Published in final edited form as:
Nat Genet. 2007 July ; 39(7): 857–864. doi:10.1038/ng2068.

Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes

John A Todd[1], **Neil M Walker**[1,9], **Jason D Cooper**[1,9], **Deborah J Smyth**[1,9], **Kate Downes**[1], **Vincent Plagnol**[1], **Rebecca Bailey**[1], **Sergey Nejentsev**[1], **Sarah F Field**[1], **Felicity Payne**[1], **Christopher E Lowe**[1], **Jeffrey S Szeszko**[1], **Jason P Hafler**[1], **Lauren Zeitels**[1], **Jennie H M Yang**[1], **Adrian Vella**[1,8], **Sarah Nutland**[1], **Helen E Stevens**[1], **Helen Schuilenburg**[1], **Gillian Coleman**[1], **Meeta Maisuria**[1], **William Meadows**[1], **Luc J Smink**[1], **Barry Healy**[1], **Oliver S Burren**[1], **Alex A C Lam**[1], **Nigel R Ovington**[1], **James Allen**[1], **Ellen Adlem**[1], **Hin-Tak Leung**[1], **Chris Wallace**[2], **Joanna M M Howson**[1], **Cristian Guja**[3], **Constantin Ionescu-Tirgoviste**[3], **GET1FIN**[4], **Matthew J Simmonds**[5], **Joanne M Heward**[5], **Stephen CL Gough**[5], **The Wellcome Trust Case Control Consortium**[6], **David B Dunger**[7], **Linda S Wicker**[1], and **David G Clayton**[1]

[1]Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 0XY, UK. [2]Department of Clinical Pharmacology, William Harvey Research Institute, Bart's and The London School of Medicine and Dentistry, Charterhouse Square, London, EC1M 6BQ, UK. [3]Clinic of Diabetes, Institute of Diabetes, Nutrition and Metabolic Disease 'N. Paulescu', Bucharest 79811, Romania. [4]Genetics of Type 1 Diabetes in Finland Diabetes Unit, Department of Health Promotion and Chronic Disease Prevention, National Public Health Institute, Helsinki, Finland. [5]Division of Medical Sciences, University of Birmingham, Birmingham.B15 2TT, UK. [6]Wellcome Trust Case Control Consortium (members listed in Supplementary Note). [7]Department of Paediatrics, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 0XY, UK.

Correspondence should be addressed to J.A.T. (john.todd@cimr.cam.ac.uk)..
[8]Current address: Endocrine Research Unit, Mayo Clinic College of Medicine, Rochester, Minnesota, 55905, USA.
[9]These authors contributed equally to this work.

AUTHOR CONTRIBUTIONS

J.A.T. participated in the conception, design and coordination of the study; data analysis and drafting of the manuscript. N.M.W. managed the data and helped coordinate the study. J.D.C. analyzed data and drafted the manuscript. D.J.S. genotyped the nsSNP study and contributed to follow-up genotyping of the nsSNP and WTCCC studies, sequencing and genotyping of *IL2* and *SOCS1*, data analysis and drafting of the manuscript. K.D. contributed to follow-up genotyping of the nsSNP and WTCCC studies, sequencing and genotyping of *PTPN2* and data analysis. V.P. developed the nsSNP scoring algorithm and contributed to data analysis. R.B. genotyped the nsSNP study and contributed to follow-up genotyping of the nsSNP study. S.N. sequenced *KIAA0350* and contributed to its bioinformatics analysis, participated in the follow-up genotyping of the WTCCC study and genotyped and analyzed 12q24 SNPs. S.F.F. genotyped the nsSNP study and contributed to follow-up genotyping of the nsSNP study. F.P. sequenced and genotyped *CIITA.* C.E.L. sequenced and genotyped *IL21.* J.S.S. genotyped and analyzed the gvSNPs. J.P.H. genotyped *CD226* SNPs. L.Z. contributed to follow-up genotyping of the WTCCC scan and bioinformatics analysis. J.Y. contributed to follow-up genotyping of the WTCCC study. A.V. genotyped the *IL2RB* tag SNPs. S. Nutland, H.E.S., H.S., G.C., M.M. and W.M. were responsible for the DNA. L.J.S., B.H., O.S.B. and A.L. provided bioinformatics support. N.R.O. managed subject exclusions and SNP exclusions and the database for the nsSNP study. J.A. and E.A. provided T1DBase support. H.L. and C.W. produced Supplementary Figure 1 and provided statistical support. J.M.M.H. performed statistical analysis; C.G. and C.T. collected the Romanian families; Jaakko Tuomilehto, Leena Kinnunen, Eva Tuomilehto-Wolf, Valma Harjutsalo and Timo Valle of GET1FIN collected the Finnish families; M.J.S., J.M.H. and S.C.L.G. provided the Graves' disease cases and genotyping of rs1990760; WTCCC carried out the 500,000-SNP GWA study; D.B.D collected the T1D cases; L.S.W. discovered the *CD226* nsSNP splice sequence alterations and contributed to the overall planning of the study and D.G.C. participated in the conception, design and coordination of the study; data analysis and drafting of the manuscript.

Note: Supplementary information is available on the Nature Genetics website.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

Published online at http://www.nature.com/naturegenetics

Europe PMC Funders Author Manuscripts

Abstract

The Wellcome Trust Case Control Consortium (WTCCC) primary genome-wide association (GWA) scan[1] on seven diseases, including the multifactorial, autoimmune disease, type 1 diabetes (T1D), shows significant association ($P < 5 \times 10^{-7}$ between T1D and six chromosome regions: 12q24, 12q13, 16p13, 18p11, 12p13 and 4q27. Here, we attempted to validate these and six other top findings in 4,000 individuals with T1D, 5,000 controls and 2,997 family trios that were independent of the WTCCC study. We confirmed unequivocally the associations of 12q24, 12q13, 16p13 and 18p11 ($P_{\text{follow-up}} \leq 1.35 \times 10^{-9}$; $P_{\text{overall}} \leq 1.15 \times 10^{-14}$), leaving eight regions with small effects or false-positive associations with T1D. We also obtained evidence for chromosome 18q22 ($P_{\text{overall}} = 1.38 \times 10^{-8}$) from a genome-wide association study of nonsynonymous SNPs. Several regions, including 18q22 and 18p11, showed association with autoimmune thyroid disease. This study increases the number of T1D loci with compelling evidence from six to at least ten.

There is convincing evidence for association of six loci with T1D: the first, discovered over 25 years ago and having by far the largest effect, are the *HLA* class II genes on chromosome 6p21 in the major histocompatibility complex (MHC). Other loci include the gene encoding insulin (*INS*) on 11p15, *CTLA4* on 2q33, *PTPN22* on 1p13, the interleukin-2 receptor α chain (*IL2RA*, also known as *CD25*) region on 10p15 and, most recently, the *IFIH1* (also known as *MDA5*) region on 2q24 (ref. [2], ref. [3]). These genes explain only some of the familial clustering of T1D (Supplementary Table 1 online). We have assumed for T1D[3] the classical model of a small number of genes with large effects and a large number of genes with small effects[4,5]. If this genetic model is correct, notwithstanding a major role for (unknown) environmental factors[6,7], there should be many more new genes (and pathways) to be discovered, provided sample sizes, study design and genotyping technology suffice[2,3,8-13].

Here, we followed up on the most statistically significant results from two GWA studies: a nonsynonymous SNP (nsSNP) case-control study of 13,378 SNPs in 3,400 affected individuals and 3,300 controls and the WTCCC study using an Affymetrix 500K Mapping Array GWA GeneChip on 2,000 cases and 3,000 controls[1]. There was a substantial overlap of samples (1,834 cases and 1,134 controls) between these studies, but we still had independent samples available for follow-up (up to 4,000 affected individuals and 5,000 controls available from the same DNA collections and 2,997 parent-child trios).

