Protein Tyrosine Phosphatase-1B Gene *PTPN1*

Selection of Tagging Single Nucleotide Polymorphisms and Association With Body Fat, Insulin Sensitivity, and the Metabolic Syndrome in a Normal Female Population

Nicola J. Spencer-Jones,¹ Xiaoling Wang,² Harold Snieder,²,³ Tim D. Spector,³ Nicholas D. Carter,¹ and Sandra D. O’Dell¹

Protein tyrosine phosphatase-1B negatively regulates leptin and insulin signaling, potentially contributing to hormonal resistance. We selected six tagging single nucleotide polymorphisms (SNPs) representing 18 common variants in the protein tyrosine phosphatase-1B gene (*PTPN1*) and tested their effect on serum leptin, body fat, and measures of insulin sensitivity and the metabolic syndrome in a large sample of normal female Caucasian twins (n = 2,777; mean age, 47.4 ± 12.5 years) from the St. Thomas’ U.K. Adult Twin Registry. SNP rs718049 was significantly associated with waist circumference (P = 0.008) and central fat (P = 0.035) and also with Avignon’s insulin sensitivity index (SiM) (P = 0.007), fasting insulin (P = 0.004), fasting glucose (P = 0.022), triglyceride (P = 0.023), and systolic blood pressure (P = 0.046). SNPs rs2282146 and rs1885177 were associated with SiM (P = 0.049 and P = 0.013, respectively), and 1484insG was associated with triglyceride (P = 0.029). A risk haplotype (7.3%) was associated with lower SiM (P = 0.036) and a protective haplotype (5.2%) with higher SiM (P = 0.057), with mean values in homozygotes differing by >1 SD (P = 0.003). The protective haplotype also showed lower triglyceride (P = 0.045) and lower systolic blood pressure (P = 0.006). Fine mapping analyses predicted significant associations with SiM and fasting insulin for several ungenotyped SNPs. *PTPN1* variants appear to contribute to central fat and metabolic syndrome traits, secondary to their effect on insulin sensitivity. Diabetes 54:3296–3304, 2005
in an attempt to locate potential functional sites (19).

Finally, we performed a fine mapping analysis that uses the larger sample with those found in previous studies. For ovarian endometriosis, we compared any population studied by Palmer et al. (15) and a 50% larger sample (16) that is almost three times the size of the Hispanic CR Hendrickson-McCreadie twins study (20). We used the package genassoc (21) to select a tSNP set that predicts the minimal set of haplotypes that covers 90% of existing haplotypes and sets of tSNPs resolving the common haplotypes were selected at a significance level of 0.05. We used two approaches to identify an optimal subset of tSNPs. The first one, devised by Strom et al. (20), selects a subset of SNPs as tSNPs by optimizing their prediction of common haplotypes of all SNPs genotyped in the 94 unrelated individuals. It considers a measure of association (\(R^2\)) between the true and predicted number of copies (0, 1, or 2) of each possible common haplotype carried by a randomly sampled subject, in which the prediction is based on knowledge of the tSNPs. The program tgenopoly was used to calculate the following measures: the number of SNPs, the number of exons, the number of haplotypes, and the number of times a particular haplotype was observed in the dataset. The latter is based on prediction of single SNP loci.

### SNPs (tSNPs) effectively capturing information of all common variants by taking into account patterns of linkage disequilibrium across the gene (19,20). We therefore used a multistep design to select a set of tSNPs representing all considered common variants in PTPN1 and tested the effect of these tSNPs, individually and as haplotypes, on a range of variables characterizing body composition, insulin sensitivity and the metabolic syndrome in a sample of 50% larger than the size of the Hispanic population studied by Palmer et al. (15) and 50% larger than the combined Chinese and Japanese cohorts of Olivier et al. (17). Where possible, we compared any specific single SNP and haplotype associations in our larger sample with those found in previous studies. Finally, we performed a fine mapping analysis that uses the tSNPs to predict potential associations between the remaining (un-genotyped) SNPs and the outcome variables, in an attempt to locate potential functional sites (19).

