HHS Public Access

Author manuscript

J Immunol. Author manuscript; available in PMC 2017 April 01.

Published in final edited form as:

J Immunol. 2016 April 1; 196(7): 3043–3053. doi:10.4049/jimmunol.1502056.

Systematic Evaluation Of Genes And Genetic Variants Associated With Type 1 Diabetes Susceptibility

Ramesh Ram[*,†], **Munish Mehta**[*,†], **Quang T. Nguyen**[*,†], **Irma Larma**[*,†], **Bernhard O. Boehm**[‡], **Flemming Pociot**[§], **Patrick Concannon**[¶], and **Grant Morahan**[*,†]

[*]Centre for Diabetes Research, Harry Perkins Institute of Medical Research, Nedlands, WA, Australia

[†]Centre of Medical Research, University of Western Australia, Nedlands, WA, Australia

[‡]Division of Endocrinology and Diabetes, Department of internal medicine, Ulm University Medical Centre, Germany

[§]Department of Pediatrics, Herlev and Gentofte Hospital, Herlev Ringvej, Denmark

[¶]Genetics Institute and Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, Florida, USA

Abstract

Genome-wide association studies (GWAS) have found over 60 loci that confer genetic susceptibility to Type 1 diabetes (T1D). Many of these are defined only by anonymous SNPs: the underlying causative genes, and the molecular bases by which they mediate susceptibility, are not known. Identification of how these variants affect the complex mechanisms contributing to the loss of tolerance is a challenge. We performed systematic analyses to characterize these variants. First, all known genes in strong linkage disequilibrium (LD) ($r^2 > 0.8$) with the reported SNPs for each locus were tested for commonly occurring non-synonymous variations. We found only a total of 22 candidate genes at 16 T1D loci with common non-synonymous alleles. Next, we performed functional studies to examine the effect of non-HLA T1D risk alleles on regulating expression levels of genes in four different cell types: EBV- transformed B cell lines (resting and 6h PMA stimulated); purified CD4+ and CD8+ T cells. We mapped *cis*-acting expression quantitative trait loci (eQTL) and found 24 non-HLA loci that affected the expression of 31 transcripts significantly in at least one cell type. Additionally, we observed 25 loci that affected 38 transcripts in *trans*. In summary, our systems genetics analyses defined the effect of T1D risk alleles on levels of gene expression and provide novel insights into the complex genetics of T1D, suggesting most of the T1D risk alleles mediate their effect by influencing expression of multiple nearby genes.

Keywords

Type 1 Diabetes; eQTL; Gene expression; Genome-Wide Association Studies

Corresponding author contact details: Prof Grant Morahan, Centre for Diabetes Research, Harry Perkins Institute of Medical Research, QQ Block, QEII Medical Centre, 6 Verdun Street, Nedlands, WA, 6009 Australia, Tel. +61 8 9224 0397, Fax. +61 8 9224 0322, grant.morahan@uwa.edu.au.

Online Resource: http://www.sysgen.org/T1DGCSysGen/

INTRODUCTION

Type 1 diabetes (T1D) affects approximately 30 million people worldwide (1). It is a complex autoimmune disease causing the destruction of pancreatic β cells. The largest genetic studies of T1D have been carried out by the Type 1 Diabetes Genetics Consortium (T1DGC) (2–4). These and other reports have now defined genetic variants associated with T1D in over 60 different chromosomal regions (see ref 5 for review).

There is a need to identify the causative variants that are in linkage disequilibrium (LD) with the single nucleotide polymorphisms (SNPs) found by such association studies, and to define the molecular bases by which they contribute to disease susceptibility. The challenge of post-genome wide association studies (GWAS) functional studies (6–8) is in finding ways to translate genetic associations into clinically useful information. The strong genetic association of the disease with HLA class II genes of the major histocompatibility complex (MHC) is well established (9) but the identity of the genes associated with many of the non-HLA loci remains largely unknown, especially with respect to those associated SNPs located in non-coding regions of the genome (2, 5). Therefore, this study focuses on characterizing the non-HLA T1D risk loci.

In principle, most genetic variants could plausibly affect biological processes by changing amino acid residues in encoded proteins or by changing their levels of expression in particular tissues. Various DNA sequence repositories allow identification of commonly occurring non-synonymous (missense) variations in genes, and amino-acid substitution polymorphisms could be characterized for their potential to affect biological processes (10). Expression quantitative trait locus (eQTL) analyses can identify genes whose variation in expression is associated with specific SNP markers. For example, sequence variation in promoters or enhancer elements could result in differential *cis* regulation. Genetic variants can also regulate expression of genes at greater distances from, or on different chromosomes than, the regulatory element, i.e., *trans* regulation (11). The mechanisms involved in *trans* regulation could include indirect genetic effects, e.g. by means of variation in encoded proteins such as transcription factors, or by other effects, such as steric (11). Some loci could exert both *cis* and *trans* effects.

In the present study, we performed systems genetics (12) analyses of the 55 loci (2, 13–25) (Table I) showing highest evidence of association with T1D, using data generated by the T1DGC (2) and Immunochip projects (13). Additionally, four new SNPs (*rs6691977*, *rs4849135*, *rs2611215* and *rs11954020*) that showed strong associations ($P < 5 \times 10^{-8}$) with T1D in (13) were included in our study. SNPs at these loci were assessed for disease gene candidacy. Expression data of 47,323 high-quality transcripts (Illumina, HT-12 V4) were correlated with SNPs reported in T1D loci adjusting for confounding factors such as population structure.

MATERIALS AND METHODS

Study Samples

The Type 1 Diabetes Genetics Consortium (T1DGC) study has been described elsewhere, including phenotypic and extensive genetic characterization of over 4,000 affected sib-pair families (3). Upon joining the T1DGC, family members provided blood samples. Peripheral blood mononuclear cells (PBMC) were isolated and aliquots were used to provide DNA samples; to derive EBV-transformed B lymphoblastoid cell lines (LCL) (26–27); and frozen for later use. EBV-B cells from 202 European subjects from the T1DGC family collection were studied here. These samples consisted of 46 unaffected subjects and the rest were T1D cases. EBV-B cells were either unstimulated, or treated with phorbol-12-myristate-13-acetate (PMA) (28) for 6h (26–27). PMA stimulated samples consisted of 49 unaffected subjects. Cell lines were stimulated on a second occasion to provide a duplicate sample. SNPs were genotyped using the Immunochip (13) platform.

Frozen PBMC samples from 113 T1DGC family members were thawed, cultured overnight, stained and separated into CD4+ and CD8+ T cell populations by flow-sorting. Sufficient RNA was obtained from 102 CD4+ T cell samples and 84 CD8+ T cell samples to perform microarrays. Sex, HLA-DR and autoantibody statuses of the affected subjects are summarized in Suppl. Table I.(i).

Microarray Analyses

After cell culture or flow-sorting, RNA was extracted using TRIZOL® Reagent (Invitrogen) following the manufacturer's instructions. The RNA quantity was measured by NanoDrop 1000 Spectrophotometer (Thermo Scientific) and RNA quality was checked on Agilent 2100 Bioanalyser (Agilent). Samples with RIN (RNA Integrity Number) number of equal or greater than 8 were biotin labeled using Illumina TotalPrep RNA Amplification kit (Ambion) as per manufacturer's instructions. The biotin- labelled samples were hybridized onto Illumina HumanHT-12 v4.0 expression beadchips and beadchips were scanned by Beadarray Reader (Illumina) following manufacturer's instructions. Raw data was finally exported by GenomeStudio software (Illumina) for analysis.

Microarray and eQTL analysis

Genome-wide gene expression values from GenomeStudio (Illumina) for each of 47,323 probes were subjected to background correction using control probe profile, variance stabilizing transformation (VST) and RSN (robust spline normalization) normalization using lumi package (29) in R. We then removed from the analysis 95 transcripts that are *ERCC*-spike in controls (having gene symbols starting with '*ERCC*'). Four separate gene expression datasets were created. Upon examining initial PCA plots, batch effects were evident. For correction within each cell type, normalized expression data for each gene was centred by batch and centred again after merging batches. The batch correction was validated by PCA analysis (Suppl. Figure 1. (A and B)) and pair plots of PCs 1–4 did not reveal any further batch effects. BLASTn software was to identify probesets with unique sequences.

The data generated in this study have been deposited in the NCBI Gene Expression Omnibus (GEO, Series accession number GSE77350) and is publically accessible via this URL. (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77350)

To assess association between SNP genotype and gene expression, the MatrixEQTL R (30) package was used. To adjust for unknown confounders in the expression, two correction methods were used and results were compared.

RUV-2 correction

First association of all T1D SNPs with normalized uncorrected data was performed and p-value association of every SNP-gene pair was obtained. For each SNP, the top 5000 associated genes ranked by p-value were excluded and the rest were treated as empirical controls for RUV-2 correction (31–32) using *naiveRandRUV* method, parameter *k* set to 20. After correction, the same SNP was tested against the corrected set and p-value association of SNP-gene pair was recorded. This procedure was repeated for all SNPs and finally Benjamini FDR correction was applied to the set of recorded nominal *P*-values.

PCA correction

PCs were derived from individual whole expression sets and tested against whole genome Immunochip SNPs (200K). The PCs that showed no or weak genome association (i.e. min SNP-PC association FDR $P > 0.001$) were chosen as un-associated PCs (33). These PCs were incrementally added in their order of precedence as covariates to assess SNP-gene associations with an aim to maximize the number of significant *cis* gene detections (at FDR $P < 0.001$) for the 77 T1D SNPs tested. Based on analysis shown in Suppl. Figure 1 (E and F), the four gene expression datasets were corrected as follows: 7 PCs: 1–6 and 8 were removed from EBV-B basal cell line samples, 3 PCs: 1, 4 and 9 were removed for PMA stimulated EBV-B cell line samples, 4 PCs: 1–4 were removed for CD4+ samples and 2 PCs: 1 and 2 were removed for CD8+ samples.

