Yan Ge,[1] Suna Onengut-Gumuscu,[2,3] Aaron R. Quinlan,[4,5] Aaron J. Mackey,[2,3] Jocyndra A. Wright,[1] Jane H. Buckner,[6] Tania Habib,[6] Stephen S. Rich,[2,3] and Patrick Concannon[1,7]

Targeted Deep Sequencing in Multiple-Affected Sibships of European Ancestry Identifies Rare Deleterious Variants in *PTPN22* That Confer Risk for Type 1 Diabetes

GENETICS/GENOMES/PROTEOMICS/METABOLOMICS

Despite finding more than 40 risk loci for type 1 diabetes (T1D), the causative variants and genes remain largely unknown. Here, we sought to identify rare deleterious variants of moderate-to-large effects contributing to T1D. We deeply sequenced 301 protein-coding genes located in 49 previously reported T1D risk loci in 70 T1D cases of European ancestry. These cases were selected from putatively high-risk families that had three or more siblings diagnosed with T1D at early ages. A cluster of rare deleterious variants in *PTPN22* was identified, including two novel frameshift mutations (ss538819444 and rs371865329) and two missense variants (rs74163663 and rs56048322). Genotyping in 3,609 T1D families showed that rs56048322 was significantly associated with T1D and that this association was independent of the T1D-associated common variant rs2476601. The risk allele at rs56048322 affects splicing of *PTPN22*, resulting in the production of two alternative *PTPN22* transcripts and a novel isoform of LYP (the protein encoded by *PTPN22*). This isoform competes with the wild-type LYP for binding to CSK and results in hyporesponsiveness of CD4+ T cells to antigen stimulation in T1D subjects. These findings demonstrate that in addition to common variants, rare deleterious variants in *PTPN22* exist and can affect T1D risk.

Type 1 diabetes (T1D) results from the autoimmune destruction of insulin-producing β-cells in the pancreatic islets, culminating in complete insulin deficiency and life-long reliance on administration of exogenous insulin. Although there exist rare monogenic forms of T1D (1,2), the common form is regarded as genetically complex, arising from the actions, and possible interactions, of multiple genetic and environmental risk factors. The human leukocyte antigen (HLA) region on chromosome 6p21 accounts for approximately half of the genetic risk for T1D (3,4). Genome-wide association studies (GWAS) and meta-analyses have identified more than 40 additional, non-HLA T1D risk loci (5–13), most of which have modest individual effects on disease risk (14).

Despite the identification of numerous T1D-associated genomic regions from GWAS, the specific causative variants for T1D located in these and other regions of the genome, as well as the genes they impact, remain largely unknown. Linkage disequilibrium within disease-associated regions limits the power to pinpoint by association mapping the causative variants among a number of tightly correlated and common alleles at different polymorphic sites. Additional difficulties are posed by associations detected at sites that are neither coding nor obviously regulatory and, therefore, without predicted biological functions.

One solution to these challenges may lie in the identification of infrequent or rare protein-altering variants, which might be expected to cluster in the gene or

[1]Department of Pathology, Immunology and Laboratory Medicine, University of Florida, Gainesville, FL
[2]Center for Public Health Genomics, University of Virginia, Charlottesville, VA
[3]Department of Public Health Sciences, University of Virginia, Charlottesville, VA
[4]Department of Human Genetics, University of Utah, Salt Lake City, UT
[5]Department of Biomedical Informatics, University of Utah, Salt Lake City, UT
[6]Translational Research Program, Benaroya Research Institute at Virginia Mason, Seattle, WA
[7]Genetics Institute, University of Florida, Gainesville, FL

Corresponding author: Patrick Concannon, patcon@ufl.edu.

genes relevant to pathogenesis in a disease-associated region and, having more readily discernible effects on gene function, could provide insights into the role of the gene in pathogenesis. In a prior study of a subset of T1D-associated loci, the exons and splice sites of 10 candidate genes in T1D case and healthy control subjects were deeply sequenced (15). Four infrequent variants (minor allele frequency [MAF] <3%) in *IFIH1* were found to be independently associated with lower risk for T1D, providing important clues to the role of *IFIH1* in T1D pathogenesis.

In the current study, we searched for rare deleterious variants of moderate-to-high penetrance by deep sequencing the coding regions of 301 genes in 49 non-HLA T1D risk loci in T1D cases from families with three or more siblings diagnosed with T1D at early ages. Such families are rare and putatively at high risk, given that the risk to siblings of a T1D case is approximately 6% (16). We identified a cluster of rare deleterious variants in *PTPN22* and found that one of them, rs56048322, affected the splicing of *PTPN22* transcripts and the function of PTPN22 in T cells.

RESEARCH DESIGN AND METHODS

Human Subjects

Genomic DNA samples from T1D-affected sibling pairs (ASP) and trio families were obtained from two sources: The Human Biological Data Interchange (HBDI) (375 ASP families, *n* = 1,840) (17,18) and the Type 1 Diabetes Genetics Consortium (T1DGC) (3,234 ASP/trio families, *n* = 12,424) (19). DNA and peripheral blood mononuclear cells (PBMCs) were obtained from the Benaroya Research Institute Translational Research Program's Immune Mediated Diseases Registry. Research protocols were approved by the Benaroya Research Institute Institutional Review Board.

Selection Criteria

Potential high-risk T1D families were identified based on the following criteria: *1*) the parents of the T1D cases were of European ancestry and had three or more children diagnosed with T1D at or after 1 year of age (to exclude neonatal diabetes) and prior to 21 years of age and *2*) neither of the parents had T1D (to reduce the risk of selecting subjects with dominantly inherited maturity-onset diabetes of the young).

