

PTPN22 Trp⁶²⁰ Explains the Association of Chromosome 1p13 With Type 1 Diabetes and Shows a Statistical Interaction With HLA Class II Genotypes

Deborah J. Smyth, Jason D. Cooper, Joanna M.M. Howson, Neil M. Walker, Vincent Plagnol, Helen Stevens, David G. Clayton, and John A. Todd

OBJECTIVE—The disease association of the common 1858C>T Arg620Trp (rs2476601) nonsynonymous single nucleotide polymorphism (SNP) of protein tyrosine phosphatase; nonreceptor type 22 (*PTPN22*) on chromosome 1p13 has been confirmed in type 1 diabetes and also in other autoimmune diseases, including rheumatoid arthritis and Graves' disease. Some studies have reported additional associated SNPs independent of rs2476601/Trp⁶²⁰, suggesting that it may not be the sole causal variant in the region and that the relative risk of rs2476601/Trp⁶²⁰ is greater in lower risk by HLA class II genotypes than in the highest risk class II risk category.

RESEARCH DESIGN AND METHODS—We resequenced *PTPN22* and used these and other data to provide >150 SNPs to evaluate the association of the *PTPN22* gene and its flanking chromosome region with type 1 diabetes in a minimum of 2,000 case subjects and 2,400 control subjects.

RESULTS—Due to linkage disequilibrium, we were unable to distinguish between rs2476601/Trp⁶²⁰ ($P = 2.11 \times 10^{-87}$) and rs6679677 ($P = 3.21 \times 10^{-87}$), an intergenic SNP between the genes putative homeodomain transcription factor 1 and round spermatid basic protein 1. None of the previously reported disease-associated SNPs proved to be independent of rs2476601/Trp⁶²⁰. We did not detect any interaction with age at diagnosis or sex. However, we found that rs2476601/Trp⁶²⁰ has a higher relative risk in type 1 diabetic case subjects carrying lower risk HLA class II genotypes than in those carrying higher risk ones ($P = 1.36 \times 10^{-4}$ in a test of interaction).

CONCLUSIONS—In our datasets, there was no evidence for allelic heterogeneity at the *PTPN22* locus in type 1 diabetes, indicating that the SNP rs2476601/Trp⁶²⁰ remains the best candidate in this chromosome region in European populations. The

heterogeneity of rs2476601/Trp⁶²⁰ disease risk by HLA class II genotype is consistent with previous studies, and the joint effect of the two loci is still greater in the high-risk group. ***Diabetes* 57: 1730–1737, 2008**

To date, there are 10 loci with confirmed evidence for association with type 1 diabetes: the major histocompatibility complex (MHC) HLA class I and II genes (1,2), insulin (3,4), the *CTLA4* locus (5,6), protein tyrosine phosphatase; nonreceptor type 22 (*PTPN22*; 7), interleukin-2 receptor-2 α chain (*IL2RA*) (8–10), interferon induced with helicase C domain 1 (*IFIH1*) (11), and the regions on chromosomes 12q24, 12q13, 16p13, and 18p11 (12–14). With the recent capability to scan hundreds of thousands of single nucleotide polymorphisms (SNPs) across the genome in thousands of samples, the identification of complex disease genes and regions has greatly been enhanced (12). However, the localization of the causal variant(s) in the chromosome regions identified by genome-wide association (GWA) studies requires comprehensive resequencing, extensive genotyping, and statistical analysis in large sample sets. With the exceptions of the MHC (2), cytotoxic T-lymphocyte-associated protein 4 (*CTLA4*)-*ICOS* (6), *IL2RA* (8), and *INS* (4), which have been subjected to further resequencing and genotyping to try to narrow down candidates for the causal variant(s), the other type 1 diabetes loci have not yet been studied in any great detail (6,11,13). For *PTPN22*, there has been extensive focus on the single nsSNP rs2476601/Arg620Trp (7,15–18). Functional studies suggest that the Trp⁶²⁰ allele has gain-of-function, immunosuppressive effects compared with the more common allele Arg⁶²⁰ (19–21). However, there have been attempts to assess whether there are other variants that confer risk of disease that are independent of rs2476601/Trp⁶²⁰ (22,23).

One of the largest studies thus far to characterize disease association with *PTPN22* resequenced the coding regions of the gene in 48 North American individuals and genotyped 37 SNPs in or near *PTPN22* in up to 1,136 rheumatoid arthritis case subjects and 1,797 control subjects (set 1: 475 case subjects and 475 control subjects; set 2: 661 case subjects and 1,322 control subjects) (24). They reported two SNPs (rs1310182 and rs3789604) on a common haplotype that were associated with rheumatoid arthritis independently of rs2476601/Trp⁶²⁰ (rs1310182, $P = 0.002$ and 0.052 for sample sets 1 and 2, respectively; rs3789604, $P = 0.002$ and 0.014 for sample sets 1 and 2, respectively). In another rheumatoid arthritis study, Steer et al. (22) genotyped 45 SNPs from the *PTPN22* region using Affymetrix and Illumina GWA technology on pools of 250 rheumatoid arthritis case subjects and 250 control

From the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Addenbrooke's Hospital, Cambridge, U.K.

Corresponding author: John A. Todd, Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Addenbrooke's Hospital, Cambridge, U.K. E-mail: john.todd@cimr.cam.ac.uk

Received for publication 13 August 2007 and accepted in revised form 23 February 2008.

Published ahead of print at <http://diabetes.diabetesjournals.org> on 27 February 2008. DOI: 10.2337/db07-1131.

Additional information for this article can be found in an online appendix at <http://dx.doi.org/10.2337/db07-1131>.

AP4B1, adaptor-related protein complex 4, $\beta 1$ subunit; C1orf178, chromosome 1 open reading frame 178; CEPH, Centre d'Etude du Polymorphisme Humain; DCLRE1B, DNA cross-link repair 1B; GWA, genome-wide association; MAF, minor allele frequency; MAGI3, membrane-associated guanylate kinase, WW, and PDZ domain containing 3; MHC, major histocompatibility complex; MIP, molecular inversion probe; PHTF1, putative homeodomain transcription factor 1; *PTPN22*, protein tyrosine phosphatase; nonreceptor type 22; SNP, single nucleotide polymorphism; WTCCC, Wellcome Trust Case-Control Consortium.

© 2008 by the American Diabetes Association.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

subjects. They reported a SNP (rs1343125) in the gene membrane-associated guanylate kinase, WW, and PDZ domain containing 3 (*MAGI3*) that provided some evidence for an association independent of rs2476601/Trp⁶²⁰ ($P = 0.03$).

For type 1 diabetes, two studies proposed that there is evidence for allelic heterogeneity. Onengut-Gumuscu et al. (23) resequenced the coding regions in 94 type 1 diabetic case subjects and genotyped 11 SNPs in 374 type 1 diabetic families. They identified a rare nsSNP K750N, which generates a premature stop codon, and presented some evidence of association with type 1 diabetes ($P = 0.026$, minor allele frequency [MAF] = 0.006). An Asian population study (25) did not detect the rs2476601/Arg620Trp variant in 1,690 Japanese or 180 Korean subjects but found a promoter SNP -1123G/C (rs2488457) to be associated with type 1 diabetes ($P = 0.0105$; odds ratio [OR] 1.41 [95% CI 1.09–1.82]). They also genotyped this SNP in 95 families from the Diabetes U.K. Warren 1 Repository and found the associations to be the same for -1123G/C (rs2488457) and rs2476601/Trp⁶²⁰ ($P = 0.019$, relative risk (RR) 1.46 [95% CI 1.06–2.01]; and $P = 0.046$, 1.46 [1.01–2.13], respectively) (25). Furthermore, a Norwegian rheumatoid arthritis study of 861 case subjects and 559 control subjects (26) found they could not distinguish between -1123G/C (rs2488457) and rs2476601/Trp⁶²⁰ because of linkage disequilibrium.