We compared numbers of *cis*- and *trans*- regulated genes detected in each cell type using two methods (Suppl. Table I. (ii)). The RUV-2 method of correction yielded more significant results than PCA methods.

Statistical Analysis

Differential gene expression analysis was performed using the Limma package written for R (34). TDT (sibship) tests were performed using the software package UNPHASED (35–36).

Enrichment Analysis

Candidate gene names were converted to Entrez gene ids and were analysed using the DAVID (37–38) function annotation tool (http://david.abcc.ncifcrf.gov/). Further pathway and network analysis were performed using GATHER (gather.genome.duke.edu) (39) and GENEMANIA (www.genemania.org) (40) respectively.

RESULTS

Systematic evaluation of non-synonymous SNPs in genes in T1D-associated regions

First, we searched for commonly occurring non-synonymous (ns) SNPs in linkage disequilibrium (LD) ($r^2 > 0.8$) with the T1D SNPs (2, 13–25) in the 1000 genomes and HAPMAP (41) CEU datasets. All amino acid substitutions were subject to prediction of the effect of these changes, evaluated as 'benign', 'probably damaging' or 'possibly damaging' by PolyPhen-2 (10). This search returned 25 nsSNPs in strong LD with only 16 of the 60 non-HLA T1D loci. These SNPs occurred in a total of 22 unique genes. The seven potentially damaging effects were found in two genes, *SULT1A2* and *GSDMB*. Prediction status does not affect candidacy per se, so all genes listed in Table II should be evaluated in further studies. In addition, none of the four SNPs recently discovered in (13) were in strong LD ($r^2 > 0.8$) with any nsSNPs. Among the LD SNPs, there were three splice-region variants and one stop-gain variant (summarized in Suppl. Table I.(iii)).

Next, we searched whether any nsSNPs showed better association with T1D than the reported SNP itself. For this, we performed a transmission disequilibrium test (TDT – sibship test) using UNPHASED (35–36) on a dataset of 2,676 nuclear families with unaffected parents and two or more affected sibs. Results are presented in Table II. Association *P*-values for SNPs not included in the Immunochip genotyping were derived from (2). At six T1D loci, the nsSNPs were the reported best SNPs. From those nsSNPs that were genotyped by Immunochip, *rs7498665* associated with *SH2B1* showed slightly better association than the reported *rs4788084* ($\Delta P = 0.1$, where $\Delta P = P_{ns\ SNP} / P_{reported\ SNP}$). Two other ns SNPs (*rs2305480* and *rs229527*) also showed very small ($\Delta P > 0.1$) improvement in association compared to the reported T1D SNP. Most of the T1D loci did not have associated nsSNPs in nearby genes.

Gene Expression Analyses

EBV-transformed B cell lines (referred to hereafter as EVB-B) were produced from blood samples obtained from T1DGC family members (3). RNA was extracted from 202 available EBV-B cell lines that were cultured under basal conditions and stimulated with PMA. We also purified CD4+ and CD8+ T cells from peripheral blood samples provided by 113 subjects. None of these subjects overlapped with the donors of the 202 EBV samples. After quality control, sufficient high quality RNA to perform microarrays was obtained from 102 CD4+ T cell samples and 84 CD8+ T cell samples. The EVB-B cell samples were derived from both T1D cases and unaffected subjects. The unaffected controls included were first-degree relatives of the subset of case samples and islet autoantibody status was not determined for these unaffected subjects. The details regarding the autoantibody, sex and HLA-DR status of the affected subjects are summarized in Suppl. Table I.(i). As expected, there were no significant differences in the gene expression between cases and unaffected subjects nor between cases and unaffected first-degree relatives (Suppl. Figure 1.C and 1.D) so all samples were used to search for eQTLs.

These RNA samples were hybridized to Illumina microarrays (HT-12v4). Data processing was carried out as described in Methods. Batch effects were corrected for each cell type by

centering the normalized gene expression data by batch and centering again after merging batches. The batch correction was validated by PCA analysis (Suppl. Figure 1 (A & B)). To eliminate probesets with potential cross-hybridization problems, a BLAST search of each probe sequence was carried out on a custom database of all 47,323 Illumina probeset sequences and 38,500 probes that had a single hit were retained. In doing so, probes associated with two known T1D candidates *RPS26* (due to sequence similarity with probes associated with *RPS26* pseudogenes) and *DEXI* (due to sequence similarity with a probe associated with *LOC653752*) were removed. There were 95 *ERCC* spike-in controls in the probeset, which were also excluded from analysis. We also performed a search for SNPs within probeset coordinates and excluded any probes that contained SNPs from further analysis. We performed differential expression analysis of unstimulated EBV-B cells and after 6 hour PMA stimulation. The negative $\log_{10}$ (adjusted *P*-value) of each probe showing differential expression was plotted against the $\log_2$ fold change in a 'volcano plot' (Figure 1). Adjusted $P < 0.0001$ was selected as a cut-off for differential expression. A total of 1,465 genes were differentially expressed at this threshold with at least a modest fold change (absolute $\log_2 FC > 0.3$). Genes with the highest fold changes in expression included *CCL3, CCL4, EGR1, EGR2, DUSP21, PIP4K2C, ILDR1* and *IL9R*.

Parameters for Systems Genetics Analyses

Genotypes of T1DGC subjects were previously determined (2–4, 13) at 77 SNPs in 55 of 60 T1D risk loci (Table I). Based on the risk allele's code at each T1D SNP, an additive recode [0,1,2] was applied so that the risk allele's effect on gene expression could be determined. Separate analyses were performed for each of the four expression sets (EBV-B basal, EBV-B 6h PMA stimulated, CD4+ and CD8+). For these analyses, we conservatively defined a *cis* transcript as being from a gene whose transcription start or end site was located within 1 Mbp from the T1D SNP. A *trans* regulated transcript was defined as a gene located elsewhere in the genome. For each set, 3,672 *cis* interactions pairs were tested; ~2.9M *trans* interactions pairs were tested; and false discovery rate (FDR) *P*-value corrections were applied separately for *cis* and *trans* eQTLs. The MatrixEQTL R package (30) was used to perform these eQTL tests. Due to unknown confounding factors that could limit the power of detecting significantly differentially expressed genes, we performed two methods of correction independently: (a) removing unwanted variation (RUV-2) (31–32); (b) and adding genome wide un-associated expression derived principle components (PCs) as covariates (described in Methods).

All transcripts with FDR $P < 0.05$ for each T1D SNP were followed up with enrichment analysis using the DAVID bioinformatics resource (37–38). Additional pathway and network analysis was performed using GATHER (39) and GENEMANIA (40) respectively. The results from these analyses are summarized in Tables III – VII and are described below. Boxplots of eQTL associations can be accessed online through our web resource (42) where we compare effects explained by raw normalized gene expression against RUV-2 and PCA corrected gene expression sets. A screenshot of the user interface is shown in Figure 2.

Effect of T1D-associated non-HLA SNPs on neighboring gene expression in EBV-B Cell lines

We examined *cis* genes in EBV-B basal cell line samples at various FDR *P* value thresholds. At $P < 0.001$, 15 T1D SNPs were associated with differences in expression of 20 genes (Table III). Using lower thresholds of adjusted *P* [< 0.05], an additional 13 T1D SNPs affected the expression of a further 20 genes (Suppl. Table II.(i)). Hence, 28 T1D SNPs were found to be associated with changes in a total of 40 significant *cis* genes. Of these, three SNPs (*rs10877012*, *rs4788084* and *rs2290400*) showed strong *cis* effects with multiple nearby genes that were either up- or down- regulated by the corresponding risk allele. In testing the four newly discovered T1D SNPs, (13) we observed that the risk allele associated with *rs2611215* reduced expression of *TMEM192* (FDR $P = 0.008$) (Suppl. Table II.(i)).

Next, we tested 6h PMA stimulated EBV-B cell line samples. Results confirmed the *cis* effects associated with 22 of 40 candidate genes identified in unstimulated EBV-B cells (at minimum FDR $P < 0.05$) and the effect directions were consistent. *IFNGR1*, *SUOX*, *SPNS1* and *UBASH3A* were among genes that showed regulatory effects in basal cells but not after PMA stimulation. In addition, 17 T1D SNP genotypes significantly regulated the expression of 16 new candidate genes (FDR $P < 0.05$). Of these, genes *INO80B* and *LYRM2* were detected highly significant at FDR $P < 0.001$ (Table III). The expression of candidate genes *IKZF1* and *TSFM* showed decreased association with their corresponding T1D SNPs after stimulation, compared to basal condition (refer to 42). The rest of these results are presented in Suppl. Table II.(i).

In summary, 31 T1D SNPs affected the expression of a total of 38 candidate *cis* genes, 22 of which had shown evidence of *cis* effects in unstimulated EBV-B cells while the remaining nine showed association after PMA stimulation, thus suggesting genes that may play a role after immune activation.

