HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2010 September 21.

Published in final edited form as:

Nat Genet. 2009 June ; 41(6): 703–707. doi:10.1038/ng.381.

Genome-wide association study and meta-analysis finds over 40 loci affect risk of type 1 diabetes

Jeffrey C. Barrett[1], **David Clayton**[1], **Patrick Concannon**[2,3,*], **Beena Akolkar**[4], **Jason D. Cooper**[1], **Henry A. Erlich**[5], **Cécile Julier**[6], **Grant Morahan**[7], **Jørn Nerup**[8], **Concepcion Nierras**[9], **Vincent Plagnol**[1], **Flemming Pociot**[8], **Helen Schuilenburg**[1], **Deborah J. Smyth**[1], **Helen Stevens**[1], **John A. Todd**[1], **Neil M. Walker**[1], **Stephen S. Rich**[3,10], and **the Type 1 Diabetes Genetics Consortium**[11]

[1]Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Addenbrooke's Hospital, Cambridge, CB2 0XY, UK.

[2]Department of Biochemistry and Molecular Genetics, University of Virginia, Charlottesville, VA, USA.

[3]Center for Public Health Genomics, University of Virginia, Charlottesville, VA, USA.

[4]Division of Diabetes, Endocrinology, and Metabolic Diseases, The National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institutes of Health, Bethesda, MD, USA

[5]Roche Molecular Systems, Pleasanton, CA, USA

[6]Inserm U730, Centre National de Génotypage, Evry, FR

[7]Centre for Diabetes Research, The Western Australian Institute for Medical Research, and Centre for Medical Research, University of Western Australia, Perth, WA, AUSTRALIA

[8]Steno Diabetes Center, Gentofte, DK

[9]Juvenile Diabetes Research Foundation, New York, NY, USA

[10]Department of Public Health Sciences, Division of Biostatistics and Epidemiology, University of Virginia, Charlottesville, VA, USA.

Abstract

*Correspondence to: Patrick Concannon, Center for Public Health Genomics, University of Virginia, P.O. Box 800717, Charlottesville, VA 22908 USA, Tel: 434-982-3288, Fax: 434-924-1312, e-mail: patcon@virginia.edu.

AUTHOR CONTRIBUTIONS JCB and DC helped design the study, and carried out the statistical analyses. JCB, DC and PC drafted the manuscript. PC managed the editing of the manuscript. FP, CJ, JAT, BA, HAE, GM, JN, CN, PC and SSR (Chair) are members of the T1DGC Steering Committee, and contributed to the general planning and design of the study, and to the writing of the manuscript. FP and JN also coordinated the inclusion of the Danish case-control samples. NMW and H. Schuilenburg provided data services. JDC and VP assisted with data analyses. DJC carried out genotyping, and H. Stevens provided DNA sample coordination and quality control. The T1DGC provided biospecimens and data from T1D families.

[11]Membership list appears in Supplementary Note

URLs Further information about all T1D loci, including those newly reported here can be found in T1DBase: http://www.t1dbase.org/

Type 1 diabetes (T1D) is a common autoimmune disorder that arises from the action of multiple genetic and environmental risk factors. We report the findings of a new genome-wide association study of T1D, combined in a meta-analysis with two previously published studies. The total sample set included 7,514 cases and 9,045 reference samples. Forty-one distinct genomic locations provided evidence for association to T1D in the meta-analysis ($P < 10^{-6}$). After excluding previously reported associations, 27 regions were further tested in an independent set of 4,267 cases, 4,463 controls and 2,319 affected sib-pair (ASP) families. Of these, 18 regions were replicated ($P < 0.01$; overall $P < 5 \times 10^{-8}$) and four additional regions provided nominal evidence of replication ($P < 0.05$). The many new candidate genes suggested by these results include *IL10, IL19, IL20, GLIS3, CD69* and *IL27*.

Results from linkage and association studies in T1D have long supported a model in which the major risk factor for T1D resided in the HLA region on chromosome 6p21. Candidate gene studies carried out over a number of years identified four non-HLA T1D risk loci: *INS*, *CTLA4*, *PTPN22*, and *IL2RA*1-4. Recently, the application of genome-wide SNP typing technology to large sample sets and comparisons with results from other immune-mediated diseases have provided convincing support for 19 additional T1D loci5-13, all with allelic odds ratios (OR's) of less than 1.3.

In order to have adequate power to detect additional T1D risk loci with ORs in the range of 1.1 to 1.3, we performed a new genome-wide association scan using British cases and controls and used this dataset in a meta-analysis which included 7,514 cases and 9,045 reference samples (Table 1). The other datasets included in the meta-analysis were from the Wellcome Trust Case Control Consortium (WTCCC) study7 and a study12 that utilized T1D cases from the Genetics of Kidneys in Diabetes (GoKinD) study of diabetic nephropathy14, and reference samples from the National Institute of Mental Health (NIMH) study15.