Recently, Chelala et al. (27) genotyped the rheumatoid arthritis-associated SNP rs3789604, which marks a haplotype that may be associated with rheumatoid arthritis independently of rs2476601/Trp⁶²⁰, and the Japanese-associated SNP rs2488547 but failed to find evidence of an independent association with type 1 diabetes in 528 French, Danish, and American multiplex families (27). Heward et al. (28) selected five common rheumatoid arthritis-associated SNPs to use as tag SNPs (rs2488458, rs12730735, rs1310182, rs1217413, and rs3811021) and reported no evidence of association with Graves' disease in 768 case subjects, 768 control subjects, and 313 families (minimum $P = 0.292$). Using the five tag SNPs and rs2476601/Trp⁶²⁰, they did, however, obtain evidence for a predisposing haplotype (haplotype 2, $P = 6.77 \times 10^{-8}$) and a protective haplotype (haplotype 3, $P = 3.7 \times 10^{-5}$) (28). Dissection of the association of these haplotypes revealed that rs2476601/Trp⁶²⁰ was responsible for the predisposing effect observed with haplotype 2 but could not explain the apparent protective effect obtained for haplotype 3 (28).

Two previous studies have explored heterogeneity in the disease-predisposing effect of rs2476601/Trp⁶²⁰ with regard to HLA class II genotype. Hermann et al. (29) observed that the effect of rs2476601/Trp⁶²⁰ was more pronounced in subjects with non-DR4-DQ8/low-risk HLA genotypes ($P = 0.0004$) from a dataset of 546 case subjects, 538 control subjects, and 245 nuclear families from Finland (mean age at diagnosis of case subjects 8.2 ± 4.1 years). They also reported some evidence that boys carrying the rs2476601 Trp⁶²⁰ allele were at higher risk of disease than girls ($P = 0.021$; 29). In Steck et al. (30), it was observed that after stratification by high-risk HLA-DR3/4 genotype, the effect of rs2476601/Trp⁶²⁰ was greater in the non-DR3/4 subgroup (OR 2.07 [95% CI 1.54–2.79]; $P < 0.0001$) compared with DR3/4-positive case subjects (1.44 [0.85–2.45]; $P = 0.18$), using 690 non-Hispanic white case subjects and 515 non-Hispanic white control subjects (mean age at diagnosis of case subjects 11.2 years; range 0.3–54 years).

In the current study, we assessed as comprehensively as

possible given current knowledge of the polymorphism content of the region, from our resequencing and public databases, whether rs2476601/Trp⁶²⁰ is the sole type 1 diabetes susceptibility variant in *PTPN22* and then whether there was any evidence for additional disease risk alleles or haplotypes in the entire 400-kb linkage disequilibrium block containing *PTPN22* and at least six other known genes: *MAGI3*; putative homeodomain transcription factor 1 (*PHTF1*); round spermatid basic protein 1 (*RSBN1*); adaptor-related protein complex 4, $\beta 1$ subunit (*AP4B1*); DNA cross-link repair 1B (*DCLRE1B*); and chromosome 1 open reading frame 178 (*C1orf178*) (Supplementary Fig. 1, available in an online appendix at <http://dx.doi.org/10.2337/db07-1131>).

RESEARCH DESIGN AND METHODS

The type 1 diabetic individuals were recruited as part of the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory British case collection (Genetic Resource Investigating Diabetes), which is a joint project between the Department of Pediatrics, Addenbrooke's Hospital, University of Cambridge, and the Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge. Most affected individuals were <16 years of age at the time of collection (mean age at diagnosis 7.5 years; range 0.5–16 years) and all resided in the U.K. The control samples were obtained from the British 1958 Birth Cohort (B58C; <http://www.b58c.gene.sgu.ac.uk/>), an ongoing study of all people born in the U.K. during 1 week in 1958. All case and control subjects were of self-reported white ethnicity. The relevant research ethics committees approved the study, and written informed consent was obtained from the participants or their parents/guardian for those too young to consent. A total of 2,400 control subjects and 2,800 case subjects had been genotyped as part of the nsSNP GWA study (31), and 3,000 control subjects and 2,000 case subjects were genotyped as part of the Wellcome Trust Case-Control Consortium (WTCCC) GWA study (12), with 1,100 control subjects and 1,700 case subjects being common to both studies. For TaqMan fine-mapping genotyping, a maximum of 7,200 control subjects and 7,600 case subjects were used, encompassing all of the samples from both GWA studies.

PTPN22 sequencing. Polymorphisms in *PTPN22* were identified by resequencing 32 Centre d'Etude du Polymorphisme Humain (CEPH) DNA samples (from Utah residents with northern and western European ancestry) in common with HapMap (www.hapmap.org; 32). The sequencing reactions were performed using Applied Biosystems BigDye chemistry (version 3.1), and the sequences were resolved using an ABI 3700 Genetic Analyzer. Analyses of the sequence traces were performed using the Staden package. All 14 newly discovered SNPs have been submitted to dbSNP (<http://www.ncbi.nlm.nih.gov/SNP>).

Genotyping methods. Information for the *PTPN22* region was taken from the WTCCC study, which used the GeneChip 500K Mapping Array Set (Affymetrix chip) (12), with molecular inversion probe (MIP) technology (Affymetrix) as part of our nsSNP GWA study (31,33) and TaqMan (Applied Biosystems) genotyping, which was carried out in accordance with the manufacturers' protocols. All genotyping data were scored blind to case-control status; TaqMan genotyping was double scored by a second operator to minimize error. We used the clustering method developed by Plagnol et al. (34) to address bias due to differential misclassification in genotyping in large-scale association studies to call genotypes of SNPs incorporated in our nsSNP GWA study (13). No SNPs deviated from Hardy-Weinberg equilibrium ($P > 0.05$). The genotype concordance between TaqMan genotyping and MIP technology for rs2476601/Arg620Trp was 99.67% for control samples (2,383 of 2,391) and 99.07% for case samples (2,986 of 3,014), and the genotype concordance between TaqMan genotyping and GeneChip 500K (Affymetrix) results for rs6679677 was 99.78% for control samples (1,380 of 1,383) and 99.35% for case samples (1,826 of 1,838).

Initially, despite almost perfect linkage disequilibrium ($r^2 = 0.994$ in control subjects) between rs2476601 and rs6679677, we found that we could apparently differentiate between the effects of the two SNPs in stepwise logistic regression analyses. The P value for adding rs2476601 to rs6679677 was 0.010 and, in the reverse regression analysis, adding rs6679677 to rs2476601 ($P = 0.497$), suggesting that rs2476601 was sufficient to model the association. This result was unexpected given the almost perfect linkage disequilibrium, and consequently, when we then merged data from three TaqMan genotyping assays of duplicate genotyping of the case-control collection, genotyped at different times for rs2476601, we identified 36 individuals who were discordant between assays. After resequencing these 36 individuals

for both rs2476601 and rs6679677, we identified 15 of the 36 individuals who had genuinely different genotypes at rs2476601 and rs6679677. The remaining 21 differences were due to random genotyping error (<0.5%) and were excluded. On repeating the logistic regression analysis, as expected, no effects were observed at $P < 0.05$. These results highlight the sensitivity of the logistic regression analysis to genotyping error and missing data.

