Human Molecular Genetics, 2009, Vol. 18, No. 2 **358–365**
doi:10.1093/hmg/ddn342
Advance Access published on October 16, 2008

IFIH1 polymorphisms are significantly associated with type 1 diabetes and *IFIH1* gene expression in peripheral blood mononuclear cells

Siyang Liu[1,†], Hongjie Wang[1,†], Yulan Jin[1,†], Robert Podolsky[1,2], MV Prasad Linga Reddy[1], Jennifer Pedersen[1,2], Bruce Bode[5], John Reed[6], Dennis Steed[6], Steve Anderson[7], Ping Yang[1], Andy Muir[1,3], Leigh Steed[1], Diane Hopkins[1], Yihua Huang[1,4], Sharad Purohit[1], Cong-Yi Wang[1,4], Andrea K. Steck[8], Annalisa Montemari[1], George Eisenbarth[8], Marian Rewers[8] and Jin-Xiong She[1,4,*]

[1]Center for Biotechnology and Genomic Medicine, [2]Department of Medicine, [3]Department of Pediatrics, and [4]Department of Pathology, Medical College of Georgia, Augusta, GA 30912, USA, [5]Atlanta Diabetes Associates, Atlanta, GA, USA, [6]Southeast Endocrine Clinics, Atlanta, GA, USA, [7]Pediatric Diabetes, Atlanta, GA, USA and [8]Barbara Davis Center for Childhood Diabetes, University of Colorado Health Sciences Center, Mail Stop A140, PO Box 6511, Aurora, CO 80045-6511, USA

Received September 18, 2008; Revised September 18, 2008; Accepted October 14, 2008

Genome-wide association (GWA) studies revealed a number of single nucleotide polymorphisms (SNPs) significantly associated with type 1 diabetes (T1D). In an attempt to confirm some of these candidate associations, we genotyped 2046 Caucasian patients and 2417 normal controls from the United States for SNPs in five genomic regions. While no evidence was obtained for four genomic regions (rs2929366/*NM_144715* on chromosome 3, rs9127/*Q7Z4C4* on chromosome 5, rs1445898/*CAPSL* on chromosome 5 and rs2302188/*NM_033543* on chromosome 19), we provide strong evidence for association between T1D and multiple SNPs in the *IFIH1* linkage disequilibrium (LD) block on chromosome 2q. Among the 10 SNPs genotyped for the 2q region, four SNPs located within the *IFIH1* gene or at the 5′ region of *IFIH1* showed significant association with T1D in the Georgia population [odds ratio (OR) = 1.7–1.9] with the best *P*-value found at SNP rs1990760 ($P = 8 \times 10^{-8}$ and OR = 1.9). Several SNPs outside of the *IFIH1* gene also showed significant but weaker associations. Furthermore, *IFIH1* gene expression levels in peripheral blood mononuclear cells are significantly correlated with *IFIH1* genotypes, and higher *IFIH1* levels are found in individuals with the susceptible genotypes ($P = 0.005$). Thus, both genetic association and gene expression data suggest that *IFIH1* is the most plausible candidate gene implicated in T1D in this LD block.

INTRODUCTION

Mapping complex disease genes has been a daunting task partly due to the large number of genes involved in each disease and the small effect of each gene. Advances in high throughput and affordable genotyping technologies for large numbers of single nucleotide polymorphisms (SNPs) have made it possible for genome-wide association (GWA) that is revolutionizing studies of complex diseases. This approach has recently been attempted for several diseases including type 1 diabetes (T1D), type 2 diabetes and several other common diseases (1–7). However, an intrinsic problem with GWA is the expected high false-positive rates due to the large number of markers analyzed by GWA (8). It is therefore essential that the putative associations identified by GWA are confirmed in multiple large cohorts.

*To whom correspondence should be addressed at: Center for Biotechnology and Genomic Medicine, Medical College of Georgia, 1120 15th Street, Augusta, GA 30912, USA. Tel: +1 7067213410; Fax: +1 7067213688; Email: jshe@mail.mcg.edu
†These authors contributed equally to this study.

Table 1. Associations for SNPs in the *IFIH1* LD block in the Georgia and Colorado Caucasian populations

Gene	SNP	Major allele	Allelic freq.		BB versus bb		BB + Bb versus bb	
			T1D	Ctrl	OR	*P*-value	OR	*P*-value
Georgia								
KCNH7	rs12185647	G	0.457	0.455	1.0	0.9765	1.1	0.4312
KCNH7	rs2068330	G	0.674	0.640	1.4	0.0069	1.3	0.0199
Intergenic	rs984971	A	0.673	0.637	1.4	0.0043	1.2	0.0129
GCA	rs3788964	A	0.867	0.852	1.4	0.178	1.1	0.1414
IFIH1	**rs3747517**	**G**	**0.763**	**0.735**	**1.7**	**0.0006**	**1.7**	**0.00062**
IFIH1	**rs1990760**	**A**	**0.667**	**0.604**	**1.8**	**8×10^{-8}**	**1.7**	**0.000001**
Intergenic	**rs2111485**	**G**	**0.662**	**0.606**	**1.7**	**4×10^{-6}**	**1.5**	**0.00007**
Intergenic	**rs13422767**	**G**	**0.858**	**0.837**	**1.9**	**0.0082**	**1.9**	**0.0105**
FAP	rs2075302	C	0.492	0.457	1.3	0.0045	1.2	0.0099
DPP4	rs2052401	T	0.612	0.583	1.4	0.0011	1.5	0.00005
Denver								
KCNH7	rs12185647	G	0.472	0.458	1.1	0.5332	1.1	0.5893
KCNH7	rs2068330	G	0.680	0.661	1.2	0.2856	1.2	0.2502
Intergenic	rs984971	A	0.677	0.642	1.4	0.0669	1.4	0.0881
GCA	rs3788964	A	0.844	0.854	0.7	0.4557	0.7	0.4721
IFIH1	**rs3747517**	**G**	**0.736**	**0.738**	**1.3**	**0.3424**	**1.3**	**0.2056**
IFIH1	**rs1990760**	**A**	**0.630**	**0.608**	**1.3**	**0.1822**	**1.3**	**0.1444**
Intergenic	**rs2111485**	**G**	**0.639**	**0.611**	**1.3**	**0.1760**	**1.5**	**0.2216**
Intergenic	**rs13422767**	**G**	**0.831**	**0.842**	**1.0**	**0.9374**	**1.1**	**0.8509**
FAP	rs2075302	C	0.492	0.471	1.2	0.3601	1.3	0.0552
DPP4	rs2052401	T	0.620	0.572	1.5	0.0143	1.5	0.0124