The two earlier studies (WTCCC and GoKinD/NIMH) used Affymetrix 500K platforms while the new (T1DGC) study used the Illumina 550K platform. Of the 841,622 SNPs genotyped in these studies which had minor allele frequencies (MAF) exceeding 1% and passed our quality control standards, 328,044 were only genotyped by the Affymetrix platform, 437,739 only by the Illumina platform, and 75,839 were genotyped by both platforms. Since only 9% of SNPs are shared between these platforms, imputation was used to combine results across studies. To develop imputation rules, we took advantage of the fact that 1,422 of the original WTCCC controls which were included in the T1DGC study had been genotyped on both platforms (Methods).

An analysis using Mantel's extension to the 1 degree of freedom (1 df) Cochran-Armitage trend test which combined comparisons over the three studies yielded 41 distinct genomic locations with *P*-values $< 10^{-6}$ (Figure 1) (Individual plots for each study are in Supplementary Figure 1). Fifteen of these sites were in regions where there were prior reports of association to T1D (Table 2). The remaining 26 of these locations along with one weaker association on the X chromosome, were chosen for further analysis. To address the possible effects of population structure, the analyses were stratified by geographical region in the case of the British studies and by a "propensity score" based on principal components

analysis on the US study. This was only partially successful in reducing the over-dispersion of test statistics, a large part of which derived from the US data (Table 3). If the residual over-dispersion were due to population structure, there would be a strong case for correcting the *P*-values (as shown in Table 3). However, the modest effect of the stratified analysis on over-dispersion, taken together with the absence of any over-dispersion in case-only interaction tests (see below) suggests that it is more likely due to differential genotyping errors. In this case, correction of the most significant *P*-values would be over-conservative since we have carefully checked all genotyping cluster plots for associated SNPs. The genomic control corrected *P*-values are nevertheless shown in Supplementary Table 1. The strongest associations tended to become somewhat less significant, but the choice of regions for follow-up, based on the criteria of $P < 10^{-6}$, was not affected. We also carried out, for SNPs with minor allele frequency exceeding 10%, 2 df "genotype" tests which would be more sensitive to associations showing marked dominance (deviation from an additive model, on the log scale). Significance was notably increased, by 3 to 4 orders of magnitude, at three SNPs, but was less significant than the corresponding 1 df tests otherwise (Supplementary Table 1) yielding no additional findings at $P < 10^{-6}$. The results of both simple and stratified 1 df tests of these SNPs, separated by study, are shown in Supplementary Tables 3 and 4. Quantile-quantile plots for tests in our new (T1DGC) study, and in the meta-analysis, after removal of tests for SNPs in linkage disequilibrium (LD) regions surrounding known and putative associations, are shown in Supplementary Figure 2a and 2b.

The most significantly T1D associated SNPs from each of the 27 novel regions selected for replication were genotyped in a further 4,267 cases and 4,670 controls and in 4,342 trios from 2,319 T1DGC families with multiple affected offspring. Genotype data passed design and quality control criteria for 25 of these SNPs. Eighteen regions replicated with $P < 0.01$ and showed genome-wide significant ($P < 5 \times 10^{-8}$) association in the joint analysis of the genome scans and replication samples (Table 4, individual scan data in Supplementary Table 2). A further three of the remaining seven SNPs also showed $P < 0.01$ in the replication studies, and a fourth had $P < 0.05$, but these failed to reach overall $P < 5 \times 10^{-8}$ (Table 4). This study, therefore, adds 18 T1D risk loci to the existing 24, and provides suggestive support for four more. As expected, nearly all of these loci have OR < 1.2, as larger effects would likely have been discovered in earlier studies. Two of the new associations (10q23 and 16q23) contradict this trend and highlight the disparity between genomic coverage of the older Affymetrix 500K chip and the newer Illumina 550K: these loci do not have a good proxy on the Affymetrix chip, explaining why they were not previously identified despite relatively large effect sizes (OR ∼ 1.3).

The families utilized for replication were derived from affected sib-pair linkage studies. One consequence of ascertainment on the basis of at least two affected siblings was a high frequency of high risk HLA genotypes16. It has been reported that relative risks for several non-HLA loci are reduced in subjects carrying high risk HLA genotypes17, 18, reflecting deviation from a multiplicative model for joint effects, and this would lead us to expect reduced effect sizes in multiple-case families. Indeed, the results of the replication study were generally less convincing in the family data than in the case-control data reflecting

smaller effect sizes in the families. One potential explanation for these different effect sizes lies in possible statistical interaction among risk loci leading to a less-than-multiplicative accumulation of risk in samples (such as those from multiplex families) with a large number of risk variants. This hypothesis is difficult to test because power to detect interaction terms is much less than that to find equivalent sized main effects and is doubly compounded when specific causal variants (rather than tag SNPs from a GWA scan) are not known.