Statistical analyses. All statistical analyses were performed in either Stata (<http://www.stata.com/>) or the R environment (<http://www.r-project.org/>). Logistic regression models were used for all case-control association tests. We stratified the analysis by place of collection for the case subjects and place of birth for the control subjects for 12 geographical regions of England, Scotland, and Wales (Southwestern, Southern, Southeastern, London, Eastern, Wales, Midlands, North Midlands, Northwestern, East and West Riding, Northern, and Scotland) to exclude the possibility of confounding by geography with little loss of power, given how well the case and control subjects were matched geographically (31). In the logistic regression analysis of a SNP, we performed a 1-d.f. likelihood ratio test to determine whether a 1-d.f. multiplicative allelic effects model or a 2-d.f. genotype effects model (no specific mode of inheritance assumed) was more appropriate (35). We assumed a multiplicative allelic effects model because it was not significantly different from the genotype model. SNPs were modeled as a numerical indicator variable coded 0, 1, or 2, representing the number of occurrences of the minor allele. In the forward logistic regression analysis, we started by assessing the evidence against the most significant SNP being the sole variant in the region (in other words, whether this SNP alone was sufficient to model the association). For the purposes of this analysis, we assumed the genotype (2-d.f.) parameterization for the most associated SNP (A>a) or for any additional SNP with significant independent effects on type 1 diabetes, so genotype risks of A/A and A/a were modeled relative to the a/a genotype. We then used a 1-d.f. test for adding each of the remaining SNPs to the model by assuming multiplicative allelic effects for the additional SNPs.

To correct the most significant P value (observed test statistic $\chi_1^2 = \hat{\theta}$) for the addition of rs735074 in the forward logistic regression analysis for multiple testing, we performed a permutation test to estimate the probability of observing a test statistic at least as large as $\hat{\theta}$ when the null hypothesis is true (that is, when no additional SNP significantly adds to the model). We stratified by rs735074 genotype, randomly assigning case-control status within these strata and then adding in each SNP to the model, including the most associated. After recording the largest test statistic achieved under the null hypothesis, we repeated the permutation of disease status and analyzed 300 times. The corrected P value then equals the number of times that the largest test statistic achieved under the null hypothesis is at least as large as $\hat{\theta}$ divided by the number of permutations.

The six SNP haplotypes were assigned to subjects (case and control subjects separately) using SNPHAP (<http://www-gene.cimr.cam.ac.uk/clayton/software/>) and analyzed using logistic regression models for each subject. Haplotype assignments were weighted by their posterior probabilities, and the Huber/White/Sandwich estimate of variance was used to allow for the nonindependence between the possible haplotype assignments to each individual. In addition to testing the overall significance of the six SNP haplotypes, we tested whether any of the SNPs alone were sufficient to model the association or, in addition to the five SNP haplotypes of the remaining SNPs, were independently associated, the latter providing evidence of allelic heterogeneity. We performed this analysis for each SNP by dropping the SNP from the six SNP haplotypes and testing whether the five SNP haplotypes were independently associated when added to the SNP. We also performed the reverse analysis, testing whether the SNP was independently associated when added to the five SNP haplotypes.

Statistical interaction between loci can be tested using a case-only analysis within a regression model (15,36–38). The *PTPN22* nsSNP, rs2476601/Arg620Trp was put in a regression model as the dependent variable, the SNP genotypes as predictor variables, and geographical region included as strata. A case-only interaction was performed for age at diagnosis and sex. However, for HLA, the number of HLA class II genotypes (>200) that exist in our collection means there are too many parameters to estimate, and therefore we grouped the *HLA-DRB1* and *HLA-DQB1* genotypes together using three different strategies. The first is a 12-group risk-based model derived using recursive partitioning: individuals are categorized into groups according to their class II genotype such that each group is as homogeneous with respect to case-control status as possible (2). The second grouping classifies individuals as high, medium, and low risk according to the haplotype risk estimates published in Table 3 of Koeleman et al. (39): Risks >1 with 95% CIs that did not exceed 1 were classed as high risk; risks with 95% CIs exceeding 1 were medium risk; and risks <1 with 95% CIs that did not exceed 1 were classed low risk (2,39). The third method groups together individuals with *HLA-DRB1*03/HLA-DRB1*04* genotypes versus those without. Genotypes positive

for *HLA-DRB1*0403* and *HLA-DQB1*0301* alleles were put in the non-*DRB1*03/HLA-DRB1*04* group.

We used a log-linear model to estimate the joint effects of *PTPN22* and HLA class II in the logistic regression while maintaining the efficiency of the case-only estimate of the interaction terms (40). We assumed that the HLA and rs2476601 genotypes are conditionally independent in control subjects given geographical region as strata. This is a four-way contingency table of disease status (D; coded 0, control subjects; 1, case subject), rs2476601 genotype (G; coded as a three-level factor—one for each genotype), HLA group (M; coded as a factor with levels for each group, e.g., two levels for the DR3/4 versus non-DR3/4 grouping) and geographical region (S; coded as a 12-level factor). We fitted the model

$$\log(E(F)) = b_0 + b_1 M + b_2 G + b_3 S + b_4 D + b_5 MS + b_6 GS + b_7 DS + b_8 DM + b_9 MDG$$

Here, $E(F)$ is the expected frequency, using a generalized linear model with Poisson family and log link function (40). $b_0, b_1, b_2, b_3, b_5,$ and b_6 parameterize the independent relationship between HLA group and rs2476601 genotype within control subjects, allowing for strata (geographical region). Specifically, b_1 is the log odds of a particular HLA group in control subjects and b_5 gives this within strata. b_2 is the log odds of having rs2476601-susceptible genotype in control subjects, and b_6 gives this within strata. A logistic model is assumed for disease risk such that b_8 and b_9 assess the HLA and HLA**PTPN22* interaction effects; i.e., b_9 gives the log OR of rs2476601 for each HLA group and is reported in Table 4.

Measures of linkage disequilibrium, D' and r^2 , were calculated using the Haploview package (<http://www.broad.mit.edu/mpg/haploview/>) and were subsequently generated and displayed through TIDBase (www.tidbase.org) for Supplementary Fig. 1.

RESULTS

We resequenced all 21 exons of *PTPN22* and 3 kb of 5' and 3' sequence of the gene using DNA from 32 CEPH individuals (32), identifying 33 SNPs and one indel, 14 of which were novel when compared with genome build 36. Twenty-three SNPs had an MAF >0.05 (Supplementary Table 1). Twenty-four SNPs from dbSNP and 22 SNPs discovered from resequencing were incorporated into a GWA study of nsSNPs using the MIP technology (31,33) for genotyping in up to 2,800 type 1 diabetic case samples and 2,400 control samples (Table 1). Nineteen SNPs had $P \leq 1.86 \times 10^{-8}$, with rs2476601/Arg620Trp being the most associated SNP ($P = 6.19 \times 10^{-37}$, OR for minor allele Trp⁶²⁰ 1.98 [95% CI 1.78–2.21]) (Table 1; Supplementary Fig. 1).

