The Journal of Immunology

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

J Immunol 2016; 196:3043-3053; Prepublished online 24 February 2016;
doi: 10.4049/jimmunol.1502056
http://www.jimmunol.org/content/196/7/3043

The Journal of Immunology is published twice each month by
The American Association of Immunologists, Inc.,
1451 Rockville Pike, Suite 650, Rockville, MD 20852

Print ISSN: 0022-1767 Online ISSN: 1550-6606.

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[*,†]

Genome-wide association studies have found >60 loci that confer genetic susceptibility to type 1 diabetes (T1D). Many of these are defined only by anonymous single nucleotide polymorphisms: the underlying causative genes, as well as 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. In this study, we performed systematic analyses to characterize these variants. First, all known genes in strong linkage disequilibrium ($r^2 > 0.8$) with the reported single nucleotide polymorphisms for each locus were tested for commonly occurring nonsynonymous variations. We found only a total of 22 candidate genes at 16 T1D loci with common nonsynonymous 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 6 h PMA stimulated) and purified $CD4^+$ and $CD8^+$ T cells. We mapped *cis*-acting expression quantitative trait loci 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 that most of the T1D risk alleles mediate their effect by influencing expression of multiple nearby genes. *The Journal of Immunology*, **2016, 196: 3043–3053.**

Type 1 diabetes (T1D) affects ∼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 >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 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 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 noncoding 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 nonsynonymous (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, that is, *trans* regulation (11). The mechanisms involved in *trans* regulation could include indirect genetic effects, for example, by means of variation in encoded proteins such as transcription factors, or by other effects, such as steric hindrance (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.

*Centre for Diabetes Research, Harry Perkins Institute of Medical Research, Nedlands, Western Australia 6009, Australia; †Centre of Medical Research, University of Western Australia, Nedlands, Western Australia 6009, Australia; ‡Lee Kong Chian School of Medicine, Nanyang Technological University, Singapore 636921; §Ulm University Medical Centre, Department of Internal Medicine I, Ulm University, 89081 Ulm, Germany; ¶Department of Pediatrics, Herlev and Gentofte Hospital, 2730 Herlev, Denmark; ‖Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL 32610; and #Genetics Institute, University of Florida, Gainesville, FL 32610

ORCIDs: 0000-0002-4827-4778 (R.R.); 0000-0003-1955-6976 (M.M.); 0000-0003-2177-0488 (I.L.); 0000-0003-3274-5448 (F.P.); 0000-0002-8562-7325 (G.M.).

Received for publication September 24, 2015. Accepted for publication January 25, 2016.

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

The data presented in this article have been submitted to the National Center for Biotechnology Information Gene Expression Omnibus (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77350) under accession number GSE77350.

Address correspondence and reprint requests to 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. E-mail address: grant.morahan@uwa.edu.au

The online version of this article contains supplemental material.

Abbreviations used in this article: Chr, chromosome; EBV-B, EBV-transformed B; eQTL, expression qualitative trait locus; FDR, false discovery rate; GO, Gene Ontology; LD, linkage disequilibrium; nsSNP, nonsynonymous SNP; PC, principal component; PCA, principal component analysis; SNP, single nucleotide polymorphism; T1D, type 1 diabetes; T1DGC, Type 1 Diabetes Genetics Consortium.

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 T1DGC study has been described elsewhere, including phenotypic and extensive genetic characterization of >4000 affected sibling-pair families (3). Upon joining the T1DGC, family members provided blood samples. PBMCs were isolated and aliquots were used to provide DNA samples, to derive EBV-transformed B (EBV-B) lymphoblastoid cell lines (26, 27), and they were frozen for later use. EBV-B cells from 202 European subjects from the T1DGC family collection were examined in the present study. These samples consisted of 46 unaffected subjects and the rest were T1D cases. EBV-B cells were either unstimulated or treated with PMA (28) for 6 h (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 Supplemental Table Ii.

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 a NanoDrop 1000 spectrophotometer (Thermo Scientific), and RNA quality was checked on an Agilent 2100 Bioanalyzer (Agilent Technologies). Samples with an RNA integrity number of ≥8 were biotin labeled using an Illumina TotalPrep RNA amplification kit (Ambion) as per the manufacturer's instructions. The biotin-labeled samples were hybridized onto Illumina HumanHT-12 v4.0 expression beadchips and beadchips were scanned by a BeadArray Reader (Illumina) following the manufacturer's instructions. Raw data were 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, and robust spline normalization using the 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 principal component (PC) analysis (PCA) plots, batch effects were evident. For correction within each cell type, normalized expression data for each gene were centered by batch and centered again after merging batches. The batch correction was validated by PCA (Supplemental Fig. 1A, 1B), and pair plots of PCs 1–4 did not reveal any further batch effects. BLASTN software was used to identify probesets with unique sequences.

The data generated in this study have been deposited in the National Center for Biotechnology Information Gene Expression Omnibus (series accession no. GSE77350) and are available at http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE77350.

