Original Article

A Haplotype-Based Analysis of the *PTPN22* Locus in Type 1 Diabetes

Suna Onengut-Gumuscu,[1,2] Jane H. Buckner,[3] and Patrick Concannon[1,2]

A recent addition to the list of widely confirmed type 1 diabetes risk loci is the *PTPN22* gene encoding a lymphoid-specific phosphatase (Lyp). However, evidence supporting a role for *PTPN22* in type 1 diabetes derives entirely from the study of just one coding single nucleotide polymorphism, 1858C/T. In the current study, the haplotype structure of the *PTPN22* region was determined, and individual haplotypes were tested for association with type 1 diabetes in family-based tests. The 1858T risk allele occurred on only a single haplotype that was strongly associated with type 1 diabetes ($P = 7.9 \times 10^{-5}$). After controlling for the effects of this allele, two other haplotypes were observed to be weakly associated with type 1 diabetes ($P < 0.05$). Sequencing of the coding region of *PTPN22* on these haplotypes revealed a novel variant (2250G/C) predicted to result in a nonsynonymous amino acid substitution. Analysis of *PTPN22* transcripts from a subject heterozygous for this variant indicated that it interfered with normal mRNA splicing, resulting in a premature termination codon after exon 17. These results support the conclusion that the 1858C/T allele is the major risk variant for type 1 diabetes in the *PTPN22* locus, but they suggest that additional infrequent coding variants at *PTPN22* may also contribute to type 1 diabetes risk. *Diabetes* 55:2883–2889, 2006

Type 1 diabetes is a complex disorder of glucose homeostasis that arises from the autoimmune destruction of the insulin-secreting β-cells of the pancreas, resulting in a lifelong dependence on exogenously administered insulin. The increased concordance rates for type 1 diabetes observed in monozygotic twins and other relatives of type 1 diabetic probands suggest that some fraction of the risk for type 1 diabetes is inherited (1,2). Numerous attempts have been made to identify type 1 diabetes risk loci through both genome-wide linkage approaches and candidate gene association studies (rev. in 3,4). Although published reports have provided evidence supporting the involvement of many different chromosomal regions or genes in determining type 1 diabetes risk, only a relatively limited number of these loci have been consistently and widely replicated in follow-up studies (5).

Regions of the genome that are widely accepted as conferring risk for type 1 diabetes include the HLA region on chromosome 6p21, which harbors multiple loci whose products both act and interact to increase risk (6); the insulin (*INS*) gene region on chromosome 11p15, where alleles of a variable-number tandem repeat polymorphism located 5′ to the insulin gene are differentially associated with type 1 diabetes (7–9); and the *CTLA4* gene region, where alternative splicing associated with a single nucleotide polymorphism (SNP) located 3′ of the gene influences type 1 diabetes risk (10,11). Although initially identified based on allelic associations with type 1 diabetes, each of these three regions has also been detected in genome-wide scans carried out in affected sibpair families as sites where there is suggestive to significant evidence of linkage to type 1 diabetes (12). However, no single locus in any of these regions can fully account for the magnitude of the evidence of linkage observed in that region, suggesting that the linkage peaks may represent the contributions of multiple, clustered loci (5,12), as has been observed in animal models of autoimmunity (13,14).

A recent addition to the list of generally replicated type 1 diabetes risk loci is the *PTPN22* locus on chromosome 1p13 (15–17). *PTPN22* encodes a 110-kDa lymphoid-specific phosphatase (Lyp). The NH_2 terminus of the protein contains a catalytically active phosphatase domain, whereas the COOH terminus contains several proline-rich repeats (18). The murine homologue of Lyp, Pep, binds to the SH3 domain of c-Src tyrosine kinase (Csk) via its most NH_2-terminal proline-rich sequence (P1) in in vitro binding assays, and this interaction plays an important role in the downregulation of T-cell receptor signaling (19,20). Pep acts as a negative regulator of T-cell activation by dephosphorylating the T-cell receptor activation–dependent kinases Lck, Fyn, and Zap70 (20). In humans, a nonsynonymous C-to-T SNP at nucleotide position 1858 of the *PTPN22* gene (rs2476601) results in an amino acid substitution at codon 620 from arginine (CGG) to tryptophan (TGG). This substitution is located in the proline-rich P1 sequence of Lyp and likely affects its ability to interact with Csk in vivo because reduced levels of Csk are coimmunoprecipitated by the 620W variant, compared with the 620R form, when these isoforms are separately expressed as epitope-tagged constructs in T-cell lines (21).

Bottini et al. (15) first noted an association between the 1858T (620W) allele in the *PTPN22* gene and autoimmune disease. They observed that the 1858T allele was more frequent among type 1 diabetes case compared with

From the [1]Molecular Genetics Program, Benaroya Research Institute, Seattle, Washington; the [2]Department of Immunology, University of Washington School of Medicine, Seattle, Washington; and the [3]Translational Research Program, Benaroya Research Institute, Seattle, Washington.

Address correspondence and reprint requests to Patrick Concannon at Molecular Genetics Program, Benaroya Research Institute, 1201 Ninth Ave., Seattle, WA 98101-2795. E-mail: patcon@benaroyaresearch.org.