Effect of T1D-associated non-HLA SNPs on neighboring gene expression in CD4+ and CD8+ T cells

Tests of CD4+ T cell samples revealed 16 T1D SNP genotypes regulated the expression of 20 genes significantly. Of these genes, eleven (*SMARCE1*, *LOC728734*, *SUOX*, *FAM119B*, *C16ORF75*, *GSDMB*, *IKZF1*, *ADCY3*, *ORMDL3*, *SKAP2* and *IKZF3*) were found to be *cis* regulated in both EBV-B and CD4+ T cells by the same T1D SNPs (Table III). In particular, the risk allele of *rs2290400* (T) affected nearby genes *ORMDL3*, *GSDMB* and *IKZF3* similar to that observed in EBV-B cells. The effect directions between the cell types for the 11 shared genes were consistent, except for gene *C16ORF75* where the risk allele increased expression in CD4+ cells but decreased it in EBV-B cells (42). We also noted expression of candidate gene *SUOX* showed a clear increase in the significance of association (i.e. lower p-value) with T1D risk allele *rs705704* (T) in the CD4+ cells compared to EBV-B cells. In addition, there were nine newly identified candidate genes associated with nine T1D SNPs. Five of these SNPs had showed *cis* effects in the EBV-B cells, but had affected a different set of genes. Among these 9 new candidate genes, *CLECL1* was the most significantly associated (Table III). The *cis* genes detected at lower FDR thresholds of 0.01 and 0.05 are

presented in the Suppl. Table II.(i). These results suggest that the effects of the T1D risk SNPs on gene expression vary between cell types.

Finally, we performed analyses of the CD8+ T cell samples and identified 17 T1D SNP genotypes regulated the expression of 19 genes across all samples tested. Excepting *ADCY3*, ten candidate genes were found *cis* regulated in EBV-B, CD4+ T cells and CD8+ cells. Thirteen of the 19 candidate genes were *cis* regulated in both CD4+ and CD8+ T cells and the effect directions were consistent. The remaining six that were neither differentially regulated in EBV-B cells nor in CD4+ cells were associated with six T1D SNPs in CD8+ cells. Of these, T1D SNP *rs2292239* regulated the expression of candidate gene *ERBB3* most significantly (FDR $P < 0.001$) (Table III). The rest of the results are presented in Suppl. Table II.(i).

In summary, 24 T1D SNP genotypes regulated the expression of 31 candidate genes highly significantly at FDR $P < 0.001$ (Table III). Using lower FDR adjusted *P*-value thresholds ($P < 0.05$), 43 T1D SNP genotypes regulated the expression of 71 candidate genes. Using even lesser stringent suggestive threshold of nominal un-adjusted $P < 0.001$ for evidence of *cis* effect, we could define up to 85 candidate genes that were affected by 50 T1D SNPs in the four cell types tested.

T1D-associated SNPs associated with changes in expression of distant genes

Next, we investigated whether T1D loci showed *trans*-regulatory effects. After performing ~2.9M tests for each cell type and appropriate statistical correction, we identified 38 genes that were highly significantly associated with 25 T1D SNPs at FDR $P < 0.001$ (Table IV). Five of these SNPs (*rs1534422, rs1990760, rs11571291, rs9585056* and *rs425105*) did not show any *cis* effect on nearby genes in the cell types tested. *Trans*-regulated genes shared between B and T cells were detected at only one T1D locus (defined by T1D SNP *rs705704*) and the effect direction was consistent. Except for *ZMYM5*, *GRAMD1B* and *LOC389386*, all significant *trans* genes were detected in the EBV-B cells. Upon characterizing the function of 38 *trans* genes in DAVID (37–38), we identified two clusters: *CD276, ST6GAL1, CCL5* and *IRF8* were associated with immune response and a further two genes (*ID2* and *IRF8)* were associated with immune system and hemopoietic (lymphoid) organ development. Eight T1D SNPs (Table IV: highlighted in bold) showed highly significant *cis* as well as *trans* regulatory interactions in one or more cell types tested, suggesting co-regulation between *cis* and *trans* genes. We describe tests for meaningful relationships between these genes in the next section.

In summary, in addition to the loci that affected genes in *cis*, we could identify five loci that exclusively affected genes in *trans*. Of the T1D loci that were not associated with expression changes in any of the four cell types, three loci contained non-synonymous SNPs defined in Table II. The *trans* regulatory effects detected at lower threshold levels are presented in Suppl. Table II (ii) and Suppl. Table III.

Enrichment analysis of genes associated with T1D susceptibility

We investigated the function of the genes whose expression was changed by individual risk SNPs. The DAVID enrichment analysis software (37–38) tests whether sets of genes are enriched for terminology referenced by UniProt Protein Information Resource (PIR) keywords, Gene Ontology (GO) and KEGG Pathways. First, we performed analysis to explore for enrichment between the highly significant (FDR $P < 0.001$) *cis* and *trans* gene candidates for the eight T1D SNPs highlighted in Table IV. For three of these SNPs, the candidate genes shared a common keyword (Table V). Second, using the list of 86 candidate genes derived from Tables II – IV, we performed pathway and enrichment analysis using GATHER (39) and we report results obtained with high confidence (unadjusted $P < 0.001$) in Table VI. In these results, we found that the cytokine-cytokine receptor interaction pathway received the highest significance. Third, we performed network analysis using GENEMANIA (40) for the same list of 86 candidate genes. The significant functional findings are presented in Table VI. The full GENEMANIA report can be accessed online (www.sysgen.org/T1DGCSysGen/genemania.pdf). Finally, we analyzed the list of *cis* and *trans* genes detected at FDR $P < 0.05$ for every T1D SNP separately. We identified 21 enrichment terms (excluding Gene Ontology cellular component terms) that were significantly enriched at Benjamini $P < 0.05$ for 10 T1D SNPs. These results are summarized in Table VII and below.

The term ‘Lectin’ was highly enriched for the T1D locus defined by *rs10466829* since it affected expression of five c-type lectin genes (*CLEC1A, CLEC2B, CLEC2D, CLECL1* and *CD69*) in the cell types tested. The T1D locus defined by *rs17696736* was highly enriched for ‘response to virus’ and ‘anti-viral defense’ due to changes in expression of 7 *trans* genes (*EIF2AK2, IFI16, IFNGR1, MX1, MX2, PLSCR1* and *STAT1*). Furthermore, genes *MX1* and *MX2* are also known inflammatory and immune response genes. In addition, the T1D SNP *rs416603* showed significant enrichment for ‘*IL10*-anti inflammatory signaling pathway’ and ‘intestinal immune network IgA production pathway’ through its regulation of three genes (*IL10, IL10RA* and *STAT5A*). We also noted that two risk SNPs (*rs2476601* and *rs679574*) showed association in *trans* with genes in the MHC (*HLA-F, G, H*, and *DRB4*), which gave positive enrichment for terms such ‘antigen processing and presentation’. These results provide insights into the functions of genes whose expression is affected by the T1D loci.

Validation of *trans*-regulatory gene interactions

To confirm our results, we searched using the blood eQTL browser (43) for the *trans* regulatory associations we identified at significance threshold FDR $P < 0.05$. Since not all T1D SNPs may be present in this browser, we allowed a 100Kb window for the search of the expression SNP. Two *trans* genes were validated: *UBE2L6* (EBV-B +/− PMA) associated with *rs3184504* and *STAT1* (EBV-B basal) with *rs17696736*. Secondly, we searched in the *trans* regulatory interactions reported by Fairfax *et al*. (44) and validated a further three gene interactions reported in their study: *LOC728823, IP6K2* and *LOC389386* all associated with the T1D SNP *rs705704*. Although many *cis* gene effects were clearly defined from our datasets, validating *trans* genes poses a challenge warranting further investigation.

DISCUSSION

Our results provide a potential molecular basis for disease association at 46 of the 59 identified T1D loci (Table I). Sixteen of these loci contained non-synonymous SNPs in strong LD with the T1D SNP. Thirty-six of the loci showed *cis* effects on 75 nearby genes. The remainder showed statistically significant *trans* regulatory interactions that were substantiated by significant enrichment results (Tables V–VII). These candidate genes can be the focus for further studies. For example, a systems genetics study (45) into candidate gene *CTSH* whose expression was affected by T1D SNP *rs3825932*, supported its product as a novel therapeutic target.

Onengut-Gumuscu *et al*. (13) recently confirmed several previously reported T1D associated SNPs (2, 5) in addition to the identification of four additional new T1D risk SNPs of which one SNP (*rs2611215*) had high significance ($P = 1.817 \times 10^{-11}$) while *P* values of the rest only just exceeded the significance threshold ($P < 5 \times 10^{-8}$). This study found that the associated SNPs localized to enhancer sequences active in thymus, T and B cells, and CD34+ stem cells. Of the four new T1D associated SNPs (13) we were able to establish likely candidacy for *rs2611215* as *TMEM192*.

An important conclusion from our study is that the cell type was important in characterizing T1D SNP function, i.e. eQTLs are cell type-specific. For example, the candidate gene *ERBB3* was highly significantly *cis* regulated in CD8+ T cells but its variation effect was largely undetectable in other cell types. The risk allele associated with *rs4788084* reduced expression of candidate gene *TUFM* exclusively in the CD4+ cells. Similarly, *CLECL1* did not show any effect in EBV-B cell lines but showed highly significant effects in both T cell types tested. Among the weakly detected effects, there was evidence that suggested the risk allele associated with *rs231727* reduced expression in *cis* of a well known candidate (*CTLA4)* exclusively in the CD8+ cells (unadjusted *P*=0.0003, FDR *P* = 0.04) (Suppl. Table II. (i)). Our CD4+/ CD8+ cell type data also assisted in mapping candidate genes at otherwise anonymous T1D SNPs; the most significant of these candidates included *SLC11A1* (*rs3731865*), *C6Orf173* (*rs9388489*) and *C10orf59* (*rs10509540*).