We tested for deviation from the model of multiplicative effects with HLA, on a genome-wide basis, by first calculating predictive risk scores using SNPs in the MHC region on each platform, and testing for association between this score and every other SNP in the remainder of the genome. These tests are "case-only" tests for statistical interaction reflecting variation of allelic relative risks with the level of HLA-attributable risk. As noted earlier, these test statistics did not show the over-dispersion which would have been indicative of population stratification (Supplementary Figure 2c). However, the subset of these tests concerning established T1D susceptibility loci tended to have larger chi-squared values than expected by chance (Supplementary Figure 2d). In the majority of cases (31/45), the interaction tests took the opposite sign from the main effect test, consistent with high MHC risk leading to lower risk for other loci. Of the five interactions which reached $P < 0.05$, four were of this type (loci near 2q24.2/*IFIH1*, 1p13.2/*PTPN22*, 17p13.1 and 2q33.2/*CTLA4*). We carried out a further test by calculating a T1D risk score using all associated loci excluding the MHC region and testing, in cases only, for correlation between this score and the MHC risk score. We found a weak, but significant (P=0.0007) negative correlation, again indicating that risk from HLA and non-HLA sources accumulates at a rate less than expected based on the model of multiplicative effects, so that there is a general tendency for relative risks for non-HLA loci to be reduced when HLA-related risk is high.

Several of the 18 regions identified here contain genes of possible functional relevance to T1D. These include the region 1q32.1 containing the potent immunoregulatory cytokine genes, *IL10*, *IL19* and *IL20*. The region of strong LD at 9p24.2 contains only a single gene, *GLIS3*. Mutations in *GLIS3* have been reported in children from three different consanguineous families with permanent neonatal diabetes associated with congenital hypothyroidism and other clinical complications19. The region on 12p13.31 harbors a number of immunoregulatory genes including *CD69*, which is induced by activation of T cells and functions in thymic egress20. Several other members of the calcium-dependent (C-type) lectin (CLEC) domain family with immune functions also map to this region. Overall, our results provide a rich new source of candidate genes, but until further genotyping, re-sequencing and functional studies are performed, it is not possible to be more specific in regard to which genes might be causal.

Methods

Subjects

The WTCCC study has been described elsewhere7. Cases were recruited from pediatric and adult diabetes clinics at 150 National Health Service Hospitals across Great Britain as part of the Genetic Resource for Investigating Diabetes (GRID) collection (www.childhood-diabetes.org.uk/grid.shtml) of the JDRF/WT DIL9. Half of the controls were drawn from the

British 1958 Birth Cohort21 and half from a group of blood donors recruited by the WTCCC in collaboration with the UK Blood Services7. The former group was subsequently genotyped on the Illumina 550K platform and was used as controls in the new T1DGC study reported here. Since the removal of this group from the WTCCC study left it somewhat short of controls, we used a group of 1,868 patients with bipolar disorder as additional reference samples — a group conspicuous in the WTCCC studies in its lack of significant differences from control allele frequencies7.

Our new study added approximately 2,500 new controls from the British 1958 Birth Cohort to the 1,500 described above, and compared these with a new group of approximately 4,000 British cases from the JDRF/WT DIL collection. All cases and controls were resident in Great Britain. To minimize the effects of population structure, the case-control comparisons in the WTCCC and T1DGC studies have been stratified by the 12 regions of Great Britain5,7. Sample exclusions in the genome-wide studies are discussed in Supplementary Methods.

Replication studies were carried out in two groups of cases and control as well as 2,319 affected sib-pair families previously recruited and characterized by the T1DGC6. The British cases were from the JDRF/WT DIL, and the controls were drawn from the British 1958 Birth Cohort, and the UK Blood Service controls of the WTCCC. The second set of cases and controls from Denmark were recruited from a nationwide registry. All cases (49% females) were diagnosed before age 18 years and the mean age at onset 9.02 years. Control subjects were randomly selected from the Inter99 study22.

Genotyping

For the T1DGC study, the 4,000 T1D case and 2,500 control DNA samples were selected based on no prior use in a prior genome wide association study and migration as a high molecular weight band of genomic DNA, ∼23 kb, by electrophoresis on a 0.75% agarose gel. All DNA samples were extracted using a chloroform-based method and quantified in triplicate using Picogreen®. Once selected, the case and control DNA were randomized by columns into a 96 well plate format.

For the T1DGC study, genotyping was performed on the Illumina 550K Infinium platform and, for comparability, all genotypes were re-scored using the *ILLUMINUS* algorithm23. The WTCCC study used the Affymetrix GeneChip Human Mapping 500K Array set, while the GoKinD/NIMH study used genotype data generated with the Affymetrix Genome-wide Human SNP Array 5.0. The 5.0 array incorporates all of the SNPs on the earlier 500K array but on a single chip along with an additional 420K non-polymorphic probes. Details of the scoring of genotypes may be found in the original publications7, 12. The criteria for discarding some SNPs from the analysis are discussed in Supplementary Methods.