All SNPs are in Hardy–Weinberg equilibrium except rs2052401 in patient population ($P = 0.003$). Odds ratios (OR) were calculated for 'bb' (minor allele homozygous genotypes) versus 'BB' (major allele homozygous genotypes) as well as 'bb' versus heterozygous genotypes 'Bb' plus 'BB' genotype. The major alleles for SNPs with allelic frequencies near 0.5 were designated in such a way that the ORs are in the same direction as the other SNPs. The allelic frequencies (Freq) are presented for the designated 'major' allele of each SNP. Data for the IFIH1 SNPs are bolded.

The first GWA for T1D (1) analyzed 6500 nonsynonymous SNPs. Although it did not cover the entire human genome, this interim analysis revealed, outside of the HLA region, 10 SNPs that showed reasonable to strong association evidence ($P = 3.2 \times 10^{-4}$—7.2×10^{-17}). Among these top 10 SNPs, four had poor genotyping quality and were not considered to be reliable (1) and another one is within the *PTPN22* gene that is well known to be associated with T1D and other autoimmune diseases in multiple populations (9–16). Therefore, we evaluate the five other SNPs that have acceptable genotyping quality in the Caucasian population ascertained from the United States and report here highly significant evidence for association between T1D and genetic markers in the *IFIH1* linkage disequilibrium (LD) block that contains four genes: fibroblast activation protein, alpha (*FAP*); interferon induced with helicase C, domain 1 (*IFIH1*); grancalcin (*GCA*); and potassium voltage-gated channel, subfamily H, member 7 (*KCNH7*). Furthermore, we evaluated and identified a significant genotype–expression correlation between the best candidate gene (*IFIH1*) and four SNPs within and at the 3′ end of the *IFIH1* gene.

RESULTS

In an attempt to confirm the associations suggested by the first GWA study, we initially genotyped five SNPs in five separate genomic regions. The association between T1D and SNP rs1990760 (exon 15 of *IFIH1*) on chromosome 2 was demonstrated in multiple data sets, whereas the associations with the other four SNPs including rs2929366 (*NM_144715*) on chromosome 3, rs9127 (*Q7Z4C4)* on chromosome 5, rs1445898 (*CAPSL*) on chromosome 5 and rs2302188 (*NM_033543*) on chromosome 19 have only been reported for the GWA data set. In our study population, there was no evidence for association between T1D and these four SNPs (Supplementary Material, Table S1). No further investigation was carried out for these four regions. Consistent with the strong evidence for association for the rs1990760 (*IFIH1*) SNP in the initial report (1), the rs1990760 SNP is significantly associated with T1D in the Georgia data set [odds ratio (OR) = 1.8, $P = 8 \times 10^{-8}$] (Table 1).

After obtaining strong evidence for association with the rs1990760 SNP in the Georgia population, we analyzed nine additional SNPs in the *IFIH1* genomic region (~450 kb). These SNPs were selected based on their high degrees of polymorphisms and spacing to cover the entire 450 kb region. On the basis of data for the GA population, 8 of the 10 SNPs (SNP 2–9) have some limited degree of LD and they form two LD sub-blocks (Fig. 1). The first LD sub-block includes six SNPs of which four are within or at the 3′ end of the IFIH1 gene, whereas the second LD sub-block contains only two SNPs with strong LD ($R^2 = 0.96$). The two SNPs in the first LD sub-block (rs1990760 and rs2111485) also have strong LD ($R^2 = 0.85$). The allelic frequencies for each SNP were shown for both T1D and control subjects (Table 1). The ORs were calculated using two commonly used approaches. In the first approach, the two homozygous genotypes for each SNP were compared with each other. These analyses revealed that the major allele homozygous genotypes (BB) are associated with higher T1D risk, while the minor allele homozygous genotypes (bb) are associated with lower T1D

