A Human Type 1 Diabetes Susceptibility Locus Maps to Chromosome 21q22.3

Patrick Concannon,1,2 Suna Onengut-Gumuscu,2,3 John A. Todd,4 Deborah J. Smyth,4 Flemming Pociot,5 Regina Bergholdt,5 Beena Akolkar,6 Henry A. Erlich,7 Joan E. Hiiñer,8 Cécile Julier,9 Grant Morahan,10 Jørn Nerup,5 Concepcion R. Nierras,11 Wei-Min Chen,2,12 Stephen S. Rich,2 and the Type 1 Diabetes Genetics Consortium

OBJECTIVE—The Type 1 Diabetes Genetics Consortium (T1DGC) has assembled and genotyped a large collection of multiplex families for the purpose of mapping genomic regions linked to type 1 diabetes. In the current study, we tested for evidence of loci associated with type 1 diabetes utilizing genome-wide linkage scan data and family-based association methods.

RESEARCH DESIGN AND METHODS—A total of 2,496 multiplex families with type 1 diabetes were genotyped with a panel of 6,090 single nucleotide polymorphisms (SNPs). Evidence of association to disease was evaluated by the pedigree disequilibrium test. Significant results were followed up by genotyping and analyses in two independent sets of samples: 2,214 parent–affected child trio families and a panel of 7,721 case and 9,679 control subjects.

RESULTS—Three of the SNPs most strongly associated with type 1 diabetes localized to previously identified type 1 diabetes risk loci: INS, IFIH1, and KIAA0350. A fourth strongly associated SNP, rs876498 (P = 1.0 × 10−6), occurred in the sixth intron of the UBAH3A locus at chromosome 21q22.3. Support for this disease association was obtained in two additional independent sample sets: families with type 1 diabetes (odds ratio [OR] 1.06 [95% CI 1.00–1.11]; P = 0.023) and case and control subjects (1.14 [1.09–1.19]; P = 7.5 × 10−8).

CONCLUSIONS—The T1DGC 6k SNP scan and follow-up studies reported here confirm previously reported type 1 diabetes associations at INS, IFIH1, and KIAA0350 and identify an additional disease association on chromosome 21q22.3 in the UBAH3A locus (OR 1.10 [95% CI 1.07–1.13]; P = 4.4 × 10−12).

This gene and its flanking regions are now validated targets for further resequencing, genotyping, and functional studies in type 1 diabetes. Diabetes 57:2858–2861, 2008

Genome-wide linkage and association studies focused on candidate genes or interval mapping in type 1 diabetes have previously identified a handful of risk loci that have been confirmed by multiple replications. Of these, the major risk factor(s) reside within the HLA region on chromosome 6p21. Other, non-HLA type 1 diabetes loci have comparatively small effects on disease risk relative to HLA but comparable effect sizes to risk loci identified in other common disorders. These include the insulin (INS) locus, where variation is thought to impact the transcription and expression of insulin, mediating thymic selection of T-cells specific for this autoantigen (1,2). Variants at the cytotoxic T-lymphocyte antigen 4 (CTLA4) locus are implicated in type 1 diabetes risk, potentially by altering the production of differentially spliced products from the locus that affects T-cell activation (3,4). Single nucleotide polymorphisms (SNPs) in the interleukin-2 receptor α (IL2RA) region, which encodes one chain of the heterodimeric interleukin-2 cytokine receptor, are associated with type 1 diabetes (5). Finally, a nonsynonymous coding region SNP in the protein tyrosine phosphatase, nonreceptor type 22 (PTPN22) gene has been associated with type 1 diabetes as well as a number of other autoimmune disorders (6–9). This SNP, which predicts an Arg to Trp substitution at position 620, increases the activity of the encoded phosphatase LYP, resulting in hyporesponsiveness of T-cells (10,11).