Targeted Deep Sequencing and Bioinformatic Analyses

For sequencing, 10 genomic DNA samples were combined into each of seven pools, and libraries were constructed using the Paired-End Sample Preparation kit (Illumina). Target regions were enriched using the SureSelect Paired-End Target Enrichment System (Agilent Technologies). Seven target-enriched libraries were sequenced on an Illumina Genome Analyzer II, and 63-bp paired-end reads were generated.

The sequencing reads were aligned to the UCSC hg18 human reference genome with the Burrows-Wheeler Aligner (version 0.5.9) (20). PCR duplicate reads were removed

using SAMtools, and one sorted Binary Alignment/Map file was generated and indexed for each sequencing library (21). Single nucleotide polymorphisms (SNPs) and indels were identified using FreeBayes (version 0.5.1) and the Genome Analysis Toolkit (version 2.4–9) (22), respectively. Sequencing depths were calculated using the Picard package (version 1.87). Variants were annotated using the Variant Effect Predictor (version 71) (23). The impacts of missense variants on protein functions were predicted using SIFT (version 5.0.2) (24) and PolyPhen (version 2.2.2) (25). Variants with the following consequences were considered deleterious: frameshift, in-frame insertion/deletion, stop codon lost or gained, predicted damaging missense by both SIFT and PolyPhen, or occurring in a splice donor or acceptor region. A variant was regarded as rare if its MAF was ≤0.00036 in European populations; this corresponded to a probability of 0.05 of detecting a variant allele in 70 randomly sampled individuals.

Genotyping

The MGB Eclipse system (ELITechGroup, Epoch Biosciences) was used to genotype 3,609 T1D ASP/trio families consisting of 14,264 participants collected by the HBDI and T1DGC, according to the manufacturer's protocol.

Cell Isolation, Stimulation, and Culture

Primary B cells, T cells, and monocytes were purified from PBMCs by positive selection with CD19, CD3, and CD14 magnetic beads, respectively (Miltenyi). For quantitative PCR, CD4$^+$ T cells were purified from PBMCs by negative selection using the human CD4$^+$ T-cell isolation kit (Miltenyi) and were stimulated with 2.5 μg/mL plate-bound anti-CD3 antibody (clone OKT3) and 0.25 μg/mL soluble anti-CD28 antibody for 0, 6, 18, and 24 h. Jurkat cells and B-lymphoblastoid cell lines (B-LCLs) were cultured in RPMI 1640 medium supplemented with 10% FBS (HyClone; Thermo Scientific), 2.05 mmol/L L-glutamine, 100 units/mL penicillin, 100 μg/mL streptomycin, and 1 mmol/L sodium pyruvate (Life Technologies) at 37°C with 5% CO_2.

Quantitative RT-PCR

Total RNA was extracted using the RNeasy Mini kit (QIAGEN), and contaminating genomic DNA was removed with Ambion DNA-free DNase Treatment and Removal Reagents (Life Technologies). First-strand cDNA was synthesized using random hexamer primers and SuperScript II Reverse Transcriptase (Life Technologies). PCRs containing SYBR Green I and ROX reference dye were performed on an ABI 7900HT Fast Real-Time PCR System (Life Technologies). All samples were tested in triplicate. Relative gene expression levels were calculated using the comparative C_T method (26), where ΔC_T = mean C_T (*PTPN22*) − mean C_T (*GAPDH*). The primers used for the assay are provided in Supplementary Table 1. The heat map in Fig. 1*C* was generated using the $2^{-\Delta CT}$ values and the R packages, gplots, and RColorBrewer.
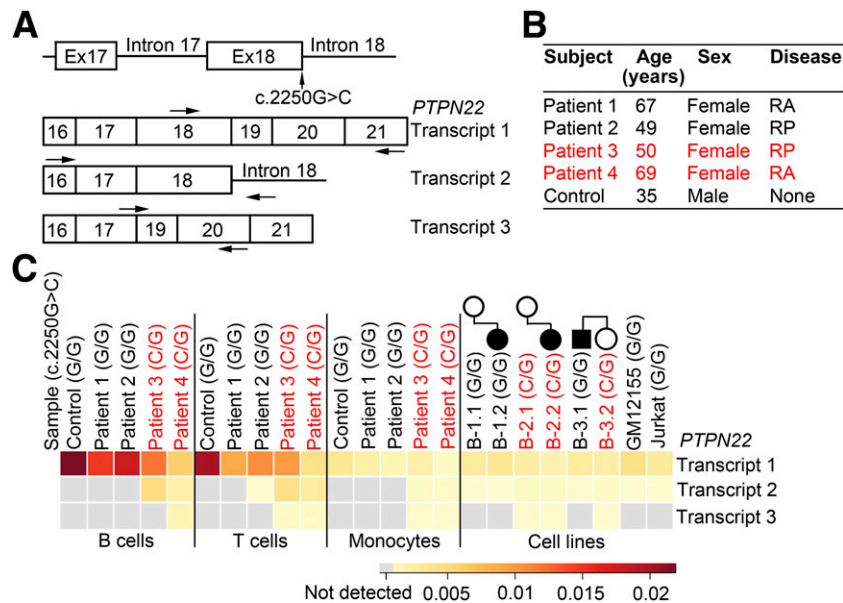
Figure 1—Relative expression levels of three *PTPN22* transcripts. *A*: Schematic diagrams of the locations of c.2250G>C (rs56048322) and primers (indicated by arrows) used for quantitative real-time PCR. Transcript 1 depicts full-length *PTPN22*, transcript 2 is an alternative *PTPN22* transcript that extends into intron 18, and transcript 3 is an alternative *PTPN22* transcript that skips exon 18 by splicing exon 17 to exon 19. *B*: Characteristics of human subjects. *C*: A heat map of the relative expression levels of the three *PTPN22* transcripts. ΔC_T values were calculated by subtracting the mean C_T of *GAPDH* from the mean C_T of *PTPN22*. The $2^{-\Delta CT}$ values are represented by colors: dark red indicates high relative expression levels, light yellow indicates low relative expression levels, and gray indicates values below the detection limit of the assay. The pedigree information of the B-LCLs is shown (white circle, unaffected female; black circle, T1D-affected female; black square, T1D-affected male). c.2250G/C heterozygous samples are marked in red. All the subjects and cell lines are homozygous for the major nonrisk allele of rs2476601 in *PTPN22*.

