total of 1347 pathways were analyzed in this analysis.

Identify Candidate Causal SNPs and Pathways (ICSNPathway)

The ICSNPathway web server implements a two-stage analysis [22]: in the first stage, the candidate causal SNPs are pre-selected according to linkage disequilibrium (LD) analysis and functional SNP annotation based on the most significant SNPs of GWAS to represent the gene; in the second stage, the biological mechanisms for the pre-selected candidate causal SNPs are annotated by the Improved Gene Set Enrichment Analysis (i-GSEA) [56], which also needs to calculate ES. SNP label permutations, however, were implemented instead of phenotype label permutations. Based on all the distributions of ESs generated by permutation, FDR is used for multiple testing corrections. In short, ICSNPathway integrates LD analysis, functional SNP annotation, and pathway-based analysis to identify candidate causal SNPs and their corresponding candidate causal pathways from GWAS data sets.

In this analysis, SNP P-values were obtained from binary association analyses by PLINK. We collected all SNPs for the next test. Other main parameters included: (1) LD cutoff: $r^2 > 0.6$, (2) distance for searching LD neighborhoods: 200 kb, (3) rule of mapping SNPs to genes: 20 kb upstream and downstream of gene, and (4) gene set database: same to the GenGen gene sets files, with pathways containing <5 or >200 genes ignored. Of the 2571 pathways, 1359 passed the filtering criteria.

Supporting Information

S1 Table. Associations ($P<1x10^{-4}$) for binary GWAS for triglyceride.
(DOC)

S2 Table. The binary pathway-based association analysis for TG by GenGen. (empirical $P<0.05$).
(DOCX)

S3 Table. Genes in triglyceride-related pathways.
(DOC)

S4 Table. The URL of the result analyzed by ICSNPathway.
(DOC)

Acknowledgments

Author Contributions

References

1. Maher B. Personal genomes: The case of the missing heritability. Nature. 2008; 456(7218):18–21. doi: 10.1038/456018a PMID: 18987709

2. Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, Hunter DJ, et al. Finding the missing heritability of complex diseases. Nature. 2009; 461(7265):747–53. doi: 10.1038/nature08494 PMID: 19812666

3. Wang LL, Jia PL, Wolfinger RD, Chen X, Zhao ZM. Gene set analysis of genome-wide association studies: Methodological issues and perspectives. Genomics. 2011; 98(1):1–8. doi: 10.1016/j.ygeno.2011.04.006 PMID: 21565265

4. Yaspan BL, Veatch OJ. Strategies for pathway analysis from GWAS data. Current protocols in human genetics / editorial board, Jonathan L Haines [et al]. 2011;Chapter 1:Unit1 20.

5. Elbers CC, van Eijk KR, Franke L, Mulder F, van der Schouw YT, Wijmenga C, et al. Using genome-wide pathway analysis to unravel the etiology of complex diseases. Genetic epidemiology. 2009; 33 (5):419–31. doi: 10.1002/gepi.20395 PMID: 19235186

6. Wang K, Li M, Hakonarson H. Analysing biological pathways in genome-wide association studies. Nature reviews Genetics. 2010; 11(12):843–54. doi: 10.1038/nrg2884 PMID: 21085203

7. Lee D, Lee GK, Yoon KA, Lee JS. Pathway-Based Analysis Using Genome-wide Association Data from a Korean Non-Small Cell Lung Cancer Study. PloS one. 2013; 8(6).

8. Schadt EE. Molecular networks as sensors and drivers of common human diseases. Nature. 2009; 461 (7261):218–23. doi: 10.1038/nature08454 PMID: 19741703

9. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences of the United States of America. 2005; 102(43):15545–50. PMID: 16199517

10. Wang K, Li M, Bucan M. Pathway-based approaches for analysis of genomewide association studies. American journal of human genetics. 2007; 81(6):1278–83. PMID: 17966091

11. Wang K, Zhang H, Kugathasan S, Annese V, Bradfield JR, Russell RK, et al. Diverse Genome-wide Association Studies Associate the IL12/IL23 Pathway with Crohn Disease. American journal of human genetics. 2009; 84(3):399–405. doi: 10.1016/j.ajhg.2009.01.026 PMID: 19249008