Received for publication 16 February 2006 and accepted in revised form 5 July 2006.

CEPH, Centre d'Etude du Polymorphisme Humain; DHPLC, denaturing high-performance liquid chromatography; FBAT, family-based association test; IBD, identical-by-descent; SNP, single nucleotide polymorphism.

DOI: 10.2337/db06-0225

control subjects ascertained in North America and Sardinia. Begovich et al. (21) similarly reported that the allele frequency of the 1858T allele was elevated in rheumatoid arthritis case subjects in two separate case-control populations from North America. Kyogoku et al. (22) observed that the 1858T (620W) allele was also associated with increased risk of systemic lupus erythematosis and reported evidence for a dosage effect suggesting that the 1858T (620W) allele might be a loss of function variant. Smyth et al. (16) reported the frequency of the 1858T allele to be elevated in Graves' disease patients from the U.K. compared with control subjects. These studies suggested the possibility that *PTPN22* R620W might have a generalized predisposing effect on autoimmune risk. However, subsequent studies have revealed several autoimmune disorders, including multiple sclerosis, psoriasis, and Crohn's disease, where no association with alleles at the *PTPN22* 1858C/T polymorphism is detected (23–25).

The 1858C/T SNP has been extensively analyzed in type 1 diabetes case-control and family-based studies (15–17,26–30). Although these studies uniformly indicate that the 1858T allele is associated with type 1 diabetes, they do not address the issue of whether 1858T is the sole allele at *PTPN22* that influences risk for type 1 diabetes. In the current study, a comprehensive approach was taken to analyze the coding and intervening sequences of *PTPN22*. The coding sequence of *PTPN22* was screened for the presence of additional coding variants in type 1 diabetes case subjects, and the common SNPs identified through this analysis, along with tagging SNPs selected from the HapMap database, were genotyped in type 1 diabetes–affected families and used in haplotype-based tests of association.

RESEARCH DESIGN AND METHODS

Type 1 diabetes–affected sibpair families (n = 374) were obtained from the Human Biological Data Interchange (HBDI) repository. These samples had been previously described and consist of nuclear families, all with both parents and at least two affected offspring available for genotyping (31). All of the Human Biological Data Interchange families that were genotyped for the current study were ascertained in the U.S. and were of European ancestry. An additional 243 affected sibpair families from the British Diabetes Association (BDA) were used in association studies for the rare coding variants. Although these British Diabetes Association families include multiple affected siblings, only a single affected sib, chosen at random, was genotyped along with both parents.

Mutation screening. All exons and flanking intronic sequences of *PTPN22* were screened for variation in 94 type 1 diabetic case subjects homozygous for the 1858C allele. Subjects for screening were chosen based on the extent of alleles shared identical-by-descent (IBD) among affected sibpairs on chromosome 1. The program Merlin (32) was used to estimate IBD status among sibpairs from a prior genome scan of 389 U.S. type 1 diabetes–affected families (31). Based on IBD scores, an affected sib was chosen from each of 16 families that scored 2, 26 families with a score of 1, and 28 families where IBD could not be determined because of the low information content of the microsatellites in the area but where case subjects were homozygous for the wild-type allele at position 1858. An additional 24 sporadic type 1 diabetes case subjects without prior genome scan data were included, bringing the total number of samples to 94.

Screening primers were designed using the program Primer 3 (33) and were situated in introns ~40 bp away from the intron-exon boundaries based on the reference sequences NM_015967.3 (*PTPN22*), NM_001010922.1 (C1orf178), and NT_019273 (*PTPN22*). Exon 13 was amplified in three overlapping PCR products, whereas exons 6 and 7 were amplified as a single product (383 bp). Amplifications were performed in a total volume of 25 μl containing 40 ng DNA. A touchdown protocol from 65 to 55°C was used for amplification.

Heterozygous individuals were detected by denaturing high-performance liquid chromatography (DHPLC) on a Transgenomic WAVE DNA fragment analysis system (Transgenomic, Omaha, NE). Elution patterns were visually analyzed to identify samples yielding heteroduplexes indicative of nucleotide mismatches. Heteroduplex samples were sequenced along with a control sample that had a homoduplex profile. Sequencing was performed using BigDye Terminator v3.1 (Applied Biosystems) on an ABI3100 gene analyzer.

Tagging SNPs. HapMap Centre d'Etude du Polymorphisme Humain (CEPH) family genotype data (CEU database) was downloaded to Haploview, and linkage disequilibrium patterns in and around *PTPN22* were established (34). Seven haplotype-tagging SNPs were chosen from HapMap SNPs with minor allele frequencies >10% (rs3811021, rs2476599, rs1217388, rs2476601, rs1217414, rs2488457, and rs1235005). All seven selected tagging SNPs were genotyped in 374 type 1 diabetes–affected sibpair families. Haplotype association testing was performed using the HBAT command in the family-based association test (FBAT) (35,36).