Coimmunoprecipitation and Immunoblotting

Cells were lysed in EBC lysis buffer (50 mmol/L Tris-HCl, pH 7.5, 120 mmol/L NaCl, 0.5% NP-40, and 1 mmol/L EDTA) containing protease and phosphatase inhibitors (cOmplete Mini and PhosSTOP, Roche). One milligram of lysate in 1 mL of EBC lysis buffer was precleared with GammaBind Plus Sepharose beads (GE Healthcare) for 2 h at 4°C and then incubated overnight at 4°C with 0.6 μg rabbit anti–tyrosine-protein kinase CSK (CSK) antibody (#4980; Cell Signaling Technology) or 0.6 μg normal rabbit IgG (SouthernBiotech). Immune complexes were captured by incubating the lysate with 20 μL GammaBind Plus Sepharose beads for 2 h at 4°C. The beads were washed four times with 1 mL EBC lysis buffer containing 1 mmol/L phenylmethanesulfonyl fluoride and 1 mmol/L sodium orthovanadate. The immunoprecipitates were eluted by heating the beads for 5 min at 95°C in 25 μL lithium dodecyl sulfate sample buffer (Life Technologies) containing 50 mmol/L dithiothreitol.

Whole-cell lysates and immunoprecipitates were resolved on 7% NuPAGE Novex Tris-Acetate gels (Life Technologies). Immunoblotting was performed using antibodies to lymphoid protein tyrosine phosphatase (LYP) (AF3428, 0.3 μg/mL; R&D Systems), CSK (H00001445-M01, 1 μg/mL; Abnova), and γ-tubulin (T6557, 1:10,000 dilution; Sigma-Aldrich).

Calcium Flux Measurement

Previously frozen PBMCs were incubated with the cell-permeant dye indo-1, AM (Molecular Probes) for 45 min at 37°C and then surface-stained with anti–CD8 PE/Cy5 (BioLegend), anti–CD27-APC (BD Biosciences), anti–CD45RO-PE (BD Biosciences), anti–CD45RA-FITC (BioLegend), anti–CD4-PerCP/Cy5.5 (BioLegend), and anti–CD19-PE/Cy7 (BioLegend). The indo-1 fluorescence ratio was acquired as a function of time on an LSR II flow cytometer, and kinetics curves were generated using the FlowJo software. Collection of a 30-s or 1-min baseline was followed by stimulation with 10 μg/mL soluble anti-CD3 (clone UCHT1, BioLegend), 20 μg/mL anti-κ + anti-λ F(ab')$_2$ (SouthernBiotech), or 1 μg/mL ionomycin. All the calcium flux profiles were generated using a standard protocol.

Statistical Analyses

The Family-Based Association Test (FBAT) program (version 2.0.4) was used for single-marker association tests and haplotype analyses, as well as for checking Mendelian inconsistency in genotyping (27). MAFs were estimated using PLINK (version 1.07) (28). Conditional association analyses were conducted using the UNPHASED program (version 3.1) (29).

RESULTS

Identification of Rare Deleterious Variants by Targeted Sequencing of T1D Cases From Multiple-Affected Families of European Ancestry

We applied our selection criteria for putatively high-risk families (see RESEARCH DESIGN AND METHODS) and chose 70 T1D cases of European ancestry from the T1D-affected

families collected by the T1DGC. In these 70 cases, the coding regions of all 301 known protein-coding genes located in 49 T1D-associated chromosomal regions, excluding the HLA region, were captured (1.12 Mb per subject) (Supplementary Table 2) and sequenced in pools of 10 subjects each.

For each pool, on average, 97.64% of the target regions were covered by at least two reads (ranging from 97.56 to 97.74%), and 93.4% of the target bases were covered at $\geq 30\times$ (ranging from 92.9 to 94.0%). All pools had mean target coverage of $>660\times$ (ranging from $661\times$ to $1,018\times$). Fifty-one genes were identified that harbored two or more rare (MAF ≤ 0.00036) predicted deleterious variants. In the T1D-associated region on chromosome 1p13.2, 10 genes were sequenced; only two rare (MAF ≤ 0.00036) predicted deleterious variants were identified, and both occurred in *PTPN22*—a 4-bp deletion in exon 11 (ss538819444) and a 1-bp deletion in exon 13 (rs371865329). Two infrequent (MAF <0.01) coding missense variants with predicted deleterious effects, rs74163663 (p.Arg748Gly) and rs56048322 (p.Lys750Asn), were also identified, and were similarly clustered in *PTPN22*. A simulation in which 10,000 sets of 70 T1D cases were randomly selected from the initial study population showed that the probability of identifying all four of the deleterious variants seen in *PTPN22* was 1.3×10^{-4}, suggesting that the criteria used to define high-risk T1D families had effectively enriched for infrequent variants with predicted deleterious effects in *PTPN22*. Of note, the selection criteria did not enrich for the most common genetic risk variants for T1D, which are high-risk HLA DRB1-DRA1-DQB1 haplotypes (Supplementary Table 3).

PTPN22 is an established risk locus for T1D and several other autoimmune disorders (30–34). Risk at *PTPN22* is associated with the minor allele at a common nonsynonymous coding region SNP, rs2476601, which results in the missense substitution R620W.

