Association of the T-cell regulatory gene *CTLA4* with susceptibility to autoimmune disease

Hironori Ueda¹, Joanna M. M. Howson¹*, Laura Esposito¹*, Joanne Heward²*, Hywel Snook¹, Giselle Chamberlain¹, Daniel B. Rainbow¹, Kara M. D. Hunter¹, Annabel N. Smith¹, Gianfranco Di Genova¹†, Mathias H. Herr¹†, Ingrid Dahlman¹†, Felicity Payne¹, Deborah Smyth¹, Christopher Lowe¹, Rebecca C. J. Twells¹, Sarah Howlett¹, Barry Healy¹, Sarah Nutland¹, Helen E. Rance¹, Vin Everett¹, Luc J. Smink¹, Alex C. Lam¹, Heather J. Cordell¹, Neil M. Walker¹, Cristina Bordin¹†, John Hulme¹, Costantino Motzo³, Francesco Cucca³, J. Fred Hess⁴, Michael L. Metzker⁴†, Jane Rogers⁵, Simon Gregory⁵, Amit Allahabadia²†, Ratnasingam Nithiyananthan², Eva Tuomilehto-Wolf⁶, Jaakko Tuomilehto^{6,7}, Polly Bingley⁸, Kathleen M. Gillespie⁸, Dag E. Undlien⁹†, Kjersti S. Rønningen¹⁰, Cristian Guja¹¹, Constantin Ionescu-Tirgovişte¹¹, David A. Savage¹², A. Peter Maxwell¹³, Dennis J. Carson¹⁴, Chris C. Patterson¹⁵, Jayne A. Franklyn², David G. Clayton¹, Laurence B. Peterson¹⁶, Linda S. Wicker¹, John A. Todd¹ & Stephen C. L. Gough²

*These authors contributed equally to this work

Genes and mechanisms involved in common complex diseases, such as the autoimmune disorders that affect approximately 5% of the population, remain obscure. Here we identify polymorphisms of the cytotoxic T lymphocyte antigen 4 gene (*CTLA4*)—which encodes a vital negative regulatory molecule of the immune system—as candidates for primary determinants of risk of the common autoimmune disorders Graves' disease, autoimmune hypothyroidism and type 1 diabetes. In humans, disease susceptibility was mapped to a non-coding 6.1 kb 3' region of *CTLA4*, the common allelic variation of which was correlated with lower messenger RNA levels of the soluble alternative splice form of *CTLA4*. In the mouse model of type 1 diabetes, susceptibility was also associated with variation in CTLA-4 gene splicing with reduced production of a splice form encoding a molecule lacking the CD80/CD86 ligand-binding domain. Genetic mapping of variants conferring a small disease risk can identify pathways in complex disorders, as exemplified by our discovery of inherited, quantitative alterations of *CTLA4* contributing to autoimmune tissue destruction.

An understanding of the mechanisms underlying common diseases, which are caused by the combined effects of many genetic and environmental factors, will contribute to the development of future preventative and disease-modifying therapies. DNA variants of the human genome, which alter physiological pathways involved in susceptibility to multifactorial disease, are by definition primary causal risk factors. Progress, however, has been limited. Relatively few disease gene variants have been identified¹, probably due to the small risks associated with individual variants and an underestimation of how large studies have to be to obtain reliable results^{1,2}. The common autoimmune diseases are presumed to result from a failure in the mechanisms of the immune system that establish and maintain non-responsiveness or tolerance to self³. Our study implicates common allelic variation in the expression levels of alternative splice forms of a key negative regulator of the T lymphocyte immune response, CTLA-4 (refs 4, 5), as a primary determinant of susceptibility to autoimmune disease.