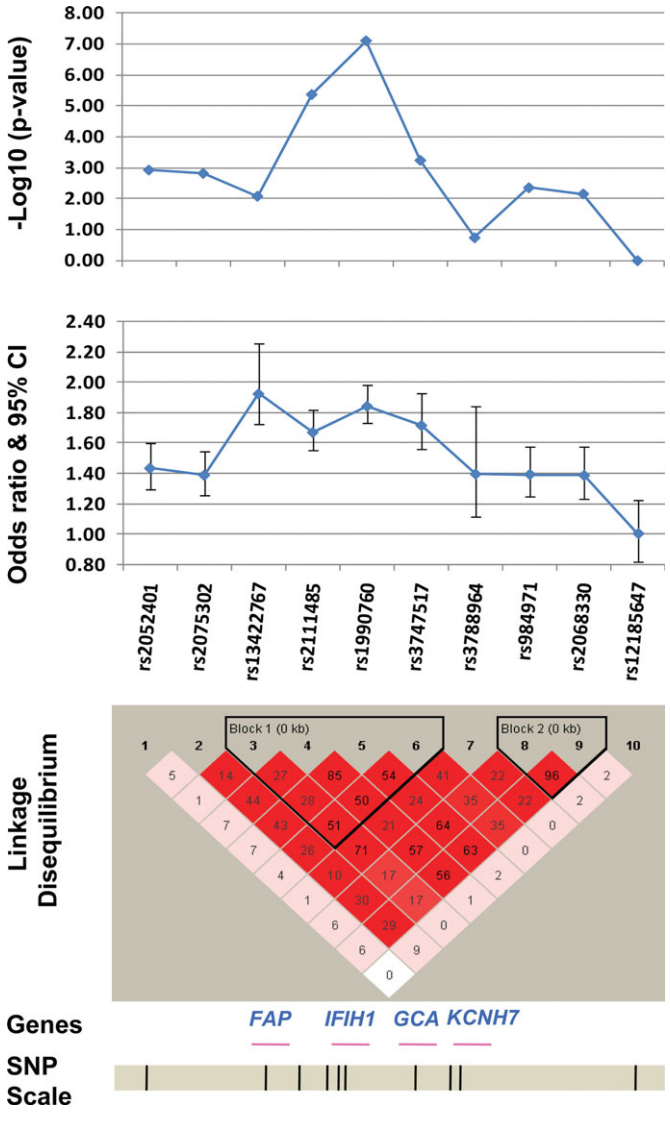

Figure 1. Genetic association between T1D and the *IFIH1* genomic region in the Georgia population. Shown are *P*-values, odds ratios and their 95% confidence interval (CI) estimated for the major allele homozygous genotypes (BB) using minor allele homozygous genotypes (bb) as reference. The whole region appears to be in LD. The R^2 values are presented between pairs of SNPs.

Table 2. Associations with SNPs in early onset, late onset and combined cases from Georgia using sex of subjects as a covariate

SNPs	OR Estimate	LCL	UCL	*P*-values SNP	SNP*sex
All cases ($n = 1434$)					
rs12185647	1.02	0.83	1.25	0.8617	0.954
rs2068330	1.48	1.16	1.88	0.0013	0.568
rs984971	1.40	1.11	1.76	0.0042	0.768
rs3788964	1.60	0.95	2.69	0.0775	0.417
rs3747517	**1.78**	**1.30**	**2.45**	**0.0003**	**0.862**
rs1990760	**1.87**	**1.49**	**2.35**	**6.6E−08**	**0.694**
rs2111485	**1.68**	**1.35**	**2.10**	**4.3E−06**	**0.529**
rs13422767	**2.31**	**1.35**	**3.95**	**0.0022**	**0.482**
rs2075302	1.33	1.09	1.62	0.0054	0.856
rs2052401	1.44	1.16	1.80	0.001	0.104
Early onset cases ($n = 893$)					
rs12185647	1.04	0.82	1.32	0.7169	0.861
rs2068330	1.41	1.06	1.86	0.0163	0.732
rs984971	1.33	1.02	1.74	0.0358	0.992
rs3788964	1.33	0.75	2.36	0.3347	0.598
rs3747517	**1.59**	**1.10**	**2.29**	**0.0136**	**0.377**
rs1990760	**1.84**	**1.41**	**2.40**	**6.6E−06**	**0.756**
rs2111485	**1.55**	**1.20**	**2.00**	**0.0007**	**0.637**
rs13422767	**1.99**	**1.09**	**3.63**	**0.0241**	**0.676**
rs2075302	1.23	0.98	1.56	0.0774	0.987
rs2052401	1.64	1.26	2.13	0.0002	0.087
Late onset cases ($n = 541$)					
rs12185647	0.98	0.74	1.31	0.9137	0.729
rs2068330	1.64	1.16	2.34	0.0057	0.455
rs984971	1.56	1.12	2.19	0.0092	0.476
rs3788964	–	–	–	–	–
rs3747517	**2.72**	**1.56**	**4.75**	**0.0004**	**0.120**
rs1990760	**2.01**	**1.43**	**2.82**	**5.3E−05**	**0.192**
rs2111485	**2.01**	**1.44**	**2.82**	**4.2E−05**	**0.280**
rs13422767	–	–	–	–	–
rs2075302	1.52	1.15	1.99	0.0029	0.776
rs2052401	1.22	0.90	1.66	0.1915	0.304

ORs are estimated using logistic regression with sex as a covariate. *P*-values for SNPs are for associations between T1D and SNPs. *P*-values for SNP*sex measure the interaction between sex and SNP association. In the late onset subset, the numbers of subjects are too small for two SNPs (with '–') and the analyses were not carried out. Data for IFIH1 SNPs are bolded.

risk. Eight of the 10 SNPs showed significant associations (Table 1). In the second approach, the minor allele homozygous genotypes (bb) were compared with the major allele homozygous and heterozygous genotypes (BB + Bb). These analyses suggest that the heterozygous genotypes also have high risk than the minor allele homozygous genotypes. Both ORs and the *P*-values indicate that the four SNPs within *IFIH1* or at the 3′ side of *IFIH1* and in the middle of the genomic region (rs13422767, rs2111485, rs1990760 and rs3747517) (OR = 1.7–1.9) are SNPs that are more strongly associated with T1D than the remaining SNPs in the region (OR = 1.1–1.4) (Table 1). The most significant *P*-values were obtained for two SNPs within the *IFIH1* gene ($P = 8 \times 10^{-8}$ and 4×10^{-6} for rs1990760 and rs2111485, respectively) (Fig. 1). These SNPs also showed significant associations when the minor allele homozygotes are compared with the other two genotypes (bb versus BB + Bb) (Table 1). Our data therefore provide strong evidence for association between the *IFIH1* LD block and T1D in the Georgia population.

