ARTICLE

Genetic Mapping at 3-Kilobase Resolution Reveals Inositol 1,4,5-Triphosphate Receptor 3 as a Risk Factor for Type 1 Diabetes in Sweden

Jared C. Roach, Kerry Deutsch, Sarah Li, Andrew F. Siegel, Lynn M. Bekris, Derek C. Einhaus, Colleen M. Sheridan, Gustavo Glusman, Leroy Hood, Åke Lernmark, and Marta Janer, on behalf of the Swedish Childhood Diabetes Study Group and the Diabetes Incidence in Sweden Study Group

We mapped the genetic influences for type 1 diabetes (T1D), using 2,360 single-nucleotide polymorphism (SNP) markers in the 4.4-Mb human major histocompatibility complex (MHC) locus and the adjacent 493 kb centromeric to the MHC, initially in a survey of 363 Swedish T1D cases and controls. We confirmed prior studies showing association with T1D in the MHC, most significantly near HLA-DR/DQ. In the region centromeric to the MHC, we identified a peak of association within the inositol 1,4,5-triphosphate receptor 3 gene (*ITPR3*; formerly *IP3R3*). The most significant single SNP in this region was at the center of the *ITPR3* peak of association ($P = 1.7 \times 10^{-4}$ for the survey study). For validation, we typed an additional 761 Swedish individuals. The *P* value for association computed from all 1,124 individuals was 1.30×10^{-6} (recessive odds ratio 2.5; 95% confidence interval [CI] 1.7–3.9). The estimated population-attributable risk of 21.6% (95% CI 10.0%–31.0%) suggests that variation within *ITPR3* reflects an important contribution to T1D in Sweden. Two-locus regression analysis supports an influence of *ITPR3* variation on T1D that is distinct from that of any MHC class II gene.

Type 1 diabetes (T1D [MIM 222100]) is the most common chronic disease in children and affects ~1% of the population in high-incidence countries, such as those of Scandinavia. The incidence of T1D is increasing worldwide by 2%–5% per year.[1] The disease is due to an autoimmune destruction of pancreatic islet β cells and is associated with cellular and humoral immunological abnormalities.[2] Patients become insulin dependent for life; often develop late complications that include angiopathy, nephropathy, retinopathy, and neuropathy; and have a shortened life expectancy. The genetic etiology of T1D is strongly associated and genetically linked with the HLA-DR/DQ locus.[3] Numerous investigations have identified the contribution of major histocompatibility complex (MHC) genes. These have included both association and linkage studies.[4–6] The relative risk of siblings with associated allele sharing at the HLA-DR/DQ locus (also known as "*IDDM1*") has been estimated to be 2.5–3.4.[7] As a result of the association of the MHC locus with T1D, Concannon et al.[3] called for MHC studies employing SNPs at a high density.

In addition to the MHC class II genes, other genes in or near the MHC locus have long been suspected to contribute to T1D risk.[8] Four lines of evidence have fueled these suspicions. First, genes currently known to be associated with T1D risk account for only a portion of the genetic risk, implying that more genes remain to be identified. Second, despite extensive genomewide scans for risk, the only span consistently identified with a high LOD score is centered in the MHC locus.[9,10] This singular high but broad peak of association has led to the suspicion that at least one additional risk gene, in addition to the MHC class II genes, lies within this region. Third, higher-resolution linkage analysis in some MHC studies has shown multiple local maxima, suggesting that multiple genes might contribute.[11] Fourth, dissection of the diabetes phenotype of the nonobese diabetic mouse has demonstrated distinct and separable genetic loci contributing to diabetes risk. In particular, Deyrutter et al.[12] used congenic mice to isolate and refine the localization of *Idd16* to a 3.1-Mb region. This locus predisposes to diabetes in the mouse distinctly from the MHC class II association. In humans, the region syntenic to *Idd16* lies immediately centromeric to the MHC locus. This region was the focus of our current study.

In previous studies, the extended linkage disequilibrium (LD) in the MHC has made it difficult to prove that other associations in or near the MHC are independent and not explained by LD with HLA-DR/DQ. There are two reasons for this limit on resolution: (1) markers were spaced too far apart to resolve separate peaks of association, and (2) LD broadens and alters peaks of association.[13] In our study, we have overcome the first limitation but not the second. Therefore, we still are unable to resolve separate etiologic influences within the MHC, if indeed there is more than

From the Institute for Systems Biology (J.C.R.; K.D.; S.L.; D.C.E.; C.M.S.; G.G.; L.H.; M.J.) and Departments of Statistics (A.F.S.), Management Science and Finance and Business Economics (A.F.S.), and Medicine (L.M.B.; Å.L.), University of Washington, Seattle

Received June 2, 2006; accepted for publication July 18, 2006; electronically published August 30, 2006.

Address for correspondence and reprints: Dr. Jared C. Roach, Institute for Systems Biology, 1441 North 34th Street, Seattle, WA 98103. E-mail: jroach@systemsbiology.org

Am. J. Hum. Genet. 2006;79:614–627. 0002-9297/2006/7904-0004$15.00

one. However, we can now resolve peaks of association that lie near the MHC, as long as they lie outside the span of strong LD of the MHC locus, and we have identified one such peak in the inositol 1,4,5-triphosphate receptor 3 gene (*ITPR3* [MIM 147267]).

This article first reports the results of our survey study of 363 individuals typed for 2,260 markers covering both the MHC locus (4,419,819 bp; 2,114 SNPs) and the region immediately centromeric (493,457 bp; 146 SNPs). The survey study unexpectedly identified a novel association in the centromeric region within the *ITPR3* gene, so our exposition next presents analyses of the statistical significance of this association based only on the survey data. After considering these data, our team made the decision to test this association in a verification study of an additional 761 individuals. To increase the power of the verification study, it was performed only on markers within the initially identified *ITPR3* peak of association. Our exposition concludes with an analysis of the data from our verification study and the conclusion that *ITPR3* is associated with T1D.

Table 1. Overview of Samples, Originally Collected in DIS or SCDS

Study and Measure	Cases	Controls
Survey:		
No. of individuals:	181	182
Females	79	92
Males	102	90
Age at diagnosis[a] (years):		
Range	1.83–34.58	1.25–35.33
Mean	16.5	17.5
Median	15.2	14.9
No. of DIS individuals	92	90
No. of SCDS individuals	89	92
Survey and validation combined:		
No. of individuals:	651	483
Females	268	236
Males	383	247
Age at diagnosis[a] (years):		
Range	1.25–34.58	1.25–35.33
Mean	15.9	16.0
Median	13.8	13.6
No. of DIS individuals	281	179
No. of SCDS individuals	370	304

[a] Age at diagnosis for controls refers to age at diagnosis of the case patient to whom the control was birth matched.

Material and Methods

Study Selection

DNA samples for both our survey and verification studies are a subset of the samples collected and analyzed previously for the Diabetes Incidence in Sweden Study (DIS) and the Swedish Childhood Diabetes Study (SCDS).[4,14,15] DIS was a case-control study of diabetes of all types in 15–34-year-olds from January 1987 to December 1988. DIS included 474 patients with T1D and 279 controls. We did not select any patients without T1D from DIS for analysis in the present study (i.e., we did not include those with type 2 diabetes [T2D {MIM 125853}]). SCDS was a matched case-control study of T1D in 0–14-year-olds from September 1986 to December 1987 and included 498 subjects with T1D and 423 controls. The majority (88%) of the controls in SCDS were matched by date of birth, region of birth, and sex. Further details for SCDS and DIS are as described by Graham et al.[16] The homogeneity of the Swedish population makes it unlikely that population stratification or admixture confounded our study. For our study, only samples over a threshold of available micrograms of DNA were selected for analysis, to maximally preserve the DNA banked resource. Because each original matched control sample was not always available, we selected our controls from those available, with the primary criterion of matching them by age to the cases. As a result, we introduced a slight sex differential between our cases and controls. For our survey study, 368 samples were selected for typing with the Illumina MHC SNP set. Typing failed for 5 of our 368 samples, so data analysis was performed on the remaining 363 samples (table 1). We do not suspect any bias influencing which samples failed. For our verification study, we typed 771 distinct additional samples for 40 SNPs. Of these, 761 were successfully typed for SNP *rs2296336,* which is representative of the success rate for our Sequenom verification genotyping of these samples. During verification, we also retyped most of the samples previously typed with Illumina technology and were thus able to cross-validate the results of the two technologies. The union of the sets of samples we typed with the Sequenom technology or the Illumina technology totals 1,134 samples. Samples that failed typing by Illumina but were successfully typed by Sequenom technology were included in the set of 771 distinct additional samples. The number of failures of Illumina typing at each locus was typically zero to three samples but, in some instances, was as high as eight samples.

