Fine Mapping, Gene Content, Comparative Sequencing, and Expression Analyses Support Cila4 and Nramp1 as Candidates for Idd5.1 and Idd5.2 in the Nonobese Diabetic Mouse

Linda S. Wicker, Giselle Chamberlain, Kara Hunter, Dan Rainbow, Sarah Howlett, Paul Tiffen, Jan Clark, Andrea Gonzalez-Munoz, Anne Marie Cumiskey, Raymond L. Rosa, Joanna M. Howson, Luc J. Smink, Amanda Kingsnorth, Paul A. Lyons, Simon Gregory, Jane Rogers, John A. Todd and Laurence B. Peterson

*J Immunol* 2004; 173:164-173; doi: 10.4049/jimmunol.173.1.164

http://www.jimmunol.org/content/173/1/164

**References**

This article cites 28 articles, 13 of which you can access for free at: http://www.jimmunol.org/content/173/1/164.full#ref-list-1

**Subscription**

Information about subscribing to *The Journal of Immunology* is online at:

http://jimmunol.org/subscription

**Permissions**

Submit copyright permission requests at:

http://www.aai.org/About/Publications/JI/copyright.html

**Email Alerts**

Receive free email-alerts when new articles cite this article. Sign up at:

http://jimmunol.org/alerts
Fine Mapping, Gene Content, Comparative Sequencing, and Expression Analyses Support Cila4 and Nramp1 as Candidates for Idd5.1 and Idd5.2 in the Nonobese Diabetic Mouse


At least two loci that determine susceptibility to type 1 diabetes in the NOD mouse have been mapped to chromosome 1, Idd5.1 (insulin-dependent diabetes 5.1) and Idd5.2. In this study, using a series of novel NOD.B10 congenic strains, Idd5.1 has been defined to a 2.1-Mb region containing only four genes, Cila4, Icos, Als2cr19, and Nrpl2 (neuropilin-2), thereby excluding a major candidate gene, Cd28. Genomic sequence comparison of the two functional candidate genes, Cila4 and Icos, from the B6 (resistant at Idd5.1) and the NOD (susceptible at Idd5.1) strains revealed 62 single nucleotide polymorphisms (SNPs), only two of which were in coding regions. One of these coding SNPs, base 77 of Cila4 exon 2, is a synonymous SNP and has been correlated previously with type 1 diabetes susceptibility and differential expression of a CTLA-4 isoform. Additional expression studies in this work support the hypothesis that this SNP in exon 2 is the genetic variation causing the biological effects of Idd5.1. Analysis of additional congenic strains has also localized Idd5.2 to a small region (1.52 Mb) of chromosome 1, but in contrast to the Idd5.1 interval, Idd5.2 contains at least 45 genes. Notably, the Idd5.2 region still includes the functionally polymorphic Nramp1 gene. Future experiments to test the identity of Idd5.1 and Idd5.2 as Cila4 and Nramp1, respectively, can now be justified using approaches to specifically alter or mimic the candidate causative SNPs. The Journal of Immunology, 2004, 173: 164–173.

The Idd5 interval (insulin-dependent diabetes 5) from diabetes-resistant C57BL/10 (B10) or C57BL/6 (B6) mice provides protection from type 1 diabetes (T1D) when introgressed onto the NOD background (1, 2). Idd5 is located on mouse chromosome 1 and has been shown by congeneric strain analysis to consist of at least two loci, Idd5.1 and Idd5.2, positioned near the proximal and distal ends, respectively, of a 9.4-cM interval (1). Idd5.1 was defined as a 1.5-cM B10-derived resistance interval containing the candidate genes Casp8, Cflar (FLIP), Cil28, and Cilad (1). Diabetes protection was also observed in NOD.B6 congenic strains developed in the Idd5.1 region (2). The proximal boundary of Idd5.1 was more precisely defined in this second study, and Casp8 and Cflar were eliminated as candidates. In addition, Lamhamedi-Cherradi et al. (2) provided evidence that another T cell regulatory molecule, ICOS (Icos), is encoded immediately distal of Cila4 and is included in the Idd5.1 region.

The Idd5.2 region was localized previously to a 5.1-cM portion of the Idd5 interval using NOD.B10 congenic strains (1). Among the many candidate genes present in the large Idd5.2 region, Nramp1 (which has recently been renamed to Slc11a1) which denotes the fact that the NRAMP1 protein is a member of the solute carrier family 11) is the most compelling because there is a known functional nonsynonymous polymorphism (Gly169 > Asp169) distinguishing the NOD and B10 Nramp1 alleles. The NOD NRAMP1 protein allotype, a divalent cation transporter in phagosomes, is wild-type and mediates protection from certain infectious diseases, and the diabetes-resistant B10 allotype is a functional knockout (3). It has been proposed that NRAMP1-mediated reduction of iron in the phagosomes reduces the ability of some microbes to survive in this intracellular environment (4).

The B10-derived Idd5 interval alone causes only a partial reduction in T1D frequency when present on the NOD background because 30% of NOD.B10 Idd5 females are diabetic by 7 mo of age as compared with 80% of NOD females. There is nearly complete protection from diabetes (<2% disease frequency in females) when a disease-resistant Idd5 interval is combined with a diabetes-protective allele at Idd3 on chromosome 3 (1, 5). This combination of Idd5 and Idd3 also provides significant protection from the development of insulitis (1, 5) and of insulin autoantibodies (5).