We next test whether the associations between SNPs in the *IFIH1* LD block is dependent on other covariates such as gender of the subjects, age of onset of diabetes or HLA-DQB1 genotypes using logistic regression. As shown in Table 2, the *IFIH1* SNPs are significantly associated with T1D in both early and late onset T1D data sets with similar OR estimates from logistic regression. Furthermore, no significant interaction between SNP and sex for any of the SNPs in the early onset, late onset or combined data sets (Table 2), suggesting that the associations are not dependent on sex of the subjects (ORs are not heterogeneous). To assess the potential effect of HLA on the observed associations, SNP/T1D associations were examined for subject groups with different HLA-DQB1 genotypes. All subjects were classified into two subsets, one with the high-risk DQB1 genotypes (0201/0201, 0302/0302 or 0201/0302) and the second subset with

Table 3. Associations for SNPs in the *IFIH1* region after stratification by HLA DQB1 genotypes

SNP	Major allele	Low-risk DQB1		High-risk DQB1		Het
		OR	*P*-value	OR	*P*-value	*P*-value
Georgia						
rs12185647	G	1.1	0.5217	1.0	0.8561	0.6699
rs2068330	G	1.9	0.0001	1.2	0.5748	0.1316
rs984971	A	1.7	0.0006	1.2	0.5422	0.1451
rs3788964	A	1.3	0.372	1.8	0.2607	0.8978
rs3747517	**G**	**1.9**	**0.001**	**1.5**	**0.1930**	0.4785
rs1990760	**A**	**2.1**	**3×10^{-7}**	**1.4**	**0.2260**	0.0911
rs2111485	**G**	**2.0**	**3×10^{-6}**	**1.3**	**0.3539**	0.0660
rs13422767	**G**	**2.4**	**0.009**	**1.9**	**0.1434**	0.7216
rs2075302	C	1.5	0.0025	0.8	0.4355	0.0183
rs2052401	T	1.5	0.0032	1.3	0.3518	0.6879
Denver						
rs12185647	G	0.8	0.4484	1.0	0.9700	0.7012
rs2068330	G	1.1	0.6538	1.1	0.7288	0.8183
rs984971	A	1.3	0.2651	1.5	0.1761	0.7881
rs3788964	A	0.8	0.3586	0.8	0.8398	0.7246
rs3747517	**G**	**1.2**	**0.6594**	**1.5**	**0.2767**	0.6159
rs1990760	**A**	**1.1**	**0.7338**	**1.5**	**0.1861**	0.4304
rs2111485	**G**	**1.1**	**0.6455**	**1.5**	**0.1860**	0.4638
rs13422767	**G**	**0.8**	**0.5646**	**1.6**	**0.3753**	0.2982
rs2075302	C	1.2	0.4581	1.0	0.8833	0.7184
rs2052401	T	1.5	0.0901	1.6	0.1032	0.9014

All subjects in each population were separated into high-risk versus low-risk subsets according to their HLA DQB1 genotypes as discussed in the text. Odds ratios (OR) presented here were calculated for 'BB' (major allele homozygous genotypes) versus 'bb' (minor allele homozygous genotypes). Het *P*-values were derived from heterogeneity test using the Breslow–Day test. Data for the IFIH1 SNPs are bolded.

low-risk DQB1 genotypes (all others). These analyses revealed highly significant associations in the low-risk HLA subset, but not in the high-risk HLA subset (Table 3). Since the *P*-values for these tests are heavily influenced by the smaller number of control subjects having the high-risk HLA genotypes, we performed heterogeneity tests to determine whether the ORs significantly differ between the high- and low-risk subsets. Although some tests had *P*-values near 0.05 (Table 3), there is no conclusive evidence that the ORs for the SNPs differ between HLA–DQB1 risk subsets in the GA population.

To further assess the difference in association between high- and low-risk HLA subsets, we use a control population that has more subjects with high-risk HLA genotypes. For this purpose, we analyzed a sample set from Colorado that contains an unbiased patient population but a control population that was enriched for HLA high-risk genotypes with the understanding that the HLA bias may have an impact on the association between *IFIH1* SNPs and T1D. This data set should allow more reliable estimate of the contribution of the *IFIH1* SNPs to T1D risk in the subjects with high risk HLA genotypes. As shown in Table 1, the SNP frequencies in the entire patient group and control group from the Colorado are similar to those observed in the Georgia data set. Three *IFIH1* SNPs (rs3747517, rs1990760 and rs2111485) had similar ORs in the Colorado data set compared with the Georgia data set, although the SNPs did not reach statistical significance (Table 1). Logistic regression analyses using sex and age of onset did not find any significant relations and did not significantly improve associations between T1D and the *IFIH1* SNPs (data not shown). Stratification by HLA risk revealed that ORs for the *IFIH1* SNPs in the high-risk HLA subset is very similar to those observed in the GA data set (Table 3). In contrast to the Georgia data set, the *IFIH1* SNPs did not show any difference between patients and controls with low-risk HLA genotypes (Table 3). This difference between Colorado and Georgia may partially explain why no significant association is observed between *IFIH1* SNPs and T1D in the Colorado data set.

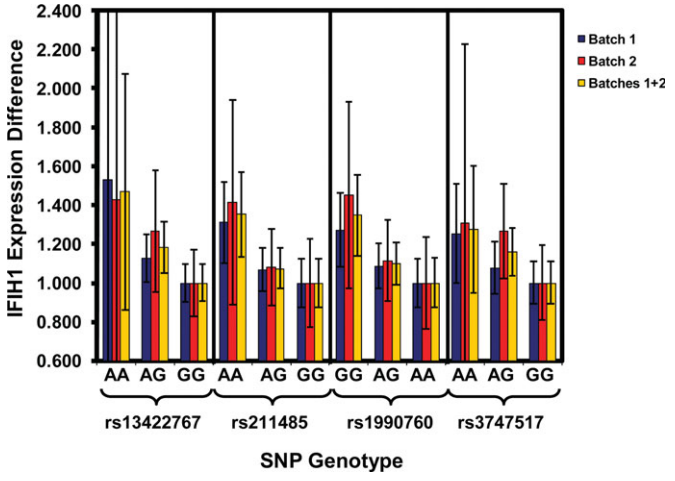

Figure 2. Correlation between *IFIH1* expression level and *IFIH1* genotypes. Expression levels of *IFIH1* are presented for the three genotypes at each of the four IFIH1 SNPs. The mean expression level was arbitrarily set to 1 for the minor allele homozygous genotype at each SNP.

