CONCISE REPORT

Investigation of genetic variation across the protein tyrosine phosphatase gene in patients with rheumatoid arthritis in the UK

Anne Hinks, Steve Eyre, Anne Barton, Wendy Thomson, Jane Worthington

Ann Rheum Dis 2007;66:683–686. doi: 10.1136/ard.2006.060459

Objective: To investigate single-nucleotide polymorphisms (SNPs) across the *PTPN22* gene region in a UK cohort of patients with rheumatoid arthritis (RA), to look for evidence of disease associations independent of the well-characterised R620W variant (rs2476601).

Methods: 951 RA cases in the UK satisfying American Rheumatism Association (ARA) criteria and 448 population controls were genotyped for 11 SNPs across the *PTPN22* gene region using the Sequenom MassArray MassEXTEND technology. Allele, genotype and estimated haplotype frequencies of cases and controls were compared.

Results: In addition to the R620W (rs2476601) SNP, three SNPs were associated with RA in this study. The sole haplotype on which the associated T allele of R620W occurred was associated with RA; no other haplotypes showed a significant difference in frequencies between RA cases and controls.

Conclusion: In contrast with a study of American patients with RA no evidence of association with *PTPN22* independent of the well-characterised R620W variant was found, suggesting that in these patients this variant alone explains the association with the *PTPN22* gene.

The importance of both genetic and environmental risk factors for common human diseases is well accepted and different approaches are being used to identify and characterise how such factors influence the disease process. For diseases with an autoimmune pathology, such as rheumatoid arthritis (RA), the association with genes in the human leucocyte antigen (HLA) region have been established for many years. Although this region seems to harbour the major susceptibility locus for many of these diseases, it is also recognised that other non-HLA genes may influence many aspects of the disease process from susceptibility to severity and outcome to treatment responsiveness.

A large number of candidate genes have been investigated for autoimmune diseases and many associations reported; however, the vast majority of these have turned out to be false-positive results that have not been replicated in subsequent investigations. The finding that emerged recently of an autoimmune susceptibility polymorphism in the protein tyrosine phosphatase gene (*PTPN22*) is, however, quite different. A functional polymorphism (R620W) in *PTPN22*, a negative regulator of T-cell activation, was first associated with type 1 diabetes (T1D) by researchers investigating a number of tyrosine phosphatases.¹ At much the same time, a pioneering large-scale candidate gene investigation in RA screened large numbers of non-synonymous single-nucleotide polymorphisms (SNPs), and revealed the strongest association with the same R620W *PTPN22* variant as that described in the study by Bottini *et al.*² This study contained evidence of replication, as the same association was observed in two large and independent cohorts of

RA cases. Subsequently the R620W association has been replicated many times over in cohorts of patients with RA from different countries around the world (reviewed in Hinks *et al.*³). The only failures to replicate have come from Asian populations, in which this particular variant is not seen.⁴

The significance of the *PTPN22* R620W variant extends beyond RA and T1D to a number of other autoimmune and rheumatological conditions, including systemic lupus erythematosus (SLE), juvenile idiopathic arthritis (JIA), autoimmune thyroid disease and Addison's disease. No association has been observed with multiple sclerosis, psoriasis, psoriatic arthritis or Crohn's disease.⁵ What is perhaps surprising is that, with just two exceptions, all these studies have restricted their investigation of the *PTPN22* gene to a single polymorphism (R620W), even those that failed to detect an association with this variant. This approach has been justified by the fact that the R620W appears to have functional significance, in that, the variant form results in reduced binding to c-src tyrosine kinase (Csk), as shown in both the early RA and T1D studies. It was initially suggested that this would result in loss of negative regulation of T-cells and lead to a hyper-responsiveness; however, the most recent experiments seem to suggest quite the opposite; that the R620W form results in a gain of function resulting in enhanced suppression of T-cell activation.⁶