Recently, the approach of genome-wide association scanning has been applied to type 1 diabetes (12–14). The combination of high-density SNP coverage and large sample sizes for both initial screening and follow-up genotyping in these studies has both confirmed known type 1 diabetes loci and identified a number of novel candidate genes and regions. New associations with type 1 diabetes have been identified and replicated for SNPs in two regions on chromosome 12, 12q24 near C12orf30 and 12q13 near erythroid blast leukemia viral oncogene homolog 3 (ERBB3); a region on 16p13 near C-type lectin domain family 16, member A, (CLEC16A/KIAA0350); a region on 18p11 near protein tyrosine phosphatase, non-receptor type 2 (PTPN2); and a region on 2q24 near interferon induced with helicase C domain 1 (IFIH1) (15,16). Detailed information about each of these type 1 diabetes–associated loci can be found at the T1DBase Web site (www.t1dbase.org).
Most of the non-HLA type 1 diabetes risk loci are characterized by ORs in the 1.15–1.3 range, and despite the recent increase in the number of such loci identified and confirmed, it is unlikely that these known loci can account for all of the genetic risk for type 1 diabetes (15). Therefore, in the current study, we took advantage of a recently completed genome-wide scan for linkage in multiplex families with type 1 diabetes to explore whether alleles at any of the 6,090 SNPs genotyped displayed evidence of association with type 1 diabetes.

RESEARCH DESIGN AND METHODS

Families used for initial screening were selected from nine different sources: the four Type 1 Diabetes Genetics Consortium (T1DGC) networks (Asia-Pacific [n = 228], Europe [n = 585], North America [n = 370], and the U.K. [n = 124]), the Diabetes U.K. Warren Repository (n = 429), the Human Biological Data Interchange repository (n = 424), the Joslin Diabetes Center (n = 112), the Steno Diabetes Center (n = 146), and Sardinia (n = 78). Details on recruitment sites and the numbers of families contributed for each of the T1DGC networks are accessible from the public side of the T1DGC Web site (www.t1dgc.org). Minimum entry criteria for families included the presence of at least one affected sibling pair, and availability of one or both biological parents was preferred. Eligibility requirements for affected individuals included documented type 1 diabetes with onset earlier than 35 years of age, insulin use within 6 months of diagnosis, and the absence of any concomitant disease or disorder associated with diabetes. In total, 2,496 families were genotyped. Both parents were available for 72.9% of families; 18.8% had only a single parent. Most families (91.7%) had exactly two affected full siblings; 4.1% had three and 0.3% had four or more affected siblings. The remaining families had pairs that were either incomplete or were half siblings.

The family replication set of samples consisted of 2,214 parent-affected child trio families derived from two sources. The Genetics of Kidneys in Diabetes (GoKinD) study of diabetic nephropathy contributed 578 families. Affected offspring in these families had type 1 diabetes diagnosed before 31 years of age, had initiated insulin therapy within 1 year of diagnosis, and had continued insulin treatment uninterrupted since initiation. The remaining 1,636 families were selected from the Danish Society of Childhood Diabetes and were all of Danish Caucasoid origin. Proband families in these families were diagnosed with type 1 diabetes before 15 years of age and had continued insulin treatment since diagnosis. The case-control replication set consisted of 17,400 subjects, including 7,721 type 1 diabetic case subjects recruited as part of the Juvenile Diabetes Research Foundation International/Wellcome Trust British cohort (21) and 9,679 control subjects from the U.K. 1968 Birth Cohort and the Wellcome Trust Case Control Blood Service (13,17).

Genotyping. Families included in the original genome-wide linkage scan were all genotyped at the Center for Inherited Disease Research using the Illumina Human Linkage-12 Genotyping Beadchip consisting of 6,090 SNPs. Genotyping of SNP rs876498 in replication samples was performed using an Eclipse SNP Genotyping Assay (Nanogen). As quality control for accuracy of this assay, a blinded resampling and genotyping of 366 families with type 1 diabetes from the initial genome-wide scan was performed.