For the replication studies, genotyping was performed in a fully blinded fashion using Taqman assays as previously described9.

Statistical methods

One degree of freedom tests are Cochran-Armitage tests for trend alternatives, extended to pool information across multiple studies or across multiple strata within a single study by the method described by Mantel24. The two degree of freedom tests follow similar principles. Testing for association with SNPs on the X chromosome was carried out using the method proposed by Clayton25. More details are given in Supplementary Methods.

The meta-analysis involved studies that used different platforms, necessitating the use of imputation. Since we had a substantial sample typed on both platforms, we used a simple linear regression approach to imputation26. Details of this, and other methods used in the meta-analysis, are given in Supplementary Methods. Supplementary Figure 3 shows the distribution of the quality of imputation, as measured by the coefficient of determination, R^2.

Analysis of the replication case-control studies was carried out in a similar manner, by 1 df comparisons of allele frequencies with Danish and UK studies treated as separate strata. The family study was analyzed by the transmission/disequilibrium test (TDT).

The MHC risk score was derived by an adaption of the lasso approach27 to logistic regression of case/control status versus all SNPs in the MHC region (defined as spanning from 24.7 Mb to 34.0 Mb on chromosome 6). This was applied to the combined Affymetrix data, with a dummy variable in the regression to differentiate WTCCC and GoKinD/NIMH studies and, separately to the T1DGC Illumina data. The coefficients for the selected regression equations are shown in Supplementary Table 3. The degree of risk prediction, as demonstrated by the receiver operating curves (Supplementary Figure 4) was very similar in the three study groups.

A case-only test for statistical interaction between each SNP and MHC risk score was carried out by a 1 df test based on the covariance between MHC risk score and the SNP genotype coded 0, 1 or 2. These tests were stratified within study by geographical region or by principal component score, and information pooled across strata and studies as described above. A 2 df test for association, possibly modified by MHC, was calculated by adding the chi-squared interaction test on 1 df to the 1 df chi-squared statistic for the stratified association test.

The lasso analysis of the MHC risk prediction was carried out in the lasso2 package in the R statistical system28. All the remaining analysis was carried out in the snpMatrix package from the bioConductor project 29.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

This research utilizes resources provided by the Type 1 Diabetes Genetics Consortium, a collaborative clinical study sponsored by the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK), National Institute of Allergy and Infectious Diseases (NIAID), National Human Genome Research Institute (NHGRI), National Institute of Child Health and Human Development (NICHD), and Juvenile Diabetes Research Foundation

International (JDRF) and supported by U01 DK062418. Further support was provided by a grant from the NIDDK (DK46635) to PC and a joint JDRF and Wellcome Trust grant to the Diabetes and Inflammation Laboratory at Cambridge, which also received support from the National Institute for Health Research Cambridge Biomedical Research Centre. DC is the recipient of a Wellcome Trust Principal Research Fellowship.

We acknowledge the contributions of the following individuals: Julie Alipaz, Anna Simpson, Judy Brown and Joe Garsetti for assistance with project management; Matt Hardy and Kate Downes for genotyping; Meeta Maisuria for DNA sample coordination and quality control; Joan Hilner and June Pierce for managing T1DGC Data and DNA resources; James Allen, Nigel Ovington, Vin Everett, Geoff Dolman, and Mark Brown for data services and computing; and Luc Smink, Oliver Burren, Joe Mychaleckyj, and Nat Goodman for bioinformatics support.

We gratefully acknowledge the following groups and individuals who provided biological samples or data for this study. We obtained DNA samples from the British 1958 Birth Cohort collection, funded by the Medical Research Council and the Wellcome Trust. We thank The Avon Longitudinal Study of Parents and Children laboratory in Bristol and the British 1958 Birth Cohort team, including S. Ring, R. Jones, M. Pembrey, W. McArdle, D. Strachan and P. Burton for preparing and providing the control DNA samples. We thank the Human Biological Data Interchange and Diabetes UK for providing DNA samples from USA and UK multiplex families, respectively. Danish patients were from the Danish Society of Childhood Diabetes (DSBD) and control DNA samples were kindly provided by Drs. T. Hansen, O. Pedersen, K. Borch-Johnsen, and T. Joergensen. This study makes use of data generated by the Wellcome Trust Case Control Consortium, funded by Wellcome Trust award 076113, and a full list of the investigators who contributed to the generation of the data is available from www.wtccc.org.uk.

We gratefully acknowledge the Genetics of Kidneys in Diabetes (GoKinD) study for generously allowing the use of their sample SNP allele intensity and genotype data, which was obtained from the Genetic Association Information Network (GAIN) database found at http://view.ncbi.nlm.nih.gov/dbgap/ through dbGaP accession number phs000018.v1.p1 12.