Graves' disease affects 0.5% of western populations, and results

¹ Juvenile Diabetes Research Foundation/Wellcome Trust Diabetes and Inflammation Laboratory, Cambridge Institute for Medical Research, University of Cambridge, Wellcome Trust/MRC Building, Cambridge, CB2 2XY, UK; ² Division of Medical Sciences, University of Birmingham, Birningham, B15 2TT, UK; ³ Dipartimento di Scienze Biomediche e Biotecnologie, Universita di Cagliari, Cagliari, Italy; ⁴ Merck Research Laboratories, West Point, Pennsylvania 19486, USA; ⁵ Wellcome Trust Sanger Institute, Hinxton, Cambridge, CB10 ISA, UK; ⁶ Diabetes and Genetic Epidemiology Unit, National Public Health Institute, University of Helsinki, Helsinki, Finland; ⁷ Department of Public Health, University of Bristol, Bristol, B510 SNG, UK; ⁹ Institute of Immunology, Rikshospitalet University Hospital and Institute of Medical Genetics, Ulleval University Hospital, University of Oslo, Oslo, Norway; ¹⁰ Laboratory of Molecular Epidemiology, Division of Epidemiology, Norwegian Institute of Public Health, Oslo, Norway; ¹¹ Clinic of Diabetes, Institute of Diabetes, Nutrition and Metabolic Diseases ¹N. Paulescu; Bucharest, Romania; ¹² Department of Medical Genetics, Queen's University Belfast, Belfast City Hospital, Belfast, BT9 7AB, UK; ¹³ Regional Nephrology Unit, Belfast City Hospital, Belfast, BT9 7AB, UK; ¹³ Regional Nephrology Unit, Belfast City Hospital, Belfast, Belfast from the presence of autoantibodies to the thyrotropin receptor, leading to over-activity of the thyroid gland. Autoimmune hypothyroidism (AIH; 1% prevalence) is caused by a T-lymphocyte-mediated inflammatory destruction of thyrocytes within the thyroid gland, leading to thyroid hormone deficiency. Type 1 diabetes (T1D; 0.4% prevalence) is caused by insulin deficiency, associated with a T-cell inflammation of the insulin-producing cells of the pancreas.

Because Graves' disease, T1D, AIH and other autoimmune diseases such as rheumatoid arthritis commonly cluster in the same families, it is likely that they share some genetic background. Indeed, risk of Graves' disease, AIH and T1D has been associated with the same region on chromosome 2q33, which contains the T-lymphocyte regulatory genes *CD28*, *CTLA4* and inducible co-stimulator (*ICOS*)^{6–9}. Moreover, T1D susceptibility in the spontaneous mouse model of the disease, the nonobese diabetic (NOD) strain, has been mapped to the equivalent region of the mouse genome, on chromosome 1, containing the orthologues of

UK; ¹⁴ Department of Child Health, Queen's University Belfast, Belfast, BT12 6BJ, UK; ¹⁵ Department of Epidemiology and Public Health, Queen's University Belfast, Belfast, BT12 6BJ, UK; ¹⁶ Department of Pharmacology, Merck Research Laboratories, Rahway, New Jersey 07065, USA.

† Present addresses: HIT Group, Tenovus Research Laboratories, Southampton General Hospital, Tremona Road, Southampton, SO16 6YD, UK (G.D.G.); Department of Paediatrics, Otto-Heubner-Centrum, Charité, Humboldt-University of Berlin, Germany (M.H.H.); Karolinska Institute, Department of Medicine, Huddinge University Hospital, Stockholm, Sweden (L.D.); Hutchison/MRC Research Centre, Hills Road, Cambridge, CB2 2XZ, UK (C.B.); Department of Molecular & Human Genetics, Baylor Genome Sequencing Center, Baylor College of Medicine, N1409, One Baylor Plaza, Houston, Texas 77030, USA (M.L.M.); Division of Clinical Sciences, University of Sheffield, Northern General Hospital, Sheffield, S5 7AU, UK (A.A.); Institute of Medical Genetics, University of Oslo, N-0315 Oslo, Norway (D.E.U.).

