Localization of a Type 1 Diabetes Locus in the *IL2RA/CD25* Region by Use of Tag Single-Nucleotide Polymorphisms

Adrian Vella,[1] Jason D. Cooper,[1] Christopher E. Lowe,[1] Neil Walker,[1] Sarah Nutland,[1] Barry Widmer,[2] Richard Jones,[3] Susan M. Ring,[3] Wendy McArdle,[3] Marcus E. Pembrey,[3,4] David P. Strachan,[5] David B. Dunger,[2] Rebecca C. J. Twells,[1] David G. Clayton,[1] and John A. Todd[1]

[1]Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, and [2]Department of Paediatrics, Addenbrooke's Hospital, University of Cambridge, Cambridge, United Kingdom; [3]Department of Community-Based Medicine, Bristol University, Bristol, United Kingdom; and [4]Clinical and Molecular Genetics Unit, Institute of Child Health, University College London, and [5]Department of Community Health Sciences, St. George's Hospital Medical School, London

As part of an ongoing search for genes associated with type 1 diabetes (T1D), a common autoimmune disease, we tested the biological candidate gene *IL2RA* (*CD25*), which encodes a subunit (IL-2Rα) of the high-affinity interleukin-2 (IL-2) receptor complex. We employed a tag single-nucleotide polymorphism (tag SNP) approach in large T1D sample collections consisting of 7,457 cases and controls and 725 multiplex families. Tag SNPs were analyzed using a multilocus test to provide a regional test for association. We found strong statistical evidence in the case-control collection ($P = 6.5 \times 10^{-8}$) for a T1D locus in the *CD25* region of chromosome 10p15 and replicated the association in the family collection ($P = 7.3 \times 10^{-3}$; combined $P = 1.3 \times 10^{-10}$). These results illustrate the utility of tag SNPs in a chromosome-regional test of disease association and justify future fine mapping of the causal variant in the region.

Introduction

Despite hundreds of association studies, few have been consistently replicated (Dahlman et al. 2002; Hirschhorn et al. 2003; Ioannidis et al. 2003; Lohmueller et al. 2003). In type 1 diabetes (T1D [MIM 222100]), only four loci have been identified and successfully replicated: the HLA class II genes on chromosome 6p21 (Cucca et al. 2001), the insulin gene (*INS*) on chromosome 11p15 (Bell et al. 1984; Barratt et al. 2004), the CTLA-4 gene on chromosome 2q33 (Nisticó et al. 1996; Ueda et al. 2003), and the recently associated PTPN22 gene on chromosome 1p13 (Bottini et al. 2004; Smyth et al. 2004). It is now generally accepted that large numbers of individuals and more stringent criteria for interpreting association studies are required to ensure reliable detection of association (Dahlman et al. 2002; Ioannidis et al. 2003; Lohmueller et al. 2003; Thomas and Clayton 2004; Wacholder et al. 2004; Freimer and Sabatti 2004; Smyth et al. 2005; Wang et al. 2005).

Most cases of T1D result from immune-mediated destruction of the insulin-producing β cells of the pancreas in an inflammatory process that involves many cell types of the immune system, including T lymphocytes. The four identified T1D loci underpin these known features of the disease, since they are all involved in T cell development (for *INS*, in terms of thymic tolerance of the insulin molecule), activation, expansion, and regulation (Todd and Wicker 2001; Ueda et al. 2003; Anjos and Polychronakos 2004). As part of an ongoing search for genes associated with T1D, we tested the biological candidate gene *CD25* (MIM 147730). In common with the T1D loci identified elsewhere, *CD25* is central to immune regulation. *CD25* expression on regulatory T cells is essential for their function in suppressing T cell immune responses and autoimmune disease (Salomon et al. 2000; Malek and Bayer 2004; Viglietta et al. 2004). Additionally, in humans, a rare mutation of *CD25* caused severe autoimmune disease (Sharfe et al. 1997).

We adopted a linkage disequilibrium (LD)-mapping approach to test for an association between T1D and the *CD25* region, using tag SNPs (Johnson et al. 2001; Chapman et al. 2003; Clayton et al. 2004) in large case-control and family collections. Elsewhere, we have shown that the use of tag SNPs can reduce genotyping costs by approximately two-thirds (Chapman et al. 2003; Clayton et al. 2004; Lowe et al. 2004).

