

Confirmation of the association of the R620W polymorphism in the protein tyrosine phosphatase PTPN22 with type 1 diabetes in a family based study

H Qu, M-C Tessier, T J Hudson, C Polychronakos

J Med Genet 2005;**42**:266–270. doi: 10.1136/jmg.2004.026971

Genetic susceptibility to the autoimmune B cell destruction that leads to Type 1 diabetes mellitus (T1D) is a complex trait.¹ In recent years, many T1D associations have been reported, but only three (major histocompatibility complex, insulin, and cytotoxic T lymphocyte associated protein 4) have been confirmed in several independent studies.^{2,3} Independent confirmation is essential to eliminate artefacts of publication bias, multiple hypothesis testing, and, in findings of case-control studies, population stratification.⁴

Bottini *et al* recently found, by a case-control design in two independent populations, a novel association of T1D with a single nucleotide polymorphism (SNP) that caused a R620W aminoacid substitution (dbSNP rs2476601) in the lymphoid protein tyrosine phosphatase, non-receptor type 22 (*PTPN22*) gene.⁵ *PTPN22* encodes LYP, a non-receptor tyrosine phosphatase involved in lymphocyte function.

This paper leaves two potential questions unanswered. Firstly, results of case-control studies are potentially artefacts of population stratification, no matter how well matched are the two groups. Secondly, although Bottini *et al* show that the T allele encodes a protein unable to bind to its important Csk partner, and postulate this as a very attractive candidate mechanism for the genetic effect, they did not address the question of the haplotype structure of the locus and the possibility that the association is due to linkage disequilibrium (LD) with another variant. Here we present the results of a study that confirms this association in a design impervious to population stratification, and in a preliminary step towards addressing the second question, we define the LD block that encompasses the PTPN22-R620W SNP and present a computationally generated list of potentially functional SNPs within the block.

MATERIALS AND METHODS

Subjects

Genomic DNA was obtained after informed consent from 588 nuclear families with at least one T1D affected child and two parents. The research ethics board of the Montreal Children's Hospital approved the study. Most probands attended the diabetic clinic at the Montreal Children's Hospital. Ethnic backgrounds were of mixed European descent, with the largest single group being of Quebec French-Canadian origin. All patients were diagnosed before the age of 18 years and required insulin treatment continuously from the time of diagnosis.

For the LD structure studies, we genotyped for PTPN22-R620W DNA from the 30 family trios from the Centre de l'Étude du Polymorphisme Humain (CEPH) used in the International HapMap Project⁶ and combined the results with genotype data from the HapMap website (www.hapmap.org) for LD block analysis.

Key points

- Recently, an association between a functional R620W polymorphism in protein tyrosine phosphatase PTPN22 and type 1 diabetes has been found by a case-control study. Because results of case-control studies may potentially be artefacts of population stratification, replication study is essential.
- To validate the association in a design free of population stratification and explore the possibility that the association is due to linkage disequilibrium (LD) with another variant.
- R620W was genotyped in 588 white nuclear families with at least one affected child and in the 30 European families used in the International HapMap Project.
- Highly significant transmission disequilibrium ($p=1.7\times 10^{-5}$) was observed, confirming the case-control study. However, R620W maps to a 293 kb LD block containing numerous polymorphisms, raising the possibility that other potentially functional polymorphisms may be responsible for the association with T1D instead of, or in addition to, R620W. A computationally generated list of potentially functional SNPs within the block was presented.
- The newly discovered association of PTPN22 with T1D was confirmed. However, for the association within an extended LD block, pinpointing the functional variant may require further studies in populations with different haplotype structure.

