ORIGINAL ARTICLE

Evidence for two independent associations with type 1 diabetes at the 12q13 locus

KL Keene[1,2], AR Quinlan[1,2,3], X Hou[2,3], IM Hall[1], JC Mychaleckyj[2,3], S Onengut-Gumuscu[2,4] and P Concannon[1,2]

[1]Department of Biochemistry and Molecular Genetics, UVA School of Medicine, University of Virginia, Charlottesville, VA, USA; [2]Center for Public Health Genomics, UVA School of Medicine, University of Virginia, Charlottesville, VA, USA; [3]Department of Public Health Sciences, UVA School of Medicine, University of Virginia, Charlottesville, VA, USA and [4]Division of Endocrinology and Metabolism, UVA School of Medicine, University of Virginia, Charlottesville, VA, USA

Genome-wide association studies have identified associations between type 1 diabetes and single-nucleotide polymorphisms (SNPs) at chromosome 12q13, surrounding the gene ERBB3. Our objective was to fine map this region to further localize causative variants. Re-sequencing identified more than 100 putative SNPs in an 80-kb region at 12q13. By genotyping 42 SNPs, spanning ~214 kb, in 382 affected sibling pair type 1 diabetes families, we were able to genotype or tag 67 common SNPs (MAF ≥ 0.05) identified from HapMap CEU data and CEU data from the 1000 Genomes Project, plus additional rare coding variants identified from our re-sequencing efforts. In all, 15 SNPs provided nominal evidence for association (P ≤ 0.05), with type 1 diabetes. The most significant associations were observed with rs2271189 (P = 4.22 × 10⁻⁵), located in exon 27 of the ERBB3 gene, and an intergenic SNP rs11171747 (P = 1.70 × 10⁻⁴). Follow-up genotyping of these SNPs in 2740 multiplex type 1 diabetes families validated these findings. After analyzing variants spanning more than 200 kb, we have replicated associations from previous GWAS and provide evidence for novel associations with type 1 diabetes. The associations across this region could be entirely accounted for by two common SNPs, rs2271189 and rs11171747.
Genes and Immunity (2012) 13, 66–70; doi:10.1038/gene.2011.56; published online 18 August 2011

Introduction

Type 1 diabetes is an autoimmune disease resulting from the actions of both genetic and environmental risk factors.[1] Genome-wide association studies (GWAS) have identified a substantial number of loci contributing to type 1 diabetes disease susceptibility,[2–7] including a location on chromosome 12q13, near the gene ERBB3.

The first report of association between single-nucleotide polymorphisms (SNPs) at chromosome 12q13 and type 1 diabetes identified a highly significant association with SNP rs11171739, located between the RPS26 and ERBB3 genes.[2] Subsequent studies, including a meta analysis, have replicated the reported association at rs11171739 and identified a second, more significantly associated SNP, rs2292239, located in intron 7 of ERBB3.[6–8] On the basis of CEU data from the International HapMap project,[9] alleles at rs11171739 and rs2292239 are correlated (r² = 0.71) and are located within a block of significant linkage disequilibrium that includes a number of additional SNPs. Although these

two SNPs in or near ERBB3 provide the strongest evidence for association with type 1 diabetes, the strong linkage disequilibrium across flanking genes leaves unresolved which genes in the 12q13 region may be involved in type 1 diabetes risk. The candidate genes include ERBB3, IKZF4, RPS26 and PA2G4.

ERBB3 encodes a member of the epidermal growth factor receptor family of receptor tyrosine kinases that is capable of ligand binding, but lacks an active kinase domain. Heterodimerization of the ERBB3 protein with other members of the epidermal growth factor family results in the activation of pathways resulting in cell proliferation and differentiation. ERBB3 is also known to interact with several proteins in lymphocytes that are involved in cell signaling, such as the Syk kinases, ZAP70 and SH2B3.[10] SH2B3 is, itself, also a candidate type 1 diabetes susceptibility gene based on GWAS data.[8] Previous studies suggest that insulin deficiencies increase ERBB3 mRNA and protein levels.[11] The protein encoded by IKZF4 is a member of the Ikaros family of DNA-binding proteins that are required for correct development of B and T lymphocytes.[12] RPS26 gene expression has been suggested to influence risk for developing type 1 diabetes.[13,14] The product of the fourth gene in this region, PA2G4, is ERBB3-binding protein 1 (EBP1), which interacts with the cytoplasmic domain of the ERBB3 receptor and may contribute to the transduction of growth regulatory signals.[15]