To assess the evidence against rs2476601/Arg620Trp being the sole causal variant in the region, we used a logistic regression analysis to test whether rs2476601/Arg620Trp alone was sufficient to model the association. We assumed no specific mode of inheritance for rs2476601/Arg620Trp and used a 1-d.f. test for adding each SNP to the model. We found that when the number of tests was considered, no SNP convincingly added to rs2476601/Arg620Trp in the logistic regression analysis (minimum [uncorrected] $P = 0.0423$ ss73688598; Table 1). We also performed the reverse analysis, adding rs2476601/Arg620Trp to each of the other 46 SNPs; we found that rs2476601/Arg620Trp significantly added to all of the SNPs ($P_{\max} = 1.57 \times 10^{-34}$). This is consistent with rs2476601/Arg620Trp being the sole causal variant in the region.

To assess further the evidence against rs2476601/Arg620Trp being the sole causal variant in the wider *PTPN22* region, we selected 111 SNPs from the 400-kb linkage disequilibrium block containing *PTPN22* and from 200-kb regions on either side of the linkage disequilibrium block that had been genotyped in 2,000 type 1 diabetic case subjects and 3,000 control subjects as part of the WTCCC GWA study (12). We noted that although the nsSNP rs2476601/Arg620Trp was not included on the GeneChip 500K Mapping Array Set (Affymetrix chip) used for geno-

TABLE 1

SNPs from a 140-kb region of chromosome 1p13 containing five genes, including *PTPN22*, genotyped using molecular inversion probe genotyping technology as part of the nsSNP study in 2,800 type 1 diabetic case subjects and 2,400 control subjects

| SNP | Gene | Build 36 | MAF control subjects | MAF case subjects | <i>P</i> value | OR (95% CI) | <i>P</i> value: add to rs2476601 | <i>P</i> value: rs2476601 add to SNP | r^2 |
|---------------|-------------------|-----------|----------------------|-------------------|------------------------|------------------|----------------------------------|--------------------------------------|-------|
| rs1935836A>G | <i>RSBN1</i> | 114110695 | 0.34 | 0.30 | 9.06×10^{-4} | 0.87 (0.80–0.94) | 0.968 | 1.49×10^{-31} | 0.05 |
| rs2040041C>T | <i>RSBN1</i> | 114114146 | 0.26 | 0.23 | 6.24×10^{-4} | 0.85 (0.78–0.93) | 0.536 | 1.67×10^{-31} | 0.04 |
| rs3789598C>A | <i>RSBN1</i> | 114120250 | 0.36 | 0.43 | 1.58×10^{-10} | 1.30 (1.20–1.41) | 0.625 | 5.24×10^{-25} | 0.17 |
| rs3789599T>C | <i>RSBN1</i> | 114124841 | 0.34 | 0.31 | 6.47×10^{-4} | 0.86 (0.79–0.94) | 1 | 2.28×10^{-31} | 0.04 |
| rs2797409T>C | <i>RSBN1</i> | 114131725 | 0.23 | 0.31 | 1.12×10^{-16} | 1.46 (1.33–1.60) | 0.63 | 6.91×10^{-19} | 0.32 |
| rs3789600T>C | <i>RSBN1</i> | 114135745 | 0.42 | 0.48 | 8.05×10^{-9} | 1.26 (1.17–1.37) | 0.661 | 1.12×10^{-26} | 0.13 |
| rs3827733A>G | <i>RSBN1</i> | 114140112 | 0.20 | 0.18 | 0.021 | 0.89 (0.80–0.98) | 0.964 | 1.69×10^{-32} | 0.02 |
| rs1217379G>A | <i>RSBN1</i> | 114145606 | 0.42 | 0.48 | 5.30×10^{-9} | 1.27 (1.17–1.37) | 0.614 | 1.80×10^{-26} | 0.13 |
| rs1217375C>T | <i>RSBN1</i> | 114148427 | 0.28 | 0.25 | 0.00483 | 0.88 (0.81–0.96) | 0.894 | 3.10×10^{-32} | 0.04 |
| rs6537798C>A | <i>RSBN1</i> | 114153229 | 0.43 | 0.48 | 1.86×10^{-8} | 1.26 (1.16–1.36) | 0.77 | 5.83×10^{-27} | 0.13 |
| rs3789604T>G | <i>RSBN1</i> | 114156465 | 0.20 | 0.18 | 0.0052 | 0.86 (0.77–0.95) | 0.401 | 3.82×10^{-31} | 0.02 |
| ss73688607C>G | <i>PTPN22</i> | 114157602 | 0.02 | 0.02 | 0.727 | 1.05 (0.81–1.36) | 0.269 | 2.35×10^{-34} | 0 |
| rs3761935A>C | <i>PTPN22</i> | 114157619 | 0.19 | 0.18 | 0.0349 | 0.89 (0.81–0.99) | 0.882 | 5.57×10^{-33} | 0.02 |
| rs1217411C>T | <i>PTPN22</i> | 114157648 | 0.42 | 0.48 | 4.53×10^{-9} | 1.27 (1.17–1.37) | 0.583 | 2.19×10^{-26} | 0.13 |
| ss73688604A>G | <i>PTPN22</i> | 114157682 | 0.01 | 0.01 | 0.776 | 1.05 (0.74–1.49) | 0.33 | 1.14×10^{-32} | 0 |
| ss73688606A>G | <i>PTPN22</i> | 114158186 | 0.19 | 0.17 | 0.0294 | 0.89 (0.80–0.99) | 1 | 6.04×10^{-33} | 0.02 |
| rs1217412A>G | <i>PTPN22</i> | 114158734 | 0.23 | 0.31 | 3.48×10^{-16} | 1.45 (1.32–1.59) | 0.77 | 2.27×10^{-19} | 0.32 |
| rs1217413A>G | <i>PTPN22</i> | 114159273 | 0.21 | 0.28 | 1.67×10^{-16} | 1.47 (1.34–1.62) | 0.792 | 4.81×10^{-19} | 0.37 |
| rs2476599G>A | <i>PTPN22</i> | 114164982 | 0.28 | 0.25 | 0.00377 | 0.88 (0.80–0.96) | 1 | 3.75×10^{-32} | 0.05 |
| rs3789607T>C | <i>PTPN22</i> | 114167957 | 0.30 | 0.27 | 2.15×10^{-4} | 0.85 (0.77–0.92) | 0.574 | 4.35×10^{-31} | 0.04 |
| rs2476600A>G | <i>PTPN22</i> | 114171257 | 0.42 | 0.48 | 4.62×10^{-9} | 1.27 (1.17–1.37) | 0.586 | 1.97×10^{-26} | 0.13 |
| rs3811021T>C | <i>PTPN22</i> | 114174051 | 0.19 | 0.18 | 0.0327 | 0.89 (0.80–0.99) | 0.896 | 7.87×10^{-33} | 0.02 |
| rs1217395A>G | <i>PTPN22</i> | 114175958 | 0.24 | 0.31 | 2.83×10^{-15} | 1.43 (1.31–1.56) | 0.903 | 4.02×10^{-20} | 0.31 |
| rs2797415C>T | <i>PTPN22</i> | 114178616 | 0.23 | 0.30 | 2.17×10^{-15} | 1.44 (1.31–1.57) | 0.975 | 4.38×10^{-20} | 0.32 |
| rs1970559T>C | <i>PTPN22</i> | 114178671 | 0.28 | 0.25 | 0.00452 | 0.88 (0.80–0.96) | 0.928 | 3.53×10^{-32} | 0.04 |
| rs2476601C>T | <i>PTPN22</i> | 114179091 | 0.09 | 0.17 | 6.19×10^{-37} | 1.98 (1.78–2.21) | — | — | — |
| rs12730735C>T | <i>PTPN22</i> | 114182980 | 0.30 | 0.27 | 2.75×10^{-4} | 0.85 (0.78–0.93) | 0.608 | 3.45×10^{-31} | 0.04 |
| rs1775759A>G | <i>PTPN22</i> | 114190327 | 0.26 | 0.23 | 8.22×10^{-4} | 0.86 (0.78–0.94) | 0.596 | 1.31×10^{-31} | 0.04 |
| rs3765598C>T | <i>PTPN22</i> | 114195986 | 0.19 | 0.18 | 0.0266 | 0.89 (0.80–0.99) | 0.977 | 7.00×10^{-33} | 0.02 |
| ss73688587C>T | <i>PTPN22</i> | 114196212 | 0.02 | 0.02 | 0.718 | 1.05 (0.80–1.37) | 0.261 | 3.40×10^{-34} | 0 |
| rs2476602G>A | <i>PTPN22</i> | 114198478 | 0.27 | 0.25 | 0.00495 | 0.88 (0.80–0.96) | 0.913 | 2.92×10^{-32} | 0.04 |
| ss73688593A>G | <i>PTPN22</i> | 114202113 | 0.02 | 0.02 | 0.563 | 1.08 (0.83–1.41) | 0.215 | 3.23×10^{-34} | 0 |
| rs1217417C>G | <i>PTPN22</i> | 114202427 | 0.27 | 0.23 | 0.00107 | 0.86 (0.78–0.94) | 0.518 | 3.12×10^{-31} | 0.03 |
| rs1217418A>G | <i>PTPN22</i> | 114202754 | 0.43 | 0.48 | 1.37×10^{-8} | 1.26 (1.16–1.36) | 0.726 | 7.60×10^{-27} | 0.13 |
| rs1217419T>G | <i>PTPN22</i> | 114203427 | 0.43 | 0.48 | 8.54×10^{-9} | 1.26 (1.17–1.37) | 0.196 | 1.26×10^{-26} | 0.13 |
| rs2488457G>C | <i>intergenic</i> | 114216891 | 0.19 | 0.27 | 3.63×10^{-17} | 1.50 (1.36–1.65) | 0.826 | 1.47×10^{-18} | 0.38 |
| ss73688598A>G | <i>intergenic</i> | 114217321 | 0.02 | 0.02 | 0.133 | 1.30 (0.92–1.82) | 0.0423 | 2.36×10^{-34} | 0.02 |
| ss73688602C>T | <i>intergenic</i> | 114218588 | 0.01 | 0.01 | 0.996 | 1.00 (0.62–1.62) | 0.727 | 5.67×10^{-34} | 0 |
| rs1235005C>G | <i>intergenic</i> | 114218960 | 0.40 | 0.46 | 7.22×10^{-10} | 1.29 (1.19–1.39) | 0.482 | 1.03×10^{-25} | 0.14 |
| rs1217384G>C | <i>C1orf178</i> | 114221283 | 0.31 | 0.28 | 5.58×10^{-5} | 0.84 (0.77–0.91) | 0.408 | 1.29×10^{-30} | 0.05 |
| rs1746860T>C | <i>C1orf178</i> | 114230863 | 0.32 | 0.29 | 0.00234 | 0.87 (0.80–0.95) | 0.976 | 6.29×10^{-32} | 0.04 |
| rs1217394A>G | <i>intergenic</i> | 114235182 | 0.33 | 0.29 | 7.80×10^{-5} | 0.84 (0.77–0.92) | 0.456 | 1.11×10^{-30} | 0.05 |
| rs10745340T>C | <i>intergenic</i> | 114238493 | 0.35 | 0.41 | 4.94×10^{-11} | 1.32 (1.21–1.43) | 0.416 | 1.23×10^{-24} | 0.16 |
| rs1217401A>G | <i>AP4B1</i> | 114240474 | 0.32 | 0.29 | 9.37×10^{-4} | 0.88 (0.82–0.95) | 0.922 | 1.57×10^{-34} | 0.04 |
| rs3789613C>G | <i>AP4B1</i> | 114244558 | 0.35 | 0.42 | 2.61×10^{-11} | 1.32 (1.22–1.43) | 0.371 | 2.14×10^{-24} | 0.16 |
| rs3761936T>C | <i>DCLRE1B</i> | 114251185 | 0.16 | 0.24 | 2.77×10^{-19} | 1.58 (1.43–1.75) | 0.612 | 2.36×10^{-16} | 0.4 |

