Brief Genetics Report

A Polymorphism in the TCF7 Gene, C883A, Is Associated With Type 1 Diabetes

Janelle A. Noble,[1] Amy M. White,[2] Laura C. Lazzeroni,[3] Ana M. Valdes,[2] Daniel B. Mirel,[2] Rebecca Reynolds,[2] Andrew Grupe,[4] Dee Aud,[5] Gary Peltz,[5] and Henry A. Erlich[2]

Type 1 diabetes is an autoimmune disease with a Th1 phenotype in which insulin-producing β-cells in the pancreas are destroyed. The T-cell–specific transcription factor TCF7 activates genes involved in immune regulation and is a candidate locus for genetic susceptibility to type 1 diabetes. A nonsynonymous single nucleotide polymorphism (SNP) (Pro to Thr) in the TCF7 gene, C883A, was examined in samples from 282 Caucasian multiplex type 1 diabetic families. HLA-DRB1 and -DQB1 genotypes were previously determined for these samples, allowing data stratification based on HLA-associated risk. The transmission disequilibrium test showed significant overtransmission of the A allele from fathers (64.1%, $P < 0.007$) and nonsignificant overtransmission (57.4%, $P < 0.06$) of the A allele to patients who do not carry the highest-risk HLA-DR3/DR4 genotype. Elliptical sib pair analysis showed significant associations of the A allele with type 1 diabetes in paternal transmissions ($P < 0.03$), transmissions to male children ($P < 0.04$), and in the non-DR3/DR4 group ($P < 0.04$). These data also suggest that TCF7 C883A may affect age of disease onset. Analysis of genotype data from surrounding SNPs suggests that this TCF7 polymorphism may itself represent a risk factor for type 1 diabetes. *Diabetes* 52:1579–1582, 2003

Type 1 diabetes is an autoimmune disease that involves destruction of insulin-producing β-cells of the pancreas. Genetic susceptibility to type 1 diabetes is multifactorial; the strongest contribution to genetic susceptibility comes from the HLA region of chromosome 6 (1). Whole-genome screens have been performed on large numbers of affected sib pairs in an attempt to identify other genes responsible for disease susceptibility (2,3). Results of these studies produced many linkage regions. The HLA region (IDDM1) and insulin promoter (IDDM2) were replicated between studies; however, some other candidate regions were not. These results suggest that effects of other loci are probably modest and may be confounded by heterogeneity of HLA-associated risk. An alternative approach, case/control genotyping with TDT analysis of polymorphisms in candidate genetic loci, has produced several reported associations with type 1 diabetes.

T-cell–mediated immune responses are generally categorized into two groups, Th1 responses, important in cell-mediated immunity, and Th2 responses, important in humoral immunity. The two responses are characterized by the production of distinct sets of cytokines, whose expression is regulated in part by transcription factors. The TCF7 gene product is a transcription factor that affects expression of genes important to Th1 responses, such as the interleukin (IL)-12R β2 subunit (4) (A.G., D.A., L. Gong, M. Mulkins, P. Gater, C. Satjawatcharaphong, A. Ahene, P. Thana, S. Chang, R. Reynolds, G. Zangenberg, W. Cookson, M. Moffat, M. Guler, K. Murphy, G.P, unpublished observations). TCF7 (alternate name TCF-1, GenBank accession no. X63901) is a member of the HMG box transcription factor family, which also includes TCF-4, a locus implicated in some colorectal cancers (5). TCF7 expression is limited to T- and NK-cells (6). Knockout experiments in mice have shown that TCF7 is necessary for T-cell development in the thymus (7).

Because type 1 diabetes is a Th1-mediated disease, TCF7 might reasonably be expected to influence type 1 diabetes, either by affecting the T-cell repertoire in developing thymocytes or by affecting the Th1 versus Th2 status of the autoimmune response, or both. The TCF7 gene maps to the cytokine gene cluster on chromosome 5q31 and contains a nonsynonymous (Pro to Thr) coding region polymorphism, referred to in this report as C883A, in exon 2. Samples from 282 Caucasian multiplex type 1 diabetic families were genotyped for TCF7 C883A to test for association with type 1 diabetes.