Table 1 PLINK TDT association results for the 15 significantly associated SNPs in the screening panel of affected sibling pair families

Marker	Position	Gene	Minor allele frequency	Transmitted minor allele count	Untransmitted allele count	TDT P-value	Odds ratio (95% CI)	rs2271189 adjusted P-value	rs11171747 adjusted P-value
rs773107	54655773	RAB5B	0.32	348	282	0.0086	1.23 (1.05–1.44)	0.8712	0.5505
rs1701704	54698754	Intergenic	0.34	358	286	0.0046	1.25 (1.07–1.46)	0.7606	0.4306
rs2456973	54703195	IKZ4	0.34	357	283	0.0034	1.26 (1.08–1.47)	0.7048	0.3831
ERBB3rsx29	54714511	IKZ4	0.09	134	99	0.0219	1.35 (1.04–1.76)	0.9678	0.5237
rs705704	54721679	Intergenic	0.34	264	196	0.0015	1.35 (1.12–1.62)	0.4748	0.2061
rs11171739	54756892	Intergenic	0.42	386	321	0.0145	1.20 (1.04–1.39)	0.2563	0.4860
rs2271194	54763961	ERBB3	0.42	378	314	0.0150	1.20 (1.04–1.40)	0.3413	0.6014
rs10876870	54764269	ERBB3	0.42	390	322	0.0108	1.21 (1.05–1.40)	0.2511	0.5115
ERBB3rsx44	54766741	ERBB3	0.09	127	93	0.0219	1.37 (1.05–1.78)	0.9782	0.4762
rs2292239	54768447	ERBB3	0.35	350	288	0.0141	1.22 (1.04–1.42)	0.8289	0.5070
rs705708	54775180	ERBB3	0.49	411	315	3.67×10^{-4}	1.31 (1.13–1.51)	0.8417	0.2842
rs2292238	54780089	ERBB3	0.45	351	274	0.0021	1.28 (1.10–1.50)	0.0824	0.2518
rs2271189	54781258	ERBB3	0.43	388	282	4.22×10^{-5}	1.38 (1.18–1.60)	1.0000	0.0721
rs11171747	54804675	Intergenic	0.36	278	374	1.70×10^{-4}	0.74 (0.64–0.87)	0.1132	1.0000
rs7960225	54851077	SMARCC2	0.34	264	333	0.0047	0.79 (0.67–0.93)	0.1690	0.6075

Abbreviation: SNP, single-nucleotide polymorphisms.
Adjusted P-values are calculated using the main effects option implemented in the program UNPHASED. SNPs denoted with ERBB3rsx represent novel SNPs identified in 1000 Genomes data, not present in dbSNP.

The distribution of confirmed SNPs in the *ERBB3* region is atypically sparse. Therefore, to identify potentially disease-associated SNPs from the 12q13 region, we carried out re-sequencing of the region in type 1 diabetes patients, and extracted and evaluated sequence and SNP data from both HapMap and 1000 Genomes Project[16] testing the SNPs identified for association with type 1 diabetes.

Results

Re-sequencing of an 80-kb region at 12q13 identified 113 SNPs, of which 68 SNPs were present in dbSNP (Build 129). Among those SNPs not present in dbSNP, 10 were also detected in the 1000 Genomes project data, while 35 were unique to this study (Supplementary Table 1). Of the 113 variants identified by sequencing, only 5 were coding variants (rs2271189, rs773123, Novel *ERBB3* Exon 6 (A/G), rs2229045 and rs2229046).

Collectively, 78 SNPs were present in data from either HapMap or the 1000 Genomes project. Of these, 26 SNPs were located in repetitive sequence or ambiguously mapped to the reference genome. Three of the known SNPs (rs66911160, rs61003310 and rs72247105) were insertions/deletions (indels). These indels are present but not validated in dbSNP, also mapped to repetitive sequence, and were detected only as discrepancies observed in assembling the reference sequence. These indels were not amenable to SNP-based genotyping, therefore they were excluded from further analyses.