Based on the WTCCC GWA study[1], we initially genotyped 11 SNPs with TaqMan technology that had shown association with $P \leq 1.64 \times 10^{-5}$ (with five having P values $< 5 \times 10^{-7}$) from 11 chromosome regions not previously associated with T1D. We genotyped samples from 4,000 affected individuals and 5,000 controls and from 2,997 parent-child trios that were independent of the WTCCC study (Table 1 and Supplementary Table 2 online). Four of these regions showed convincing evidence of disease association: chromosomes 12q24, 12q13, 16p13 and 18p11 in independent cases and controls ($P \leq 1.82 \times 10^{-6}$), in families ($P = 5.23 \times 10^{-3}$ to 1.07×10^{-6}) and overall ($P = 1.15 \times 10^{-14}$ to 1.52×10^{-20}) (Table 1, Supplementary Table 2 and Supplementary Fig. 1 online). Results from SNPs in the T1D-associated MHC region will be presented elsewhere and were excluded from the analyses presented here.

We developed and applied a strategy for follow-up genotyping as a first step toward defining the disease association of the region. Our aims were to explore in a preliminary way (i) whether there were SNPs even more strongly associated with T1D in a region, (ii) whether the T1D association was due to one or more causal variants and (iii) more precisely where those variants might be within the region (Supplementary Note online).

On chromosome 18p11, the 114-kb region of strong linkage disequilibrium (LD)[14] contained only one gene: *PTPN2* (encoding T-cell protein tyrosine phosphatase) (Supplementary Fig. 1). We selected 11 SNPs from this interval for genotyping based on their pattern of LD with the original SNP found to be associated in the WTCCC study (rs2542151); two SNPs in introns 3 (rs1893217) and 7 (rs478582) of *PTPN2* were more associated with T1D than the original WTCCC SNP and were independently associated with disease (Supplementary Table 3 online). We also resequenced nine of the ten exons of *PTPN2* and 3 kb of each of the 3′ and 5′ regions, uncovering 19 new SNPs and 7 new deletion-insertion polymorphisms. We did not identify any coding variants or obvious splice mutations (Supplementary Note). However, noncoding variants could alter expression of the alternative PTPN2 45-kDa isoform, which is known to dephosphorylate STAT1 (signal transducer and activator of transcription), a major regulator of immune signaling, including in the IL-2 pathway[15].

On chromosome 12q24, the most WTCCC-associated SNP , , rs17696736 (ref. [1]), is located within a large (>1.2-Mb) LD block[14] that contains several genes of possible functional relevance to T1D (Table 1 and Supplementary Fig. 1 (ref. [1])). We genotyped four SNPs for which the LD r^2 values with rs17696736 ranged from 0.59 to 0.82; rs3184504, an nsSNP in exon 3 of *SH2B3* encoding a pleckstrin homology domain (R262W), had the highest association ($P = 1.73 \times 10^{-21}$; odds ratio (OR) = 1.33, 95% confidence interval (c.i.) = 1.26–1.42). This single nsSNP was sufficient to model the association of the entire region (Supplementary Table 3).

In the 16p13 region, SNP rs12708716, which was found to be associated with T1D in the WTCCC study (ref. [1]), remained the most associated after genotyping of additional SNPs (Supplementary Note). LD between HapMap SNPs and rs12708716 localized the association to intron 18 of *KIAA0350* (Supplementary Fig. 1). The *KIAA0350* LD block is flanked by two strong functional candidate genes, *CIITA* (activator of the MHC class II gene transcription) and *SOCS1* (suppressor of cytokine signaling). We resequenced exonic and flanking sequences and genotyped SNPs from these two genes, but neither was responsible for the observed association in *KIAA0350* (Supplementary Note). We resequenced the 24 exons and potentially regulatory 5′ and 3′ sequences of *KIAA0350* and found 12 new SNPs, none of which were an obvious functional candidate (Supplementary Note). We also note that the dexamethasone-induced transcript (*DEXI)* may also be in the LD-defined region; further resequencing and genotyping of the entire region is required.

KIAA0350 is a widely expressed and highly conserved transcript of unknown function with a recognized putative C-type lectin domain encoded by exon 14 (according to Ensembl; see URL below). However, alignment of the domain across species suggested that this domain cannot be considered functional based on homology alone[16]. Further bioinformatics analyses showed that exon 12 may encode an immunoreceptor tyrosine-based activation motif (ITAM) (Supplementary Fig. 2 and the T1DBase PosterPages (see URL below)). ITAMs bind proteins such as SH2B3 (SH2B adaptor protein 3) (also known as LNK, Lymphocyte adaptor protein) that contain SH2 signaling domains. We also noted that SH2B3 binds ERBB3 (v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)) (ref. [17]), which has the highest association with T1D in the other chromosome 12 region, 12q13. Therefore, we identified potential functional links between the new candidate genes and previously identified loci in interactions between antigen presenting cells (for example, dendritic cells) and T lymphocytes during T cell repertoire formation and immune inflammatory events leading to autoimmune pancreatic β-cell destruction in T1D[13].

Of the remaining loci, two are probably false positives (2p13 and 1q32; $P > 0.7$ in the new 4,000 cases and 5,000 controls), and five could be true effects ($P < 0.05$; Table 1). We

followed up on the SNPs in chromosome regions 4q27, 5q14, 2q11, 10p11 and 12p13 in the families, obtaining weak ($P \leq 0.0307$) or no ($P \geq 0.0653$ for 5q14 and 12p13) support for disease association (Table 1 and Supplementary Table 2).

We carried out further genotyping of the 4q27 region (Supplementary Fig. 1) because (i) it contains the *IL2* gene, which has been identified as a susceptibility gene in the nonobese diabetic (NOD) mouse model of T1D[9]; (ii) the chromosome 10p15 region, containing the gene encoding the IL-2 receptor, *IL2RA*, is associated with T1D[18] and autoimmune thyroid disease[19] and (iii) using imputation, the WTCCC study reported a SNP (rs6534347) in the 4q27 region with an apparently strong association with T1D (OR = 1.30, 95% c.i. = 1.10–1.55; $P = 4.48 \times 10^{-7}$)[1]. We resequenced the region encompassing genes *IL2* through *IL21* and found 178 new SNPs but observed neither *IL2* and *IL21* coding variants nor obvious regulatory or splice variants (Supplementary Note).Follow-up genotyping provided some support for association of this region with T1D, but finer localization within the 200-kb region on chromosome 4q27 was not possible owing to strong LD (Supplementary Fig. 1). We did not obtain support for the presence of an effect as large as OR = 1.3 in the region from *IL2* to *IL21*; our most associated SNP was rs3136534, 3′ of *IL2* (OR = 1.11, 95% c.i. = 1.05–1.18; $P_{\text{all cases and controls}} = 1.62 \times 10^{-4}$; Supplementary Note).

The IL-2 receptor, which is critical for immune function and regulation, is a trimeric molecule of α (IL2RA), β (IL2RB, also known as CD122) and γ (IL2RG) chains. We noted that SNP rs3218253 in intron 1 of *IL2RB* shows evidence of T1D association in the WTCCC study[1] ($P = 1.59 \times 10^{-4}$), but we found no convincing support for T1D association (Supplementary Note). This suggests that the WTCCC result was a false positive, emphasizing, along with other findings presented here, the fact that most results in a GWA study at $P < 10^{-6}$ are false positives, even in a sample as large as that used in the WTCCC study.