Received December 13, 2004; accepted for publication February 21, 2005; electronically published March 17, 2005.

Address for correspondence and reprints: Prof. John A. Todd, Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, University of Cambridge, Cambridge Institute for Medical Research, Wellcome Trust/Medical Research Council Building, Addenbrooke's Hospital, Hills Road, Cambridge, CB2 2XY, United Kingdom. E-mail: john.todd@cimr.cam.ac.uk

0002-9297/2005/7605-0006$15.00

Subjects and Methods

Subjects

The resequencing panel consisted of 32 CEPH individuals (Utah residents with ancestry from northern and western Europe) (Fondation Jean Dausset–CEPH).

The 3,527 cases were recruited as part of the United Kingdom Genetic Resource Investigating Diabetes (U.K. GRID) study, which is a joint project between the University of Cambridge Department of Paediatrics and the Department Medical Genetics at the Cambridge Institute for Medical Research and is funded by the Juvenile Diabetes Research Foundation and the Wellcome Trust. The eventual aim of the project is to collect 8,000 cases with T1D for comparison with 8,000 controls from the 1958 British Birth Cohort (1958 BBC), to allow well-powered genetic association studies. The 1958 BBC is an ongoing follow-up of all persons born in Great Britain during one week in 1958 (National Child Development Study), including a recent biomedical assessment during 2002–2004 at which blood samples and informed consent were obtained for creation of a genetic resource. All cases were white, and at least 97% of the controls were of white ethnicity.

All families were of white European descent and were composed of two parents and at least one affected child. The population studied consisted of 472 multiplex families from the Diabetes United Kingdom Warren collection and 268 multiplex families from the (U.S.) Human Biological Data Interchange. The characteristics and inclusion criteria for each family collection have been described elsewhere (Vella et al. 2004), and these and reference to the case-control samples can be obtained from the Juvenile Diabetes Research Foundation/Wellcome Trust Web site. All DNA samples were collected with approval of the relevant research ethics committees, and written informed consent was obtained from the participants.

Identifying Polymorphisms

To identify polymorphisms in the CD25 gene, the exons, exon/intron boundaries, and up to 3 kb of 3′ and 5′ flanking sequence were resequenced in DNA samples from 32 CEPH DNA samples by use of nested PCR products; we also sequenced the regions −9,000 to −8,000, −4,000 to −3,000, and +3,000 to +4,000 bases, numbered relative to the transcription start site, to encompass the CD28 response element (CD28rE), the Positive Regulatory Region (PRR) III, and PRR IV, respectively (Toledano et al. 1990; Kim and Leonard 2002). In total, 15 kb was sequenced from the *CD25* region, spanning ~60 kb. The sequencing reactions were performed using the Applied Biosystems (ABI) BigDye (version 3.1) chemistry, and the sequences were analyzed using an ABI 3700 capillary sequencer. Analysis of the sequence traces was performed using the Staden package (Bonfield et al. 1998) and was double scored by a second operator. All sequence information and primer locations are provided at the T1DBase Web site.

Tag SNPs

As described elsewhere (Chapman et al. 2003; Clayton et al. 2004), we used the resequencing genotype data to investigate the ability of smaller subsets of SNPs to predict the genotypes of the remainder. Predictive performance was assessed using a locus R^2 measure (coefficient of determination), which measures the ability to predict each known SNP genotype by linear regression on the tag SNP genotypes (Chapman et al. 2003). We considered only SNPs with a minor-allele frequency (MAF) $\geqslant 5\%$ and required that the subset of tag SNPs predict the remaining SNPs with a minimum R^2 of 0.8.

We selected an optimal set of tag SNPs, using a mixture of step-up, step-down, and exhaustive subset search algorithms. Since the exhaustive subset search procedure can be slow, we initially identified a set of tag SNPs selected by both step-up and step-down searches, and we determined the best additional set of tag SNPs by exhaustive subset search of the remaining SNPs (Lowe et al. 2004). The programs for the selection and analysis of tag SNPs are implemented in STATA and can be downloaded from D.G.C.'s Web site.

Genotyping

Tag SNPs were genotyped using either Taqman (Applied Biosystems) or Invader (Third Wave Technologies) technologies on a British case-control collection (3,527 cases and 3,930 controls), in accordance with the manufacturers' protocols. All genotyping data were double scored to minimize error.

