# Further Evidence of a Primary, Causal Association of the *PTPN22* 620W Variant With Type 1 Diabetes

**Magdalena Zoledziewska,[1,2] Chiara Perra,[2†] Valeria Orrù,[1,3] Loredana Moi,[2] Paola Frongia,[4] Mauro Congia,[2] Nunzio Bottini,[3] and Francesco Cucca[1]**

**OBJECTIVE—**The minor allele of the nonsynonymous single nucleotide polymorphism (SNP) +1858C>T within the *PTPN22* gene is positively associated with type 1 diabetes and other autoimmune diseases. Genetic and functional data underline its causal effect, but some studies suggest that this polymorphism does not entirely explain disease association of the *PTPN22* region. The aim of this study was to evaluate type 1 diabetes association within this gene in the Sardinian population.

**RESEARCH DESIGN AND METHODS—**We resequenced the exons and potentially relevant portions of *PTPN22* and detected 24 polymorphisms (23 SNPs and 1 deletion insertion polymorphism [DIP]), 8 of which were novel. A representative set of 14 SNPs and the DIP were sequentially genotyped and assessed for disease association in 794 families, 490 sporadic patients, and 721 matched control subjects.

**RESULTS—**The +1858C>T variant, albeit rare in the general Sardinian population (allele frequency 0.014), was positively associated with type 1 diabetes ($P_{\text{one tail}} = 3.7 \times 10^{-3}$). In contrast, the background haplotype in which this mutation occurred was common (haplotype frequency 0.117) and neutrally associated with disease. We did not confirm disease associations reported in other populations for non +1858C>T variants (rs2488457, rs1310182, and rs3811021), although they were present in appreciable frequencies in Sardinia. Additional weak disease associations with rare variants were detected in the Sardinian families but not confirmed in independent case-control sample sets and are most likely spurious.

**CONCLUSIONS—**We provide further evidence that the +1858C>T polymorphism is primarily associated with type 1 diabetes and exclude major contributions from other purportedly relevant variants within this gene. ***Diabetes* 57:229–234, 2008**

A large body of experimental evidence suggests that most cases of type 1 diabetes result from the autoimmune attack of T-cells on insulin-producing pancreatic β-cells. The lifetime disease risk for a monozygotic twin of an affected patient is ~50% (1,2) and shows a quick fall-off with decreased genetic relatedness (3). These data and the increasing incidence of disease reported in most European and European-derived populations over the last few decades (4) suggest that the chance of occurrence of this autoimmune process depends on the complex interplay between a polygenic trait and unknown environmental factors influencing penetrance of susceptibility genes.

The major histocompatibility complex/human leukocyte antigen (HLA) region on chromosome 6p21 contains the major component of disease-inherited risk (5,6). A second disease locus has been located (7) and fine mapped on chromosome 11p15.5 to a minisatellite (variable number of tandem repeats) locus in the insulin gene promoter region (8). Over the last few years, several new non-HLA, non-*INS* genes and variants have been found to be associated with type 1 diabetes (9,10). Among these, a +1858C>T polymorphism, localized in exon 14 of the *PTPN22* gene (1p13.3), shows consistent remarkable size effects for type 1 diabetes susceptibility (odds ratio [OR] constantly >1.5 in all populations in which it is present at appreciable frequencies) (9,11,12). This variant is also positively associated with some other autoimmune diseases, including rheumatoid arthritis (13), Graves' disease (14), and systemic lupus erythematosus (15), but not with other autoimmune disorders such as multiple sclerosis, celiac disease, Crohn's disease, and psoriasis (16,17).

Although the reasons for the different association with distinct autoimmune traits are still unclear, the available data pose the +1858C>T variant as one of the few non-HLA polymorphisms with relevant size effects in the inherited risk for type 1 diabetes. This variant causes an amino acid change from arginine (R) to tryptophan (W) in codon 620 of the lymphoid protein tyrosine phosphatase (Lyp), which has a critical negative role on T-cell activation (18). Experimental evidence suggests that the Lyp-W620 variant causes a gain of physiological function of the wild-type phosphatase; that is, $W^{620}$ has more negative regulatory activity than the more common wild-type $R^{620}$ allele (19,20).