Genotyping

The SNP was genotyped by the AcycloPrime-FP SNP detection kit. PCR primers, designed by Primer3 Webtool,⁷ and fluorescence polarisation (FP) probes are listed in table 1. Reagents for each reaction included 12 ng DNA, 2 mmol/l MgCl₂, 25 μmol/l dNTPs, and 0.025 U of AmpliTaq Gold (Applied Biosystems, Foster City, CA, USA). Amplification primers were used at 100 nmol/l each (all concentrations are final). PCR was performed in a Dual 384 well GeneAmp PCR system 9700 in clear 384 well microplates (Greiner Labortechnik, Germany). PCR conditions were: 95°C for 12 minutes; 5 cycles at 97°C for 30 seconds, 58°C for 30 seconds, and 72°C for 30 seconds, then 45 cycles at 95°C for 30 seconds, 58°C for 30 seconds, and 72°C for 30 seconds; and 72°C for 6 minutes in a final volume of 8 μl. Unincorporated primers and dNTPs were removed and final extension was performed in the same PCR system in black

Table 1 Primers and probes used for PTPN22 R620W genotyping

Primers
 Sense: 5'- ACCAGTTCCTCAACCACAA -3'
 Antisense: 5'- AAGAATTCCTTTGGATTGTCT -3'

Probes
 Forward: 5'- CCACAATAATGATTGAGGTGCC -3'
 Reverse: 5'- AAATCCCCCTCCACTTCCTGTA -3'

Table 2 Association analysis between rs2476601 and T1D

R620W*	Frequency	Family number†	Z value	P
Allele				
A (W)	0.135	175	4.303	1.7×10 ⁻⁵
G (R)	0.865	175	-4.303	1.7×10 ⁻⁵
Genotype				
A/A	0.022	29	3.795	0.000148
A/G	0.227	175	1.890	0.058782
G/G	0.751	172	-3.342	0.000830

*Results of the anti-sense probe are reported here, hence the alleles are identified as A and G; †number of nuclear families informative for (with a non-zero contribution to) FBAT analysis.

microplates (MJ Research, Waltham, MA, USA). The extension reaction was performed at 40 cycles at 95 °C for 10 seconds and then 55°C for 30 seconds. Final detection of the SNP was done by the Criterion Analyst HT System (Molecular Devices, Sunnyvale, CA, USA).

Statistics

Hardy-Weinberg equilibrium of parent genotype distribution was tested by Transposer software.⁸ Association was tested by the transmission disequilibrium test, using the Family Based Association Test (FBAT) software (www.biostat.harvard.edu/~fbat/fbat.htm) under the default additive genetic model.⁹ Computation of transmission ratio and haplotype analysis was performed by Haploview software (www.broad.mit.edu/personal/jcbarret/haploview).

Computational prediction of SNP function

The NCBI dbSNP database (www.ncbi.nlm.nih.gov/SNP) and BLAST webtool (www.ncbi.nlm.nih.gov/BLAST), Celera refSNP database (www.celeradiscoverysystem.com), and Genomatix SNP analysis webtool (www.genomatix.de/) were used to perform this task. Candidate SNPs were first identified based on the evolutionary conservation of human–mouse sequence alignment (noted by Celera refSNP as HMCS). For SNPs not included in Celera refSNP, local conservation of a DNA region were predicted by at least 80% sequence identity in a 20 bp sliding window spanning the SNP site.¹⁰ Further function prediction of an amino acid substitution was based on the location in a functional domain of the protein, and the evolutionary distance of amino acid substitution (Gonnet score),¹¹ and potential regulatory role was evaluated by predicted creation/abolition of known transcription factor response and other regulatory elements.

RESULTS AND DISCUSSION

Mendelian error was zero and parent genotypes were in Hardy-Weinberg equilibrium (p = 0.32). There was no discrepancy between the sense and antisense genotyping assays. A highly significant (p = 1.7×10⁻⁵) excess transmission of the A allele (T on the sense strand) from heterozygous parents to affected children (transmitted 131 times, not transmitted 79 times) confirmed the previously reported excess of this allele in affected individuals compared to normal controls (table 2). From the genotype association analysis, we can see that the T1D association is independent of genetic model. The highly significant association also exists under both dominant and recessive genetic models.

To address the question of other potentially functional genetic variants in LD with this SNP, DNA from the 30 CEPH family trios whose SNP genotypes at 5 kb resolution are publicly available (International HapMap Project, 10th release, July 2004) was genotyped for PTPN22-R620W. Combined LD analysis of the R620W results along with other SNPs at that locus was performed stepwise to make sure that all marker SNPs in a LD block were included. The result shows that PTPN22-R620W maps to a 293 kb block of 41 marker SNPs with a solid spine of LD (fig 1). The D' value between the SNPs that define the two ends is 1.0, suggesting that the genetic effect could be caused (or contributed to) by another SNP located anywhere in that segment.