### RESEARCH DESIGN AND METHODS

The St. Thomas’ U.K. Adult Twin Registry (Twins U.K.) comprises unselected, mostly female volunteers ascertained from the general population through national media campaigns in the U.K. (21). Means and ranges of quantitative variables are as described by de Lange et al. (22). Serum insulin was measured by immunoassay (Abbott Laboratories, Maidenhead, U.K.) and glucose on an Ektachem 700 multichannel analyzer using an enzymatic colorimetric slide assay (Johnson and Johnson Clinical Diagnostics Systems, Amersham, U.K.). A subsample of twins underwent an oral glucose tolerance test for which glucose and insulin levels were measured before and 2 h after a 75-g oral glucose load. Levels of HDL cholesterol and triglycerides were measured using a Cobas Farma machine (Roche Diagnostics) (23). SBP and diastolic blood pressure (DBP) were measured twice using an automated cuff sphygmomanometer (OMRON HEM713C) and averaged over the two readings. Informed consent was obtained from participants before they entered the study. The protocol was approved by the local research ethics committee. We used two approaches to identify an optimal subset of tSNPs. The first one, devised by Strom et al. (20), selects a subset of SNPs as tSNPs by optimizing their prediction of common haplotypes of all SNPs genotyped in the 94 unrelated individuals. It considers a measure of association (\(R^2\)) between the true and predicted number of copies (0, 1, or 2) of each possible common haplotype carried by a randomly sampled subject, in which the prediction is based on knowledge of the tSNPs. The program tgenopoly was used to calculate the following measures: the number of SNPs, the number of exons, the number of haplotypes, and the number of times a particular haplotype was observed in the dataset. The latter is based on prediction of single SNP loci.

### Genotyping in cohort.

Six tSNPs (rs6067484, rs1851717, rs2282146, rs718050, rs3787339, and rs1854485) were selected and genotyped in the complete cohort by Pyrosequencing (Biotage, Uppsala, Sweden). Genotyping accuracy was assessed by inclusion of duplicates (pairs of monozygotic twins) in the arrays and negative controls (water blanks) on each plate. Primers and PCR conditions for SNP validation and tSNP selection are given in online appendix 2 (available at http://diabetes.diabetesjournals.org). The rs number and relative position of these 19 SNPs are shown in Fig. 1.

### Association analysis.

Regular association analyses were performed using GEEs. Analyses were done separately for each of the SNPs and followed up by haplotype analyses. For individual SNP association analyses, we first tested the following gene-environment interaction models: M(SNP) × environmental factor, M(SNP) × environment (SNP), M(SNP) × environmental factor (SNP), and M(SNP) × environmental factor (SNP). In the presence of a positive interaction, the full model was tested by adding each of the interaction terms.

### Analytical approach.

The main purposes of our analyses were to select a set of tSNPs representing the common variants in PTPN1 and to test the effect of these tSNPs, individually and/or as haplotypes, on a range of variables characterizing body composition, insulin sensitivity and the metabolic syndrome in a sample of 50% larger than the size of the Hispanic population studied by Palmer et al. (15) and 50% larger than the combined Chinese and Japanese cohorts of Olivier et al. (17). Where possible, we compared any specific single SNP and haplotype associations in our larger sample with those found in previous studies. Finally, we performed a fine mapping analysis that uses the tSNPs to predict potential associations between the remaining (un-genotyped) SNPs and the outcome variables, in an attempt to locate potential functional sites (19).

### Table 1

General characteristics of subjects

<table>
<thead>
<tr>
<th>Variable</th>
<th>n</th>
<th>Means ± SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age (years)</td>
<td>2,777</td>
<td>47.4 ± 12.5</td>
</tr>
<tr>
<td>Postmenopausal (%)</td>
<td>2,454</td>
<td>47.7</td>
</tr>
</tbody>
</table>

Obesity-related variables

- Leptin (ng/ml) 2,777 16.5 ± 11.9
- BMI (kg/m²) 2,758 24.7 ± 4.4
- Weight (kg) 2,759 65.3 ± 11.8
- Waist circumference (cm) 2,707 78.3 ± 10.2
- Total fat (kg) 2,718 23.4 ± 8.8
- Total fat (%) 2,677 35.6 ± 8.0
- Central fat (kg) 2,695 1.33 ± 0.73
- Central fat (%) 2,695 31.2 ± 11.5

Insulin sensitivity

- Fasting glucose (mmol/l) 2,319 4.49 ± 0.56
- Fasting insulin (mmol/l) 1,902 6.90 ± 5.40
- 2-h glucose (mmol/l) 759 5.18 ± 1.10
- 2-h insulin (μU/ml) 759 33.3 ± 23.1
- SiM (10⁻¹¹ L pmol⁻¹ mmol⁻¹) 759 89.5 ± 68.5

Other components of the metabolic syndrome

- Triglyceride (mmol/l) 2,485 1.27 ± 0.80
- HDLc (mmol/l) 2,606 1.55 ± 0.39
- SBP (mmHg) 2,586 120.9 ± 17.6
- DBP (mmHg) 2,586 76.1 ± 11.1

*Number of subjects (842 monozygotic and 1,935 dizygotic) with leptin data and genotype data on at least one tSNP.

