

Research Article

Integrating Genome-Wide Association and eQTLs Studies Identifies the Genes and Gene Sets Associated with Diabetes

Xiao Liang, Awen He, Wenyu Wang, Li Liu, Yanan Du, Qianrui Fan, Ping Li, Yan Wen, Jingcan Hao, Xiong Guo, and Feng Zhang

Key Laboratory of Trace Elements and Endemic Diseases of National Health and Family Planning Commission, School of Public Health, Health Science Center, Xi'an Jiaotong University, Xi'an, China

Correspondence should be addressed to Feng Zhang; fzhxjtu@xjtu.edu.cn

Received 29 March 2017; Accepted 24 May 2017; Published 28 June 2017

Academic Editor: Rosaria Scudiero

Copyright © 2017 Xiao Liang et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Aim. To identify novel candidate genes and gene sets for diabetes. **Methods.** We performed an integrative analysis of genome-wide association studies (GWAS) and expression quantitative trait loci (eQTLs) data for diabetes. Summary data was driven from a large-scale GWAS of diabetes, totally involving 58,070 individuals. eQTLs dataset included 923,021 cis-eQTL for 14,329 genes and 4,732 trans-eQTL for 2,612 genes. Integrative analysis of GWAS and eQTLs data was conducted by summary data-based Mendelian randomization (SMR). To identify the gene sets associated with diabetes, the SMR single gene analysis results were further subjected to gene set enrichment analysis (GSEA). A total of 13,311 annotated gene sets were analyzed in this study. **Results.** SMR analysis identified 6 genes significantly associated with fasting glucose, such as C11ORF10 (p value = 6.04×10^{-8}), MRPL33 (p value = 1.24×10^{-7}), and FADS1 (p value = 2.39×10^{-7}). Gene set analysis identified HUANG_FOXA2_TARGETS_UP (false discovery rate = 0.047) associated with fasting glucose. **Conclusion.** Our study provides novel clues for clarifying the genetic mechanism of diabetes. This study also illustrated the good performance of SMR approach and extended it to gene set association analysis for complex diseases.

1. Introduction

Diabetes is a group of metabolic diseases, mainly characterized by raised blood glucose over a prolonged period. Without effective treatments, diabetes will lead to serious secondary disorders, such as heart disease, stroke, chronic kidney failure, and foot ulcers. During the past decades, the prevalence of diabetes continues to increase, caused by aging, obesity, smoking, and other unhealthy lifestyle factors [1]. It was estimated that 334 million individuals would suffer diabetes in 2025 [1]. Diabetes has become one of the major public health problems, bringing heavy economic burden to the society.

Genetic factors contribute greatly to the development of diabetes. Extensive genetic studies have been conducted and identified a group of susceptibility genes for diabetes, such as PTEN [2], SREBF1 [3], JAZF1 [4], BCL2 [5], and FAM19A2 [5]. However, the genetic risk of diabetes explained by the identified loci was limited, suggesting the existence

of undiscovered susceptibility loci for diabetes. The missing heritability can partly be attributed to the regulatory genetic variants, which are mostly locating outside genes and ignored by traditional genetic studies.

Expression quantitative trait loci (eQTLs) are a group of important regulatory loci, which can regulate gene expression levels. The disease-associated SNPs identified by GWAS are significantly enriched in eQTLs, supporting the implication of eQTLs in the pathogenesis of complex diseases [6]. Through genome-wide detecting associations between gene transcript abundance and genomic polymorphisms, a large amount of eQTLs has been identified in human genome [7, 8]. Recently, summary data-based Mendelian randomization (SMR) analysis was proposed to utilize extensive published GWAS as well as eQTLs data. SMR is capable of integrating GWAS summary and eQTLs annotation data to identify novel causal genes, the expression levels of which are associated with target diseases [9]. SMR showed a high power for identifying novel causal genes of complex diseases [9].

In this study, we conducted a genome-wide single gene and gene sets expression association analysis for diabetes. SMR was first applied to a large-scale GWAS data for screening novel susceptibility genes of diabetes. To gain insight into the biological significance of identified genes, we extended SMR to gene set enrichment analysis (GSEA). SMR gene-level analysis results were subjected to GSEA for identifying diabetes associated gene sets with known functional information.