As the genetic association data indicated that *IFIH1* may be the best candidate gene for this LD block that contains four annotated genes (*FAP*, *IFIH1*, *GCA* and *KCNH7*), we further tested this hypothesis using functional studies. We decided to search for potential correlation between *IFIH1* genotype and *IFIH1* gene expression. Owing to limitation of sample availability, we were only able to examine gene expression in the peripheral blood mononuclear cells (PBMCs) purified from patient and control subjects. Initially, we determined whether these four genes can be detected in PBMCs using real-time RT–PCR and found that the expression of *FAP* and *KCNH7* were below the reliable detection level but *IFIH1* and *GCA* were readily detectable in PBMCs without any stimulation. Therefore, we analyzed the basal expression levels of *IFIH1* and *GCA* in the PBMC of 384 subjects. The samples were analyzed in two different batches, each containing 192 samples. Within each batch, 96 T1D and 96 controls were analyzed. Our primary goal of this experiment was to examine the possible correlation between *IFIH1* with a secondary goal to examine potential differences between T1D and control subjects. Using data from both batches, we first used logistic regression to examine the correlation between SNPs and gene expression (both *IFIH1* and *GCA*) using T1D status as covariate. As no significant interaction was revealed between T1D and SNP/expression correlation, T1D and AbN subjects can be pooled to estimate the differences in *IFIH1* or *GCA* expression between the SNP genotypes. As shown in Fig. 2, the mean *IFIH1* levels in the heterozygous

Table 4. Estimated differences of *IFIH1* and GCA expression according to *IFIH1* genotypes and between T1D and control subjects in human PBMCs

Genes	SNP	ΔC_T between homozygous genotypes					ΔC_T between AbN and T1D				T1D*SNP
		Estimate	LCL	UCL	*P*-value	Adj. *P*	Estimate	LCL	UCL	*P*-value	*P*-value
IFIH1	rs13422767	0.571	0.172	0.970	0.005	0.011	0.211	0.006	0.416	0.121	0.743
IFIH1	rs1990760	0.407	0.120	0.694	0.006	0.011	0.213	0.011	0.414	0.267	0.560
IFIH1	rs2111485	0.398	0.110	0.687	0.007	0.011	0.198	−0.005	0.400	0.226	0.762
IFIH1	rs3747517	0.409	0.080	0.739	0.015	0.015	0.195	−0.011	0.400	0.203	0.734
IFIH1	rs3788964	0.533	0.136	0.930	0.009	0.011	0.207	0.006	0.409	0.089	0.946
GCA	rs13422767	0.279	−0.086	0.645	0.133	0.167	0.120	−0.069	0.310	0.331	0.857
GCA	rs1990760	0.213	−0.047	0.473	0.107	0.167	0.135	−0.050	0.320	0.745	0.292
GCA	rs2111485	0.255	−0.004	0.515	0.054	0.167	0.111	−0.073	0.296	0.736	0.425
GCA	rs3747517	0.170	−0.126	0.467	0.259	0.259	0.115	−0.074	0.303	0.584	0.500
GCA	rs3788964	0.320	−0.044	0.685	0.085	0.167	0.119	−0.066	0.305	0.287	0.958

IFIH1 and *GCA* expression levels were measured by RT–PCR (2) in 384 subjects (192 T1D and 192 controls). Data were analyzed using logistic regression. T1D*SNP *P*-values indicate that there is no interaction between the disease phenotype and the SNPs. Therefore, the correlation between SNP genotype and gene (*IFIH1* or *GCA*) expression was assessed using both T1D and control subjects. The expression differences between homozygous genotypes are expressed as delta (Δ) C_T (cycle threshold) between C_T for bb (minor allele homozygotes) and C_T for BB (the major allele homozygotes), e.g. $\Delta C_T = C_T$ for bb $- C_T$ for BB. Therefore, a positive ΔC_T indicate higher expression level for the major allele homozygotes compared with minor allele homozygotes. For example, the mean *IFIH1* expression in subjects with rs13422767 GG homozygotes (major allele) is higher (by 0.57 cycle or $2^{0.57} = 1.5$-fold) than the minor allele homozygotes. Similarly, positive values for 'ΔC_T between AbN and T1D' indicates higher expression in T1D subjects. LCL, lower confidence limit; UCL, upper confidence limit. Adj. *P*, adjusted *P*-value.

genotypes are between the mean levels in the two homozygous genotypes. Therefore, gene expression differences between SNP genotypes were further examined using a regression model in which the association between gene expression and SNP genotype was additive. The estimated mean expression differences between homozygous genotypes are consistent in the two batches of samples (Fig. 2), and the estimates for the combined data set are presented in Table 4. These analyses indicated that the expression of *IFIH1* was significantly higher for the major allele homozygous genotypes (BB), which are associated with higher T1D risk, compared with the minor allele homozygous genotypes (bb), which are associated with lower T1D risk. For example, the mean expression level of *IFIH1* was 43% (or 0.57 C_T) higher for subjects with the high-risk GG homozygotes at rs13422767 than subjects with the low-risk homozygous genotype. These results suggest that higher *IFIH1* expression may be associated with increased risk for T1D. Indeed, the mean *IFIH1* expression was slightly higher in T1D subjects than controls even though the differences did not reach the statistical significance. This potential difference should be further examined in larger data sets. The *GCA* expression data were analyzed in the same fashion but no significant genotype–expression correlation was found (Table 4), further strengthening the argument that *IFIH1* may be the best candidate gene. Of course, the lack of gene expression difference alone does not exclude this gene as a candidate gene.

DISCUSSION

Since the initial report on *IFIH1* association with T1D, significant association has been reported between *IFIH1* and two other autoimmune diseases including Grave's disease (17) and multiple sclerosis (18) but no association was observed in two other studies (19,20). However, confirmatory evidence for *IFIH1* association with T1D is still lacking despite an independent report in a family-based study that provided marginal association evidence ($P = 0.03$) (21). In this study, we used a relatively large case/control data set with 10 SNPs in the *IFIH1* region and we obtained strong evidence for association between T1D and the *IFIH1* genomic region. Therefore, our study provides the urgently needed evidence supporting the *IFIH1* LD block as a T1D susceptibility interval.