12. Liu YJ, Guo YF, Zhang LS, Pei YF, Yu N, Yu P, et al. Biological Pathway-Based Genome-Wide Association Analysis Identified the Vasoactive Intestinal Peptide (VIP) Pathway Important for Obesity. Obesity. 2010; 18(12):2339–46. doi: 10.1038/oby.2010.83 PMID: 20379146

13. Perry JR, McCarthy MI, Hattersley AT, Zeggini E, Weedon MN, Frayling TM. Interrogating type 2 diabetes genome-wide association data using a biological pathway-based approach. Diabetes. 2009; 58 (6):1463–7. doi: 10.2337/db08-1378 PMID: 19252133

14. Miller M. The epidemiology of triglyceride as a coronary artery disease risk factor. Clinical cardiology. 1999; 22(6 Suppl):II1–6. PMID: 10376190

15. Patel A, Barzi F, Jamrozik K, Lam TH, Ueshima H, Whitlock G, et al. Serum triglycerides as a risk factor for cardiovascular diseases in the Asia-Pacific region. Circulation. 2004; 110(17):2678–86. PMID: 15492305

16. Wei EH, Ali YB, Lyon J, Wang HJ, Nelson R, Dolinsky VW, et al. Loss of TGH/Ces3 in Mice Decreases Blood Lipids, Improves Glucose Tolerance, and Increases Energy Expenditure. Cell metabolism. 2010; 11(3):183–93. doi: 10.1016/j.cmet.2010.02.005 PMID: 20197051

17. Teslovich TM, Musunuru K, Smith AV, Edmondson AC, Stylianou IM, Koseki M, et al. Biological, clinical and population relevance of 95 loci for blood lipids. Nature. 2010; 466(7307):707–13. doi: 10.1038/nature09270 PMID: 20686565

18. Willer CJ, Schmidt EM, Sengupta S, Peloso GM, Gustafsson S, Kanoni S, et al. Discovery and refinement of loci associated with lipid levels. Nature genetics. 2013; 45(11):1274–83. doi: 10.1038/ng.2797 PMID: 24097068

19. Kathiresan S, Melander O, Guiducci C, Surti A, Burtt NP, Rieder MJ, et al. Six new loci associated with blood low-density lipoprotein cholesterol, high-density lipoprotein cholesterol or triglycerides in humans. Nature genetics. 2008; 40(2):189–97. doi: 10.1038/ng.75 PMID: 18193044

20. Kooner JS, Chambers JC, Aguilar-Salinas CA, Hinds DA, Hyde CL, Warnes GR, et al. Genome-wide scan identifies variation in MLXIPL associated with plasma triglycerides. Nature genetics. 2008; 40 (2):149–51. doi: 10.1038/ng.2007.61 PMID: 18193046

21. Shen GQ, Li L, Wang QK. Genetic variant R952Q in LRP8 is associated with increased plasma triglyceride levels in patients with early-onset CAD and MI. Annals of human genetics. 2012; 76(3):193–9. doi: 10.1111/j.1469-1809.2012.00705.x PMID: 22404453

22. Zhang K, Chang S, Cui S, Guo L, Zhang L, Wang J. ICSNPathway: identify candidate causal SNPs and pathways from genome-wide association study by one analytical framework. Nucleic acids research. 2011; 39(Web Server issue):W437–43. doi: 10.1093/nar/gkr391 PMID: 21622953

23. Wang K, Li WD, Zhang CK, Wang Z, Glessner JT, Grant SF, et al. A genome-wide association study on obesity and obesity-related traits. PloS one. 2011; 6(4):e18939. doi: 10.1371/journal.pone.0018939 PMID: 21552555

24. Wilson PM, Fryer RH, Fang Y, Hatten ME. Astn2, a novel member of the astrotactin gene family, regulates the trafficking of ASTN1 during glial-guided neuronal migration. The Journal of neuroscience: the official journal of the Society for Neuroscience. 2010; 30(25):8529–40.