2. Methods

2.1. GWAS Summary Datasets. A large-scale GWAS meta-analysis summary data of diabetes was used in this study [10]. Briefly, this GWAS comprised 58,070 individuals from 29 studies involved in the Meta-Analysis of Glucose and Insulin related traits Consortium. Fasting glucose and fasting insulin were measured from whole blood, plasma, or serum samples. Detailed information of measurements of fasting glucose and fasting insulin is summarized in Supplementary Table S1 and Table S2 in Supplementary Material available online at <https://doi.org/10.1155/2017/1758636>. Commercial platforms were used for genome-wide SNP genotyping, such as Affymetrix 500K SNP array, Illumina 550K, and Perlegen 600K. Imputation was conducted by MACH [11] or IMPUTE [12] against the HapMap CEU reference genome (build 36). The GWAS meta-analysis was conducted by joint meta-analytical approach [13]. Detailed information of cohorts, genotyping, imputation, meta-analysis, and quality control approaches can be found in the published studies [10].

2.2. SMR Single Gene Analysis. The GWAS meta-analysis summary data of diabetes was input into SMR for single gene expression association analysis of fasting glucose and insulin resistance. SMR is capable of integrating GWAS results with eQTLs annotation information to evaluate the relationships between gene expression levels and complex traits [9]. We applied the eQTLs annotation dataset built by Westra et al. [14]. Briefly, these eQTLs datasets were driven from a meta-analysis of 5,311 peripheral blood samples and replicated in another 2,775 samples. Illumina whole-genome Expression BeadChips were used for gene expression profiling. SNP genotyping was conducted using commercial platforms, such as Illumina 610K quad arrays and Illumina HumanHap300 arrays. Imputation was conducted using MACH [11] or IMPUTE [12] against the HapMap 2 reference panels. 923,021 cis-eQTL for 14,329 gene expression probes and 4,732 trans-eQTL for 2,612 gene expression probes were identified at false discovery rate (FDR) < 0.05 [14]. An expression association testing p value for each gene was calculated by SMR. After Bonferroni correction, the genes with SMR p values < 9.28×10^{-6} (0.05/5389) were considered as significant genes in our study.

2.3. Gene Set Enrichment Analysis. To reveal the functional significance of identified genes, the SMR single gene expression association testing results were further subjected to GSEA [15]. The gene set annotation database (msigdb.v5.1)

TABLE 1: List of candidate genes identified by SMR for fasting glucose.

Gene	Top SNP	MAF	SMR	
			β	p value
C11ORF10	rs174547	0.331	-0.059	6.04×10^{-8}
MRPL33	rs3736594	0.258	-0.118	1.24×10^{-7}
FADS1	rs174548	0.301	-0.067	2.39×10^{-7}
ACP2	rs901746	0.297	-0.050	1.74×10^{-6}
NRIH3	rs901746	0.297	-0.051	1.78×10^{-6}
SNX17	rs1260320	0.392	-0.072	2.19×10^{-6}

Note. MAF, minor allele frequency.

was obtained from the GSEA Molecular Signatures Database (<http://software.broadinstitute.org/gsea/msigdb/index.jsp>). 5,000 permutations were conducted to calculate the FDR adjusted p value of each gene set [16]. Significant gene sets were identified at FDR adjusted p value < 0.05. Detailed GSEA procedures can be found in our previous studies [17].

3. Results

3.1. SMR Single Gene Expression Association Analysis. A total of 5,389 genes with both GWAS summary and eQTLs data were analyzed in this study. After strict Bonferroni correction, SMR identified 6 genes significantly associated with fasting glucose (Table 1), including C11ORF10 (p value = 6.04×10^{-8}), MRPL33 (p value = 1.24×10^{-7}), FADS1 (p value = 2.39×10^{-7}), ACP2 (p value = 1.74×10^{-6}), NRIH3 (p value = 1.78×10^{-6}), and SNX17 (p value = 2.19×10^{-6}).