RESEARCH DESIGN AND METHODS

Subjects. Data were collected from genomic DNA from members of 282 families in the Human Biological Data Interchange repository (Philadelphia, PA). All families had at least two children affected by type 1 diabetes. A total of 21 families had three affected children, 1 family had four affected children, and 1 family had five affected children. One father was affected, all other parents were unaffected. The average age of onset in the affected children was 12.06 (SD 8.53), with a range of 6 months to 36 years.

From the [1]Children's Hospital Oakland Research Institute, Oakland, California; [2]Roche Molecular Systems, Alameda, California; [3]Department of Health Research and Policy, Stanford University, Palo Alto, California; [4]Celera Diagnostics, Alameda, CA; and [5]Roche Bioscience, Palo Alto, California.

Address correspondence and reprint requests to Dr. Janelle Noble, Children's Hospital Oakland Research Institute, 5700 Martin Luther King, Jr. Way, Oakland, CA 94609. E-mail: jnoble@chori.org.

Received for publication 24 October 2002 and accepted in revised form 10 March 2003.

A.G. was employed by Roche Bioscience during the course of this study.

ESP, elliptical sib pair; IBD, identity by descent; IL, interleukin; LD, linkage disequilibrium; SNP, single nucleotide polymorphism; TDT, transmission disequilibrium test.

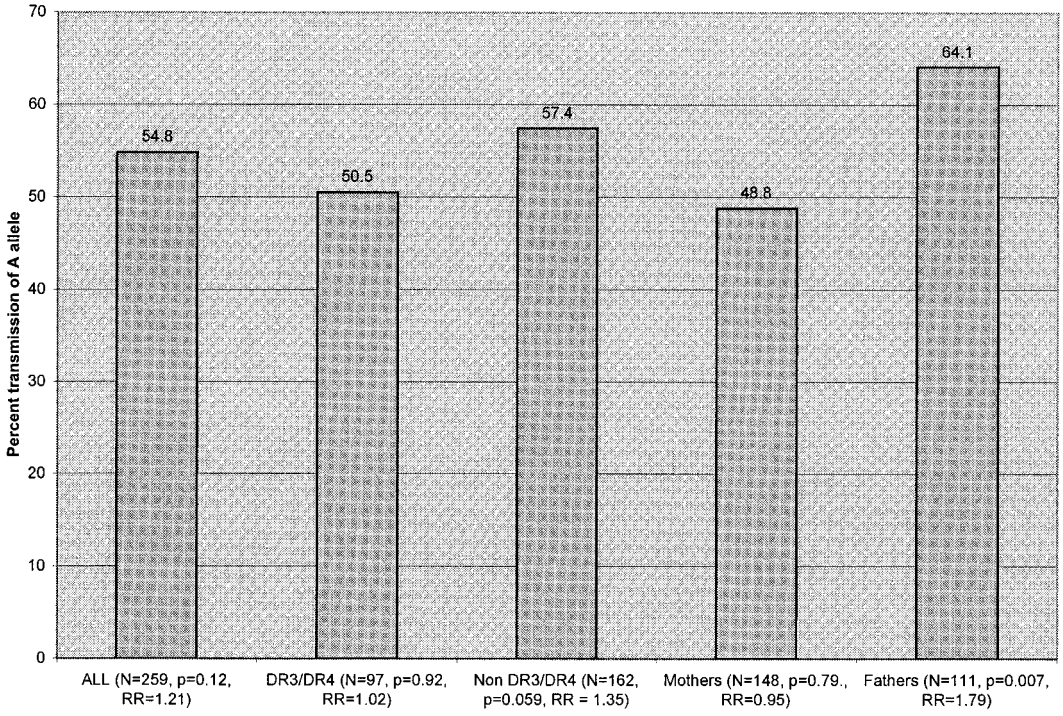

FIG. 1. Results of TDT analysis of the TCF7 C883 genotype data. Bars represent the percent transmission of the A allele in different groups of patients. Significance is indicated by uncorrected *P* values. Only the value for fathers (*P* = 0.007) remains significant after Bonferroni correction (*P* = 0.035). Relative risk (RR) of A to C was estimated as percent transmission of A divided by percent transmission of C.