Multilocus Test for Association

Chapman et al. (2003) suggested the use of a multivariate test statistic in the analysis of a tagged region. Essentially, the test contrasts the profile of tag SNP allele frequencies between cases and controls by use of Hotelling's T^2 test (Xiong et al. 2002; Chapman et al. 2003; Fan and Knapp 2003). The test does not assume Hardy-Weinberg equilibrium in cases and controls; since no imputation of haplotype phase is required, variance and covariances of genotypes are estimated empirically. In the analysis of the case-control collection, the multilocus test was stratified by broad geographical region within Great Britain to exclude the possibility of confounding by geography. For each of 12 regions, we computed the vector of contrasts between case and control allele frequencies. The final test was based on a weighted sum of these contrast vectors, with weights inversely proportional to variance. This procedure is a multivariate generali-

Table 1

Polymorphisms Identified in *CD25*

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

zation of standard methods for control of confounding in epidemiological studies (Breslow and Day 1980; Clayton and Hills 1993).

In the analysis of the family collection, parent-child trios were analyzed by a very closely related procedure. Each trio contributed a vector of transmitted allele pairs and a vector of untransmitted allele pairs. These can be thought of as "cases" and matched "pseudocontrols," respectively. Each pair was then scored 0, 1, or 2, and the case and control profiles were compared using a paired Hotelling's T^2 test (Chapman et al. 2003).

These tests are most powerful when the effect of the causal variant is codominant. A recessive mode of inheritance results in reduced heterozygosity at the causal locus, but this may be reflected only weakly at any one tag SNP. Chapman et al. (2003) suggested a method for incorporation of this information into the multilocus test, but we did not judge it to be necessary in our study.

The evidence from case-control and family collections was combined in the same manner as was used to amalgamate the evidence from different geographical regions in the analysis of the case-control collection; a vector of contrasts comparing case and control allele frequencies for the set of tag SNPs was contributed from both studies, and a weighted mean of this vector was computed and tested against 0 (appendix A).

Imputation of Missing Tag SNP Genotypes

The multilocus test takes account of correlations between genotype for different tags and, therefore, requires that a complete set of scores be entered for each subject. Even a modest genotyping failure rate can result in a substantial attrition of subjects for such "complete case" analyses and substantial loss of power. We have avoided this by imputation of missing values.

Imputation of missing genotypes was performed using linear regression; that is, the missing tag SNP, $t_i (i = 1,2,\ldots,n)$, genotypes (scored 0, 1, or 2) were predicted from the regression of t_i on the set of complete tag SNP genotypes, excluding t_i. Clayton et al. (2004) justify this procedure for high-LD regions. Imputation was performed under the null hypothesis so that, for the same genotypes at observed loci, a missing locus would be assigned the same score whether the subject were a case or a control. The effect of this is to shrink case-control differences toward zero, but, since their variances and covariances are estimated empirically, the size of the test is preserved.

We evaluated the effect of imputation on type 1 error rates and on power in a simulation study. A number of scenarios, defined by the number of tag SNPs (maximum 20) and the percentage of missing at-random tag SNP genotypes (5% or 10%), in both case-control samples (1,000 cases and 1,000 controls) and families (1,000 parent-child trios), were considered. The effect of imputation on type 1 error rates in data generated under the null hypothesis of no association was evaluated on the basis of how often the null hypothesis was rejected when a significance test with a critical P value of P_1 was applied. The null hypothesis should be rejected with probability P_1. The power, with and without the imputation of missing tag SNP genotypes, was evaluated on the basis of how often the null hypothesis was rejected when applied to data generated under the alternative hypothesis.

Results

The resequencing of *CD25* for 32 CEPH individuals identified 55 polymorphisms (table 1), 54 of which were SNPs; 13 of these SNPs were novel when compared with dbSNP build 123, and 1 polymorphism, *ss35031434,* was a novel G insertion/deletion. Sixteen SNPs had an MAF <5% and were consequently not included in the tag SNP selection. From the 39 common SNPs (MAF $\geq$5%), 20 tag SNPs were selected and genotyped in the case-control collection (table 2). All tag SNP genotypes in cases and controls were in Hardy-Weinberg equilibrium.