SNP genotyping. The MGB-Eclipse system (Nanogen) was used to genotype the two coding and nine intronic variants in *PTPN22*. Genotyping assays were performed using 10 ng of DNA in a 5-μl reaction volume according to the manufacturer's protocol and analyzed on an ABI HT7900. Pedcheck was used to test for Mendelian inconsistency in genotyping (37). FBAT (version 1.5.5) was used for association studies in the type 1 diabetes multiplex families (35). FBAT was run under an additive model. Because multiplex type 1 diabetes–affected families were analyzed, the –e flag option was used in FBAT to account for the correlation of sibling genotypes when linkage is present (38).

cDNA analysis. Frozen peripheral blood mononuclear cells used for cDNA analysis were derived from subjects participating in studies under the auspices of the JDRF (Juvenile Diabetes Research Foundation) Center for Translational Research. Informed consent was obtained from all subjects according to institutional review board approval protocols at Children's Hospital and Regional Medical Center and Benaroya Research Institute. RNA was extracted from frozen peripheral blood mononuclear cells using Trizol (Invitrogen) in accordance with the manufacturer's protocols. First-strand cDNA was synthesized using SuperScript II RT (Invitrogen) with random primers according to the manufacturer's instructions. Primer sequences for cDNA amplifications are as follows: rPTPN22 18F-5′TCCTGACACCATG GAAAATTCA3′, 21R-5′CAGGTGTACTTGCAGCCCATATT3′, 15F-5′GGGTG GAACATCTGAACCAAAG3′, and 20R-5′CCAAAATTCAGAAATGAGCTGGA3′. Before sequencing, amplified fragments were gel purified using a Qiagen Gel Extraction kit (Qiagen).

RESULTS

The *PTPN22* gene consists of 21 exons spanning ~58 kb. To detect possible coding variants besides the 1858T allele, 94 type 1 diabetic case subjects homozygous for the 1858C allele were screened. These subjects were selected from among affected sibpairs in multiplex type 1 diabetic families based on increased allele sharing in the region of the *PTPN22* gene. A total of 11 variants were identified after sequencing all samples that displayed heteroduplex formation by DHPLC analysis (Table 1). Three of these SNPs had relatively frequent minor alleles and were already present in the dbSNP database (rs1217418, rs2797415, and rs3761935); the remaining eight SNPs were novel variants. The three common SNPs detected were included in the subsequent linkage disequilibrium and haplotype analyses. Three of the novel variants were predicted to result in amino acid substitutions, one each in exons 8 (658A/G, S220G), 13 (1508A/G, Y503C), and 18 (2250G/C, K750N).

For a comprehensive genetic analysis of the coding and intervening sequences of *PTPN22*, tagging SNPs from the HapMap database were used in conjunction with the three common SNPs identified by DHPLC analysis in family-based single-marker and haplotype association tests. HapMap markers spanning *PTPN22* were in strong linkage disequilibrium with each other in the CEU dataset, making a tagSNP approach feasible. Seven haplotype-tagging SNPs (rs3811021, rs2476599, rs1217388, rs2476601, rs1217414, rs2488457, and rs1235005) were selected, which accounted for 95.3% of the chromosomes derived from the 30 CEPH family trios genotyped at *PTPN22* in the HapMap database. The rs2476601 SNP is the nonsynonymous 1858C/T SNP previously shown to be associated

TABLE 1
Variants identified in 94 type 1 diabetic case subjects who were homozygous for the 1858C allele in *PTPN22*

SNP	Position	Nucleotide change*	Amino acid change	dbSNP	Exon	Observed on DHPLC†
IVS2–4	114113746	A>G	—	—	—	1
IVS3–36	114113273	C>T	—	rs1217418	—	42
IVS3–32	114113269	A>C	—	—	—	1
658	114109596	A>G	S220G	—	8	1
IVS9+112	114109023	T>A	—	—	—	2
1508	114092556	A>G	Y503C	—	13	1
IVS14–32	114089135	C>T	—	rs2797415	—	35
IVS17+43	114084570	A>C	—	rs3761935	—	30
IVS17+47	114084566	T>C	—	—	—	4
2250	114084256	G>C	K750N	—	18	2
IVS20+81	114074160	T>A	—	—	—	3

*Nucleotide changes were identified by sequencing samples that showed a heteroduplex elution pattern on DHPLC analysis; †number of times a heteroduplex elution pattern was observed in 94 samples that were screened for the indicated exon.

with type 1 diabetes (17) in the multiplex type 1 diabetes–affected families studied here. These seven SNPs, along with the three common intronic SNPs detected by DHPLC screening, were genotyped in 374 type 1 diabetes multiplex families (Table 2). Single-marker association tests indicated that the minor alleles of rs2476601, rs1235005, rs2488457, rs1217418, rs2797415, and rs1217388 were preferentially transmitted to affected offspring from heterozygote parents, whereas the remaining four SNPs provided no evidence of association to type 1 diabetes in the families genotyped.