In the current study, we describe congenic strains that significantly reduce the limits of the Idd5.1 and Idd5.2 intervals. A physical map of both intervals has been assembled based on the primary DNA sequence, and the gene content was determined. A

*Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Department of Medical Genetics, Cambridge Institute for Medical Research, University of Cambridge, Cambridge, United Kingdom; †Wellcome Trust Sanger Institute, Hinxton, Cambridge, United Kingdom; and ‡Department of Pharmacology, Merck Research Laboratories, Rahway, NJ 07065

Received for publication October 1, 2003. Accepted for publication March 26, 2004.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked advertisement in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

1 L.S.W., P.A.L., and J.A.T. are supported by grants from the Juvenile Diabetes Research Foundation and the Wellcome Trust. The availability of NOD congenic mice through the Taconic Farms Emerging Models Program has been supported by grants from the Merck Genome Research Institute, National Institute of Allergy and Infectious Diseases, and the Juvenile Diabetes Research Foundation.

2 Address correspondence and reprint requests to Dr. Linda S. Wicker, Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, Wellcome Trust/MRC Building, Addenbrooke’s Hospital, Cambridge, CB2 2XY, U.K. E-mail address: linda.wicker@cmr.cam.ac.uk

3 Abbreviations used in this paper: Idd, insulin-dependent diabetes; BAC, bacterial artificial chromosome; flCTLA, full-length CTLA; flICOS, full-length ICOS; hCTLA, ligand-independent CTLA; NRAMP1, natural resistance-associated macrophage protein 1; NRPL2, neuropilin-2; qPCR, quantitative PCR; scCTLA, soluble CTLA; sICOS, soluble ICOS; SNP, single nucleotide polymorphism; T1D, type 1 diabetes.
comparison of the B6 and NOD genome sequences from the Idd5.1 region containing the two primary candidate genes, Cita4 and Icos, showed a moderately high level of polymorphism, suggesting that the B10 and NOD alleles belong to different ancestral haplotypes. Expression of mRNAs for full-length and soluble ICOS, as well as full-length CTLA-4 (flCTLA-4) and two additional CTLA-4 isoforms, ligand-independent (lCTLA-4) and soluble (sCTLA-4), was assessed in Idd5.1 congenic mice. The observed patterns of gene expression support the hypothesis (6) that a synonymous single nucleotide polymorphism (SNP) in exon 2 of Cita4 affects the expression of lCTLA-4 mRNA and may be the genetic basis of Idd5.1.

Materials and Methods

Congenic mouse strains and assessment of diabetes

The breeding and genotyping strategies for the development of the congenic mouse strains protected from T1D have been reviewed (7). All of the new strains developed for this study were derived from one of two diabetes-resistance strains, NOD.B10 Idd5/R8 (R8) and NOD.B10 Idd5/R444 (R444), described previously (1). New strains (Fig. 1) were produced by backcrossing R8 or R444 to NOD/Mrtac (Taconic Farms, Germantown, NY) and selecting for informative recombinants with subsequent fixing to homozygosity. NOD.B10 Idd5/R974 (R974) and NOD.B10 Idd5/R426 (R426) were derived from R8; NOD.B10 Idd5/R444 (R444a) was derived from R444; whereas NOD.B10 Idd5/R193 (R193) and NOD.B10 Idd5/R46 (R46) were derived from R444. NOD.B10 Idd5/R467 (R467) and NOD.B10 Idd5/R82 (R82) have been described previously (1). All congenic strains have been backcrossed to NOD 10–16 times, and nonconomochrome 1 markers examined in founder lines are NOD in origin (1). Elevated urinary glucose was detected using Diastix (Miles, Elkhart, IN). Animals were considered diabetic when urinary glucose was >500 mg/dl. Diabetic mice also exhibited polydipsia, polyuria, and weight loss. The frequency of diabetes was compared between strains with the Kaplan-Meier log-rank test using Prism (GraphPad, San Diego, CA) software. The R974, R46, R2, and R444 strains are available from Taconic Farms through the Emerging Models Program (lines 974, 2193, 1092, and 1094, respectively). The derivation and disease characterization of the diabetes-resistant NOD congenic strains noted in Fig. 6 as Idd5/1, Idd5/2, Idd9/10/18, and Idd5/10/18 have been described (5).

Identification of new microsatellite markers and genotyping

Mice were initially genotyped by PCR using primers to the previously published microsatellite markers. To map the recombination points more precisely within these strains, further microsatellites were identified and characterized. Sequences were obtained from the Ensembl mouse genome database (www.ensembl.org). Repeat sequences were then identified using RepeatMasker 4-Apr-2000 version (http://ftp.genome.washington.edu/RM/RepeatMasker.html) modified by D. Beare, Welcome Trust Sanger Institute, and L. Smink, Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, University of Cambridge. Sequences containing microsatellite repeats and ~500 bp of flanking sequence on either side were then extracted. PCR primers were designed for these microsatellite sequences using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi). All primers were ordered from Sigma-Genosys (The Woodlands, TX). Forward primers labeled at the 5’ end with the fluorescent dye HEX. PCR were optimized using B10 and NOD DNA templates. Markers that were polymorphic between these two strains were used in further genotyping. All PCR were performed using AmpliTaq Gold with buffer II (Applied Biosystems, Foster City, CA). When PCR product sizes were not distinguishable between B10 and NOD samples by 4% agarose gel electrophoresis, genotyping was performed using an ABI Prism 3100 genetic analyzer (Applied Biosystems). Primer sequences for novel microsatellite markers used in this study are shown in Table I.