Table 1 Association of CTLA4 SNPs in 3,671 type 1 diabetic families					
SNP/allele	RR	95% confidence intervals	<i>P</i> value	Transmission to affected offspring (%)	P _{TDT}
CTAF343/T	1.18	1.08–1.27	0.00009	54.0	0.00009
rs1863800/C	1.16	1.09-1.25	0.00001	53.8	0.00001
MH30/G	1.16	1.09-1.24	8.0×10^{-6}	53.7	8.3×10^{-6}
+49/G	1.09	1.02-1.16	0.013	52.1	0.013
CT60/G	1.14	1.07-1.21	0.00006	53.3	0.00006
JO31/G	1.12	1.05-1.20	0.00090	52.9	0.00091
JO30/G	1.18	1.10-1.27	5.7×10^{-6}	54.2	5.8×10^{-6}
JO27_1/T	1.17	1.09-1.25	0.00003	53.8	0.00003
CTIC154_1/C	1.01	0.89–1.14	0.91	50.2	0.91

Relative risk (RR) was calculated from the coefficients of the regression equation⁴⁵, and *P* values are for the null hypothesis of no association of the marker. For comparison, the *P*_{TDT} (where TDT is the transmission/disequilibrium test) probability values⁴⁶ are provided along with percentage transmission from heterozygous parents to affected offspring.

CD28, *CTLA4* and *ICOS*^{10–12}. However, the causal variants in neither humans nor mice have been localized or identified. All three genes (*CD28*, *CTLA4* and *ICOS*) are strong candidates for the source of the disease risk associated with this chromosome region and its mouse orthologue, because they have critical roles in T-cell activation and in autoimmune disease^{4,13–16}.

Human genetics

Initially, the sequence of most of the 300-kilobase (kb) *CD28–CTLA4–ICOS* region was unknown. Hence, the possibility of the presence of other candidate genes could not be excluded, and the polymorphism content was also unknown. We sequenced 330 kb of human DNA containing *CD28*, *CTLA4* and *ICOS*, and where possible re-sequenced segments from 32 individuals to identify the polymorphisms (Supplementary Information A). We also sequenced the orthologous mouse *Ctla4–Icos* region to help establish the gene content and to aid our search for the mouse T1D gene (*Idd5.1*; refs 10–12) that exists in this region of mouse chromosome 1 (see below).

CD28, *CTLA4* and *ICOS* were the only functional genes in the region (Supplementary Information A). No polymorphisms that altered amino acids were found in any of the exons of the three genes, except for the known +49G>A single nucleotide polymorphism (SNP) (Thr17Ala) in exon 1 of *CTLA4* (Supplementary Information A). A total of 108 SNPs and the *CTLA4* (AT)_n-3' untranslated region (UTR) marker were genotyped in 384 cases of Graves' disease and 652 controls to determine their allele frequen-



Figure 1 Association of the *CD28–CTLA4–ICOS* region with Graves' disease. The log *P* values (*y* axis) of the differences in allele frequencies of 108 SNPs and the *CTLA4* (AT)_n-3' UTR microsatellite marker between cases of Graves' disease (n = 384) and controls (n = 652) were plotted against physical distance (*x* axis). Vertical lines indicate the two boundaries of the LD blocks as shown in Supplementary Fig. B1. Lines are drawn between points only for the sake of clarity, as the disease association between markers is not linear.

cies (Fig. 1; see also Supplementary Information A and B). If a marker allele is more or less frequent in cases compared to controls it can be categorized as being associated with disease. This is due to association, or linkage disequilibrium (LD), of the marker allele with an allele of the causal variant, or to the marker itself being a causal variant. Significant association was observed in the region (Fig. 1). CTLA4 and the 5' part of ICOS were contained within a 100-kb subregion, or block¹⁷, of strong intermarker LD flanked by two localized regions of low LD, one between CD28 and CTLA4, the other in the first intron of ICOS (Fig. 1; see also Supplementary Fig. B1). The causal variant was most probably located within the 100-kb region of LD containing CTLA4 and the 5' region of ICOS, with the association reaching a maximum in a region immediately 3' of the CTLA-4 structural gene ($P = 1.6 \times 10^{-6}$; odds ratio of 1.56). Outside this LD block, disease association declined rapidly. None of the five SNPs tested in CD28 or the 14 SNPs in ICOS were associated with disease (Fig. 1; see also Supplementary Figs B2-B4 and Table B7).