All 10 SNPs spanning *PTPN22* were in strong linkage disequilibrium with each other ($D' > 0.9$) and captured seven haplotypes, which accounted for 98.7% of the founder chromosomes in the population studied (Table 3). Five of these haplotypes were relatively common, with frequencies >10%. The 1858T allele (rs2476601) occurred on only a single haplotype (H4), and this haplotype was strongly associated with type 1 diabetes ($P = 7.9 \times 10^{-5}$) in the 374 type 1 diabetes–affected families screened. The H5 haplotype differed from H4 only at position 1858 of the coding region (rs2476601) and provided no significant evidence of association with type 1 diabetes. A second haplotype (H2) was undertransmitted to affected offspring from informative parents ($P = 0.0022$). When the haplotype association test was conditioned on type 1 diabetic case subjects who were homozygous for the 1858C allele at rs2476601, two haplotypes, H1 ($P = 0.031$) and H3 ($P = 0.041$), displayed increased transmission rates to affected offspring that were at least marginally significant. These associations might reflect the presence of additional rare variants of *PTPN22* occurring on these haplotypes that confer some increased risk for type 1 diabetes.

The DHPLC screening of the *PTPN22* coding region identified three novel but rare nonsynonymous coding variants (Table 1). To evaluate the possible role of these variants in type 1 diabetes susceptibility, they were evaluated for their likelihood of affecting protein structure, using the programs PolyPhen and SIFT (39,40). Two of the variants, 1508A/G and 2250G/C, had minor alleles that scored as possibly damaging to *PTPN22* in this analysis. These variants were genotyped in 617 type 1 diabetes multiplex families and tested for association with type 1 diabetes. The 1508A/G variant was observed in only two families (predicted minor allele frequency 0.0008), and thus no conclusion could be drawn regarding its relation to type 1 diabetes risk. The *2250C* variant was observed in 15 affected sibpair families (minor allele frequency 0.006) and was preferentially transmitted to affected offspring in these families (21 transmitted vs. 7 nontransmitted, $P = 0.026$). The 2250C allele occurred exclusively on the H1 haplotype. When 2250G/C status was used to split this haplotype, the increased transmission of H1 to affected offspring noted above and in Table 3 was accounted for entirely by the 15 families in which the 2250C allele was segregating.

TABLE 2
Single-marker family-based association analysis of SNP markers in the *PTPN22* gene

SNP number	cDNA position	dbSNP number	Genomic position*	Allele†	MAF	FBAT	
						z‡	P‡
1	—	rs1235005	114129479	C/G	0.444	2.806	0.005015
2	—	rs2488457	114127410	G/C	0.240	3.157	0.001596
3	—	rs1217414	114124709	C/T	0.274	0.014	0.989227
4	—	rs1217418	114113273	C/T	0.463	2.482	0.013070
5	1858	rs2476601	114089610	C/T	0.126	4.155	0.000033
6	—	rs2797415	114089135	C/T	0.280	3.087	0.002023
7	—	rs3761935	114084570	A/C	0.182	−0.301	0.763082
8	2250	—	114084256	G/C	0.006	2.214	0.026857
9	—	rs1217388	114076518	T/C	0.276	3.009	0.002619
10	—	rs2476599	114075501	C/T	0.272	0.014	0.989156
11	3′ UTR	rs3811021	114068705	T/C	0.181	−0.299	0.765138

*Positions are according to NCBI (National Center for Biotechnology Information) Build 35, and the SNPs are listed from the 5′ end of *PTPN22* to the 3′ end; †common allele is listed first; ‡calculated for the minor allele. MAF, minor allele frequency; UTR, untranslated region.

TABLE 3
Family-based haplotype association analysis of *PTPN22*

Haplotype	SNP number										Frequency*	HBAT	
	1	2	3	4	5	6	7	9	10	11		*z*	*P*
H1	C	G	T	C	C	C	A	T	T	T	0.270	−0.253	0.80052
H2	C	G	C	C	C	C	A	T	C	T	0.264	−3.063	0.00219
H3	G	G	C	T	C	C	C	T	C	C	0.176	0.213	0.83110
H4	G	C	C	T	T	T	A	C	C	T	0.124	3.946	0.000079
H5	G	C	C	T	C	T	A	C	C	T	0.115	0.460	0.645415
H6	G	G	C	T	C	T	A	C	C	T	0.023	0.430	0.667337
H7	C	G	C	T	C	T	A	C	C	T	0.015	−0.186	0.852684

*Haplotypes with frequencies <0.01 were not listed.

To confirm that the 2250C allele was expressed, cDNA was synthesized from a subject heterozygous for 2250G/C and a control subject homozygous for the common 2250G allele. Sequencing of RT-PCR products spanning exons 18 to 21 from individuals heterozygous for 2250G/C revealed only the 2250G allele product. Position 2250 in *PTPN22* corresponds to the last nucleotide position of exon 18. Nucleotide usage at this position in mammalian exons is highly biased, with a G residue occurring in the −1 position in 78% of exons and C being present in <4% of mammalian exons (41). To determine whether the 2250G/C variant affected splicing of exon 18, primers in exons 15 and 20 were used to amplify *PTPN22* from cDNA. As shown in Fig. 1, the expected 393-nucleotide fragment was obtained from an individual who was homozygous for 2250G, whereas two fragments, one of the expected size and a shorter fragment, were obtained from a 2250G/C heterozygote. Nucleotide sequencing of these products revealed that the larger fragment corresponded to the correctly spliced *PTPN22* product containing the G-allele at position 2250, whereas the shorter fragment corresponded to a product in which exon 17 had been spliced directly to exon 19, with exon 18, containing position 2250, excluded. There was no evidence of the 2250C allele in any RT-PCR product generated in this analysis, suggesting that the presence of C at this position results in a high frequency of missplicing and deletion of exon 18. The result of this aberrant splicing event is the creation of a premature termination codon precisely at the exon 17-to-19 junction (Fig. 1).