SNP Typing

The MHC Panel Set implemented by Illumina was used to type all the samples in our survey study (but not in the validation study). The Illumina call-rate statistics for each locus typed are available at T1DBase—ITPR3. Development of SNPs selected for this panel leveraged both the MAGMA software for SNP selection and algorithms for maximizing linkage information.[17–19] The panel is described in the MHC Panel Set Datasheet (available at the Illumina: MHC Mapping Panel Web site) and includes (1) 1,293 SNPs evenly spaced across the MHC region, (2) 1,228 SNPs in or near coding regions, and (3) additional SNPs designed to tag haplotype blocks (tSNPs). The sequences and mapped locations of these SNPs are available at T1DBase—ITPR3. The distribution of SNPs includes ~30% that are >10 kb from known coding regions, 20% in introns, and ~40% <10 kb from coding regions; the remaining SNPs are in transcribed regions, including both coding regions and UTRs. SNPs were typed with bead-based assays, with each bead affixed to the end of a glass fiber, to enable optical readout of a proprietary extension/ligation reaction (details available at the Illumina: Technology and Applications Web site). The total region covered by this set of SNPs was selected by Illumina to have broad applicability for a number of purposes but was designed in part to implement fine-scale association mapping of the MHC, with specific consideration of the T1D candidate region in the MHC. The set of SNPs covers a span of the MHC from the gene *RFP* (MIM 602165) to the gene *KNSL2* (MIM 603763), which has been termed "the extended MHC locus," and includes additional SNPs centromeric to the extended MHC locus, for a total of 2,360 SNPs. The total of 2,360 was determined in

part by the quanta imposed by the implementation of Illumina technology. In particular, the centromeric bound on the region covered by our SNPs lies just beyond the *ITPR3* gene. There are at least two limitations of the Illumina MHC SNP Set. First, despite the generally uniform high tiling density, there are some large spacings between SNPs. Second, some particularly important spacings are fairly long. For example, the spacing containing HLA-DQB1 is 8,381 bp. We did not extend typing to more-centromeric SNPs because that was the limit of the Illumina MHC SNP Set and because we did not expect to find an association beyond this limit.

We searched for evidence of assay-specific technical failures in our Illumina survey data by computing *P* values for Hardy-Weinberg equilibrium in our control population of 182 Swedish individuals. The distribution of *P* values showed no outliers; in particular, of the 2,360 *P* values, 52 were <.05 and 12 were <.01.

In determining which SNPs to display in figure 1*C*, we required that each of the three possible genotypes be seen at least 10 times in our survey data set of 363 individuals. This filter protects against the possibility of error introduced by technical issues specific to the typing of that particular SNP (e.g., high optical noise) but might exclude some informative SNPs with low minor-allele frequencies (MAFs). However, these excluded SNPs have low power because of their low MAF. As a result of this filter, 1,696 SNPs are plotted in figure 1*C*.

For our verification study, typing was performed for 40 SNPs with the MassArray genotyping system (Sequenom), which is centered on a chip-based matrix-assisted laser desorption/ionization time-of-flight (MALDI-TOF) mass spectrometer. The SNPs were selected to span the 180,552-bp interval containing the *ITPR3* gene. Multiplex SNP assays were designed using SpectroDESIGNER 3.0. The 40 SNPs were genotyped in four sets. Two sets were typed using IPLEX chemistry, and two sets were typed using Homogeneous MassEXTEND (HME) chemistry. In both cases, 384-well plates containing 5–10 ng of DNA in each well were PCR amplified per Sequenom's specifications. After PCR, Arctic shrimp alkaline phosphatase was added to samples to prevent further incorporation and interference with the primer-extension assay. Allele-discrimination reactions with the use of HME chemistry were conducted by addition of extension primers, DNA polymerase, and a cocktail mixture of deoxynucleotide triphosphates (dNTPs) and dideoxynucleotide triphosphates (ddNTPs) to each well.[20] For IPLEX reactions, the dNTP-ddNTP mixture was substituted with a mixture of modified ddNTPs.[21] After the primer-extension step, MassEXTEND clean resin was added to the mixture to remove extraneous salts that could interfere with MALDI-TOF analysis. Genotypes were determined by spotting 15 nl of each sample onto a 384 SpectroCHIP, which was subsequently read by the MALDI-TOF mass spectrometer. Of the 40 attempted SNPs, 28 were successful in that they both (1) had a technically successful assay, as indicated by distinct clustering of the signals for the three different genotypes, and (2) had a nonzero MAF in our samples. Of these 28 SNPs, 18 were present in the Illumina MHC SNP set employed in the survey study and 10 were not (table 2). SNPs *rs6912568, rs2296333, rs9461899, rs9469559,* and *rs2229641* were invariant in our samples, and SNPs *rs11963294* and *rs2229633* had a low MAF—these SNPs are not shown in table 2.

SNP Mapping

All SNPs were selected from or have been assigned identifiers in dbSNP. SNPs were localized to all known genomic sequences derived from human MHC loci (listed at T1DBase—ITPR3). These genomic sequences included: (1) the published sequences of two haplotypes of the MHC[22]; (2) all clones sequenced by the Sanger Centre as part of the MHC haplotype project, in accordance with the assignments provided by the Sanger Centre to one of eight haplotypes: *PGF, COX, APD, DBB, MANN, SSTO, QBL,* and *MCF*; (3) all genomic sequences >10 kb with significant BLAST scores matching the Sanger MHC assembly; (4) the two alternative versions of the MHC region in the University of California–Santa Cruz (UCSC) May 2004 human assembly (chr6_hla_hap1 and chr6_hla_hap2) (MGC/ORFenomes Genome Browser); and (5) the UCSC–National Center for Biotechnology Information (NCBI) consensus sequence. The complete sequences of *PGF* and *COX* are reported by Stewart et al.[22] SNPs were also mapped to known variations in IMGT/HLA. The use of multiple reference sequences allowed SNPs to be assigned to different positions in different haplotypes, when haplotype information was known. Some SNPs may not be present in some haplotypes (e.g., the contextual sequence is missing or altered in such a way that the assay, depending on implementation, always provides a negative or ambiguous result for that haplotype).

The SNP *rs701831* maps to locations 3625544, 3687731, and 3660864, with respect to the origin shown in figure 1. This reflects the repeat structure of the region. Because these positions are fairly close, we portray this SNP only once, at position 3660864, to avoid creating a visual illusion of more statistical significance than our data justify. The SNP *rs3131635* maps to two locations: 2614465 and 2519595. The score of *rs3131635* correlates much better with its neighbors at position 2614465 than at 2519595, so we reject the positioning at 2519595 on the basis of this evidence and portray only position 261446 in figure 1. The SNP *rs3130914* maps to only one location on the canonical reference sequence but maps to two locations on most haplotypes, including the Sanger MHC reference sequence (The MHC Haplotype Project Web site), *CX, DB, MC,* and *SS*. If one were to arbitrarily assign it to the single location it exactly matches on the canonical reference sequence, it would represent a local maximum and possibly alter the interpretation of results. However, assigning it to its other location places it below an inferred dynamic baseline, and thus it appears to be a less interesting marker. In figure 1, we plot this SNP at this less interesting location, just 5′ of *MICB* (MIM 602436). Thus, an understanding of the sequence variations that influence the apparent genetic position of SNPs is important for the interpretation of this SNP and will become increasingly important as more haplotypic sequence information becomes known for the MHC locus.

Statistical Analysis

Hardy-Weinberg equilibrium and LD statistics were computed with HWE and LD (Cran—Package genetics), packages by Warnes and Leisch implemented in The R Project for Statistical Computing. Single-point association scores were calculated from 3 × 2 genotype contingency tables by use of Fisher's test, as implemented in R. Odds ratios (ORs) and 95% CIs were computed with the R package EpiTools (Aragón). Population-attributable risks (PARs) were computed as follows:

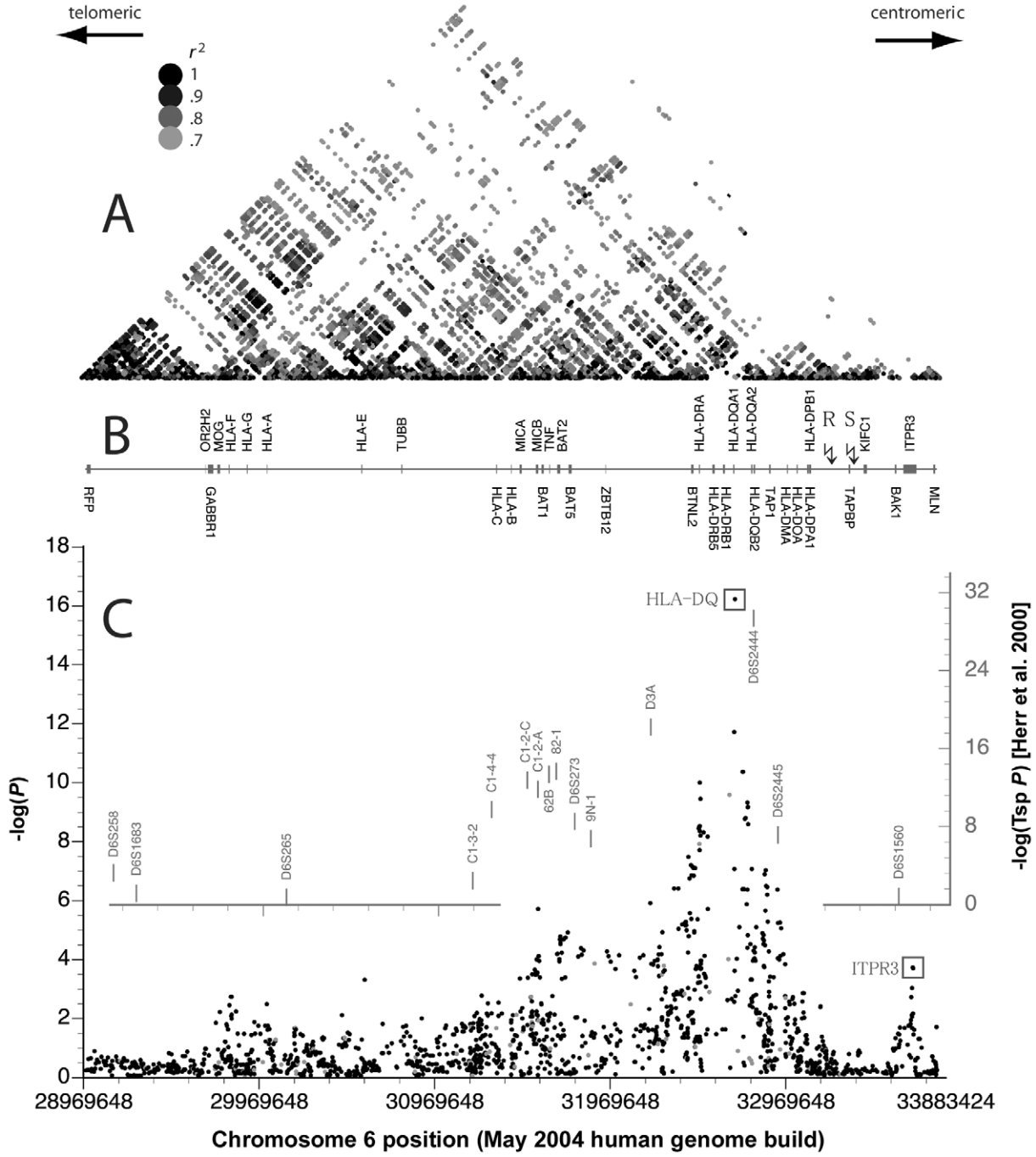

Figure 1. Association with T1D at *ITPR3*. *A,* Long-range LD is plotted as a function of r^2 between all pairs of SNPs (only $r^2 \geqslant 0.7$ is displayed); the resulting LD plot derived from our study population controls is similar to plots from other studies. The r^2 value between a pair of SNPs is plotted at the apex of an isosceles right triangle defined by the baseline connecting the SNPs. *B,* Locations of genes in and near the MHC, provided for reference. Not all genes are shown. R = recombination hotspot; S = syntenic breakpoint between mouse and other mammalian genomes. *C,* Single-point associations between T1D and SNPs in the Illumina MHC Panel. If at least one allele of a SNP exactly matches this canonical reference, it is blackened; otherwise, the SNP is shaded gray. The *Y*-axis is in units of $-\log_{10}(P)$. SNPs inside a box are selected to be illustrative of a peak of association. The Tsp linkage statistic is described by Herr et al.[11] Gray bars indicate microsatellites typed by Herr et al.[11] (the microsatellite *D6S1629* with Tsp $-\log(P) = .61$ is ~13 kb centromeric to the right margin of the fig.). The *X*-axis is the same for all graphs and is indexed to base pairs of the May 2004 human genome build.

Table 2. Verification of *ITPR3* SNP Genotypes

Locus	Typed in Survey Study	Covered by Sequencing	Chromosome 6 Position	*P* for All Data	*P* for Survey Data[a]	Cases AA	Cases AB	Cases BB	Controls AA	Controls AB	Controls BB
rs10947415	No	No	33699761	.237	ND	443	163	14	309	129	17
rs4711332	Yes	No	33703376	.912	.236	12	149	466	8	117	349
rs10947418	No	No	33712995	.364	ND	15	163	442	16	130	309
rs4713646	Yes	No	33714500	.00200	.0851	51	255	336	22	159	297
rs12205634	Yes	No	33719644	.0368	.0186	385	203	25	265	177	31
rs6457738	Yes	No	33725441	.505	.195	249	304	83	173	226	72
rs1570760	Yes	No	33730910	.0897	.0158	30	221	374	33	185	256
rs5875456	No	No	33732282	.106	ND	29	225	373	31	178	245
rs2296343	Yes	No	33734694	2.95×10^{-4}	.00182	264	265	93	241	193	37
rs2296337	Yes	No	33739673	1.98×10^{-4}	9.12×10^{-4}	92	267	281	34	194	248
rs1536036	Yes	No	33739991	.0149	.00891	195	285	148	110	237	127
rs6901411	No	No	33741504	6.24×10^{-4}	ND	266	261	91	234	183	36
rs9357163	No	Yes	33742614	.120	ND	456	142	16	308	128	13
rs12195707	No	Yes	33744024	.189	ND	16	149	453	20	118	318
rs9380372	No	Yes	33744210	2.12×10^{-4}	ND	4	147	455	14	142	298
rs2296336	Yes	Yes	33744638	1.30×10^{-6}	1.70×10^{-4}	107	296	240	35	217	229
rs2077163	Yes	Yes	33744884	5.30×10^{-4}	.00692	215	321	96	123	241	110
rs2229634	Yes	Yes	33746157	6.69×10^{-6}	1.82×10^{-4}	100	314	236	36	216	228
rs2274198	No	Yes	33746635	1.69×10^{-6}	ND	116	293	207	41	212	204
rs3736893	Yes	Yes	33747737	.00138	.0372	327	256	62	191	225	62
rs2274197	Yes	No	33753092	.125	.0257	28	226	373	33	178	260
rs2229638	Yes	No	33756205	.448	.603	472	117	6	377	76	5
rs12528378	No	No	33761463	.0182	ND	45	51	516	56	39	355
rs3818521	Yes	No	33765223	.123	.871	157	319	151	96	242	132
rs2229642	No	No	33767449	.112	ND	137	327	171	122	234	106
rs3818527	Yes	No	33769012	.181	.891	154	336	152	100	239	134
rs2281919	Yes	No	33771148	.220	.240	453	161	18	319	142	16
rs2281816	Yes	No	33880313	3.39×10^{-4}	.0204	269	285	83	231	213	29

NOTE.—These data confirmed the results of our survey study—that *ITPR3* is associated with T1D. An index of the allele sequences is provided at T1DBase—ITPR3.

[a] ND = no data.

Type	Cases	Controls
Risk	*a*	*b*
Protective	*c*	*d*

$$P_{\text{con}} \approx \frac{b}{b+d}$$

$$\text{OR}' = \frac{ad}{bc}$$

$$\text{PAR}\% \approx \frac{P_{\text{con}}(\text{OR}'-1)}{P_{\text{con}}(\text{OR}'-1)+1} \times 100\ ,$$

where P_{con} is the probability of a control having the risk genotype, OR′ is a formula for computing the OR, and PAR% is the PAR percentage.

ORs approximate genotype relative risks. For example, the OR for RR:PP (where R is the risk allele and P is the protective allele) approximates the genotype relative risk appropriate for a recessive model. The PAR we compute is for both the heterozygous and homozygous risk genotypes (e.g., *a* in the above equations is the sum of the number of cases that are homozygous for risk alleles and the number of cases that are heterozygous).

Multiple-Test Correction

No multiple-test correction has been applied to the *P* values of the 1,696 plotted SNP scores in figure 1*C*. To estimate the probability of type I error associated with identification of the peak of association in *ITPR3* in our survey study, we implemented a permutation multiple-test correction for the region centromeric to the MHC.[23] We defined a significant peak as a contiguous set of nine SNPs, of which seven had a single-point association $P < .01$ and three had $P < .001$. This reflects the characteristics of the peak we identified at *ITPR3,* using standard statistical *P* value cutoffs. We then estimated the probability of type I error by permuting the case-control labels on our survey data 1,000 times and enumerating the number of iterations in which such a peak was observed. No correction was applied to the validation study, since all 28 validation SNPs were selected because they lie within the peak of association encompassing *ITPR3* and are thus highly correlated and because an a priori decision was made for the primary analysis to be based on the typing of the single SNP at the peak of the survey association, *rs2296336*.

Sequencing

After evaluation by 1% agarose gel electrophoresis, templates were purified with Millipore filtration. Sequencing reactions were performed in the Institute for Systems Biology core sequencing facility by use of ABI 3730 capillary sequencers. Chromatograms were assembled and visualized with VectorNTI (Invitrogen).

Bioinformatics

Our convention for genomic coordinates for this article is the May 2004 human genome assembly. We interpreted our data in the context of reference information on T1D genetics found in the confederated database T1DBase.[24] Input files for Haploview that permit exploration of the characteristics of LD in our study population are available at T1DBase—ITPR3.[25] Files are also available at T1DBase—ITPR3 that provide coding for MHC class II alleles as biallelic markers, permitting Haploview visualization of the amount of LD between specific MHC class II alleles and *ITPR3* in our verification control population. We employed PHASE v2.1.1 to infer haplotypes.[26] Regression analysis was performed with LRASSOC 1.1.[27] The full model implemented in LRASSOC is of the form $\log[r/(1-r)] = \mu + a_1x_1 + d_1z_1 + a_2x_2 + d_2z_2 + i_{aa}x_1x_2 + i_{ad}x_1z_2 + i_{da}z_1x_2 + i_{dd}z_1z_2$. The probability r that a subject with the given genotype represents a case rather than a control is modeled from the dummy variables x and z that specify genotype. The parameters to be estimated consist of the mean term μ, terms to model additive effects a_1 and a_2 at each locus, terms to model dominance effects d_1 and d_2, and terms to model interactive effects i_{aa}, i_{ad}, i_{da}, and i_{dd}. In addition to the full model, LRASSOC fits eight models with fewer parameters: MEAN (μ), ADD1 (μ, a_1), ADD2 (μ, a_2), ADD (μ, a_1, a_2), DOM1 (μ, a_1, d_1), DOM2 (μ, a_2, d_2), DOM (μ, a_1, d_1, a_2, d_2), and ADDDOM (μ, a_1, d_1, a_2, d_2, i_{aa}, i_{ad}, i_{da}).

Resolution

We mapped SNPs to 2,402 genomic positions with 2,401 corresponding spacings. A few of our 2,360 SNPs mapped to more than one position. The mean spacing was 2,046 bp, the median was 1,347 bp, and the minimum was 5 bp. The size of the longest spacing was 59,292 bp, between *rs389512* and *rs6474*; this spacing contains *C4A* (MIM 120810) in tandem duplication. There was a total of 31 other spacings >10 kb, which ranged from 10,030 bp to 39,868 bp. The use of a reference haplotype other than the May 2004 assembly could alter some of the reported spacings. We refer to the resolution of our methodology as "3-kb." Given the actual value of the mean spacing, we may have been justified in claiming a 2-kb resolution. However, given that not all SNPs are equally powerful and that the mean spacing does not alone summarize the spacing distribution, we refer to our resolution more conservatively as a 3-kb resolution.