Within the CTLA4 LD block the pattern of disease association had three main peaks (from left to right in Fig. 1) with 23 out of 78 (29.5%) SNPs showing no association (P > 0.05). The first peak $(P \approx 10^{-5})$ was 20–35 kb 5' of the CTLA4 ATG codon (including the marker MH30). The second ($P \approx 10^{-6}$) was 0.2–6.3 kb 3' of the previously reported end of the CTLA4 transcript and included SNPs CT60, JO31, JO30 and JO27_1. The third $(P \approx 10^{-4})$ was 49–55 kb 3' of the CTLA4 transcript end and 8-13 kb 5' of the ICOS ATG codon, and comprised seven markers, including CTBC217_1. Regression modelling¹⁸ indicated that the third peak was an unlikely location for the causal variant because once the most-associated marker from the middle peak (CT60) was included in the model, the most-associated marker from the third did not improve the model (Supplementary Figs B2–B4). Furthermore, none of the previously known polymorphisms, -319C>T (CT44), +49G>A, +1,822T>C (CT55)⁷ and (AT)_n-3' UTR, could explain the association of the region with Graves' disease (Supplementary Table B2).

To distinguish between the remaining two association peaks, the seven SNPs associated most closely with Graves' disease along with the +49G>A SNP were genotyped in a larger data set (a further 288 Graves' disease cases and 192 controls; Supplementary Table B3). Regression analyses showed that the first peak was less likely to harbour the causal variant than the second peak (Supplementary Information B). The most-associated marker remained CT60 (odds ratio of 1.51, 95% confidence interval of 1.31-1.75; $P = 2.72 \times 10^{-8}$) but it could not be distinguished in the regression analyses from its neighbouring SNPs: JO31, JO30 and JO27_1. CT60 is a common SNP with 63.4% of 1,316 Graves' disease patient chromosomes and 53.2% of 1,646 control chromosomes having the susceptible G allele. Compared with the protective CT60 A/A genotype, the A/G and G/G genotypes had odds ratios of 1.59 (1.19-2.13) and 2.32 (1.71-3.15), respectively. In controls the A/A, A/G and G/G genotypes had frequencies of 22.8%, 48.0% and 29.2%, and in cases, 13.7%, 45.7% and 40.6%, respectively. Conversely, relative to the disease-predisposing G/G genotype, A/G and

A/A had odds ratios of 0.68 (0.54–0.86) and 0.43 (0.32–0.59), respectively (Supplementary Information B). Regression analysis of haplotypes provided no evidence for haplotype-specific effects; there are two very common CTLA4 haplotypes, one susceptible for Graves' disease, the other, protective (Supplementary Table B4). This result along with the single-point-association analysis (Fig. 1 and Supplementary Table B1), a Graves' disease family-based study (Supplementary Table B1) and the two-locus regression results (Supplementary Figs B2–B4) suggested that a single common variant, one of the SNPs CT60, JO31, JO30 or JO27_1 in the 6.1-kb region 3' of *CTLA4*, is responsible for the disease association of the region.