†Nonfasting sample, mostly female volunteers ascertained from the general population through national media campaigns in the U.K. (21). Means and ranges of quantitative variables are as described by de Lange et al. (22). Serum insulin was measured by immunoassay (Abbott Laboratories, Maidenhead, U.K.) and glucose on an Ektachem 700 multichannel analyzer using an enzymatic colorimetric slide assay (Johnson and Johnson Clinical Diagnostics Systems, Amersham, U.K.). A subsample of twins underwent an oral glucose tolerance test for which glucose and insulin levels were measured before and 2 h after a 75-g oral glucose load. Levels of HDL cholesterol and triglycerides were measured using a Cobas Farma machine (Roche Diagnostics) (23). SBP and diastolic blood pressure (DBP) were measured twice using an automated cuff sphygmomanometer (OMRON HEM713C) and averaged over the two readings. Informed consent was obtained from participants before they entered the study. The protocol was approved by the local research ethics committee.

### Genotyping for SNP validation and tSNP selection.

Nineteen validated SNPs from the National Center for Biotechnology Information database (http://www.ncbi.nlm.nih.gov/SNP) and one SNP, 1484insG, reported by Di Paolo et al. (13) were selected for validation in the Twins U.K. cohort by genotyping eight random unrelated subjects. Among these 20 SNPs, only rs718050 was not polymorphic. The other 19 SNPs were further genotyped in a sample of 94 unrelated subjects for tSNP selection, using PCR and restriction. This sample size is three times that suggested by one permutation study (26), which indicates that genotyping 25–32 unphased individuals is sufficient to select tSNPs. SNP rs7187339 (minor allele frequency [MAF] of 1%) turned out to be uninformative and was excluded from all further analyses. Primers and PCR conditions for SNP validation and tSNP selection are given in online appendix 1 (available at http://diabetes.diabetesjournals.org). The rs number and relative position of these 19 SNPs are shown in Fig. 1.
as outlined by Zaykin et al. (29), with the probabilities of haplotype pairs estimated by PHASE 2.0 software (30). Obesity-related variables included leptin, weight, BMI, waist circumference, total fat mass, percent total fat, central fat mass, and percent central fat. Factor analysis was used to combine strongly correlated indexes of obesity into two measures: one for general obesity (serum leptin, BMI, weight, total fat mass, and percent total fat) and one for central obesity (waist circumference, central fat mass, and percent central fat). To reduce the likelihood of identifying false-positive associations, results of single variables characterizing obesity were confirmed by these two combined scores. We used two indexes of insulin sensitivity: fasting insulin and Avignon’s insulin sensitivity index (SiM). SiM is calculated based on both fasting and 2-h insulin and glucose data (31) and was calculated according to the following formulae: SiM = \frac{10^8}{\text{fasting insulin} \times \text{fasting glucose}} \times VD; \text{SIH2} = \frac{10^8}{\text{2-h insulin} \times \text{2-h glucose}} \times VD) and VD = 150 ml/kg × body wt. Both fasting insulin (r = 0.68) and SiM (r = 0.92) are highly correlated with insulin sensitivity in the normal population. SiM is also a good predictor of diabetes, especially in Caucasians (32). For the metabolic syndrome measures, we included five variables: triglycerides, HDL cholesterol, waist circumference, fasting glucose, and SBP, based on the clinical definition of the insulin receptor substrate 9 (metabolic syndrome) as specified in the third report of the Adult Treatment Panel (ATPIII) of high blood cholesterol (16). To control for population stratification, dizygotic twin pairs discordant for genotype were also used in sibling transmission-disequilibrium test (TDT) association analysis to confirm the results of regular association tests as described elsewhere (27). Preliminary analyses were performed using STATA 8 (StataCorp, College Station, TX). Phenotypes significantly (P < 0.05) deviating from normal were log transformed to obtain normal distributions before analysis. Hardy-Weinberg equilibrium was tested by a χ² test with 1 df in one twin of each pair chosen