Sixteen transcripts (twelve in *cis*, four in *trans*) were significantly associated with T1D SNPs in both EBV-B and the T cell types tested. Of these, a novel uncharacterized *cis* transcript *LOC728734* (nuclear pore complex interacting protein family, member B8) was identified to be associated with T1D SNP *rs4788084* (chr 16p11.2) where the risk allele decreased expression in all four cell types. The effect direction of *cis* and *trans* regulation by T1D SNPs on genes detected across multiple cell types were found consistent for all SNPs except *C16ORF75*. We also noted that probes associated with candidate genes *DEXI* and *RPS26* also showed strong *cis* regulatory effect in association with T1D risk SNPs *rs12708716* and *rs705704*, respectively, in one or more cell types. However, due to quality control procedures relevant to cross hybridisation problems described in the previous section, these probes were excluded from further analysis. Non-synonymous SNPs may also affect gene expression in *trans*. We found two examples of these: *rs1990760* (chr 2q24.2) in *IFIH1* also affected the expression of *LOC643997* in *trans*; similarly *rs2304256* (chr 19p13.2) in *TYK2* also affected the expression of *ZNF280D* in *trans*.

Pathway analysis identified the “Cytokine-cytokine receptor interaction” pathway with highest confidence. The “Sulfur metabolism” pathway also scored high significance because two genes *SUOX* (*cis*) and *SULT1A2* (non-synonymous) involved in this pathway were identified as candidates in this study. It is also well known that sulfur plays an important role in insulin production (see 46, for review). Furthermore, DAVID enrichment analysis of locus specific *cis* and *trans* transcript perturbations revealed significant enrichment of 48 category terms in 15 of the T1D regions at FDR $P < 0.05$. Among the best enriched terms were ‘response to virus’, ‘acetylation’, ‘lectin’ and ‘IL10-anti-inflamatory pathway’. From the enrichment analysis for genes associated with each T1D SNP, upon examination T1D risk SNP *rs17696736* (chr 12q24.12) was notably associated with ‘response to virus’ and ‘antiviral defence’ due to *trans* genes that are involved in pro-inflamatory response (such as *MX1*, *MX2*) in the salmonella infection pathway (KEGG Pathway: 05132). In contrast, chemokine gene *CCL5* was highly significantly associated with diabetes loci associated with T1D SNP *rs425105* (chr 19q13.32). These results support evidence found in a recent work (47) suggesting salmonella and chemokine vaccines can prove clinically useful in diabetes management and prevention.

In conclusion, our results confirm systems genetics (12) as a powerful tool for investigating the genetic architecture of complex diseases such as T1D. Many genes were identified whose expression levels were influenced by SNPs associated with T1D susceptibility. These nsSNPs, *cis* and *trans* regulated genes we identified are important candidates for further investigation. So that other researchers can extend the work reported here, we have implemented a web interface (42) allowing users to browse boxplots for the eQTL interactions reported below.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by NIDDK, National Institute of Allergy and Infectious Diseases, National Human Genome Research Institute, National Institute of Child Health and Human Development and Juvenile Diabetes Research Foundation International.

This work was supported by Program Grants 53000400 and 37612600 from the National Health and Medical Research Council of Australia, by the Diabetes Research Foundation (WA) and by grant 1DP3DK085678 from the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). R.R. is supported by MACA Ride to Conquer Cancer in association with Harry Perkins Institute of Medical research. B.O.B. is supported by the Deutsche Forschungsgemeinschaft and a grant from the Boehringer Ingelheim Ulm University BioCenter.

REFERENCES

1. Danaei G, Finucane MM, Lu Y, Singh GM, Cowan MJ, Paciorek CJ, Lin JK, Farzadfar F, Khang YH, Stevens GA, Rao M, Ali MK, Riley LM, Robinson CA, Ezzati M. National, regional, and global trends in fasting plasma glucose and diabetes prevalence since 1980: systematic analysis of health examination surveys and epidemiological studies with 370 country-years and 2.7 million participants. Lancet. 2011; 378:31–40. [PubMed: 21705069]
2. Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA, Julier C, Morahan G, Nerup J, Nierras C, Plagnol V, Pociot F, Schuilenburg H, Smyth DJ, Stevens H, Todd JA, Walker

NM, Rich SS. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet. 2009; 41:703–707. [PubMed: 19430480]

3. Rich SS, Akolkar B, Concannon P, Erlich H, Hilner JE, Julier C, Morahan G, Nerup J, Nierras C, Pociot F, Todd JA. Overview of the Type I Diabetes Genetics Consortium. Genes Immun. 2009; 10(Suppl 1):S1–S4. [PubMed: 19956093]
4. Morahan G, Mehta M, James I, Chen WM, Akolkar B, Erlich HA, Hilner JE, Julier C, Nerup J, Nierras C, Pociot F, Todd JA, Rich SS. Tests for genetic interactions in type 1 diabetes: linkage and stratification analyses of 4,422 affected sib-pairs. Diabetes. 2011; 60:1030–1040. [PubMed: 21266329]
5. Morahan G. Insights into type 1 diabetes provided by genetic analyses. Curr Opin Endocrinol Diabetes Obes. 2012; 19:263–270. [PubMed: 22732486]
6. Editorial. On beyond GWAS. Nat Genet. 2010; 42:551. [PubMed: 20581872]
7. Freedman ML, Monteiro AN, Gayther SA, Coetzee GA, Risch A, Plass C, Casey G, De Biasi M, Carlson C, Duggan D, James M, Liu P, Tichelaar JW, Vikis HG, You M, Mills IG. Principles for the post-GWAS functional characterization of cancer risk loci. Nat Genet. 2011; 43:513–518. [PubMed: 21614091]
8. Sacks DB, Arnold M, Bakris GL, Bruns DE, Horvath AR, Kirkman MS, Lernmark A, Metzger BE, Nathan DM. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. Diabetes Care. 2011; 34:e61–e99. [PubMed: 21617108]
9. Morahan, G.; Varney, M. The Genetics of Type 1 Diabetes. In: Mehra, NK.; Kaur, G.; McCluskey, J.; Christiansen, FT.; Claas, FHJ., editors. The HLA Complex in Biology and Medicine: A Resource Book. 1st. New Delhi: Jaypee Brothers Medical Publishers (P) Ltd; 2010. p. 205-218.
10. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. A method and server for predicting damaging missense mutations. Nat Methods. 2010; 7:248–249. [PubMed: 20354512]
11. Veyrieras JB, Kudaravalli S, Kim SY, Dermitzakis ET, Gilad Y, Stephens M, Pritchard JK. High-resolution mapping of expression-QTLs yields insight into human gene regulation. PLoS Genet. 2008; 4:e1000214. [PubMed: 18846210]
12. Morahan G, Williams RW. Systems genetics: the next generation in genetics research? Novartis Found Symp. 2007; 281:181–188. discussion 188-91, 208-9. [PubMed: 17534074]
13. Onengut-Gumuscu S, Chen WM, Burren O, Cooper NJ, Quinlan AR, Mychaleckyj JC, Farber E, Bonnie JK, Szpak M, Schofield E, Achuthan P, Guo H, Fortune MD, Stevens H, Walker NM, Ward LD, Kundaje A, Kellis M, Daly MJ, Barrett JC, Cooper JD, Deloukas P, Todd JA, Wallace C, Concannon P, Rich SS. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. Nat Genet. 2015; 47:381–386. [PubMed: 25751624]
14. Cooper JD, Smyth DJ, Smiles AM, Plagnol V, Walker NM, Allen JE, Downes K, Barrett JC, Healy BC, Mychaleckyj JC, Warram JH, Todd JA. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat Genet. 2008; 40:1399–1401. [PubMed: 18978792]
15. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, C. Ionescu-Tîrgovişte, Genetics of Type 1 Diabetes in Finland; Simmonds MJ, Heward JM, Gough SC, Wellcome Trust Case Control Consortium. Dunger DB, Wicker LS, Clayton DG. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet. 2007; 39:857–864. [PubMed: 17554260]
16. Smyth DJ, Plagnol V, Walker NM, Cooper JD, Downes K, Yang JH, Howson JM, Stevens H, McManus R, Wijmenga C, Heap GA, Dubois PC, Clayton DG, Hunt KA, van Heel DA, Todd JA. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N Engl J Med. 2008; 359:2767–2777. [PubMed: 19073967]
17. Bradfield JP, Qu HQ, Wang K, Zhang H, Sleiman PM, Kim CE, Mentch FD, Qiu H, Glessner JT, Thomas KA, Frackelton EC, Chiavacci RM, Imielinski M, Monos DS, Pandey R, Bakay M, Grant

SF, Polychronakos C, Hakonarson H. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet. 2011; 7:e1002293. [PubMed: 21980299]