Statistical analysis. Before statistical genetic analyses, genotype data were evaluated for Mendelian errors using PedCheck (18). PREST (19) was used to estimate the likelihood of each specified relationship in pedigrees given the genotyping information. Based on these analyses, 43 families were removed from further analysis due to the presence of a duplicate family, a duplicate sample within the family, nonresolvable family errors, or missing genotype information. Allelic association with type 1 diabetes within multiplex families was assessed using the pedigree disequilibrium test (PDT) (20), which provides a valid test of association in the presence of linkage in families with multiple affected individuals. The PDT examines the discrepancy between the alleles of heterozygote parents and those transmitted to affected offspring, as well as the allelic difference between affected and unaffected siblings. Both PDT-SUM and PDT-AVE tests (21) were carried out. Because these tests are asymptotically equivalent, the PDT-AVE test, in which counts of trios and pairs for each pedigree are weighted, is primarily reported here. PLINK (v.1.02) (22) was used to assess allelic association to type 1 diabetes in trio families, via the transmission/disequilibrium test (23), and in case-control samples. A logistic regression model was used to estimate effect sizes, with familial correlations being taken into account by the generalized estimating equation method (24).

To combine results of the original genome scan and those of follow-up studies, a weighted Z score–based fixed effects meta-analysis method was used. In brief, a Z statistic summarizing the magnitude of the P value for

| Table 1 | PDT results from 6k SNP scan at P < 0.001 |
|---|---|
| Chromosome | SNP | Position* | P | Locus† |
| 11 | rs1004446 | 2126719 | 2.5 × 10⁻⁹ | INS |
| 2 | rs1990760 | 16294558 | 4.1 × 10⁻⁵ | IFIH1 |
| 16 | rs887864 | 11066386 | 7.1 × 10⁻⁴ | KIAA0350 |
| 21 | rs876498 | 42714896 | 1.0 × 10⁻⁴ | UBAH3A |
| 11 | rs2076837 | 14931161 | 2.5 × 10⁻⁴ | |
| 10 | rs87788 | 72959546 | 3.6 × 10⁻⁴ | |
| 2 | rs6767 | 5021488 | 4.4 × 10⁻⁴ | |
| 10 | rs942434 | 7277013 | 5.8 × 10⁻⁴ | |
| 6 | rs169679 | 28964551 | 6.1 × 10⁻⁴ | HLA |
| 22 | rs885987 | 17572750 | 6.4 × 10⁻⁴ | |
| 2 | rs1504 | 36977669 | 6.8 × 10⁻⁴ | |
| 5 | rs860732 | 179360006 | 7.0 × 10⁻⁴ | |
| 6 | rs1011094 | 28883061 | 7.9 × 10⁻⁴ | HLA |
| 7 | rs1543851 | 64250422 | 8.4 × 10⁻⁴ | |
| 8 | rs1872283 | 107880067 | 9.0 × 10⁻⁴ | |
| 3 | rs353087 | 15165591 | 9.9 × 10⁻⁴ | |

*Nucleotide position on indicated chromosome; †most proximal locus to indicated SNP.

Association and direction of effect was generated for each study. An overall Z statistic was then computed as a weighted average of the individual statistics, and a corresponding P value for that statistic was computed. The weights were proportional to the square root of the number of individuals in each study and scaled such that the squared weights summed to 1. For the meta-analysis of the effect size, the inverse variance was used as the weight for each study.

RESULTS

After data cleaning, PDT was performed on the data from 5,943 markers genotyped in the family collection. Table 1 lists all SNPs with P < 0.001 in this analysis. Only a single marker, rs1004446, displayed evidence of association with type 1 diabetes at a genome-wide significance level (P = 2.5 × 10⁻⁹). This SNP is located in the INS-IGF2 region on chromosome 11p15 and has previously been reported to be associated with type 1 diabetes (14). Indeed, among the 16 SNPs listed in Table 1, 5 corresponded to previously confirmed type 1 diabetes loci identified through genome-wide scanning: rs1004446 (INS-IGF2), a nonsynonymous coding SNP rs1909760 (IFIH1), rs887864 (intron 18 of CLEC16A/KIAA0350), and two SNPs, rs169679 and rs1011094, that flank the HLA region.