Using 2,700 case and 3,500 control follow-up samples, we genotyped 14 out of 7,446 nsSNPs from the nsSNP GWA study that had minor allele frequencies (MAF) ≥ 0.01 and *P* values $<1 \times 10^{-3}$ (Table 2 and Supplementary Table 2).In addition to the previously reported *PTPN22* and *IFIH1* region associations[2,20,21], we found one other locus with consistent statistical support for a T1D association: rs763361 in the T lymphocyte costimulation gene *CD226* (ref. [22]) on chromosome 18q22 ($P_{\text{follow-up}} = 9.46 \times 10^{-6}$ and $P_{\text{overall}} = 1.38 \times 10^{-8}$; Table 2 and Supplementary Fig. 1). The *CD226* nsSNP could alter splicing of exon 7 of the gene (Supplementary Note).

In addition to *CD226*, we found evidence ($P_{\text{all cases and controls}} \leq 8.25 \times 10^{-4}$) for nsSNPs rs1445898 (in *CAPSL* on 5p13), rs380421 (in *C20orf168* on 20q13), rs3194051 and rs6897932 (in *IL7R* on 5p13) and rs213950 (in *CFTR* on 7q31) (Table 2).In the family collection (2,997 parent-child trios), we obtained consistent evidence of disease association for all of these nsSNPs (that is, $P < 0.05$ and allelic ORs in the same direction as the original study) except rs1445898 (in *CAPSL*; $P = 0.0885$) (Table 2 and Supplementary Table 2). Confirmation of these potential associations will require further studies.

The *SH2B3* nsSNP rs3184504 was originally excluded from our nsSNP GWA analysis, as the genotype clustering was of marginal quality[2,8]. Recently, we attempted to recover additional poorly clustered nsSNPs from the nsSNP GWA study by identifying for each nsSNP the batches of cases or controls lowering the quality of the fluorescent signal and excluding them. Although it reduced the sample size, this exclusion improved the clustering of nsSNP rs3184504 in *SH2B3* ($P = 2.0 \times 10^{-12}$; OR = 1.30, 95% c.i. = 1.20–1.39 in 3,712 cases and 2,682 controls), making it the nsSNP with the second highest association with T1D in the study, after *PTPN22* (Table 2).

One other outcome of the nsSNP scan analyses, regarding geographical variability in nsSNP allele frequencies, pertains to two potential questions. First, does population structure increase the false-positive rate in case-control association studies[1,8] (see Methods)? Second, are the allelically variable regions of genes responsible for host resistance to infectious disease, which are subject to selection pressures, also candidate susceptibility loci for autoimmune disease[13]. For example, the *IFIH1* nsSNP rs1990760 (ref. [2]), which is associated with T1D[2] and autoimmune thyroid disease (**Tables** 2 and 3), showed some variability in frequency across Great Britain (3.11% from north to south; $P = 6.33 \times 10^{-4}$; Supplementary Table 4) and is known to function as the pathogen recognition receptor (PRR) for picornavirus and enterovirus RNA molecules[23]. We analyzed the nsSNPs for allele frequency differences between geographical regions of Great Britain (Supplementary Table 4). The most geographically variable nsSNP was in the PRR Toll-like receptor 1 (*TLR1*), N248S (rs4833095) on chromosome 4p14 (Supplementary Table 4). This region also showed extreme geographical variation in the WTCCC study[1]. The *TLR1* nsSNP was even more stratified than the three nsSNPs analyzed in the well-established geographically variable lactase persistence gene (*LCT*) (Supplementary Table 4, which has been under recent selection to allow adult consumption of cows' milk[24]. In a heterodimeric receptor with *TLR2*, *TLR1* recognizes lipopeptides from *Mycobacteria*, the causes of leprosy and tuberculosis (Supplementary Note). The SNP in *TLR1* causing the N248S variantand/or other variants in LD with it in the neighboring *TLR6* and *TLR10* genes could have been under selection for resistance to these and other infectious diseases (Supplementary Note and Supplementary Table 4), thus helping to explain their extreme geographical variation across Great Britain and Europe and between the major ethnic groups (Supplementary Note). However, the SNP causing the *TLR1* N248S variant (rs4833095) was not associated in any convincing way with T1D (Supplementary Table 4).

As the autoimmune thyroid disease Graves' disease is known to share genetic susceptibility with T1D[13,19,21], we genotyped 13 T1D-associated SNPs in 2,200 individuals with Graves' disease. We found some evidence of association for 2q11 (rs9653442, intergenic *AFF3* to *LOC150577*), 4q27 (rs17388568, in *Tenr-IL2*), 5p13 (rs1445898, in *CAPSL*), 18p11 (rs1893217 and rs478582, in *PTPN2*) and 18q22 (rs763361, in *CD226*) (Table 3 and Supplementary Table 5 online). Except for the SNP in the *Tenr-IL2* region, all alleles were associated in the same direction as in T1D. We note that the *IFIH1* nsSNP, rs1990760 (ref. [2]), also showed some evidence of association with Graves' disease (Table 3). These data suggest that these genes may be acting as more generalized susceptibility loci for autoimmune disease.

Some[25], but not all[10], authors predict that in human association studies, the distribution of genotypes between unlinked disease loci will deviate from a multiplicative model, and hence, statistical power could be improved in the detection of novel loci using gene-gene interaction analyses[25]. In case-only gene-gene interaction analyses between the new candidates and the known T1D loci, we did not find any evidence of deviation from the model of multiplicative (random) effects, sex effects or age-at-diagnosis effects (Table 4). We can model that the previously identified and newly associated SNPs account for approximately 48% of familial clustering of T1D, compared with an estimated 41% for the MHC region alone. Together, and estimating an environmental contribution of approximately 20% (ref. [6]), about one-third remains unexplained. This residual could be due to numerous as-yet-undetected susceptibility loci, which we expect to range in relative risk effect size up to 20 %, consistent with the expected and emerging L-shaped distribution of allelic effect sizes for the ten loci so far confirmed (Fig. 1 and Supplementary Table 1). Rare causal variants will also have a role.

Our results place the genetic basis of T1D in a genome-wide context.The known genes and the new candidates (such as *PTPN2* and *CD226*) indicate that T1D is caused, in a permissive environment[6,7], by a combination of immune recognition of pancreatic islet antigens (including insulin), T cell repertoire development, immune regulation[13] and other unknown pathways (for example, a pathway including the potential candidate KIAA0350 protein) that have common functional variation.

METHODS

Subjects

The 6,800 affected individuals were recruited as part of the Juvenile Diabetes Research Foundation/Wellcome Trust (JDRF/WT) Diabetes and Inflammation Laboratory's JDRF/WT British case collection (Genetic Resource Investigating Diabetes), which is a joint project between the University of Cambridge Departments of Paediatrics at the Addenbrooke's Hospital and Medical Genetics at the Cambridge Institute for Medical Research. Most affected individuals were <16 years of age at the time of collection; all were under age 17 years at diagnosis and all resided in Great Britain. The 7,000 control samples were obtained from the British 1958 Birth Cohort (B58C), an ongoing study of all people born in Great Britain during one week in 1958 (see URL below). All cases and control were of self-reported white ethnicity, with the exception of 18 cases for whom the WTCCC study found genotype evidence for non-white ethnic group status[1].

All families were of reported or self-reported white ethnicity and of European descent, with two parents and at least one affected child. The family collection consisted of 458 families from the UK Diabetes UK Warren 1 repository, 328 families from USA Human Biological Data Interchange, 250 families from Northern Ireland, 951 Finnish families, 360 Norwegian families, 412 Romanian families and 80 families from Yorkshire, UK (Supplementary Table 6). All DNA samples were collected after approval from the relevant research ethics committees, and written informed consent was obtained from the participants or their guardians.