On the basis of our genetic association data (OR and *P*-value) in the Georgia population, the two SNPs within the *IFIH1* coding region (rs1990760 and rs3747517) and two other SNPs (rs13422767 and rs2111485) in the 3′ intergenic region of *IFIH1* (23 and 13 kb from the end of *IFIH1* gene) had the best *P*-value and ORs. These data suggest that *IFIH1* may be the best candidate gene. To further substantiate this hypothesis, we took an approach with the hope to reveal potential correlation between *IFIH1* expression and *IFIH1* genotypes. Using real-time RT–PCR, we analyzed the basal expression levels of *IFIH1* in the PBMC and revealed a significant correlation between *IFIH1* expression level and genotypes at four SNPs within or near *IFIH1*. The expression results were obtained in PBMCs without *in vitro* stimulation and therefore reflect the basal expression levels. After viral infections, the expression level of *IFIH1* is significantly up-regulated. It is possible that the *IFIH1* expression differences between *IFIH1* genotypes are further increased upon up-regulation by viral infection. This hypothesis should be addressed in future studies.

Since *IFIH1* expression level is correlated with genotypes at four SNPs, an interesting question is which SNP(s) might be responsible for the expression difference. This is a difficult question because the SNPs are in LD and we have not identified/analyzed all SNPs in the region. Two of the four SNPs (rs13422767 and rs2111485) are located at the 3′ end of the *IFIH1* gene (23 and 13 kb from the end of the 3′-UTR of the *IFIH1* gene, respectively). These SNPs do not change any known transcription factor binding sites and it is unknown whether they contribute to the regulation of *IFIH1* expression. The two other SNPs (rs1990760 and rs3747517) are both non-synonymous polymorphism within the *IFIH1*

coding region. The rs1990760 SNP is located within an HNF-3b binding site and rs3747517 is within an AP-1 binding site. These SNPs are located in the coding region that are 45–50 kb from the start codon and it remains to be determined whether these SNPs may play a role in regulating *IFIH1* expression.

IFIH1 is a functionally plausible candidate gene for T1D. The innate immune system has developed two pathways for the recognition of dsRNA produced during viral replication or directly from dsRNA viruses. The widely known pathway for dsRNA is TLR3, which allows cells to detect dsRNA that is phacocytosed from the extracellular space and allow the detection of dsRNA viruses that are internalized from the extracellular space through receptor-mediated endocytosis. The second pathway for detecting dsRNA is mediated by cytoplasmic dsRNA sensors (IFIH1 and RIG-I), which allow cells to directly detect intracellular viral infection. IFIH1 and RIG-I are cytoplasmic helicases. IFIH1 or RIG-I signaling activates interferon-regulatory factors (IRFs) and other transcription factors (NFkB) that induce INF-α/β and interferon-inducible genes. The dsRNA recognition pathways have overlapping but non-redundant functions. Some viruses and dsRNA analogs (including polyI:C) may be detected by both TLR3 and cytoplasmic RNA helicase pathways, while others may be sensed primarily through one pathway. The two helicases also discriminate among different dsRNA ligands, even though they share similar signaling features and structural homologies. The signaling pathways have profound impact on the production of the types and amounts of cytokines, surface molecules and DC function. For example, *IFIH1*$^{-/-}$ mice treated with polyI:C have decreased levels of the proinflammatory cytokine IL-6 and monocytes chemotractant protein 1 (MCP-1), in addition to abrogation of type I IFN production (22,23). One recent report also suggested that signaling through cytoplasmic helicases/IFIH1 resulted in markedly higher capacity of human monocyte-derived DC in stimulating CD4^{+} T-cell proliferation (24). These results suggest that signaling through IFIH1 may result in immunogenic dendritic cells that may promote the development of diabetes.

The *IFIH1* expression pattern observed in our study is consistent with the potential pathogenic mechanism. Individuals with *IFIH1* susceptible genotypes are associated with higher *IFIH1* levels while the protective genotypes are associated with lower *IFIH1* levels. In individuals with higher *IFIH1* levels, infections by viral triggers such as Coxsackie viruses may be primarily recognized by the IFIH1 pathway, leading to enhanced stimulating capacity of dendritic cells and production of proinflammatory cytokines. This hypothesis is highly plausible but requires further functional prove. If our hypothesis is proved true, IFIH1 would provide a molecular link between genetic susceptibility, viral infections and the innate immune response, which are all believed to contribute to T1D pathogenesis.

MATERIALS AND METHODS

Study populations

The main study population consists of Caucasian T1D patients and healthy controls from Georgia. The data set includes 1434 patients and 1865 controls. All diabetic patients were diagnosed using the criteria of the American Diabetes Association. Most of the cases (893) developed type 1 diabetes at or before the age of 17 years and this group of patients were referred to as early onset patients. A good proportion of patients (541) developed diabetes after 17 years of age and these patients were referred to as late onset diabetes for data analysis purposes. The controls are normal subjects without a family history for T1D, and they are ascertained from the same geographic area as the patients. We have attempted to include only those subjects who have genetic heritage of European Caucasians based on the family history of the subjects. The second data set includes 612 non-Hispanic white (NHW) patients with T1D (526 early onset and 86 late onset) seen at the Barbara Davis Center in Denver and 552 NHW healthy control subjects from the same geographic area. The Colorado and Georgia data sets were analyzed separately due to ascertainment difference and potential geographic differences.

SNP genotyping methods

The SNPs were genotyped mainly using the TaqMan-assay method (25). All assays (primers and probes) used in this study were designed and validated by Applied Biosystems. Amplification reactions were performed in a 5 μl final volume in optical 384-well plates. PCR was carried out with 2 min at 50°C, 10 min at 95°C followed by 40 cycles of 15 s at 95°C and 1 min at 60°C using an ABI9700 real-time PCR system (Applied Biosystems). To validate the TaqMan assays, five SNPs were also genotyped using amplified restriction fragment length polymorphism. The PCR amplification mixture contains 2 μl (20 ng) of genomic DNA, 1.5 μl of 10× PCR buffer, 1.5 μl of dNTP (1.25 mmol/l), 0.75 μl of DMSO, 0.3 μl of each primer (20 pmol/μl) and 0.1 μl of *Taq* polymerase in a 12-μl reaction volume. Samples were initially denatured for 2 min at 94°C followed by 35 cycles of 94°C for 30 s, 30 s at 58–60°C, and 30 s at 72°C and an additional extension of 2 min at 72°C. Amplified products (12 μl) were digested using 10 units of enzyme. Digested products were electrophoresed on a 2.5–3.5% agarose gel.