The multilocus test P value for the case-control collection was 6.5×10^{-8} (3,521 case and 3,930 control genotypes; $F_{20,7419} = 3.7$). The multilocus test was stratified by 12 broad geographic regions, to minimize any confounding due to variation in allele frequencies across Great Britain (see the "Subjects and Methods" section) (unstratified P value was 1.4×10^{-8}).

We proceeded to genotype the tag SNPs in a large family collection (472 British and 268 U.S. multiplex families with T1D). tag SNP genotypes in parents and affected offspring were all in Hardy-Weinberg equilibrium. We replicated the association in the family collection, with a multilocus test P value of 7.3×10^{-3} (parent-child trio genotypes = 1,378; $\chi^2_{20} = 38.7$) (table 3), thus providing independent evidence of an association between T1D and *CD25*. Figure 1 shows the striking agreement between the odds ratios and the relative risks (transmission ratios) for the minor alleles of the tag SNPs genotyped in the case-control and family collections. Consequently, when results from both studies were combined, the multilocus test (Smyth et al. 2004)

Table 2

Summary of Tag SNPs for the Case-Control Collection

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

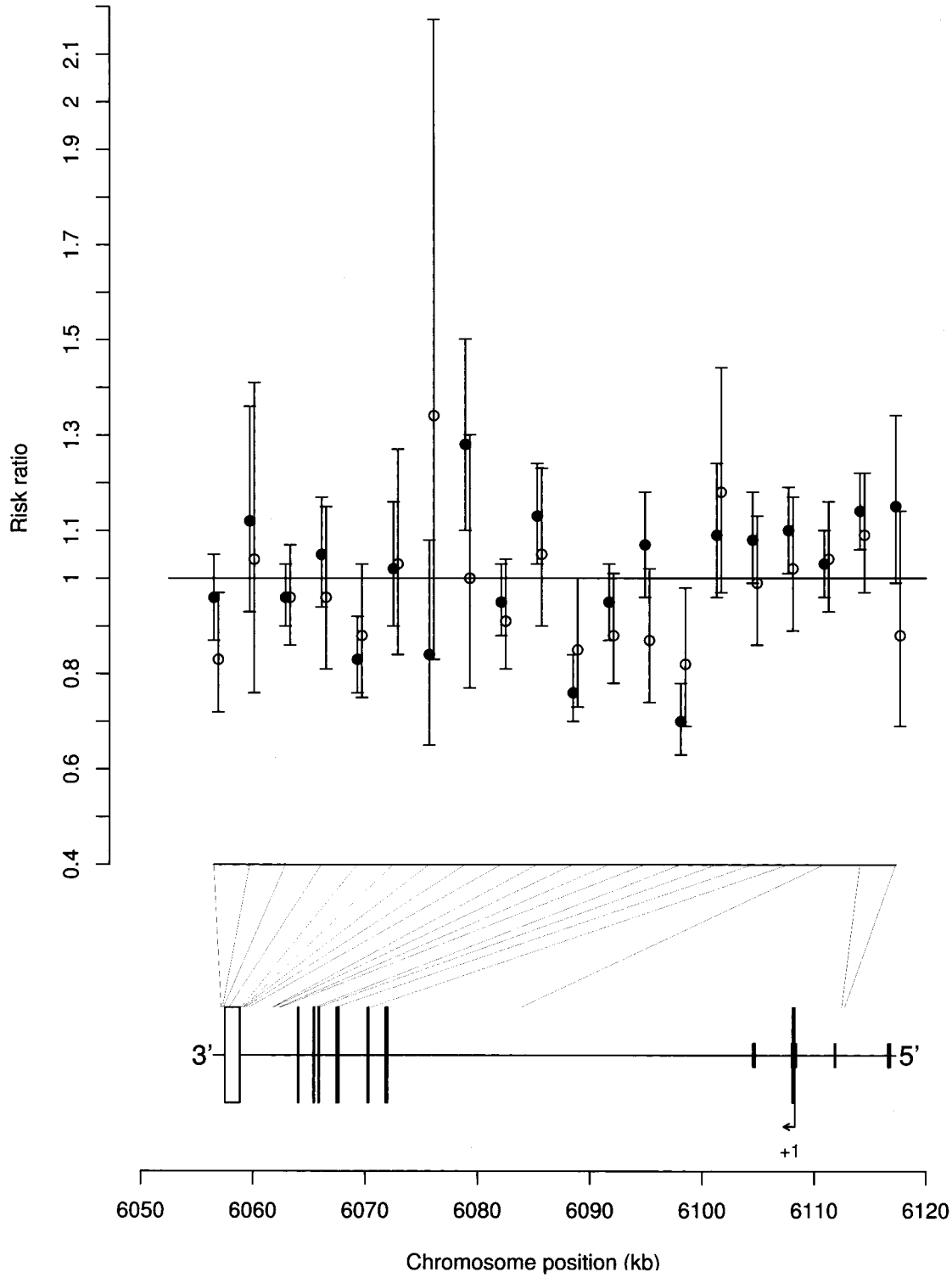

Figure 1 *Upper panel*, Odds ratios and transmission ratios for the minor allele of *CD25* tag SNPs genotyped in the case-control (*filled circles*) and family (*open circles*) collections, respectively. Vertical lines indicate 95% CIs. *Lower panel*, Chromosome position of *CD25* tag SNPs. Open long rectangle indicates UTR, filled long rectangles indicate exons, filled short rectangles indicate regulatory regions, and the arrow labeled "+1" indicates the transcript start site. A version of this figure can be viewed at the T1DBase Web site.

Table 3

Summary of Tag SNPs for the Family Collection

The table is available in its entirety in the online edition of *The American Journal of Human Genetics*.

(appendix A) provided strong statistical support for a T1D locus in the *CD25* region of chromosome 10p15.1 ($\chi^2_{20} = 88.6$; $P = 1.3 \times 10^{-10}$). There was no suggestion of reduced heterozygosity in cases or affected offspring, an indication of recessive inheritance.