As part of the AITD Autoimmune thyroid disease (AITD) UK National Collection, 2,200 unrelated, reported white individuals with Graves' disease were recruited. Participants were recruited from centers across the UK, including Birmingham, Bournemouth, Cambridge, Cardiff, Exeter, Leeds, Newcastle and Sheffield (Supplementary Table 6). Affected individuals were defined by the presence of biochemical hyperthyroidism together with at least one of the following: (i) a diffuse goiter on a scan, (ii) positive autoantibodies to the thyrotropin receptor (TSHR), (iii) diffuse goiter on palpation, along with thyroglobulin or thyroid peroxidase autoantibodies or (iv) thyroid eye disease (NOSPECS classification score of 2–6).

Sequencing

Polymorphisms in *MHC2TA*, *SOCS1*, *KIAA0350* and *PTPN2* were identified by resequencing 32 CEPH DNA samples (from Utah residents with northern and western European ancestry) in common with HapMap[14]. The sequencing reactions were performed using Applied Biosystems' BigDye (version 3.1) chemistry and the sequences resolved using an ABI 3700 Genetic Analyzer. Analyses of the sequence traces were performed using the Staden package, and traces were scored independently by a second operator by hand. Annotations for *MHC2TA*, *SOCS1*, *KIAA0350* and *PTPN2* are available from T1DBase (available only from the UK mirror site; see URLs below), together with sequence and polymorphism data the T1DBase PosterPages (see URL below) . For *IL2* and *IL21* and the

flanking regions, polymorphisms were identified by resequencing samples from 32 individuals with T1D.

Genotyping

Follow-up SNPs in the nsSNP and WTCCC studies were genotyped using TaqMan (Applied Biosystems). All genotyping data were scored twice to minimize error; the second operator was unaware of case-control status or and family structure. Concordance data between the two GWA studies and TaqMan genotyping are shown in Supplementary Table 7. All SNPs genotyped in controls did not significantly deviate from Hardy-Weinberg disequilibrium.

Statistical analyses

All statistical analyses were performed in the Stata or R statistical systems (see URLs below) and information about the R package SNP Matrix can be found in ref. [26].

Genome-wide association nsSNP genotyping

In the nsSNP GWA study, we developed and used a clustering method to call genotypes automatically[27]. As two research and development chips had been used in the study, we analyzed 7,446 nsSNPs (MAF ≥ 0.01) that had been on both chips or introduced on the second chip, as these had been attempted in at least 2,908 case and 2,664 control samples. We excluded 172 HLA nsSNPs from this study. Poor clustering was defined as a cluster quality score <2.8 (ref. [27]) or extreme deviation from Hardy-Weinberg equilibrium ($\chi_1^2 > 16$; 165 SNPs dropped)[8]. GWA study data were analyzed using the R package snpMatrix[26], and follow-up analyses used Stata.

Logistic regression analyses

Logistic regression models were used for all case-control association tests. As the T1D cases and controls were chosen to be well matched geographically, we were able to stratify by the 12 subregions of England, Scotland and Wales to exclude the possibility of confounding by geography with little loss of power. We note that the WTCCC study shows that SNPs with significant geographical variation are limited to a small numbers of chromosome regions[1], including the TLR region on chromosome 4p14 described in the present report.

In the logistic regression analysis of a SNP, we performed a one–degree of freedom (1-d.f.) likelihood ratio test to determine whether a 1-d.f. multiplicative allelic effects model or a 2-d.f. full genotype model was more appropriate[28]. We assumed a multiplicative allelic effects model, as it was not significantly different from the full genotype model, except for rs2666236 (*NRP1*). In the forward logistic regression analysis, we started by assessing the evidence against the most significant SNP being the sole variant in the region (in other words, whether this SNP alone was sufficient to model the association). For the purposes of this analysis, we did not assume any specific mode of inheritance for the most associated SNP (A>a) or for any additional SNP with significant independent effects on T1D, so genotype risks of A/A and A/a were modeled relative to the a/a genotype. We then used a 1-d.f. test for adding each of the remaining SNPs to the model by assuming multiplicative allelic effects for the additional SNPs.

2-d.f. locus-based test for pairs of SNPs

To estimate the joint effects of the two independently associated *PTPN2* SNPs from the 18p11 region (rs1893217 and rs478582), we performed a 2-d.f. test by simply entering both genotypes into the logistic regression model as numerical indicator variables coded 0, 1 or 2 (in other words, as multiplicative allelic effects), representing the number of occurrences of

Europe PMC Funders Author Manuscripts

the minor alleles A and A. When compared with the basic model, this 2-d.f. likelihood ratio test corresponds to the 'locus-based' score test described in ref. [11].

3-d.f. haplotype-based test

To test for a haplotype-specific effect, we compared a 3-d.f. haplotype-based test with the 2-d.f. locus-based test. The 3-d.f. haplotype-based test was performed by adding a numerical indicator variable for the 'interaction' term to the 2-d.f. locus-based model: coding the indicator variable as 0, 1 or 2, representing the number of occurrences of the G.G haplotype. However, this interaction term often depends on the (unobserved) haplotype phase, so for the case-control analysis, we replace this indicator variable by its expectation under the null hypothesis, $\theta / (1 + \theta)$, where θ is the odds ratio measure of association between the rs1893217(G>A) and rs478582(G>A).

In the 3-d.f. haplotype-based test, the haplotype phase required by the interaction term was resolved in cases and controls together—consistent with the null hypothesis that case and control haplotypes were drawn from the same population. The interaction term was estimated using the EM algorithm without the imputation of missing genotypes.

Combined test

A score test was used to combine evidence from cases, controls and families[21].

Gene-gene interaction

The case-only gene-gene interaction analysis, defined as deviation from the multiplicative model for the joint effects of the two genotypes, was performed using a regression model as a score test for association between genotypes in case subjects[21]. Affected sib pairs were not tested, as they are not independent. The HLA class II loci were grouped according to their genotypes using a risk-based method, rpart (S.N., J.M.M.H. and J.A.T., unpublished data; see URL below).

Geographically variable SNPs

To test for allele frequency differences between geographical regions, we used the R function snp.lhs.tests, which is part of the snpMatrix package and described in ref. [26]. The SNP genotype was treated as the dependent variable (a binominal variate with two 'trials'). Case-control status was fitted as a covariate, and region, the term to be tested, was fitted as a factor. This results in an 11-d.f. test for allele frequency differences between geographical regions.

Linkage disequilibrium

Measures of linkage disequilibrium, D′ and r^2, were calculated using the Haploview package, and the plots were subsequently generated and displayed through gbrowse (URLs given below) within T1DBase[29].