Rea-time RT–PCR assay for gene expression

Peripheral blood (2.5 ml for each subject) was immediately preserved in the PaxGene RNA tubes. After 2 h at room temperature, the tubes are frozen at −80°C freezers and total RNA is extracted within a few weeks using a Qiagen kit specially developed for the PaxGene RNA tubes. Extracted RNA was stored at −80°C freezers till use. An aliquot of total RNA (2 μg per sample) were arrayed in 96-well plates and then converted to cDNA using a High Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and a PTC-100^{TM} programmable thermal controller (MJ Research, Inc.). The cDNA products are diluted and an aliquot of cDNA equivalent of 10 ng total RNA was used for quantitative real-time PCR performed using ready-to-use TaqMan gene expression assays from the Applied Biosystems. Assays for IFIH1, GCA, FAP and KCNH7 are Hs00223420_m1, Hs 00201854_m1, Hs00990806_m1 and Hs00264454_m1,

respectively. The 18S rRNA (Hs 99999901_s1) was used as endogenous control for normalizing RNA concentration. Real-time PCR was performed in 384-well plates with the ABI 7900HT Fast Real-Time PCR System. Standard thermal cycling condition (10 min at 95°C, 40 cycles for 15 s at 95°C, 1 min at 60°C) was used for all genes. All 192 samples were analyzed on the same plate, and each sample was analyzed in duplicate. Cycle threshold (C_T) values for each test gene and 18S rRNA were obtained for each sample using the SDS2.3 and analyzed with RQ Manager 1.2 software. Differences in C_T values between a test gene and 18srRNA (ΔC_T) for each sample were calculated and used for statistical analyses.

Statistical analysis for SNP association

Association between each SNP and T1D was assessed by calculating the ORs using two different approaches. In the first approach, the two homozygous genotypes (the minor allele bb homozygotes and major allele BB homozygotes) were compared between T1D and control populations. In the second approach, two of the three genotypes at each SNP locus are collapsed into one group and then compared with the third genotype (e.g. BB + Bb versus bb). *P*-values were calculated using the χ^2 test. Furthermore, the association between each SNP and phenotype (patient versus control) was tested using logistic regression to fit a linear or additive model, and only the results from the linear models are presented. We also used logistic regression to examine the associations using sex as a covariate. We used the Breslow–Day test to examine heterogeneity in the ORs between low- and high-risk HLA-DQB1 subsets. The early and late onset patients were analyzed separately and jointly to examine the potential influence of age of onset on the SNP associations.

Statistical analysis for correlation between gene expression and SNPs

IFIH1 and *GCA* expression in PBMCs were assessed in two batches of samples, each containing 192 subjects (96 with T1D and 96 controls). Gene expression differences between SNP genotypes were first examined separately for each of the two batches of samples using a regression model in which the association between gene expression and SNP genotype was additive, which is equivalent to asking whether a trend exists in the association between gene expression and SNP genotype. In such models, the estimated mean for the heterozygote will be equal to the average of the homozygotes. The estimated effects of these models were compared, with the estimates being similar between batches. As such, we used a mixed model analysis of variance in which the additive effect of SNP genotype was one factor and the random effect of batch was the second factor. We also examined the association between gene expression and SNP genotype between T1D subjects and controls using mixed-model analyses of variance. All mixed-model analyses were conducted using the Proc Mixed procedure of SAS, and models without random effects were conducted using the Proc GLM procedure.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at *HMG* Online.

ACKNOWLEDGEMENTS

We thank a number of physicians (David Brown, N. Spencer Welch, Paul Davidson, Joseph Johnson, David Robertson, Constance Baldwin, Melissa Carlucci, Mark Rappaport, Robert Schultz, Vijayasudha Gunna) and staff members (Kim English, Jessica Leggett, Katherine Guthrie, Melanie Brown, Deana McFeely, Nellie Jenkins, Tami Kinnersley and Ann Simmonds) for their contribution in sample collection. We are very grateful to all patients and other volunteers who participated in this study.

Conflict of Interest statement. None declared.

FUNDING

This study was partly supported by a grant from the National Institutes of Health (2RO1HD37800, 4R33DK069878, 4R33HD050196) (JXS), P30 DK 57516 (MJR) and Juvenile Diabetes Research Foundation International (JDRF 1-2004-661) (JXS). Purohit S. is recipient of an Advanced Postdoctoral Fellowship from JDRF (10-2006-792).

REFERENCES

1. Smyth, D.J., Cooper, J.D., Bailey, R., Field, S., Burren, O., Smink, L.J., Guja, C., Ionescu-Tirgoviste, C., Widmer, B., Dunger, D.B. *et al.* (2006) A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat. Genet.*, **38**, 617–619.
2. Sladek, R., Rocheleau, G., Rung, J., Dina, C., Shen, L., Serre, D., Boutin, P., Vincent, D., Belisle, A., Hadjadj, S. *et al.* (2007) A genome-wide association study identifies novel risk loci for type 2 diabetes. *Nature*, **445**, 881–885.
3. Todd, J.A., Walker, N.M., Cooper, J.D., Smyth, D.J., Downes, K., Plagnol, V., Bailey, R., Nejentsev, S., Field, S.F., Payne, F. *et al.* (2007) Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat. Genet.*, **39**, 857–864.
4. Hinney, A. (2007) Genome wide association (GWA) study for early onset extreme obesity supports the role of fat mass and obesity associated gene (FTO) variants. *PLoS One*, **2**, e1361.
5. Diabetes Genetics Initiative of Broad Institute of Harvard and MIT, Lund University, and Novartis Institutes of BioMedical ResearchSaxena, R., Voight, B.F., Lyssenko, V., Burtt, N.P., de Bakker, P.I.W., Chen, H., Roix, J.J., Kathiresan, S., Hirschhorn, J.N. *et al.* (2007) Genome-wide association analysis identifies loci for type 2 diabetes and triglyceride levels. *Science*, **316**, 1331–1336.
6. Hakonarson, H., Grant, S.F.A., Bradfield, J.P., Marchand, L., Kim, C.E., Glessner, J.T., Grabs, R., Casalunovo, T., Taback, S.P., Frackelton, E.C. *et al.* (2007) A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature*, **448**, 591–594.
7. The 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.
8. Todd, J.A. (2006) Statistical false positive or true disease pathway? *Nat. Genet.*, **38**, 731–733.
9. Bottini, N., Musumeci, L., Alonso, A., Rahmouni, S., Nika, K., Rostamkhani, M., MacMurray, J., Meloni, G.F., Lucarelli, P., Pellecchia, M. *et al.* (2004) A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat. Genet.*, **36**, 337–338.