Simulations of data generated under the null hypothesis indicated that imputation did not affect type 1 error rates; that is, the null hypothesis was rejected with probability P_1 when a significance test with a critical P value of P_1 was applied. For example, the null hypothesis was rejected in 483, 102, and 9 of 10,000 simulations of 1,000 cases and 1,000 controls genotyped in 20 tag SNPs, with 5% of tag SNP genotypes missing at critical P values of 0.05, 0.01, and 0.001, respectively. Simulations also indicated that imputation partially recaptures the loss of power incurred by restricting the analysis to subjects with a complete set of tag SNP genotypes (table 4). For example, if a genomic region has 20 tag SNPs that are genotyped in 1,000 cases and 1,000 controls with 5% of tag SNP genotypes missing at random, by imputing missing genotypes, power to detect a causal variant with an odds ratio of 1.5 (causal allele frequency of 0.35 in controls) at a significance level of $P = .05$ increases from 46.5% to 96.0% in 1,000 simulations. In the equivalent exercise in families (1,000 parent-child trios), power increases from 12.9% to 99.1%. Imputation is particularly important for parent-child trios and for candidate genes with a relatively large number of tag SNPs, since, in those cases, restriction of the analysis to complete cases is particularly damaging. The multilocus test P values for the case-control and family collections without imputation were 2.1×10^{-6} (2,812 case and 2,981 control genotypes; $F_{20,5761} = 3.2$) and .052 (parent-child trio genotypes = 558; $\chi^2_{20} = 31.2$), respectively.

Discussion

In the present study, we have assessed only the polymorphisms located in or close to exons and known regulatory regions, as well as up to 3 kb of 3′ and 5′ flanking sequence. Since we have found strong statistical evidence for a T1D locus in the *CD25* region, we have now started to resequence an extended and, where possible, contiguous *CD25* chromosome region of ~190 kb including *CD25* to identify potential causal variants. Fine mapping of an extended region is required, since the *CD25* tag SNPs could be in LD with a causal variant beyond *CD25*. For example, *IL15RA*, a strong functional candidate (Fehniger and Caligiuri 2001), is ~30 kb from the 3′ flanking sequence of *CD25*.