The primary role of the +1858C>T variant in type 1 diabetes is also supported by a detailed haplotype analysis in a familial type 1 diabetes sample set of northern European origin (21). However, this study described some preliminary evidence of disease association of the very rare (minor allele frequency [MAF] = 0.006) +2250G>C, K750N variant. This variant was independent of +1858C>T

From the [1]Dipartimento di Scienze Biomediche, University of Sassari, Sassari, Italy; the [2]Laboratorio di Immunogenetica, Ospedale Microcitemico, Cagliari, Italy; the [3]Institute for Genetic Medicine, University of Southern California, Los Angeles, California; and the [4]Divisione Pediatrica, Ospedale Brotzu, Cagliari, Italy.

Address correspondence and reprint requests to Francesco Cucca, Cattedra di Genetica Medica, Dipartimento di Scienze Biomediche, Università di Sassari, Viale S. Pietro, 07100 Sassari, Italy. E-mail: fcucca@uniss.it.

Received for publication 1 March 2007 and accepted in revised form 7 October 2007.

Published ahead of print at http://diabetes.diabetesjournals.org on 12 October 2007. DOI: 10.2337/db07-0289.

M.Z., C.P., and V.O. contributed equally to this work.

†C.P. is deceased.

AFBAC, affected family–based control; DIP, deletion insertion polymorphism; HLA, human leukocyte antigen; Lyp, lymphoid protein tyrosine phosphatase; MAF, minor allele frequency; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test.

and resulted in an RNA product in which exon 17 had been spliced directly to exon 19, with exon 18 containing position 2250 excluded (21).

Furthermore, another study did not detect the +1858C>T variant in large Asian sample sets and suggested that another polymorphism located in the promoter region (rs2488457) was associated with the classical form of type 1 diabetes in both Asian- and European-derived family sample sets (22). After comparing the strength of disease association of the +1858C>T and rs2488457 single nucleotide polymorphisms (SNPs) in a European collection of families, the authors concluded that disease association was stronger at rs2488457 and suggested that this SNP was a more likely causative variant within the *PTPN22* gene (22). Finally, after a comprehensive resequencing of the gene and conditional association analysis, Carlton et al. (23) also suggested that +1858C>T by itself is unlikely to account for all association observed between *PTPN22* and rheumatoid arthritis. Three SNPs, located on intron 16 (ss38346943), in the 3′-untranslated (rs3811021) region and 1,496 bases downstream of *PTPN22* at the 5′ end of the nearby *RSBN1* gene (rs3789604), were associated with rheumatoid arthritis independently of +1858C>T (23).

Previous work found the +1858C>T mutation to be relatively rare in the general Sardinian population (9). This raises the possibility that other mutations in *PTPN22* could be mediating type 1 diabetes risk in Sardinia. Here, we report a fine-mapping association analysis of the *PTPN22* gene in type 1 diabetes sample sets from Sardinia. The aim of the study was to clarify the results of previous studies and determine whether additional, non +1858C>T variants are independent contributors to the inherited risk of type 1 diabetes in this island population.

## RESEARCH DESIGN AND METHODS

We used a primary sample set consisting of 794 type 1 diabetic families, all of Sardinian origin. These families included 693 simplex families with one affected child and 101 multiplex families with at least two affected children. The average age of the patients at disease onset was 8.8 ± 5.5 years (±SD) with a range of 0.5–41 years. We also used an additional sample set of 490 Sardinian sporadic patients and 721 ethnically matched control subjects. These control subjects were healthy blood donor volunteers. The study was approved by the ethic committees of the Universities of Cagliari and Sassari. All participating individuals and their parents or legal guardians signed a statement of informed consent.

**Resequencing.** A commercial kit from Applied Biosystems was used to resequence the coding segments of the *PTPN22* gene in DNA samples from 64 parents of type 1 diabetic patients. Thirty-two of these parents were randomly chosen, while 32 were selected after an initial step of genotyping and haplotype analysis of the +1858C>T SNP (rs2476001) and four additional markers (SNPs rs1235005 and rs2476599, microsatellites (GGAA)n repeat, contig position 10479010–10478941 and (CA)n repeat, contig position 10428350–10428304) in 631 type 1 diabetic families. More specifically, these latter 32 patients were chosen to be homozygous for the +1858C wild-type variant and heterozygous for any of two haplotypes, defined by the four markers reported above, which showed some marginal evidence of positive transmission to affected children from heterozygous parents (data not shown). This resequencing strategy, on one hand, allowed us to obtain an unbiased representation of the common variants present in the Sardinian population and at the same time increased the chance of detecting non +1858C>T disease susceptibility variants. PCRs, purifications, and sequencing reactions (BigDye Terminators v1.3) were performed according to the Applied Biosystems resequencing protocol. The sequences were run on an ABI PRISM 3100 Sequencing Analyzer. Electropherograms were analyzed using SeqScape v2.1.1 (Applied Biosystems).