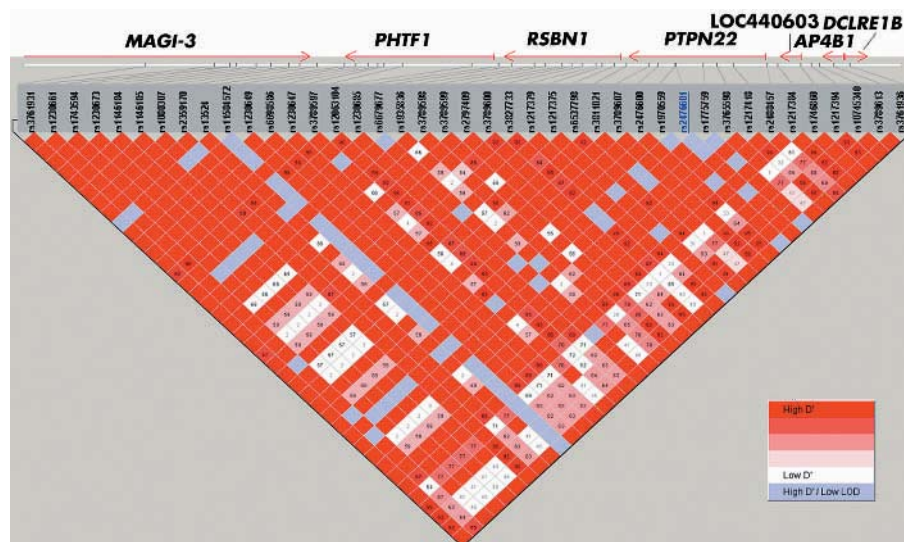


Figure 1 PTPN22 R620W polymorphism rs2476601 (underlined in blue) maps to a solid LD block. The genotyping data, except for rs2476601, are from the 10th release data of International HapMap Project. The haplotype map is made by Haploview v2.0.3 software. D' values (%) are shown in the boxes. D' = 100% for the empty boxes. High D'/low LOD can be seen when there is no/very little recombination evidence between two SNPs, but one SNP is much rarer than the other.