18. Yang JH, Downes K, Howson JM, Nutland S, Stevens HE, Walker NM, Todd JA. Evidence of association with type 1 diabetes in the *SLC11A1* gene region. BMC Med Genet. 2007; 12:59. [PubMed: 21524304]
19. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]
20. Swafford AD, Howson JM, Davison LJ, Wallace C, Smyth DJ, Schuilenburg H, Maisuria-Armer M, Mistry T, Lenardo MJ, Todd JA. An allele of IKZF1 (Ikaros) conferring susceptibility to childhood acute lymphoblastic leukemia protects against type 1 diabetes. Diabetes. 2011; 60:1041–1044. [PubMed: 21270240]
21. Evangelou M, Smyth DJ, Fortune MD, Burren OS, Walker NM, Guo H, Onengut-Gumuscu S, Chen WM, Concannon P, Rich SS, Todd JA, Wallace C. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. Genet Epidemiol. 2014; 38:661–670. [PubMed: 25371288]
22. Bailey R, Cooper JD, Zeitels L, Smyth DJ, Yang JH, Walker NM, Hyppönen E, Dunger DB, Ramos-Lopez E, Badenhoop K, Nejentsev S, Todd JA. Association of the vitamin D metabolism gene CYP27B1 with type 1 diabetes. Diabetes. 2007; 56:2616–2621. [PubMed: 17606874]
23. Huang J, Ellinghaus D, Franke A, Howie B, Li Y. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 Data. Eur J Hum Genet. 2012; 20:801–805. [PubMed: 22293688]
24. Heinig M, Petretto E, Wallace C, Bottolo L, Rotival M, Lu H, Li Y, Sarwar R, Langley SR, Bauerfeind A, Hummel O, Lee YA, Paskas S, Rintisch C, Saar K, Cooper JD, Buchan R, Gray EE, Cyster JG, Cardiogenics Consortium. Erdmann J, Hengstenberg C, Maouche S, Ouwehand WH, Rice CM, Samani NJ, Schunkert H, Goodall AH, Schulz H, Roider HG, Vingron M, Blankenberg S, Münzel T, Zeller T, Szymczak S, Ziegler A, Tiret L, Smyth DJ, Pravenec M, Aitman TJ, Cambien F, Clayton D, Todd JA, Hubner N, Cook SA. A trans-acting locus regulates an anti-viral expression network and type 1 diabetes risk. Nature. 2010; 467:460–464. [PubMed: 20827270]
25. Wallace C, Smyth DJ, Maisuria-Armer M, Walker NM, Todd JA, Clayton DG. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. Nat Genet. 2010; 42:68–71. [PubMed: 19966805]
26. Burster T, Boehm BO. Processing and presentation of (pro)-insulin in the MHC class II pathway: the generation of antigen-based immunomodulators in the context of type 1 diabetes mellitus. Diabetes Metab Res Rev. 2010; 26:227–238. [PubMed: 20503254]
27. Rosinger S, Nutland S, Mickelson E, Varney MD, Boehm BO, Olsem GJ, Hansen JA, Nicholson I, Hilner JE, Perdue LH, Pierce JJ, Akolkar B, Nierras C, Steffes MW. Collection and processing of whole blood for transformation of peripheral blood mononuclear cells and extraction of DNA: the Type 1 Diabetes Genetics Consortium. Clin Trials. 2010; 7:S65–S74. [PubMed: 20595244]
28. Nomura N, Nomura M, Sugiyama K, Hamada J. Phorbol 12-myristate 13-acetate (PMA)-induced migration of glioblastoma cells is mediated via p38MAPK/Hsp27 pathway. Biochem Pharmacol. 2007; 74:690–701. [PubMed: 17640620]
29. Du P, Kibbe WA, Lin SM. lumi: a pipeline for processing Illumina microarray. Bioinformatics. 2008; 24:1547–1548. [PubMed: 18467348]
30. Shabalin AA. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. Bioinformatics. 2012; 28:1353–1358. [PubMed: 22492648]
31. Risso D, Ngai J, Speed TP, Dudoit S. Normalization of RNA-seq data using factor analysis of control genes or samples. Nat Biotechnol. 2014; 32:896–902. [PubMed: 25150836]
32. Jacob L, Gagnon-Bartsch JA, Speed TP. Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed. Biostatistics. 2015
33. Flutre T, Wen X, Pritchard J, Stephens M. A Statistical Framework for Joint eQTL Analysis in Multiple Tissues. PLoS Genet. 2013; 9:e1003486. [PubMed: 23671422]
34. Smyth GK. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol. 2004; 3 Article3.

35. Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum Hered. 2008; 66:87–98. [PubMed: 18382088]
36. Dudbridge F, Holmans PA, Wilson SG. A flexible model for association analysis in sibships with missing genotype data. Ann Hum Genet. 2011; 75:428–438. [PubMed: 21241274]
37. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc. 2009; 4:44–57. [PubMed: 19131956]
38. Huang da W, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res. 2009; 37:1–13. [PubMed: 19033363]
39. Chang JT, Nevins JR. GATHER: A Systems Approach to Interpreting Genomic Signatures. Bioinformatics. 2006; 22:2926–2933. [PubMed: 17000751]
40. Warde-Farley D, Donaldson SL, Comes O, Zuberi K, Badrawi R, Chao P, Franz M, Grouios C, Kazi F, Lopes CT, Maitland A, Mostafavi S, Montojo J, Shao Q, Wright G, Bader GD, Morris Q. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. Nucleic Acids Res. 2010; 38:214–220.
41. http://www.hapmap.org/hapmart.html.en.
42. http://www.sysgen.org/T1DGCSysGen/.
43. Westra HJ, Peters MJ, Esko T, Yaghootkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, Zhernakova A, Zhernakova DV, Veldink JH, Van den Berg LH, Karjalainen J, Withoff S, Uitterlinden AG, Hofman A, Rivadeneira F, 't Hoen PA, Reinmaa E, Fischer K, Nelis M, Milani L, Melzer D, Ferrucci L, Singleton AB, Hernandez DG, Nalls MA, Homuth G, Nauck M, Radke D, Volker U, Perola M, Salomaa V, Brody J, Suchy-Dicey A, Gharib SA, Enquobahrie DA, Lumley T, Montgomery GW, Makino S, Prokisch H, Herder C, Roden M, Grallert H, Meitinger T, Strauch K, Li Y, Jansen RC, Visscher PM, Knight JC, Psaty BM, Ripatti S, Teumer A, Frayling TM, Metspalu A, van Meurs JB, Franke L. Systematic identification of trans eQTLs as putative drivers of known disease associations. Nat Genet. 2013; 45:1238–1243. [PubMed: 24013639]
44. Fairfax BP, Makino S, Radhakrishnan J, Plant K, Leslie S, Dilthey A, Ellis P, Langford C, Vannberg FO, Knight JC. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. Nat Genet. 2012; 44:502–510. [PubMed: 22446964]
45. Floyel T, Brorsson C, Nielsen LB, Miani M, Bang-Berthelsen CH, Friedrichsen M, Overgaard AJ, Berchtold LA, Wiberg A, Poulsen P, Hansen L, Rosinger S, Boehm BO, Ram R, Nguyen Q, Mehta M, Morahan G, Concannon P, Bergholdt R, Nielsen JH, Reinheckel T, von Herrath M, Vaag A, Eizirik DL, Mortensen HB, Storling J, Pociot F. CTSH regulates beta-cell function and disease progression in newly diagnosed type 1 diabetes patients. Proc Natl Acad Sci U S A. 2014; 111:10305–10310. [PubMed: 24982147]
46. Wallace JL, Wang R. Hydrogen sulfide-based therapeutics: exploiting a unique but ubiquitous gasotransmitter. Nat Rev Drug Discov. 2015; 14:329–345. [PubMed: 25849904]
47. Husseiny MI, Rawson J, Kaye A, Nair I, Todorov I, Hensel M, Kandeel F, Ferreri K. An oral vaccine for type 1 diabetes based on live attenuated Salmonella. Vaccine. 2014; 32:2300–2307. [PubMed: 24631074]

ABBREVIATIONS

T1D	Type 1 Diabetes
HLA	Human Leukocyte Antigen
MHC	Major histocompatibility complex
eQTL	expression Qualitative trait locus
eSNP	expression Single Nucleotide Polymorphism

TDT	Transmission Disequilibrium Test of Association
T1DGC	Type 1 Diabetes Genetics Consortium
LD	Linkage Disequilibrium
PBMC	Peripheral blood mononuclear cells
EBV	Epstein-Barr virus
PMA	phorbol. 12-myristate 13-acetate
GWAS	Genome Wide Association Study
WGA	Whole Genome Association