We gratefully acknowledge the National Institute of Mental Health for generously allowing the use of their control CEL and genotype data. Control subjects from the National Institute of Mental Health Schizophrenia Genetics Initiative (NIMH-GI), data and biomaterials are being collected by the “Molecular Genetics of Schizophrenia II” (MGS-2) collaboration. The Investigators and co-investigators are: ENH/Northwestern University, Evanston, IL, MH059571, Pablo V. Gejman, M.D. (Collaboration Coordinator: PI), Alan R. Sanders, M.D.; Emory University School of Medicine, Atlanta, GA, MH59587, Farooq Amin, M.D. (PI); Louisiana State University Health Sciences Center; New Orleans, Louisiana, MH067257, Nancy Buccola APRN, BC, MSN (PI); University of California-Irvine, Irvine, CA, MH60870, William Byerley, M.D. (PI); Washington University, St. Louis, MO, U01, MH060879, C. Robert Cloninger, M.D. (PI); University of Iowa, Iowa, IA, MH59566, Raymond Crowe, M.D. (PI), Donald Black, M.D.; University of Colorado, Denver, CO, MH059565, Robert Freedman, M.D. (PI); University of Pennsylvania, Philadelphia, PA, MH061675, Douglas Levinson M.D. (PI); University of Queensland, Queensland, Australia, MH059588, Bryan Mowry, M.D. (PI); Mt. Sinai School of Medicine, New York, NY MH59586, Jeremy Silverman, Ph.D. (PI). The samples were collected by V L Nimgaonkar’s group at the University of Pittsburgh, as part of a multi-institutional collaborative research project with J Smoller, M.D. D.Sc. and P Sklar, M.D. Ph.D. (Massachusetts General Hospital) (grant MH 63420).

References

1. Bell GI, Horita S, Karam JH. A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes. 1984; 33:176–183. [PubMed: 6363172]
2. Bottini N, et al. A functional variant of lymphoid tyrosine phosphatase is associated with type I diabetes. Nat. Genet. 2004; 36:337–338. [PubMed: 15004560]
3. Nistico L, et al. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Hum. Mol. Genet. 1996; 5:1075–1080. [PubMed: 8817351]
4. Lowe CE, et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. Nat. Genet. 2007; 39:1074–1082. [PubMed: 17676041]
5. 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]
6. Concannon P, et al. A human type 1 diabetes susceptibility locus maps to chromosome 21q22.3. Diabetes. 2008; 57:2858–2861. [PubMed: 18647951]
7. Wellcome Trust Case Control Consortium Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. Nature. 2007; 447:661–678. [PubMed: 17554300]

8. Hakonarson H, et al. A genome-wide association study identifies KIAA0350 as a type 1 diabetes gene. Nature. 2007; 448:591–594. [PubMed: 17632545]
9. Todd JA, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. Nat. Genet. 2007; 39:857–864. [PubMed: 17554260]
10. Smyth DJ, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. N. Engl. J. Med. 2008; 359:2767–2777. [PubMed: 19073967]
11. Fung E, et al. Analysis of 17 autoimmune disease-associated variants in type 1 diabetes identifies 6q23/TNFAIP3 as a susceptibility locus. Genes and Immunity. 2008 in press.
12. Cooper JD, et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes risk loci. Nat. Genet. 2008; 40:1399–1401. [PubMed: 18978792]
13. Cooper JD, et al. Analysis of 50 autoimmune disease and type 2 diabetes loci: further confirmation of chromosomes 4q27, 12q13 and 12q24 as a type 1 diabetes loci, and support for a new locus, 12q13/KIF5A. Genes and Immunity. 2009 in press.
14. Mueller PW, et al. Genetics of Kidneys in Diabetes (GoKinD) study: a genetics collection available for identifying genetic susceptibility factors for diabetic nephropathy in type 1 diabetes. J. Am. Soc. Nephrol. 2006; 17:1782–1790. [PubMed: 16775037]
15. Baum AE, et al. A genome-wide association study implicates diacylglycerol kinase eta (DGKH) and several other genes in the etiology of bipolar disorder. Mol. Psychiatry. 2008; 13:197–207. [PubMed: 17486107]
16. Nejentsev S, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes HLA-B and HLA-A. Nature. 2007; 450:887–892. [PubMed: 18004301]
17. Smyth DJ, et al. PTPN22 Trp620 explains the association of chromosome 1p13 with type 1 diabetes and shows a statistical interaction with HLA class II genotypes. Diabetes. 2008; 57:1730–1737. [PubMed: 18305142]
18. Bjornvold M, et al. Joint effects of HLA, INS, PTPN22 and CTLA4 genes on the risk of type 1 diabetes. Diabetologia. 2008; 51:589–596. [PubMed: 18292987]
19. Senee V, et al. Mutations in GLIS3 are responsible for a rare syndrome with neonatal diabetes mellitus and congenital hypothyroidism. Nat. Genet. 2006; 38:682–687. [PubMed: 16715098]
20. Shiow LR, et al. CD69 acts downstream of interferon-alpha/beta to inhibit S1P1 and lymphocyte egress from lymphoid organs. Nature. 2006; 440:540–544. [PubMed: 16525420]
21. Power C, Elliott J. Cohort profile: 1958 British birth cohort (National Child Development Study). Int. J. Epidemiol. 2006; 35:34–41. [PubMed: 16155052]
22. Glumer C, Jorgensen T, Borch-Johnsen K. Prevalences of diabetes and impaired glucose regulation in a Danish population: the Inter99 study. Diabetes Care. 2003; 26:2335–2340. [PubMed: 12882858]
23. Teo YY, et al. A genotype calling algorithm for the Illumina BeadArray platform. Bioinformatics. 2007; 23:2741–2746. [PubMed: 17846035]
24. Mantel N. Chi-square tests with one degree of freedom: extension of the Mantel-Haenszel procedure. J. Am. Stat. Assoc. 1963; 58:690–700.
25. Clayton D. Testing for association on the X chromosome. Biostatistics. 2008; 9:593–600. [PubMed: 18441336]
26. 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]
27. Tibshirani R. Regression shrinkage and selection via the lasso. J. R. Statist. Soc. B. 1996; 58:267–288.
28. R Development Core Team. R: A language and environment for statistical computing. R Foundation for Statistical Computing; Vienna, Austria: 2008. Ref Type: Computer Program
29. Gentleman RC, et al. Bioconductor: open software development for computational biology and bioinformatics. Genome Biol. 2004; 5:R80. [PubMed: 15461798]
30. Hakonarson H, et al. A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. Diabetes. 2008; 57:1143–1146. [PubMed: 18198356]