Table 3 Potential functional SNPs in the LD block around *PTPN22**

NCBI dbSNP ID (Celera ref SNP ID)	Relative position from rs2476601 (nucleotide)	Heterozygosity†	Genome position	Amino acid substitution‡	Potential functional elements altered by SNP
rs3761932 (hCV440662)	-220 242	NA	Intron 7 of MAGI-3/NM_152900	-	Ecotropic viral integration site 1 encoded factor Interferon regulatory factor 7 (IRF-7) Octamer binding factor 1 Serum responsive factor
rs2208673 (hCV16162717)	-192 647	NA	Exon 12 of MAGI-3/NM_020965	-	
rs1141016 (hCV8701607)	-137 606	NA	3'UTR of PHTF1/NM_006608	-	Ecotropic viral integration site 1 encoded factor Tax/CREB complex Muscle specific Mt binding site
rs1141015 (hCV8701602)	-137 605	NA	3'UTR of PHTF1/NM_006608	-	Ecotropic viral integration site 1 encoded factor GATA binding factor 3 Muscle specific Mt binding site
rs1141014 (hCV8701601)	-128 971	0	Exon 13 of PHTF1/NM_006608	Trp529Leu (Gonnet score -0.7)	Multifunctional c-Abl src type tyrosine kinase
rs2295853 (hCV1329357)	-123 676	0.480	Intron 10 of PHTF1/NM_006608	-	<i>Xenopus</i> homeodomain factor Xvent-2 Octamer binding factor 1 COMP1 Pancreatic and intestinal lim homeodomain factor Steroidogenic factor 1 Pbx1/Meis1 heterodimer Homeobox protein MEIS1 binding site Special AT rich sequence binding protein 1 Hepatic nuclear factor 1
rs3789597 (hCV1329356)	-121 429	0.222	Intron 7 of PHTF1/NM_006608	-	Ikaros 3, potential regulator of lymphocyte differentiation
rs2273757 (hCV16179758)	-96 787	0.080	Exon 5 of PHTF1/NM_006608	-	Octamer binding factor 1, POU specific domain Retinoic acid receptor site, member of nuclear receptors RAR related orphan receptor alpha1 site MEF3 binding site
rs3789600 (hCV453456)	-43 346	0.304	Intron 3 of FLJ11220/NM_018364	-	Ecotropic viral integration site 1 encoded factor Hepatic nuclear factor 1 Pax-4 homeodomain binding site, together with PAX-6 involved in pancreatic development
rs958007 (hCV8696660)	-20 860	0.196	Intergenic region between FLJ11220 and PTPN22	-	Promyelocytic leukaemia zinc finger (TF with nine Krueppel-like zinc fingers) protein
rs1217413 (hCV1900170)	-19 818	0.474	Intron 21 of PTPN22/NM_015967	-	Cut-like homeodomain protein Ecotropic viral integration site 1 encoded factor Pituitary homeobox 1
rs1217418 (hCV8701537)	23 663	0.500	Intron 4 of PTPN22/NM_015967	-	Myocyte enhancer factor GATA binding factor 1
rs11349 (hCV8701513)	60 169	0	Exon 12 of AP4B1/NM_006594	Phe724Ile (Gonnet score 1.0)	Cdx-2 mammalian caudal related intestinal transcription factor POU-factor Tst-1/Oct-6
rs1217401 (hCV8701512)	61 383	0.471	Exon 10 of AP4B1/NM_006594	Ser480Leu (Gonnet score -2.1, locates in a protein functional motif of chain B, Ap2 clathrin adaptor core)	TCF/LEF-1, involved in the Wnt signal transduction pathway
rs3789613 (hCV25619041)	65 467	0.407	Intron 6 of AP4B1/NM_006594	-	Interferon regulatory factor (IRF) related protein (NF-EM5, PIP, LSIRF, ICSAT) Ecotropic viral integration site 1 encoded factor Fkh domain factor FKHL1 (FOXO) IRF-7
rs12022378 (hCV25619066)	70 821	NA	Exon 1 of DCLRE1B/NM_022836	Tyr61His (Gonnet score 2.2)	TCF11/MafG heterodimers, binding to subclass of AP1 sites cAMP responsive element binding protein E4BP4, bZIP domain, transcriptional repressor v-Myb, AMV v-myb Pax-3 paired domain protein, expressed in embryogenesis, mutations correlate to Waardenburg Syndrome
rs11102699 (hCV31004540)	72 045	NA	Intron 1 of DCLRE1B/NM_022836	-	Fork head related activator-4 (FOXO1) E2F, involved in cell cycle regulation, interacts with Rb p107 protein Signal transducer and activator of transcription 6 (STAT6)
rs11102700 (hCV31004541)	72 054	NA	Exon 2 of DCLRE1B/NM_022836	Phe65Ser (Gonnet score -2.8)	CCAAT/enhancer binding protein beta

Table 3 Continued

NCBI dbSNP ID (Celera ref SNP ID)	Relative position from rs2476601 (nucleotide)	Heterozygosity†	Genome position	Amino acid substitution‡	Potential functional elements altered by SNP
rs11552448 (no Celera ID)	72 074	NA	Exon 2 of DCLRE1B/ NM_022836	Met72Leu (Gonnet score 2.8)	Cut-like homeodomain protein
rs3761936 (hCV27478865)	72 094	0.339	Exon 2 of DCLRE1B/ NM_022836	–	AhR nuclear translocator homodimers

*All the SNPs show conservation by human–mouse protein (for nsSNP) or DNA sequence alignment; †NA, not available in NCBI dbSNP database build 122; ‡for non-synonymous SNP, amino acid substitution is followed by Gonnet score, and protein functional motif (if there is one neighboring the amino acid substitution) shown in the brackets. Gonnet score ranges from +14.2 (highly conserved) to –5.2 (drastic substitution).