![Genomic map of PTPN1 gene with locations and MAFs of the validated SNPs genotyped in 94 subjects. Exons; □, untranslated regions. MAFs are indicated next to corresponding rs numbers.](image-url)
The same set of six tSNPs, rs6067484, rs1885177, rs2282146, and rs718049 had a frequency of 4.2%, and three showed a frequency of 1.9%. Accurately predicted both common haplotypes and unmeasured loci. Minimum values of $R^2_H$ and $R^2_L$ were 0.884 for haplotype 3 and 0.882 for rs10485614, respectively.

Table 3 shows the genotype and allele frequencies of the six tSNPs in the whole cohort. The total number of subjects genotyped for each polymorphism varied slightly and was somewhat lower than 2,356, the total number of twins genotyped (i.e., one monozygotic and both dizygotic twins of each pair). This was due to unsuccessful amplification of the target sequences for some samples. None of the loci showed deviation from Hardy-Weinberg equilibrium. The inferred haplotype frequencies of the six tSNPs in the whole cohort are also shown in Table 3. The same seven most common haplotypes observed in the 94 subjects (Table 2) were also identified by the six tSNPs, although the actual estimates of haplotype frequencies and corresponding order varied somewhat between the full cohort and the tSNP selection sample ($n = 94$).

Table 4 presents the results of individual SNP analyses. For the SNPs of lowest frequency, rs2282146 and rs718049, only dominant models were tested, whereas for the four more common SNPs, the 2-df codominant model was first tested, followed by a dominant and a recessive model in the presence of a significant association. With the exception of the effect of rs718049 on waist circumference, none of the main effects of the SNPs on leptin and other obesity-related variables reached statistical significance. We therefore did not include these variables in Table 4. Regular association ($P = 0.074$) and sibling TDT tests ($P = 0.025$) for the combined central obesity score confirmed the significant associations with waist circumference (data not shown). No significant effects of haplotypes on any of the obesity-related variables were found (data not shown). Carriers of the minor allele of rs718049 had significantly higher waist circumference, lower SiM, and higher fasting insulin and glucose. Excepting glucose, these associations were maintained in sibling TDT. SNP rs2282146 was significantly associated with SiM ($P = 0.049$), and this association changed to borderline significance in sibling TDT ($P = 0.062$), with carriers of the minor allele having lower SiM. Under the recessive model, carriers of the minor allele of rs1885177 had significantly lower SiM ($P = 0.013$) and maintained significance in sibling TDT ($P = 0.039$). Two SNPs showed significant associations with other components of the metabolic syndrome. Again, the strongest relationship was shown with rs718049, associated with serum triglyceride and SBP, with all except triglyceride remaining significant in sibling TDT. SNP 1484insG was associated with triglyceride ($P = 0.029$) only in the regular test.

Table 5 shows haplotype frequencies and results of haplotype association tests. Although the overall test did not reach significance ($P = 0.085$), haplotype 4 was significantly associated with lower SiM ($P = 0.036$), i.e., confers risk of insulin resistance, and haplotype 5 was borderline-significantly associated with higher SiM ($P = 0.057$), i.e., is protective. The mean levels of SiM were 0.46 SD lower for haplotype 4 homozygotes and 0.56 SD higher for haplotype 5 homozygotes compared with those homozygous for the most common haplotype 1, a difference of $\geq 1$ SD ($P = 0.003$). The explained variance of SiM by the PTNP1 haplotypes was 1.92%. The risk and protective haplotypes are discriminated by rs2282146 and rs718049, reflecting the strongest individual significant associations.