To assess association between SNP genotype and gene expression, the Matrix eQTL 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 the naiveRandRUV method, with parameter k set to 20. After correction, the same SNP was tested against the corrected set and p value association of the SNP–gene pair was recorded. This procedure was repeated for all SNPs and finally Benjamini false discovery rate (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 (200,000). The PCs that showed no or weak genome association (i.e., minimum SNP-PC association FDR $p > 0.001$) were chosen as unassociated 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 Supplemental Fig. 1E and 1F, the four gene expression datasets were corrected as follows: seven PCs (1–6 and 8) were removed from EBV-B basal cell line samples, three PCs (1, 4, and 9) were removed for PMA-stimulated EBV-B cell line samples, four PCs (1–4) were removed for $CD4^+$ samples, and two 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 (Supplemental Table Iii). The RUV-2 method of correction yielded more significant results than did PCA methods.

Statistical analysis

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

Enrichment analysis

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

Results

Systematic evaluation of nonsynonymous SNPs in genes in T1D-associated regions

First, we searched for commonly occurring nonsynonymous SNPs (nsSNPs) in 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. Additionally, none of the four SNPs recently discovered in Onengut-Gumuscu et al. (13) was 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 Supplemental Table Iiii).

Next, we searched whether any nsSNPs showed better association with T1D than did the reported SNP itself. For this, we performed a transmission disequilibrium test (sibship) using UNPHASED (35, 36) on a dataset of 2676 nuclear families with unaffected parents and two or more affected siblings. Results are presented in Table II. Association p values for SNPs not included in the Immunochip genotyping were derived from Barrett et al. (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 $\Delta p = p_{\text{ns SNP}}/p_{\text{reported SNP}}$). Two other ns SNPs (*rs2305480* and *rs229527*) also showed very small ($\Delta p > 0.1$) improvement in association compared with the reported T1D SNP. Most of the T1D loci did not have associated nsSNPs in nearby genes.