### Table I. Microsatellite markers used to define Idd5.2

<table>
<thead>
<tr>
<th>Marker</th>
<th>Size (bp)</th>
<th>Allele Size (bp) (B10/NOD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>D1pt23_8</td>
<td>179/175</td>
<td></td>
</tr>
<tr>
<td>D1pt252131_A</td>
<td>207/219</td>
<td></td>
</tr>
<tr>
<td>D1pt252131_B</td>
<td>207/205</td>
<td></td>
</tr>
</tbody>
</table>

### Construction of a bacterial artificial chromosome (BAC) contig across the Idd5.1 interval.

Markers were identified from between D1Mit479 and D1Mit300 according to radiation hybrid-mapping data from the MGI database (www.informatics.jax.org) and the Whitehead Institute. These markers were used to screen the RPCI-23 (B6) library and the DIL NOD BAC library for positive clones. Fingerprint data obtained by the Wellcome Trust Sanger Institute were also used to ascertain which clones overlapped with other clones in the region. A BAC contig was constructed across the originally defined Idd5.1 interval for the B6 strain, and across portions of this same interval for NOD. Shotgun BAC sequencing was performed at the Wellcome Trust Sanger Institute.

### Gene identification.

The repeat masked BAC contig sequence was subjected to blast analysis against vertebrate mRNA, dbEST, and exofish blast databases. Expressed sequence tag hits were used to identify genes within the Idd5.1 region. Mouse Ensembl and the orthologous region in human Ensembl confirmed the gene content of Idd5.1. The gene content of Idd5.2 was obtained from the shotgun sequence available through mouse Ensembl. The orthologous human region in human Ensembl was also examined to confirm gene order.

### Delayed-type hypersensitivity assay.

Mice were immunized i.v. with 1 × 10^9 SRBC (TCS Biosciences, Buckingham, U.K.), and 4 days later were challenged with 5 × 10^8 SRBC in one hind footpad. At various times, the challenged foot was measured with calipers (Rabone Chesterman, Birmingham, U.K.). Twenty-four hours after challenge, the foot was removed and frozen immediately in liquid nitrogen. Control mice were untreated. Feet were pulverized while frozen in liquid nitrogen, homogenized, and RNA extracted in TRIzol (Invitrogen, San Diego, CA). Poly(A) RNA was prepared using a GeneHelix mRNA MiniPrep kit (Sigma-Aldrich, St. Louis, MO) and quantitated with an Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). PCRs were performed after reverse transcription of RNA (RT-PCR) using Superscript II (Invitrogen). Quantitative PCRs (qPCRs) were performed using primers and probes purchased from Applied Biosystems. Real-time PCR product detection and analysis were performed on the TaqMan 7700 (Applied Biosystems). Results are normalized to GAPDH or mouse β2-microglobulin (6). Primer and probe sequences for IFN-γ cDNA (8) and flCTLA-4, lCTLA-4, and sCTLA-4 cDNAs (6) have been described previously. Data between groups were analyzed for significance using an unpaired t test.

### Icos splice variant.

RT-PCR amplification of ICOS isoforms was performed using the 5’ primer GCACCTGAGGAGAAGACTG and 3’ primer AGTCCATGCTGTTTCCCTCTGTG. As with the expected full-length isoform, other faint products were noted when the products were analyzed on an agarose gel. To test the hypothesis that one of these variants may be analogous to either the ligand-independent or soluble forms of ICOS, RT-PCR was performed using a forward primer (CTCACCAGACCAAGGGAGAAG) in exon 2 and a reverse primer (GGATCCGATATTTCATAAATATGC, probe TGACCCACCTCCTTTTCAAATCAAACTTGAATGATAAACTT) over the boundary of exons 2 and 4, a forward primer in the 5’ untranslated region (TGACACCACATCAACCTCCA) and reverse primers over the boundary of exons 1 and 3 (CAGAGCTGGGATCTGTTAAGAAAGT) and the boundary of exons 1 and 4 (GGATCCTGATTTCTGTTAAGAACTG). No RT-PCR products were seen when using the primers over the exon 1/3 or 1/4 boundaries. However, a product of the expected size was seen when using the primers in exon 2 and over the exon 2/4 boundary. This product was then sequenced and found to have the normal splice site at the end of exon 2, no exon 3, which has the transmembrane domain, then the normal beginning of exon 4. This skipping of exon 3 changes the reading frame in exon 4, resulting in a premature stop codon (see Table II). Full-length and soluble ICOS primers and probes used for qPCR are forward TACTCTCTGACCCTGTCCCAT, reverse CACGAGACGCTGGATTACATA, probe TGACCCACATCCTTCTCTGTACCAAGAA, and forward TACTCTCTGACCCTGTCCCAT, reverse GGA TCCGATTTCTCATAAAATCTAG, probe TGACACCTCCTTCTTCAAGAA, respectively.

Downloaded from http://www.jimmunol.org/ by guest on July 1, 2017
Activation of T cells in vivo and in vitro

T cells were activated in vivo by injecting 5 μg of anti-CD3 (clone 145-2C11; American Type Culture Collection, Manassas, VA) and 2 μg of anti-CD28 (clone 37.51; BD Pharmingen, San Diego, CA) i.v. Six hours following injection, T cells were purified from isolated splenocytes by depletion of B cells, macrophages, and class II-positive cells. In some experiments, purified T cells were further separated into CD45RB high cells (naive) and CD45RB low cells (previously activated) with anti-CD45RB (clone 16A; BD Pharmingen) and anti-rat magnetic beads from Qiagen (Valencia, CA). In vitro stimulation of T cells was performed using 1 μg/ml anti-CD3 and 1 μg/ml anti-CD28. Poly(A) RNA purification and qPCR assays were performed, as above. IL-2 qPCR primers and probe were obtained from Applied Biosystems. Data between groups were analyzed for significance using an unpaired t test.