Although resequencing in individuals selected for the absence of the established type 1 diabetes risk allele 1858T in *PTPN22* can potentially identify novel risk alleles occurring on other haplotypes, it leaves open the question of whether 1858T is the sole variant on the H4 haplotype responsible for its disease association. To address this question, 24 type 1 diabetic case subjects heterozygous for the 1858C/T variant were screened by DHPLC and sequencing to identify any additional coding variants occurring on the same haplotype. No other variants were detected on the H4 haplotype in these samples. An additional rare coding variant in exon 1 (77A/G, N26S) was identified on a chromosome carrying the 1858C allele in one of the heterozygous 1858C/T case samples screened. However, given that the 77A/G variant was observed only once in the 24 1858C/T heterozygous samples screened, along with the 94 1858C homozygous samples screened previously, and because Polyphen and SIFT analysis predicted the amino acid change was not likely to be damaging for the protein structure, the variant was not further analyzed in this study (39,40). An examination of the data derived from CEPH families in HapMap suggests that there is a haplotype block >300 kb in size spanning the *PTPN22* locus. Thus, the contribution of variants in either noncoding regions or adjacent genes cannot be ruled out as playing some role in the association between the H4 haplotype and type 1 diabetes. An examination of the genes flanking *PTPN22* on chromosome 1p13 revealed only one with any obvious potential to affect immune responses, C1orf178, which encodes *BFK*, a Bcl-2 family member with modest proapoptotic effects that is expressed in both spleen and thymus (42). The coding region of C1orf178 was screened by nucleotide sequencing in two type 1 diabetic case subjects homozygous for either the 1858T or 1858C allele. No coding variants were observed.

DISCUSSION

In the current study, a systematic approach was taken to elucidate whether the reported 1858T allele was the sole susceptibility allele in *PTPN22* or whether there were additional susceptibility alleles at this locus that can also

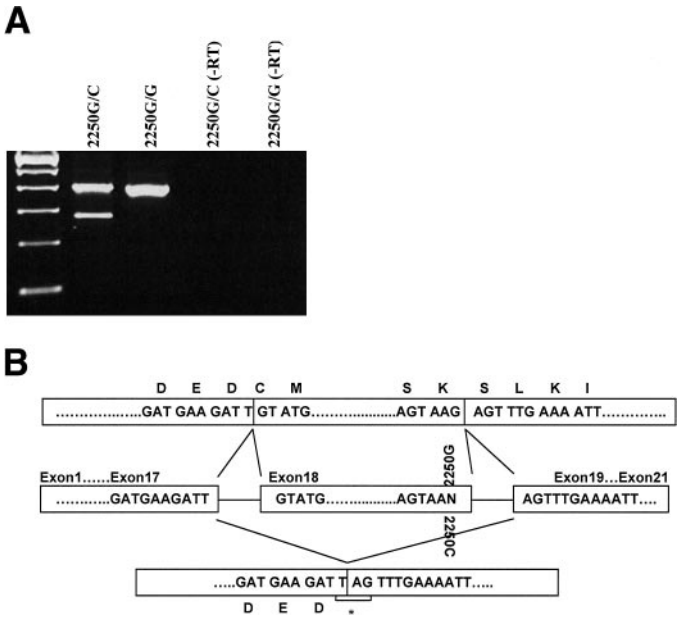

FIG. 1. The 2250C allele leads to missplicing and deletion of exon 18. *A*: cDNA amplification of exons 15–20. Genotypes of the samples amplified are listed above each lane. Lanes marked with "-RT" indicate negative controls in which no reverse transcriptase was added to the reaction. *B*: A schematic diagram of the effect of the coding SNP 2250G/C on splicing. Presence of the 2250C allele at the last position of exon 18 results in the skipping of exon 18. Exon 17 joins directly to exon 19, forming an immediate stop codon. The wild-type 2250G allele allows regular splicing of exon 18. *Immediate stop codon.

play a role in type 1 diabetes development. We considered two hypotheses. First, the possibility that there might exist additional type 1 diabetes risk variants within the *PTPN22* coding region was evaluated by screening all exons of the *PTPN22* gene for variation in type 1 diabetic case subjects homozygous for the wild-type 1858C allele. These results were extended by genotyping both the common SNPs identified in the screening and tag SNPs culled from the HapMap database in a collection of 374 multiplex type 1 diabetes–affected families to assess both single-marker and haplotype associations with disease. Among the common SNPs genotyped (i.e., those with minor allele frequencies $\geq$0.1), six displayed some evidence of association with type 1 diabetes ($P < 0.05$) when considered as single markers. As noted by Carlton et al. (43) in their study of rheumatoid arthritis patients, the 1858T allele was observed to occur on a single haplotype, designated here as H4. Our haplotype analysis revealed a second haplotype (H5) that is identical with the overtransmitted haplotype H4, except at 1858C/T. Because neither H5 nor any of the other remaining common haplotypes have displayed any evidence of positive association with type 1 diabetes, we conclude that the markers rs1235005, rs2488457, rs1217418, rs2797415, and rs1217388, which showed significant association to type 1 diabetes in the single-marker tests, are most likely not independent risk variants for type 1 diabetes.