<table>
<thead>
<tr>
<th>TABLE 2 Continued</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs1885177</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
</tbody>
</table>

at random to prevent inflated significance. Pairwise linkage disequilibrium coefficients were calculated using GOLD and reported as $D^{\prime}$ and $r^2$ (33).
TABLE 3
Genotype, allele, and haplotype distributions

<table>
<thead>
<tr>
<th>tSNP*</th>
<th>Genotype</th>
<th>Total</th>
<th>Hardy-Weinberg test†</th>
<th>MAF (95% CI) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>11</td>
<td>12</td>
<td>22</td>
</tr>
<tr>
<td>rs6067484</td>
<td>1,093</td>
<td>743</td>
<td>126</td>
<td>1,962</td>
</tr>
<tr>
<td>rs1885177</td>
<td>619</td>
<td>1,109</td>
<td>485</td>
<td>2,213</td>
</tr>
<tr>
<td>rs2282146</td>
<td>1,713</td>
<td>286</td>
<td>12</td>
<td>2,011</td>
</tr>
<tr>
<td>rs718049</td>
<td>813</td>
<td>951</td>
<td>282</td>
<td>2,046</td>
</tr>
<tr>
<td>rs3787348</td>
<td>516</td>
<td>998</td>
<td>410</td>
<td>1,924</td>
</tr>
<tr>
<td>1484insG</td>
<td>1,697</td>
<td>185</td>
<td>7</td>
<td>1,889</td>
</tr>
</tbody>
</table>

Haplotype‡ | rs6067484 | rs1885177 | rs2282146 | rs718049 | rs3787348 | 1484insG | Frequency (SE)

1 | 1 | 1 | 1 | 1 | 2 | 1 | 0.432 (0.0019)
2 | 2 | 2 | 1 | 1 | 1 | 1 | 0.242 (0.0016)
3 | 1 | 1 | 1 | 1 | 1 | 1 | 0.075 (0.0019)
4 | 1 | 2 | 2 | 1 | 1 | 1 | 0.073 (0.0015)
5 | 1 | 2 | 1 | 1 | 1 | 1 | 0.052 (0.0024)
6 | 1 | 2 | 1 | 1 | 1 | 2 | 0.052 (0.0017)
7 | 1 | 2 | 1 | 2 | 1 | 1 | 0.028 (0.0021)
8 | 1 | 1 | 1 | 2 | 2 | 1 | 0.016 (0.0010)
9 | 2 | 2 | 1 | 1 | 1 | 1 | 0.008 (0.0010)

*Includes one monozygotic and both dizygotic twins for each pair. †Tested in one of each monozygotic pair (one of each dizygotic pair and all singleton dizygotic twins). ‡Includes both dizygotic pairs and one of monozygotic pairs with genotype data on at least three tSNPs (n = 2,320).

with SiM in the regular and sibling TDT analyses. The protective effects of haplotype 5 were also observed for SBP ($P = 0.006$) and triglyceride ($P = 0.045$) levels. The $\beta$-coefficient for haplotype 5 on SBP was $-7.2$, and on triglycerides was $-0.17$, i.e., SBP levels were 7.2 mmHg lower and triglyceride levels 0.17 mmol/l lower for protective haplotype 5 homozygotes compared with common haplotype 1 homozygotes. The explained percentages of variance in SBP and triglyceride levels by the $PTPN1$ haplotypes were 0.33% and 0.63%, respectively.

The approach we used to select an optimal set of tSNPs (19) can also predict which of those SNPs not typed in the full cohort might also show association with the trait. Figure 2 shows the fine mapping results (including the ungenotyped SNPs) based on prediction of single SNP loci from the six tSNPs. Associations of both tSNPs and predicted values for unmeasured SNPs with SiM, fasting insulin, and fasting glucose under the additive genetic model are indicated by $-\log(P)$. Predicted associations for lipids, SBP, and waist circumference were found to be uninformative and are not shown.

DISCUSSION

Our previous work (21) and previous work by others (35) have shown that variation in leptin levels, total adiposity, and central abdominal fat mass is under strong genetic influence. However, in this study, neither the single SNP analyses nor the haplotype analyses of $PTPN1$ showed any significant effect on leptin, weight, BMI, or total fat variables. That is, our results do not support a major role for $PTPN1$ variants in body weight regulation, because downstream effects of the hypothesized impairment in leptin signal transduction such as increases in serum leptin and measures of total body fat were not observed. This conclusion is especially important because the design of our study allows us to exclude the most likely alternative explanations for this largely negative result. First, our study did not suffer from a lack of statistical power to detect SNPs with small effects. The current study had 80% ($\alpha = 0.05$) power to detect a biallelic quantitative trait locus explaining as little as 0.5% of the variance (36). Second, strong linkage disequilibrium across $PTPN1$ allowed us to effectively capture common variation in this gene by the selected set of tSNPs, making it unlikely that we may have missed any major variants that are either causal or in strong linkage disequilibrium with a causal locus. The third advantage of our study is the comprehensive and accurate measurement of the phenotypes of interest: general and central obesity. In addition to serum leptin levels and anthropometric variables, central and total body fat was assessed by dual energy X-ray absorptiometry, which is a more objective and reliable method of assessing adiposity than, for example, BMI because it allows discrimination between fat, muscle mass, bone, and vital organs (23). Information on all phenotypes was subsequently used to generate one combined score for general obesity and one for central obesity.