MAF in control samples; OR for minor allele. Map positions from NCBI build 36. Last three columns, single locus test *P* value when SNP added to rs2476601, single locus test *P* value when rs2476601 added to SNP in logistic regression analyses, and pairwise r^2 of SNP with rs2476601.

typing by the WTCCC GWA study, rs2476601/Arg620Trp was in perfect linkage disequilibrium ($r^2 = 1$) with rs6679677, an intergenic SNP between the genes *PHTF1* and *RSBN1*, in the 3,000 British control subjects; this near-perfect linkage disequilibrium has been reported by others (12,17). We genotyped rs6679677 and rs2476601/Arg620Trp in the full case-control collection, with 7,500 case subjects and 7,200 control subjects. We found that we could not distinguish between the effects of the two SNPs on type 1 diabetes (Table 2). We found that only 15 of 14,487 samples had discordant genotypes at both SNPs,

giving an r^2 of 0.996 in case subjects and 0.994 in control subjects; we ruled out random genotyping error by resequencing these individuals and verifying their genotypes (see RESEARCH DESIGN AND METHODS).

From the WTCCC GWA study, 13 SNPs had $P < 4.28 \times 10^{-8}$, with rs6679677 being the most associated ($P = 1.32 \times 10^{-23}$, OR for minor allele A 1.88 [95% CI 1.66–2.13]) (Supplementary Table 2; Supplementary Fig. 1). We performed a forward logistic regression analysis on these 110 WTCCC SNPs from the wider *PTPN22* region to test whether there was any evidence for a variant associated

TABLE 2
Single-locus test results for rs2476601/Arg620Trp and rs6679677 in the case-control collection

| | Case subjects | Control subjects | OR (95% CI) | P value |
|-----------|----------------|------------------|------------------|------------------------|
| rs2476601 | | | | |
| C | 12,225 (82.22) | 12,754 (90.42) | | |
| T | 2,643 (17.78) | 1,352 (9.58) | 2.01 (1.89–2.19) | 2.11×10^{-87} |
| C/C | 5,030 (67.67) | 5,762 (81.70) | | |
| C/T | 2,165 (29.12) | 1,230 (17.44) | 2.00 (1.85–2.18) | |
| T/T | 239 (3.21) | 61 (0.86) | 4.47 (3.35–5.96) | 4.33×10^{-86} |
| rs6679677 | | | | |
| C | 12,232 (82.27) | 12,756 (90.43) | | |
| A | 2,636 (17.73) | 1,350 (9.67) | 2.03 (1.89–2.19) | 3.21×10^{-87} |
| C/C | 5,033 (67.70) | 5,764 (81.72) | | |
| C/A | 2,166 (29.14) | 1,228 (17.41) | 2.01 (1.85–2.18) | |
| A/A | 235 (3.16) | 61 (0.87) | 4.45 (3.33–5.93) | 6.74×10^{-86} |

Data are *n* (%). OR for minor allele. $r^2 = 0.994$ between these two SNPs in control subjects.

with type 1 diabetes independently of rs6679677. After correcting the most associated additional SNP, there was no evidence of any additional independent effects (minimum uncorrected $P = 0.00854$ for rs735074 located in the flanking linkage disequilibrium block; corrected $P = 0.243$), including the previously associated *MAGI3* SNP rs1343125 (22) (uncorrected $P = 0.907$), the Japanese-associated SNP rs2488457 (25) (uncorrected $P = 0.866$), and the rheumatoid arthritis-associated SNP rs3789604 (24) (uncorrected $P = 0.401$) (Table 1; Supplementary Table 2).