A total of 42 SNPs were genotyped in 1744 individuals from 382 type 1 diabetes multiplex families. Excluding SNPs that were not designable on Illumina (Illumina Inc., San Diego, CA, USA) or Eclipse (Epoch Biosciences, Bothell, WA, USA) genotyping platforms, we were able to genotype or tag 67 common variants (MAF⩾0.05) identified from HapMap CEU data and CEU data from the 1000 Genomes Project, plus additional rare coding variants identified from our re-sequencing efforts. After

data cleaning, four of these SNPs (rs10876864, rs1131017, rs3759094 and rs773121) were eliminated from further analyses owing to excessive numbers of Mendelian errors, indicative of a failed assay. The remaining 38 SNPs were tested for association with type 1 diabetes. Evidence of association with type 1 diabetes was observed for 15 SNPs, with P-values ranging from 0.0219–4.22×10^{-5} (Table 1). The most significant associations were observed with rs2271189 ($P = 4.22 \times 10^{-5}$), a synonymous coding SNP located in exon 27 of *ERBB3*, and an intergenic SNP, rs11171747 ($P = 1.70 \times 10^{-4}$), located ∼21 kb 3′ of *ERBB3*.

Conditional analysis, using the 'main effects' test of Cordell and Clayton,[17] as implemented in the program UNPHASED,[18] was used to evaluate the independence of the observed associations with type 1 diabetes. Upon conditioning on either rs2271189 or rs11171747, no additional SNPs provided residual evidence of association with type 1 diabetes (Table 1). To replicate our initial observations we further tested rs2271189 and rs11171747 for association in our complete set of 2740 type 1 diabetes families. The evidence for association with type 1 diabetes for both SNPs was, again, significant (rs2271189 $P = 1.06 \times 10^{-9}$, OR = 1.23 (1.15–1.32); rs11171747 $P = 4.24 \times 10^{-11}$, OR = 0.79 (0.74–0.85)) (Table 2). Conditional analyses in this larger dataset indicate that the two SNPs are independently associated with type 1 diabetes.

Discussion

We have investigated genetic variation across ∼214 kb surrounding the *ERBB3* gene at 12q13, in 382 affected sibling pair type 1 diabetes families from the Type 1 Diabetes Genetics Consortium in order to fine map potential risk variants. The association of this region with type 1 diabetes risk is already well established at genome-wide significance levels, and there is similar evidence for association in this region with rheumatoid

Table 2 PLINK TDT association results for rs2271189 and rs11171747 in the complete data set (2740 type 1 diabetes families)

Marker	Minor allele frequency	Transmitted minor allele count	Untransmitted allele count	TDT P-value	Odds ratio (95% CI)	rs2271189 adjusted P-value	rs11171747 adjusted P-value
rs2271189	0.43	1871	1516	1.06×10^{-9}	1.23 (1.15–1.32)	1.0000	0.0110
rs11171747	0.36	1437	1813	4.24×10^{-11}	0.79 (0.74–0.85)	0.0301	1.0000