Results

Refinement of the proximal boundary of the Idd5.1 interval

Idd5.1 was initially defined using a panel of chromosome 1 congenic strains, including R8, R2, R467, and R444 (Fig. 1), as a 5.1-cM B10-derived resistance interval containing the candidate genes Casp8, Cflar (FLIP), Cd28, and Icld4 (1). We now report the development of the R974 strain (Fig. 1), which exhibits the same level of protection from T1D as the R444 strain (p = 0.15; Fig. 2A). Thus, the R974 and R444 strains have resistance alleles at both Idd5.1 and Idd5.2. Delineation of the centromeric recombination point in R974 required the harvesting of additional NOD/B10 polymeric microsatellites between and flanking the Cd28 and Icld4 genes (Fig. 3). This search was greatly facilitated by the availability of both B6 and NOD genomic sequence in this region (see below). The centromeric boundary of Idd5.1 is now defined as between D1Mit249 and D228distal3 (Fig. 3), thus eliminating Cd28, as well as Casp8 and Cflar (FLIP) (Fig. 1) as candidates for Idd5.1. These findings are consistent with those of Lamhamedi-Cherradi et al. (2), in which Casp8 and Cflar were excluded from the Idd5.1 region. Importantly, Lamhamedi-Cherradi et al. based their Idd5.1 mapping on NOD/B6 congenic strains rather than NOD/B10 congenic strains as used in the current study; thus, the similar findings strongly suggest that the highly related B6 and B10 strains share a T1D resistance allele at Idd5.1. This has important implications because the major public sequencing effort of the mouse genome has made use of B6 DNA.

Refinement of the distal boundary of the Idd5.1 interval: discovery of a cold spot of recombination within Idd5.1

The distal boundary of Idd5.1 was first defined by the R2 strain (Fig. 1), which has a recombination breakpoint between D1Mit303 and D1Mit300 (1). We now report the development of a recombinant congenic strain, R46, which is protected from T1D as compared with the NOD strain (p = 0.00006), but not to the extent seen in R444 (p = 0.0002, R46 vs R444) (Fig. 2). Consistent with the protection seen previously with the R467 recombinant strain (Fig. 1) that was shown to have a resistant allele at Idd5.1, but not at Idd5.2 (1), the R46 recombinant strain has the degree of protection conferred by the Idd5.1 region alone (Fig. 2). The R46 strain defines the distal boundary of Idd5.1 within the Nrp2 (neuropilin-2) gene (Fig. 3) showing that the Idd5.1 locus is up to 2.1 Mb in length. The R426 and R193 strains demonstrated the same protection from disease as R46 (p = 0.61 and p = 0.13 vs R46, respectively), consistent with the presence of only the B10 allele of Idd5.1 in both of these new recombinant strains.

As the recombination breakpoints near the Idd5.1 boundaries were analyzed, we found that three recombination events had occurred in 2194 meioses between markers D1Mit249 (~76 kb centromeric of Cfl4) and D1Mit303 (Fig. 3). Because D1Mit303 is ~2 Mb distal of D1Mit249, 22 recombinant mice would have been expected rather than the three observed. In addition, of the six mice with samples available for testing (nine recombinants total) that had recombinated between D1Mit303 and D1Mit300 (a distance of ~1 Mb), all six recombined in a 280-kb region between AL645534GA and D1Mit300. This nonlinear recombination is not atypical (9), but does present practical difficulties when it is desirable to construct a congenic strain carrying a specific recombination point that is relatively close to, but not at, the hot spot of recombination.

Gene content of the Idd5.1 region: assessment of functional candidacy

Analysis of the Idd5.1 gene content from the B6 BAC-based genomic sequence (Fig. 3) indicates that the 2.1-Mb region contains only four genes, Cfl4, Icos, Als2cr19, and Nrp2. Whereas Cfl4 and Icos are clearly strong functional candidates for Idd5.1 due to their important functions in the immune system (10, 11), Als2cr19 (12) and Nrp2 (13) are less likely candidates for influencing autoimmune diabetes because they have no recognized function in either β cell development or the immune system. Human NRP2 is a receptor for both semaphorins, which regulate neuronal guidance, and vascular endothelial growth factor, an angiogenic molecule. Disruption of the mouse NRP2 gene impairs yolk sac development and embryonic angiogenesis. It is noted, however, that NRP1 has a central role in T cell activation (14), and it is possible that NRP2 has a yet undiscovered role in the immune system. Als2cr19 is a partitioning defective-like protein. Partitioning defective proteins were first discovered in Caenorhabditis elegans and are essential for asymmetric cell division and polarized growth. Als2cr19 has 23 exons extending over 1 Mb and Nrp2 has 17 exons over 112 kb. The R46 distal recombination point is between exons 7 and 17 of Nrp2. Thus, the small number of genes in the Idd5.1 region is accounted for, in part, by the existence of one, very large gene (Als2cr19) extending over nearly one-half of the 2.1-Mb interval. Finally, from the currently known functions of the four genes in the interval, and results from genetic analysis of their

<table>
<thead>
<tr>
<th>Table II. Protein alignment of Icos and putative soluble splice variant</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ICOS</strong></td>
</tr>
<tr>
<td>Soluble ICOS</td>
</tr>
<tr>
<td>Soluble ICOS</td>
</tr>
<tr>
<td>Icos</td>
</tr>
<tr>
<td>Soluble Icos</td>
</tr>
<tr>
<td>Icos</td>
</tr>
<tr>
<td>Soluble Icos</td>
</tr>
<tr>
<td>Icos</td>
</tr>
<tr>
<td>Bold = transmembrane domain, underline = signal peptide, italics = different OFR.</td>
</tr>
</tbody>
</table>
orthologues in human T1D (6), Cita4 and Icos remain the favored candidates by this functional criterion.