URLs

Ensembl: http://www.ensembl.org; British 1958 Birth Cohort: http://www.b58cgene.sgul.ac.uk/; T1DBase: http://t1dbase.org (and UK mirror site, http://dil.t1dbase.org); Stata: http://www.stata.com/; R: http://www.r-project.org/; rpart: http://cran.r-project.org/; David Clayton's Software: http://www-gene.cimr.cam.ac.uk/clayton/software/; Haploview: http://www.broad.mit.edu/mpg/haploview/; gbrowse: http://www.gmod.org/. T1DBase PosterPages : https://dil.t1dbase.org/page/PosterAdhoc

Europe PMC Funders Author Manuscripts

Accession codes

All genes are referred to by their HUGO symbol, except for *Tenr* on 4q27 (Entrez GeneID 132612, alias FLJ32741) and DEXI on 16p13 (Entrez GeneID 28955, alias MYLE)

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

This work was funded by the Juvenile Diabetes Research Foundation International and the Wellcome Trust. We gratefully acknowledge the participation of all the patients, control subjects and family members and thank the Human Biological Data Interchange and Diabetes UK for the USA and UK multiplex families, respectively, the Norwegian Study Group for Childhood Diabetes for the collection of Norwegian families (D. Undlien and K. Rønningen), D. Savage, C. Patterson, D. Carson and P. Maxwell for the Northern Irish samples. GET1FIN (J. Tuomilehto, L. Kinnunen, E. Tuomilehto-Wolf, V. Harjutsalo and T. Valle) thank the Academy of Finland, the Sigrid Juselius Foundation and the JDRF for funding. We acknowledge use of the DNA from the 1958 British Birth Cohort collection, funded by the Medical Research Council and Wellcome Trust, and we thank D. Strachan and P. Burton for their help. We also thank The Avon Longitudinal Study of Parents and Children laboratory in Bristol, including S. Ring, R. Jones, M. Pembrey and W. McArdle for preparing and providing the control DNA samples. We thank colleagues at Affymetrix for help and advice in genotyping and T. Willis, M. Faham and P. Hardenbol for the molecular inversion probe technology. We thank the Wellcome Trust for funding the AITD UK national collection; all doctors and nurses in Birmingham, Bournemouth, Cambridge, Cardiff, Exeter, Leeds, Newcastle and Sheffield for recruitment of patients and J. Franklyn, S. Pearce (Newcastle) and P. Newby (Birmingham) for preparing and providing DNA samples on Graves' disease patients. We thank V. Everett, G. Scholz and G. Dolman for information technology support. T1D DNA samples were prepared by K. Bourget, S. Duley, M. Hardy, S. Hawkins, S. Hood, E. King, T. Mistry, A. Simpson, S. Wood, P. Lauder, S. Clayton, F. Wright and C. Collins. We thank L. Peterson for helpful discussions. C.W. is supported by the British Heart Foundation. S. Nejentsev is a Diabetes Research and Wellness Foundation Non-Clinical Fellow.

References

1. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. Jun 6.2007 advance online publication. doi: 10.1038/nature05911.
2. Smyth DJ, et al. A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (*IFIH1*) region. Nat. Genet. 2006; 38:617–619. [PubMed: 16699517]
3. Wang WY, Barratt BJ, Clayton DG, Todd JA. Genome-wide association studies: theoretical and practical concerns. Nat. Rev. Genet. 2005; 6:109–118. [PubMed: 15716907]
4. Fisher RA. Correlation between relatives on the supposition of Mendelian inheritance. Trans. R. Soc. Edinb. 1918:399–433.
5. Barton NH, Keightley PD. Understanding quantitative genetic variation. Nat. Rev. Genet. 2002; 3:11–21. [PubMed: 11823787]
6. Hyttinen V, Kaprio J, Kinnunen L, Koskenvuo M, Tuomilehto J. Genetic liability of type 1 diabetes and the onset age among 22,650 young Finnish twin pairs: a nationwide follow-up study. Diabetes. 2003; 52:1052–1055. [PubMed: 12663480]
7. Todd JA. A protective role of the environment in the development of type 1 diabetes? Diabet. Med. 1991; 8:906–910. [PubMed: 1838040]
8. Clayton DG, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. Nat. Genet. 2005; 37:1243–1246. [PubMed: 16228001]
9. Yamanouchi J, et al. Interleukin-2 gene variation impairs regulatory T cell function and causes autoimmunity. Nat. Genet. 2007; 39:329–337. [PubMed: 17277778]
10. Todd JA. Statistical false positive or true disease pathway? Nat. Genet. 2006; 38:731–733. [PubMed: 16804532]

Europe PMC Funders Author Manuscripts

11. Chapman JM, Cooper JD, Todd JA, Clayton DG. Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. Hum. Hered. 2003; 56:18–31. [PubMed: 14614235]
12. Lowe CE, et al. Cost-effective analysis of candidate genes using htSNPs: a staged approach. Genes Immun. 2004; 5:301–305. [PubMed: 15029236]
13. Ueda H, et al. Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease. Nature. 2003; 423:506–511. [PubMed: 12724780]
14. The International HapMap Consortium. A haplotype map of the human genome. Nature. 2005; 437:1299–1320. [PubMed: 16255080]
15. ten Hoeve J, et al. Identification of a nuclear Stat1 protein tyrosine phosphatase. Mol. Cell. Biol. 2002; 22:5662–5668. [PubMed: 12138178]
16. Zelensky AN, Gready JE. The C-type lectin-like domain superfamily. FEBS J. 2005; 272:6179–6217. [PubMed: 16336259]
17. Jones RB, Gordus A, Krall JA, MacBeath G. A quantitative protein interaction network for the ErbB receptors using protein microarrays. Nature. 2006; 439:168–174. [PubMed: 16273093]
18. Vella A, et al. Localization of a type 1 diabetes locus in the *IL2RA/CD25* region by use of tag single-nucleotide polymorphisms. Am. J. Hum. Genet. 2005; 76:773–779. [PubMed: 15776395]
19. Brand OJ, et al. Association of the interleukin-2 receptor alpha (IL-2Ra)/CD25 gene region with Graves' disease using a multilocus test and tag SNPs. Clin. Endocrinol. 2007; 66:508–512. [PubMed: 17371467]
20. 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. [PubMed: 16697661]
21. Smyth D, et al. Replication of an association between the lymphoid tyrosine phosphatase locus (*LYP/PTPN22*) with type 1 diabetes, and evidence for its role as a general autoimmunity locus. Diabetes. 2004; 53:3020–3023. [PubMed: 15504986]
22. Dardalhon V, et al. *CD226* is specifically expressed on the surface of Th1 cells and regulates their expansion and effector functions. J. Immunol. 2005; 175:1558–1565. [PubMed: 16034094]
23. Kato H, et al. Differential roles of *MDA5* and *RIG-I* helicases in the recognition of RNA viruses. Nature. 2006; 441:101–105. [PubMed: 16625202]
24. Bersaglieri T, et al. Genetic signatures of strong recent positive selection at the lactase gene. Am. J. Hum. Genet. 2004; 74:1111–1120. [PubMed: 15114531]
25. Brassat D, et al. Multifactor dimensionality reduction reveals gene-gene interactions associated with multiple sclerosis susceptibility in African Americans. Genes Immun. 2006; 7:310–315. [PubMed: 16625214]
26. Clayton D, Leung H, An R. Package for analysis of whole-genome association studies. Hum. Hered. 2007; 64:45–51. [PubMed: 17483596]
27. Plagnol V, Cooper JD, Todd JA, Clayton DG. A method to address differential bias in genotyping in large scale association studies. PLoS Genet. Apr 5.2007 in the press. 2007. 10.1371.
28. Cordell HJ, Clayton DG. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. Am. J. Hum. Genet. 2002; 70:124–141. [PubMed: 11719900]
29. Hulbert EM, et al. T1DBase: integration and presentation of complex data for type 1 diabetes research. Nucleic Acids Res. 2007; 35:D742–D746. [PubMed: 17169983]
30. Hardenbol P, et al. Highly multiplexed molecular inversion probe genotyping: over 10,000 targeted SNPs genotyped in a single tube assay. Genome Res. 2005; 15:269–275. [PubMed: 15687290]