TCF7 genotyping. The TCF7 single nucleotide polymorphism (SNP) at position 883 in GenBank accession sequence X63901 was genotyped by two methods. The gel-based method utilized a four-primer PCR with two opposing allele-specific primers flanked by two general primers. The general primers, LS045 (TCCAGGTCCTTCCCCTAAAA) and GZ348 (GCGGGGTCCACTTAC CA), generated a 315-bp PCR product and served as an internal control. Each allele-specific primer, GZ351 (CATGCATTACCCACCCC) for the C allele and GZ374 (CCTGCTCCCGAGGT) for the A allele, could pair with one of the general primers to generate an allele-specific product. The TCF7 C883A C allele–specific product (LS045-GZ351) was 275-bp long, and the A allele–specific product (GZ374-GZ348) was 75 bp. PCR amplifications were carried out in 2.4 mmol/l MgCl2, 60 mmol/l KCl, 12 mmol/l Tris-HCl (pH 8.3), 0.2 mmol/l each dATP, dCTP, dGTP, and dTTP, and 5 units/reaction AmpliTaq GOLD (Perkin-Elmer) in a total volume of 50 μl per reaction. Product size was analyzed by electrophoresis on 3% NuSieve (FMC, Rockland, ME) 1% Agarose. Heterozygous samples produced all three bands; homozygous samples produced the general band plus one or the other allele-specific band.

The second, more robust, genotyping method, an immobilized probe/linear array assay, included probes specific for the TCF7 C883A SNP and nine surrounding SNPs, including IL-4 C582T, IL-13 C4045T, IL-13 G4166A, CSF2 T2600C, IL-9 C4244T, ADRB2 A1633G, ADRB2 C1666G, CD14 C2232T, and a second TCF7 SNP, A383T, that sits at position 383 of accession sequence AF163776.1. Relative to the TCF7 transcription start site, the A383T SNP sits at position −1,479, and the C883A SNP sits at position 368. For each SNP, two probes (one per allele) were immobilized onto a sheet of backed nylon membrane in an array of parallel lines. Membrane sheets were cut crosswise into strips and then hybridized to biotinylated PCR amplification products from individual samples. Biotinylated primer pairs were used in a single-tube, multiplex PCR to generate a pool of labeled PCR products. Hybridization of biotinylated PCR products to target probes on the array was detected by incubation of the strips with streptavidin–horseradish peroxidase followed by a TMB (tetramethylbenzidine) solution. Similar assays have been previously described (8,9).

Statistical methods. The transmission disequilibrium test (TDT) was performed by the method of Spielman and Ewens (10).

Elliptical sib pair test for association. During this study, we developed the elliptical sib pair (ESP) test (11). The ESP test is similar to the TDT in that it relies only on the transmission of alleles from heterozygous parents to affected children. It can be shown that whenever one allele is transmitted more frequently than the other, identity by descent (IBD) in sibship data must exceed 50%. The ESP association test is a modified, more powerful version of the TDT that exploits some of the IBD information in sibship data. The ESP test uses exact calculations to choose the optimal form of the test given the number of heterozygous parents to calculate *P* values.

Association with age of onset. Patients were stratified into two groups: those with either an HLA-DR3 or -DR4 haplotype (including only DRB1*0401, DRB1*0402, DRB1*0404, and DRB1*0405 haplotypes carrying DQB1*0302 or DQB1*02) and those with neither (DRX/DRX). Within strata, *t* tests were utilized to compare average age of onset for individuals with at least one copy of the TCF7 C883A A allele (A+) with that of individuals homozygous for the C allele (CC). For the ESP analysis, patients with age of onset ≤10 years (the median) were considered "early onset." Age of onset was unavailable for 14 patients; 6 of these had a parent heterozygous at the C883A SNP. These six were grouped with the late-onset category.

RESULTS

TCF7 SNP screening. Observed Caucasian allele frequencies in an initial sample of 177 subjects are 0.88 for the C allele and 0.12 for the A allele (A.G., D.A., L. Gong, M. Mulkins, P. Gater, C. Satjawatcharaphong, A. Ahene, P. Thana, S. Chang, R. Reynolds, G. Zangenberg, W. Cookson, M. Moffat, M. Guler, K. Murphy, G.P., unpublished observations). All parent samples, and all children of parents with at least one A allele, were genotyped by the gel-based method. Data were confirmed, and 0.9% of called genotypes were corrected, by genotyping all samples with the linear array assay. The 282 families had the following distribution based on parental genotypes: 167 CC/CC, 8 AA/CC, 93 AC/CC, and 14 AC/AC. Parent frequencies of the two alleles matched previously observed frequencies very closely: (f) C = 0.879, (f) A = 0.121. The distribution of parental genotypes was in Hardy-Weinberg equilbrium (not shown). Frequency of the A allele for the affected children [n = 590, f(A) = 0.134] was only slightly higher than expected.