The possibility that other polymorphisms in *PTPN22* may be important in RA, either independently or in combination with R620W, has been investigated in only one study to date. Carlton *et al.*⁷ identified 15 novel SNPs in *PTPN22* and genotyped these together with an additional 22 SNPs in two cohorts of RA cases and controls. The R620W variant was found to reside on a single haplotype, but this haplotype did not explain fully the difference between RA cases and controls. Two SNPs on a different common haplotype were found to be independently associated with RA, suggesting that the association between *PTPN22* and RA may be more complex than was initially realised. However, the novel associations did not reach the levels of statistical significance seen for R620W and require validation in more cohorts. Thus, in this study we have investigated 11 SNPs from across the *PTPN22* gene region, including those associated with RA by Carlton *et al.*, in a large cohort of RA cases in the UK, in which we have previously reported association with the *PTPN22* R620W variant.⁸

MATERIALS AND METHODS

Subjects

DNA was available for 951 RA cases obtained from two main sources: the Arthritis and Rheumatism Council (ARC) National

Abbreviations: HLA, human leucocyte antigen; LD, linkage disequilibrium; *PTPN22*, protein tyrosine phosphatase gene; RA, rheumatoid arthritis; RF, rheumatoid factor; SNP, single-nucleotide polymorphism; T1D, type 1 diabetes

Figure 1 Linkage disequilibrium plot across the *PTPN22* gene in controls.

Repository for families with RA, and clinics within the Greater Manchester area of northern England. For patients obtained through the national repository, only one affected case per family was selected at random for investigation. All RA cases had a disease that satisfied the 1987 American College of Rheumatology (ACR) criteria⁹ modified for genetic studies,¹⁰ and 73% were female. Rheumatoid factor (RF) status was ascertained using a particle agglutination test, and a positive result was classified as a titre of 1 in >40. In all, 75% of the cases were RF positive and 67% had erosive joint disease. *HLA-DRB1* genotypes were determined using a commercially available semiautomated PCR-sequence specific oligonucleotide probe typing technique (INNO-LiPA; Abbott Laboratories, Maidenhead, UK). It was found that 47% and 31% carried one or two copies of the shared epitope, respectively. This frequency reflects those observed in a hospital-based series of RA cases.

DNA was available for 448 control subjects, who were recruited from blood donors and from general practitioners (GP) registers. All patients and controls were of UK Caucasoid ethnic origin, were recruited with ethical committee approval and provided informed consent.

SNPs selected for genotyping

As there was some redundancy due to strong linkage disequilibrium (LD) between the 37 SNPs genotyped by Carlton *et al.*,⁷ we chose to investigate the seven haplotype-tagging SNPs (SNPs 1, 18, 20, 22, 27, 35 and 36), three SNPs (SNP 2, 32 and 37) that were potentially functional, and one SNP (SNP 21) that showed the strongest allelic association in the Carlton sample set 1. In total, 11 SNPs from the Carlton paper were selected for genotyping.

Genotyping protocol

SNPs were genotyped using the Sequenom MassArray genotyping technology, according to the manufacturer's instructions (<http://www.sequenom.com/seq-genotyping.html>).

Statistical analysis

All SNPs were tested for Hardy–Weinberg equilibrium in cases and controls. Association of the *PTPN22* SNPs was tested using the χ^2 test implemented in STATASTATA V.9.

Pairwise LD measures of D' and LD correlation coefficient r^2 were calculated and plotted on a graph. Haplotypes were inferred using the expectation–maximisation algorithm which was implemented in HelixTree V 4.1 (Golden Helix, Bozeman, Montana, USA). Haplotype frequencies were compared between cases and controls using the χ^2 test implemented in STATA. Each haplotype was compared with all the others combined.

RESULTS

Allele and genotype frequencies for all the *PTPN22* SNPs were in Hardy–Weinberg equilibrium in the RA cohort and in the control population.

As we have shown previously, the rs2476601 T allele was associated with RA (allelewise $p = 5 \times 10^{-8}$). Three additional SNPs showed evidence of association: rs2488458 (SNP2), rs1217388 (SNP32) and rs1217413 (SNP35) (table 1); however all these SNPs are in LD with rs2476601 (pairwise $r^2 > 0.5$) (fig 1), suggesting that these are not independent associations. This result contrasts with the data presented by Carlton *et al.*, in which the six SNPs in addition to R620W associated with RA (SNPs 2, 18, 21, 27, 32, 35) were not strongly correlated with the R620W variant ($r^2 < 0.5$).