At least six other known genes map to this LD block besides PTPN22: the round spermatid basic protein 1 (*RSBN1*, also known as *FLJ11220*), putative homeodomain transcription factor 1 (*PHTF1*), the 3' part of the membrane associated guanylate kinase related gene (*MAGI-3*), *LOC440603* (similar to Gm566 protein), adaptor related protein complex 4, beta 1 subunit (*AP4B1*), and the 5' part of DNA cross link repair 1B (*DCLRE1B*). The block contains 625 SNPs listed in NCBI dbSNP build 122; eight of these are non-synonymous (ns). Besides rs2476601 (R620W), only one nsSNP, rs1217401, maps to exon 10 of *AP4B1* and has an appreciable reported minor allele frequency (0.380) at NCBI dbSNP. This nsSNP is also predicted to have potential functional effects (table 3). However, disease association may be due to non-coding SNPs that affect expression levels, as is the case in two of the three established T1D associations with insulin¹² and CTLA4.¹³ Among the 625 SNPs, 20 (including six nsSNPs) show high human–mouse conservation (an indication of functional importance) and potential regulatory effects (table 3). Of these 20 SNPs, six (rs3789597, rs2273757, rs3789600, rs1217418, rs3789613, and rs3761936), are included in the present HapMap. Except for rs2273757, which is nonpolymorphic in 30 CEPH family trios; the other five SNPs have $D' = 1$ with R620W. R^2 values, a better indicator of whether one SNP may account for an effect observed in another, were 0.06, 0.16, 0.16, 0.26, and 0.58, respectively.

In addition to the strength of the LD, the extent to which one SNP may account for the genetic effect of another in the same block depends on allele frequencies. In this respect, rs6679677 is of particular interest among available HapMap SNPs, because its allele frequency is identical to that of PTPN22-R620W ($R^2 = 1$), which makes it impossible to distinguish the effects of these two SNPs on the basis of genetic data alone. This SNP is located in the intergenic region between *PHTF1* and *FLJ11220*, in a sequence with no human/mouse conservation. The Genomatix SNP analysis web tool predicts the destruction of a binding site of promoter CCAAT binding factors, the destruction of a site of enhancer CCAAT binding factors, and the generation of a binding site of the transcription factor Sox-5. Because of the limited resolution of HapMap, most of the 625 known SNPs in the block are not included in the database and it is therefore not known how many more may be in complete or near complete LD (R^2 close to 1) with PTPN22-R620W.

Thus it cannot be stated with absolute certainty that PTPN22-R620W is the functional variant responsible (or solely responsible) for the genetic effect. However, a very compelling case can be made for PTPN22 and its variant R620W on functional grounds. LYP is a well established suppressor of T cell activation,¹⁴ and targeted disruption of PEP, its mouse orthologue, results in significant enhancement of

memory T cell numbers.¹⁵ Moreover, disruption of binding to Csk with an induced mutation that mimics the effect of R620W abolishes the inhibitory effect of PEP on TCR signalling.¹⁶

The newly discovered association of PTPN22 with T1D and, more recently, other autoimmune disorders^{17–18} highlights an issue that is likely to become common in the elucidation of complex disorders: once association with an extended LD block is established, pinpointing the functional variant may require, in addition to functional studies of the type reported by Bottini *et al*, genetic studies in populations with different haplotype structure.¹⁹ Ultimate proof may require the generation of animal models that carry (or exactly mimic) the effects of the human polymorphism.

ACKNOWLEDGEMENTS

This work was funded by Genome Canada and the Juvenile Diabetes Research Foundation International. We thank R Grabs, F Bacot and R Fr chette for genotyping. Genotyping was performed using the facilities of the McGill University/Genome Quebec Innovation Center, with advice and help of A Sammak and Y Renaud. T J Hudson is supported by a Clinician Scientist Award in Translational Research by the Burroughs Wellcome Fund and an Investigator Award from the Canadian Institutes of Health Research.