Since the causal variant(s) in the *CD25* region remains unknown, to replicate the association with T1D reported in the present study, a tag SNP approach would be required, genotyping either the same or a new set of tag SNPs. A new selection could be required—when the population in a subsequent study has a different pattern of LD in the *CD25* region, for instance. The temptation to ignore the foundations of the reported association, the set of tag SNPs, and the pattern of LD behind their

Table 4

Effect of Imputation on Power of a Simulated Case-Control Study and of a Family Study

No. of Tag SNPs and P	Percentage of Tag SNP Genotypes Missing at Random							
	In Case-Control Study with				In Family Study with			
	No Imputation		Imputation		No Imputation		Imputation	
	5%	10%	5%	10%	5%	10%	5%	10%
4:								
.05	99.9	98.8	100.0	99.8	97.3	78.6	99.8	99.9
.01	98.6	95.4	99.6	99.3	90.8	62.5	98.7	97.8
.001	94.3	84.8	98.1	96.6	75.2	35.3	95.5	92.0
12:								
.05	89.7	57.5	99.7	99.5	38.9	3.3	99.3	98.6
.01	74.2	33.2	97.8	97.3	19.0	.1	95.8	93.6
.001	48.7	13.1	92.4	91.0	3.3	.0	85.1	79.7
20:								
.05	46.5	16.9	96.0	94.7	12.9	.0	99.1	97.7
.01	24.1	5.2	88.1	85.3	2.1	.0	95.3	92.4
.001	7.9	.6	67.7	65.2	.1	.0	80.9	78.0

Note.—Power was estimated from 1,000 simulations of each type of study (1,000 each, cases, controls, and parent-child trios). The causal variant had an OR of 1.5 and a causal-allele frequency of ~35% in controls and parents.

selection and to genotype only the most-associated tag SNP may well lead to false-negative results, since the power of the present study is based on the complete set of tag SNPs.

Acknowledgments

We gratefully acknowledge the participation of all patients with T1D, their family members, and control individuals. We thank Helen Rance, Tasneem Hassanali, Jayne Hutchings, Gillian Coleman, Sarah Field, Trupti Mistry, Kirsi Bourget, Sally Clayton, Matthew Hardy, Jennifer Keylock, Pamela Lauder, Meeta Maisuria, William Meadows, Meera Sebastian, and Sarah Wood, for preparing DNA samples; the Avon Longitudinal Study of Parents and Children Laboratory in Bristol, for preparing DNA samples; Luc Smink, Oliver Burren, and Alex Lam, for bioinformatic support. We acknowledge use of DNA from the 1958 BBC collection, funded by Medical Research Council grant G0000934 and Wellcome Trust grant 068545/Z/02. We acknowledge Diabetes UK and the Human Biological Data Interchange, for multiplex family collections, and the Wellcome Trust, the Juvenile Diabetes Research Foundation, Diabetes UK, and the Medical Research Council, for financial support. A.V. is a Mayo Foundation Scholar, and D.G.C. is a Juvenile Diabetes Research Foundation/Wellcome Trust Principal Research Fellow.

Appendix A

In an earlier study (Smyth et al. 2004), we used and defined a score test to combine single-locus tests from family- and population-based studies; in the present article, we used a multivariate version of that test. Chapman et al. (2003) defined the multilocus test as a multivariate score test, $T^2 = \mathbf{U}^T\mathbf{V}^\theta\mathbf{U}$, where $\mathbf{U}$ is a score vector—with one element for each tag SNP, contrasting allele frequencies in cases and controls or, in family studies, frequencies of transmitted and untransmitted alleles. $\mathbf{V}$ is the estimated variance of the score statistic, and θ denotes a generalized inverse. The test statistic is asymptotically distributed as χ^2, with degrees of freedom equal to the rank of $\mathbf{V}$, which is equal to the number of tag SNPs. When combining results from family- and population-based studies, we first calculate the $\mathbf{U}$ vector and $\mathbf{V}$ matrix for each study. We then calculate an overall $\mathbf{U}$ and $\mathbf{V}$ by summation of the contributions from each study—$\mathbf{U} = \mathbf{U}_1 + \mathbf{U}_2$ and $\mathbf{V} = \mathbf{V}_1 + \mathbf{V}_2$—and calculate the T^2 test statistic as before.