For fasting insulin, SMR detected suggestive association signals for 7 genes (Table 2), including ATRIP (p value = 9.68×10^{-5}), MRPL33 (p value = 9.75×10^{-6}), ATRIP (p value = 1.90×10^{-4}), POLRIE (p value = 2.60×10^{-4}), AMT (p value = 3.44×10^{-4}), TNFSF13 (p value = 4.55×10^{-4}), and POLRIE (p value = 7.82×10^{-4}).

3.2. Gene Set Enrichment Analysis. A total of 10,987 annotated gene sets were analyzed in this study. GSEA observed significant association between HUANG_FOXA2_TARGETS_UP gene ontology (GO) term and fasting glucose (FDR adjusted p value = 0.047). For fasting insulin, GSEA detected suggestive association signal for chr8p23 GO term (FDR adjusted p value = 0.063).

4. Discussion

It is a challenge to reveal the biological significances of identified loci by GWAS, especially a large part of significant loci locating outside genes [9]. To better understand the genetic basis and make full use of published GWAS data of diabetes, we conducted an eQTL-based single gene and gene set expression association analysis for diabetes. We identified multiple genes and gene sets associated with fasting glucose or fasting insulin.

TABLE 2: List of candidate genes identified by SMR for fasting insulin.

Gene	Top SNP	MAF	SMR	
			β	<i>p</i> value
ATRIP	rs2228561	0.129	-0.070	9.68×10^{-5}
MRPL33	rs3736594	0.258	-0.067	9.75×10^{-5}
ATRIP	rs2228561	0.129	-0.084	1.90×10^{-4}
POLRIE	rs10758435	0.166	-0.026	2.60×10^{-4}
AMT	rs1050088	0.429	0.031	3.44×10^{-4}
TNFSF13	rs9898876	0.193	-0.037	4.55×10^{-4}
POLRIE	rs10973396	0.168	-0.028	7.82×10^{-4}

Note. MAF, minor allele frequency.

SMR analysis observed the most significant association between fasting glucose and CIIORF10. CIIORF10 is close to another significant gene FADS1 identified by SMR. It has been demonstrated that CIIORF10 played an important role in fatty acid and glucose metabolism [18]. Zabaneh and Balding reported that CIIORF10 and FADS1 were significantly associated with metabolic syndrome [19]. Powell et al. observed that FADS1 knockout mice presented less glucose and insulin excursions during oral glucose tolerance tests along with lower fasting glucose, insulin, triglyceride, and total cholesterol levels [20]. Yao et al. suggested that FADS1-FADS2 gene cluster was significantly associated with type 2 diabetes [21]. Cormier et al. observed that FADS gene cluster could modulate plasma fasting glucose and fasting insulin levels in response to n-3 polyunsaturated fatty acids supplementation [22].

SNX17 is another notable gene associated with fasting glucose. SNX17 encodes sorting nexin 17, which involves receptor binding and phosphatidylinositol binding. It has been demonstrated that the eQTLs of SNX17 was significantly associated with glucometabolic phenotypes [23]. Adachi and Tsujimoto found that SNX17 directly interacted with FEEL-1/stabilin-1, which was implicated in the development of diabetes [24].

TNFSF13 is significantly associated with fasting insulin in this study. Gao et al. reported that the TNFSF13 level in serum was significantly associated with the diabetic status of patients with pancreatic ductal adenocarcinoma-associated diabetes [25].

Besides confirming functional relevance of previously reported candidate genes with diabetes, SMR analysis also identified several novel candidate genes for diabetes, such as MRPL33, ACP2, and NR1H3. To the best of our knowledge, few efforts have been paid to investigate the potential roles of these genes in the development of diabetes. Further biological studies are warranted to confirm our finding and clarify the potential roles of novel candidate genes in the pathogenesis of diabetes.

Gene set analysis found that HUANG_FOXA2_TARGETS_UP GO term was significantly associated with fasting glucose. HUANG_FOXA2_TARGETS_UP comprises 45 genes, some of which have been suggested to be implicated in the development of diabetes, such as KAT2B and TNFAIP3. Rabhi et al. found that disruption of KAT2B

led to impaired insulin secretion and glucose intolerance in mice [26]. They suggested that KAT2B was a key transcriptional regulator in maintaining normal function of adaptive β cell [26]. TNFAIP3 was suggested to be associated with type 1 diabetes [27].