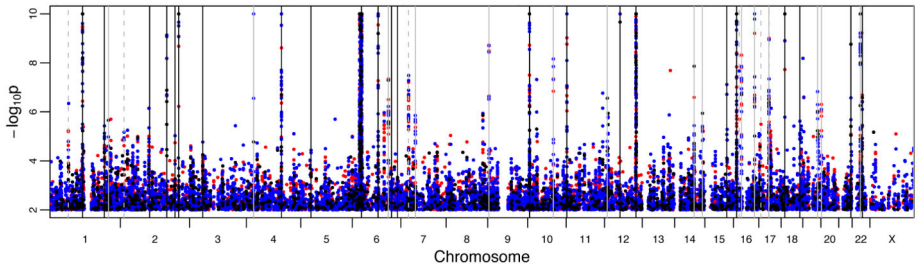

Figure 1.

Genome-wide plots of -$\log_{10}$ *P*-values from stratified 1 df tests combining results from all three studies. Values of -$\log_{10}$ *P* greater than 10 are plotted at 10. SNPs only present on the Illumina chip are shown in blue, those only present on the Affymetrix chip in red, and those present on both chips are shown in black. Points are plotted in the order red, blue, black. Previously known disease susceptibility loci are marked by vertical black lines, while new findings from the current analysis are marked by vertical grey lines (solid lines for convincingly replicated loci and dashed lines for nominally replicated results).

Table 1

Samples from three genome-wide association analyses of type 1 diabetes used in this analysis.

Subjects[a]	GWA Meta-analysis				Replication Study				Total
	T1DGC	GoKinD/ NIMH	WTCCC	Combined	T1DGC	UK	Danish	Combined	
Cases	3,983	1,601	1,930	7,514	-	2,499	1,768	4,267	11,781
Reference	3,999	1,704	3,342	9,045	-	2,690	1,980	4,670	13,715
Totals	7,982	3,305	5,272	16,559	-	5,189	3,748	8,937	25,496
Trios[b]	-	-	-	-	4,342	-	-	-	4,342

[a] The derivation of subjects from the various indicated studies is described in detail in the Methods section.

[b] From 2,319 affected sib-pair families.

Table 2

Results for locations of known susceptibility loci for type 1 diabetes.