Rare Allele of rs56048322 in *PTPN22* Has an Independent Effect on T1D Risk

We genotyped ss538819444, rs371865329, rs74163663, and rs56048322 in 14,264 individuals from 3,609 T1D ASP/trio families. Single-marker FBATs (27) were performed to examine the association between variants in *PTPN22* and T1D. Positive and negative Z scores of FBAT indicate risk and protection, respectively. The single-marker FBAT showed that the minor alleles of rs56048322 (P = 0.0014), rs74163663 (P = 0.028), and rs2476601 ($P = 8.62 \times 10^{-36}$) were each associated with T1D (Table 1). Comparable results were obtained by assessing the sharing of alleles at these markers among ASPs by the transmission disequilibrium test. No association with T1D was detected for ss538819444 or rs371865329, but their rarity in the population (each occurred in fewer than three families) limited the power to assess their effects on T1D risk.

Six haplotypes for ss538819444, rs371865329, rs74163663, rs56048322, and the common T1D risk variant, rs2476601, were identified in the T1D ASP/trio families (Table 2). Differences in the numbers of informative families shown in Tables 1 and 2 are due to missing genotypes. The H2 haplotype carrying the minor 1858T risk allele at rs2476601 was significantly associated with T1D ($P = 1.81 \times 10^{-34}$), as expected. The H3 haplotype, which carried the 2250C risk allele at rs56048322, was also associated with T1D (P = 0.0014) in the absence of the common risk allele at rs2476601 (Table 2). The minor allele at rs56048322 still exhibited a significant association with T1D (P = 0.0003) after conditioning on rs2476601. The minor allele of rs74163663 co-occurred with the 1858T risk allele at rs2476601 (Table 2) and showed a marginally significant association with T1D (P = 0.06) after conditioning on rs2476601.

Minor Allele of rs56048322 Is Associated with the Production of Alternative *PTPN22* Transcripts

The SNP rs56048322 alters the last nucleotide of exon 18 of *PTPN22*. In a prior report, PBMCs from one individual heterozygous at rs56048322 produced an alternative *PTPN22* transcript in which exon 17 was spliced to exon 19 creating a premature stop codon at the exon 17-to-19 junction (Fig. 1A) (35). In B-LCLs derived from subjects heterozygous at rs56048322, both this previously reported alternative transcript and a second more common alternative *PTPN22* transcript, resulting from run-on transcription into intron 18, were detected (Fig. 1C).

Table 1—Single-marker FBAT analysis of variants in the *PTPN22* gene

Exon	rsID	cDNA change	Amino acid change	22	12	11	MAF	Family (*n*)	Z score	*P* value
11	ss538819444	c.878_881delAGAT	p.Gln293Argfs*45	14,210	5	0	0.00018	2	0.447	0.65
13	rs371865329	c.1040delA	p.Ile348Serfs*14	14,184	6	0	0.000088	1	1.000	0.32
14	rs2476601	c.1858C>T	p.Arg620Trp	10,600	3,303	341	0.12	1,023	12.489	8.62E-36
18	rs74163663	c.2242A>G	p.Arg748Gly	14,154	22	0	0.00053	7	2.191	0.028
18	rs56048322	c.2250G>C	p.Lys750Asn	13,943	284	0	0.0087	93	3.190	0.0014

cDNA and amino acid changes are based on NM_015967.5 and NP_057051.3, respectively. Z scores and *P* values are calculated for the minor alleles. 22, homozygous for the major allele; 12, heterozygous; 11, homozygous for the minor allele; *n*, number of informative families.

Table 2—Family-based haplotype association analysis of *PTPN22*

Haplotype	ss538819444	rs371865329	rs2476601	rs74163663	rs56048322	Frequency	Family (*n*)	Z score	P value
H1	2	2	2	2	2	0.872	1,072	−12.98	1.68E-38
H2	2	2	1	2	2	0.119	1,012	12.24	1.81E-34
H3	2	2	2	2	1	0.008	92	3.19	0.0014
H4	2	2	1	1	2	0.001	7	2.19	0.028
H5	1	2	2	2	2	<0.001	2	0.45	0.65
H6	2	1	2	2	2	<0.001	1	1.00	0.32

1, minor allele; 2, major allele; *n*, number of informative families.

These alternative transcripts and the full-length *PTPN22* transcript were quantified in primary T cells, B cells, and monocytes from five subjects, including one healthy control, two rheumatoid arthritis (RA), and two relapsing polychondritis (RP) case subjects (Fig. 1*B*). The use of samples from subjects with autoimmune disorders other than T1D were necessitated both by the rarity of carriers of the minor allele at rs56048322 and the availability of comparably treated, viably frozen PBMCs from these subjects. As shown in Fig. 1*C*, the presence of the risk allele was associated with a substantial enrichment for the two alternative *PTPN22* transcripts in all three primary cell types and in B-LCLs derived from T1D families (Fig. 1*C*). The average relative expression level of the intron-18–run-on transcript in the heterozygous B-LCLs was 72 times as high as that in the B-LCLs homozygous for the major allele (Fig. 1*C*). The exon-18–skipping transcript was detected only at low levels in individuals heterozygous at rs56048322 (Fig. 1*C*). In B-1.1 and B-1.2, which were heterozygous for c.878_881delAGAT, transcripts carrying the 4-bp deletion were detected by Sanger sequencing of PCR products generated from cDNA, but the relative expression level of full-length *PTPN22* was not evidently lower in these cells than in the wild-type B-LCLs (Fig. 1*C*).