TDT. Transmission of TCF7 alleles from heterozygous parents to affected children was assessed for significant deviation from the expected value of 50% for each allele. Of 1,180 total transmissions, 259 (22%) were informative, i.e., from a heterozygous parent. TDT results (Fig. 1) show an overall nonsignificant excess of transmission of the A allele. Examination of parent of origin indicates that the excess is entirely due to a significant excess of paternal

TABLE 1
Observed associations of TCF7 A allele with T1D: ESP test results.

Group	*P* value for A allele association	Number of transmissions
All	0.098	259
Stratified by single criterion		
DR3/DR4	0.702	97
Non-DR3/DR4	0.040	162
From mothers	0.244	148
From fathers	0.029	111
To girls	0.784	103
To boys	0.043	156
To early onset	0.090	140
To late onset	0.650	113
Stratified by multiple criteria		
From fathers to DR3/DR4	0.588	48
From fathers to non-DR3/DR4	0.039	63
From fathers to early-onset non-DR3/DR4	0.010	38
From fathers to late-onset non-DR3/DR4	0.842	25

transmissions of the A allele ($P = 0.007$). Maternal transmissions of A and C alleles were nearly equivalent.

The HLA region is the major contributor to type 1 diabetes genetic susceptibility. To detect a potentially modest contribution of TCF7 to type 1 diabetes susceptibility, patients with high-risk HLA-associated genotype, defined as HLA-DR3/DR4, and present in ~40% of patients in this dataset, were analyzed separately from lower-risk patients. Those DR3/DR4 patients whose DR4 haplotype contained either DRB1*0403 or DQB1*0301, both considered protective alleles, were grouped with the non-DR3/DR4 category. The non-DR3/DR4 group showed overtransmission, although not statistically significant, of the A allele (57.4%, $P = 0.059$).

ESP analysis. Because the families in the dataset were multiplex, the genotyping data generated information on not only transmission of alleles but also allele sharing between siblings. The ESP test was developed to increase the sensitivity of the TDT by incorporating information on IBD allele sharing. Table 1 shows the results of ESP analysis of the TCF7 SNP genotyping data. Data were stratified into the same categories as for the TDT analysis above, and the additional categories of age of onset and child sex were examined. As with the TDT, the overall data suggest overtransmission of the A allele with type 1 diabetes ($P = 0.098$) and show significant overtransmission of the A allele from fathers ($P = 0.029$). Overtransmissions of the A allele to male children ($P = 0.043$) and to non-DR3/DR4 children ($P = 0.040$) also appear to be significant. Analysis of age of type 1 diabetes onset shows a nonsignificant increased transmission of the A allele to children with early age of onset, at or before age 10 years ($P = 0.090$). Combinations of these categories tend to increase statistical significance despite the reduction in sample size. Examples include transmission from fathers to early-onset boys ($P = 0.0156$) or transmission from fathers to non-DR3/DR4 early-onset children ($P = 0.0095$). The concordance among the results of the TDT and ESP test support the idea that the observed effect reflects association of TCF7 C883A with type 1 diabetes rather than linkage disequilibrium (LD) of TCF7 C883A with an alternate predisposing locus.

TABLE 2
Age of onset among A^+ (Thr^+) and CC (Pro-Pro) type 1 diabetic patients stratified by DR-DQ genotype

$DR3^+$ or $DR4^+$ genotypes			DRX/DRX genotypes		
TCF7 genotype	Age of onset (years)	*P*	TCF7 genotype	Age of onset (years)	*P*
A+	11.23 ± 0.71		A+	10.50 ± 1.42	
CC	11.46 ± 0.40		CC	15.03 ± 1.52	
		0.79(NS)*			<0.036*

Data are means ± SE. *P values were derived by t test. NS, not significant.