Gene expression analyses

EBV-B cell lines 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

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	Chr 1	63881359	*PGM1*	4.00×10^{-7}	(2)	II, S
2	1p13.2	rs2476601	Chr 1	114179091	*PTPN22*	8.50×10^{-85}	(2)	II
3	1q31.2	rs2816316	Chr 1	190803436	*RGS1*	3.10×10^{-5a}	(2)	
4	1q32.1	rs3024493	Chr 1	205010591	*IL10*	1.90×10^{-9}	(2)	
5	2p24.3	rs1534422	Chr 2	12558192	*Intergenic*	2.00×10^{-6}	(2)	IV
6	2p23.3	rs2165738	Chr 2	24546313	*Intergenic*	4.00×10^{-6}	(14)	III, S
7	2p13.1	rs363609^	Chr 2	74756380	*DQX1*	8.53×10^{-6b}	(15)	II, III, S
8	2q11.2	rs9653442	Chr 2	100191799	*AFF3*	5.00×10^{-6}	(15)	S
9	2q12.1	rs6543134	Chr 2	102416890	*IL18RAP*	8.03×10^{-5a}	(16)	III, S
10	2q24.2	rs1990760	Chr 2	162832297	*IFIH1*	6.60×10^{-9}	(2)	II, IV
	2q24.2	rs3747517	Chr 2	162837070	*IFIH1*	4.70×10^{-7}	(2)	
11	2q33.2	rs11571291	Chr 2	204429377	*CTLA4*	1.19×10^{-12}	(17)	IV
	2q33.2	rs3087243	Chr 2	204447164	*CTLA4*	1.20×10^{-15}	(2)	
	2q33.2	rs231727	Chr 2	204449795	*CTLA4*	2.13×10^{-18}	(13)	II, S
12	2q35	rs3731865	Chr 2	218958247	*SLC11A1*	1.55×10^{-6}	(18)	III
13	3p21.31	rs11711054	Chr 3	46320615	*CCR5*	1.70×10^{-5a}	(2)	IV
14	4p15.2	rs10517086	Chr 4	25694609	*Intergenic*	4.60×10^{-10}	(2)	
15	4q27	rs4505848	Chr 4	123351942	*IL2*	4.70×10^{-13}	(2)	
	4q27	rs17388568	Chr 4	123548812	*IL2*	3.00×10^{-6}	(19)	
	4q27	rs2069763	Chr 4	123596932	*IL2*	1.91×10^{-10}	(13)	
16	5p13.2	rs6897932	Chr 5	35910332	*IL7R*	8.00×10^{-6}	(15)	II
17	6q15	rs597325	Chr 6	91059215	*BACH2*	3.38×10^{-10}	(17)	S
	6q15	rs56297233	Chr 6	91070750	*BACH2*	5.40×10^{-8}	(2)	III
18	6q22.32	rs9388489	Chr 6	126740412	*C6orf173*	4.20×10^{-13}	(2)	III
19	6q23.3	rs2327832	Chr 6	138014761	*TNFAIP3*	1.60×10^{-4a}	(2)	S
	6q23.3	rs10499194	Chr 6	138044330	*TNFAIP3*	3.00×10^{-4a}	(2)	III, IV
20	6q25.3	rs1738074	Chr 6	159385965	*TAGAP*	7.59×10^{-9}	(16)	S
21	7p15.2	rs7804356	Chr 7	26858190	*Intergenic*	5.30×10^{-9}	(2)	III, IV
22	7p12.2	rs10272724	Chr 7	50444707	*IKZF1*	4.80×10^{-9}	(20)	III
23	9p24.2	rs7020673	Chr 9	4281747	*GLIS3*	5.40×10^{-12}	(2)	IV, S
	9p24.2	rs10758593	Chr 9	4282083	*GLIS3*	1.18×10^{-8}	(17)	S
24	10p15.1	rs12251307	Chr 10	6163501	*IL2RA*	1.30×10^{-13}	(2)	IV
25	10p15.1	rs947474	Chr 10	6430456	*PRKCQ*	4.00×10^{-9}	(14)	IV, S
	10p15.1	rs11258747	Chr 10	6512897	*PRKCQ*	1.20×10^{-7}	(2)	
26	10p11.22	rs722988	Chr 10	33466153	*NRP1*	4.88×10^{-8}	(21)	S
27	10q23.31	rs10509540	Chr 10	90013013	*C10orf59*	1.30×10^{-28}	(2)	III, S
28	11p15.5	rs7928968	Chr 11	2006875	*INS*	2.78×10^{-14}	(17)	
	11p15.5	rs3842727	Chr 11	2141424	*TH*	4.89×10^{-196}	(13)	IV
	11p15.5	rs7111341	Chr 11	2169742	*INS*	4.40×10^{-48}	(2)	
29	12p13.31	rs3764021	Chr 12	9724895	NR	5.00×10^{-8}	(19)	III, S
	12p13.31	rs10466829	Chr 12	9767358	*CLECL1*	9.19×10^{-9}	(17)	III, S
	12p13.31	rs4763879	Chr 12	9801431	*CD69*	1.90×10^{-11}	(2)	
30	12q13.2	rs705704	Chr 12	54721679	*ERBB3*	4.31×10^{-31}	(17)	III, IV, S
	12q13.2	rs11171739	Chr 12	54756892	*ERBB3*	1.00×10^{-11}	(19)	
	12q13.2	rs2292239	Chr 12	54768447	*ERBB3*	2.20×10^{-25}	(2)	III, S
31	12q13.3	rs3809114	Chr 12	56134906	NR	6.90×10^{-4a}	(2)	
32	12q14.1	rs10877012	Chr 12	56448352	*CYP27B1*	3.80×10^{-6}	(22)	II, III, S
33	12q24.12	rs1265565	Chr 12	110199580	*CUX2*	1.00×10^{-16}	(23)	IV, S
	12q24.12	rs3184504	Chr 12	110368991	*SH2B3*	2.80×10^{-27}	(2)	II
34	12q24.13	rs17696736	Chr 12	110971201	*C12orf30*	1.73×10^{-13}	(15)	IV, S
35	13q32.3	rs9585056	Chr 13	98879767	*GPR183*	5.20×10^{-9}	(24)	IV
36	14q24.1	rs1465788	Chr 14	68333352	*Intergenic*	1.80×10^{-12}	(2)	
37	14q32.2	rs4900384	Chr 14	97568704	*Intergenic*	3.70×10^{-9}	(2)	
38	14q32.2	rs56994090	Chr 14	100376200	*DLK1*	1.62×10^{-10}	(25)	
39	15q14	rs17574546	Chr 15	36689768	*RASGRP1*	3.35×10^{-8}	(13)	III
	15q14	rs12908309	Chr 15	36715969	*RASGRP1*	4.31×10^{-8}	(17)	IV
40	15q25.1	rs3825932	Chr 15	77022501	*CTSH*	7.70×10^{-8}	(2)	II–IV
41	16p13.13	rs12708716	Chr 16	11087374	*CLEC16A*	2.20×10^{-16}	(2)	
	16p13.13	rs12927355	Chr 16	11102272	*DEXI*	1.91×10^{-16}	(17)	
	16p13.13	rs416603	Chr 16	11271580	*C16orf75*	3.00×10^{-6}	(14)	II–IV
42	16p11.2	rs4788084	Chr 16	28447349	*IL27*	2.60×10^{-13}	(2)	II, III, S
	16p11.2	rs9924471	Chr 16	28499031	*IL27*	1.21×10^{-11}	(17)	II
43	16q23.1	rs7202877	Chr 16	73804746	*Intergenic*	3.10×10^{-15}	(2)	
	16q23.1	rs8056814	Chr 16	73809828	NR	1.13×10^{-7}	(17)	
44	17p13.1	rs16956936	Chr 17	7574417	*Intergenic*	5.00×10^{-7}	(2)	
45	17q12	rs2290400	Chr 17	35319766	*ORMDL3*	5.50×10^{-13}	(2)	II–IV, S
46	17q21.2	rs7221109	Chr 17	36023812	*Intergenic*	1.30×10^{-9}	(2)	III, IV, S
47	18p11.21	rs1893217	Chr 18	12799340	*PTPN2*	3.60×10^{-15}	(2)	IV, S
48	18q22.2	rs763361	Chr 18	65682622	*CD226*	1.56×10^{-8}	(16)	II, III, IV
49	19p13.2	rs2304256	Chr 19	10336652	*TYK2*	4.13×10^{-9}	(25)	II, IV, S
50	19q13.32	rs425105	Chr 19	51900321	*Intergenic*	2.70×10^{-11}	(2)	IV
51	19q13.33	rs679574	Chr 19	53897920	*FUT2*	4.30×10^{-18}	(16)	IV, S