Because the selection of SNPs for the Illumina MHC SNP set included an additional focus on coding regions, the coding regions have an increased density of SNPs. In figure 1*C*, one visual effect of this higher density is an increase in the range of SNP *P* values for coding regions but not in the underlying distribution of *P* values. The opposite effect is true for highly polymorphic coding regions, where, despite increased scrutiny, no suitable universal primers could be found.[28]

Results

We applied a high-density SNP panel to a collection of carefully selected T1D cases and controls from DIS and SCDS, to test the hypothesis that genetic factors besides HLA-DR/DQ may contribute to the risk of T1D.[4,15] The tiling of our SNPs was designed to be at least an order of magnitude denser than the expected minimal spans of LD, to (1) permit increased levels of confidence for the associations found by leveraging association trends among SNPs in LD and (2) use increased resolution to distinguish separate peaks of association. A particular goal was to narrow the search for etiologies of T1D genetically coded in or centromeric to the MHC. In our initial survey, we typed 181 Swedish T1D cases and 182 controls for the 2,360 densely spaced SNP markers that comprise the Illumina MHC SNP set.

We computed single-marker associations between each of our typed SNPs and T1D by calculating *P* values with Fisher's exact method from 3 × 2 contingency tables of case and control genotypes (fig. 1). These are also provided as a table of markers, locations, and *P* values at T1DBase—ITPR3. We identified two peaks in this region that might represent independent associations with T1D. One of these peaks, HLA-DR/DQ, is a reconfirmation of a well-recognized association.[3] As can be seen in graphics of long-range LD (fig. 1*A* and a Haploview analysis of the same data at T1DBase—ITPR3), the nearly 3-Mb telomeric tail of this peak covering the MHC locus can be explained by LD. The second peak is contained within the *ITPR3* gene. It is neither within the MHC nor in a region of extended LD. It is sharp and constrained to a short interval.

In our initial survey, the SNP with the most significant association (survey $P = 2.4 \times 10^{-17}$) lies in an intron of HLA-DQA1 (MIM 146880). The dbSNP accession number for this SNP is *rs9272723*. The *ITPR3* peak is represented most significantly by the SNP *rs2296336* (survey $P = 1.7 \times 10^{-4}$). This SNP is located at chromosome 6 position 33744638, in an intron of *ITPR3*. We applied a permutation multiple-test correction to the *ITPR3* peak and observed a peak meeting the significance criteria in a single iteration of 1,000, suggesting a corrected *P* value for the peak of ~.001 (fig. 2). We computed ORs to measure the strength of association for each SNP in our survey (available in a table at T1DBase—ITPR3). The survey OR for *rs9272723* was 5.8 (95% CI 3.8–9.1). The survey OR for *rs2296336* was 1.8 (95% CI 1.3–2.5). After all global analyses were performed, we separately analyzed the male and female subgroups; in both instances, we observed a peak of association at the *ITPR3* gene (data not shown).

We employed Monte Carlo analysis to estimate the power of our survey study of 363 individuals to detect an effect similar to the effect we observed for *rs2296336*. We simulated 1,000 sets of data for a single locus, with the case genotype frequencies drawn from a trinomial distribution reflecting the observed case frequencies, the observed control frequencies, or nine intermediate frequency sets (table 3). In all cases, the frequencies of the simulated control genotypes were the frequencies of the observed control genotypes for all rows. These simulations suggest that our survey study was adequately powered to observe an effect of *ITPR3* on T1D of the strength observed in Sweden but may not have had the power to observe much weaker effects.

Associations between T1D and non–class II genes within the MHC have a low likelihood of representing a causal

association, as judged by many researchers evaluating reports of such associations that have been subsequently been explained by conditioning on MHC class II genotypes.[13] Therefore, for this article, we must address inference for causal association from two perspectives: (1) that *ITPR3* lies near enough to the MHC to be in the category of candidate genes that must be examined with exceptional care for linkage to an MHC haplotype and (2) that *ITPR3* lies in a genomic region distinct from the MHC. The first perspective requires additional tests to be performed, so it is reasonable to separate discussion of these perspectives for statistical as well as expository reasons.

It is known from prior studies that the MHC span of LD ends telomeric to *ITPR3,* with the region of extended LD in the MHC ending just centromeric to HLA-DPB1 (MIM 142858).[28,29] We also verified this result in the Swedish population we studied. In figure 1*A,* we plot r^2 for all control individuals in our study. Two other observations are consistent with this bound on LD. First, there is a known recombination hotspot in this region ("R" in fig. 1*B*).[30] Second, presumably, a lack of selection pressure to maintain linkage allowed the mouse genome a syntenic breakpoint ("S" in fig. 1*B*) centromeric to *TAPBP* (MIM 601962) with respect to other mammalian genomes. Together, these lines of evidence suggest that it is unlikely that LD plays a major role between HLA-DR/DQ and genes cen-

Table 3. Power of Survey Study of 363 Individuals to Detect an Effect Similar to That Observed at *ITPR3,* as Computed by Monte Carlo Analysis

Subject Group	Case Genotype Ratios Determining Frequencies	No. of Simulated Survey Studies with $P <$				
		.05	.01	.001	.000168	.0001
Cases	39:81:61	993	958	790	558	469
Intermediate 1	36:81:63	981	880	573	321	257
Intermediate 2	34:82:66	944	751	364	154	120
Intermediate 3	31:82:68	816	570	184	51	38
Intermediate 4	29:82:71	682	348	65	11	6
Intermediate 5	26:83:73	449	168	23	4	1
Intermediate 6	23:83:75	263	62	2	0	0
Intermediate 7	21:83:78	85	8	0	0	0
Intermediate 8	18:83:80	29	3	0	0	0
Intermediate 9	16:84:83	4	0	0	0	0
Controls	13:84:85	4	0	0	0	0

NOTE.—Each row represents the result of 1,000 simulations of data for a single locus, with the case genotype frequencies drawn from a trinomial distribution reflecting the observed case frequencies (top row of data), the observed control frequencies (bottom row), or nine intermediate frequency sets. The frequencies of the simulated control genotypes were the frequencies of the observed control genotypes for all rows. As expected, approximately half of the time (558 of 1,000 simulations), we would have observed a more significant *P* value than we actually did for *rs2296336* ($P = .000168$), if the true genotype frequencies were our observed frequencies.

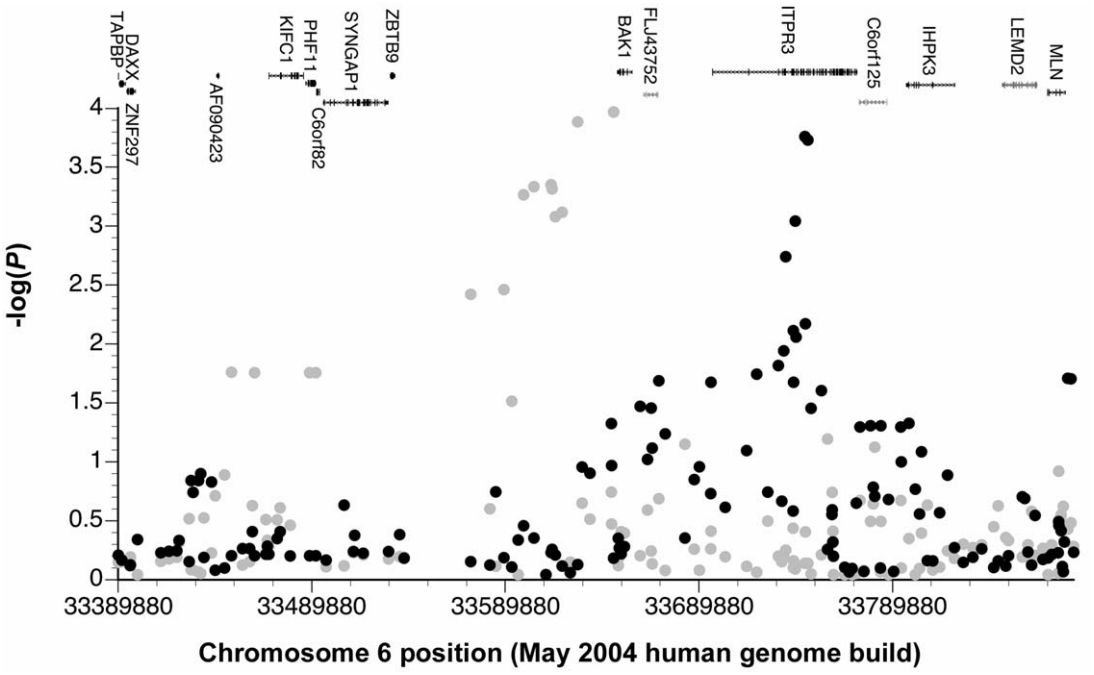

Figure 2. T1D association in the region centromeric to the MHC locus. The most significant association lies within an intron of *ITPR3*. The *X*-axis is indexed to base pairs of the May 2004 human genome build. Gray dots are values from the single most significant of the 1,000 permutations of our survey data (the single iteration was the only 1 of 1,000 that produced such a peak). This visualization permits an appreciation of the contours of the peak of significance, which are much crisper in the peak derived from real data than in the most significant peak derived from the permutations. Although difficult to quantify, this peak shape increases the significance of the survey result. Known genes from the UCSC genome browser are indicated at the top of the figure.

tromeric to *TAPBP*. In particular, our data do not support strong LD between *ITPR3* and HLA-DR/DQ (e.g., $r^2 = 0.022$ between *rs2296336* and *rs9272723*). These lines of evidence suggest that it is appropriate to take the perspective that *ITPR3* is in a region of the genome that is distinct from the MHC locus.

