Human Immunology 76 (2015) 753–758

Contents lists available at ScienceDirect

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

# Functional relevance for type 1 diabetes mellitus-associated genetic variants by using integrative analyses

Ying-Hua Qiu [a,b], Fei-Yan Deng [a,b], Zai-Xiang Tang [a,b], Zhen-Huan Jiang [c,*], Shu-Feng Lei [a,b,*]

[a] Center for Genetic Epidemiology and Genomics, School of Public Health, Soochow University, Suzhou, Jiangsu 215123, PR China
[b] Jiangsu Key Laboratory of Preventive and Translational Medicine for Geriatric Disease, School of Public Health, Soochow University, Suzhou 215123, Jiangsu, PR China
[c] Department of Orthopedics, the Affiliated Yixing Hospital of Jiangsu University, Yixing 214200, Jiangsu, PR China

ARTICLE INFO

*Article history:*
Received 21 October 2014
Revised 16 July 2015
Accepted 27 September 2015
Available online 30 September 2015

*Keywords:*
Functional relevance
Immunogenetics
Integrative analyses
Type 1 DM

ABSTRACT

*Aim:* Type 1 diabetes mellitus (type 1 DM) is an autoimmune disease. Although genome-wide association studies (GWAS) and meta-analyses have successfully identified numerous type 1 DM-associated susceptibility loci, the underlying mechanisms for these susceptibility loci are currently largely unclear.
*Methods:* Based on publicly available datasets, we performed integrative analyses (i.e., integrated gene relationships among implicated loci, differential gene expression analysis, functional prediction and functional annotation clustering analysis) and combined with expression quantitative trait loci (eQTL) results to further explore function mechanisms underlying the associations between genetic variants and type 1 DM.
*Results:* Among a total of 183 type 1 DM-associated SNPs, eQTL analysis showed that 17 SNPs with cis-regulated eQTL effects on 9 genes. All the 9 eQTL genes enrich in immune-related pathways or Gene Ontology (GO) terms. Functional prediction analysis identified 5 SNPs located in transcription factor (TF) binding sites. Of the 9 eQTL genes, 6 (TAP2, HLA-DOB, HLA-DQB1, HLA-DQA1, HLA-DRB5 and CTSH) were differentially expressed in type 1 DM-associated related cells. Especially, rs3825932 in CTSH has integrative functional evidence supporting the association with type 1 DM.
*Conclusions:* These findings indicated that integrative analyses can yield important functional information to link genetic variants and type 1 DM.


## 1. Introduction

Type 1 diabetes mellitus (type 1 DM) is an autoimmune disease characterized by destruction of pancreatic beta cells, leading to absolute insulin deficiency. Diabetes increases the risk of micro- and macro-vascular complications [1], and premature death in the general population and results in a huge economic burden for society [2]. It accounts for 5–10% of total cases of diabetes worldwide [3] and ~3% increase in the incidence of globally per year [4]. Genetic factors play an important role in the pathogenesis of type 1 DM as evidenced by high concordance among monozygotic twins [5] and has increased risk in families and sib pairs [6].

Elucidation of genetic basis of type 1 DM remains one of huge challenges in the field of human genetics largely due to complex nature of genetic determination for type 1 DM. Currently, genome-wide association studies (GWAS) and meta-analysis have successfully identified a long list of type 1 DM-associated susceptibility loci with relatively high statistical power [7–9]. Unfortunately, most of the studies have only reported statistical association evidence for the loci, but not reported functional mechanisms for the associations. For a better understanding of genetic basis of type 1 DM, we performed comprehensive integrative analyses to gain integrative evidences from multiple levels (i.e., DNA, RNA) to ascertain potential function mechanisms of these associations.

The integrative analyses used in this study included gene relationships among implicated loci (GRAIL), differential gene expression analysis, functional prediction and functional annotation clustering analysis and the expression quantitative trait loci (eQTL) analysis. Finally, we found some interesting eQTL effect SNPs located in the potentially function regions, whose corresponding eQTL genes have significant differential expression and tend to enrich in immune-related pathways or GO (Gene Ontology) terms.

* Corresponding author at: Center for Genetic Epidemiology and Genomics, School of Public Health, Soochow University, Suzhou, Jiangsu 215123, PR China.
*E-mail addresses:* jiangzh1018@126.com (Z.-H. Jiang), leisf@suda.edu.cn (S.-F. Lei).

http://dx.doi.org/10.1016/j.humimm.2015.09.033
0198-8859/

## 2. Materials and methods

### 2.1. Type 1 DM-associated SNP selection

Phenotype-Genotype Integrator (PheGenI) (http://www.ncbi.nlm.nih.gov/gap/phegeni/) is a bioinformatics web tool that merges National Human Genome Research Institute (NHGRI) GWAS catalog data with several databases, including Gene, dbGaP, OMIM, GTEx and dbSNP. This phenotype-oriented resource allows users to efficiently identify and prioritize high-risk SNPs according to the phenotype and *P*-value. A total of 183 type 1 DM-associated SNPs with $P < 10E{-}08$ were downloaded from PheGenI. In total, we found 151 unique SNPs (only detected in one study) and the corresponding 108 candidate genes.