B6-NOD sequence comparison of genes in the Idd5.1 region

We have obtained NOD-derived BAC-based sequence from portions of the Idd5.1 region, including the interval containing Cita4 and Icos as well as one exon and portions of two introns of Als2cr19 (Fig. 3). Table III details the polymorphisms seen in the regions in which sequences can be directly compared. Numerous SNPs are observed in intronic and 3’ and 5’ regulatory regions of Cita4 and Icos and indicate that both genes are within a region that is ancestrally distinct between NOD and B6 mice (15, 16). Although many SNPs in these regions were observed, no variants were present in either gene within any of the 5’ or 3’ splice sites or branch points. There are two exonic SNPs, one each in Cita4 and Icos. The exonic SNP present in Icos causes a nonconservative
Because the vast majority of Ctla4 splice forms of were examined (18, 19). In addition, we also searched for novel groups, but whose physiologic function is poorly understood (6), CTLA-4 and sCTLA-4, a form of CTLA-4 observed by many congenic strains of mice. mRNAs encoding full-length ICOS and Ctla4 and differed by only 2 nt between NOD and B6. Idd5.1 expression differences in T cells from NOD and NOD.B10 and putative regulatory regions, we tested for steady state mRNA genotyped were polymorphic and only 1 (Table III). This possibility is substantiated by the observation that this particular region as compared with the nearby interval containing or activated T cells (Fig. 4). However, as shown in the accompanying study (20), genotype-related expression differences in ICOS protein on the cell surface of T cells have been observed. Increased expression of mRNA encoding all three CTLA-4 isoforms in CD45RB low T cells Because a genotype-dependent production of liCTLA-4 was observed previously, it is important to determine whether this is due to differential expression depending on the maturation state of the cells examined. It has been demonstrated previously that liCTLA-4 expression at the protein level is lower in CD45RB high (naive) compared with CD45RB low (memory) T cells (21, 22). We, therefore, isolated CD45RB high and CD45RB low T cells from NOD and R444 spleens and tested them for expression of liCTLA-4, liCTLA-4, and sCTLA-4 (Fig. 5). For all three isoforms, expression was 16- to 64-fold (4 – 6 qPCR cycles) higher in memory CD45RB low CD4+ T cells as compared with naive CD45RB high CD4+ T cells. A similar, albeit smaller (2 – 4 qPCR cycles more in memory cells), difference was observed for ICOS and sICOS mRNA (data not shown). Additionally, the liCTLA-4 was differentially expressed in both subsets of cells, although the quantitative detection of liCTLA-4 mRNA was problematic in naive NOD T cells due to its very low abundance in this cell type. Genetic control of liCTLA-4 expression Strains of mice having either the NOD or B6 allele at Ctla4 were assessed for liCTLA-4 mRNA expression. Spleen cells were tested following in vivo activation with anti-CD3/CD28. To account for individual differences in the numbers of CD45RB low memory T cells, the cell subset in which the majority of CTLA-4 expression is found (Fig. 5), the liCTLA-4 mRNA level in each mouse was normalized against its own CD45RB low (memory) T cells as compared with naive CD45RB high CD4+ T cells. Results from congenic strains recombining near the CTLA-4 locus were consistent with the hypothesis that an altered splicing preference is determined by the exon 2 SNP: R974 mice (B10 allele at Idd5.1) had high levels of liCTLA-4 mRNA, whereas R2 mice (NOD allele at Idd5.1) had low levels (Fig. 6). Finally, we tested several other congenic strains not having the Idd5.1 B10 allele, including some having combinations of Id4 loci that confer profound protection from T1D such as congenic strains with disease-resistant alleles at Idd3/10/18 and Idd9.1/9.2/9.3 (5), and observed that they produced low levels of liCTLA-4 mRNA. In contrast, when B10 Idd5.1 is present together with the B10 Idd5.2 and B6 Idd3 loci in Idd3/5 mice, a strain in which only 1% of the mice develop T1D and <10% develop more than minimal insulitis or

**FIGURE 2.** Frequency of diabetes in female Idd5 congenic mice.

Although the Ctla4 SNP in exon 2 does not cause an amino acid change, we have recently reported that this silent B6/NOD SNP at residue 77 in exon 2, the exon encoding the ligand binding domain of CTLA-4, is a critical nucleotide in the creation of a novel splice variant (6). This isoform is referred to as liCTLA-4 (6). The variation at residue 77 results in a 4-fold higher expression of liCTLA-4 mRNA in B10 Idd5.1 T cells as compared with NOD T cells (6) (also see Figs. 4, 5, 6, and 7). The B6/NOD SNP in exon 2 of Ctla4 alters the sequence of a putative exonic splicing silencer motif, thereby changing the efficiency of producing mRNA for liCTLA-4, but not affecting the production of liCTLA-4 or sCTLA-4 (6).

Potential splice variants of ICOS were assessed by RT-PCR, as described in Materials and Methods. Unlike CTLA-4, in which four mRNA species are present (6), only a single splice variant encoding a putative soluble form of ICOS, sICOS, could be documented (see Materials and Methods). In addition, no reproducible, genotype-related expression differences of either full-length ICOS (liICOS) or sICOS mRNA have been observed in unstimulated or activated T cells (Fig. 4). However, as shown in the accompanying study (20), genotype-related expression differences in ICOS protein on the cell surface of T cells have been observed.

Ctla4 and Icos splice variants

Because the vast majority of Ctla4 and Icos SNPs are in introns and putative regulatory regions, we tested for steady state mRNA expression differences in T cells from NOD and NOD.B10 Idd5.1 congenic strains of mice. mRNAs encoding full-length ICOS and CTLA-4 and sCTLA-4, a form of CTLA-4 observed by many groups, but whose physiologic function is poorly understood (6), were examined (18, 19). In addition, we also searched for novel splice forms of Ctla4 and Icos in normal and activated T cells.