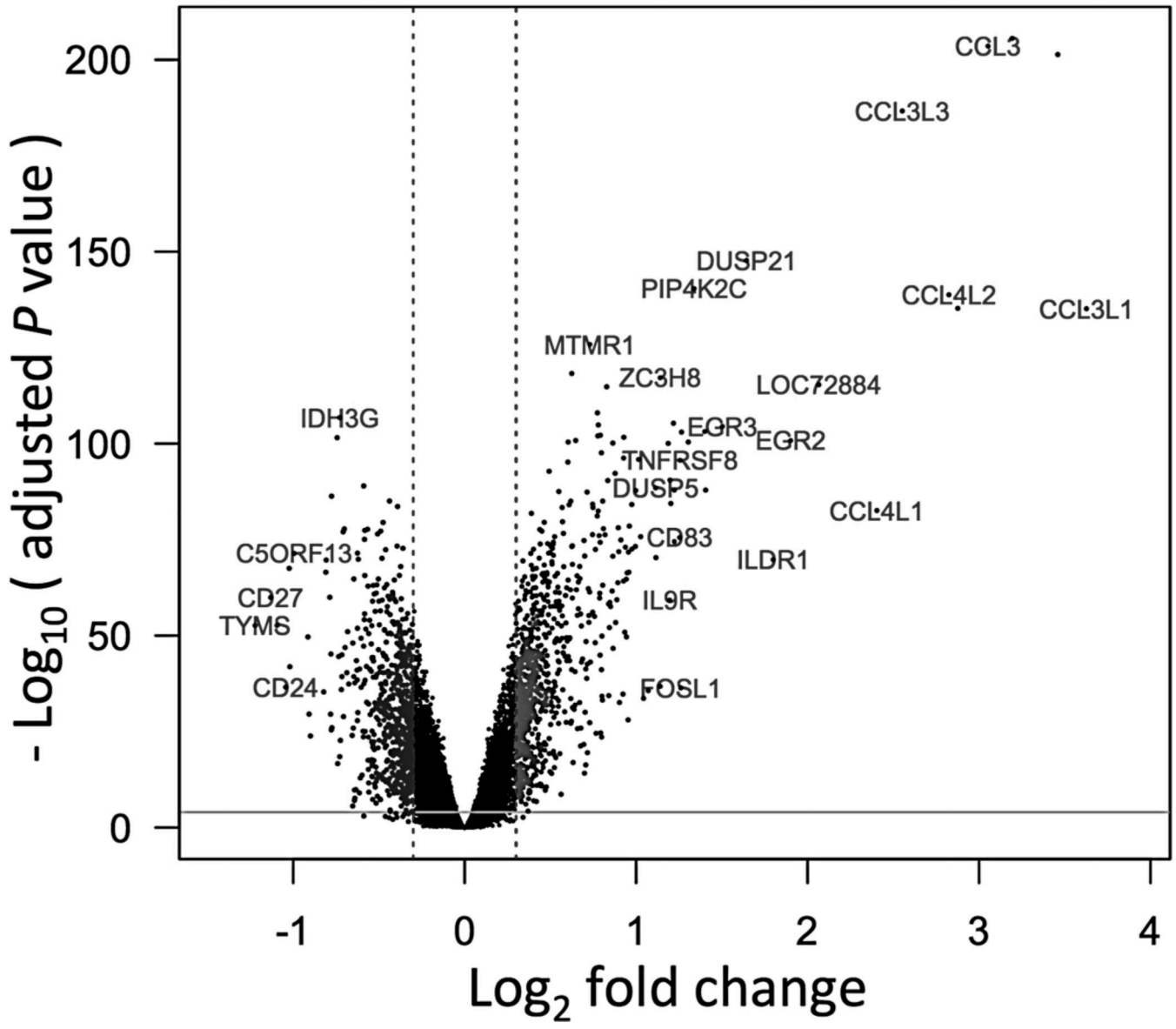

Figure. 1.

Comparison of gene expression in EBV-B cells between basal and 6 hour PMA stimulated samples. Differences (log_2 fold change) in gene expression are shown on the X axis; Y axis shows $-log_{10}$ (adjusted P values).

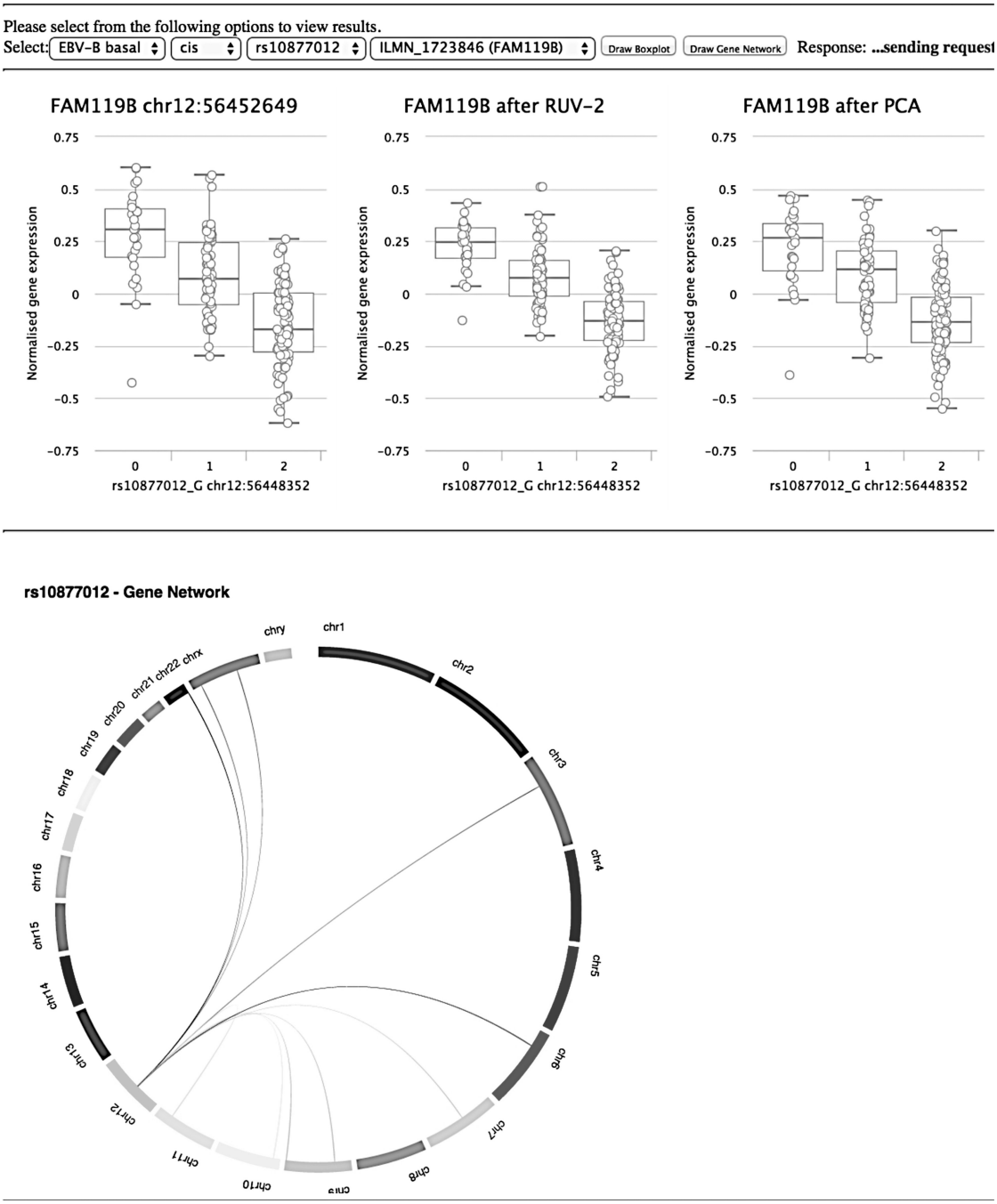

Figure. 2.

Screen capture of the Web interface for browsing box-plots and gene networks. URL: http://www.sysgen.org/T1DGCSysGen/

Table I

List of reported T1D SNPs located in 59 non HLA T1D loci.

ID	Locus	T1D SNP	CHR	BP	Gene	P-value	Ref.	Tables
1	1p31.3	rs2269241	chr1	63881359	*PGM1*	4.00×10^{-7}	2	II, S
2	1p13.2	rs2476601	chr1	114179091	*PTPN22*	8.50×10^{-85}	2	II
3	1q31.2	rs2816316	chr1	190803436	*RGS1*	3.10×10^{-5}*	2	
4	1q32.1	rs3024493	chr1	205010591	*IL10*	1.90×10^{-9}	2	
5	2p24.3	rs1534422	chr2	12558192	*Intergenic*	2.00×10^{-6}	2	IV
6	2p23.3	rs2165738	chr2	24546313	*Intergenic*	4.00×10^{-6}	14	III, S
7	2p13.1	rs363609^	chr2	74756380	*DQX1*	8.53×10^{-6}^	15	II, III, S
8	2q11.2	rs9653442	chr2	100191799	*AFF3*	5.00×10^{-6}	15	S
9	2q12.1	rs6543134	chr2	102416890	*IL18RAP*	8.03×10^{-5}*	16	III, S
10	2q24.2 2q24.2	rs1990760 rs3747517	chr2 chr2	162832297 162837070	*IFIH1* *IFIH1*	6.60×10^{-9} 4.70×10^{-7}	2 2	II, IV
11	2q33.2 2q33.2 2q33.2	rs11571291 rs3087243 rs231727	chr2 chr2 chr2	204429377 204447164 204449795	*CTLA4* *CTLA4* *CTLA4*	1.19×10^{-12} 1.20×10^{-15} 2.13×10^{-18}	17 2 13	IV II, S
12	2q35	rs3731865	chr2	218958247	*SLC11A1*	1.55×10^{-6}	18	III
13	3p21.31	rs11711054	chr3	46320615	*CCR5*	1.70×10^{-5}*	2	IV
14	4p15.2	rs10517086	chr4	25694609	*Intergenic*	4.60×10^{-10}	2	
15	4q27 4q27 4q27	rs4505848 rs17388568 rs2069763	chr4 chr4 chr4	123351942 123548812 123596932	*IL2* *IL2* *IL2*	4.70×10^{-13} 3.00×10^{-6} 1.91×10^{-10}	2 19 13	
16	5p13.2	rs6897932	chr5	35910332	*IL7R*	8.00×10^{-6}	15	II
17	6q15 6q15	rs597325 rs56297233	chr6 chr6	91059215 91070750	*BACH2* *BACH2*	3.38×10^{-10} 5.40×10^{-8}	17 2	S III
18	6q22.32	rs9388489	chr6	126740412	*C6orf173*	4.20×10^{-13}	2	III
19	6q23.3 6q23.3	rs2327832 rs10499194	chr6 chr6	138014761 138044330	*TNFAIP3* *TNFAIP3*	1.60×10^{-4}* 3.00×10^{-4}*	2 2	S III, IV
20	6q25.3	rs1738074	chr6	159385965	*TAGAP*	7.59×10^{-9}	16	S
21	7p15.2	rs7804356	chr7	26858190	*Intergenic*	5.30×10^{-9}	2	III, IV