Given that the top three most associated SNPs (ranked by statistical significance in Table 1) correspond to confirmed type 1 diabetes risk loci, we examined the fourth-ranked SNP, rs876498, for which there was also substantial evidence of association with type 1 diabetes (P = 1.0 × 10⁻⁴ for PDT-AVE test and P = 1.0 × 10⁻³ for PDT-SUM test). Further impetus for follow-up study of this SNP came from its location within the sixth intron of the gene UBAH3A, which is expressed predominantly in T-cells and is involved in the regulation of signaling from the antigen receptor (25). The putative disease-associated allele was relatively common, with a minor allele frequency of 0.45 in unaffected founders. Blinded validation genotyping was carried out in 366 of the original 2,496 T1DGC families by a second methodology to confirm the initial finding. A concordance rate of 100% was observed between the original and the validation study. To confirm and extend the initial finding, rs876498 was genotyped in additional independent populations, both family based and case control (Table 2). Overall, there was highly significant evidence of association between the minor
allele at rs876498 and type 1 diabetes (OR 1.10 [95% CI 1.07–1.13]; \( P = 4.4 \times 10^{-12} \)).

**DISCUSSION**

In the current study, we took advantage of available SNP genotyping data from a linkage study of 2,496 multiplex families with type 1 diabetes-affected sibling pairs. The ability to use family-based testing for association allowed the combined analysis of subjects selected from multiple different geographic areas while reducing concerns regarding the potential effects of population stratification. Two approaches were used as an additional check for potential heterogeneity from family collection in the nine T1DGC regions. First, a meta-analysis combining the nine individual PDT results provided a \( P = 2.8 \times 10^{-3} \), the same magnitude of significance as our combined data analysis. Second, a Q statistic based on the association effect and SE was computed. The association heterogeneity test (Q 5.15, 7 d.f. [one group has missing estimates]; \( P = 0.64 \)) indicates no statistically significant heterogeneity across the regions of the T1DGC data. The validity of our test for association can also be shown from the genomic control number 1.02 obtained when the whole-genome scan results were examined (26).

Although SNP coverage was extremely sparse for genome-wide association testing, an advantage was that the SNP selection in the linkage panel of markers differs from the higher density panels used in previous genome-wide association studies. Three of the top four markers, ranked by \( P \) value, corresponded to previously established type 1 diabetes risk loci. The fourth-ranked SNP mapped to chromosome 21q22.3; no locus associated at genome-wide significance levels with type 1 diabetes or any other autoimmune disorder has previously been mapped to this chromosome. The region had, however, been previously linked to type 1 diabetes in Scandinavian families (27), and a region \( \sim 10 \) Mb centromeric to rs876498 has been fine-mapped in Danish families (28). Furthermore, a type 1 diabetes risk locus on chromosome 21 might have been anticipated given the increased frequency of Downs syndrome, caused by trisomy for chromosome 21, among type 1 diabetic patients (29). The findings reported here for rs876498 were validated by a blinded 15% resample using an alternative genotyping approach and then replicated in independent sets of parent-affected child trio families and case-control subjects. These studies provide compelling evidence for association of the minor allele at this SNP, rs876498, with type 1 diabetes.