**Genotyping.** Of the 24 variants detected during the resequencing stage, 13 potentially informative SNPs (rs2488457, rs1217419, rs1217418, rs3789609, rs3789608, rs2476602, ss38346945, rs1970559, rs2797415, rs3761935, ss38346943, rs1217412, and rs3811021) and one 4-base deletion insertion polymorphism (DIP) were initially chosen. Selection criteria depended on their position within *PTPN22*, to cover the exons and the promoter region and on their frequency in the resequenced samples. More specifically, polymorphisms that were seen more than four times in the 128 resequenced chromosomes were selected to be genotyped in families. Three of the selected markers (rs3879609, rs2797415, and rs1217412) were found to be problematic to genotype because of apparent design problems (rs1217412) or because of an anomalous pattern of the plot with real time PCR (rs3879609) or because they provided genotypes of marginal quality (rs2797415), and thus they were removed from the analysis. However, in public databases, variants rs1217412 and rs2797415 were found in the same haplotype block ($r^2 > 0.8$) of variant rs2488457, which was successfully genotyped. Furthermore, three additional SNPs (rs1235005, rs2476601, and rs2476599) were typed before the resequencing stage, and one supplementary variant (rs1310182) of prior interest (23) that was not included in the segments of the gene we resequenced was also genotyped. Overall, 15 variants (14 SNPs and a 4-bp DIP) were selected and successfully genotyped. All 14 SNPs were genotyped using TaqMan MGB probes ABI PRISM 7000 System (Applied Biosystems), whereas the 4-bp DIP was genotyped using the MegaBace 1000 Sequencing Analyzer (Amersham Biosciences).

Partial results of the transmission disequilibrium test (TDT) analysis for the +1858C>T variant in a subset of the type 1 diabetic families have been reported elsewhere (20).

**Statistical analysis.** Single-point disease association of individual variants was evaluated using the TDT and monitoring transmission only to one affected individual, the proband, per family to ensure statistical independence of the data (24). To compute the frequency of assessed variants in the general Sardinian population, we computed the affected family–based control (AFBAC) for each allele and haplotype, as described by Thomson (25). AFBAC frequencies are based on the alleles and haplotypes that are never transmitted from the parents to affected children and provide allelic frequencies comparable with those detected at the general population level in the absence of population stratification (26). The statistical power of our sample set has been computed according to Knapp (27).

Parental haplotypes were reconstructed using the best option of the Merlin package, version 1.1-α3 (28). The basis of the algorithm for haplotype reconstruction is from the software SNPHAP by David Clayton (http://www-gene.cimr.cam.ac.uk:/clayton/software/). The statistical significance of variation in haplotype frequency between AFBACs and affected individuals was performed using permutation testing implemented in the Haplo.Stats package (29). In families with more than one affected child, only the proband was analyzed.

To determine the genetic architecture of the *PTPN22* gene in the Sardinian population, the linkage disequilibrium measure $r^2$ was calculated on parental genotypes between every pair of markers (30,31). The measure was built in a standard fashion from the basic pairwise-disequilibrium coefficient, *D*, computed using the HAPLOXT program in the GOLD software package (www.s-ph.umich.edu/csg/abecasis/GOLD). Intermarker pairwise $r^2$ ranges from 0 to 1, with a value of 0 reflecting full independence between alleles at the two loci compared and a value of one occurring only when the alleles at the two marker loci always coincide.

## RESULTS

Twenty-three SNPs and one DIP in the promoter region were detected after resequencing of the exons, the exon-intron boundary, and the region upstream and downstream of the gene in 64 parents of type 1 diabetic patients (Table 1). Eight of these variants (and the DIP) are still not present in public databases.

Fifteen markers spanning a region of 60,774 bp of the *PTPN22* gene region were initially genotyped in 694 type 1 diabetic families. These 15 variants included 11 of the polymorphisms detected with the resequencing of the gene (rs2488457, ACTC DIP, rs1217419, rs1217418, rs3789608, rs2476602, ss38346945, rs1970559, rs3761935, ss38346943, and rs3811021), three additional SNPs that were genotyped before the resequencing stage (rs1235005, rs2476601, and rs2476599), and one variant (rs1310182) of prior interest (23) not included in the segments of the gene we resequenced. TDT analysis was performed for the 15 genotyped markers, and 4 of them were found to be positively associated with type 1 diabetes at a 5% level of significance (Table 2). These included the established