Europe PMC Funders Author Manuscripts

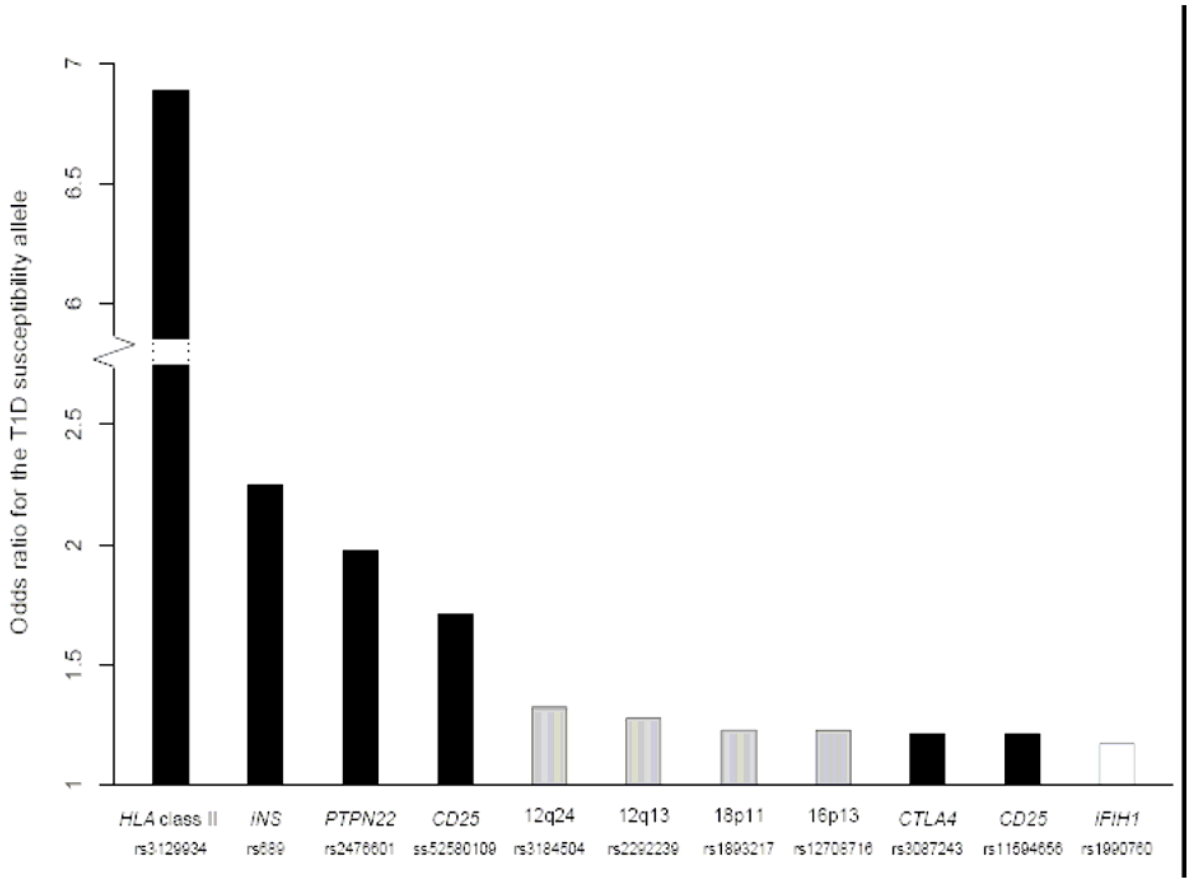

Figure 1.

Odds ratios for the susceptibility allele for the ten independent type 1 diabetes associated genes or regions. The filled black bars indicate previously known associated genes and regions. The open bar indicates the *IFIH1* region identified by the nsSNP genome scan[2] (Table 2), and the filled gray bars were identified by the WTCCC Affymetrix 500K scan[1] and confirmed by the studies reported here (Table 1). The HLA class II SNP (rs3129934) was the marker with the highest association with T1D in the MHC 25-35 Mb region in the WTCCC study[1].

Europe PMC Funders Author Manuscripts

Nat Genet. Author manuscript; available in PMC 2008 July 31.

Table 1

Follow up analysis of the Wellcome Trust Case Control Consortium genome-wide association study of 500,000 random SNPs in type 1 diabetes

				Cases and controls						Families		Overall *P* values	
				WTCCC results (2,000 affected individuals, 3,000 controls)		Follow-up (4,000 affected individuals, 5,000 controls)		All (6,000 affected individuals, 6,200 controls[b])		Follow-up (2,997 parent-child trio samples)		Follow-up samples	All samples
Chr	Gene region[a]	SNP	MAF	*P* (1-d.f. test)	OR (95% c.i.)	*P* (1-d.f. test)	OR (95% c.i.)	*P* (1-d.f. test)	OR (95% c.i.)	TDT *P*[c]	RR (95% c.i.)	*P*	*P*
1p13	*PHTF1–PTPN22*	rs6679677[d] C>A	0.0962	8.03×10^{-24}	1.89 (1.67–2.13)	Confirmed previously							
12q24	*C12orf30*	rs17696736 A>G	0.423	7.27×10^{-14}	1.37 (1.27–1.49)	1.82×10^{-6}	1.16 (1.09–1.23)	1.73×10^{-13}	1.22 (1.15–1.28)	9.20×10^{-5}	1.16 (1.07–1.25)	1.35×10^{-9}	2.31×10^{-16}
12q13	*ERBB3*[e]	rs2292239 C>A	0.34	1.49×10^{-9}	1.30 (1.20–1.42)	1.89×10^{-14}	1.28 (1.20–1.36)	6.46×10^{-19}	1.28 (1.21–1.35)	2.33×10^{-4}	1.15 (1.06–1.24)	3.83×10^{-16}	1.52×10^{-20}
16p13	*KIAA0350*	rs12708716 A>G	0.322	1.28×10^{-8}	0.77 (0.70–0.84)	7.07×10^{-9}	0.83 (0.78–0.89)	7.43×10^{-14}	0.81 (0.77–0.86)	1.07×10^{-6}	0.82 (0.76–0.89)	20.8×10^{-13}	2.57×10^{-18}
18p11	*PTPN2*	rs2542151 A>C	0.163	8.40×10^{-8}	1.33 (1.20–1.49)	3.36×10^{-10}	1.29 (1.19–1.40)	1.49×10^{-14}	1.30 (1.22–1.40)	5.23×10^{-3}	1.13 (1.03–1.25)	1.23×10^{-10}	1.15×10^{-14}
11p15	*INS*[f]	rs3741208 C>T	0.379	2.28×10^{-7}	1.25 (1.15–1.35)	Confirmed previously							
4q27	*Tenr–IL2*	rs17388568 G>A	0.261	6.35×10^{-7}	1.27 (1.15–1.39)	0.0231	1.08 (1.01–1.15)	2.94×10^{-4}	1.11 (1.05–1.18)	0.0307	1.08 (1.00–1.17)	3.32×10^{-3}	5.57×10^{-5}
5q14	*Q8WY63*	rs7722135 G>A	0.241	4.24×10^{-6}	0.79 (0.71–0.87)	0.0474	0.92 (0.86–1.00)	1.23×10^{-3}	0.90 (0.85–0.96)	0.0653	0.94 (0.06–1.02)	0.0149	7.64×10^{-4}
2q11	*AFF3–LOC150577*	rs9653442 A>G	0.458	4.78×10^{-6}	1.21 (1.11–1.32)	0.0213	1.07 (1.01–1.14)	6.23×10^{-5}	1.11 (1.05–1.17)	0.0139	1.09 (1.01–1.18)	1.43×10^{-3}	5.04×10^{-6}
2p13	*DQX1*	rs6546909 *T*>A	0.132	8.53×10^{-6}	1.31 (1.16–1.47)	0.71	1.02 (0.93–1.11)	0.0537	1.07 (1.00–1.16)	N/A			
10p11	*NRP1*	rs2666236 G>A	0.414	1.05×10^{-5}	1.21 (1.11–1.31)	0.126	1.05 (0.99–1.12)	1.76×10^{-4}	1.10 (1.05–1.16)	8.89×10^{-3}	1.10 (1.02–1.19)	7.78×10^{-3}	9.77×10^{-6}
	2-d.f. test[g]					3.07×10^{-3}		6.08×10^{-6}		0.0236		1.46×10^{-4}	6.83×10^{-8}
1q32	*PIK3C2B*	rs12061474 G>A	0.122	1.64×10^{-5}	0.75 (0.65–0.85)	0.934	1.00 (0.91–1.10)	0.0316	0.91 (0.84–0.99)	N/A			
10p15	*RBM17–CD25*	rs12251307 C>T	0.123	3.73×10^{-5}	0.75 (0.66–0.86)	Confirmed previously							
2q33	*CTLA4*	rs3087243 G>A	0.446	8.89×10^{-5}	0.85 (0.78–0.92)	Confirmed previously							