**a**

<table>
<thead>
<tr>
<th>Age (days)</th>
<th>% Nonobdiecant</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>90</td>
</tr>
<tr>
<td>60</td>
<td>80</td>
</tr>
<tr>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>120</td>
<td>60</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
</tr>
<tr>
<td>180</td>
<td>40</td>
</tr>
<tr>
<td>210</td>
<td>30</td>
</tr>
<tr>
<td>240</td>
<td>20</td>
</tr>
</tbody>
</table>

**b**

<table>
<thead>
<tr>
<th>Age (days)</th>
<th>% Nonobdiecant</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>100</td>
</tr>
<tr>
<td>30</td>
<td>90</td>
</tr>
<tr>
<td>60</td>
<td>80</td>
</tr>
<tr>
<td>90</td>
<td>70</td>
</tr>
<tr>
<td>120</td>
<td>60</td>
</tr>
<tr>
<td>150</td>
<td>50</td>
</tr>
<tr>
<td>180</td>
<td>40</td>
</tr>
<tr>
<td>210</td>
<td>30</td>
</tr>
<tr>
<td>240</td>
<td>20</td>
</tr>
</tbody>
</table>

**Ctla4 and Icos splice variants**

Although the Ctla4 SNP in exon 2 does not cause an amino acid change, we have recently reported that this silent B6/NOD SNP at residue 77 in exon 2, the exon encoding the ligand binding domain of CTLA-4, is a critical nucleotide in the creation of a novel splice variant (6). This isoform is referred to as liCTLA-4 (6). The variation at residue 77 results in a 4-fold higher expression of liCTLA-4 mRNA in B10 Idd5.1 T cells as compared with NOD T cells (6) (also see Figs. 4, 5, 6, and 7). The B6/NOD SNP in exon 2 of Ctla4 alters the sequence of a putative exonic splicing silencer motif, thereby changing the efficiency of producing mRNA for liCTLA-4, but not affecting the production of liCTLA-4 or sCTLA-4 (6).

Potential splice variants of ICOS were assessed by RT-PCR, as described in Materials and Methods. Unlike CTLA-4, in which four mRNA species are present (6), only a single splice variant encoding a putative soluble form of ICOS, sICOS, could be documented (see Materials and Methods). In addition, no reproducible, genotype-related expression differences of either full-length ICOS (liICOS) or sICOS mRNA have been observed in unstimulated or activated T cells (Fig. 4). However, as shown in the accompanying study (20), genotype-related expression differences in ICOS protein on the cell surface of T cells have been observed.

Increased expression of mRNA encoding all three CTLA-4 isoforms in CD45RB low T cells

Because a genotype-dependent production of liCTLA-4 was observed previously, it is important to determine whether this is due to differential expression depending on the maturation state of the cells examined. It has been demonstrated previously that liCTLA-4 expression at the protein level is lower in CD45RB high (naive) compared with CD45RB low (memory) T cells (21, 22). We, therefore, isolated CD45RB high and CD45RB low T cells from NOD and R444 spleens and tested them for expression of liCTLA-4, liCTLA-4, and sCTLA-4 (Fig. 5). For all three isoforms, expression was 16- to 64-fold (4 – 6 qPCR cycles) higher in memory CD45RB low CD4+ T cells as compared with naive CD45RB high CD4+ T cells. A similar, albeit smaller (2 – 4 qPCR cycles more in memory cells), difference was observed for ICOS and sICOS mRNA (data not shown). Additionally, the liCTLA-4 was differentially expressed in both subsets of cells, although the quantitative detection of liCTLA-4 mRNA was problematic in naive NOD T cells due to its very low abundance in this cell type.

Genetic control of liCTLA-4 expression

Strains of mice having either the NOD or B6 allele at Ctla4 were assessed for liCTLA-4 mRNA expression. Spleen cells were tested following in vivo activation with anti-CD3/CD28. To account for individual differences in the numbers of CD45RB low memory T cells, the cell subset in which the majority of CTLA-4 expression is found (Fig. 5), the liCTLA-4 mRNA level in each mouse was normalized against its own liCTLA-4 mRNA level. Results from congenic strains recombining near the CTLA-4 locus were consistent with the hypothesis that an altered splicing preference is determined by the exon 2 SNP: R974 mice (B10 allele at Idd5.1) had high levels of liCTLA-4 mRNA, whereas R2 mice (NOD allele at Idd5.1) had low levels (Fig. 6). Finally, we tested several other congenic strains not having the Idd5.1 B10 allele, including some having combinations of Id4 loci that confer profound protection from T1D such as congenic strains with disease-resistant alleles at Idd3/10/18 and Idd9.1/9.2/9.3 (5), and observed that they produced low levels of liCTLA-4 mRNA. In contrast, when B10 Idd5.1 is present together with the B10 Idd5.2 and B6 Idd3 loci in Idd3/5 mice, a strain in which only 1% of the mice develop T1D and <10% develop more than minimal insulitis or
FIGURE 3. Fine mapping of Idd5.1. The proximal and distal boundaries defined by the R974 and R46 strains, respectively, define the four genes encoded within the Idd5.1 interval: Ctla4, Icos, Als2cr19, and Nrp2. These genes are also represented in the human orthologous region 2q33.3-q34. Dark gray clones with the prefix RP23 are B6 clones and all have complete sequence. Light gray clones with the prefix DN are NOD clones and all have complete sequence. Microsatellite markers that are not polymorphic between NOD and B10 are shown to the left of the B6 contig.
anti-insulin autoantibodies (5), the higher level of liCTLA-4 mRNA expected from the B10 allele at Idd5.1 was observed (Fig. 6). B6 mice congenic for the NOD Idd5.1 allele produced low levels of liCTLA-4 mRNA compared with B6 or B10 mice having their normal Idd5.1 B6 or B10 allele (Fig. 6), an expected result if differential expression of liCTLA-4 is due to a cis-DNA variation intrinsic to the allele. The activated splenocytes described in Fig. 6 were also examined for sCTLA-4, ICOS, and sICOS mRNA levels by qPCR, and no genotype-related expression differences were observed (data not shown).