In human primary CD4$^+$ T cells stimulated with anti-CD3 and anti-CD28 antibodies, the amounts of the full-length *PTPN22* transcript peaked at 6 h post stimulation and then decreased to below basal levels by 24 h, regardless of the genotype at rs56048322 (Fig. 2*A* and *B*). The intron-18–run-on *PTPN22* transcript was expressed at low levels in CD4$^+$ T cells from subjects homozygous for the major allele of rs56048322, and the expression levels decreased upon stimulation (Fig. 2*A* and *C*). In contrast, in CD4$^+$ T cells from carriers of the risk allele of rs56048322, the amounts of the intron-18–run-on transcript, which were markedly higher than those in the noncarriers, increased upon stimulation and peaked at 6 h (Fig. 2*A* and *C*) as the full-length transcript did. The exon-18–skipping *PTPN22* transcript was detected only in carriers of the risk allele of rs56048322, and the expression levels decreased upon stimulation (Fig. 2*A* and *D*).

Rare Allele of rs56048322 Produces a Novel LYP Isoform That Can Bind to CSK

PTPN22 encodes the intracellular LYP, which regulates signaling through the T- and B-cell receptors. LYP forms a complex with CSK, a negative regulator of T-cell receptor (TCR)–mediated signaling. This complex of LYP and CSK is reported to exert synergistic inhibition on TCR signaling (36). LYP has four proline-rich motifs (P1-P4) at the C-terminus, of which, the P1 motif interacts with the SH3 domain of CSK (37,38). The tryptophan residue introduced at position 620 by the risk allele of rs2476601 is located in the P1 motif and decreases the binding of LYP to CSK (30). It has been proposed that this effect on the interaction between LYP and CSK is responsible for the risk for autoimmunity associated with the risk allele at rs2476601. In whole-cell lysates derived from B-LCLs heterozygous at rs56048322, two LYP isoforms were detected, one corresponding to the predicted size for full-length LYP (91.7 kDa) and the second to the predicted size for the intron-18–run-on product (87.2 kDa) (Fig. 3*A*). Both isoforms could be coimmunoprecipitated with CSK at comparable levels (Fig. 3*B*), indicating comparable ability to bind to CSK.

Rare Allele of rs56048322 Impairs TCR-Mediated Calcium Flux in CD4$^+$ T Cells

Although the association of the minor allele at rs2476601 with autoimmunity is well established, there are conflicting reports as to whether this allele results in loss or gain of function of LYP (30,39–43). The availability of a second independent risk allele affecting the coding region of *PTPN22* provided an opportunity to further explore the mechanism whereby *PTPN22* mediates risk for autoimmunity. To determine the effects of the independent T1D risk allele of rs56048322, intracellular calcium flux induced by TCR cross-linking in CD4$^+$ T cells from the peripheral blood of a healthy subject and age-matched case subjects with T1D or RA was compared (Fig. 4). All subjects were homozygous for the nonrisk 1858C allele at rs2476601. Compared with the subjects homozygous for the major allele of rs56048322, CD4$^+$ T cells from subjects heterozygous at rs56048322 exhibited decreased TCR-mediated intracellular calcium flux (Fig. 4), consistent with a gain of function. However, the risk allele of rs56048322 did not impair calcium flux in TCR-stimulated CD4$^+$ T cells from nonautoimmune control subjects (Supplementary Fig. 1) or in human primary B cells stimulated via B-cell receptors (Supplementary Fig. 2).
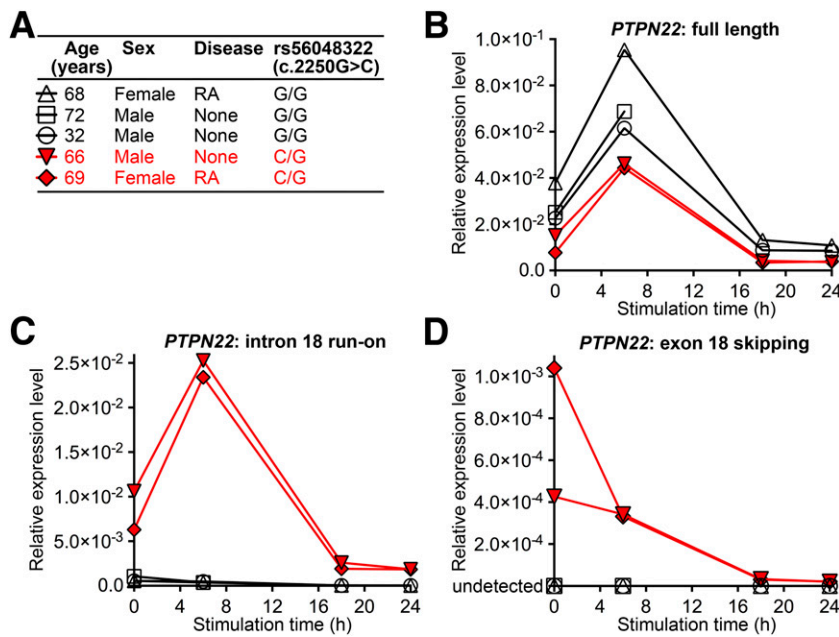
Figure 2—Relative expression levels of three *PTPN22* transcripts in human primary CD4[+] T cells upon TCR-mediated stimulation. *A*: Information on the human subjects in panels *B–D*. Carriers of the minor risk allele of rs56048322 are shown in red. All the subjects are homozygous for the major nonrisk allele of rs2476601 in *PTPN22*. *B–D*: Primary CD4[+] T cells from five human subjects were stimulated with anti-CD3 and anti-CD28 antibodies for 0, 6, 18, and 24 h. The relative expression levels of three *PTPN22* transcripts (depicted in Fig. 1*A*) in the stimulated CD4[+] T cells are shown. ΔC_T values were calculated by subtracting the mean C_T of *GAPDH* from the mean C_T of *PTPN22*, and the $2^{-\Delta CT}$ values are shown.