Electronic-Database Information

The URLs for data presented herein are as follows:

dbSNP Home Page, http://www.ncbi.nlm.nih.gov/SNP/index.html

D.G.C.'s Web site, http://www-gene.cimr.cam.ac.uk/clayton/software/ (for the programs used for the selection and analysis of tag SNPs)

Fondation Jean Dausset–CEPH, http://www.cephb.fr/ (for information about the individuals used in the sequencing panel)

Juvenile Diabetes Research Foundation/Wellcome Trust, http://www-gene.cimr.cam.ac.uk/todd/dna-refs.shtml (for references to DNA collections used at the Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory)

National Center for Biotechnology Information (NCBI), http://www.ncbi.nlm.nih.gov/

National Child Development Study, http://www.cls.ioe.ac.uk/Cohort/Ncds/mainncds.htm (for information about the control collection)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for T1D and *CD25*)

STATA, http://www.stata.com (for statistical software used)

T1DBase, http://www.t1dbase.org/cgi-bin/welcome.cgi (for further information about the *CD25* region)

U.K. GRID, http://www-gene.cimr.cam.ac.uk/ucdr/grid.shtml (for information about the case collection)

References

Anjos S, Polychronakos C (2004) Mechanisms of genetic susceptibility to type I diabetes: beyond HLA. Mol Genet Metab 81:187–195

Barratt BJ, Payne F, Lowe CE, Hermann R, Healy BC, Harold D, Concannon P, Gharani N, McCarthy MI, Olavesen MG, McCormack R, Guja C, Ionescu-Tirgoviste C, Undlien DE, Ronningen KS, Gillespie KM, Tuomilehto-Wolf E, Tuomilehto J, Bennett ST, Clayton DG, Cordell HJ, Todd JA (2004) Remapping the insulin gene/*IDDM2* locus in type 1 diabetes. Diabetes 53:1884–1889

Bell GI, Horita S, Karam JH (1984) A polymorphic locus near the human insulin gene is associated with insulin-dependent diabetes mellitus. Diabetes 33:176–183

Bonfield JK, Rada C, Staden R (1998) Automated detection of point mutations using fluorescent sequence trace subtraction. Nucleic Acids Res 26:3404–3409

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

Breslow NE, Day NE (1980) Statistical methods in cancer research: volume 1—the analysis of case-control studies. IARC Scientific Publications, Lyon

Chapman JM, Cooper JD, Todd JA, Clayton D (2003) Detecting disease associations due to linkage disequilibrium using haplotype tags: a class of tests and the determinants of statistical power. Hum Hered 56:18–31

Clayton D, Chapman J, Cooper J (2004) Use of unphased multilocus genotype data in indirect association studies. Genet Epidemiol 27:415–428

Clayton D, Hills M (1993) Statistical models in epidemiology. Oxford Science Publications, Oxford, United Kingdom

Cucca F, Lampis R, Congia M, Angius E, Nutland S, Bain SC, Barnett AH, Todd JA (2001) A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes

and the structure of their proteins. Hum Mol Genet 10:2025–2037

Dahlman I, Eaves IA, Kosoy R, Morrison VA, Heward J, Gough SC, Allahabadia A, Franklyn JA, Tuomilehto J, Tuomilehto-Wolf E, Cucca F, Guja C, Ionescu-Tirgoviste C, Stevens H, Carr P, Nutland S, McKinney P, Shield JP, Wang W, Cordell HJ, Walker N, Todd JA, Concannon P (2002) Parameters for reliable results in genetic association studies in common disease. Nat Genet 30:149–150

Fan R, Knapp M (2003) Genome association studies of complex diseases by case-control designs. Am J Hum Genet 72: 850–868

Fehniger TA, Caligiuri MA (2001) Interleukin 15: biology and relevance to human disease. Blood 97:14–32

Freimer N, Sabatti C (2004) The use of pedigree, sib-pair and association studies of common diseases for genetic mapping and epidemiology. Nat Genet 36:1045–1051

Hirschhorn JN (2003) Genetic epidemiology of type 1 diabetes. Pediatr Diabetes 4:87–100

Ioannidis JPA, Trikalinos TA, Ntzani EE, Contopoulos-Ioannidis DG (2003) Genetic associations in large versus small studies: an empirical assessment. Lancet 361:567–571