Expression of liCTLA-4 in an in vivo Ag-specific response

To examine the expression of liCTLA-4 mRNA under more physiological conditions, NOD and R444 mice were primed with SRBC and later challenged in the footpad with SRBC. Twenty-four hours after challenge, mRNA extracted from the footpads demonstrated a reproducible differential level of expression between NOD and R444 of liCTLA-4, but not of sCTLA-4, sCTLA-4, ICOS, IFN-γ, IL-2, and IL-10 (Fig. 7 and data not shown). By qPCR, sICOS mRNA was not detected in these samples. There was no difference in the degree of the delayed-type hypersensitivity response as measured by footpad swelling between the two groups of mice in either experiment (data not shown).

Refinement of the proximal and distal boundaries of the Idd5.2 interval

Idd5.2 was previously defined to be in a 5.1-cM region distal to Idd5.1, and is, in combination with Idd5.1, critical for the full protection mediated by the Idd5 region (1). New congenic strains were developed to further define the Idd5.2 interval: R444s, R193, and R426 (Figs. 1, 2, and 8). The frequency of diabetes does not differ between R444 and R444s (Fig. 2 B, p = 0.66). However, R193 is less protected from disease than R444 (Fig. 2 A, p = 0.001), indicating that R193 has lost the protective allele at Idd5.2. Supporting this hypothesis is the observation that disease protection in R193 is indistinguishable from R426 (p = 0.92) and R46 (p = 0.56). All three strains, R193, R426, and R46, are protected as compared with the NOD parental strain (p < 0.0006 for each comparison), and all, therefore, have the Idd5.1 allele. The level of disease protection observed in the R193, R426, and R46 congenic strains is very similar to that reported for the R1, R39, and R67 Idd5.1 NOD.B6 strains (2), consistent with B6 and B10 sharing a T1D-resistant allele in Idd5.1.

Interrogation of the Idd5.2 region between the proximal and distal boundaries as defined by the R193 and R444s strains (Figs. 1 and 8) was performed using the B6 sequence available from mouse EnsEMBL. In stark contrast to the gene-poor Idd5.1 region detailed above, an extremely gene-rich Idd5.2 region was found: at least 45 genes are present within the 1.52-Mb interval (Fig. 8). Nramp1, known to be functionally polymorphic between B6/B10 (expressing a nonfunctional allotype) and NOD (expresses the functional protein), remained within the Idd5.2 interval.

The gene content in Idd5.1 and Idd5.2 is conserved with human 2q33.3–q34 and 2q35, respectively

A comparative analysis of draft sequence from the B6 BAC tile path with the orthologous human sequence revealed 4 and 45 orthologous genes within the 2.1-Mb Idd5.1 and 1.52-Mb Idd5.2 intervals, respectively (Figs. 3 and 8). The 4 genes in Idd5.1 were examined for the extent of identity at the amino acid level between the mouse and human sequences. As has been described previously, amino acid identity for CTLA-4 and ICOS with their human counterparts is 76 and 72% (23, 24), respectively, whereas the

<table>
<thead>
<tr>
<th>Region proximal to Cita4</th>
<th>Size in B6 (bp)</th>
<th>SNPs</th>
<th>Coding Changes</th>
<th>in/dels</th>
<th>SNPs/kb</th>
</tr>
</thead>
<tbody>
<tr>
<td>5’ UTR</td>
<td>1,049</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Exon 1</td>
<td>109</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Intron 1</td>
<td>3,140</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Exon 2</td>
<td>348</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Intron 2</td>
<td>448</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Exon 3</td>
<td>110</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Intron 3</td>
<td>1,266</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Exon 4</td>
<td>105</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>3’ UTR</td>
<td>4,050</td>
<td>5</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Region between Cita4 and Icos</td>
<td>57,810</td>
<td>153</td>
<td>47</td>
<td>2.6</td>
<td></td>
</tr>
</tbody>
</table>

Table III. Genetic variation between NOD and B6 in the Idd5.1 region

* The number of SNPs and indels between NOD and B6 is shown for the portions of the Idd5.1 interval having complete sequence from both strains. Of the indels proximal to Cita4, one is an L1 repeat (7400 bp) and one is a mouse endogenous retroviral sequence (6441 bp); both are only present in NOD. Of the indels distal to Cita4, one is an L1 repeat (7025 bp), also only present in NOD.

* UTR, untranslated region.
ALS2CR19 and NRP2 genes have 85 and 94% identity with the equivalent human genes, respectively.