SNP[a]	Chromosome	LD region	GWA *p*-value	Gene of Interest[c]	References
rs2476601	1p13.2	113.62-114.46	8.5×10^{-85}	*PTPN22*	2
rs2816316	1q31.2	190.73-190.82	3.1×10^{-5}	*RGS1*	10
rs917997	2q12.1	102.22-102.58	0.067[b]	*IL18RAP*	10
rs1990760	2q24.2	162.67-163.10	6.6×10^{-9}	*IFIH1*	5
rs3087243	2q33.2	204.38-204.53	1.2×10^{-15}	*CTLA4*	3
rs11711054	3p21.31	45.96-46.63	1.7×10^{-5}	*CCR5*	10
rs4505848	4q27	123.13-123.83	4.7×10^{-13}	*IL2*	9,12
rs6897932	5p13.2	35.84-36.07	0.026	*IL7R*	9
rs9268645	6p21.32	24.70-34.00	$<< 10^{-100}$	*MHC*	16
rs11755527	6q15	90.86-91.10	5.4×10^{-8}	*BACH2*	12
rs2327832	6q23.3	137.80-138.40	0.0003	*TNFAIP3*	11
rs1738074	6q25.3	159.13-159.62	0.006	*TAGAP*	10
rs12251307	10p15.1	6.07-6.24	1.3×10^{-13}	*IL2RA*	4
rs11258747	10p15.1	6.48-6.59	1.2×10^{-7}	*PRKCQ*	12
rs7111341	11p15.5	2.02-2.26	4.4×10^{-48}	*INS*	1
rs2292239	12q13.2	54.64-55.09	2.2×10^{-25}	*ERBB3*	9, 30
rs3809114	12q13.3	55.23-57.27	0.002	*multiple*	13
rs3184504	12q24.12	109.77-111.72	2.8×10^{-27}	*SH2B3*	9
rs3825932	15q25.1	76.77-77.05	7.7×10^{-8}	*CTSH*	12
rs12708716	16p13.13	10.92-11.56	2.2×10^{-16}	*CLEC16A*	8, 9
rs1893217	18p11.21	12.73-12.92	3.6×10^{-15}	*PTPN2*	9
rs763361	18q22.2	65.63-65.72	1.2×10^{-5}	*CD226*	9
rs11203203	21q22.3	42.68-42.76	1.7×10^{-9}	*UBASH3A*	6
rs229541	22q13.1	35.90-36.00	2.1×10^{-7}	*C1QTNF6*	12

[a] Focal SNP in each region was taken from the referenced studies.

[b] 2d.f. test, as this effect does not conform to a multiplicative model.

[c] The gene of interest choice was based on known expression or function in the immune system, association results from other immune-mediated diseases, the extent of the region of LD based on recombination frequencies from HapMap data, and the location of the SNPs with the highest T1D association; this selection does not infer that this is the causal gene in the region. Other genes, recombination frequency and summary association results are shown in T1DBase.

Table 3

Over-dispersion factors (λ) of 1 df association tests

Study	Simple tests	Stratified tests		
	λ	λ	$p=10^{-6}$	$p=10^{-8}$
WTCCC	1.077	1.062	2.1×10^{-6}	2.7×10^{-8}
GoKinD/NIMH	1.196	1.150	5.1×10^{-6}	9.1×10^{-8}
T1DGC	1.066	1.055	1.9×10^{-6}	2.4×10^{-8}
GB studies[a]	1.105	1.092	3.2×10^{-6}	5.0×10^{-8}
Combined[b]	1.136	1.119	3.8×10^{-6}	6.0×10^{-8}

For the stratified test λ values, the effect of genomic control correction of *p*-values of 10^{-6} and 10^{-8} are also shown.

(a) Values are shown for each study separately and for meta-analyses of both GB studies (WTCCC and T1DGC)

(b) Values are shown for each study separately and for meta-analyses of all three studies.