Authors' affiliations

H Qu, M-C Tessier, C Polychronakos, The McGill University Health Center (Montreal Children's Hospital), Montr al, Quebec, Canada
T J Hudson, McGill University and Genome Quebec Innovation Centre, Montr al, Quebec, Canada

Competing interests: none declared

Correspondence to: Dr C Polychronakos, The McGill University Health Center (Montreal Children's Hospital), 2300 Tupper, Montr al, Quebec H3H 1P3, Canada; constantin.polychronakos@mcgill.ca

REFERENCES

- Pociot F, McDermott MF. Genetics of type 1 diabetes mellitus. *Genes Immun* 2002;3:235–49.
- Undlien DE, Lie BA, Thorsby E. HLA complex genes in type 1 diabetes and other autoimmune diseases. Which genes are involved? *Trends Genet* 2001;17:93–100.
- Anjos S, Polychronakos C. Mechanisms of genetic susceptibility to type 1 diabetes: beyond HLA. *Mol Genet Metab* 2004;81:187–95.
- Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, Allahabadia A, Franklyn JA, Tuomilehto J, Tuomilehto-Wolf E, Cucca F, Guja C, Ionescu-Tirgoviste C, Stevens H, Carr P, Nutland S, McKinney P, Shield JP, Wang W, Cordell HJ, Walker N, Todd JA, Concannon P. Parameters for reliable results in genetic association studies in common disease. *Nat Genet* 2002;30:149–50.
- Bottini N, Musumeci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D, Mustelin T. A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes. *Nat Genet* 2004;36:337–8.
- The International HapMap Consortium. The International HapMap Project. *Nature* 2003;426:789–96.

- 7 **Rozen S**, Skaletsky HJ. *Primer3 on the WWW for general users and for biologist programmers*. Totowa, NJ: Humana Press, 2000.
- 8 **Cox DG**, Canzian F. Genotype transposer: automated genotype manipulation for linkage disequilibrium analysis. *Bioinformatics* 2001;**17**:738–9.
- 9 **Horvath S**, Xu X, Laird NM. The family based association test method: strategies for studying general genotype–phenotype associations. *Eur J Hum Genet* 2001;**9**:301–6.
- 10 **Loots GG**, Ovcharenko I. rVISTA 2.0: evolutionary analysis of transcription factor binding sites. *Nucleic Acids Res* 2004;**32**:W217–21.
- 11 **Gonnet GH**, Cohen MA, Benner SA. Exhaustive matching of the entire protein sequence database. *Science* 1992;**256**:1443–5.
- 12 **Vafiadis P**, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C. Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. *Nat Genet* 1997;**15**:289–92.
- 13 **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* 2003;**423**:506–11.
- 14 **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* 2002;**30**:237–44.
- 15 **Hasegawa K**, Martin F, Huang G, Tumas D, Diehl L, Chan AC. PEST domain-enriched tyrosine phosphatase (PEP) regulation of effector/memory T cells. *Science* 2004;**303**:685–9.
- 16 **Cloutier JF**, Veillette A. Cooperative inhibition of T-cell antigen receptor signaling by a complex between a kinase and a phosphatase. *J Exp Med* 1999;**189**:111–21.
- 17 **Begovich AB**, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Spoerke JM, Conn MT, Chang M, Chang SY, Saiki RK, Catanese JJ, Leong DU, Garcia VE, McAllister LB, Jeffery DA, Lee AT, Batiwalla F, Remmers E, Criswell LA, Seldin MF, Kastner DL, Amos CI, Sninsky JJ, Gregersen PK. A missense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 2004;**75**:330–7.
- 18 **Kyogoku C**, Langefeld CD, Ortmann WA, Lee A, Selby S, Carlton VE, Chang M, Ramos P, Baechler EC, Batiwalla FM, Novitzke J, Williams AH, Gillett C, Rodine P, Graham RR, Ardlie KG, Gaffney PM, Moser KL, Petri M, Begovich AB, Gregersen PK, Behrens TW. Genetic association of the R620W polymorphism of protein tyrosine phosphatase PTPN22 with human SLE. *Am J Hum Genet* 2004;**75**:504–7.
- 19 **Glazier AM**, Nadeau JH, Aitman TJ. Finding genes that underlie complex traits. *Science* 2002;**298**:2345–9.