We genotyped an additional five SNPs with TaqMan because they were not covered by either GWA scan: two SNPs (rs1310182 and rs2488458) associated with rheumatoid arthritis from Carlton et al. (24), the rare nsSNP (ss73688585/K750N) associated with type 1 diabetes from Onengut-Gumuscu et al. (23), and two SNPs (rs3789608 and ss73688608) found from our resequencing. Although we found evidence for associations between type 1 diabetes and rs1310182 ($P = 9.33 \times 10^{-13}$), rs2488458 ($P = 3.46 \times 10^{-34}$), and ss73688608 ($P = 1.80 \times 10^{-7}$), these

results were not independent of rs2476601/Arg620Trp (uncorrected $P = 0.840$, 0.343, and 0.702, respectively). No evidence of an association was found with either rs3789608 or ss73688585/K750N (uncorrected $P = 0.0615$ and 0.117, respectively).

We also tested the six SNP haplotypes from the Graves' study by Heward et al. (28) (rs2488458, rs12730735, rs2476601, rs1310182, rs1217413, and rs3811021) in 3,129 case subjects and 3,633 control subjects for which we obtained complete data for all six SNPs. The haplotypes that we identified correlated with the 10 common haplotypes identified in the Graves' study (28) and also the previous rheumatoid arthritis study (24) (Table 3). We found that the susceptible haplotype (haplotype 2), the only haplotype to carry the T/Trp allele of rs2476601/Arg620Trp, was associated with disease, consistent with the two previous studies (Table 3). We did not, however, obtain any evidence for protective haplotypes in contrast to the Graves' study, in which haplotype 3 was protective (28), and in the rheumatoid arthritis study, haplotypes 5 and 6 were protective (24) (Table 3). In our study, the

TABLE 3
Six marker *PTPN22* haplotype frequencies in 3,129 type 1 diabetic case subjects and 3,633 control subjects

| Number* | Haplotype | Frequency in case subjects | Frequency in control subjects | OR (95% CI) |
|---------|-----------|--|-------------------------------|--|
| 1 | AACTCT | 0.109 | 0.123 | 1.00 (0.88–1.15) |
| 2 | AATCTC | 0.173 | 0.086 | 2.25 (1.95–2.59) |
| 3 | AACTTT | 0.025 | 0.022 | 1.31 (1.00–1.72) |
| 4 | GACTTC | 0.180 | 0.190 | 1.04 (0.92–1.17) |
| 5 | GGCCTT | 0.269 | 0.303 | 1.00 (reference) |
| 6–10 | GACCTT | 0.245 | 0.271 | 1.03 (0.92–1.15) |
| Number* | Haplotype | <i>P</i> value adding SNP to haplotype | | <i>P</i> value adding haplotype to SNP |
| 1 | _ACTCT | (rs2488458) 0.469 | | 5.36×10^{-19} |
| 2 | A_TTCT | (rs12730735) 0.669 | | 1.28×10^{-28} |
| 3 | AA_TTT | (rs2476601) 9.15×10^{-24} | | 0.184 |
| 4 | GAC_TC | (rs1310182) 0.784 | | 1.37×10^{-25} |
| 5 | GGCC_T | (rs1217413) 0.120 | | 3.51×10^{-20} |
| 6–10 | GACCT_ | (rs3811021) 0.397 | | 1.63×10^{-31} |

*Haplotype number adheres to that reported by Carlton et al. (24) and Heward et al. (28) and is presented in the same order. The order of the SNPs in the haplotypes are rs2488458, rs12730735, rs2476601, rs1310182, rs1217413, and rs3811021. The haplotypes were assigned to case and control subjects separately. Haplotype 3 was the one reported by Heward et al. (28) that was protective for Graves' disease and could not be explained by the alleles of rs2476601/Arg620Trp. Haplotype 5 was the most frequent and taken as reference. The haplotype frequencies do not add to 1 because of unlisted, rare haplotypes.

TABLE 4
Joint effects of HLA class II genes and the nsSNP rs2476601/Arg620Trp of *PTPN22*

| HLA risk | Case subjects* | RR for rs2476601 and HLA class II | | | OR for T allele of rs2476601 |
|---|----------------|-----------------------------------|-------|-------|------------------------------|
| | | C/C | C/T | T/T | |
| Model A, rpart HLA class II risk groups ($P_{\text{Interaction}} = 1.36 \times 10^{-4}$) | | | | | |
| High | 1,705 (70.78) | 1.000 | 1.71 | 2.923 | 1.710 |
| | 289 (12.00) | 0.281 | 0.554 | 1.093 | 1.973 |
| | 105 (4.36) | 0.111 | 0.369 | 1.225 | 3.322 |
| Medium | 88 (3.65) | 0.062 | 0.137 | 0.305 | 2.228 |
| | 53 (2.20) | 0.055 | 0.116 | 0.244 | 2.108 |
| | 30 (1.25) | 0.051 | 0.082 | 0.132 | 1.604 |
| Low | 38 (1.58) | 0.016 | 0.045 | 0.125 | 2.817 |
| | 45 (1.87) | 0.005 | 0.016 | 0.046 | 2.949 |
| Model B, HLA class II genotype risk groups ($P_{\text{Interaction}} = 0.0017$) | | | | | |
| High | 1,500 (72.29) | 1.000 | 1.763 | 3.109 | 1.763 |
| Medium | 542 (26.12) | 0.101 | 0.233 | 0.540 | 2.314 |
| Low | 33 (1.59) | 0.006 | 0.020 | 0.064 | 3.258 |
| Model C, HLA-DRB1*03/HLA-DRB1*04 genotypes ($P_{\text{Interaction}} = 8.04 \times 10^{-4}$) | | | | | |
| DR3/4 | 851 (34.61) | 1.000 | 1.592 | 2.527 | 1.592 |
| Non-DR3/4 | 1,608 (65.39) | 0.043 | 0.091 | 0.19 | 2.100 |

Data are n (%) unless otherwise indicated. Three modeling strategies for the HLA loci were adopted: A, a 12-group model derived using recursive partitioning (rpart) (2); B, a 3-group model in which HLA class II genotypes were classified as high risk, neutral, and low risk for type 1 diabetes (2,39); and C, a 2-group model in which all non-DR3/4 genotypes (including HLA-DRB1*0403 and HLA-DQB1*0301) were compared with DR3/4. We report the RRs of the joint effect of *PTPN22* (rs2476601) and HLA class II. Note that the effect of *PTPN22* is greater in the low-risk HLA groups, e.g., the non-DR3/4 group (OR 2.10), compared with high-risk HLA groups, e.g., the DR3/4 group (OR 1.59). Interaction of HLA class II with *PTPN22* (rs2476601) was tested in the case subjects ($P_{\text{interaction}}$). *Frequencies do not add to 100% because rare groups are not listed.