(*Table continues*)

Table I. (*Continued*)

ID	Locus	T1D SNP	Chr	BP	Gene	*p* Value	Ref.	Tables
52	20p13	rs2281808	Chr 20	1558551	*Intergenic*	1.20×10^{-11}	(2)	II
53	21q22.3	rs11203203	Chr 21	42709255	*UBASH3A*	1.70×10^{-9}	(2)	III, S
	21q22.3	rs876498	Chr 21	42714896	*UBASH3A*	7.06×10^{-9}	(13)	IV
54	22q12.2	rs5753037	Chr 22	28911722	*Intergenic*	2.60×10^{-16}	(2)	III, S
55	22q12.3	rs229541	Chr 22	35921264	*C1QTNF6*	2.10×10^{-7}	(2)	II, S
				Four Newly Discovered Loci				
56	1q32.1	rs6691977	Chr 1	200814959	NR	4.30×10^{-8}	(13)	
57	2q13	rs4849135	Chr 2	111615079	NR	4.40×10^{-8}	(13)	
58	4q32.3	rs2611215	Chr 4	166574267	NR	1.80×10^{-11}	(13)	S
59	5p13.2	rs11954020	Chr 5	35883251	*IL7R*	4.40×10^{-8}	(13)	

In the Tables column, II, III, IV, and S correspond to tables where SNPs are featured; that is, II, nonsynonymous LD SNPs; III, *cis*- interacting genes; IV, *trans*-interacting genes; S, supplementary *cis*-interacting genes (see Supplemental Table IIi).

[a]The *p* value is derived from table II in Barrett et al. (2).

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

BP, NCBI36 chromosome positions; ID, unique T1D loci identifier; NR, no gene of interest was reported at the locus; Ref., publication reference.

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 EBV-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 Supplemental Table Ii. As expected, there were no significant differences in the gene expression between cases and unaffected subjects or between cases and unaffected first-degree relatives (Supplemental Fig. 1C, 1D), 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 *Materials and 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 (Supplemental Fig. 1A, 1B). 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 h 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 (Fig. 1). Adjusted $p < 0.0001$ was selected as a cut-off for differential expression. A total of 1465 genes were differentially expressed at this threshold with at least a modest fold change (absolute $\log_2$ fold change > 0.3). Genes with the highest fold

Table II. nsSNPs found in LD ($r^2 > 0.8$) with 16 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^{-7a}	1.6×10^{-6a}
2	rs2476601	*PTPN22*	rs2476601	1.35×10^{-27}	1.35×10^{-27}
7	rs6546909	*MOGS*	rs2268416, rs1063588	6.3×10^{-3a}	8.7×10^{-3a}
		MRPL53	rs1047911		
		TTC31	rs6707475		
		LBX2	rs17009998		
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*	rs180743	0.0006, 0.44	6.44×10^{-5}
		SULT1A2	rs1059491, rs10797300		
		SH2B1	rs7498665		
45	rs2290400	*ZPBP2*	rs11557467	0.007	0.006
		GSDMB	rs2305480, rs2305479		
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

For genes *SULT1A2* and *GSDMB* (in bold type), the SNPs cause deleterious changes. The best NS SNP is underlined where possible.

[a]Association *p* values are derived from Barrett et al. (2).

ID, T1D loci identifiers as in Table I; n/a, SNP genotypes not available.

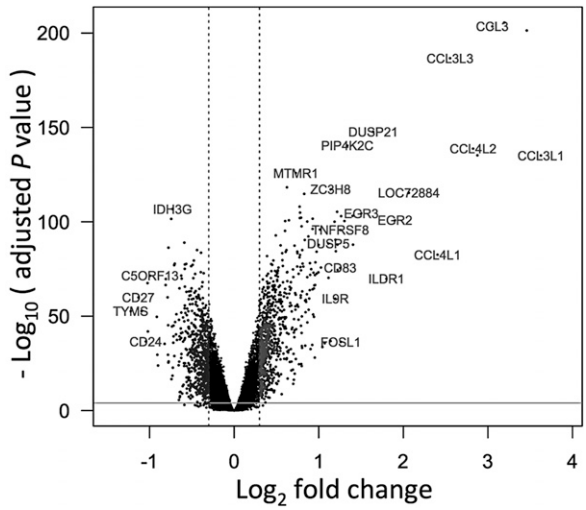

FIGURE 1. Comparison of gene expression in EBV-B cells between basal and 6 h PMA-stimulated samples. Differences ($\log_2$ fold change) in gene expression are shown on the *x*-axis; *y*-axis shows $-\log_{10}$ (adjusted *p* values).