TABLE 1
Allelic frequencies of *PTPN22* variants detected after resequencing 64 parents of type 1 diabetic patients

| Position | Polymorphism | Localization | MAF | |
|---|---|---|---|---|
| | | | 32 selected parents | 32 random parents |
| C10501483G | rs2488457 | 5′ flanking | 0.12 | 0.16 |
| C10501426T | Unknown | 5′ flanking | 0.00 | 0.02 |
| 10500885–10500882 | ACTC DIP | 5′ flanking | 0.17 | 0.45 |
| A10488019C | rs1217419 | IVS2 | 0.2 | 0.38 |
| G10487819A | Unknown | IVS2 | 0.00 | 0.02 |
| T10487346C | rs1217418 | IVS3 | 0.2 | 0.43 |
| G10485405A | Unknown | IVS5 | 0.05 | 0.00 |
| A10485278G | Unknown | IVS6 | 0.03 | 0.00 |
| A10483914G | rs3789609 | IVS7 | 0.17 | 0.52 |
| A10483903G | rs3789608 | IVS7 | 0.05 | 0.09 |
| C10483377T | Unknown | IVS8 | 0.02 | 0.00 |
| A10483096T | ss38346946 | IVS9 | 0.05 | 0.00 |
| T10483070C | rs2476602 | IVS9 | 0.64 | 0.12 |
| A10480804G | ss38346945 | Exon 10 R263Q | 0.09 | 0.02 |
| A10467503G | Unknown | IVS11 | 0.02 | 0.00 |
| C10466692T | Unknown | Exon 13 V482A | 0.00 | 0.02 |
| G10463263A | rs1970559 | IVS14 | 0.36 | 0.12 |
| T10463208C | rs2797415 | IVS14 | 0.16 | 0.23 |
| C10458643A | rs3761935 | IVS17 | 0.05 | 0.16 |
| C10458639T | ss38346943 | IVS17 | 0.12 | 0.02 |
| A10448233T | ss38346942 | IVS20 | 0.05 | 0.02 |
| T10448107A | Unknown | IVS20 | 0.02 | 0.00 |
| C10443326T | rs1217412 | 3′ flanking | 0.16 | 0.23 |
| C10442778T | rs3811021 | 3′ flanking | 0.05 | 0.14 |

Positions are according to genomic contig NT_019273. Polymorphisms are oriented according to transcript NM_012411, which is the reverse complement of the genomic contig sequence. Polymorphisms are listed from the 5′ end to the 3′ end of the gene. The names of known polymorphisms are according to the dbSNP. Polymorphisms not reported in public databases are described as unknown. The last two columns report the MAF detected in 64 parents of the patients selected as described in RESEARCH DESIGN AND METHODS.

+1858C>T polymorphism in exon 14 (marker 9 on Table 2) and three additional variants: allele A of ss38346945 (marker 8 on Table 2), allele C of ss38346943 (marker 13 on Table 2), and allele T of rs2476599 (marker 14 on Table 2), localized in exon 10 (+789G>A), in intron 17 (IVS17 +47T>C), and in intron 19 (IVS19 −1183C>T), respectively. The rs2488457/−1123G SNP (marker 2 on Table 2), found to be associated with type 1 diabetes in previous studies (21,22), is present and common in the Sardinian population (AFBAC frequency 0.128). Despite having >78% power to detect a primary association even as weak as an OR = 1.3, at a significance level of 5%, there was no evidence of association in our sample set (Table 2). Similarly, allele T of rs1310182 and allele C of rs3811021 (markers 11 and 15 on Table 2), which were found to be associated with rheumatoid arthritis in another study (23), were not associated with type 1 diabetes in our sample set (Table 2), although the powers to detect association at $P = 0.05$ assuming an OR ≥ 1.3 were 96 and 80%, respectively, for the two variants.

TABLE 2
Association analysis of 15 variants with type 1 diabetes using the TDT in 694 Sardinian families