Arg⁶²⁰ protective allele-bearing haplotype 3 was slightly increased in frequency in the case subjects, but this is unlikely to be a true result given the multiplicity of tests (OR 1.31 [95% CI 1.00–1.72], $P = 0.048$). More importantly, we noted that the control frequency for haplotype 3 in the Graves' study is unexpectedly higher (0.062) than for the control frequency in our study (0.022) or the rheumatoid arthritis study (0.034 and 0.032) (24) and may account for the apparent protective haplotype effect reported previously (28) (Supplementary Table 3). Furthermore, in contrast to the Graves' study but consistent with the rheumatoid arthritis study, the SNPs used in the haplotype analysis (rs2476601, rs2488458, rs12730735, rs1310182, rs1217413, and rs3811021) were all associated with type 1 diabetes. None of these SNP associations were found to be independent, as expected, of rs2476601/Arg620Trp (Table 3).

Finally, we tested for evidence of a statistical interaction of rs2476601/Arg620Trp with age at type 1 diabetes diagnosis and sex in the case subjects, obtaining no convincing support for interaction in 7,443 case subjects ($P = 0.106$ and 0.219, respectively). We did, however, find evidence for statistical interaction with the HLA class II genotypes for all three MHC class II groupings considered in up to 2,459 case subjects (Table 4), with a minimum $P = 1.36 \times 10^{-4}$. Specifically, the relative risk of rs2476601/Trp⁶²⁰ is higher in low-risk HLA genotypes, e.g., non-DR3/4 (RR 2.10), than the high-risk DR3/4 (1.59).

DISCUSSION

After extensive resequencing and genotyping of the *PTPN22* region, we have found that we are unable to distinguish statistically between two associated SNPs, rs2476601/Trp⁶²⁰ and rs6679677, located 73 kb apart. rs6679677 is located in the intergenic region between

PHTF1 and *RSBN1* and does not have an obvious potential functional effect, and it does not reside in a transcriptional factor binding site, CpG island, or highly conserved sequence block. The TFMATRIX web tool (<http://www.cbrc.jp/htbin/nph-tfsearch>), however, predicts the disruption of a binding site of promoter CCAAT binding factors by this SNP. The nsSNP rs2476601/Arg620Trp changes the amino acid residue 620 from Arg to Trp in the encoded LYP protein, a well-established suppressor of T-cell activation (7,41). Recently, it has been shown that Arg620Trp is a functional residue in the Pro-rich motif in LYP that binds the SH3 domain of CSK; Trp⁶²⁰ fails to bind CSK (19). Trp⁶²⁰ may, however, lead to a gain of function (more negative regulation of the immune system) such that cells in vitro were reported to be less responsive in individuals with the susceptibility allele rather than the expected overactivation due to loss of *PTPN22* function with this minor, disease-associated allotype.

We have provided evidence of statistical interaction between HLA class II genotypes and a non-MHC HLA susceptibility locus in agreement with two previous studies from Hermann et al. (29) and Steck et al. (30). This interaction, which is less than multiplicative, cannot be interpreted biologically (42), and all that we are able to conclude from it is that despite the relative risk of *PTPN22* being higher in low-risk HLA genotypes, the combined risk of diabetes for individuals carrying high-risk HLA and *PTPN22* genotypes is higher than that for those carrying low risk HLA genotypes and high-risk *PTPN22* genotypes. This finding is also supported by the previous observation that non-MHC loci had smaller relative risks in affected sibpairs (who are "DR3/4" rich) than in single affected sibling families or case/control collections and that epistatic models could explain such observations (2,43). In

comparison with *PTPN22*, we note that for *INS*, there is no evidence for a statistical interaction with HLA class II genotypes ($P \geq 0.0976$) for all three MHC class II groupings considered in the same sample set (not shown).

Our results support the hypothesis that rs2476601/Arg620Trp is the most likely causal variant for type 1 diabetes and that for this disease, in large sample sizes, there is no evidence for allelic heterogeneity. However, the possibility still remains that an alternative or as yet unidentified variant(s) could have a role in disease susceptibility, as it is not known how many more SNPs are in the region because the entire linkage disequilibrium region has not been resequenced. For example, 73 kb telomeric of rs2476601/Arg620Trp, SNP rs6679677 is in perfect linkage disequilibrium with the *PTPN22* functional candidate SNP and genetically and statistically, therefore, is as good a positional candidate as *PTPN22* rs2476601/Arg620Trp. Our failure to obtain any evidence for allelic heterogeneity or haplotype effects in type 1 diabetes for this chromosome region suggests that for type 1 diabetes, previous studies (23,25,26) could have been false positives, especially because much smaller sample sizes were studied. We cannot comment on the validity of additional variants in rheumatoid arthritis and Graves' disease (22,24,28), except that any result should be replicated in additional datasets. Nevertheless, it does appear that the haplotype effect in Graves' disease (28) may be due to fluctuation in control haplotype frequencies rather than a genuine difference between case and control frequencies.

With the recent availability of next-generation sequencing technologies (44–46), the essential next step in the localization and disease association analysis of all of the variation in this region of chromosome 1p13 will be made more feasible, being less labor-intensive and costly. It will not be surprising if additional variants are found that are in near perfect linkage disequilibrium with rs2476601/Arg620Trp, as is the case for SNP rs6679677, making them genetically plausible candidates for this type 1 diabetes locus. Furthermore, it is possible that an unknown variant exists in the region that is not in complete linkage disequilibrium with rs2476601 and is more strongly associated with type 1 diabetes. Further genetic studies, including populations with different genetic diversity from European populations, combined with targeted functional studies will confirm unequivocally rs2476601/Arg620Trp as the sole type 1 diabetes locus in the region. Notwithstanding this possibility, our results thus far provide no evidence for allelic heterogeneity and support an etiological role for the rs2476601/Trp⁶²⁰ allele.

ACKNOWLEDGMENTS

This work was funded by the Juvenile Diabetes Research Foundation International, the Wellcome Trust, and the National Institute for Health Research, Cambridge Biomedical Centre.

We gratefully acknowledge the participation of all of the patients and control subjects. We acknowledge use of the DNA from the British 1958 Birth Cohort collection, funded by the Medical Research Council and Wellcome Trust. We also thank the Avon Longitudinal Study of Parents and Children laboratory in Bristol and the British 1958 Birth Cohort team, including S. Ring, R. Jones, M. Pembrey, W. McArdle, D. Strachan, and P. Burton, for preparing and providing the control DNA samples. We thank colleagues at Affymetrix for help and advice in genotyping and T.

Willis, M. Faham, and P. Hardenbol for the MIP technology. We thank Oliver Burren and Barry Healy for bioinformatics support. DNA samples were prepared by K. Bourget, S. Duley, M. Hardy, S. Hawkins, S. Hood, E. King, T. Mistry, A. Simpson, S. Wood, P. Lauder, S. Clayton, F. Wright, and C. Collins.

This study makes use of data generated by the WTCCC. A full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk.