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 6 h 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, 3672 *cis* interactions pairs were tested, ~2.9 million *trans* interactions pairs were tested, and FDR *p* value corrections were applied separately for *cis* and *trans* eQTLs. The Matrix eQTL R package (30) was used to perform these eQTL tests. Owing to unknown confounding factors that could limit the power of detecting significantly differentially expressed genes, we performed two methods of correction independently: 1) removing unwanted variation (RUV-2) (31, 32), and 2) adding genome-wide unassociated expression–derived PCs as covariates (described in *Materials and 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 analyses were performed using GATHER (39) and GENEMANIA (40), respectively. The results from these analyses are summarized in Tables III–IV and are described below. Box plots 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 Fig. 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

Table III. *cis* Genes associated with 24 T1D SNPs with minimum 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	↓ *
		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

For FDR, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.
↓, risk (effect) allele reduces expression; ↑, risk (effect) allele increases expression (determined using β coefficient); ID, T1D loci identifiers as in Table I.

associated with differences in expression of 20 genes (Table III). Using lower thresholds of adjusted p values (<0.05), an additional 13 T1D SNPs affected the expression of a further 20 genes (Supplemental Table IIi). 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 downregulated 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 $p = 0.008$) (Supplemental Table IIi).

Next, we tested 6 h 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. Additionally, 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 with basal condition (refer to Ref. 42). The rest of these results are presented in Supplemental Table IIi.

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, whereas the remaining 9 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 that 16 T1D SNP genotypes regulated the expression of 20 genes significantly. Of these genes, 11 (*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 that 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 with EBV-B cells. Additionally, there were nine newly identified candidate genes associated with nine T1D SNPs. Five of these SNPs

Table IV. Transgenes 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	↑ *	↑ *	↑ ***	↑ ****
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

T1D SNPs in bold also showed strong *cis* regulatory effects at FDR $p < 0.001$.
For FDR, *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.0001$.
↓, risk (effect) allele reduces expression; ↑, risk (effect) allele increases expression (determined using β coefficient). ID, T1D loci identifiers as in Table I.

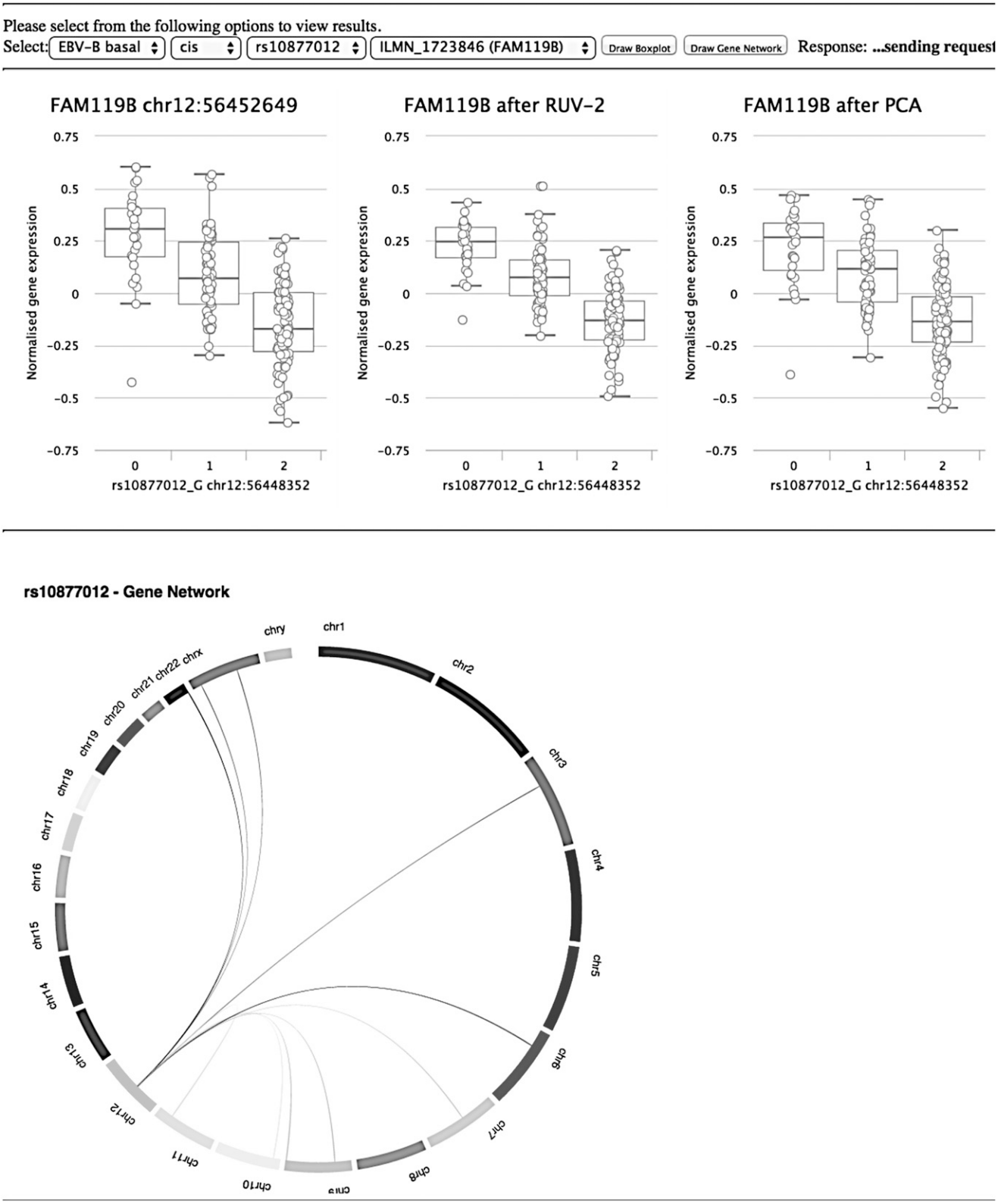

FIGURE 2. Screen capture of the Web interface for browsing box plots and gene networks (available at: http://www.sysgen.org/T1DGCSysGen/).

had shown *cis* effects in the EBV-B cells but had affected a different set of genes. Among these nine 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 Supplemental Table IIi. 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 that regulated the expression of 19 genes across all samples tested. Excepting *ADCY3*, 10 candidate genes were found *cis* regulated in EBV-B cells, $CD4^+$ 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 not differentially

regulated in EBV-B cells or 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 Supplemental Table IIi.