As in the case of rheumatoid arthritis, conditioning on the haplotype containing the 1858T allele (H4) revealed modest evidence for excess transmission of two other haplotypes to affected offspring in type 1 diabetes–affected families. Because the SNPs used in the current study differ from those used by Carlton et al. (43) in their study of rheumatoid arthritis, it is not possible to make a direct comparison between the haplotypes. However, the excess transmission observed in both studies suggests the possibility that there might be additional variants in *PTPN22* that increase the risk of autoimmunity. In the current study, a plausible candidate on the H1 haplotype was identified in the rare variant 2250G>C. Although segregating in only 15 families, the 2250C allele was modestly associated with type 1 diabetes ($P = 0.026$), and, because RNA analysis indicated that it generates a premature termination codon, the variant allele might be expected to affect *PTPN22* function. However, given the infrequency of the 2250C allele, further association studies, both in type 1 diabetes and in other autoimmune diseases where a role for *PTPN22* variation has been previously demonstrated, are clearly necessary before a role for this variant in autoimmunity can be established.

The prematurely terminated coding sequence of the 2250C allele is predicted to encode a 711–amino acid protein lacking the last two COOH-terminal proline-rich repeats. The role of these two proline-rich repeats in Lyp is not well characterized, but they may play a role in binding to proteins with SH3 domains. If so, expression of such a truncated protein could potentially exert a dominant interfering effect on the wild-type allele product in heterozygotes. However, proteins containing premature termination mutations are often subject to rapid turnover and degradation, in which case the 2250C allele might only affect *PTPN22* function by reducing the overall cellular levels of the protein. Functional studies of 2250G/C heterozygotes will be necessary to resolve these two possibilities and to reconcile the effects of this variant with the reported gain of function observed with the 1858T allele (44).

In the first part of this study, we focused on identifying risk variants for type 1 diabetes within the *PTPN22* gene. Haplotype analysis identified the H4 haplotype, which carries the 1858T allele, as the major haplotype that contributes to type 1 diabetes susceptibility in *PTPN22*. However, these results leave open the possibility that additional variants might lie on this haplotype outside of the *PTPN22* gene that could affect type 1 diabetes risk. An examination of the CEU database in HapMap suggests that significant linkage disequilibrium may extend for several hundred kilobases surrounding *PTPN22*, potentially including other genes. When the functions or expression patterns of genes flanking *PTPN22* were considered, there were few obvious candidates for a gene that might impact autoimmunity. *RSBN1* is expressed exclusively in germ cells (45). *DCLRE1B* is expressed ubiquitously, but its function in the cellular response to DNA intrastrand cross-links does not suggest a role in immune responses (46). On the other hand, C1orf178 is a member of the Bcl-2 family with some proapoptotic function and is expressed in both spleen and thymus (42). To scan for additional coding variants that might be in linkage disequilibrium with the 1858T allele, the coding sequences of *PTPN22* and C1orf178 were screened in individuals that carried the H4 haplotype. The 1858T allele was the only coding variant identified in *PTPN22* and C1orf178 on the H4 haplotype.

In summary, a haplotype-based analysis of the role of the *PTPN22* locus in type 1 diabetes indicates that the 1858T (620W) allele occurs on a single haplotype that is overtransmitted to affected individuals in type 1 diabetes–affected families. No other coding variants were identified on this haplotype in either *PTPN22* or the flanking *C1orf178* gene. Within the *PTPN22* gene, four rare coding variants were identified, one of which also leads to a prematurely terminated protein. Although infrequent, this variant was associated with type 1 diabetes in a family-based analysis. Establishing whether this rare coding variant contributes to type 1 diabetes risk will require confirmation in a larger dataset and in functional studies to determine its impact on Lyp function.

ACKNOWLEDGMENTS

This work was supported by grants from the National Institutes of Health (DK46635) and the JDRF (Center for Translational Research at Benaroya Research Institute).

The authors thank Eric Olson for assistance with nucleotide sequencing and Sharon Teraoka for assistance with the WAVE platform. We acknowledge investigators and staff of the JDRF Center for Translational Research at Children's Hospital and Regional Medical Center in Seattle and the BRI Diabetes Clinical Research for subject recruitment. We also thank the Translational Research Clinical Core for sample processing and handling.