| Marker | Polymorphism | Position | Minor allele | Affected children | | | | AFBAC frequency (%) |
|---|---|---|---|---|---|---|---|---|
| | | | | T | NT | T (%) | *P* value TDT | |
| 1 | rs1235005 | C10503552G | G | 249 | 274 | 47.6 | $2.7 \times 10^{-01}$ | 38.1 |
| 2 | rs2488457 | C10501483G | C | 140 | 138 | 50.4 | $9.0 \times 10^{-01}$ | 12.8 |
| 3 | ACTC DIP | 10500885–10500882 | long | 199 | 191 | 51.1 | $6.9 \times 10^{-01}$ | 39.0 |
| 4 | rs1217419 | A10488019C | A | 260 | 285 | 47.7 | $2.8 \times 10^{-01}$ | 38.4 |
| 5 | rs1217418 | T10487346C | T | 256 | 280 | 47.8 | $3.0 \times 10^{-01}$ | 38.5 |
| 6 | rs3789608 | A10483903G | A | 144 | 143 | 50.2 | $9.5 \times 10^{-01}$ | 14.8 |
| 7 | rs2476602 | T10483070C | T | 185 | 151 | 55.1 | $6.4 \times 10^{-02}$ | 18.1 |
| 8 | ss38346945 | A10480804G | A | 38 | 21 | 64.4 | $2.7 \times 10^{-02}$ | 1.8 |
| 9 | rs2476601 | T10463683C | T | 34 | 17 | 66.7 | $1.7 \times 10^{-02}$ | 1.3 |
| 10 | rs1970559 | G10463263A | G | 188 | 153 | 55.1 | $5.8 \times 10^{-02}$ | 18.3 |
| 11 | rs1310182 | T10459618C | T | 243 | 266 | 47.7 | $3.1 \times 10^{-01}$ | 38.6 |
| 12 | rs3761935 | C10458643A | C | 126 | 125 | 50.2 | $9.5 \times 10^{-01}$ | 13.4 |
| 13 | ss38346943 | C10458639T | C | 49 | 25 | 66.2 | $5.2 \times 10^{-03}$ | 2.1 |
| 14 | rs2476599 | T10449574C | T | 193 | 154 | 55.6 | $3.6 \times 10^{-02}$ | 18.2 |
| 15 | rs3811021 | C10442778T | C | 135 | 132 | 50.6 | $8.5 \times 10^{-01}$ | 13.9 |

Marker 9 is the canonically associated +1856C>T polymorphism. AFBAC frequency, allele frequency in AFBAC; NT, nontransmitted; T, transmitted.

TABLE 3
Association analysis of four variants with type 1 diabetes using the TDT in 794 Sardinian families

| Marker | Polymorphism | Position | Minor allele | Affected children | | | |
|---|---|---|---|---|---|---|---|
| | | | | T | NT | T (%) | *P* value TDT |
| 8 | ss38346945 | A10480804G | A | 42 | 22 | 65.6 | $1.2 \times 10^{-02}$ |
| 9 | rs2476601 | T10463683C | T | 38 | 18 | 67.8 | $7.5 \times 10^{-03}$ |
| 13 | ss38346943 | C10458639T | C | 53 | 27 | 66.2 | $3.6 \times 10^{-03}$ |
| 14 | rs2476599 | T10449574C | T | 220 | 176 | 55.6 | $2.7 \times 10^{-02}$ |

Marker 9 is the +1856C>T polymorphism. The numbers of the markers in column 1 are as those reported in Table 2. NT, nontransmitted; T, transmitted.

We then genotyped the four variants showing evidence of association in our survey study in the Sardinian population in an additional 100 type 1 diabetic families (380 individuals) collected during the course of the study. Evidence of association at the 5% level was still detected for all four alleles in this enlarged sample set (Table 3).

We next performed a haplotype analysis in all 15 variants genotyped in families. Considering haplotypes with a frequency >0.008 in the parental chromosomes, we detected seven major haplotypes accounting for >95% of haplotypes present in the familial sample set. Allele T of marker 9 (+1858C>T) was contained in just one haplotype (haplotype 6 in Table 4), which was positively associated with disease ($P = 0.036$). The background wild-type haplotype (haplotype 1 in Table 4) in which the +1858C>T mutation apparently occurred was common in this Sardinian sample set (AFBAC frequency 0.117) and was neutrally associated with type 1 diabetes. Finally, also a haplotype containing markers 8, 13, and 14 (haplotype 7 in Table 4) showed some degree of significant association with type 1 diabetes in this familial sample set ($P = 0.016$). Allele T of marker 14 was only associated with disease when present together with allele A of marker 8 and allele C of marker 13 on the same haplotype.

Genetic relationships between the variants at these 15 markers are also illustrated by the pairwise linkage disequilibrium of their MAFs measured by $r^2$ values (Fig. 1). The critical allele T of marker 9 was independent from all other genotyped variants. Moreover, allele A of marker 8 and allele C of marker 13 on the one hand, and allele T of marker 7, allele G of marker 10 and allele T of marker 14 on the other hand, constituted distinct haplotype blocks ($r^2 > 0.8$). The same patterns of linkage disequilibrium were also detected on samples of northern European origin (23).