25. Vrijenhoek T, Buizer-Voskamp JE, van der Stelt I, Strengman E, Sabatti C, Geurts van Kessel A, et al. Recurrent CNVs disrupt three candidate genes in schizophrenia patients. American journal of human genetics. 2008; 83(4):504–10. doi: 10.1016/j.ajhg.2008.09.011 PMID: 18940311

26. Kathiresan S, Manning AK, Demissie S, D'Agostino RB, Surti A, Guiducci C, et al. A genome-wide association study for blood lipid phenotypes in the Framingham Heart Study. BMC Med Genet. 2007; 8 Suppl 1:S17. PMID: 17903299

27. Zeggini E, Scott LJ, Saxena R, Voight BF, Marchini JL, Hu T, et al. Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes. Nat Genet. 2008; 40(5):638–45. doi: 10.1038/ng.120 PMID: 18372903

28. Fogarty MP, Xiao R, Prokunina-Olsson L, Scott LJ, Mohlke KL. Allelic expression imbalance at high-density lipoprotein cholesterol locus MMAB-MVK. Human molecular genetics. 2010; 19(10):1921–9. doi: 10.1093/hmg/ddq067 PMID: 20159775

29. Kondos SC, Hatfaludi T, Voskoboinik I, Trapani JA, Law RHP, Whisstock JC, et al. The structure and function of mammalian membrane-attack complex/perforin-like proteins. Tissue Antigens. 2010; 76 (5):341–51. doi: 10.1111/j.1399-0039.2010.01566.x PMID: 20860583

30. Adams NC, Tomoda T, Cooper M, Dietz G, Hatten ME. Mice that lack astrotactin have slowed neuronal migration. Development. 2002; 129(4):965–72. PMID: 11861479

31. Lesch K-P, Timmesfeld N, Renner TJ, Halperin R, Röser C, et al. (2008) Molecular genetics of adult ADHD: converging evidence from genome-wide association and extended pedigree linkage studies. Journal of Neural Transmission 115: 1573–1585. doi: 10.1007/s00702-008-0119-3 PMID: 18839057

32. Uno K, Katagiri H, Yamada T, Ishigaki Y, Ogihara T, Imai J, et al. Neuronal pathway from the liver modulates energy expenditure and systemic insulin sensitivity. Science. 2006; 312(5780):1656–9. PMID: 16778057

33. Schwartz MW. Progress in the search for neuronal mechanisms coupling type 2 diabetes to obesity. The Journal of clinical investigation. 2001; 108(7):963–4. PMID: 11581296

34. Kim K-Z, Min J-Y, Kim K, Sung J, Cho S-I. Exploring Trans-acting regulators of gene expression associated with metabolic syndrome: a coupled application of factor analysis and linkage analysis. Genes & Genomics. 2013; 35(1):59–67.

35. Junyent M, Parnell LD, Lai CQ, Lee YC, Smith CE, Arnett DK, et al. Novel variants at KCTD10, MVK, and MMAB genes interact with dietary carbohydrates to modulate HDL-cholesterol concentrations in the Genetics of Lipid Lowering Drugs and Diet Network Study. The American journal of clinical nutrition. 2009; 90(3):686–94. doi: 10.3945/ajcn.2009.27738 PMID: 19605566

36. Li WD, Dong C, Li D, Zhao H, Price RA. An obesity-related locus in chromosome region 12q23-24. Diabetes. 2004; 53(3):812–20. PMID: 14988268

37. Maiti N, Widjaja L, Banerjee R. Proton transfer from histidine 244 may facilitate the 1,2 rearrangement reaction in coenzyme B(12)-dependent methylmalonyl-CoA mutase. The Journal of biological chemistry. 1999; 274(46):32733–7. PMID: 10551831

38. Clarke R, Armitage J. Vitamin supplements and cardiovascular risk: review of the randomized trials of homocysteine-lowering vitamin supplements. Seminars in thrombosis and hemostasis. 2000; 26 (3):341–8. PMID: 11011852

39. Refsum H, Ueland PM, Nygard O, Vollset SE. Homocysteine and cardiovascular disease. Annual review of medicine. 1998; 49:31–62. PMID: 9509248

40. Werstuck GH, Lentz SR, Dayal S, Hossain GS, Sood SK, Shi YY, et al. Homocysteine-induced endoplasmic reticulum stress causes dysregulation of the cholesterol and triglyceride biosynthetic pathways. The Journal of clinical investigation. 2001; 107(10):1263–73. PMID: 11375416