### 2.2. Gene relationship across implicated loci (GRAIL)

The first step of detection functional mechanisms underlying the significant association is to identify the candidate genes associated with the significant SNPs. A commonly used approach is to select nearby genes that are physically located at two sides of each associated SNP as susceptible candidate genes for further functional studies. GRAIL [10] is another efficient method used in prioritizing and identifying susceptible candidate genes (called GRAIL genes late on), which could automatically assesses the degree of functional connectivity with the other susceptible candidate genes using 250,000 PubMed abstracts, then prioritizes and identifies the best candidate gene around associated SNP. We use $P < 0.05$ as significance thresholds to filter GRAIL genes.

### 2.3. eQTL analysis

As we know, variants at the DNA level may have allele-specific influence on the transcriptional regulation (such as RNA level) and/or gene expression, which subsequently result in variations in end-point phenotypes (such as susceptibility to type 1 DM). eQTL data could help to elucidate mechanisms underlying associations between gene variants and quantifiable intermediate phenotypes (such as candidate gene RNA level), which is summarized and archived in web browsers (eQTL Browser: http://eqtl.uchicago.edu/cgi-bin/gbrowse/eqtl/). We used the eQTL data to evaluate whether the aforementioned detected type 1 DM-associated SNPs influence transcript level of corresponding genes. We only explored the eQTL effects of the above selected associations in immune-related cells (including lymphoblastoid cell lines (LCLs), monocytes and T cells). For the identified eQTL effect SNPs, only their regulated genes (defined as eQTL genes) are the same as aforementioned candidate genes, we will further explore their functional mechanisms.

### 2.4. Differential expression analysis for eQTL genes

To detect the functional relevance between the eQTL genes and type 1 DM, differential expression analysis was performed for the eQTL genes in type 1 DM-related cells. First, we downloaded three publicly available expression data sets from GEO Datasets (www.ncbi.nlm.nih.gov/geo) (GSE NO.: GSE35725, GSE33440, and GSE29142). These studies were performed with original purposes of identifying transcriptional signatures as a disease-specific and predictive inflammatory biomarker underlying type 1 DM in peripheral blood mononuclear cells (PBMCs) or monocytes [11,12]. Details on sample quality control and experimental procedures were previously described in the original publication [11–14]. Then, we performed *T*-test to identify differential expression of the eQTL genes by comparing mean gene expression signals in peripheral blood mononuclear cells (PBMCs) and monocytes between type 1 DM cases and controls. The significance level of $P < 0.05$ was used.

### 2.5. Functional prediction for eQTL SNPs

RegulomeDB is a novel analysis tool that guides interpretation of regulatory variants in non-coding regions of the human genome and is a comprehensive variant annotation tool that makes use of functional sources [15]. We searched the RegulomeDB (“http://regulome.stanford.edu/”) to investigate whether the eQTL effect SNPs are functional variants and have putative regulation effect. RegulomeDB score, ranging from 1 to 6, is set up to assess the degree of confidence to support a functional variant. Lower scores indicate strong evidence for a variant to be located in a functional region. Details on RegulomeDB score explanations were described in the original publication [15].

### 2.6. Functional annotation clustering analysis

Here, to gain the functional similarity of these association genes, we tested the probability of the eQTL genes clustering into a specific GO (Gene Ontology) term and particular biological pathway by using the GO [16] project and KEGG (Kyoto Encyclopedia of Genes and Genomes) database, respectively. The functional annotation clustering analysis was performed by using the Database for Annotation, Visualization and Integrated Discovery (DAVID) integrated database query tools (http://david.abcc.ncifcrf.gov/) [17]. The Fisher exact test was calculated to quantitatively measure the enrichment of GO term or pathway. Bonferroni correction was adopted for multiple testing [18].

## 3. Results

The overall design of analysis process and general results are shown in Fig 1. We gain 183 type 1 DM associations results with $P < 10E8$ from the PheGenI (Table S1). These associations collected from 11 published type 1 DM-association studies. The 183 type 1 DM associations corresponded to 151 unique SNPs (only detected in one study) and 21 replication SNPs (detected in at least two studies, e.g., rs2476601 was significantly associated with type 1 DM ($P = 2.00E{-}111$, $P = 9.00E{-}85$ and $P = 2.00E{-}80$) in three independent studies). The most significant SNPs is rs3129871 ($P = 1.15E{-}299$) in HLA-DRA gene. As we expected, the majority of these associations (>100 associations) were mapped to the HLA region, according to the physical location at chromosome 6p21.3. As shown in Table S1, we observed 108 unique candidate genes in the columns “Gene 1” and “Gene 2”, which listed the nearest genes physically located at two sides of the SNP, respectively. Based on the GRAIL analysis, about 44 “susceptible” genes were detected as listed in the column “GRAIL genes” of Table S1. Among the 44 “susceptible” genes, 14 are novel detected. Thus a total of 122 unique “candidate” genes were identified using the above two methods, which were subject to further analyses.

According to eQTL analysis results, about 73 SNPs have potential eQTL effect either on the candidate genes or other genes (Table S2). Among the 73 SNPs, 26 associations between 17 unique SNPs and expression of 9 “candidate” genes were detected in monocytes and/or LCLs (Table 1). Thus, the 17 unique SNPs and 9 “candidate” genes were subject to further analyses as eQTL effect SNPs and eQTL genes.

Among the 17 eQTL effect SNPs, three were located at non-HLA regions (e.g., rs7221109, rs11171739 and rs3825932). The entire 17 eQTL effect SNPs act as cis-effect regulators of the 9 “candidate” genes (eQTL genes). Especially, six SNPs around HLA-DRA gene