Discussion

The challenge of defining genes underlying complex genetic traits, including those causing disease, remains formidable. In defining genes in a model of disease, such as the NOD mouse that develops T1D, key tools include the development of congenic strains that are protected from disease because of a genetic region derived from a disease-resistant strain. Ultimately, the best that can be obtained from such strains is the narrowing of a region containing the disease gene to 0.5–3 Mb, depending on the frequency of recombination in the region under study. At this point, a systematic gene identification strategy is to obtain a complete sequence of the defined region, determine the gene content, and, from comparative sequencing of the same region in the other parental strain, identify all SNPs. Unfortunately, these steps remain time consuming and expensive and only represent the beginning of the search for the causative SNP in a polygenic disease. It will be rare for a susceptibility locus to be identified from sequence information alone; one case would be if the susceptible and resistant strains are identical by descent in the defined genetic interval and there is a functional de novo mutation that has occurred in one of the two strains being compared. Unfortunately, as is the case for Idd5.1 and Idd5.2, most genes that influence a complex disease will be one of many linked genes within a distinct ancestral haplotype (15, 16) that differs between the susceptible and resistant parental strains. Therefore, many DNA variants are present (>20 SNPs per 10 kb), making the identification of the causal SNP or SNPs much more difficult.

FIGURE 4. Correlation of genotype with levels of liCTLA-4 mRNA. Purified T cells were stimulated with anti-CD3 (1 μg/ml) and anti-CD28 (1 μg/ml) in vitro and then assessed at various times for mRNA levels by reverse transcription and qPCR (TaqMan). Values are normalized to β2-microglobulin mRNA with ΔCT, representing the difference in CT (cycle threshold) values between the experimental and control genes. Error bars represent the SD of the mean of duplicate samples. The dashed and solid lines represent NOD and R444 (Idd5.1/5.2) T cells, respectively. In the case of CTLA-4 and ICOS, triangles and squares represent the full-length and soluble isoforms, respectively. Circles represent the liCTLA-4 isoform.

FIGURE 5. CD45RB low cells are enriched for fCTLA-4, liCTLA-4, and sCTLA-4 mRNA expression. ΔCT (cycle threshold) values were determined as detailed in Fig. 4.

FIGURE 6. The Idd5.1 region controls expression levels of liCTLA-4 in congenic strains. ΔCT (cycle threshold) values for fCTLA-4 and liCTLA-4 were determined as in Fig. 4. The levels of liCTLA-4 were normalized, as described in the text, by subtracting the fCTLA-4 ΔCT from the liCTLA-4 ΔCT. The number of mice tested per group ranged from 3 to 10. The mean and SE for each group of mice are shown. *, These strains are described in detail in Robles et al. (5). The B6.NOD congenic strains were developed by Wakeland and colleagues (27). The B10.H2b7 strain has been described previously (28).
The disease-causative Idd intervals described in this study, Idd5.1 and Idd5.2, represent two contrasting consequences of using the gene identification strategy outlined above. The larger 2.1-Mb Idd5.1 region contains only four genes, two of which are strong functional candidates, whereas the physically smaller 1.65-Mb Idd5.2 region contains 45 genes. Because both of these regions have already been delineated using thousands of meiotic events to select specific recombinant mice, it is clear that even at the end of the congenic strain development phase, the investigator is left with regions containing many genes containing hundreds or thousands of SNPs, nearly all of which will be irrelevant for the disease phenotype, even though some of the variants may have functional effects. Although we propose that Idd5.1 is Ctla4 and the molecular basis of the phenotypic change is caused by the exon 2 SNP that determines the production of liCTLA-4, it remains possible that a SNP (or SNPs) in the ICOS gene is responsible for the disease phenotype, or that variants at Ctla4 and Icos together are causative. Although no genotype-dependent expression or splicing difference was noted for ICOS mRNA under the conditions tested in this study, it may well be that under different stimulation conditions or if expression is assessed in specific cell types (see Ref. 20), ICOS will be differentially expressed in NOD vs Idd5.1 mice due to one or more SNPs in Icos. It is also possible that the SNP present in the leader sequence of the ICOS molecule could cause a change in protein expression that would not be associated with an alteration in the steady state level of Icos transcription. An additional complexity of the Idd5.1 locus is that a difference in ICOS expression could be a downstream consequence of a genetic variant in Ctla4. Riley et al. (25) found that CTLA-4 engagement reduced expression of ICOS on the cell surface. Thus, the higher expression of liCTLA-4 in activated T cells having the
B10 allele at Idd5.1 would increase negative signaling, thereby decreasing the up-regulation of ICOS on the cell surface. This hypothesis could explain some of the observations reported in the accompanying study by Greve et al. (20), in which higher cell surface expression of ICOS on activated T cells is correlated with the NOD haplotype at Idd5.1, a genotype that confers lower liCTLA-4 expression.

To prove that a variant gene within a disease-modifying genetic interval is the disease-causing allele requires that the SNP change the course of the disease. This is true even for a gene with alleles having a compelling functional difference. To determine whether the altered splicing preference of CTLA-4 is in fact ultimately responsible for changing disease susceptibility, we are developing an Idd5.1 congenic mouse strain in which the donor segment has been derived from an inbred strain in which a historical recombinemation event has occurred between Ctxa4 and Icos, thereby breaking apart the extended haplotypes present in B10 and NOD mice. If data from this novel congenic strain support a causative role for residue 77, the next experiment would be to knockin the B10 exon 2 Ctxa4 SNP into the NOD genome to determine whether a single SNP is sufficient to not only alter liCTLA-4 expression, but also ICOS expression and disease frequency. To this end, we have developed (NOD × 129) embryonic stem cells to target the NOD genome (26).

Another argument for the likelihood that Ctxa4 rather than lcos functions as Idd5.1 is the observation that the Ctxa4 knockout mouse displays a much more severe phenotype than the lcos knockout mouse (10, 11). This suggests that quantitative variation of Ctxa4 would be more likely to confer a selective pressure than quantitative variation of lcos. The association of human autoimmune diseases with CTLA4, but not lcos, is consistent with this hypothesis (6).

References