Although we observed a small but significant effect of one $PTPN1$ variant on central obesity, we suggest this effect is mediated through the inhibition of insulin signaling by PTP-1B, because central fat is a major determinant of insulin sensitivity whereas total fat is the major determinant of leptin levels (23). We found that carriers of the minor alleles of rs1885177, rs2282146, and rs718049 had significantly lower SiM and higher fasting insulin and glucose, associations maintained in sibling TDT. Effects of $PTPN1$ on insulin sensitivity and type 2 diabetes risk have been reported by two recent studies (15,37). Palmer et al. (15) found that 20 SNPs with MAFs $>0.1$ in a single haplotype block covering the $PTPN1$ genomic sequence showed significant association with the insulin sensitivity index $S_i$ in Hispanic Americans ($P = 0.003–0.044, n = 811$, based on codominant models). Although the study of Palmer et al. is a family study of 55 pedigrees, unlike us, no TDT results were reported to exclude the possibility of false-positive associations due to stratification or admixture. Four of our tSNPs were among those genotyped by Palmer et al.: rs1885177, rs718049, rs3787348, and 1484insG.
Our results and those of Palmer et al. are consistent in that carriage of the A allele of rs1885177 [rare allele (2) in our cohort and marginally the more frequent allele (1) in that of Palmer et al.] was associated with lower Sim and S5 as was the allele (2) of rs718049. SNP rs2282146, associated with Sim by us (P = 0.049), had an MAF of ~6% in the Hispanic population and did not show a significant influence on S5. Palmer et al. (15) found that their 20 SNPs were all significantly associated with fasting glucose (P = <0.001–0.029), however, we only found an association with rs718049 (P = 0.045).

Palmer et al. (15) analyzed eight SNPs tagging haplotypes with frequency >10%, having completed genotyping of all 35 SNPs in their 811 subjects. We reduced effort considerably by first genotyping candidate tSNPs in a small subsample, which confirmed the strong linkage disequilibrium in the region that emerged from the larger dataset of Palmer et al. and enabled tSNP selection. The eight tSNPs of Palmer et al. included three of our tSNPs, rs718049, rs3787348, and 1484insG, and another three that we had already ascertained were tagged by rs718049: rs3787345, rs7545118, and rs2282147. We had not attempted to validate their remaining two SNPs. Palmer et al. (15) defined one risk, one protective, and one neutral haplotype with respect to S5. We can only directly compare the haplotypes constructed by common SNPs rs718049, rs3787348, and 1484insG in the Hispanic sample and our sample with respect to S5 and Sim. In fact, rs718049 and rs3787348 alone discriminate the Hispanic haplotypes. Haploype rs718049-rs3787348-1484insG 1-1-1 is protective in the Hispanics and in our twin population. The Hispanic neutral haplotype is also identical to our neutral (most common) haplotype: 1-2-1. The Hispanic risk haplotype 2-2-1 is slightly different; in the twins it is 2-1-1. Therefore,
overall, our haplotype results are consistent with those of Palmer et al. with respect to insulin sensitivity. However, unlike them, we found no haplotypes associated with fasting glucose. The same haplotypes associated with \( S_i \) and fasting glucose by Palmer et al. (15) have independently been shown to be associated with type 2 diabetes risk and protection in Caucasian type 2 diabetic subjects (37).

Olivier et al. (17) found that of six SNPs in strong linkage disequilibrium spanning \( PTPN1 \), one SNP (rs16995294, not genotyped by us) was significantly associated with hypertension in 1,553 Chinese and Japanese subjects from 672 families. In a TDT, they identified two 6-SNP haplotypes significantly overtransmitted in hypertensive subjects (\( P < 0.0001 \)). Again, although direct comparisons are not possible, our finding of a protective haplotype for SBP in Caucasians provides further evidence in support of an influence of \( PTPN1 \) on development of hypertension. In the absence of an obvious etiological role, this effect is likely to be secondary to the influence of the gene on insulin sensitivity.