However, in any given generation, *ITPR3* and HLA-DR/DQ haplotypes are fairly tightly maintained. Estimates of linkage from NCBI Map Viewer include 0.96 cM (deCODE), 2.2 cM (Genethon), and 0.55 cM (Marshfield). Over the course of the history of the Swedish population, this proximity may have permitted genetic mechanisms, possibly including currently unpostulated genetic mechanisms, to create a statistical correlation between these loci that is unrelated to T1D causality. Simulations of pedigree linkage studies—but not association studies—have produced associations at a distance from a selected variation in pedigree data sets with coarse marker density.[31] Therefore, even though our data consist of high-density SNP markers typed in a case-control association study in a homogeneous population, caution is appropriate; it is reasonable to take the perspective that *ITPR3* should be considered under the umbrella of the MHC locus. Considering all arguments together, we determined to proceed with a validation study and then to perform two-locus logistic regression between disease status and genotype predictor variables, with both *rs2296336* and each of the MHC class II alleles included as predictor variables in the model, but to limit the number of comparisons to only those MHC class II alleles that associated more significantly with T1D than did *rs2996336* in our study population.

As part of our verification study, sequencing was performed on tiled PCR products across the region of peak association,[32] to detect all variation common to the risk haplotype. We sequenced a 4,375-bp region from chromosome 6 position 33743599 to position 33747974, in 10 individuals: 9 from the SCDS/DIS and 1 unaffected non-Scandinavian white for reference DNA. The nine SCDS/DIS samples included six cases and three controls. All nine were predicted to be homozygous for the *ITPR3* risk haplotype. In the sequenced 4,375-bp region, we identified no variable nonsynonymous SNPs. We did not identify any SNPs that were not previously reported in dbSNP. However, we did identify SNPs in the risk haplotype that had not been typed in our survey study. We therefore included these in our verification study (table 2).

To confirm the association of *ITPR3* with T1D, we genotyped an additional 462 cases and 299 controls for 18 SNPs from our survey study as well as 10 additional SNPs (table 2). The additional subjects were also from DIS and SCDS and were distinct from the 363 individuals included in our survey study.[4,15] This verification confirmed our initially observed associations. For the most significantly associated SNP from our survey study, *rs2296336*, the *P* value computed from only the new data was 1.89×10^{-3}, and the *P* value from all combined 1,124 individuals was 1.30×10^{-6} (table 4).

If the peak of association in *ITPR3* represents a causal influence, such causality could be due a single SNP (or other variation), either at the peak of association or within a nearby region of LD, or due to multiple linked variations in this region. We therefore employed PHASE (fastPHASE Web site) to infer haplotypes among all survey cases and controls.[26] To limit computational complexity, we constrained our haplotype inferences to the eight survey SNPs in *ITPR3* with the most-significant associations; these SNPs span 12 kb. We saw 14 of 256 possible haplotypes; of these, 5 make up 95% of all observed haplotypes (table 5). Only one of these five haplotypes includes the risk allele of *rs2296336*. Construction of a maximum parsimony cladogram divides these haplotypes into two clear sets (fig. 3). The Hamming distance (a measure of the shortest mutational path between two strings) between the closest two haplotypes in separate sets is 5. The two most frequent haplotypes, one of which is the risk haplotype, are complements of each other (i.e., they differ at all eight SNPs), suggesting that they may be ancestral. The diversity is more limited in the risk haplotype group than in the protective haplotype group. If we define a biallelic "synthetic genotype" based on whether an individual's predicted *ITPR3* haplotypes belong to haplotype set 1 (risk) or set 2 (protective), the frequencies of these synthetic genotypes in a T1D risk contingency table are negligibly different from those of *rs2296336*; therefore, *rs2296336* makes an excellent tSNP for this haplotype and risk for T1D, as would be the case if a variant of this SNP were causal. We therefore present our analysis and discussion, using *rs2296336* as a proxy for this haplotype.

As reported in previous publications, we have typing information for HLA-DQA1, HLA-DQB1, and HLA-DRB1 for our samples, with major type available for nearly all samples and with subtype available for most samples.[16] We therefore computed single-point associations with

Table 4. Epidemiology of *ITPR3* T1D Risk

rs2296336 Attribute	Finding
Risk allele (R)	G
Protective allele (P)	C
No. of cases with genotype:	
R/R	107
R/P	296
P/P	240
No. of controls with genotype:	
R/R	35
R/P	217
P/P	229
OR (95% CI) of RP:PP	1.3 (1.00–1.69)
OR (95% CI) of RR:PP	2.9 (1.88–4.59)
PAR percentage (95% CI)	21.2 (9.7–30.5)

NOTE.—Counts are the combined results of the survey and validation studies. PAR is an estimate of the percentage of T1D cases in the Swedish population that would not occur if all alleles of *rs2296336* were C, the protective allele.

T1D of each major type and subtype for these MHC class II genes for the combined samples from our survey and validation studies (table 6). We chose to specifically model those with *P* values more significant than the *P* value for *rs2296336*—modeling each separately with *rs2296336* in a regression analysis to see if any one of these might explain the association with *ITPR3* (table 6). In all cases of conditioning on the 16 MHC class II predictors, the *P* value for an effect on the regression of *rs2296336* after conditioning was $\leqslant .011$. Thus, we can conclude that *rs2296336* is an independent predictor for risk of T1D. However, at least some of the risk predicted by *rs2296336* can be accounted for by MHC class II status (e.g., the regression *P* value for *rs2296336* alone is 2×10^{-6}, but, conditioned on the *DQB1*0302* indicator, this *P* value becomes .01). This observation is consistent with a hypothesis that *ITPR3* has a weaker effect on T1D risk than does MHC class II and is also consistent with a hypothesis of an interaction between the effects of risk at both of these loci, such as a particular variant that is required at both loci for maximal risk.

We fit our observed data to a standard set of regression models (those implemented in LRASSOC). For each comparison, we selected the best model as the one with lowest Akaike information content (AIC), subject to the constraint that it be at least 2 AIC units less than the next best simpler model. The regression models include ADD1 (just *rs2296336*), ADD2 (just the other predictor), ADD (both *rs2296336* and the other predictor), DOM (adding terms for homozygous effects), and ADDDOM (adding interaction terms). If *rs2296336* had nothing to do with T1D,

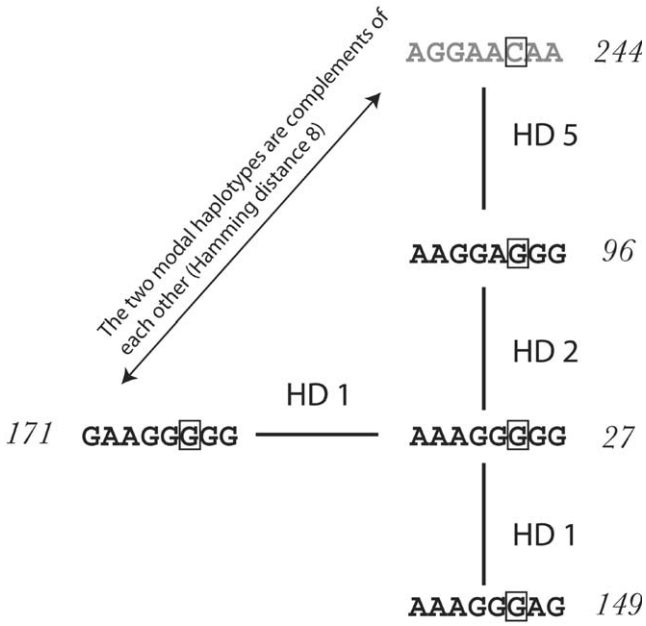

Figure 3. Haplotype Hamming distances (HDs). The risk haplotype is in gray type, and the protective haplotypes are in black type. Numbers in italics indicate the number of each haplotype inferred from PHASE analysis to be present in the our set of cases and controls. The tSNP *rs2296336* is boxed. The Illumina allele codes for each of the loci are as follows: *rs2274200,* A→A, G→B; *rs2296343,* A→A, G→B; *rs2296339,* A→A, G→B; *rs2296337,* A→A, G→B; *rs1536036,* A→A, G→B; *rs2296336,* C→A, G→B; *rs2077163,* A→A, G→B; and *rs2229634,* A→A, G→B.

Table 5. Haplotypes at the *ITPR3* Locus

	Allele at SNP								No. of Haplotypes Predicted by PHASE		
Haplotype	*rs2274200*	*rs2296343*	*rs2296339*	*rs2296337*	*rs1536036*	*rs2296336*	*rs2077163*	*rs2229634*	Total	In Patients	In Controls
1	A	A	G	G	A	C	A	A	13	11	2
2	A	A	G	A	A	C	A	A	3	1	2
3	A	G	G	G	A	C	A	A	1	0	1
4	A	G	G	A	A	C	A	A	244	144	100
5	A	G	G	A	G	C	A	A	2	0	2
6	A	A	G	G	A	C	A	G	5	2	3
7	A	A	A	G	G	G	A	G	149	70	79
8	A	A	A	G	G	C	A	G	1	1	0
9	A	A	A	G	G	G	G	G	27	14	13
10	A	A	G	G	A	G	G	G	96	43	53
11	A	G	G	G	A	G	G	G	2	1	1
12	G	A	A	G	G	G	G	G	171	69	102
13	G	A	G	G	A	G	G	G	2	1	1
14	G	A	G	G	G	G	G	G	10	5	5
Total									726	362	364

NOTE.—Predicted haplotypes inferred from the eight most significantly associated SNPs fall into two distinct sets. The "risk" set (haplotypes 1–6) consists primarily of a single haplotype (4), whereas the "protective" set (haplotypes 7–14) has several common haplotypes. Haplotype 8, predicted to occur only once, clusters with the other protective haplotypes (fig. 3), but, since it shares the *rs2296336* allele with the risk haplotype set, it could be considered a risk haplotype. This table demonstrates that a single core haplotype is responsible for most of the association observed between *ITPR3* and T1D in Sweden. This haplotype most likely represents the state of these eight SNPs in the ancestral risk chromosome. The Illumina allele codes for each of the loci are as follows: *rs2274200,* A→A, G→B; *rs2296343,* A→A, G→B; *rs2296339,* A→A, G→B; *rs2296337,* A→A, G→B; *rs1536036,* A→A, G→B; *rs2296336,* C→A, G→B; *rs2077163,* A→A, G→B; and *rs2229634,* A→A, G→B.