In summary, we conducted a genome-wide integrative analysis of GWAS and eQTLs data for diabetes. We identified several novel candidate genes and gene sets associated with the risk of diabetes. Our results provide new clues for clarifying the genetic mechanism of diabetes. We also illustrated the good performance of SMR approach and extended it to gene set association analysis for complex diseases.

Conflicts of Interest

There are no conflicts of interest regarding the publication of this article.

Authors' Contributions

Xiao Liang and Awen He contributed equally to this manuscript.

Acknowledgments

This study is supported by the National Natural Scientific Foundation of China (81472925, 81673112), the Technology Research and Development Program of Shaanxi Province of China (2013KJXX-51), and the Fundamental Research Funds for the Central Universities.

References

- [1] S. Wild, G. Roglic, A. Green, R. Sicree, and H. King, "Global prevalence of diabetes: estimates for the year 2000 and projections for 2030," *Diabetes Care*, vol. 27, no. 5, pp. 1047–1053, 2004.
- [2] L. Grønder-Hansen, R. Ribel-Madsen, J. F. P. Wojtaszewski, P. Poulsen, L. G. Grønnet, and A. Vaag, "A common variation of the PTEN gene is associated with peripheral insulin resistance," *Diabetes and Metabolism*, vol. 42, no. 4, pp. 280–284, 2016.
- [3] N. Grarup, K. L. Stender-Petersen, E. A. Andersson et al., "Association of variants in the sterol regulatory element-binding factor 1 (SREBF1) gene with type 2 diabetes, glycemia, and