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 an even less stringent suggestive threshold of nominal unadjusted $p < 0.001$ 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 ~2.9 million 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 coregulation 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 nsSNPs defined in Table II. The *trans* regulatory effects detected at lower threshold levels are presented in Supplemental Tables IIii and 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 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 (http://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 GO 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* because 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 seven *trans* genes (*EIF2AK2*, *IFI16*, *IFNGR1*, *MX1*, *MX2*, *PLSCR1*, and *STAT1*). Furthermore, genes *MX1* and *MX2* are also known inflammatory and immune response genes. Additionally, 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*, *HLA-G*, *HLA-H*, and *DRB4*), which gave positive enrichment for terms such as "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 $p < 0.05$. Because not all T1D SNPs may be present in this browser, we allowed a 100 Kb window for the search of the expression SNP. Two *trans* genes were validated: *UBE2L6* (EBV-B with/without PMA) associated with *rs3184504* and *STAT1* (EBV-B basal) with *rs17696736*. Second, 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 nsSNPs in strong LD with the T1D SNP. Thirty-six

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; PIR, Protein Information Resource.

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
GO	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	

Numbers in brackets indicate GO annotation depth.

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 ($p = 1.817 \times 10^{-11}$) whereas *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, that is, 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

Table VII. Significant enrichment terms found using the DAVID bioinformatics resource

ID	T1D SNP	Enrichment Type	Enriched Term	FDR *p* Value
2	rs2476601	GO term BP	Ag processing and presentation of peptide Ag 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
		GO term MF	Sugar binding	0.0096
		PIR keywords	Signal anchor	0.041
		GO term MF	Carbohydrate binding	0.045
33	rs3184504	Interpro	*HIN-200/IF120x*	0.0031
34	rs17696736	GO term BP	*Response to virus*	0.00049
		PIR keywords	Antiviral defense	0.0058
41	rs416603	Biocarta	*IL-10 anti-inflammatory signaling 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	Ag processing and presentation	0.0088
		KEGG pathway	Viral myocarditis	0.011
		Interpro	Ig/MHC, conserved site	0.017
		KEGG pathway	Graft-versus-host disease	0.021
		KEGG pathway	Allograft rejection	0.024

Italics indicate FDR $p < 0.005$.
BP, biological process; ID, T1D loci identifiers as in Table I; MF, molecular function; PIR, Protein Information Resource.

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 $p = 0.04$) (Supplemental Table IIi). 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 (12 in *cis*, 4 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* (chromosome [Chr] 16p11.2) where the risk allele decreased expression in all four cell types. The effect directions 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 a 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-hybridization problems described in the previous section, these probes were excluded from further analysis. nsSNPs 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* (nonsynonymous) 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 (for review, see Ref. 46). 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 defense due to *trans* genes that are involved in proinflamatory response (such as *MX1* and *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 that *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*-regulated genes, and *trans*-regulated genes we identified are important candidates for further investigation. So that other researchers can extend the work reported in the present study, we have implemented a Web interface (42) allowing users to browse box plots for the eQTL interactions reported.

Acknowledgments

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

Disclosures

The authors have no financial conflicts of interest.