DISCUSSION

In this study, we searched for rare risk variants of moderate-to-large effect sizes in chromosomal regions implicated as harboring risk loci for T1D in prior GWAS. We implemented a family-based approach, given the hypothesis that T1D families having multiple affected siblings with early ages at onset would be enriched for risk variants with large effect sizes that might be expected to be rare in the population. We focused on rare variants that had substantial predicted impacts on the encoded proteins, including protein-altering indels, stop gain/loss variants, and splice variants, with the expectation that their distribution among genes in a region might identify the relevant causative gene and that their predicted deleterious effects might provide insight into the mechanism whereby the implicated gene might act to confer risk for T1D. Multiple novel deleterious variants in different T1D risk regions were identified by this approach. Functional studies were performed for one gene, *PTPN22*, where the enrichment for rare indels and predicted deleterious missense variants supported its role as the causative gene in the region and provided an opportunity to explore the mechanism of its effect on T1D risk.

Conflicting results have been reported as to whether the risk allele of rs2476601 in *PTPN22*, which encodes an R to W amino acid substitution at position 620, results in loss (30,39) or gain (40–43) of function. Two of the deleterious variants identified by sequencing of *PTPN22* were

predicted to result in premature termination, consistent with a possible loss of function, but had MAFs of <0.0002, limiting the ability to carry out functional studies. In B-lymphoblastoid cells from carriers of these variants, no alternative *PTPN22* isoforms were observed and the wild-type isoform occurred at normal levels. A rare predicted deleterious missense variant, rs74163663, was detected but only on haplotypes also containing the common risk allele at rs2476601. The minor allele of the remaining rare deleterious variant identified in *PTPN22*, rs56048322, was associated with increased T1D risk independent of the common T1D-associated risk allele of rs2476601 (Table 2). The risk allele of rs56048322 interfered with the correct splicing of *PTPN22*, resulting in the production of two alternative *PTPN22* transcripts, one of which was degraded upon T-cell activation, whereas the other generated significant amounts of a lower–molecular weight C-terminally truncated LYP isoform. Unlike the product of the risk allele at rs2476601 (30), this LYP isoform bound CSK through its P1 domain but lacked the C-terminal P3 and P4 proline-rich domains that likely mediate additional, as-yet uncharacterized protein-protein interactions. Despite the ability of its protein product to bind CSK, the rs56048322 risk allele, similar to the risk allele at rs2476601, resulted in blunted TCR-mediated calcium flux in CD4[+] T cells, a phenotype most pronounced in the memory CD4[+] T-cell compartment (data not shown) (41). Unlike carriers of the rs2476601 risk allele, carriers of the risk allele at rs56048322 displayed more variability in
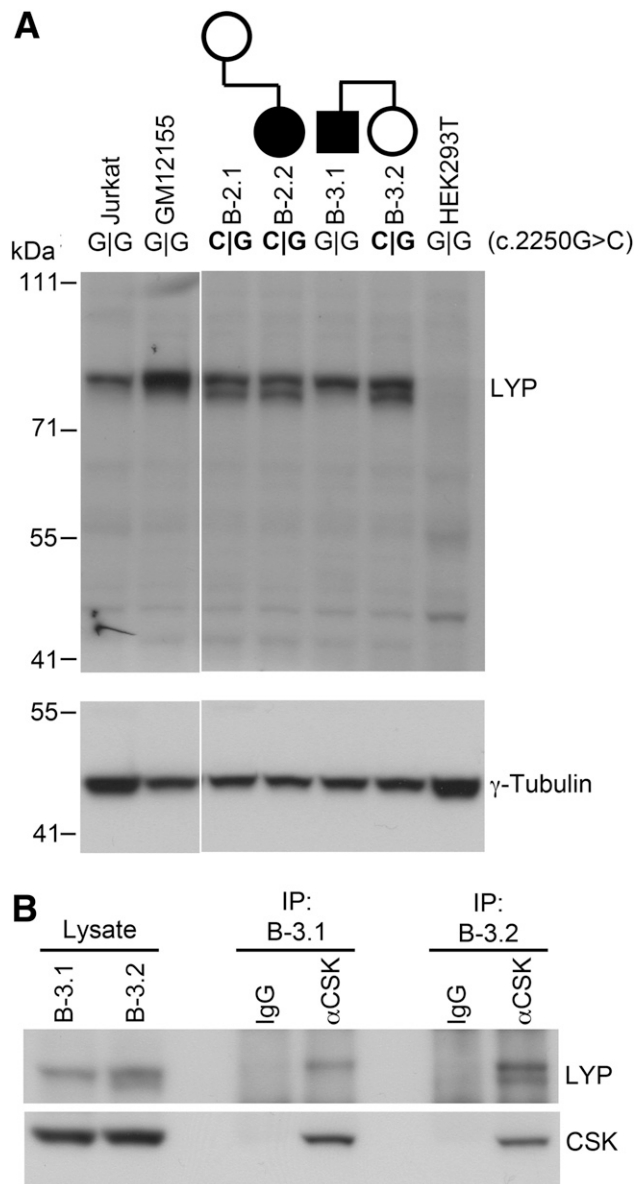
Figure 3—Characterization of the effects of c.2250G>C on LYP. *A*: Immunoblotting of whole-cell lysates from the following cell lines: Jurkat, GM12155, four B-LCLs derived from two T1D-affected families, and human embryonic kidney 293T (HEK293T). The pedigree information and genotypes for c.2250G>C (rs56048322) are shown (white circle, unaffected female; black circle, T1D-affected female; black square, T1D-affected male). The c.2250G/C heterozygous genotype is highlighted in bold. LYP was detected using an anti-LYP antibody. The blot was stripped and reprobed with an anti–γ-tubulin antibody. The vertical white line indicates that the lanes are not consecutive, although they are of the same gel. *B*: Whole-cell lysates from B-3.1 and B-3.2 were immunoprecipitated (IP) with a rabbit anti-CSK antibody or normal rabbit IgG. The immunoprecipitated samples and the whole-cell lysates were subjected to immunoblotting with an anti-LYP antibody. The blot was stripped and reprobed with an anti-CSK antibody.