Adjusted *P*-values are calculated using the main effects option implemented in the program UNPHASED.
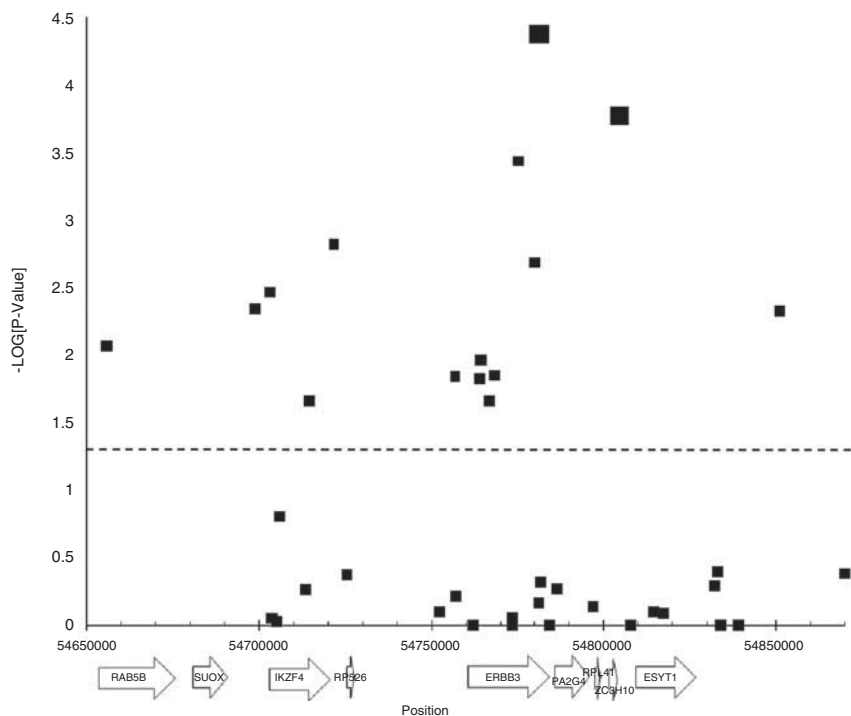

Figure 1 Association plot for single-SNP associations with type 1 diabetes. The SNP position and -LOG (*P*-value) are plotted on the *x* and *y* axis, respectively. All symbols above the dashed line represent a significant association ($P \leqslant 0.05$). SNPs rs2271189 and rs11171747 are represented by larger black squares.

arthritis and thyroiditis.[2,19,20] Evidence of association with type 1 diabetes was observed for 15 of the 42 SNPs genotyped (Figure 1). The most statistically significant associations were observed with a synonymous coding SNP in exon 27 of *ERBB3* and an intergenic SNP. Alleles at the two SNPs are common, with MAF >0.35 and are modestly correlated ($r^2 = 0.42$). Although eight additional SNPs located outside of the coding region of *ERBB3* were nominally associated with type 1 diabetes, upon conditioning on either rs2271189 or rs11171747, there was no significant residual evidence of association at these SNPs. This included rs2292239 and rs11171739 that were reported to be associated with type 1 diabetes in prior GWAS.

The distinct locations of the two implicated SNPs (23 kb apart), one within the coding region of *ERBB3* and one located in the intergenic region between *ZC3H10* and *ESYT1*, as well as the residual evidence of association at each SNP after conditioning on the other, suggest that the two SNPs have at least some independent effects on risk. Although rs2271189 is located within the coding region of *ERBB3*, it does not alter the amino acid sequence. Thus, if it is acting to affect ERBB3 function, the

most plausible explanation is via an effect on mRNA expression, splicing or stability. The second SNP is not located within a known coding region.

Previous studies searching for loci affecting the quantitative expression of genes on a genome-wide basis without particular reference to type 1 diabetes (http://eqtl.uchicago.edu) have identified 21 variants within the 12q13 region of interest, as cis-acting expression quantitative trait loci (eQTLs). Of these 21 SNPs, 16 were genotyped or tagged by SNPs present in our study. Although neither rs2271189 nor rs11171747 was among the previously identified eQTLs, there is detectable linkage disequilibrium between alleles at these SNPs and four eQTL variants ($r^2 = 0.22$–0.39). Notably, a number of these reported eQTL, while physically located within the *ERBB3* gene, affect expression of nearby flanking genes, in particular the *RPS26* gene. These reports suggest that caution is appropriate in interpreting the locations of the disease-associated SNPs identified here with regard to their possible effects on the expression of specific genes.

In summary, in the current study we evaluated variants spanning ~214 kb of the *IKZF4-RPS26-ERBB3-*

PA2G4 gene region to further refine the previously reported regional association with type 1 diabetes susceptibility. Two SNPs were identified that display significantly stronger associations with type 1 diabetes than those reported in prior GWAS studies. Alleles at these two SNPs are only modestly correlated and conditional analyses suggest that they may confer independent effects on risk for type 1 diabetes. The nature of the two SNPs, one coding but synonymous and the other intergenic, suggest that their most likely mechanism for affecting disease risk would be through effects on gene expression.