To further investigate the putative novel evidence of disease association, we then assessed allele A of marker 8 and allele C of marker 13 in an additional collection of sporadic patients and matched control subjects from the same population. However, in this independent sample set, these markers showed nearly overlapping frequencies in patients and in control subjects; allele A of marker 8 was present in 27 of 962 fully genotyped patients' chromosomes (0.028) and in 39 of 1,442 control subjects' chromosomes (0.027), whereas allele C of marker 13 was counted in 36 of 976 patients' chromosomes (0.036) and in 54 of 1,442 control subjects' chromosomes (0.037).

## DISCUSSION

Using a collection of type 1 diabetic families from Sardinia, we provide further evidence for a primary association with type 1 diabetes of the +1858C>T polymorphism within the *PTPN22* gene. This is underlined by the fact that the background haplotype in which the +1858C>T mutation occurred is retained and still relatively common in the Sardinian population but neutrally associated with type 1 diabetes. The data are in full agreement with evidence from type 1 diabetic families of the Human Biological Data Interchange and the British Diabetes Association repositories (21). These results are also consistent with a reported primary association of +1858C>T with rheumatoid arthritis in two northern European–derived case-control sample sets from the U.S. (23). Primary association of the +1858T with type 1 diabetes is further supported by experimental data indicating the relevant functional consequences of this polymorphism at the protein level (19,20). Conversely, our observations are at odds with the results of a study suggesting that association of the +1858T was in fact secondary to association of a polymorphism, rs2488457 (marker 2 on Table 2), located in the promoter region of the gene and frequent in both the Asian

TABLE 4
Association of *PTPN22* haplotypes with type 1 diabetes

| H* | Markers† | | | | | | | | | | | | | | | Frequency | | *P* value |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | Patients | AFBAC | |
| 1 | G | C | Short | C | T | G | C | G | C | A | T | A | T | C | T | 0.117 | 0.102 | $1.2 \times 10^{-01}$ |
| 2 | G | G | Short | C | T | G | C | G | C | A | T | A | T | C | T | 0.077 | 0.071 | $4.1 \times 10^{-01}$ |
| 3 | C | G | Long | A | C | G | C | G | C | A | C | A | T | C | T | 0.393 | 0.389 | $7.3 \times 10^{-01}$ |
| 4 | G | G | Short | C | T | A | C | G | C | A | T | C | T | C | C | 0.149 | 0.148 | $9.0 \times 10^{-01}$ |
| 5 | C | G | Short | A | C | G | T | G | C | G | C | A | T | T | T | 0.166 | 0.184 | $1.1 \times 10^{-01}$ |
| 6 | G | C | Short | C | T | G | C | G | T | A | T | A | T | C | T | 0.020 | 0.029 | $3.0 \times 10^{-02}$ |
| 7 | C | G | Short | A | C | G | T | A | C | G | C | A | C | T | T | 0.026 | 0.038 | $1.5 \times 10^{-02}$ |

*H refers to haplotypes; haplotypes with a frequency <0.008 in the total sample set are not listed. †The marker number corresponds to the numbers given in Table 2. Marker 9 is the +1856C>T polymorphism. Frequency refers to haplotypes counted in patients and in pseudocontrols (AFBAC), and *P* value is statistical significance of the frequency difference between the two classes as measured by the permutation test implemented in the Haplo-Stats program.

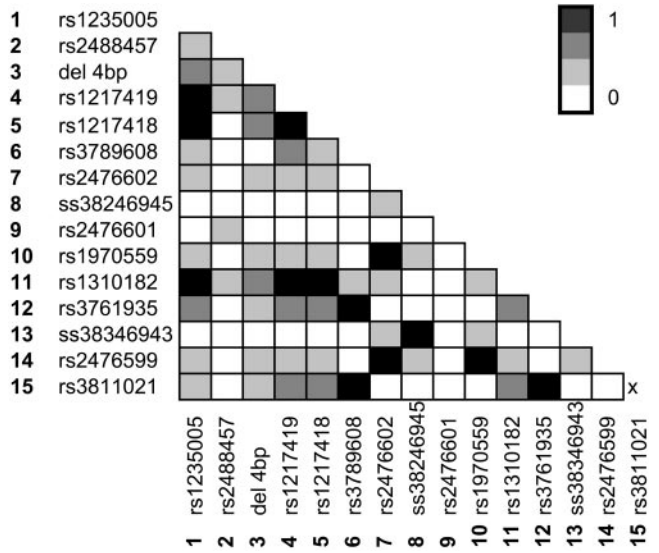


**FIG. 1. Comparison of the strength and extent of linkage disequilibrium in the *PTPN22* gene by means of distribution of pairwise linkage disequilibrium between 15 SNPs measured by $r^2$ on the parental chromosomes from Sardinian type 1 diabetic families. Marker 9 is the canonically associated +1856C>T polymorphism.**