References

1. Danaei, G., M. M. Finucane, Y. Lu, G. M. Singh, M. J. Cowan, C. J. Paciorek, J. K. Lin, F. Farzadfar, Y. H. Khang, G. A. Stevens, et al; Global Burden of Metabolic Risk Factors of Chronic Diseases Collaborating Group (Blood Glucose). 2011. 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* 378: 31–40.
2. Barrett, J. C., D. G. Clayton, P. Concannon, B. Akolkar, J. D. Cooper, H. A. Erlich, C. Julier, G. Morahan, J. Nerup, C. Nierras, et al. 2009. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat. Genet.* 41: 703–707.
3. Rich, S. S., B. Akolkar, P. Concannon, H. Erlich, J. E. Hilner, C. Julier, G. Morahan, J. Nerup, C. Nierras, F. Pociot, and J. A. Todd. 2009. Overview of the Type I Diabetes Genetics Consortium. *Genes Immun.* 10(Suppl. 1): S1–S4.
4. Morahan, G., M. Mehta, I. James, W.-M. Chen, B. Akolkar, H. A. Erlich, J. E. Hilner, C. Julier, J. Nerup, C. Nierras, et al. 2011. Tests for genetic interactions in type 1 diabetes: linkage and stratification analyses of 4,422 affected sib-pairs. *Diabetes* 60: 1030–1040.
5. Morahan, G. 2012. Insights into type 1 diabetes provided by genetic analyses. *Curr. Opin. Endocrinol. Diabetes Obes.* 19: 263–270.
6. 2010. On beyond GWAS. *Nat. Genet.* 42: 551.
7. Freedman, M. L., A. N. Monteiro, S. A. Gayther, G. A. Coetzee, A. Risch, C. Plass, G. Casey, M. De Biasi, C. Carlson, D. Duggan, et al. 2011. Principles for the post-GWAS functional characterization of cancer risk loci. *Nat. Genet.* 43: 513–518.
8. Sacks, D. B., M. Arnold, G. L. Bakris, D. E. Bruns, A. R. Horvath, M. S. Kirkman, A. Lernmark, B. E. Metzger, and D. M. Nathan, National Academy of Clinical Biochemistry, Evidence-Based Laboratory Medicine Committee of the American Association for Clinical Chemistry. 2011. Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus. *Diabetes Care* 34: e61–e99.
9. Morahan, G., and M. Varney. 2010. The genetics of type 1 diabetes. In *The HLA Complex in Biology and Medicine: A Resource Book*, 1st Ed. N. K. Mehra, G. Kaur, J. McCluskey, F. T. Christiansen, and F. H. Claas, eds. Jaypee Brothers Medical Publishers, New Delhi, India, p. 205–218.
10. Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov, and S. R. Sunyaev. 2010. A method and server for predicting damaging missense mutations. *Nat. Methods* 7: 248–249.
11. Veyrieras, J. B., S. Kudaravalli, S. Y. Kim, E. T. Dermitzakis, Y. Gilad, M. Stephens, and J. K. Pritchard. 2008. High-resolution mapping of expression-QTLs yields insight into human gene regulation. *PLoS Genet.* 4: e1000214.
12. Morahan, G., and R.W. Williams. 2007. Systems genetics: the next generation in genetics research? *Novartis Found. Symp.* 281: 181–188.
13. Onengut-Gumuscu, S., W.-M. Chen, O. Burren, N. J. Cooper, A. R. Quinlan, J. C. Mychaleckyj, E. Farber, J. K. Bonnie, M. Szpak, E. Schofield, et al. 2015. Fine mapping of type 1 diabetes susceptibility loci and evidence for colocalization of causal variants with lymphoid gene enhancers. *Nat. Genet.* 47: 381–386.
14. Cooper, J. D., D. J. Smyth, A. M. Smiles, V. Plagnol, N. M. Walker, J. E. Allen, K. Downes, J. C. Barrett, B. C. Healy, J. C. Mychaleckyj, et al. 2008. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. *Nat. Genet.* 40: 1399–1401.
15. Todd, J. A., N. M. Walker, J. D. Cooper, D. J. Smyth, K. Downes, V. Plagnol, R. Bailey, S. Nejentsev, S. F. Field, F. Payne, et al. 2007. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.* 39: 857–864.
16. Smyth, D. J., V. Plagnol, N. M. Walker, J. D. Cooper, K. Downes, J. H. Yang, J. M. Howson, H. Stevens, R. McManus, C. Wijmenga, et al. 2008. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N. Engl. J. Med.* 359: 2767–2777.
17. Bradfield, J. P., H. Q. Qu, K. Wang, H. Zhang, P. M. Sleiman, C. E. Kim, F. D. Mentch, H. Qiu, J. T. Glessner, K. A. Thomas, et al. 2011. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. *PLoS Genet.* 7: e1002293.
18. Yang, J. H., K. Downes, J. M. Howson, S. Nutland, H. E. Stevens, N. M. Walker, and J. A. Todd. 2011. Evidence of association with type 1 diabetes in the *SLC11A1* gene region. *BMC Med. Genet.* 12: 59.
19. Wellcome Trust Case Control Consortium. 2007. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447: 661–678.
20. Swafford, A. D., J. M. Howson, L. J. Davison, C. Wallace, D. J. Smyth, H. Schuilenburg, M. Maisuria-Armer, T. Mistry, M. J. Lenardo, and J. A. Todd. 2011. An allele of *IKZF1* (Ikaros) conferring susceptibility to childhood acute lymphoblastic leukemia protects against type 1 diabetes. *Diabetes* 60: 1041–1044.