- insulin resistance A study of 15,734 danish subjects,” *Diabetes*, vol. 57, no. 4, pp. 1136–1142, 2008.
- [4] E. Zeggini, L. J. Scott, and R. Saxena, “Meta-analysis of genome-wide association data and large-scale replication identifies additional susceptibility loci for type 2 diabetes,” *Nature Genetics*, vol. 40, no. 5, pp. 638–645, 2008.
- [5] G. A. Walford et al., “Genome-wide association study of the modified Stumvoll Insulin Sensitivity Index identifies BCL2 and FAM19A2 as novel insulin sensitivity loci,” *Diabetes*, vol. 65, no. 10, Article ID db160199, pp. 3200–3211, 2016.
- [6] D. L. Nicolae, E. Gamazon, W. Zhang, S. Duan, M. Eileen Dolan, and N. J. Cox, “Trait-associated SNPs are more likely to be eQTLs: Annotation to enhance discovery from GWAS,” *PLoS Genetics*, vol. 6, no. 4, Article ID e1000888, 2010.
- [7] S. Yang, Y. Liu, N. Jiang et al., “Genome-wide eQTLs and heritability for gene expression traits in unrelated individuals,” *BMC Genomics*, vol. 15, no. 1, article 13, 2014.
- [8] E. Petretto, “Single cell expression quantitative trait loci and complex traits,” *Genome Medicine*, vol. 5, no. 8, article 72, 2013.
- [9] Z. Zhu, F. Zhang, H. Hu et al., “Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets,” *Nature Genetics*, vol. 48, no. 5, pp. 481–487, 2016.
- [10] A. K. Manning, “A genome-wide approach accounting for body mass index identifies genetic variants influencing fasting glycemic traits and insulin resistance,” *Nature Genetics*, vol. 44, no. 6, pp. 659–669, 2012.
- [11] Y. Li, C. J. Willer, J. Ding, P. Scheet, and G. R. Abecasis, “MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes,” *Genetic Epidemiology*, vol. 34, no. 8, pp. 816–834, 2010.
- [12] J. Marchini, B. Howie, S. Myers, G. McVean, and P. Donnelly, “A new multipoint method for genome-wide association studies by imputation of genotypes,” *Nature Genetics*, vol. 39, no. 7, pp. 906–913, 2007.
- [13] A. K. Manning, M. LaValley, C.-T. Liu et al., “Meta-analysis of gene-environment interaction: joint estimation of SNP and SNP \times environment regression coefficients,” *Genetic Epidemiology*, vol. 35, no. 1, pp. 11–18, 2011.
- [14] H. J. Westra et al., “Systematic identification of trans eQTLs as putative drivers of known disease associations,” *Nature Genetics*, vol. 45, no. 10, pp. 1238–1243, 2013.
- [15] A. Subramanian, P. Tamayo, V. K. Mootha 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*, vol. 102, no. 43, pp. 15545–15550, 2005.
- [16] K. Wang, M. Li, and M. Bucan, “Pathway-based approaches for analysis of genomewide association studies,” *American Journal of Human Genetics*, vol. 81, no. 6, pp. 1278–1283, 2007.
- [17] Y. Wen, W. Wang, X. Guo, and F. Zhang, “PAPA: A flexible tool for identifying pleiotropic pathways using genome-wide association study summaries,” *Bioinformatics*, vol. 32, no. 6, pp. 946–948, 2015.
- [18] G. Bochenek, R. Häslar, N.-E. E. Mokhtari et al., “The large non-coding RNA ANRIL, which is associated with atherosclerosis, periodontitis and several forms of cancer, regulates ADIPOR1, VAMP3 and C11ORF10,” *Human Molecular Genetics*, vol. 22, no. 22, Article ID ddt299, pp. 4516–4527, 2013.
- [19] D. Zabaneh and D. J. Balding, “A genome-wide association study of the metabolic syndrome in Indian Asian men,” *PLoS ONE*, vol. 5, no. 8, Article ID e11961, 2010.
- [20] D. R. Powell, J. P. Gay, M. Smith et al., “Fatty acid desaturase 1 knockout mice are lean with improved glycemic control and decreased development of atheromatous plaque,” *Diabetes, Metabolic Syndrome and Obesity: Targets and Therapy*, vol. 9, pp. 185–199, 2016.
- [21] M. Yao, J. Li, T. Xie et al., “Polymorphisms of rs174616 in the FADS1-FADS2 gene cluster is associated with a reduced risk of type 2 diabetes mellitus in northern Han Chinese people,” *Diabetes Research and Clinical Practice*, vol. 109, no. 1, pp. 206–212, 2015.
- [22] H. Cormier, I. Rudkowska, E. Thifault, S. Lemieux, P. Couture, and M.-C. Vohl, “Polymorphisms in Fatty Acid Desaturase (FADS) gene cluster: Effects on glycemic controls following an omega-3 Polyunsaturated Fatty Acids (PUFA) supplementation,” *Genes*, vol. 4, no. 3, pp. 485–498, 2013.
- [23] S. P. Sajuthi, N. K. Sharma, J. W. Chou et al., “Mapping adipose and muscle tissue expression quantitative trait loci in African Americans to identify genes for type 2 diabetes and obesity,” *Human Genetics*, vol. 135, no. 8, pp. 869–880, 2016.
- [24] H. Adachi and M. Tsujimoto, “Adaptor protein sorting nexin 17 interacts with the scavenger receptor FEEL-1/stabilin-1 and modulates its expression on the cell surface,” *Biochimica et Biophysica Acta - Molecular Cell Research*, vol. 1803, no. 5, pp. 553–563, 2010.
- [25] W. Gao, Y. Zhou, Q. Li et al., “Analysis of global gene expression profiles suggests a role of acute inflammation in type 3C diabetes mellitus caused by pancreatic ductal adenocarcinoma,” *Diabetologia*, vol. 58, no. 4, pp. 835–844, 2015.
- [26] N. Rabhi, P.-D. Denechaud, X. Gromada et al., “KAT2B Is Required for Pancreatic Beta Cell Adaptation to Metabolic Stress by Controlling the Unfolded Protein Response,” *Cell Reports*, vol. 15, no. 5, pp. 1051–1061, 2016.
- [27] S. Hoffjan, A. Okur, J. T. Epplen, S. Wiczorek, A. Chan, and D. A. Akkad, “Association of TNFAIP3 and TNFRSF1A variation with multiple sclerosis in a German case-control cohort,” *International Journal of Immunogenetics*, vol. 42, no. 2, pp. 106–110, 2015.