and northern European populations (22). This latter variant was present and common in our Sardinian sample set but was not associated with type 1 diabetes. Note that in contrast with some other populations, the distribution of alleles +1858C>T and rs2488457 is informative in Sardinia, most likely owing to their distinct distribution on positively and neutrally associated haplotypes, respectively. Similar conclusions about the role of rs2488457 were recently reached by Chelala et al. (32) after analyzing a pooled French, Danish, and U.S. sample set.

Although Carlton et al. (23) suggested an association of rheumatoid arthritis with polymorphisms rs1310182, rs3811021, and rs3789604, with the latter two alleles in nearly complete linkage disequilibrium with each other, when we genotyped two of these variants, rs1310182 (marker 11) and rs3811021 (marker 15), in the Sardinian families, we found no evidence of association with type 1 diabetes. This result is consistent with other recent studies concerning the association of these variants with rheumatoid arthritis and Graves' disease (33,34). The recently described, positively associated rare variant +2250C causing the amino acid change K750N and alternative splicing with exclusion of exon 18 of *PTPN22* (21) was not detected in the Sardinian samples resequenced and hence was not genotyped in the enlarged sample set. We did, however, obtain some initial evidence that variants within the *PTPN22* gene, other than +1858C, were positively associated with type 1 diabetes independently of the +1858C>T variant. These included the minor alleles of ss38346945 (marker 8), which causes an amino acid change from arginine to glutamine in codon 263 of Lyp (R263Q) and of ss38346943 (marker 13), in strong linkage disequilibrium with each other. Nevertheless, these novel associations were borderline significant after correction for number of tests performed and were not confirmed by further analyses in independent case-control sample sets from Sardinia and from the U.K. (J.A. Todd and D. Smyth, personal communication). Furthermore, these variants did not show any significant evidence of association with rheumatoid arthritis in a previous study (23). It appears likely, therefore, that these inconsistent additional disease associations with non +1858C>T *PTPN22* variants may in fact be spurious and related to stochastic fluctuations in the frequencies or in the transmission of variants to affected children. The role of the +2250G>C (K750N) variant remains to be evaluated in larger sample sets from other populations (22).

We can conclude that a comprehensive analysis in a representative sample set from Sardinia provides further evidence that the +1858C>T is primarily associated with type 1 diabetes, even in a population in which this variant is rare. Furthermore, we excluded a major role for other potentially relevant variants within this gene. These results underline the need of the systematic replication of any evidence of disease association in sample sets offering adequate statistical power.

## ACKNOWLEDGMENTS

N.B. has received Juvenile Diabetes Research Foundation (JDRF) Grant 1-2005-342. F.C. has received Telethon-JDRF Grant GJT030477 and Ministero dell'Università e della Ricerca (MIUR) Grant 2005060193 from the Italian Ministry of Scientific Research.

We thank John Todd and Debbie Smyth for sharing unpublished data with us, Michael Whalen and John Todd for useful suggestions, Patrizia Zavattari for technical advice, and Rebecca Lewis for editorial assistance in the preparation of the manuscript.

This study is dedicated to the memory of Chiara Perra.

## REFERENCES

1. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J: Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. *Diabetes* 52:1052–1055, 2003
2. Redondo MJ, Yu L, Hawa M, Mackenzie T, Pyke DA, Eisenbarth GS, Leslie RD: Heterogeneity of type I diabetes: analysis of monozygotic twins in Great Britain and the United States. *Diabetologia* 44:354–362, 2001
3. Harjutsalo V, Podar T, Tuomilehto J: Cumulative incidence of type 1 diabetes in 10,168 siblings of Finnish young-onset type 1 diabetic patients. *Diabetes* 54:563–569, 2005
4. Dahlquist G: Can we slow the rising incidence of childhood-onset autoimmune diabetes? The overload hypothesis. *Diabetologia* 49:20–24, 2006
5. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA: The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet* 59:1134–1148, 1996
6. Cucca F, Lampis R, Congia M, Angius E, Nutland S, Bain SC, Barnett AH, Todd JA: A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum Mol Genet* 10:2025–2037, 2001
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. Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, Concannon P, Gharani N, McCarthy MI, Olavasen MG, McCormack R, Guja C, Ionescu-Tirgoviste C, Undlien D, Ronningen K, Gillespie KM, Tuomilehto-Wolf E, Tuomilehto J, Bennett ST, Clayton D, Cordell H, Todd JA: Remapping the insulin gene/IDDM2 locus in type 1 diabetes. *Diabetes* 53:1884–1889, 2004
9. 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
10. Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszko JS, Hafler JP, Zeitels L, Yang JH, Vella A, Nutland S, Stevens HE, Schuilenburg H, Coleman G, Maisuria M, Meadows W, Smink LJ, Healy B, Burren OS, Lam AA, Ovington NR, Allen J, Adlem E, Leung HT, Wallace C, Howson JM, Guja C, Ionescu-Tirgoviste C, Simmonds MJ, Heward JM, Gough SC, Dunger DB, Wicker LS, Clayton DG: Robust associations of four new chromosome

regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 39:857–864, 2007
11. 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
12. 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
13. 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
14. Velaga MR, Wilson V, Jennings CE, Owen CJ, Herington S, Donaldson PT, Ball SG, James RA, Quinton R, Perros P, Pearce SH: The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. *J Clin Endocrinol Metab* 89:5862–5865, 2004
15. 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
16. 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
17. Rueda B, Nunez C, Orozco G, Lopez-Nevot MA, de la Concha EG, Martin J, Urcelay E: C1858T functional variant of PTPN22 gene is not associated with celiac disease genetic predisposition. *Hum Immunol* 66:848–852, 2005
18. Bottini N, Vang T, Cucca F, Mustelin T: Role of PTPN22 in type 1 diabetes and other autoimmune diseases. *Semin Immunol* 18:207–213, 2006
19. Rieck M, Arechiga A, Onengut-Gumuscu S, Greenbaum C, Concannon P, Buckner JH: Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. *J Immunol* 179:4704–4710, 2007
20. 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
21. Onengut-Gumuscu S, Buckner JH, Concannon P: A haplotype-based analysis of the PTPN22 locus in type 1 diabetes. *Diabetes* 55:2883–2889, 2006
22. Kawasaki E, Awata T, Ikegami H, Kobayashi T, Maruyama T, Nakanishi K, Shimada A, Uga M, Kurihara S, Kawabata Y, Tanaka S, Kanazawa Y, Lee I, Eguchi K: Systematic search for single nucleotide polymorphisms in a lymphoid tyrosine phosphatase gene (PTPN22): association between a promoter polymorphism and type 1 diabetes in Asian populations. *Am J Med Genet A* 140:586–593, 2006
23. 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
24. Spielman RS, McGinnis RE, Ewens WJ: Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am J Hum Genet* 52:506–516, 1993
25. Thomson G: Mapping disease genes: family-based association studies. *Am J Hum Genet* 57:487–498, 1995
26. Lampis R, Morelli L, Congia M, Macis MD, Mulargia A, Loddo M, De Virgiliis S, Marrosu MG, Todd JA, Cucca F: The inter-regional distribution of HLA class II haplotypes indicates the suitability of the Sardinian population for case-control association studies in complex diseases. *Hum Mol Genet* 9:2959–2965, 2000
27. Knapp M: A note on power approximations for the transmission/disequilibrium test. *Am J Hum Genet* 64:1177–1185, 1999
28. 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, 2002
29. Schaid DJ, Rowland CM, Tines DE, Jacobson RM, Poland GA: Score tests for association between traits and haplotypes when linkage phase is ambiguous. *Am J Hum Genet* 70:425–434, 2002
30. Franklin I, Lewontin RC: Is the gene the unit of selection? *Genetics* 65:707–734, 1970
31. Hill WG, Robertson A: Linkage disequilibrium in finite populations. *Theor Appl Genet* 38:226–231, 1968
32. Chelala C, Duchatelet S, Joffret ML, Bergholdt R, Dubois-Laforgue D, Ghandil P, Pociot F, Caillat-Zucman S, Timsit J, Julier C: PTPN22 R620W functional variant in type 1 diabetes and autoimmunity related traits. *Diabetes* 56:522–526, 2007
33. Heward JM, Brand OJ, Barrett JC, Carr-Smith JD, Franklyn JA, Gough SC: Association of PTPN22 haplotypes with Graves' disease. *J Clin Endocrinol Metab* 92:685–690, 2007
34. Hinks A, Eyre S, Barton A, Thomson W, Worthington J: Investigation of genetic variation across the protein tyrosine phosphatase gene in patients with rheumatoid arthritis in the UK. *Ann Rheum Dis* 66:683–686, 2007