Table 6. Association of MHC Class II Types and Subtypes with T1D and Two-Locus Logistic Regression

Allele	Single-Point Association *P*	Best Two-Locus Model with *rs2296336*	Regression *P* for *rs2296336* (ADD1 Model)	Regression *P* for *rs2296336* Conditioned on Locus (ADD Model)
rs9272723	7.92×10^{-19}	ADDDOM	.000079	.000195
DQA1*01	3.40×10^{-43}	ADDDOM	.000001	.000083
DQB1*0302	3.33×10^{-35}	DOM	.000002	.010114
DRB1*15	4.68×10^{-34}	DOM	.000003	.000021
DQB1*0602	1.77×10^{-33}	DOM	.000002	.000014
DRB1*04	1.73×10^{-26}	DOM	.000003	.003781
DQA1*03	1.77×10^{-24}	DOM	.000001	.001643
DQA1*0102	4.98×10^{-22}	ADD	.000001	.000005
DQB1*03	3.17×10^{-20}	ADD	.000002	.000005
DRB1*03	7.31×10^{-18}	ADD	.000003	.000001
DQB1*0603	7.95×10^{-12}	ADD	.000002	.000005
DQA1*0103	2.48×10^{-11}	ADD	.000001	.000002
DQB1*0301	2.44×10^{-9}	ADD	.000002	.000003
DQB1*02	2.68×10^{-8}	DOM	.000002	.000049
DRB1*13	5.20×10^{-8}	ADD	.000003	.000013
DQA1*05	9.28×10^{-8}	DOM	.000001	<.000001
DRB1*11	5.36×10^{-7}	ADD	.000003	.000005

NOTE.—MHC class II types and subtypes with association *P* values less significant than that of *rs2296636* or with insufficient diversity in our study population to support regression analysis are not shown. Each MHC allele is treated as a biallelic locus: either present, absent, or not typed. Only individuals with unambiguous genotypes at both considered loci were included in the regression models for each row; thus, since only 363 individuals were typed for *rs9272723* in our survey, the *P* value for *rs2296336* is less significant than the *P* values for other alleles that were computed from more individuals. Regression of the most significant single-point SNP (*rs9272723*), which lies in an intron of HLA-DQA1, produces results that are consistent with regression with the MHC class II predictors. Regression with DQA1*01 (or *rs9272723*) suggests that an interactive term may be appropriate for modeling the risk associated with *ITPR3* and the MHC. The last column demonstrates that the T1D association with *rs2296336* remains significant even when conditioned on an MHC class II predictor. Tests are not independent (e.g., DQA1*0302 is a subset of DQA1*03). A more extensive version of this table is available at T1DBase.

then the best model selected by LRASSOC should be ADD2. ADD2 is never observed to be the best model, nor is ADD1. All the best models observed employ both predictors. The DOM model would be predicted for classic epistasis between the loci. The ADD model would be predicted if both loci contributed to risk independently. The ADDDOM model would be predicted if particular variants at each loci were both required for risk, as might be the case if the products of these variants interacted with each other. However, the converses of these predictions are not necessarily true (e.g., it is not the case that, if the best model is ADDDOM, then the products of the loci interact). Regression with DQA1*01, which yielded the most significant T1D association with an MHC class II predictor in single-point analysis, suggests that an interactive term may be appropriate in modeling risk when the combination of *ITPR3* and MHC effects on T1D are considered; ADDDOM was the best model. Each of these analyses is based on data summarized by a 9 × 2 contingency table and yields a linear model for predicting the proportion of those with a given multilocus genotype who represent cases rather than controls. As an example, the contingency table for DQA1*01 is provided (table 7), and a comparison of the results of the regression model with the observed ratios is shown in figure 4. In all contingency tables produced for these regression analyses, homozygosity for the risk variants of both predictors produces the highest risk, although data for this double homozygous genotype are fairly sparse. The ORs of homozygous risk genotypes compared with homozygous protective genotypes computed from these two-locus contingency tables tend to approximate the product of the two single-locus ORs computed separately (supplemental data are available at T1DBase—ITPR3).

In addition to comparisons with the MHC class II genes as typed using traditional approaches, we also tested a two-locus regression model between disease status and two genotype predictor variables: *rs2296336* and the most significant single-point SNP from our survey data (*rs9272723*, which lies in an intron of HLA-DQA1), although this required limiting the number of individuals for regression to the 363 individuals tested in our survey (table 6). Regression with *rs9272723* produces results that are consistent with regression results based on HLA-DQA1. This result suggests that typing of noncoding MHC SNPs has at least limited value in reflecting risk factors that are associated with variation in MHC class II coding regions.

To further explore evidence of independent mechanisms of action of the HLA-DR/DQ and *ITPR3* associations, we examined the correlation of the age-at-onset phenotype with genotype. Age at onset of T1D is known to correlate with HLA-DR/DQ.[4,16] In our data, for example, the

Table 7. Contingency Table for *ITPR3* and HLA-DQA1*01

ITPR3	DQA1*01	No. of Cases	No. of Controls
R/R	+/+	5	0
R/R	+/−	24	25
R/R	−/−	5	79
R/P	+/+	34	8
R/P	+/−	90	85
R/P	−/−	83	194
P/P	+/+	44	5
P/P	+/−	119	69
P/P	−/−	62	156

NOTE.—$\chi^2 = 204.6$; $P < 2.2 \times 10^{-16}$. P = protective allele; R = risk allele. The combination of homozygous risk alleles for both *ITPR3* and DQA1*01 (top line) increases the risk for T1D over other genotypes. This table is representative of the set of contingency tables for *ITPR3* and a highly associated MHC class II type or subtype (all contingency tables are available at T1DBase—ITPR3). To minimize the number of estimated variables for regression analysis, each DQA allele is coded as one of only two possibilities: the allele is DQA1*01 (+) or is not DQA1*01 (−). Separate contingency tables were computed in a similar fashion for each typed class II locus and allele.

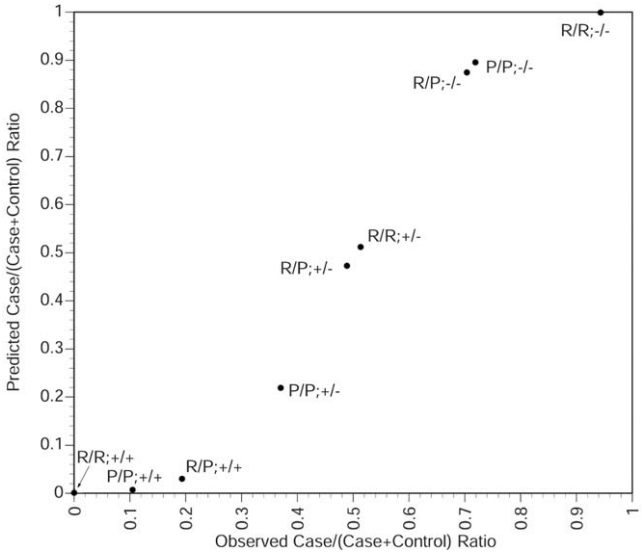

Figure 4. Regression analysis for predicting the proportion that are cases versus controls, given the genotype, with *rs2296336* and HLA-DQA1*01 status as predictor variables. P = protective allele of *ITPR3*; R = risk allele of *ITPR3*; + = DQA1*01 present; − = DQA1*01 absent.

mean age at onset in case patients with a homozygous protective HLA-DR/DQ haplotype is 21.5 years but drops to 15.7 years for the homozygous risk haplotype (P = .030). However, the association at *ITPR3* does not correlate with age at onset (P = .64). The different correlation with age supports the hypothesis that the MHC and ITPR3 have independent mechanisms contributing to T1D risk.

Discussion

The biology of ITPR3 makes *ITPR3* an excellent candidate gene. ITPRs mediate second messenger signaling by releasing Ca^{++} from intracellular stores in response to inositol triphosphate. There are three ITPR subtypes, and these are widely expressed in many tissues, consistent with their role in mediating second messenger signaling. Each of the subtypes has a unique tissue-distribution profile. *ITPR3* mRNA and ITPR3 protein are rapidly upregulated in pancreatic β cells following stimulation with glucose, and the protein is then rapidly degraded in proteasomes.[33] *ITPR3* mRNA is one of the most strongly upregulated messages in this system. Srivastava et al.[34] showed that the *Anx7*$^{+/-}$ knockout mouse model had reduced ITPR3 expression in the endoplasmic reticulum of pancreatic β cells, which caused defects in Ca^{++}-dependent insulin secretion. There is no report of the phenotype of the single *Itpr3* knockout, but the double *Itpr2/Itpr3* knockout mouse is reported to be hypoglycemic and lean and to have impaired exocrine function.[35] Thus, ITPR3 is, in a biological context, consistent with a causal association with T1D. Variation at *rs2296336* may influence risk for T1D through an effect on alternative splicing, since the most significant variations of the *rs2296336* haplotype lie within introns of *ITPR3*.