Age of onset. The hypothesis that TCF7 is associated with age of onset was further tested by calculating the average age of onset in patient groups stratified by HLA. HLA haplotypes were defined as DR3, DR4 (see RESEARCH DESIGN AND METHODS), and DRX (DRX ≠ DR3 or DR4). TCF7 status was defined as the presence of at least one C883A A allele (Thr+), which includes heterozygous (AC) and homozygous (AA) patients, or the absence of any A allele (Pro-Pro), which includes only homozygotes (CC). In low HLA-associated risk groups, age of onset tended to be lower for TCF7 Thr+ patients (not shown) than for TCF7 Pro-Pro patients. Table 2 shows that the lowest HLA-associated risk group, DRX/DRX, has significantly lower average age of onset than does the group with higher HLA-associated risk ($P < 0.05$).

Surrounding SNPs. TCF7 is located at 5q31, a region rich in immunologically relevant genes, including IL-4 and IL-13. To examine whether the TCF7 C883A SNP creates a functional change related to type 1 diabetes susceptibility, or is simply a marker for another disease-related SNP in linkage LD, nine surrounding SNPs, spanning 20.43 Mb, were genotyped and analyzed for LD with TCF7 C883A. Only TCF7 A383T was in significant LD with TCF7C883A (Table 3). TDT analysis shows that the SNP at position 383 does not have a significant association with type 1 diabetes (overall, $P = 0.558$; DR3/DR4, $P = 0.106$; non-DR3/DR4, $P = 0.667$). Of the 818 meioses that are informative for either or both SNPs, 332 are in parents carrying the (383–883) A-C and T-C haplotypes, 207 are in parents carrying the A-C and A-A haplotypes, and 89 are in parents carrying the A-A and T-C haplotypes. The T-A haplotype is absent. In 79% of meioses that are informative for position 383, the parent is an uninformative C/C homozygote at position 883. When the three observed haplotypes are analyzed as three "alleles" by TDT, data for the non-DR3/DR4 (low HLA risk) patients show that A-A is significantly overtransmitted (60.4%, $P = 0.012$). In this analysis, A-C appears significantly undertransmitted (43.7%, $P = 0.048$), probably because most of the parents who overtransmit A-A have A-C on the other chromosome. These results show that the distortion in transmission at position 883 is not due to the SNP at position 383. In addition, our TCF7 C883A data show more frequent transmission of the A allele to affected children, even among the siblings of type 1 diabetic children who received the C allele (not shown).

TABLE 3
LD analysis of SNPs surrounding TCF7 C883A

Locus	SNP	Description	Physical	Distance to TCF7 C883A (Mb)	D′ with TCF7 C883A	*P*
5 cen			46.05	97.98		
CSF2	T2600C	I117T	139.19	4.84	−0.09	0.64
IL13	C4045T	(intron 3)	141.35	2.68	−0.05	0.85
IL13	G4166A	R110Q	141.37	2.66	−0.07	0.77
IL4	C582T	(−589)	141.37	2.66	−0.13	0.53
TCF7	C883A	P19T	144.03	0.00	0.00	—
TCF7	A383T	(−1489)	144.03	0.002	−1.00	<0.0001
IL9	C4244T	T113M	146.04	2.01	0.05	0.23
CD14	C2232T	(−260)	154.62	10.59	−0.04	0.74
ADRB2	A1633G	R16G	159.62	15.59	−0.12	0.12
ADRB2	C1666G	Q27E	159.62	15.59	−0.15	0.39
5q tel			191.00	46.97		

This suggests that having the A allele is of greater importance for development of type 1 diabetes than IBD with an affected sib, which favors a functional role for TCF7 C883A. Taken together, these data support the hypothesis that TCF7 C883A may be a functional polymorphism.

DISCUSSION

These data implicate TCF7 as one locus involved in genetic susceptibility to type 1 diabetes. Prior characterization of HLA, the largest contributor to type 1 diabetes genetic risk, allowed stratification of the patients into high- and low-risk HLA groups. As expected, the modest effects of TCF7 on both disease susceptibility and age of onset were more apparent in patients whose HLA-associated risk was low. Association in paternal, and not maternal, transmissions to patients could result from imprinting of the maternal gene. The apparent effect of the TCF7 A allele with early age of disease onset suggests that the product of this allele may lead to earlier initiation or to acceleration of the autoimmune process that precedes clinically apparent disease, perhaps by causing increased expression of Th1-associated cytokines. The specifity of the effect for male patients could result from an interaction of TCF7 with an X chromosome locus such as that proposed to explain the increase in DR3 haplotypes in male compared with female patients (12).