Materials and methods

Subjects
This study was conducted under Institutional Review Board approval from the University of Virginia. DNA from 192 samples, recruited from both the Virginia Mason Medical Center and Puget Sound Blood Center (Seattle, WA, USA) was utilized for next generation DNA sequencing. Affected sibling pair families obtained from the Type 1 Diabetes Consortium and Human Biological Data Interchange repository[21] were used for subsequent association studies. Information regarding identification, clinical characteristics and recruitment of the type 1 diabetes affected sibling pair families, is accessible from the Type 1 Diabetes Consortium web site (http://www.t1dgc.org).

Re-sequencing and SNP selection
The International HapMap project (HapMap B35)[9] genotyped 25 common SNPs (MAF > 0.05), spanning ~96 kb of the 12q13 region, in the CEU population. Using a pairwise tagging approach (r² threshold = 0.8) in Haploview 3.2,[22] 9 tagging SNPs capture the 25 SNPs, with an average r² = 0.958.

Owing to the limited number of common SNPs genotyped in the HapMap CEU population, next generation sequencing was utilized to identify additional SNPs throughout the region. DNA pooling and long-range PCR (Roche Diagnostics, Indianapolis, IN, USA) approaches were used to isolate an 80-kb region encompassing *ERBB3*, *IKZF4*, *RPS26* and *PA2G4*. In all, Eight PCR primer pairs were designed to amplify overlapping fragments ~10 kb in size, spanning the 80 kb region, in 48 DNA pools, each containing DNA from 4 individuals (192 individuals total). The PCR products were then mixed in equimolar amounts, sheared by sonication and adaptors added. The resulting sequencing library had an average fragment size of 550 bp. Paired-end, 35 bp sequencing was performed using the Genome AnalyzerIIx (Illumina Inc.). The resulting paired-end sequences (readpairs) were aligned to the human reference genome (Build 36) using the program Mosaik (unpublished; http://code.google.com/p/mosaik-aligner/). To minimize false-positive SNPs, subsequent analyses were restricted to readpairs that aligned uniquely to the targeted region. SNPs were identified among the aligned readpairs using a custom script based upon the SAMTOOLS 'pileup' program.[23] For a given SNP to be included in our analyses, we required a minimum posterior SNP probability of 0.99 (that is, pileup SNP quality of at least 20) and at least 10 aligned reads at the polymorphic locus.

Additional novel SNPs were identified from 1000 Genomes Project data (denoted with the prefix ERBB3rsx). Sequencing data, derived from 36 unrelated CEU individuals, were downloaded from ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/2009_02/Pilot1/ (February 2009 data release). Custom scripts were developed, which extracted 65 common SNPs across 12q13 (that is, chromosome 12, from 54.6 Mb–54.9 Mb).

Genotyping
In all, 42 SNPs were genotyped in 382 affected sibling pair type 1 diabetes families, using Eclipse (Epoch Biosciences) and Illumina GoldenGate BeadXpress (Illumina Inc.) genotyping platforms. SNPs with poor clustering or significant deviations from Hardy–Weinberg equilibrium were excluded from further analysis. The two most significant SNPs were further genotyped for replication purposes in the complete set of 2740 type 1 diabetes families.

Statistical analyses
Families with Mendelian errors, as determined using Pedcheck,[24] were zeroed out before performing tests for association, using the TDT option in PLINK v1.07. (http://pngu.mgh.harvard.edu/purcell/plink/).[25] The 'main effects' test of Cordell and Clayton,[17] as implemented in the program UNPHASED,[18] was used to perform association tests conditioning on markers rs2271189 and rs11171747.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgements

References

1 Concannon P, Rich SS, Nepom GT. Genetics of type 1A diabetes. *N Engl J Med* 2009; **360**: 1646–1654.
2 Wellcome Trust Case Control Consortium. Genome-wide association study of 14 000 cases of seven common diseases and 3000 shared controls. *Nature* 2007; **447**: 661–678.
3 Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ et al. A genome-wide association of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. *Nat Genet* 2006; **38**: 617–619.
4 Cooper JD, Smyth DJ, Smiles AM, Plagnol V, Walker NM, Allen JE et al. Meta-analysis of genome-wide association study

data identifies additional type 1 diabetes risk loci. *Nat Genet* 2008; **40**: 1399–1401.