10. Zheng, W. and She, J.X. (2005) Genetic association between a lymphoid tyrosine phosphatase (PTPN22) and type 1 diabetes. *Diabetes*, **54**, 906–908.
11. Begovich, A.B., Carlton, V.E., Honigberg, L.A., Schrodi, S.J., Chokkalingam, A.P., Alexander, H.C., Ardlie, K.G., Huang, Q., Smith, A.M., Spoerke, J.M. *et al.* (2004) A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am. J. Hum. Genet.*, **75**, 330–337.
12. Kyogoku, C., Langefeld, C.D., Ortmann, W.A., Lee, A., Selby, S., Carlton, V.E., Chang, M., Ramos, P., Baechler, E.C., Batliwalla, F.M. *et al.* (2004) Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am. J. Hum. Genet.*, **75**, 504–507.
13. Smyth, D., Cooper, J.D., Collins, J.E., Heward, J.M., Franklyn, J.A., Howson, J.M., Vella, A., Nutland, S., Rance, H.E., Maier, L. *et al.* (2004) Replication of an association between the lymphoid tyrosine phosphatase locus (LYP/PTPN22) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. *Diabetes*, **53**, 3020–3023.
14. Steck, A.K., Liu, S.Y., McFann, K., Barriga, K.J., Babu, S.R., Eisenbarth, G.S., Rewers, M.J. and She, J.X. (2006) Association of the PTPN22/LYP gene with type 1 diabetes. *Pediatr. Diabetes*, **7**, 274–278.
15. Zoledziewska, M., Perra, C., Orru, V., Moi, L., Frongia, P., Congia, M., Bottini, N. and Cucca, F. (2008) Further evidence of a primary, causal association of the PTPN22 620W variant with type 1 diabetes. *Diabetes*, **57**, 229–234.
16. Onengut-Gumuscu, S., Buckner, J.H. and Concannon, P. (2006) A haplotype-based analysis of the PTPN22 locus in type 1 diabetes. *Diabetes*, **55**, 2883–2889.
17. Sutherland, A., Davies, J., Owen, C.J., Vaikkakara, S., Walker, C., Cheetham, T.D., James, R.A., Perros, P., Donaldson, P.T., Cordell, H.J. *et al.* (2007) Genomic polymorphism at the interferon-induced helicase (IFIH1) locus contributes to Graves' disease susceptibility. *J. Clin. Endocrinol. Metab.*, **92**, 3338–3341.
18. Martinez, A., Santiago, J.L., Cenit, M.C., de las Heras, V., de la Calle, H., Fernandez-Arquero, M., Arroyo, R., de la Concha, E.G. and Urcelay, E. (2008) IFIH1-GCA-KCNH7 locus: influence on multiple sclerosis risk. *Eur. J. Hum. Genet*, **16**, 861–864.
19. Zhao, Z.F., Cui, B., Chen, H.Y., Wang, S., Li, I., Gu, X.J., Qi, L., Li, X.Y., Ning, G. and Zhao, Y.J. (2007) The A946T polymorphism in the interferon induced helicase gene does not confer susceptibility to Graves' disease in Chinese population. *Endocrine*, **32**, 143–147.
20. Marinou, I., Montgomery, D., Dickson, M., Binks, M., Moore, D., Bax, D. and Wilson, A. (2007) The interferon induced with helicase domain 1 A946T polymorphism is not associated with rheumatoid arthritis. *Arthritis Res. Ther.*, **9**, R40.
21. Qu, H.Q., Marchand, L., Grabs, R. and Polychronakos, C. (2008) The association between the IFIH1 locus and type 1 diabetes. *Diabetologia*, **51**, 473–475.
22. Gitlin, L., Barchet, W., Gilfillan, S., Cella, M., Beutler, B., Flavell, R.A., Diamond, M.S. and Colonna, M. (2006) Essential role of mda-5 in type I IFN responses to polyriboinosinic:polyribocytidylic acid and encephalomyocarditis picornavirus. *Proc. Natl Acad. Sci.*, **103**, 8459–8464.
23. Loo, Y.M., Fornek, J., Crochet, N., Bajwa, G., Perwitasari, O., Martinez-Sobrido, L., Akira, S., Gill, M.A., Garcia-Sastre, A., Katze, M.G. and Gale, M. Jr (2008) Distinct RIG-I and MDA5 signaling by RNA viruses in innate immunity. *J. Virol.*, **82**, 335–345.
24. Groschel, S., Piggott, K.D., Vaglio, A., Ma-Krupa, W., Singh, K., Goronzy, J.J. and Weyand, C.M. (2008) TLR-mediated induction of negative regulatory ligands on dendritic cells. *J. Mol. Med.*
25. de Kok, J.B., Wiegerinck, E.T., Giesendorf, B.A. and Swinkels, D.W. (2002) Rapid genotyping of single nucleotide polymorphisms using novel minor groove binding DNA oligonucleotides (MGB probes). *Hum. Mutat.*, **19**, 554–559.