41. Shaffer AL, Yu X, He Y, Boldrick J, Chan EP, Staudt LM. BCL-6 represses genes that function in lymphocyte differentiation, inflammation, and cell cycle control. Immunity. 2000; 13(2):199–212. PMID: 10981963

42. Ye BH, Lista F, Lo Coco F, Knowles DM, Offit K, Chaganti RS, et al. Alterations of a zinc finger-encoding gene, BCL-6, in diffuse large-cell lymphoma. Science. 1993; 262(5134):747–50. PMID: 8235596

43. Glauser DA, Schlegel W. The emerging role of FOXO transcription factors in pancreatic beta cells. The Journal of endocrinology. 2007; 193(2):195–207. PMID: 17470511

44. Glauser DA, Schlegel W. The FoxO/Bcl-6/cyclin D2 pathway mediates metabolic and growth factor stimulation of proliferation in Min6 pancreatic beta-cells. Journal of receptor and signal transduction research. 2009; 29(6):293–8. doi: 10.3109/10799890903241824 PMID: 19929250

45. Barish GD, Yu RT, Karunasiri MS, Becerra D, Kim J, Tseng TW, et al. The Bcl6-SMRT/NCoR cistrome represses inflammation to attenuate atherosclerosis. Cell metabolism. 2012; 15(4):554–62. doi: 10. 1016/j.cmet.2012.02.012 PMID: 22465074

46. Jonkers IJ, Mohrschladt MF, Westendorp RG, van der Laarse A, Smelt AH. Severe hypertriglyceridemia with insulin resistance is associated with systemic inflammation: reversal with bezafibrate therapy in a randomized controlled trial. The American journal of medicine. 2002; 112(4):275–80. PMID: 11893366

47. Vogel RA, Corretti MC, Plotnick GD. Effect of a single high-fat meal on endothelial function in healthy subjects. The American journal of cardiology. 1997; 79(3):350–4. PMID: 9036757

48. Wang L, Gill R, Pedersen TL, Higgins LJ, Newman JW, Rutledge JC. Triglyceride-rich lipoprotein lipolysis releases neutral and oxidized FFAs that induce endothelial cell inflammation. Journal of lipid research. 2009; 50(2):204–13. doi: 10.1194/jlr.M700505-JLR200 PMID: 18812596

49. Lee HS, Moon KC, Song CY, Kim BC, Wang S, Hong HK. Glycated albumin activates PAI-1 transcription through Smad DNA binding sites in mesangial cells. American journal of physiology Renal physiology. 2004; 287(4):F665–72. PMID: 15198928

50. Kim YS, Kim BC, Song CY, Hong HK, Moon KC, Lee HS. Advanced glycosylation end products stimulate collagen mRNA synthesis in mesangial cells mediated by protein kinase C and transforming growth factor-beta. The Journal of laboratory and clinical medicine. 2001; 138(1):59–68. PMID: 11433229

51. Li MC, He SH. IL-10 and its related cytokines for treatment of inflammatory bowel disease. World journal of gastroenterology: WJG. 2004; 10(5):620–5. PMID: 14991925

52. Li WD, Dong C, Li D, Garrigan C, Price RA. A genome scan for serum triglyceride in obese nuclear families. Journal of lipid research. 2005; 46(3):432–8. PMID: 15604520

53. Wang K, Li WD, Glessner JT, Grant SF, Hakonarson H, Price RA. Large copy-number variations are enriched in cases with moderate to extreme obesity. Diabetes2010. p. 2690–4. doi: 10.2337/db10-0192 PMID: 20622171

54. Lahiri DK, Nurnberger JI Jr. A rapid non-enzymatic method for the preparation of HMW DNA from blood for RFLP studies. Nucleic acids research. 1991; 19(19):5444. PMID: 1681511

55. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. American journal of human genetics. 2007; 81(3):559–75. PMID: 17701901

56. Zhang K, Cui S, Chang S, Zhang L, Wang J. i-GSEA4GWAS: a web server for identification of pathways/gene sets associated with traits by applying an improved gene set enrichment analysis to genome-wide association study. Nucleic acids research. 2010; 38(Web Server):W90–W5. doi: 10. 1093/nar/gkq324 PMID: 20435672