Haplotype analysis

Out of the 10 common haplotypes reported by Carlton *et al.*, eight were identified in the UK RA and control cohorts at frequencies >1% (table 2), and the most common haplotype in UK cases and controls was haplotype 5, as in the American data. The 620W risk T allele is carried only on haplotype 2, and this was the only haplotype to occur at a different frequency in cases and controls in this study ($p = 6.6 \times 10^{-6}$). Haplotype 1, which differs from haplotype 2 only at the R620W variant, carrying the non-associated C allele, is not associated with RA ($p = 0.08$). Furthermore, unlike the Carlton study, we found no significant differences between the haplotype frequencies of the RA cases and controls other than haplotype 2. In addition, when all chromosomes carrying the risk T allele (W620) were removed and the haplotype frequencies calculated, there were no haplotypes showing a significantly different allele frequency in RA cases compared with controls (data not shown). These observations, together with the fact that the only significantly associated SNPs were those in LD with the R620W variant, suggest that, in the UK population, the association between RA and *PTPN22* is fully explained by association with the 620W (T allele) of rs2476601.

DISCUSSION

The well established association of the R620W variant of *PTPN22* with RA and other autoimmune diseases has demonstrated an important role for this gene in determining disease susceptibility. Almost all investigations to date have focused on the R620W SNP (rs2476601) and studies are currently underway to elucidate the mechanism by which this amino acid substitution influences the function of the *PTPN22* gene and its role in the disease process. It is, however, clear from many examples that once an association between disease and an allelic variant has been identified, further studies of the gene reveal additional variants, often with a more minor but nonetheless significant role (eg, *HLA-DRB1* and RA).¹¹ The study by Carlton *et al.* suggests this to be the case for *PTPN22* and RA; screening of 37 SNPs in the *PTPN22* gene region identified two SNPs on a single common haplotype that was associated with RA independently of rs2476601. The effect was, however, weaker than that observed for the R620W SNP and requires validation in independent cohorts.

Table 1 PTPN22 SNP allele frequencies and allelic association with RA

Carlton SNP id	Db SNP id	Position	Frequency in cases	Frequency in controls	OR (95% CI)	Allelic p value	Genotype frequencies in RA cases	Genotype frequencies in controls	Genotype p value
1	rs1217414	114124709	0.234	0.241	0.96 (0.79 to 1.17)	0.71	11 = 0.58 (530) 12 = 0.37 (332) 22 = 0.05 (47)	11 = 0.56 (235) 12 = 0.39 (163) 22 = 0.05 (19)	0.63
2	rs2488458	114118493	0.302	0.258	1.24 (1.02 to 1.51)	0.03	12 = 0.42 (368) 22 = 0.09 (83) 11 = 0.50 (439)	12 = 0.40 (156) 22 = 0.06 (22) 11 = 0.54 (209)	0.05
18	rs12760457	114101790	0.295	0.304	0.96 (0.79 to 1.16)	0.64	12 = 0.41 (366) 22 = 0.09 (77) 11 = 0.84 (760)	12 = 0.46 (179) 22 = 0.07 (29) 11 = 0.84 (340)	0.32
20	rs11102685	114093675	0.083	0.081	1.04 (0.76 to 1.43)	0.81	12 = 0.15 (137) 22 = 0.01 (7) 11 = 0.50 (431)	12 = 0.15 (61) 22 = 0.01 (2) 11 = 0.47 (169)	0.85
21	rs12730735	114093499	0.297	0.303	0.97 (0.8 to 1.18)	0.79	12 = 0.40 (347) 22 = 0.10 (82) 11 = 0.68 (612)	12 = 0.45 (160) 22 = 0.08 (28) 11 = 0.83 (342)	0.3
22	rs2476601*	114089610	0.171	0.090	2.08 (1.58 to 2.76)	<0.001	12 = 0.29 (259) 22 = 0.03 (23) 11 = 0.27 (221)	12 = 0.16 (66) 22 = 0.01 (4) 11 = 0.29 (101)	<0.001
27	rs1310182	114085545	0.475	0.463	1.05 (0.88 to 1.26)	0.59	12 = 0.51 (424) 22 = 0.22 (180) 11 = 0.48 (431)	12 = 0.50 (175) 22 = 0.21 (75) 11 = 0.52 (208)	0.78
32	rs1217388	114076518	0.309	0.270	1.21 (1.0 to 1.46)	0.04	12 = 0.42 (374) 22 = 0.10 (90) 11 = 0.54 (463)	12 = 0.42 (168) 22 = 0.06 (24) 11 = 0.63 (240)	0.05
35	rs1217413	114069792	0.274	0.209	1.42 (1.16 to 1.76)	<0.001	12 = 0.38 (332) 22 = 0.08 (71) 11 = 0.67 (582)	12 = 0.33 (126) 22 = 0.04 (17) 11 = 0.64 (247)	0.003
36	rs3811021	114068705	0.177	0.193	0.90 (0.72 to 1.12)	0.33	12 = 0.31 (270) 22 = 0.02 (19) 11 = 0.67 (576)	12 = 0.33 (129) 22 = 0.03 (10) 11 = 0.65 (226)	0.6
37	rs3789604	114066984	0.181	0.197	0.90 (0.72 to 1.13)	0.36	12 = 0.29 (245) 22 = 0.04 (32)	12 = 0.31 (110) 22 = 0.04 (14)	0.61