REFERENCES

- 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
- Nejentsev S, Howson JM, Walker NM, Szeszeko J, Field SF, Stevens HE, Reynolds P, Hardy M, King E, Masters J, Hulme J, Maier LM, Smyth D, Bailey R, Cooper JD, Ribas G, Campbell RD, Clayton DG, Todd JA: Wellcome Trust Case Control Consortium: Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. *Nature*, 450:887–892, 2007
- 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
- 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
- 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, Goris FK, Tosi R, Pozzilli P, Todd JA: The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes: Belgian Diabetes Registry. *Hum Mol Genet* 5:1075–1080, 1996
- 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
- Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellicchia M, Eisenbarth GS, Comings D, Mustelin T: A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes. *Nat Genet* 36:337–338, 2004
- Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget K, Plagnol V, Field S, Atkinson M, Clayton DG, Wicker LS, Todd JA: Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the *IL2RA* region in type 1 diabetes. *Nat Genet*, 39:1074–1082, 2007
- Vella A, Cooper JD, Lowe CE, Walker N, Nutland S, Widmer B, Jones R, Ring SM, McArdle W, Pembrey ME, Strachan DP, Dunger DB, Twells RC, Clayton DG, Todd JA: Localization of a type 1 diabetes locus in the *IL2RA/CD25* region by use of tag single-nucleotide polymorphisms. *Am J Hum Genet* 76:773–779, 2005
- Qu HQ, Montpetit A, Ge B, Hudson TJ, Polychronakos C: Toward further mapping of the association between the *IL2RA* locus and type 1 diabetes. *Diabetes* 56:1174–1176, 2007
- Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ, Guja C, Ionescu-Tirgoviste C, Widmer B, Dunger DB, Savage DA, Walker NM, Clayton DG, Todd JA: A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (*IFIH1*) region. *Nat Genet* 38:617–619, 2006
- WTCCC: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:661–678, 2007
- Todd JA, Walker NM, Cooper JD, Smyth DJ, Gough SC, Plagnol V, Bailey R, Nejentsev S, Field SF, Payne F, Lowe CE, Szeszeko 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
14. Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT, Grabs R, Casalunovo T, Taback SP, Frackelton EC, Lawson ML, Robinson LJ, Skraban R, Lu Y, Chiavacci RM, Stanley CA, Kirsch SE, Rappaport EF, Orange JS, Monos DS, Devoto M, Qu HQ, Polychronakos C: A genome-wide association study identifies *KIAA0350* as a type 1 diabetes gene. *Nature* 448:591–594, 2007
 15. 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
 16. 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
 17. 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
 18. Zheng W, She JX: Genetic association between a lymphoid tyrosine phosphatase (*PTPN22*) and type 1 diabetes. *Diabetes* 54:906–908, 2005
 19. 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
 20. 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
 21. Aarnisalo J, Treszl A, Svec P, Marttila J, Oling V, Simell O, Knip M, Korner A, Madacsy L, Vasarhelyi B, Ilonen J, Hermann R: Reduced CD4+T cell activation in children with type 1 diabetes carrying the *PTPN22*Lyp620Trp variant. *J Autoimmun*, 2008. DOI: 10.1016/j.jaut.2008.01.001
 22. Steer S, Abkevich V, Gutin A, Cordell HJ, Gendall KL, Merriman ME, Rodger RA, Rowley KA, Chapman P, Gow P, Harrison AA, Highton J, Jones PB, O'Donnell J, Stamp L, Fitzgerald L, Iliev D, Kouzmine A, Tran T, Skolnick MH, Timms KM, Lanchbury JS, Merriman TR: Genomic DNA pooling for whole-genome association scans in complex disease: empirical demonstration of efficacy in rheumatoid arthritis. *Genes Immun* 8:57–68, 2007
 23. Onengut-Gumuscu S, Buckner JH, Concannon P: A haplotype-based analysis of the *PTPN22* locus in type 1 diabetes. *Diabetes* 55:2883–2889, 2006
 24. 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
 25. 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
 26. Viken MK, Olsson M, Flam ST, Forre O, Kvien TK, Thorsby E, Lie BA: The *PTPN22* promoter polymorphism –1123G>C association cannot be distinguished from the 1858C>T association in a Norwegian rheumatoid arthritis material. *Tissue Antigens* 70:190–197, 2007
 27. 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
 28. 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
 29. Hermann R, Lipponen K, Kiviniemi M, Kakko T, Veijola R, Simell O, Knip M, Ilonen J: Lymphoid tyrosine phosphatase (*LYP/PTPN22*) Arg620Trp variant regulates insulin autoimmunity and progression to type 1 diabetes. *Diabetologia* 49:1198–1208, 2006
 30. Steck AK, Liu SY, McFann K, Barriga KJ, Babu SR, Eisenbarth GS, Rewers MJ, She JX: Association of the *PTPN22/LYP* gene with type 1 diabetes. *Pediatr Diabetes* 7:274–278, 2006
 31. Clayton DG, Walker NM, Smyth DJ, Pask R, Cooper JD, Maier LM, Smink LJ, Lam AC, Ovington NR, Stevens HE, Nutland S, Howson JM, Faham M, Moorhead M, Jones HB, Falkowski M, Hardenbol P, Willis TD, Todd JA: Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 37:1243–1246, 2005
 32. Consortium TIH: A haplotype map of the human genome. *Nature* 437:1299–1320, 2005
 33. Hardenbol P, Yu F, Belmont J, Mackenzie J, Bruckner C, Brundage T, Boudreau A, Chow S, Eberle J, Erbilgin A, Falkowski M, Fitzgerald R, Ghose S, Iartchouk O, Jain M, Karlin-Neumann G, Lu X, Miao X, Moore B, Moorhead M, Namsaraev E, Pasternak S, Prakash E, Tran K, Wang Z, Jones HB, Davis RW, Willis TD, Gibbs RA: Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. *Genome Res* 15:269–275, 2005
 34. Plagnol V, Cooper JD, Todd JA, Clayton DG: A method to address differential bias in genotyping in large-scale association studies. *PLoS Genet* 3: e74, 2007
 35. Cordell HJ, Clayton DG: A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 70:124–141, 2002
 36. Chapman J, Clayton D: Detecting association using epistatic information. *Genet Epidemiol* 31:894–909, 2007
 37. Cordell HJ: Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11:2463–2468, 2002
 38. Weinberg CR, Umbach DM: Choosing a retrospective design to assess joint genetic and environmental contributions to risk. *Am J Epidemiol* 152:197–203, 2000
 39. Koelman BP, Lie BA, Undlien DE, Dudbridge F, Thorsby E, de Vries RR, Cucca F, Roep BO, Giphart MJ, Todd JA: Genotype effects and epistasis in type 1 diabetes and HLA-DQ trans dimer associations with disease. *Genes Immun* 5:381–388, 2004
 40. Umbach DM, Weinberg CR: Designing and analysing case-control studies to exploit independence of genotype and exposure. *Stat Med* 16:1731–1743, 1997
 41. Hill RJ, Zozulya S, Lu YL, Ward K, Gishizky M, Jallal B: The lymphoid protein tyrosine phosphatase *Lyp* interacts with the adaptor molecule *Grb2* and functions as a negative regulator of T-cell activation. *Exp Hematol* 30:237–244, 2002
 42. Cordell HJ, Barratt BJ, Clayton DG: Case/pseudocontrol analysis in genetic association studies: a unified framework for detection of genotype and haplotype associations, gene-gene and gene-environment interactions, and parent-of-origin effects. *Genet Epidemiol* 26:167–185, 2004
 43. Howson JM, Barratt BJ, Todd JA, Cordell HJ: Comparison of population- and family-based methods for genetic association analysis in the presence of interacting loci. *Genet Epidemiol* 29:51–67, 2005
 44. Bennett ST, Barnes C, Cox A, Davies L, Brown C: Toward the 1,000 dollars human genome. *Pharmacogenomics* 6:373–382, 2005
 45. Bentley DR: Whole-genome re-sequencing. *Curr Opin Genet Dev* 16:545–552, 2006
 46. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM: Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437:376–380, 2005