Europe PMC Funders Author Manuscripts

			Cases and controls							Families		Overall *P* values	
			WTCCC results (2,000 affected individuals, 3,000 controls)			Follow-up (4,000 affected individuals, 5,000 controls)		All (6,000 affected individuals, 6,200 controls[b])		Follow-up (2,997 parent-child trio samples)		Follow-up samples	All samples
Chr	Gene region[a]	SNP	MAF	*P* (1-d.f. test)	OR (95% c.i.)	*P* (1-d.f. test)	OR (95% c.i.)	*P* (1-d.f. test)	OR (95% c.i.)	TDT *P*[c]	RR (95% c.i.)	*P*	*P*
Multilocus imputation analysis													
12p13	*CLEC2D*	rs3764021 C>T	0.471	7.19×10^{-5}	0.64 (0.56–0.72)	0.0267	0.93 (0.88–0.99)	1.85×10^{-5}	0.89 (0.86–0.94)	0.124	0.96 (0.89–1.04)	0.0246	4.77×10^{-5}
	2-d.f. test[h]			5.8×10^{-8}		0.0646		4.11×10^{-6}		0.256		0.0330	2.11×10^{-6}

The results for the 500,000-SNP scan were taken from the WTCCC[1] and are used here as reference points. We used the analyses stratified by geographical region (see Methods). Chr = chromosome; MAF = minor allele frequency in control samples; N/A = not attempted; OR = odds ratio for minor allele; 95% c.i. = 95% confidence intervals and RR = relative risk for minor allele. Confirmed previously = these regions have been published previously and thus were not regenotyped in this study.

[a]For any disease-associated region where a particular gene is named, this does not imply that this gene is causal but that it contains, based on currently available sequence and genotype information, the SNP with the highest association to disease. Further studies are required to localize causal variants with each region.

[b]All WTCCC samples, except for 1,500 blood donor samples, were used in the follow-up using TaqMan genotyping technology.

[c]TDT *P* values are based on one-tailed tests (in other words, the null hypothesis was not rejected unless the effect was in the same direction as the original study).

[d]For chromosome 1p13, it is believed that nsSNP R620W (rs2476601) is the causal variant[20] for T1D susceptibility. However, note that there is a SNP, rs6679677, that is in perfect LD with R620W ($r^2 = 1$), making it and gene *PHTF1* an additional candidate.

[e]We also genotyped rs11171739 from *ERBB3* but found it not as significant as rs2292239 ($P = 3.48 \times 10^{-16}$; OR = 1.22 (95% c.i. = 1.17–1.29).

[f]In T1D case and control subjects, *INS* rs3741208 passed the WTCCC call rate filter (call rate ≥ 0.95), but as the WTCCC required all disease and control samples to pass the call rate filter as a whole, *INS* rs3741208 did not make the $P < 5 \times 10^{-7}$ level. *INS* rs3741208 is in LD with the *INS* VNTR.

[g]For chromosome 10p11, there was a significant difference between the full genotype model and the multiplicative allelic effects model ($P = 0.00161$) for cases and controls but not for families ($P = 0.479$).

[h]For chromosome 12p13, there was a significant difference between the full genotype model and the multiplicative allelic effects model ($P = 0.0110$) for cases and controls but not for families ($P = 0.989$).

Europe PMC Funders Author Manuscripts

Table 2

Follow up analysis of the genome-wide association scan of 13,378 nonsynonymous SNPs in type 1 diabetes

			Cases and controls							Families		Overall *P*-values	
			nsSNP results (3,400 affected individuals, 3,300 controls[a])			Follow-up (2,700 affected individuals, 3,500 controls)		All (6,100 affected individuals, 6,800 controls)		Follow-up (2,997 parent-child trios)		Follow-up samples	All samples
Chr	Gene	SNP	MAF	1-d.f. test *P*	OR (95% c.i.)	1-d.f. test *P*	OR (95% c.i.)	1-d.f. test *P*	OR (95% c.i.)	TDT *P*[b]	RR (95%c.i.)	*P*	*P*
1p13	*PTPN22*	rs2476601 (R620W)	0.0939	3.26×10^{-37}	1.99 (1.79–2.22)	1.69×10^{-32}	2.04 (1.81–2.29)	2.71×10^{-63}	1.98 (1.82–2.15)	2.79×10^{-19}	1.66 (1.48–1.86)	1.33×10^{-48}	2.07×10^{-80}
5p13	*CAPSL* [a]	rs1445898 (R75Q)	0.452	1.45×10^{-8}	0.81 (0.76–0.87)	0.114	0.94 (0.87–1.01)	9.61×10^{-6}	0.89 (0.84–0.94)	0.0885	0.95 (0.88–1.02)	0.0379	8.11×10^{-6}
2q24	*IFIH1*	rs1990760 (A946T)	0.398	2.28×10^{-7}	0.82 (0.77–0.89)	3.84×10^{-3}	0.89 (0.83–0.89)	3.27×10^{-9}	0.85 (0.81–0.90)	4.66×10^{-4}	0.88 (0.81–0.95)	1.34×10^{-5}	1.77×10^{-11}
20q13	*C20orf168* [d]	rs380421[e]	0.387	8.84×10^{-7}	0.83 (0.78–0.90)	0.738	0.99 (0.91–1.07)	8.25×10^{-4}	0.91 (0.87–0.96)	7.76×10^{-3}	0.91 (0.84–0.99)	0.0554	3.99×10^{-5}
20q13	*SPINT4* [d]	rs6017667 (G73S)	0.386	1.78×10^{-6}	0.84 (0.78–0.90)	N/A							
18q22	*CD226*	rs763361 (S307G)	0.465	2.16×10^{-5}	1.17 (1.09–1.25)	1.55×10^{-5}	1.18 (1.10–1.27)	2.82×10^{-8}	1.16 (1.10–1.22)	0.0281	1.08 (1.00–1.16)	9.46×10^{-6}	1.38×10^{-8}
5p13	*IL7R* [a]	rs3194051 (I356V)	0.249	1.08×10^{-4}	1.17 (1.08–1.27)	0.0199	1.11 (1.02–1.21)	2.06×10^{-4}	1.12 (1.05–1.19)	N/A			
19p13	*PDE4A*	rs1051738 (A497E)	0.196	1.51×10^{-4}	0.84 (0.76–0.92)	0.603	0.97 (0.88–1.07)	3.55×10^{-3}	0.91 (0.85–0.97)	N/A			
1p32	*ZMYM4* [c]	rs12094543[e]	0.0122	1.92×10^{-4}	1.84 (1.32–2.55)	0.988	1.00 (0.75–1.32)	0.0132	1.27(1.05–1.55)	N/A			
7q31	*CFTR*	rs213950 (V470M)	0.396	2.18×10^{-4}	1.14 (1.06–1.23)	0.0601	1.08 (1.00–1.16)	6.93×10^{-4}	1.09 (1.04–1.15)	4.27×10^{-3}	1.11 (1.03–1.20)	1.49×10^{-3}	1.95×10^{-5}
5p13	*IL7R*[a,c]	rs6897932 (T244I)	0.285	2.19×10^{-4}	0.81 (0.72–0.91)	0.0959	0.93 (0.85–1.01)	8.07×10^{-5}	0.89 (0.84–0.94)	0.0139	0.91 (0.84–0.99)	6.54×10^{-3}	7.77×10^{-6}
3q25	*MED12L*	rs3732765 (R1210Q)	0.378	2.87×10^{-4}	0.87 (0.81–0.94)	0.298	0.96 (0.89–1.04)	1.26×10^{-3}	0.92 (0.87–0.97)	N/A			
20q11	*LBP* [c]	rs2232613 (L333P)	0.0917	4.43×10^{-4}	0.77 (0.67–0.89)	0.738	0.98 (0.85–1.12)	0.0164	0.89 (0.81–0.98)	N/A			
16q23	*WWOX* [c]	rs7499843[e]	0.303	6.33×10^{-4}	1.16 (1.07–1.27)	0.497	1.03 (0.95–1.11)	0.0354	1.06 (1.00–1.12)	N/A			