Bold entries refer to significant associations (p values <0.05).
RA, rheumatoid arthritis; SNP, single-nucleotide polymorphism.
*R620W variant.

Table 2 *PTPN22* haplotype frequencies for RA cases and controls

Carlton haplotype number												Case	Control	p Value
	1	2	18	20	21	22*	27	32	35	36	37	Haplotype frequency	Haplotype frequency	
1	C	A	G	A	A	C	T	C	C	T	A	0.095902	0.122275	0.08
2	C	A	G	A	A	T	T	C	C	T	A	0.169277	0.084716	<0.001
3	C	A	G	A	A	C	T	C	T	T	A	0.024969	0.044717	0.052
4	C	G	G	A	A	C	T	T	T	C	C	0.16553	0.190022	0.37
5	C	G	A	A	G	C	C	T	T	T	A	0.281104	0.299069	0.45
6	T	G	G	A	A	C	C	T	T	T	A	0.139789	0.143126	0.97
8	T	G	G	G	A	C	C	T	T	T	A	0.075949	0.072456	0.93

*R620W variant

We investigated 11 SNPs from across the *PTPN22* gene region in order to analyse all haplotypes defined by Carlton *et al*, and attempted to replicate their findings with RA in a cohort in which we have previously observed a strong association with rs2476601. We observed associations with rs2476601 and with three additional SNPs (rs2488458, rs1217388, rs1217413); however, all these SNPs are correlated and the associations are unlikely to be independent. The two independent associations identified by Carlton *et al* (SNP 27 (rs1310182) and the highly correlated SNP 36 (rs3811021) and SNP 37 (rs3789604)) were not associated with RA in our cohorts ($p = 0.59$, $p = 0.33$ and $p = 0.36$, respectively).

Patterns of LD across the *PTPN22* gene region were similar to those previously described, although individual correlation values were higher in the UK cohort. The haplotype frequency distributions were the same in the US and UK cohorts; however, in the UK data, only haplotype 2, characterised by the associated T allele of R620W variant, was associated with RA. Thus in contrast with Carlton *et al*, in this study, we detected no *PTPN22* SNP associations with RA independent of the well-characterised R620W.

There are no obvious differences between the RA cases used in this study and that of Carlton *et al*, although no direct comparison is possible based on published data. In both investigations, the cohorts were derived from multicase families and sporadic cases. Our study had 80% power to detect an OR = 1.45 at $p < 0.05$; therefore, we had adequate power to detect an effect similar to that observed by Carlton *et al*. However, it is possible that a more modest association with the SNPs investigated could have been missed.