Johnson GCL, Esposito L, Barratt BJ, Smith AN, Heward J, Di Genova G, Ueda H, Cordell HJ, Eaves IA, Dudbridge F, Twells RCJ, Payne F, Hughes W, Nutland S, Stevens H, Carr P, Tuomilehto-Wolf E, Tuomilehto J, Gough SCL, Clayton DG, Todd JA (2001) Haplotype tagging for the identification of common disease genes. Nat Genet 29:233–237

Kim HP, Leonard WJ (2002) The basis for TCR-mediated regulation of the IL-2 receptor α chain gene: role of widely separated regulatory elements. EMBO J 21:3051–3059

Lohmueller KE, Pearce CL, Pike M, Lander ES, Hirschhorn (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. Nat Genet 33:177–182

Lowe CE, Cooper JD, Chapman JM, Barratt BJ, Twells RCJ, Green EA, Savage DA, Guja C, Ionescu-Tîrovişte C, Tuomilehto-Wolf E, Tuomilehto J, Todd JA, Clayton DG (2004) Cost-effective analysis of candidate genes using htSNPs: a staged approach. Genes Immun 5:301–305

Malek TR, Bayer AL (2004) Tolerance, not immunity, crucially depends on IL-2. Nat Rev Immunol 4:665–674

Nisticó 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, Todd JA (1996) The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Hum Mol Genet 5:1075–1080

Salomon B, Lenschow DJ, Rhee L, Ashourian N, Singh B, Sharpe A, Bluestone JA (2000) B7/CD28 costimulation is essential for the homeostasis of the CD4+CD25+ immunoregulatory T cells that control autoimmune diabetes. Immunity 12:431–440

Sharfe N, Dadi HK, Shahar M, Roifman CM (1997) Human immune disorder arising from mutation of the α chain of the interleukin-2 receptor. Proc Natl Acad Sci USA 94:3168–3171

Smyth D, Cooper JD, Collins JE, Heward JM, Franklyn JA, Howson JM, Vella A, Nutland S, Rance HE, Maier L, Barratt BJ, Guja C, Ionescu-Tirgoviste C, Savage DA, Dunger DB, Widmer B, Strachan DP, Ring SM, Walker N, Clayton DG, Twells RC, Gough SC, Todd JA (2004) 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 53:3020–3023

Smyth D, Howson JM, Lowe CE, Walker N, Lam AC, Nutland S, Hutchings J, Tuomilehto-Wolf E, Tuomilehto J, Guja C, Ionescu-Tirgoviste C, Undlien DE, Ronningen KS, Savage DA, Dunger DB, Twells RC, McArdle W, Strachan DP, Todd JA (2005) Assessing the validity of the association between the SUMO4 M55V variant and risk of type 1 diabetes. Nat Genet 37:110–111

Thomas DC, Clayton DG (2004) Betting odds and genetic associations. J Natl Cancer Inst 96:421–423

Todd JA, Wicker LS (2001) Genetic protection from the inflammatory disease type 1 diabetes in humans and animal models. Immunity 15:387–395

Toledano MB, Roman DG, Halden NF, Lin BB, Leonard WJ (1990) The same target sequences are differentially important for activation of the interleukin 2 receptor α-chain gene in two distinct T-cell lines. Proc Natl Acad Sci USA 87:1830–1834

Ueda H, Howson JM, Esposito L, Heward J, Snook H, Chamberlain G, Rainbow DB, et al (2003) Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease. Nature 423:506–511

Vella A, Howson JM, Barratt BJ, Twells RC, Rance HE, Nutland S, Tuomilehto-Wolf E, Tuomilehto J, Undlien DE, Ronningen KS, Guja C, Ionescu-Tirgoviste C, Savage DA, Todd JA (2004) Lack of association of the Ala(45)Thr polymorphism and other common variants of the NeuroD gene with type 1 diabetes. Diabetes 53:1158–1161

Viglietta V, Baecher-Allan C, Weiner HL, Hafler DA (2004) Loss of functional suppression by CD4+CD25+ regulatory T cells in patients with multiple sclerosis. J Exp Med 199: 971–979

Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N (2004) Assessing the probability of false-positive reports in molecular epidemiology studies. J Natl Cancer Inst 96:434–442

Wang WYS, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nat Rev Genet 6:109–118

Xiong M, Zhao J, Boerwinkle E (2002) Generalized T^2 test for genome association studies. Am J Hum Genet 70:1257–1268