The combination of high-throughput genotyping, the completed genome sequences of human and mouse, the availability of cases and controls from homogeneous populations, linkage studies that permit targeting particular loci for fine-scale analysis, and mature bioinformatics software and algorithms has increasingly made the discovery of variation contributing to complex diseases possible. For example, the findings of associations of *IFIH1* (MIM 606951) with T1D, *TCF7L2* (MIM 602228) with T2D, and *INSIG2* (MIM 608660) with obesity (MIM 601665) represent examples of such success.[36–38] In association studies, P values for common variants that moderately influence risk may be less significant than those for rare variants that drastically influence risk. This phenomenon is observed in figure 1*C*, in which rare alleles of HLA-DR/DQ that confer near-absolute protection from T1D are at least somewhat responsible for a relatively significant statistical association compared with that of *ITPR3*. Nevertheless, both of these genes may have comparable influences on public health in Sweden. Detection of such genes seems to be possible with studies comprising between 300 and 1,200 individuals and with marker resolutions between 3 and 46 kb (in our study and the study of Grant et al.,[37] respectively). However, in our study, the effective haplotype block length for the *ITPR3* association is ~13 kb. The implications for future mapping projects may be that, to find more associations for many disease genes, the search must be performed at a resolution <13 kb. High-resolution mapping can (1) provide precision, possibly limiting the number of candidate genes to exactly one and possibly even identifying the causal coding SNP; (2) find associations missed by coarser association studies; and (3) resolve nearby independent associations in complex multigenic disorders that, in a more coarse study, would have been lumped as a single association. To find all genes contributing to complex disorders, it may, in fact, be necessary to perform such mapping for every candidate gene and perhaps for the entire genome.

We remain cautious in interpreting association as evidence of causality. In the event that causality between ITPR3 and T1D is ever disproved, then our results may possibly retain some interest, in that, at least within the Swedish population, genotyping of *ITPR3* together with other assays may represent a convenient method of evaluating T1D risk as a proxy for actual causal variation. If one accepts the argument that the peak of association centered in the *ITPR3* gene is due at least in part to the influence on phenotype of a nearby variation and is not due entirely to influence on phenotype from long-distance correlation with variation within the MHC locus, one might still consider rejecting the argument that the nearby phenotypic influence must necessarily be due to alteration in the function of ITPR3. If *ITPR3* is itself not causal, but yet there is a nearby casual genetic variation,

then any of the genes in the 3.1-Mb mouse interval *Idd16* would be a candidate.[12] *IHPK3* (MIM 606993), centromeric to *ITPR3,* plays a role in inositol triphosphate metabolism. *C6orf125* lies between *ITPR3* and *IHPK3*. We believe that one of the most valuable contributions of fine-scale SNP association mapping is that it provides data that can be visualized such that clearly defined and sharp peaks of association can be seen. We see such a sharp peak of association within a specific domain of the *ITPR3* gene in figure 2. This peak lies entirely within the *ITPR3* gene and points like an arrow to the *TRPC3* (MIM 602345) binding domain.[39] Variation in nearby unidentified genes or regulatory sequences may also be at play.[40] Follow-up by many labs at many levels, ultimately including clinical studies, will be necessary to confirm that *ITPR3* is more than a candidate T1D gene and to understand the possible role of nearby genes. A systems-biology analysis that more fully places ITPR3 in the contextual model for T1D, as well as typing of *ITPR3* variation in other populations, may aid such a process.

Our discovery of *ITPR3* as a T1D-associated gene in close association with HLA-DR/DQ does not represent the first time a non-MHC gene in or near the MHC locus has been associated with disease. For example, the gene associated with 21-hydroxylase deficiency (MIM 201910) was identified in the 1980s.[41] We doubt *ITPR3* will be the last such gene. The high density of genes in the MHC may be due, in part, to paired interaction of linkage and selection (also known as "hitchhiking"), which is what we theorize HLA-DR/DQ and *ITPR3* may share.[42,43] The hint of an interaction term in our regression analysis (table 6) is consistent with such a model.[44] Observed interactions between the effects of genes may be facilitated by overdominant selection or heterozygote advantage, as has been observed already in the MHC.[45]

Author Groups

The following authors are from the Diabetes Incidence in Sweden Study Group: Hans Arnqvist, Department of Internal Medicine, University of Linköping, Linköping; Elisabeth Björck, Department of Medicine, University Hospital, Uppsala; Jan Eriksson, Department of Medicine, University of Umeå, Umeå; Lennarth Nyström, Department of Epidemiology and Public Health, University of Umeå, Umeå; Lars Olof Ohlson, Sahlgrenska Hospital, University of Göteborg, Göteborg; Bengt Scherstén, Department of Community Health Sciences, Dahlby, University of Lund, Lund; and Jan Östman, Center for Metabolism and Endocrinology, Huddinge University Hospital, Stockholm.

The following authors are from the Swedish Childhood Diabetes Study Group, all from Departments of Pediatrics: M. Aili, Halmstad; L.E. Bååth, Östersund; E. Carlsson, Kalmar; H. Edenwall, Karlskrona; G. Forsander, Falun; B. W. Granström, Gällivare; I. Gustavsson, Skellefteå; R. Hanås, Uddevalla; L. Hellenberg, Nyköping; H. Hellgren, Lidköping; E. Holmberg, Umeå; H. Hörnell, Hudiksvall; Sten-A. Ivarsson, Malmö; C. Johansson, Jönköping; G. Jonsell, Karlstad; K. Kockum, Ystad; B. Lindblad, Mölndal; A. Lindh, Borås; J. Ludvigsson, Linköping; U. Myrdal, Västerås; J. Neiderud, Helsingborg; K. Segnestam, Eskilstuna; S. Sjöblad, Lund; L. Skogsberg, Boden; L. Strömberg, Norrköping; U. Ståhle, Ängelholm; B. Thalme, Huddinge; K. Tullus, Danderyd; T. Tuvemo, Uppsala; M. Wallensteen, Stockholm; O. Westphal, Göteborg; and J. Åman, Örebro.

Acknowledgments

This study was supported by National Institutes of Health grants 5K08AI056092 (from the National Institute of Allergy and Infectious Diseases [NIAID] to J.C.R.), DK26190 (from the National Institute of Diabetes and Digestive and Kidney Diseases [NIDDK] to Å.L.], and DK532004 (from NIDDK to Å.L. and M.J.). The study also received support from the Juvenile Diabetes Research Foundation (JDRF) and NIAID grant 5U54AI054523. A.F.S. holds the Grant I. Butterbaugh Professorship at the University of Washington. John Stuelpnagel and Mark Chee facilitated use of the Illumina technology. T1DBase (a JDRF project) and its staff, including Erin Helton, Burak Kutlu, David Burdick, and Nat Goodman, were invaluable. Jinko Graham, Pierre-Yves Bochud, Lee Rowen, and Alan Aderem provided knowledge, support, and mentorship.

Web Resources

The URLs for data presented herein are as follows:

dbSNP, http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=Snp
EpiTools, http://www.epitools.net/
fastPHASE, http://www.stat.washington.edu/stephens/software.html
Cran—Package genetics, http://cran.us.r-project.org/src/contrib/Descriptions/genetics.html (for R packages HWE and LD)
Illumina: MHC Mapping Panel, http://www.illumina.com/products/snp/mhc_mapping_panel.ilmn
Illumina: Technology and Applications, http://www.illumina.com/technology/tech_overview.ilmn
IMGT/HLA Database, http://www.ebi.ac.uk/imgt/hla/
MAGMA, http://snp-magma.sourceforge.net/
NCBI Map Viewer, http://www.ncbi.nlm.nih.gov/mapview/map_search.cgi?taxid=9606
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim/ (for T1D, *ITPR3,* T2D, *RFP, KNSL2, MICB, C4A,* HLA-DQA1, HLA-DPB1, *TAPBP, IFIH1, TCF7L2, INSIG2,* obesity, *IHPK3, TRPC3,* and 21-hydroxylase deficiency)
The MHC Haplotype Project, http://www.sanger.ac.uk/HGP/Chr6/MHC/
T1DBase, http://t1dbase.org/
T1DBase—ITPR3, http://t1dbase.org/downloads/ITPR3/
The R Project for Statistical Computing, http://www.r-project.org/
MGC/ORFenomes Genome Browser, http://mgc.ucsc.edu/FAQ/FAQdownloads