REFERENCES

1. Barnett AH, Eff C, Leslie RDG, Pyke DA: Diabetes in identical twins: a study of 200 pairs. *Diabetologia* 20:87–93, 1981
2. Verge CF, Gianani R, Yu L, Pietropaolo M, Smith T, Jackson RA, Soeldner JS, Eisenbarth GS: Late progression to diabetes and evidence for chronic β-cell autoimmunity in identical twins of patients with type 1 diabetes. *Diabetes* 44:1176–1179, 1995
3. Rich SS, Concannon P: Challenges and strategies for investigating the genetic complexity of common human diseases (Review). *Diabetes* 51 (Suppl. 3):S288–S294, 2002

4. Hirschhorn JN: Genetic epidemiology of type 1 diabetes. *Pediatr Diabetes* 4:87–100, 2003
5. Onengut-Gumuscu S, Concannon P: The genetics of type 1 diabetes: lessons learned and future challenges. *J Autoimmun* 25 (Suppl.):34–39, 2005
6. Thomson G, Robinson WP, Kuhner MK, Joe S, MacDonald MJ, Gottschall JL, Barbosa J, Rich SS, Bertrams J, Baur MP: Genetic heterogeneity, modes of inheritance, and risk estimates for a joint study of Caucasians with insulin-dependent diabetes mellitus. *Am J Hum Genet* 43:799–816, 1988
7. Bell GI, Horita S, Karam JH: A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. *Diabetes* 33:176–183, 1984
8. Lucassen AM, Julier C, Beressi JP, Boitard C, Froguel P, Lathrop M, Bell JI: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nat Genet* 4:305–310, 1993
9. Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, Concannon P, Gharani N, McCarthy MI, Olavesen MG, McCormack R, Guja C, Ionescu-Tirgoviste C, Undlien DE, Ronningen KS, Gillespie KM, Tuomilehto-Wolf E, Tuomilehto J, Bennett ST, Clayton DG, Cordell HJ, Todd JA: Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes* 53:1884–1889, 2004
10. Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, Bosi E, Larrad MT, Rios MS, Chow CC, Cockram CS, Jacobs K, Mijovic C, Bain SC, Barnett AH, Vandewalle CL, Schuit F, Gorus FK, Tosi R, Pozzilli P, Todd JA: The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Hum Mol Genet* 5:1075–1080, 1996
11. Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlman I, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Smink LJ, Lam AC, Cordell HJ, Walker NM, Bordin C, Hulme J, Motzo C, Cucca F, Hess JF, Metzker ML, Rogers J, Gregory S, Allahabadia A, Nithiyananthan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Maxwell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC: Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423:506–511, 2003
12. Concannon P, Erlich HA, Julier C, Morahan G, Nerup J, Pociot F, Todd JA, Rich SS: Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. *Diabetes* 54:2995–3001, 2005
13. Morel L, Blenman KR, Croker BP, Wakeland EK: The major murine systemic lupus erythematosus susceptibility locus, Sle1, is a cluster of functionally related genes. *Proc Natl Acad Sci U S A* 98:1787–1792, 2001
14. Hill NJ, Lyons PA, Armitage N, Todd JA, Wicker LS, Peterson LB: NOD Idd5 locus controls insulitis and diabetes and overlaps the orthologous CTLA4/IDDM12 and NRAMP1 loci in humans. *Diabetes* 49:1744–1747, 2000
15. Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D, Mustelin T: A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. *Nat Genet* 36:337–338, 2004
16. Smyth D, Cooper JD, Collins JE, Heward JM, Franklyn JA, Howson JM, Vella A, Nutland S, Rance HE, Maier L, Barratt BJ, Guja C, Ionescu-Tirgoviste C, Savage DA, Dunger DB, Widmer B, Strachan DP, Ring SM, Walker N, Clayton DG, Twells RC, Gough SC, Todd JA: 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, 2004
17. Onengut-Gumuscu S, Ewens KG, Spielman RS, Concannon P: A functional polymorphism (1858C/T) in the PTPN22 gene is linked and associated with type I diabetes in multiplex families. *Genes Immun* 5:678–680, 2004
18. Cohen S, Dadi H, Shaoul E, Sharfe N, Roifman CM: Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp. *Blood* 93:2013–2024, 1999
19. Cloutier JF, Veillette A: Association of inhibitory tyrosine protein kinase p50csk with protein tyrosine phosphatase PEP in T cells and other hemopoietic cells. *EMBO J* 15:4909–4918, 1996
20. Cloutier JF, Veillette A: Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. *J Exp Med* 189:111–121, 1999
21. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke JM, Conn MT, Chang M, Chang SY, Saiki RK, Catanese JJ, Leong DU, Garcia VE, McAllister LB, Jeffery DA, Lee AT, Batliwalla F, Remmers E, Criswell LA, Seldin MF, Kastner DL, Amos CI, Sninsky JJ, Gregersen PK: 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, 2004
22. Kyogoku C, Langefeld CD, Ortmann WA, Lee A, Selby S, Carlton VE, Chang M, Ramos P, Baechler EC, Batliwalla FM, Novitzke J, Williams AH, Gillett C, Rodine P, Graham RR, Ardlie KG, Gaffney PM, Moser KL, Petri M, Begovich AB, Gregersen PK, Behrens TW: Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet* 75:504–507, 2004