ID	Locus	T1D SNP	CHR	BP	Gene	P-value	Ref.	Tables
22	7p12.2	rs10272724	chr7	50444707	*IKZF1*	4.80×10^{-9}	20	III
23	9p24.2 9p24.2	rs7020673 rs10758593	chr9 chr9	4281747 4282083	*GLIS3* *GLIS3*	5.40×10^{-12} 1.18×10^{-8}	2 17	IV, S S
24	10p15.1	rs12251307	chr10	6163501	*IL2RA*	1.30×10^{-13}	2	IV
25	10p15.1 10p15.1	rs947474 rs11258747	chr10 chr10	6430456 6512897	*PRKCQ* *PRKCQ*	4.00×10^{-9} 1.20×10^{-7}	14 2	IV, S
26	10p11.22	rs722988	chr10	33466153	*NRP1*	4.88×10^{-8}	21	S
27	10q23.31	rs10509540	chr10	90013013	*C10orf59*	1.30×10^{-28}	2	III, S
28	11p15.5 11p15.5 11p15.5	rs7928968 rs3842727 rs7111341	chr11 chr11 chr11	2006875 2141424 2169742	*INS* *TH* *INS*	2.78×10^{-14} 4.89×10^{-196} 4.40×10^{-48}	17 13 2	IV
29	12p13.31 12p13.31 12p13.31	rs3764021 rs10466829 rs4763879	chr12 chr12 chr12	9724895 9767358 9801431	NR *CLECL1* *CD69*	5.00×10^{-8} 9.19×10^{-9} 1.90×10^{-11}	19 17 2	III, S III, S
30	12q13.2 12q13.2 12q13.2	rs705704 rs11171739 rs2292239	chr12 chr12 chr12	54721679 54756892 54768447	*ERBB3* *ERBB3* *ERBB3*	4.31×10^{-31} 1.00×10^{-11} 2.20×10^{-25}	17 19 2	III, IV, S III, S
31	12q13.3	rs3809114	chr12	56134906	NR	6.90×10^{-4}*	2	
32	12q14.1	rs10877012	chr12	56448352	*CYP27B1*	3.80×10^{-6}	22	II, III, S
33	12q24.12 12q24.12	rs1265565 rs3184504	chr12 chr12	110199580 110368991	*CUX2* *SH2B3*	1.00×10^{-16} 2.80×10^{-27}	23 2	IV, S II
34	12q24.13	rs17696736	chr12	110971201	*C12orf30*	1.73×10^{-13}	15	IV, S
35	13q32.3	rs9585056	chr13	98879767	*GPR183*	5.20×10^{-9}	24	IV
36	14q24.1	rs1465788	chr14	68333352	*Intergenic*	1.80×10^{-12}	2	
37	14q32.2	rs4900384	chr14	97568704	*Intergenic*	3.70×10^{-9}	2	
38	14q32.2	rs56994090	chr14	100376200	*DLK1*	1.62×10^{-10}	25	
39	15q14 15q14	rs17574546 rs12908309	chr15 chr15	36689768 36715969	*RASGRP1* *RASGRP1*	3.35×10^{-8} 4.31×10^{-8}	13 17	III IV
40	15q25.1	rs3825932	chr15	77022501	*CTSH*	7.70×10^{-8}	2	II, III, IV
41	16p13.13 16p13.13 16p13.13	rs12708716 rs12927355 rs416603	chr16 chr16 chr16	11087374 11102272 11271580	*CLEC16A* *DEXI* *C16orf75*	2.20×10^{-16} 1.91×10^{-16} 3.00×10^{-6}	2 17 14	II, III, IV

ID	Locus	T1D SNP	CHR	BP	Gene	P-value	Ref.	Tables
42	16p11.2 16p11.2	rs4788084 rs9924471	chr16 chr16	28447349 28499031	*IL27* *IL27*	2.60×10^{-13} 1.21×10^{-11}	2 17	II, III, S II
43	16q23.1 16q23.1	rs7202877 rs8056814	chr16 chr16	73804746 73809828	*Intergenic* NR	3.10×10^{-15} 1.13×10^{-7}	2 17	
44	17p13.1	rs16956936	chr17	7574417	*Intergenic*	5.00×10^{-7}	2	
45	17q12	rs2290400	chr17	35319766	*ORMDL3*	5.50×10^{-13}	2	II, III, IV, S
46	17q21.2	rs7221109	chr17	36023812	*Intergenic*	1.30×10^{-9}	2	III, IV, S
47	18p11.21	rs1893217	chr18	12799340	*PTPN2*	3.60×10^{-15}	2	IV, S
48	18q22.2	rs763361	chr18	65682622	*CD226*	1.56×10^{-8}	16	II, III, IV
49	19p13.2	rs2304256	chr19	10336652	*TYK2*	4.13×10^{-9}	25	II, IV, S
50	19q13.32	rs425105	chr19	51900321	*Intergenic*	2.70×10^{-11}	2	IV
51	19q13.33	rs679574	chr19	53897920	*FUT2*	4.30×10^{-18}	16	IV, S
52	20p13	rs2281808	chr20	1558551	*Intergenic*	1.20×10^{-11}	2	II
53	21q22.3 21q22.3	rs11203203 rs876498	chr21 chr21	42709255 42714896	*UBASH3A* *UBASH3A*	1.70×10^{-9} 7.06×10^{-9}	2 13	III, S IV
54	22q12.2	rs5753037	chr22	28911722	*Intergenic*	2.60×10^{-16}	2	III, S
55	22q12.3	rs229541	chr22	35921264	*C1QTNF6*	2.10×10^{-7}	2	II, S
Four newly discovered loci								
56	1q32.1	rs6691977	chr1	200814959	NR	4.30×10^{-8}	13	
57	2q13	rs4849135	chr2	111615079	NR	4.40×10^{-8}	13	
58	4q32.3	rs2611215	chr4	166574267	NR	1.80×10^{-11}	13	S
59	5p13.2	rs11954020	chr5	35883251	*IL7R*	4.40×10^{-8}	13	

Ref.: Publication reference. ID: unique T1D loci identifiers. BP: NCBI36 chromosome positions.

^ *rs363609* is in LD with reported SNP *rs6546909* (r^2 = 0.84) whose p-value is stated.

* P-values are derived from Barrett *et al*. (from Table 2 in (2)). 'NR' denotes no gene of interest was reported at the locus.

II, III, IV, S corresponds to Tables where SNPs are featured; II – Non-synonymous LD SNPs, III – *cis*- interacting genes, or IV – *trans*- interacting genes, S – Supplementary *cis*- interacting genes (see Suppl. Table II(i)).

J Immunol. Author manuscript; available in PMC 2017 April 01.

Table II

Non-synonymous SNPs found in LD ($r^2 > 0.8$) with 15 T1D loci.

ID	T1D SNPs	Gene	NS variants in LD ($r^2 > 0.8$)	P- value of T1D SNP	P-value of best NS SNP
1	rs2269241	*PGM1*	rs11208257	3.6×10^{-7}^	1.6×10^{-6}^
2	rs2476601	*PTPN22*	rs2476601	1.35×10^{-27}	1.35×10^{-27}
7	rs6546909	*MOGS* *MRPL53* *TTC31* *LBX2*	rs2268416, rs1063588 rs1047911 rs6707475 rs17009998	6.3×10^{-3}^	8.7×10^{-3}^
10	rs1990760	*IFIH1*	rs1990760	7.43×10^{-5}	7.43×10^{-5}
11	rs231727	*CTLA4*	rs231775	4.28×10^{-5}	4.28×10^{-5}
16	rs6897932	*IL7R*	rs6897932	0.002	0.002
32	rs10877012	*TSFM*	rs1599932	0.44	n/a
33	rs3184504	*SH2B3*	rs3184504	2.11×10^{-9}	2.11×10^{-9}
40	rs3825932	*CTSH*	rs1036938	0.19	0.13
41	rs416603	*TNP2*	rs11640138	0.01	0.02
42	rs4788084, rs9924471	*APOBR* ***SULT1A2*** *SH2B1*	rs180743 rs1059491, rs10797300 rs7498665	0.0006, 0.44	6.44×10^{-5}
45	rs2290400	*ZPBP2* ***GSDMB***	rs11557467 rs2305480, rs2305479	0.007	0.006
48	rs763361	*CD226*	rs763361	8.07×10^{-5}	8.07×10^{-5}
49	rs2304256	*TYK2*	rs2304256	0.02	0.02
52	rs2281808	*SIRPG*	rs6043409	0.0009	0.001
55	rs229541	*C1QTNF6*	rs229527	0.05	0.04

ID: T1D loci identifiers as in Table I. Genes *SULT1A2* and *GSDMB*, the SNPs cause deleterious changes. The best NS SNP is underlined where possible.

^Association p-values are derived from Barrett *et al.* (2).

n/a – SNP genotypes not available.

J Immunol. Author manuscript; available in PMC 2017 April 01.

Table III

Cis-genes associated with 24 T1D SNPs with minimum False Discovery Rate, FDR P < 0.001.