Finally, comparing our study of the 1484insG SNP with that of Di Paola et al. (13) in 477 normoglycemic women from two Italian regions, we found no significant differences in quantitative phenotypes between wild-type and 1484insG carriers, including those for which they found positive associations: SBP and DBP (higher in one group only) and elevated 1-h and 2-h insulin levels (measured in the other group). They found that 1484insG was associated with higher serum triglycerides in Italian men but not in women, however, we found an association in women, which remained significant after adjustment for BMI (\( P = 0.029 \)).

Variation in the coding region of \( PTPN1 \) is relatively rare (only two nonsynonymous changes listed on the National Center for Biotechnology Information database), and associations of the synonymous coding or intronic SNPs studied by us are likely to result from linkage disequilibrium with functional variants, rather than be causal themselves. We can use the prediction rules computed during the choice of the marker subset to target specific loci for further study (19). With regard to SiM and fasting insulin, we predict additional significant associations with rs3787345 and rs2038526 in intron 4, rs754118 in intron 5, and rs2282147 in intron 7. Although it remains difficult to predict where the causal locus or loci might be due to the strong linkage disequilibrium across this gene, Fig. 2 indicates that certain regions of the gene can be excluded and others need to be considered when looking for causal loci and, as such, provides useful direction. The large

---

**TABLE 5**

<table>
<thead>
<tr>
<th>Haplotype</th>
<th>SNP no.</th>
<th>Frequency</th>
<th>SiM (pmol/l)</th>
<th>SBP (mmHg)</th>
<th>TG (mmol/l)</th>
</tr>
</thead>
<tbody>
<tr>
<td>rs6067484</td>
<td>rs1885177</td>
<td>rs2282146</td>
<td>rs718049</td>
<td>rs3787348</td>
<td></td>
</tr>
<tr>
<td>1*</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>5</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>6</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>7</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

**Overall**

*Estimates are contrasts with this most common haplotype. TG, triglyceride.

---

**FIG. 2.** Associations of \( PTPN1 \) SNPs (both unmeasured SNPs and tSNPs) with SiM, fasting insulin, and fasting glucose under the additive genetic model. Associations of tSNPs and predicted values for unmeasured SNPs are indicated by \(-10 \log (P)\). Points above the dotted line representing \(-10 \log 0.05\) represent SNP associations found or predicted to be significant at the \( P = 0.05 \) level.
introns in PTPN1 and the flanking regions may contain unknown regulatory elements; for example, the 3′- untranslated region 1484insG variant was shown to stabilize PTP1-B mRNA in vitro (13). We are now validating SNPs in the PTPN1 5′-flanking region and will explore whether they are in strong linkage disequilibrium with the existing tSNPs (especially rs2282146), which could be the causal locus or loci. Our findings and those of others relating PTPN1 variation to insulin resistance, lipid profile, central obesity, and hypertension suggest that this gene has a significant influence on progression of the metabolic syndrome. Demonstration of an etiologic PTPN1 variant is eagerly awaited and would confirm PTP1-B as an influential factor.

ACKNOWLEDGMENTS

This study has received funding from British Heart Foundation Project Grant PG/04/028. The Twin Research and Genetic Epidemiology Unit has received support from the Wellcome Trust, Arthritis Research Campaign, the Chronic Disease Research Foundation, and the European Union 5th Framework Programme GenomEU twin no. QLG2-CT-2002-01254.

This research was conducted within the network of the London IDEAS Genetic Knowledge Park, and used the St. George’s, University of London Medical Biocentre.

REFERENCES

15. Palmer ND, Bento JL, Mychaleckyj JC, Langefeld CD, Campbell JK, Norris JM, Haffner SM, Bergman RN, Bowden DW: Insulin Resistance Atherosclerosis Study (IRAS) Family Study: Association of protein tyrosine phosphatase 1B gene polymorphisms with measures of glucose homeostasis in Hispanic Americans. The Twin Research and Genetic Epidemiology Unit has received support from the European Union 5th Framework Programme GenomEU twin no. QLG2-CT-2002-01254.
32. Tabor HK, Risch NJ, Myers RM: Candidate-gene approaches for studying