23. Begovich AB, Caillier SJ, Alexander HC, Penko JM, Hauser SL, Barcellos LF, Oksenberg JR: The R620W polymorphism of the protein tyrosine phosphatase PTPN22 is not associated with multiple sclerosis. *Am J Hum Genet* 76:184–187, 2005
24. Nistor I, Nair RP, Stuart P, Hiremagalore R, Thompson RA, Jenisch S, Weichenthal M, Abecasis GR, Qin ZS, Christophers E, Lim HW, Voorhees JJ, Elder JT: Protein tyrosine phosphatase gene PTPN22 polymorphism in psoriasis: lack of evidence for association. *J Invest Dermatol* 125:395–396, 2005
25. van Oene M, Wintle RF, Liu X, Yazdanpanah M, Gu X, Newman B, Kwan A, Johnson B, Owen J, Greer W, Mosher D, Maksymowych W, Keystone E, Rubin LA, Amos CI, Siminovitch KA: Association of the lymphoid tyrosine phosphatase R620W variant with rheumatoid arthritis, but not Crohn's disease, in Canadian populations. *Arthritis Rheum* 52:1993–1998, 2005
26. Ladner MB, Bottini N, Valdes AM, Noble JA: Association of the single nucleotide polymorphism C1858T of the PTPN22 gene with type 1 diabetes. *Hum Immunol* 66:60–64, 2005
27. Kahles H, Ramos-Lopez E, Lange B, Zwermann O, Reincke M, Badenhoop K: Sex-specific association of PTPN22 1858T with type 1 diabetes but not with Hashimoto's thyroiditis or Addison's disease in the German population. *Eur J Endocrinol* 153:895–899, 2005
28. Qu H, Tessier MC, Hudson TJ, Polychronakos C: Confirmation of the association of the R620W polymorphism in the protein tyrosine phosphatase PTPN22 with type 1 diabetes in a family based study. *J Med Genet* 42:266–270, 2005
29. Zheng W, She JX: Genetic association between a lymphoid tyrosine phosphatase (*PTPN22*) and type 1 diabetes. *Diabetes* 54:906–908, 2005
30. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, Moser KL, Begovich AB, Carlton VE, Li W, Lee AT, Ortmann W, Behrens TW, Gregersen PK: Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* 76:561–571, 2005
31. Cox NJ, Wapelhorst B, Morrison VA, Johnson L, Pinchuk L, Spielman RS, Todd JA, Concannon P: Seven regions of the genome show evidence of linkage to type 1 diabetes in a consensus analysis of 767 multiplex families. *Am J Hum Genet* 69:820–830, 2001
32. Abecasis GR, Cherny SS, Cookson WO, Cardon LR: Merlin-rapid analysis of dense genetic maps using sparse gene flow trees. *Nat Genet* 30:97–101, 2001
33. Rozen S, Skaletsky H: Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* 132:365–386, 2000
34. Barrett JC, Fry B, Maller J, Daly MJ: Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 21:263–265, 2005
35. Horvath S, Xu X, Laird NM: The family based association test method: strategies for studying general genotype–phenotype associations. *Eur J Hum Genet* 9:301–306, 2001
36. Horvath S, Xu X, Lake SL, Silverman EK, Weiss ST, Laird NM: Family-based tests for associating haplotypes with general phenotype data: application to asthma genetics. *Genet Epidemiol* 26:61–69, 2004
37. O'Connell JR, Weeks DE: PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 63:259–266, 1998
38. Lake SL, Blacker D, Laird NM: Family-based tests of association in the presence of linkage. *Am J Hum Genet* 67:1515–1525, 2000
39. Ramensky V, Bork P, Sunyaev S: Human non-synonymous SNPs: server and survey. *Nucleic Acid Res* 30:3894–3900, 2002
40. Ng PC, Henikoff S: Predicting deleterious amino acid substitutions. *Genome Res* 11:863–874, 2001
41. Shapiro MB, Senapathy P: RNA splice junctions of different classes of eukaryotes: sequence statistics and functional implications in gene expression. *Nucleic Acid Res* 15:7155–7174, 1987
42. Coultas L, Pellegrini M, Visvader JE, Lindeman GJ, Chen L, Adams JM, Huang DC, Strasser A: Bfk: a novel weakly proapoptotic member of the Bcl-2 protein family with a BH3 and a BH2 region. *Cell Death Differ* 10:185–192, 2003
43. Carlton VE, Hu X, Chokkalingam AP, Schrodi SJ, Brandon R, Alexander HC, Chang M, Catanese JJ, Leong DU, Ardlie KG, Kastner DL, Seldin MF,

Criswell LA, Gregersen PK, Beasley E, Thomson G, Amos CI, Begovich AB: PTPN22 genetic variation: evidence for multiple variants associated with rheumatoid arthritis. *Am J Hum Genet* 77:567–581, 2005

44. Vang T, Congia M, Macis MD, Musumeci L, Orru V, Zavattari P, Nika K, Tautz L, Tasken K, Cucca F, Mustelin T, Bottini N: Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet* 37:1317–1319, 2005
45. Takahashi T, Tanaka H, Iguchi N, Kitamura K, Chen Y, Maekawa M, Nishimura H, Ohta H, Miyagawa Y, Matsumiya K, Okuyama A, Nishimune Y: Rosbin: a novel homeobox-like protein gene expressed exclusively in round spermatids. *Biol Reprod* 70:1485–1492, 2004
46. Demuth I, Digweed M, Concannon P: Human SNM1B is required for normal cellular response to both DNA interstrand crosslink-inducing agents and ionizing radiation. *Oncogene* 23:8611–8618, 2004