Table 4

Replication study of new type 1 diabetes risk loci

SNP[a]	Chr	LD region[b] (Mb)	Gene of interest (#)[c]	*P*-values			Risk Allele	MAF[e]	OR (95% CI)[f]	
				GWA[d]	Replication	Combined			Case-control	Families
rs3024505	1q32.1	204.87-205.12	*IL10* (5)	2.2×10^{-6}	0.00015	1.9×10^{-9}	C	0.169	0.84 (0.77-0.91)	0.96 (0.88-1.04)
rs10517086	4p15.2	25.64-25.75	(0)	2.8×10^{-7}	0.00021	4.6×10^{-10}	A	0.299	1.09 (1.02-1.17)	1.09 (1.02-1.16)
rs9388489	6q22.32	126.48-127.46	*C6orf173* (1)	5.1×10^{-8}	1.4×10^{-6}	4.2×10^{-13}	G	0.452	1.17 (1.10^{-1}.24)	1.05 (0.99-1.12)
rs7804356	7p15.2	26.62-27.17	(10)	3.3×10^{-8}	0.0051	5.3×10^{-9}	T	0.238	0.88 (0.82-0.94)	0.99 (0.92-1.06)
rs4948088	7p12.1	50.87-51.64	*COBL* (1)	2.7×10^{-6}	0.0019	4.4×10^{-8}	C	0.047	0.77 (0.67-0.90)	0.93 (0.79-1.10)
rs7020673	9p24.2	4.22-4.31	*GLIS3* (1)	1.9×10^{-9}	0.00013	5.4×10^{-12}	G	0.502	0.88 (0.83-0.93)	0.97 (0.91-1.03)
rs10509540	10q23.31	90.00-90.27	*C10orf59* (1)	6.9×10^{-9}	4.9×10^{-24}	1.3×10^{-28}	T	0.285	0.75 (0.70-0.80)	0.81 (0.76-0.87)
rs4763879	12p13.31	9.51-9.87	*CD69* (6)	2.8×10^{-7}	1.1×10^{-5}	1.9×10^{-11}	A	0.368	1.09 (1.02-1.16)	1.12 (1.05-1.19)
rs1465788	14q24.1	68.24-68.39	(2)	1.4×10^{-8}	1.5×10^{-5}	1.8×10^{-12}	G	0.287	0.86 (0.80-0.91)	0.95 (0.89-1.02)
rs4900384	14q32.2	97.43-97.60	(0)	1.1×10^{-6}	0.00042	3.7×10^{-9}	G	0.288	1.09 (1.02-1.16)	1.08 (1.01-1.16)
rs4788084	16p11.2	28.19-28.94	*IL27* (24)	5.2×10^{-8}	8.4×10^{-7}	2.6×10^{-13}	G	0.424	0.86 (0.81-0.91)	0.94 (0.88-1.00)
rs7202877	16q23.1	73.76-74.09	(7)	5.7×10^{-11}	1.2×10^{-6}	3.1×10^{-15}	G	0.096	1.28 (1.17-1.41)	1.09 (0.99-1.20)
rs2290400	17q12	34.63-35.51	*ORMDL3* (23)	1.3×10^{-7}	8.2×10^{-7}	5.5×10^{-13}	G	0.495	0.87 (0.82-0.93)	0.92 (0.87-0.98)
rs7221109	17q21.2	35.95-36.13	(3)	9.9×10^{-10}	0.0083	1.3×10^{-9}	C	0.353	0.95 (0.89-1.01)	0.94 (0.88-1.00)
rs425105	19q13.32	51.84-52.02	(5)	1.5×10^{-7}	2.6×10^{-5}	2.7×10^{-11}	A	0.162	0.86 (0.79-0.93)	0.90 (0.82-0.98)
rs2281808	20p13	1.44-1.71	(3)	5.0×10^{-7}	4.8×10^{-6}	1.2×10^{-11}	C	0.362	0.90 (0.84-0.95)	0.90 (0.85-0.96)
rs5753037	22q12.2	28.14-29.00	(14)	1.8×10^{-14}	5.8×10^{-5}	2.6×10^{-16}	T	0.391	1.10 (1.04-1.17)	1.08 (1.02-1.15)
rs2664170	Xq28	153.48-154.10	(16)	3.0×10^{-5}	5.8×10^{-5}	7.8×10^{-9}	G	0.316	1.16 (1.07-1.24)	1.06 (0.97-1.16)
rs2269241	1p31.3	63.87-63.94	*PGM1* (1)	5.9×10^{-6}	0.0069	4.2×10^{-7}	G	0.192	1.10 (1.02-1.18)	1.05 (0.98-1.14)
rs1534422	2p25.1	12.53-12.60	(0)	6.7×10^{-6}	0.025	2.1×10^{-6}	G	0.460	1.08 (1.02-1.15)	1.01 (0.95-1.08)
rs12444268	16p12.3	20.17-20.28	(2)	2.0×10^{-6}	0.0045	1.7×10^{-7}	A	0.295	1.10 (1.03-1.17)	1.04 (0.97-1.11)
rs16956936	17p13.1	7.56-7.66	(2)	3.2×10^{-6}	0.0097	5.3×10^{-7}	C	0.135	0.92 (0.84-1.00)	0.92 (0.83-1.01)

[a] SNPs providing evidence of association at $P < 0.05$ with T1D in replication study. SNPs showing evidence of replication at $P < 0.01$ and $P < 5 \times 10^{-8}$ overall are listed by autosome 1-22 and chromosome X (n = 18), followed by those SNPs attaining evidence of association in the replication study at $P < 0.01$ (n = 3) or 0.05 (n = 1) but failing to reach $P < 5 \times 10^{-8}$ overall.

[b] To define an LD region for a given focal SNP, we extended the region to the left until either 0.1 cM had been traversed or until reaching another SNP with $p < 10^{-6}$. In the latter case we then set this new SNP as the left bound and repeated the process. The right hand boundary was defined in the same way. However, the boundaries of the region 7p12.1 (50.87-51.64 Mb), were chosen on recombination frequency (T1DBase) and the fat that this larger interval contained all of the *COBL* gene.

[c] Gene names are shown for regions with a functionally interesting candidate or for regions with only one gene. The total number of genes in each LD region are shown in parentheses.

[d] *P*-values for stratified 1 degree of freedom tests combining data from all three GWA scans in a meta-analysis.

[e] Minor allele frequency in British controls.

[f] Odds ratio (95% CI, confidence interval). Odds ratios represent the effect of a single copy of the indicated allele within the multiplicative model for allelic effects. For rs2664170, on the X chromosome, the model fitted assumes that relative risks for males reflect those between homozygous females25.