Whatever the biological mechanism, these data show that TCF7 may be another in the expanding set of genetic loci that influence genetic susceptibility to type 1 diabetes and represent another step toward increasing the predictive power of genetics for type 1 diabetes.

ACKNOWLEDGMENTS

This work was supported, in part, by an American Diabetes Association Career Development Award to J.A.N, National Institutes of Health (NIH) grants 5R01GM062628 and 5R01GM053275 to L.C.L., and NIH grants DK46626 and AI29049 to H.A.E.

The authors thank the DNA synthesis group at Roche Molecular Systems for synthesis of oligonucleotide primers and probes used for these experiments, as well as Gabriele Beer, Karen Walker, and Lori Steiner for advice and assistance.

REFERENCES

1. Noble JA, Valdes AM, Cook M, Klitz W, Thomson G, Erlich HA: The role of HLA class II genes in insulin-dependent diabetes mellitus: molecular analysis of 180 Caucasian, multiplex families. *Am J Hum Genet* 59:1134–1148, 1996
2. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SC, Jenkins SC, Palmer SM, Balfour KM, Rowe BR, Farrall M, Barnett AH, Bain SC, Todd JA: A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371:130–136, 1994
3. Concannon P, Gogolin-Ewens KJ, Hinds DA, Wapelhorst B, Morrison VA, Stirling B, Mitra M, Farmer J, Williams SR, Cox NJ, Bell GI, Risch N, Spielman RS: A second-generation screen of the human genome for susceptibility to insulin-dependent diabetes mellitus [see comments]. *Nat Genet* 19:292–296, 1998
4. Guler ML, Gorham JD, Dietrich WF, Murphy TL, Steen RG, Parvin CA, Fenoglio D, Grupe A, Peltz G, Murphy KM: Tpm1, a locus controlling IL-12 responsiveness, acts by a cell-autonomous mechanism. *J Immunol* 162: 1339–1347, 1999
5. Roose J, Huls G, van Beest M, Moerer P, van der Horn K, Goldschmeding R, Logtenberg T, Clevers H: Synergy between tumor suppressor APC and the beta-catenin-Tcf4 target Tcf1. *Science* 285:1923–1926, 1999
6. van de Wetering M, Oosterwegel M, Holstege F, Dooyes D, Suijkerbuijk R, Geurts van Kessel A, Clevers H: The human T cell transcription factor-1 gene: structure, localization, and promoter characterization. *J Biol Chem* 267:8530–8536, 1992
7. Verbeek S, Izon D, Hofhuis F, Robanus-Maandag E, te Riele H, van de Wetering M, Oosterwegel M, Wilson A, MacDonald HR, Clevers H: An HMG-box-containing T-cell factor required for thymocyte differentiation. *Nature* 374:70–74, 1995
8. Cheng S, Grow MA, Pallaud C, Klitz W, Erlich HA, Visvikis S, Chen JJ, Pullinger CR, Malloy MJ, Siest G, Kane JP: A multilocus genotyping assay for candidate markers of cardiovascular disease risk. *Genome Res* 9:936–949, 1999
9. Bugawan TL, Apple R, Erlich HA: A method for typing polymorphism at the HLA-A locus using PCR amplification and immobilized oligonucleotide probes. *Tissue Antigens* 44:137–147, 1994
10. Spielman RS, Ewens WJ: The TDT and other family-based tests for linkage disequilibrium and association. *Am J Hum Genet* 59:983–989, 1996
11. Lazzeroni LC: Allele sharing and allelic association. I. Sib pair tests with increased power. *Genet Epidemiol* 22:328–344, 2002
12. Cucca F, Goy JV, Kawaguchi Y, Esposito L, Merriman ME, Wilson AJ, Cordell HJ, Bain SC, Todd JA: A male-female bias in type 1 diabetes and linkage to chromosome Xp in MHC HLA-DR3-positive patients. *Nat Genet* 19:301–302, 1998