ID	T1D SNPs (Effect allele)	Gene	EBV-B Basal	EBV-B PMA	CD4+	CD8+
6	rs2165738 (C)	*ADCY3*	↓**	↓***	↓****	ns
7	rs363609 (C)	*INO80B*	ns	↓****	ns	ns
9	rs6543134 (C)	*IL18R1*	↓***	↓****	ns	ns
12	rs3731865 (C)	*SLC11A1*	ns	ns	↑***	ns
17	rs56297233 (T)	*LYRM2*	ns	↑***	ns	ns
18	rs9388489 (C)	*C6orf173*	ns	ns	↓**	↓***
19	rs10499194 (C)	*IFNGR1*	↑****	ns	ns	ns
21	rs7804356 (T)	*SKAP2*	↓***	↓***	↓***	↓**
22	rs10272724 (T)	*IKZF1*	↑****	↑*	↑****	↑**
27	rs10509540 (T)	*C10orf59*	ns	ns	↑***	ns
29	rs3764021 (T)	*CLEC2D*	↑****	↑****	ns	ns
	rs10466829 (T)	*CLECL1*	ns	ns	↓****	↓****
30	rs705704 (T)	*SUOX*	↓**	ns	↓****	↓*
	rs2292239 (T)	*ERBB3*	ns	ns	ns	↓****
32	rs10877012 (C)	*FAM119B*	↓****	↓****	↓****	↓****
		TSFM	↑****	↑*	ns	ns
		XRCC6BP1	↑**	↑***	ns	ns
39	rs17574546 (C)	*RASGRP1*	↑****	↑***	ns	ns
40	rs3825932 (C)	*CTSH*	↑****	↑****	ns	ns
41	rs416603 (T)	*C16orf75*	↓****	↓****	↑****	↑**
42	rs4788084 (C)	*LOC728734*	↓****	↓****	↓****	↓****
		SPNS1	↓****	ns	ns	↓*

ID	T1D SNPs (Effect allele)	Gene	EBV-B Basal	EBV-B PMA	CD4+	CD8+
		TUFM	ns	ns	↓***	ns
45	rs2290400 (T)	*ORMDL3*	↑****	↑****	↑***	↑****
		GSDMB	↑****	↑****	↑****	↑****
		IKZF3	↓****	↓****	↓**	↓***
		ZPBP2	↓****	↓****	ns	ns
46	rs7221109 (C)	*SMARCE1*	↓****	↓****	↓****	↓****
48	rs763361 (T)	*CD226*	↓****	↓****	ns	ns
53	rs11203203 (T)	*UBASH3A*	↑***	ns	ns	ns
54	rs5753037 (T)	*MTMR3*	↓****	↓****	ns	ns

ID: T1D loci identifiers as in Table I.

Following notations are used: FDR: **** P < 0.0001, *** P < 0.001, ** P < 0.01, * P < 0.05, ns – not significant, ↓ risk (effect) allele reduces expression, ↑ risk (effect) allele increases expression (determined using beta coefficient)

Table IV

Trans-genes associated with 25 T1D SNPs with minimum FDR P < 0.001.

ID	T1D SNPs (Effect allele)	Gene	EBV-B Basal	EBV-B PMA	CD4+	CD8+
5	rs1534422 (C)	*ST6GAL1*	↓****	ns	ns	ns
10	rs1990760 (T)	*LOC643997*	↓***	↓**	ns	ns
11	rs11571291 (T)	*TMUB2*	ns	↓***	ns	ns
13	rs11711054 (C)	*GRAMD1B*	ns	ns	↑***	ns
19	**rs10499194 (C)**	*TUBB6*	↓***	↓***	ns	ns
21	**rs7804356 (T)**	*POLA2*	ns	↑***	ns	ns
		RRP15	ns	↓***	ns	ns
		SEC61G	ns	↓***	ns	ns
		SLC39A8	↑***	↑**	ns	ns
		TYMS	ns	↑***	ns	ns
23	rs7020673 (C)	*FHL3*	ns	↓***	ns	ns
		PLCB2	ns	↑***	ns	ns
24	rs12251307 (C)	*DERA*	ns	↑***	ns	ns
25	rs947474 (C)	*MEIS2*	↑***	↑*	ns	ns
28	rs3842727 (T)	*CD276*	ns	↓***	ns	ns
		ID2	↑*	↑****	ns	ns
30	**rs705704 (T)**	*IP6K2*	↑****	↑****	↑****	↑****
		LOC389386	ns	ns	↑****	↑****
		LOC728873	↑****	ns	ns	ns
		LOC92659	↑*	↑*	↑****	↑****
		MIR130A	↑****	↑****	↑****	↑****
		MIR1471	↑*	↑*	↑***	↑****

J Immunol. Author manuscript; available in PMC 2017 April 01.

ID	T1D SNPs (Effect allele)	Gene	EBV-B Basal	EBV-B PMA	CD4+	CD8+
33	rs1265565 (T)	*ZMYM5*	ns	ns	↓***	ns
34	rs17696736 (C)	*IRF8*	↑**	↑***	ns	ns
		NCOA7	↑*	↑****	ns	ns
35	rs9585056 (C)	*AHCTF1*	ns	↓***	ns	ns
39	rs12908309 (C)	*FAHD1*	↓*	↓****	ns	ns
		RNF13	↓***	ns	ns	ns
40	**rs3825932 (C)**	*PCK2*	ns	↑***	ns	ns
41	**rs416603 (T)**	*SORL1*	ns	↑***	ns	ns
45	**rs2290400 (T)**	*TEX9*	↓*	↓****	ns	ns
46	**rs7221109 (C)**	*USP14*	ns	↑***	ns	ns
47	rs1893217 (C)	*MCM3AP*	ns	↓***	ns	ns
48	**rs763361 (T)**	*P2RY11*	↓*	↓****	ns	ns
49	rs2304256 (C)	*ZNF280D*	↑***	ns	ns	ns
50	rs425105 (T)	*CCL5*	↑**	↑***	ns	ns
51	rs679574 (C)	*EIF5A*	↓*	↑***	ns	ns
53	rs876498 (T)	*CAT*	ns	↓***	ns	ns

ID: T1D loci identifiers as in Table I.

Following notations are used: FDR: **** P < 0.0001, *** P < 0.001, ** P < 0.01, * P < 0.05, ns – not significant, ↓ risk allele reduces expression, ↑ risk allele increases expression. T1D SNPs highlighted also showed strong *cis*-regulatory effects at FDR P < 0.001.

Table V

Enrichment terms shared between *cis* genes and *trans* genes in association with three T1D SNPs.

ID	T1D SNP	Genes	Enrichment term
19	rs10499194	*IFNGR1 (cis), TUBB6*	Cytoskeletal part (GO:0044430)
21	rs7804356	*SKAP2 (cis), POLA2, RRP15, SLC39A8, TYMS*	Phosphoprotein (PIR Keywords)
48	rs763361	*CD226 (cis), PRRY11*	Receptor (PIR Keywords)

ID: T1D loci identifiers as in Table I.

Table VI

Network and Pathway analysis of the list of candidate genes identified in Tables II–IV using GATHER and GENEMANIA (Unadjusted $P < 0.001$).

Annotation Type	Annotation	P-value
GATHER Gene Annotation and Pathway Analysis		
Chromosome	2p13 *(TTC31, MOGS, INO80B, LBX2, MRPL53)*	< 0.0001
Transcription factor binding sites	c-Ets-2 binding sites	0.0009
KEGG Pathways	path:hsa04060: Cytokine-cytokine receptor interaction	0.0005
	path:hsa00920: Sulfur metabolism	0.0007
Gene Ontology	GO:0009607 [4]: response to biotic stimulus	0.0003
	GO:0006952 [5]: defense response	0.0004
	GO:0006955 [4]: immune response	0.0004
	GO:0009613 [5]: response to pest, pathogen or parasite	0.0009
	GO:0043207 [5]: response to external biotic stimulus	0.0009
GENEMANIA Network analysis		
Functions	regulation of leukocyte activation	5.4×10^{-5}
	regulation of lymphocyte activation	6.8×10^{-5}
	regulation of cell activation	
	T cell activation	0.0002
	mononuclear cell proliferation	0.0006
	positive regulation of cell activation	
	lymphocyte proliferation	
	positive regulation of leukocyte activation	
	regulation of T cell activation	
	leukocyte proliferation	0.0008
	regulation of lymphocyte proliferation	
	regulation of mononuclear cell proliferation	

Table VII

Significant enrichment terms found using DAVID bioinformatics resource.

ID	T1D SNP	Enrichment Type	Enriched Term	FDR *P*
2	rs2476601	GOterm BP	Antigen processing and presentation of peptide antigen via MHC class I	0.017
10	rs1990760	PIR Keywords	*Acetylation*	0.00065
15	rs17388568	PIR Keywords	Phosphoprotein	0.025
22	rs10272724	PIR Keywords	Cytoplasm	0.032
		PIR Keywords	Phosphoprotein	0.047
29	rs10466829	PIR Keywords	*Lectin*	0.002
		GOterm MF	Sugar binding	0.0096
		PIR Keywords	Signal-anchor	0.041
		GOterm MF	Carbohydrate binding	0.045
33	rs3184504	Interpro	*HIN-200/IF120x*	0.0031
34	rs17696736	GOterm BP	*Response to virus*	0.00049
		PIR Keywords	Antiviral defence	0.0058
41	rs416603	Biocarta	*IL-10 Anti-inflammatory Signalling Pathway*	0.0021
		KEGG Pathway	Intestinal immune network for IgA production	0.034
42	rs9924471	PIR Keywords	Nucleus	0.0069
51	rs679574	Smart	*IGc1*	0.0031
		KEGG Pathway	Antigen processing and presentation	0.0088
		KEGG Pathway	Viral myocarditis	0.011
		Interpro	Immunoglobulin/major histocompatibility complex, conserved site	0.017
		KEGG Pathway	Graft-versus-host disease	0.021
		KEGG Pathway	Allograft rejection	0.024

ID: T1D loci identifiers as in Table I.

PIR: Protein Information Resource, GO: Gene Ontology, BP: Biological process, MF: Molecular Function