the degree of T-cell hyporeactivity, and this phenotype was only observed in subjects with existing autoimmunity. No effect of rs56048322 was observed on B-cell receptor–mediated calcium flux in CD19⁺ B cells (Supplementary

Fig. 2), and further investigation will be required to determine whether rs56048322 affects signaling in specific subtypes of B cells as rs2476601 does (41). It should be noted that carriers of the rs56048322 risk allele are rare. Only six carriers were available for the functional studies carried out here; a comprehensive description of immune phenotypes associated with this variant will require further ascertainment and study. Nevertheless, the findings presented here suggest that the rare risk allele at rs56048322 and the more common risk allele at rs2476601 exert similar effects on T-cell activation and function but likely do so through different mechanisms. We propose that both act by dominant interference, producing a protein product either lacking or defective in selected proline-rich domains but differing in which domains are affected and, presumably, what interactions are disrupted. The evidence from these two independently disease-associated genetic variants, acting by distinct mechanisms but resulting in a common phenotype, suggests that risk alleles at *PTPN22* contribute to T1D susceptibility via blunted TCR signaling that may be either due to direct effects of the risk variants or due to compensatory alterations in this pathway in response to these variants.

In addition to providing insight into the mechanism of action of *PTPN22* in autoimmunity, the enrichment we observed for rare deleterious variants in this single gene in the 1p13.2 region raises the possibility that resequencing in high-risk families, as defined here, could be used to identify causative genes in other regions delineated by GWAS or, more broadly, to identify novel risk loci. Although we did identify additional deleterious variants in other T1D-associated regions in our study, the small number of the families studied (n = 70) limited our power to detect any enrichment for deleterious variants in single genes at additional sites. Certainly, the use of high-risk families has been instrumental in identifying highly penetrant risk variants in various cancers. Our success in the current study with *PTPN22* encourages the application of this approach on a larger scale in autoimmunity. Of note, the T1DGC family collection, which was ascertained based on the presence of at least two affected siblings, is already enriched for high-risk HLA DR-DQ haplotypes; our selection criteria for high-risk families, which required one additional affected sibling, did not further increase the frequencies of high-risk HLA DR-DQ haplotypes (Supplementary Table 3) but were effective at identifying low-frequency deleterious variants of T1D-associated loci with lower overall risk than the HLA locus. Although the current study used a targeted approach focusing on chromosomal regions previously implicated by GWAS, an exome sequencing approach would be more efficient and offer the possibility of identifying novel risk loci not previously identified by GWAS. In the case of multiple sclerosis (MS), exome sequencing was recently performed in 43 affected individuals selected from families with four or more MS-affected members in more than one generation. One novel loss-of-function variant was identified in *CYP27B1* in one of the cases, and the variant was
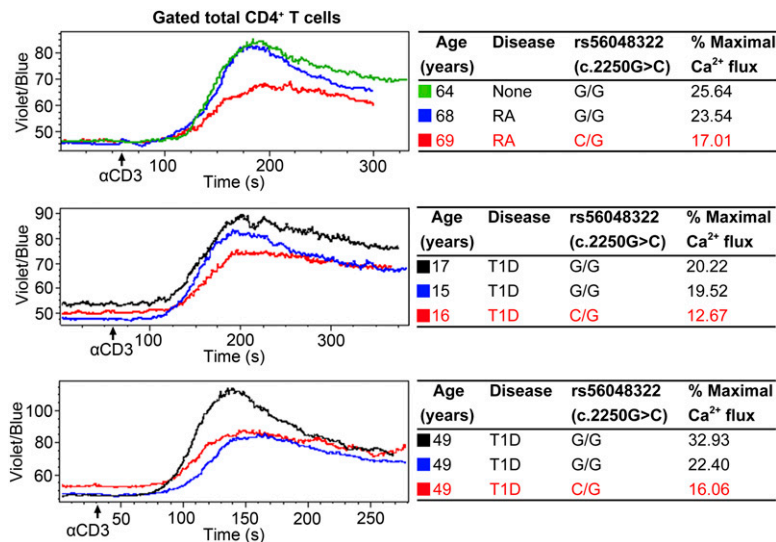
Figure 4—The risk allele of rs56048322 affects TCR-mediated intracellular calcium flux in CD4$^+$ T cells from subjects with T1D or RA. Previously frozen PBMCs from a healthy subject and from subjects with T1D or RA were loaded with cell-permeant indo-1, AM dye and stained with antibodies against T-cell surface markers. Shown are kinetics profiles depicting the mean indo-1 ratio (violet/blue) as a function of time in gated total CD4$^+$ T cells before and after stimulation with 10 µg/mL anti-CD3 antibody. Percent maximal calcium flux is expressed as [(peak anti-CD3 flux − mean baseline)/(peak ionomycin flux − mean baseline)] × 100. All the subjects are homozygous for the major nonrisk allele of rs2476601 in *PTPN22*.

significantly associated with MS, indicating a causative role for *CYP27B1* in MS (44). Results such as these, and the findings from the current study, support the examination of the human exome in families with high disease burdens as an approach for the identification of causative genes and variants for common autoimmune disorders.