SNP screening across the *PTPN22* gene region has also been carried out in T1D and psoriasis. In patients with psoriasis, a disease not associated with the 620W variant, association was reported with two haplotypes spanning the two genes adjacent to *PTPN22*: round sperm basic protein 1 (*RSBP1*) and the neighbouring gene putative homeodomain transcription factor 1 (*PHTF1*).¹² Interestingly rs3811021 (SNP36) and rs3789604 (SNP37) lie in the 3' untranslated region (3'UTR) of *PTPN22* and 5' of the neighbouring gene, *RSBP1*, respectively. A recent systematic screen of SNPs across the *PTPN22* gene region in Asian patients with T1D, who did not have the R620W polymorphism, revealed association with a promoter SNP (rs2488457). This SNP is in LD with the R620W variant in Caucasians and more strongly associated with diabetes, leading the authors to speculate that this might be the causal variant.⁴

Although we found no association with *PTPN22* SNPs in UK patients with RA independent of the R620W variant, it is intriguing that the region around *PTPN22* may harbour additional disease-associated alleles, particularly in light of the conclusion that the R620W variant does not fully explain the linkage peak on chromosome 1p13 in the North American

Rheumatoid Arthritis Consortium (NARAC) RA-affected sibling pair families.¹³

Authors' affiliations

Anne Hinks, Steve Eyre, Anne Barton, Wendy Thomson, Jane Worthington, Arthritis Research Campaign Epidemiology Unit, University of Manchester, Manchester, UK

Funding: This work was funded by the Arthritis Research Campaign.

Correspondence to: Dr A Hinks, Arthritis Research Campaign Epidemiology Unit, University of Manchester, Manchester M13 9PT, UK; a.m.hinks@manchester.ac.uk

Accepted 10 December 2006

Published Online First 14 December 2006

REFERENCES

- 1 Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, *et al*. A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes. *Nat Genet* 2004;**36**:337–338.
- 2 Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, *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.
- 3 Hinks A, Worthington J, Thomson W. The association of *PTPN22* with rheumatoid arthritis and juvenile idiopathic arthritis. *Rheumatology (Oxford)* 2006;**45**:365–368.
- 4 Kawasaki E, Awata T, Ikegami H, Kobayashi T, Maruyama T, Nakanishi K, *et al*. 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* 2006;**140**:586–593.
- 5 Bottini N, Vang T, Cucca F, Mustelin T. Role of *PTPN22* in type 1 diabetes and other autoimmune diseases. *Semin Immunol* 2006;**18**:207–213.
- 6 Vang T, Congia M, Macis MD, Musumeci L, Orru V, Zavattari P, *et al*. Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet* 2005;**37**:1317–1319.
- 7 Carlton VE, Hu X, Chokkalingam AP, Schrodi SJ, Brandon R, Alexander HC, *et al*. *PTPN22* Genetic Variation: Evidence for Multiple Variants Associated with Rheumatoid Arthritis. *Am J Hum Genet* 2005;**77**:567–581.
- 8 Hinks A, Barton A, John S, Bruce I, Hawkins C, Griffiths CE, *et al*. Association between the *PTPN22* gene and rheumatoid arthritis and juvenile idiopathic arthritis in a UK population: Further support that *PTPN22* is an autoimmunity gene. *Arthritis Rheum* 2006;**52**:1694–1699.
- 9 Arnett FC, Edworthy SM, Bloch DA, McShane DJ, Fries JF, Cooper NS, *et al*. The American Rheumatism Association 1987 revised criteria for the classification of rheumatoid arthritis. *Arthritis Rheum* 1988;**31**:315–324.
- 10 MacGregor AJ, Bamber S, Silman AJ. A comparison of the performance of different methods of disease classification for rheumatoid arthritis. Results of an analysis from a nationwide twin study. *J Rheumatol* 1994;**21**:1420–1426.
- 11 Wordsworth P, Pile KD, Buckley JD, Lanchbury JS, Ollier B, Lathrop M, *et al*. HLA heterozygosity contributes to susceptibility to rheumatoid arthritis. *Am J Hum Genet* 1992;**51**:585–591.
- 12 Huffmeier U, Steffens M, Burkhardt H, Lascorz J, Schurmeier-Horst F, Stander M, *et al*. Evidence for susceptibility determinant(s) to psoriasis vulgaris in or near *PTPN22* in German patients. *J Med Genet* 2006;**43**:517–522.
- 13 Kawasaki E, Awata T, Ikegami H, Kobayashi T, Maruyama T, Nakanishi K, *et al*. 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* 2006;**140**:586–593.
- 14 Amos CI, Chen WV, Lee A, Li W, Kern M, Lundsten R, *et al*. High-density SNP analysis of 642 Caucasian families with rheumatoid arthritis identifies two new linkage regions on 11p12 and 2q33. *Genes Immun* 2006;**7**:277–286.