References

1. Green A, Patterson CC (2001) Trends in the incidence of childhood-onset diabetes in Europe 1989–1998. Diabetologia Suppl 3 44:B3–B8
2. Atkinson MA, Eisenbarth GS (2001) Type 1 diabetes: new

perspectives on disease pathogenesis and treatment. Lancet 358:221–229
3. Concannon P, Erlich HA, Julier C, Morahan G, Nerup J, Pociot F, Todd JA, Rich SS, Type 1 Diabetes Genetics Consortium (2005) Type 1 diabetes: evidence for susceptibility loci from four genome-wide linkage scans in 1,435 multiplex families. Diabetes 54:2995–3001
4. Graham J, Hagopian WA, Kockum I, Li LS, Sanjeevi CB, Lowe RM, Schaefer JB, Zarghami M, Day HL, Landin-Olsson M, Palmer JP, Janer-Villanueva M, Hood L, Sundkvist G, Lernmark A, Breslow N, Dahlquist G, Blohme G, Diabetes Incidence in Sweden Study Group, Swedish Childhood Diabetes Study Group (2002) Genetic effects on age-dependent onset and islet cell autoantibody markers in type 1 diabetes. Diabetes 51:1346–1355
5. Pociot F, Karlsen AE, Pedersen CB, Aalund M, Nerup J (2004) Novel analytical methods applied to type 1 diabetes genome-scan data. Am J Hum Genet 74:647–660
6. Maier LM, Wicker LS (2005) Genetic susceptibility to type 1 diabetes. Curr Opin Immunol 17:601–608
7. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, Gough SCL, Jenkins SC, Palmer SM, Balfour KM, Rowe BR, Farrall M, Barnett AH, Bain SC, Todd JA (1994) A genome-wide search for human type 1 diabetes susceptibility genes. Nature 371:130–136
8. Park Y (2004) Prediction of the risk of type 1 diabetes from polymorphisms in candidate genes. Diabetes Res Clin Pract Suppl 1 66:S19–S25
9. European Consortium for IDDM Genome Studies (2001) A genomewide scan for type 1-diabetes susceptibility in Scandinavian families: identification of new loci with evidence of interactions. Am J Hum Genet 69:1301–1313
10. Onengut-Gumuscu S, Concannon P (2005) The genetics of type 1 diabetes: lessons learned and future challenges. J Autoimmun Suppl 25:34–39
11. Herr M, Dudbridge F, Zavattari P, Cucca F, Guja C, March R, Campbell RD, Barnett AH, Bain SC, Todd JA, Koeleman BP (2000) Evaluation of fine mapping strategies for a multifactorial disease locus: systematic linkage and association analysis of IDDM1 in the HLA region on chromosome 6p21. Hum Mol Genet 9:1291–1301
12. Deruytter N, Boulard O, Garchon HJ (2004) Mapping non-class II H2-linked loci for type 1 diabetes in nonobese diabetic mice. Diabetes 53:3323–3327
13. Clerget-Darpoux F (2000) The HLA component of type I diabetes. Am J Hum Genet 66:1468
14. Landin-Olsson M, Karlsson FA, Lernmark A, Sundkvist G (1992) Islet cell and thyrogastric antibodies in 633 consecutive 15- to 34-yr-old patients in the diabetes incidence study in Sweden. Diabetes 41:1022–1027
15. Landin-Olsson M, Palmer JP, Lernmark A, Blom L, Sundkvist G, Nystrom L, Dahlquist G (1992) Predictive value of islet cell and insulin autoantibodies for type 1 (insulin-dependent) diabetes mellitus in a population-based study of newly-diagnosed diabetic and matched control children. Diabetologia 35:1068–1073
16. Graham J, Kockum I, Sanjeevi CB, Landin-Olsson M, Nystrom L, Sundkvist G, Arnqvist H, Blohme G, Lithner F, Littorin B, Schersten B, Wibell L, Ostman J, Lernmark A, Breslow N, Dahlquist G (1999) Negative association between type 1 diabetes and HLA DQB1*0602-DQA1*0102 is attenuated with age at onset. Swedish Childhood Diabetes Study Group. Eur J Immunogenet 6:117–127
17. Hubley R, Zitzler E, Siegel AF, Roach J (2002) Multiobjective genetic marker selection. Advances in nature-inspired computation: the PPSN VII workshops 7:32–33
18. Hubley RM, Zitzler E, Roach JC (2003) Evolutionary algorithms for the selection of single nucleotide polymorphisms. BMC Bioinformatics 4:30
19. Carlson CS, Eberle MA, Rieder MJ, Yi Q, Kruglyak L, Nickerson DA (2004) Selecting a maximally informative set of single-nucleotide polymorphisms for association analysis using linkage disequilibrium. Am J Hum Genet 74:106–120
20. Jurinke C, van den Boom D, Cantor CR, Koster H (2002) Automated genotyping using DNA MassARRAY technology. Methods Mol Biol 187:179–192
21. Oeth P, Beaulieu M, Park C, Kosman D, del Mistro G, van den Boom D, Jurinke C (2005) iPLEX assay: increased plexing efficiency and flexibility for MassARRAY system through single base primer extension with kass-modified terminators. Sequenom application note. Document 8876–006, R01. CO 0500154
22. Stewart CA, Horton R, Allcock RJ, Ashurst JL, Atrazhev AM, Coggill P, Dunham I, et al (2004) Complete MHC haplotype sequencing for common disease gene mapping. Genome Res 14:1176–1178
23. de Bakker PI, Yelensky R, Pe'er I, Gabriel SB, Daly MJ, Altshuler D (2005) Efficiency and power in genetic association studies. Nat Genet 37:1217–1223
24. Smink LJ, Helton EM, Healy BC, Cavnor CC, Lam AC, Flamez D, Burren OS, Wang Y, Dolman GE, Burdick DB, Everett VH, Glusman G, Laneri D, Rowen L, Schuilenburg H, Walker NM, Mychaleckyj J, Wicker LS, Eizirik DL, Todd JA, Goodman N (2005) T1DBase, a community Web-based resource for type 1 diabetes research. Nucleic Acids Res Database Issue 33: D544–D549
25. Barrett JC, Fry B, Maller J, Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21:263–265
26. Stephens M, Scheet P (2005) Accounting for decay of linkage disequilibrium in haplotype inference and missing-data imputation. Am J Hum Genet 76:449–462
27. North BV, Curtis D, Sham PC (2005) Application of logistic regression to case-control association studies involving two causative loci. Hum Hered 59:79–87
28. Miretti MM, Walsh EC, Ke X, Delgado M, Griffiths M, Hunt S, Morrison J, Whittaker P, Lander ES, Cardon LR, Bentley DR, Rioux JD, Beck S, Deloukas P (2005) A high-resolution linkage-disequilibrium map of the human major histocompatibility complex and first generation of tag single-nucleotide polymorphisms. Am J Hum Genet 76:634–646
29. Lincoln MR, Montpetit A, Cader MZ, Saarela J, Dyment DA, Tiislar M, Ferretti V, Tienari PJ, Sadovnick AD, Peltonen L, Ebers GC, Hudson TJ (2005) A predominant role for the HLA class II region in the association of the MHC region with multiple sclerosis. Nat Genet 2005 37:1108–1112
30. Cullen M, Perfetto SP, Klitz W, Nelson G, Carrington M (2002) High-resolution patterns of meiotic recombination across the human major histocompatibility complex. Am J Hum Genet 71:759–776
31. Roberts SB, MacLean CJ, Neale MC, Eaves LJ, Kendler KS (1999) Replication of linkage studies of complex traits: an

examination of variation in location estimates. Am J Hum Genet 65:876–884
32. Livingston RJ, von Niederhausern A, Jegga AG, Crawford DC, Carlson CS, Rieder MJ, Gowrisankar S, Aronow BJ, Weiss RB, Nickerson DA (2004) Pattern of sequence variation across 213 environmental response genes. Genome Res 14:1821–1831
33. Lee B, Gai W, Laychock SG (2001) Proteasomal activation mediates down-regulation of inositol 1,4,5-trisphosphate receptor and calcium mobilization in rat pancreatic islets. Endocrinology 142:1744–1751
34. Srivastava M, Eidelman O, Leighton X, Glasman M, Goping G, Pollard HB (2002) Anx7 is required for nutritional control of gene expression in mouse pancreatic islets of Langerhans. Mol Med 8:781–797
35. Futatsugi A, Nakamura T, Yamada MK, Ebisui E, Nakamura K, Uchida K, Kitaguchi T, Takahashi-Iwanaga H, Noda T, Aruga J, Mikoshiba K (2005) IP3 receptor types 2 and 3 mediate exocrine secretion underlying energy metabolism. Science 309:2232–2234
36. Smyth DJ, Cooper JD, Bailey R, Field S, Burren O, Smink LJ, Guja C, Ionescu-Tirgoviste C, Widmer B, Dunger DB, Savage DA, Walker NM, Clayton DG, Todd JA (2006) A genome-wide association study of nonsynonymous SNPs identifies a type 1 diabetes locus in the interferon-induced helicase (IFIH1) region. Nat Genet 38:617–619
37. Grant SF, Thorleifsson G, Reynisdottir I, Benediktsson R, Manolescu A, Sainz J, Helgason A, et al (2006) Variant of transcription factor 7-like 2 (TCF7L2) gene confers risk of type 2 diabetes. Nat Genet 38:320–323
38. Herbert A, Gerry NP, McQueen MB, Heid IM, Pfeufer A, Illig T, Wichmann HE, Meitinger T, Hunter D, Hu FB, Colditz G, Hinney A, Hebebrand J, Koberwitz K, Zhu X, Cooper R, Ardlie K, Lyon H, Hirschhorn JN, Laird NM, Lenburg ME, Lange C, Christman MF (2006) A common genetic variant is associated with adult and childhood obesity. Science 312:279–283
39. Bezprozvanny I (2005) The inositol 1,4,5-trisphosphate receptors. Cell Calcium 38:261–272
40. Glusman G, Qin S, El-Gewely MR, Siegel AF, Roach JC, Hood L, Smit AF (2006) A third approach to gene prediction suggests thousands of additional human transcribed regions. PLoS Comput Biol 2:e18
41. New MI (1986) Molecular genetics and the characterization of steroid 21-hydroxylase deficiency. Endocr Res 12:505–522
42. Andolfatto P (2001) Adaptive hitchhiking effects on genome variability. Curr Opin Genet Dev 11:635–641
43. Smith JM, Haigh J (1974) The hitch-hiking effect of a favourable gene. Genet Res 23:23–35
44. Christiansen FB, Frydenberg O (1977) Selection-mutation balance for two nonallelic recessives producing an inferior double homozygote. Am J Hum Genet 29:195–207
45. Hughes AL, Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. Nature 335:167–170