References

1. Nagamine K, Peterson P, Scott HS, et al. Positional cloning of the APECED gene. Nat Genet 1997;17:393–398
2. Bennett CL, Christie J, Ramsdell F, et al. The immune dysregulation, polyendocrinopathy, enteropathy, X-linked syndrome (IPEX) is caused by mutations of FOXP3. Nat Genet 2001;27:20–21
3. Risch N. Assessing the role of HLA-linked and unlinked determinants of disease. Am J Hum Genet 1987;40:1–14
4. 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 1996;59:1134–1148

5. Todd JA, Walker NM, Cooper JD, et al.; Genetics of Type 1 Diabetes in Finland; Wellcome Trust Case Control Consortium. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat Genet 2007;39:857–864
6. Burton PR, Clayton DG, Cardon LR, et al.; Wellcome Trust Case Control Consortium; Australo-Anglo-American Spondylitis Consortium (TASC); Biologics in RA Genetics and Genomics Study Syndicate (BRAGGS) Steering Committee; Breast Cancer Susceptibility Collaboration (UK). Association scan of 14,500 nonsynonymous SNPs in four diseases identifies autoimmunity variants. Nat Genet 2007;39:1329–1337
7. Hakonarson H, Grant SF, Bradfield JP, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature 2007;448:591–594
8. Burton PR, Clayton DG, Cardon LR, et al.; Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature 2007;447:661–678
9. Hakonarson H, Qu H-Q, Bradfield JP, et al. A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. Diabetes 2008;57:1143–1146
10. Cooper JD, Smyth DJ, Smiles AM, et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat Genet 2008;40:1399–1401
11. Grant SF, Qu H-Q, Bradfield JP, et al.; DCCT/EDIC Research Group. Follow-up analysis of genome-wide association data identifies novel loci for type 1 diabetes. Diabetes 2009;58:290–295
12. Barrett JC, Clayton DG, Concannon P, et al.; Type 1 Diabetes Genetics Consortium. Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. Nat Genet 2009;41:703–707
13. Bradfield JP, Qu H-Q, Wang K, et al. A genome-wide meta-analysis of six type 1 diabetes cohorts identifies multiple associated loci. PLoS Genet 2011;7: e1002293
14. Concannon P, Rich SS, Nepom GT. Genetics of type 1A diabetes. N Engl J Med 2009;360:1646–1654
15. Nejentsev S, Walker N, Riches D, Egholm M, Todd JA. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 2009;324:387–389

16. Spielman RS, Baker L, Zmijewski CM. Gene dosage and suceptibility to insulin-dependent diabetes. Ann Hum Genet 1980;44:135–150

17. Lernmark A, Ducat L, Eisenbarth G, et al. Family cell lines available for research. Am J Hum Genet 1990;47:1028–1030

18. Cox NJ, Wapelhorst B, Morrison VA, et al. 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 2001;69:820–830

19. Rich SS, Akolkar B, Concannon P, et al. Overview of the Type I Diabetes Genetics Consortium. Genes Immun 2009;10(Suppl. 1):S1–S4

20. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754–1760

21. Li H, Handsaker B, Wysoker A, et al.; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics 2009;25:2078–2079

22. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297–1303

23. McLaren W, Pritchard B, Rios D, Chen Y, Flicek P, Cunningham F. Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. Bioinformatics 2010;26:2069–2070

24. Ng PC, Henikoff S. SIFT: predicting amino acid changes that affect protein function. Nucleic Acids Res 2003;31:3812–3814

25. Adzhubei IA, Schmidt S, Peshkin L, et al. A method and server for predicting damaging missense mutations. Nat Methods 2010;7:248–249

26. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2-$\Delta\Delta$CT method. Methods 2001;25(4):402–408

27. Laird NM, Horvath S, Xu X. Implementing a unified approach to family-based tests of association. Genet Epidemiol 2000;19(Suppl. 1):S36–S42

28. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559–575

29. Dudbridge F. Likelihood-based association analysis for nuclear families and unrelated subjects with missing genotype data. Hum Hered 2008;66:87–98

30. Bottini N, Musumeci L, Alonso A, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat Genet 2004;36:337–338

31. Begovich AB, Carlton VE, Honigberg LA, et al. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. Am J Hum Genet 2004;75:330–337

32. Kyogoku C, Langefeld CD, Ortmann WA, et al. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. Am J Hum Genet 2004;75:504–507

33. Velaga MR, Wilson V, Jennings CE, et al. The codon 620 tryptophan allele of the lymphoid tyrosine phosphatase (LYP) gene is a major determinant of Graves' disease. J Clin Endocrinol Metab 2004;89:5862–5865

34. Vandiedonck C, Capdevielle C, Giraud M, et al. Association of the PTPN22*R620W polymorphism with autoimmune myasthenia gravis. Ann Neurol 2006;59:404–407

35. Onengut-Gumuscu S, Buckner JH, Concannon P. A haplotype-based analysis of the PTPN22 locus in type 1 diabetes. Diabetes 2006;55:2883–2889

36. Cloutier J-F, Veillette A. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. J Exp Med 1999;189:111–121

37. Cohen S, Dadi H, Shaoul E, Sharfe N, Roifman CM. Cloning and characterization of a lymphoid-specific, inducible human protein tyrosine phosphatase, Lyp. Blood 1999;93:2013–2024

38. 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 1996;15:4909–4918

39. Zhang J, Zahir N, Jiang Q, et al. The autoimmune disease-associated PTPN22 variant promotes calpain-mediated Lyp/Pep degradation associated with lymphocyte and dendritic cell hyperresponsiveness. Nat Genet 2011;43:902–907

40. Vang T, Congia M, Macis MD, et al. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. Nat Genet 2005;37:1317–1319

41. 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 2007;179:4704–4710

42. Arechiga AF, Habib T, He Y, et al. Cutting edge: the PTPN22 allelic variant associated with autoimmunity impairs B cell signaling. J Immunol 2009;182:3343–3347

43. Vang T, Liu WH, Delacroix L, et al. LYP inhibits T-cell activation when dissociated from CSK. Nat Chem Biol 2012;8:437–446

44. Ramagopalan SV, Dyment DA, Cader MZ, et al. Rare variants in the CYP27B1 gene are associated with multiple sclerosis. Ann Neurol 2011;70:881–886