5 Hakonarson H, Grant SF, Bradfield JP, Marchand L, Kim CE, Glessner JT *et al.* A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. *Nature* 2007; **448**: 591–594.

6 Hakonarson H, Qu HQ, Bradfield JP, Marchand L, Kim CE, Glessner JT *et al.* A Novel Susceptibility Locus for Type 1 Diabetes on Chr12q13 identified by a Genome-Wide Association Study. *Diabetes* 2008; **57**: 1143–1146.

7 Barrett JC, Clayton DG, Concannon P, Akolkar B, Cooper JD, Erlich HA *et al.* Genome-wide association study and meta-analysis find that over 40 loci affect risk of type 1 diabetes. *Nat Genet* 2009; **41**: 703–707.

8 Todd JA, Walker NM, Cooper JD, Smyth DJ, Downes K, Plagnol V *et al.* Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007; **39**: 857–864.

9 International HapMap Consortium.. A haplotype map of the human genome. *Nature* 2005; **437**: 1299–1320.

10 Sithanandam G, Anderson LM. The ERBB3 receptor in cancer and cancer gene therapy. *Cancer Gene Ther* 2008; **15**: 413–448.

11 Carver RS, Mathew PM, Russell WE. Hepatic expression of ErbB3 is repressed by insulin in a pathway sensitive to PI-3 kinase inhibitors. *Endocrinology* 1997; **138**: 5195–5201.

12 Georgopoulos K, Moore DD, Derfler B. Ikaros, an early lymphoid-specific transcription factor and a putative mediator for T cell commitment. *Science* 1992; **258**: 808–812.

13 Schadt EE, Molony C, Chudin E, Hao K, Yang X, Lum PY *et al.* Mapping the genetic architecture of gene expression in human liver. *PLoS Biol* 2008; **6**: e107.

14 Plagnol V, Smyth DJ, Todd JA, Clayton DG. Statistical independence of the colocalized association signals for type 1 diabetes and RPS26 gene expression on chromosome 12q13. *Biostatistics* 2009; **10**: 327–334.

15 Yoo JY, Wang XW, Rishi AK, Lessor T, Xia XM, Gustafson TA *et al.* Interaction of the PA2G4 (EBP1) protein with ErbB-3 and

regulation of this binding by heregulin. *Br J Cancer* 2000; **82**: 683–690.

16 1000 Genomes Project Consortium.. A map of human genome variation from population-scale sequencing. *Nature* 2010; **467**: 1061–1073.

17 Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms with a gene using case/control or family data: application to HLA in type 1 diabetes. *Am J Hum Genet* 2002; **70**: 124–141.

18 Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet Epidemiol* 2003; **25**: 115–122.

19 Barton A, Thomson W, Ke X, Eyre S, Hinks A, Bowes J *et al.* Rheumatoid arthritis susceptibility loci at chromosomes 10p15, 12q13 and 22q13. *Nat Genet* 2008; **40**: 1156–1159.

20 Awata T, Kawasaki E, Tanaka S, Ikegami H, Maruyama T, Shimada A *et al.* Association of type 1 diabetes with two loci on 12q13 and 16p13 and the influence coexisting thyroid autoimmunity in Japanese. *J Clin Endocrinol Metab* 2009; **94**: 231–235.

21 Lernmark A, Ducat L, Eisenbarth G, Ott J, Permutt MA, Rubenstein P *et al.* Family cell lines available for research. *Am J Hum Genet* 1990; **47**: 1028–1030.

22 Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005; **21**: 263–265.

23 Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N *et al.* The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* 2009; **25**: 2078–2079.

24 O'Connell JR, Weeks DE. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* 1998; **63**: 259–266.

25 Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D *et al.* PLINK: a toolset for whole-genome association and population-based linkage analysis. *Am J Hum Genet* 2007; **81**: 559–575.

Supplementary Information accompanies the paper on Genes and Immunity website (http://www.nature.com/gene)