The SNP rs876498 is located within the sixth intron of a gene variously referred to as ubiquitin associated and SH3 domain containing, A (UBASH3A) (30), T-cell ubiquitin ligand (TULA) (31), and suppressor of T-cell signaling 2 (Sts-2) (32). Based on the available data in HapMap (release 23a), rs876498 is centrally located in a haplotype block of \( \sim 14 \) kb contained entirely within UBASH3A. Flanking genes in the region are not obvious candidates, on functional grounds, for involvement in type 1 diabetes etiology, and they include the trefoil domain containing family members TFF1 and TFF3; the transmembrane serine protease TMPRSS3; RSPH1, which encodes a male meiotic metaphase chromosome associated protein; and the glycerol 3-phosphate transporter SLC37A1.

UBASH3A is expressed predominantly in T-cells, where it interacts with c-CBL through its SH3 domain and binds to ubiquitin and ubiquitylated proteins via its ubiquitin-associated domain (25). In T-cells, it acts in part by inhibiting c-CBL-mediated downregulation of protein tyrosine kinases such as ZAP70 that are activated upon T-cell receptor stimulation (32). In this regard, UBASH3A is reminiscent of another identified type 1 diabetes risk locus, PTPN22, whose product, LYP, also interacts with c-CBL but directly downregulates some of the same protein tyrosine kinases by dephosphorylation (6,33). UBASH3A has other functions in T-cells, including participation in proapoptotic pathways (34). The functions of UBASH3A make it an appealing candidate to contribute to type 1 diabetes risk. While it may be premature to focus exclusively on this single gene based upon the present data, the significant and replicated findings of association reported here strongly suggest that a gene or genes in this region contributes to risk for type 1 diabetes.

**ACKNOWLEDGMENTS**

This research uses resources provided by the T1DGC, 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 P.C. 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 Centre.

**REFERENCES**

1. Vafadis P, Bennett ST, Todd JA, Nadeau J, Grabs R, Goodyer CG, Wickramasinghe S, Colle E, Polychronakos C: Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet 15:289–292, 1997
2. Pugliese A, Zeller M, Fernandez A Jr, Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD: The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet 15:293–297, 1997
3. Nistico L, Buzzetti R, Pritchard LE, Van der Auwera B, Giovannini C, Bosi E, Larrad MT, Rios MS, Chow CC, Cockram CS, Jacobs K, Mijovic C, Bain SC, Barnett AH, Vandewalle CL, Schuit F, Gorus FK, Tosi R, Pozzilli P, Wickramasinghe S, Colle E, Polychronakos C: Insulin expression in human thymus is modulated by INS VNTR alleles at the IDDM2 locus. Nat Genet 15:289–292, 1997
4. Pugliese A, Zeller M, Fernandez A Jr, Zalcberg LJ, Bartlett RJ, Ricordi C, Pietropaolo M, Eisenbarth GS, Bennett ST, Patel DD: The insulin gene is transcribed in the human thymus and transcription levels correlated with allelic variation at the INS VNTR-IDDM2 susceptibility locus for type 1 diabetes. Nat Genet 15:293–297, 1997
Todd JA, Belgian Diabetes Registry: The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. *Hum Mol Genet* 5:1075–1080, 1996

4. Ueda H, Howson JM, Esposito L, Heward J, Chamberlain G, Rainbow DB, Hunter KM, Smith AN, Di Genova G, Herr MH, Dahlme C, Payne F, Smyth D, Lowe C, Twells RC, Howlett S, Healy B, Nutland S, Rance HE, Everett V, Smink LJ, Lam AC, Cordell HJ, Walker NM, Bordwin C, Hulme J, Motzo C, Cucca F, Hess JM, Metzker ML, Rogers J, Gregory S, Allahabadia A, Nithiyanthanan R, Tuomilehto-Wolf E, Tuomilehto J, Bingley P, Gillespie KM, Undlien DE, Ronningen KS, Guja C, Ionescu-Turcu C, Savage DA,Maxwell AP, Carson DJ, Patterson CC, Franklin JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC: Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423:506–511, 2003

5. Lowe CE, Cooper JD, Brusko T, Walker NM, Smyth DJ, Bailey R, Bourget J, Plagnol V, Field S, Atkinson M, Clayton DG, Wicker LS: Todd JA: Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Rev Genet* 10:1074–1082, 2007