21. Evangelou, M., D. J. Smyth, M. D. Fortune, O. S. Burren, N. M. Walker, H. Guo, S. Onengut-Gumuscu, W. M. Chen, P. Concannon, S. S. Rich, et al. 2014. A method for gene-based pathway analysis using genomewide association study summary statistics reveals nine new type 1 diabetes associations. *Genet. Epidemiol.* 38: 661–670.
22. Bailey, R., J. D. Cooper, L. Zeitels, D. J. Smyth, J. H. Yang, N. M. Walker, E. Hyppönen, D. B. Dunger, E. Ramos-Lopez, K. Badenhoop, et al. 2007. Association of the vitamin D metabolism gene CYP27B1 with type 1 diabetes. *Diabetes* 56: 2616–2621.
23. Huang, J., D. Ellinghaus, A. Franke, B. Howie, and Y. Li. 2012. 1000 Genomes-based imputation identifies novel and refined associations for the Wellcome Trust Case Control Consortium phase 1 data. *Eur. J. Hum. Genet.* 20: 801–805.
24. Heinig, M., E. Petretto, C. Wallace, L. Bottolo, M. Rotival, H. Lu, Y. Li, R. Sarwar, S. R. Langley, A. Bauerfeind, et al. 2010. A *trans*-acting locus regulates an anti-viral expression network and type 1 diabetes risk. *Nature* 467: 460–464.
25. Wallace, C., D. J. Smyth, M. Maisuria-Armer, N. M. Walker, J. A. Todd, and D. G. Clayton. 2010. The imprinted DLK1-MEG3 gene region on chromosome 14q32.2 alters susceptibility to type 1 diabetes. *Nat. Genet.* 42: 68–71.
26. Burster, T., and B. O. Boehm. 2010. 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.* 26: 227–238.
27. Rosinger, S., S. Nutland, E. Mickelson, M. D. Varney, B. O. Boehm, G. J. Olsem, J. A. Hansen, I. Nicholson, J. E. Hilner, L. H. Perdue, et al; T1DGC. 2010. 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* 7(1, Suppl.): S65–S74.
28. Nomura, N., M. Nomura, K. Sugiyama, and J. Hamada. 2007. Phorbol 12-myristate 13-acetate (PMA)-induced migration of glioblastoma cells is mediated via p38MAPK/Hsp27 pathway. *Biochem. Pharmacol.* 74: 690–701.
29. Du, P., W. A. Kibbe, and S. M. Lin. 2008. *lumi*: a pipeline for processing Illumina microarray. *Bioinformatics* 24: 1547–1548.
30. Shabalin, A. A. 2012. Matrix eQTL: ultra fast eQTL analysis via large matrix operations. *Bioinformatics* 28: 1353–1358.
31. Risso, D., J. Ngai, T. P. Speed, and S. Dudoit. 2014. Normalization of RNA-seq data using factor analysis of control genes or samples. *Nat. Biotechnol.* 32: 896–902.
32. Jacob, L., J. A. Gagnon-Bartsch, and T. P. Speed. 2016. Correcting gene expression data when neither the unwanted variation nor the factor of interest are observed. *Biostatistics* 17: 16–28.
33. Flutre, T., X. Wen, J. Pritchard, and M. Stephens. 2013. A statistical framework for joint eQTL analysis in multiple tissues. *PLoS Genet.* 9: e1003486.
34. Smyth, G. K. 2004. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* 3: 1–25.
35. Dudbridge, F. 2008. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. *Hum. Hered.* 66: 87–98.
36. Dudbridge, F., P. A. Holmans, and S. G. Wilson. 2011. A flexible model for association analysis in sibships with missing genotype data. *Ann. Hum. Genet.* 75: 428–438.
37. Huang, W., B. T. Sherman, and R. A. Lempicki. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat. Protoc.* 4: 44–57.
38. Huang, W., B. T. Sherman, and R. A. Lempicki. 2009. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 37: 1–13.
39. Chang, J. T., and J. R. Nevins. 2006. GATHER: a systems approach to interpreting genomic signatures. *Bioinformatics* 22: 2926–2933.
40. Warde-Farley, D., S. L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C. T. Lopes, et al. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Res.* 38: W214-20.
41. International HapMap Project. Available at: http://www.hapmap.org/hapmart.html.en. Accessed: November 27, 2015.
42. T1DGC Systems Genetics Project. Available at: http://www.sysgen.org/T1DGCSysGen/. Accessed: September 23, 2015.
43. Westra, H. J., M. J. Peters, T. Esko, H. Yaghootkar, C. Schurmann, J. Kettunen, M. W. Christiansen, B. P. Fairfax, K. Schramm, J. E. Powell, et al. 2013. Systematic identification of *trans* eQTLs as putative drivers of known disease associations. *Nat. Genet.* 45: 1238–1243.
44. Fairfax, B. P., S. Makino, J. Radhakrishnan, K. Plant, S. Leslie, A. Dilthey, P. Ellis, C. Langford, F. O. Vannberg, and J. C. Knight. 2012. Genetics of gene expression in primary immune cells identifies cell type-specific master regulators and roles of HLA alleles. *Nat. Genet.* 44: 502–510.
45. Fløyel, T., C. Brorsson, L. B. Nielsen, M. Miani, C. H. Bang-Berthelsen, M. Friedrichsen, A. J. Overgaard, L. A. Berchtold, A. Wiberg, P. Poulsen, et al. 2014. CTSH regulates β-cell function and disease progression in newly diagnosed type 1 diabetes patients. *Proc. Natl. Acad. Sci. USA* 111: 10305–10310.
46. Wallace, J. L., and R. Wang. 2015. Hydrogen sulfide-based therapeutics: exploiting a unique but ubiquitous gasotransmitter. *Nat. Rev. Drug Discov.* 14: 329–345.
47. Husseiny, M. I., J. Rawson, A. Kaye, I. Nair, I. Todorov, M. Hensel, F. Kandeel, and K. Ferreri. 2014. An oral vaccine for type 1 diabetes based on live attenuated *Salmonella*. *Vaccine* 32: 2300–2307.