Europe PMC Funders Author Manuscripts

The genotyping of 13,378 nsSNPs, which was every possible nsSNP sequence across the entire genome to which a genotyping assay could be designed for, was carried out using molecular inversion probe technology[8,30], resulting in 7,446 nsSNPs with a MAF > 0.01 scored successfully[27] Chr = chromosome; MAF = minor allele frequency in control samples; N/A = not attempted; OR = odds ratio for minor allele; 95% c.i. = 95% confidence intervals; RR = relative risk for minor allele .

[a] D′ = 0.99 and r^2 = 0.13 between *IL7R* SNPs rs3194051 and rs6897932. *CAPSL* nsSNP rs1445898 is in the same region of LD as the *IL7R* nsSNPs and D′ = 0.81 and r^2 = 0.23 and D′ = 0.95 and r^2 = 0.43 between rs1445898 and *IL7R* SNPs rs3194051 and rs6897932, respectively.
IL7R is also called *CD127*. *IL7R* nsSNP rs6897932 is in the transmembrane domain and has been associated previously with multiple sclerosis (Supplementary Note).

[b] TDT *P* values are based on one-tailed tests (that is, the null hypothesis was not rejected unless the effect was in the same direction as the original study).

[c] Different numbers of samples were genotyped in the GWA study for rs12094543 (*ZMYM4*), rs2232613 (*LBP*) and rs7499843 (*WWOX*, 2,641 affected individual and 2,484 control samples) and rs6897932 (*IL7R*, 1,712 affected individual and 1,529 control samples).

[d] D′ = 1.00 and r^2 = 0.99 between rs380421 (*C20orf168*) and rs6017667 (*SPINT47*), so only rs380421 was followed up.

[e] These SNPs are no longer nsSNPs according to dbSNP build 36.

Europe PMC Funders Author Manuscripts

Table 3

Association study of type 1 diabetes associated SNPs in 2,200 individuals with Graves' disease and 3,600 geographically-matched controls

Chr	Gene region	SNP	MAF	OR (95% c.i.)	*P* (1-d.f. test)
2q11	*AFF3–LOC150577*	rs9653442 A>G	0.467	1.10 (1.01–1.19)	0.0221
2q23	*IFIH1*	rs1990760 A>G	0.398	0.91 (0.84–0.99)	0.0265
4q27	*Tenr–IL2*	rs17388568 G>A	0.293	0.87 (0.79–0.95)	1.81×10^{-3}
5p13	*IL7R*	rs6897932 G>A	0.268	0.96 (0.88–1.05)	0.363
5p13	*CAPSL*	rs1445898 G>A	0.446	0.88 (0.81–0.96)	2.72×10^{-3}
10p12	*NRP1*	rs2666236 G>A	0.421	1.03 (0.95–1.12)	0.450
12q13	*ERBB3*	rs2292239 C>A	0.352	0.99 (0.92–1.08)	0.899
12q24	*C12orf30*	rs17696736 A>G	0.438	1.04 (0.96–1.13)	0.332
12q24	*SH2B3*	rs3184504 A>G	0.488	1.07 (0.98–1.16)	0.127
16p13	*KIAA0350*	rs12708716 A>G	0.359	0.97 (0.89–1.06)	0.497
18p11	*PTPN2*	rs1893217 A>G	0.167	1.13 (1.02–1.25)	0.0251
18p11	*PTPN2*	rs478582 A>G	0.460	0.91 (0.84–0.99)	0.0239
18q22	*CD226*	rs763361 G>A	0.465	1.10 (1.02–1.20)	0.0182
11p15	*INS*	rs689 A>T	0.300	0.95 (0.87–1.05)	0.305

MAF = minor allele frequency in control samples, OR = odds ratio for minor allele, 95% c.i. = 95% confidence interval. We can conclude that these potential associations with Graves' disease are not due to the presence of T1D in a few individuals with Graves' disease, because some SNPs (for example, rs12708716 and the very strongly T1D associated SNP rs689 in *INS*) do not show any evidence of association with disease in these individuals with Graves' disease. We note that for the *Tenr–IL2* region SNP rs17388568, the minor allele is associated with reduced risk in Graves' disease but with susceptibility in T1D.

Europe PMC Funders Author Manuscripts

Nat Genet. Author manuscript; available in PMC 2008 July 31.

Table 4

Analysis of gene-gene interactions of new type 1 diabetes loci with known disease loci

		New T1D loci				
Gene	**rs number**	**12q24 *SH2B3* rs3184504**	**18p11 *PTPN2* rs1893217**	**18p11 *PTPN2* rs478582**	**12q13 *ERBB3* rs2292239**	**16p13 *KIAA0350* rs12708716**
HLA class II	(11d.f.)	0.780 (2,347)	0.811 (2,527)	0.973 (2,488)	0.307 (2,551)	0.142 (2,554)
INS	rs689	0.602 (3,839)	0.333 (4,081)	0.299 (4,077)	0.904 (3,616)	0.233 (4,174)
PTPN22	rs2476601	0.529 (4,970)	0.631 (5,270)	0.537 (5,286)	0.0240 (4,695)	0.659 (5,380)
CD25	ss52580109	0.264 (4,816)	0.638 (5,076)	0.752 (5,087)	0.747 (4,519)	0.244 (5,188)
CD25	rs11594656	0.407 (4,764)	0.341 (5,023)	0.487 (5,050)	0.471 (4,481)	0.457 (5,133)
CTLA4	rs3087243	0.600 (3,806)	0.955 (4,043)	0.989 (4,044)	0.359 (3,580)	0.922 (4,137)
IFIH1	rs1990760	0.777 (5,147)	0.362 (5,463)	0.825 (5,484)	0.877 (5,075)	0.104 (5,589)
	Sex	0.355 (5,338)	0.422 (5,674)	0.849 (5,699)	0.683 (5,371)	0.776 (5,804)
	Age at diagnosis	0.0153 (5,338)	0.450 (5,674)	0.396 (5,699)	0.139 (5,371)	0.715 (5,804)

Data represent *P* values (with the number of affected individuals in parentheses).