6. Bottini N, Musuneci L, Alonso A, Rahmouni S, Nika K, Rostamkhani M, MacMurray J, Meloni GF, Lucarelli P, Pellecchia M, Eisenbarth GS, Comings D, Mustelin T: A functional variant of lymphoid tyrosine phosphatase is associated with type 1 diabetes. *Nat Genet* 36:337–338, 2004

7. Begovich AB, Carlton VE, Honigberg LA, Schrodi SJ, Chokkalingam AP, Alexander HC, Ardlie KG, Huang Q, Smith AM, Sporer JM, Conm MT, Chang M, Chang SY, Saiik RK, Catanesi JG, Leonard DU, Garcia VE, McAllister LB, Jeffery DA, Lee AT, Batillawa F, Remmers E, Criswell LA, Seldin MF, Kastner DL, Ames CI, Sninsky JJ, Gregersen PK: A nonsense single-nucleotide polymorphism in a gene encoding a protein tyrosine phosphatase (PTPN22) is associated with rheumatoid arthritis. *Am J Hum Genet* 75:329–337, 2004

8. Kyogoku C, Langefeld CD, Ortmann WA, Lee A, Selby S, Carlton VE, Chang M, Ramos P, Raecheil EC, Batillawa FM, Novitzke J, Williams AH, Gillett C, Rodine P, Graham RR, Ardlie KG, Gaffney FM, Moser KL, Petriti M, Begovich AB, Gregersen PK, Behrens TW: Genetic Association of the R620W Polymorphism of Protein Tyrosine Phosphatase PTPN22 with Human SLE. *Am J Hum Genet* 75:504–507, 2004

9. Criswell LA, Pfeiffer KA, Lum RF, Gonzales B, Novitzke J, Kern M, Moser KL, Begovich AB, Carlton VE, Li W, Lee AT, Ortmann W, Behrens TW, Gregersen PK: Analysis of families in the multiple autoimmune disease genetics consortium (MADGC) collection: the PTPN22 620W allele associates with multiple autoimmune phenotypes. *Am J Hum Genet* 76:561–571, 2005

10. Vang T, Congia M, Macis MD, Musuneci L, Orru V, Zavattari P, Nika K, Tautu L, Tatedo K, Cucca F, Mustelin T, Bottini N: Autoimmune-associated lymphoid tyrosine phosphatase is a gain-of-function variant. *Nat Genet* 37:1317–1319, 2005

11. Rieck M, Arechiga A, Onengut-Gumuscu S, Greenbaum C, Concannon P, Buckner JH: Genetic variation in PTPN22 corresponds to altered function of T and B lymphocytes. *J Immunol* 179:1704–1710, 2007

12. Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ, Guja C, Ionescu-Turcu C, Savage DA, Maxell AP, Carson DJ, Patterson CC, Franklyn JA, Clayton DG, Peterson LB, Wicker LS, Todd JA, Gough SC: Association of the T-cell regulatory gene CTLA4 with susceptibility to autoimmune disease. *Nature* 423:506–511, 2003

13. Wellcome Trust Case Control Consortium: Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 447:671–678, 2007

14. Hakonarson H, Grant SF, Bradford JP, Marchand L, Kim CE, Glessner JT, Grahs B, Casahumovo T, Taback SP, Francelko EC, Lawson ML, Otieno FG, Santa E, Shaner JL, Smith RM, Ortyiah CC, Skraban R, Chiavacci RM, Robinson LJ, Stanley CA, Kirsch SE, DeVito M, Monos DS, Grant SF, Polychronakos C: A novel susceptibility locus for type 1 diabetes on Chr12q13 identified by a genome-wide association study. *Diabetes* 57:1143–1146, 2008

15. Power C, Elliott J: Cohort profile: 1958 British birth cohort (National Child Development Study). *Int J Epidemiol* 35:34–41, 2006