The CT60 SNP was also associated with AIH (228 cases and 844 controls) to the same degree as Graves' disease (odds ratio of 1.45 (1.17–1.80); P = 0.0005). However, the effect was much weaker in T1D. In 3,671 T1D families we analysed nine markers: CTAF343, rs1863800 and MH30 in the first Graves' disease peak; CT60, JO31, JO30 and JO27_1 in the second peak, along with +49G>A, and CTIC154 1, a SNP in the 3' UTR of ICOS. Alleles that were positively disease-associated in Graves' disease were also associated with T1D, with the most-associated markers being MH30, JO30 and CT60; however, the effect was small (odds ratios \approx 1.15; Table 1). Regression analysis failed to exclude the 5' CTLA4 markers MH30 and rs1863800. However, +49G>A was rejected as the causal SNP, as were CTAF343, JO31 and the ICOS SNP CTIC154_1 (Supplementary Information B). Increased transmission of CTLA4 alleles to siblings affected with T1D (53-54%, Table 1) was mirrored by under-transmission to unaffected siblings: for example, the G allele of CT60 was under-transmitted to unaffected offspring at 47.1% (989:1,110; P = 0.009), ruling out transmission ratio distortion and genotyping error as explanations for the association in T1D. The T1D data do raise the possibility that additional, more rare causal variants may exist in the region (Supplementary Table B5). Furthermore, neither the T1D nor Graves' disease data exclude the possibility that the common causal variant lies outside the 3' 6.1-kb region; for example, 5' of CTLA4. Nevertheless, taken together, the data suggest the presence of a common Graves' disease, T1D (designated previously IDDM12) and AIH locus in the 6.1-kb 3' region of CTLA4.

CTLA-4 gene expression and genotype

As the most disease-associated SNPs did not affect the CTLA-4 amino acid sequence, and the 6.1-kb region contained most of the CTLA-4 3' UTR (Supplementary Fig. A4), we examined the steadystate mRNA levels of the two known isoforms of CTLA-4: a fulllength isoform (flCTLA-4) encoded by exon 1 (leader peptide), exon 2 (ligand-binding domain), exon 3 (transmembrane domain) and exon 4 (cytoplasmic tail), and a soluble form (sCTLA-4)¹⁹⁻²¹, which lacks exon 3 (Fig. 2a). We used a combination of polymerase chain reaction (PCR) primers specific for the flCTLA-4 and sCTLA-4 mRNAs and allele-specific transcript quantification (ASTQ)²² to measure the relative amounts of the sCTLA-4 and flCTLA-4 mRNAs transcribed from the susceptible and protective haplotypes. Unstimulated CD4 T cells from four individuals were examined: three doubly heterozygous +49A/G-CT60 A/G individuals, all having one dose each of the protective (+49A-CT60A) and susceptible (+49G-CT60G) haplotype across the 6.1-kb region, and a +49G/A heterozygous/CT60 G/G homozygous individual, who carried a haplotype with a recombination event between +49 and CT60, corresponding to two doses of the susceptible 6.1-kb region haplotype. The sCTLA-4 mRNA level from the disease-protective haplotype (A allele at +49) was higher than that of the sCTLA-4 level transcribed from the susceptible haplotype in the samples from all three +49A/G-CT60 A/G heterozygous individuals, giving ratios of 2.0:1, 2.1:1 and 2.5:1 (Fig. 2b). These ratios were significantly different to the ratios of flCTLA-4 transcript production according to haplotype (1.3:1, 1.7:1, 1.4:1 and 1.5:1; *P* = 0.007; note that these





ratios are equivalent to a ratio of 1:1 owing to the size difference in the ASTQ PCR bands; Fig. 2b). As would be expected if allelic variation of the 6.1-kb region was the cause of the difference in splicing efficiency, the individual who was homozygous-susceptible for the region showed a sCTLA-4 ratio of 1.3.

To extend these findings to individuals with all three genotypes at CT60, we developed a real-time PCR assay that specifically detected the mRNA abundance of the sCTLA-4 isoform. To account for the fact that variable numbers of CTLA-4-positive cells would be present in the peripheral blood of different individuals, flCTLA-4 was used to normalize the sCTLA-4 values. We found that the ratio of sCTLA-4 to flCTLA-4 mRNA splice forms in unstimulated CD4 T cells was 50% lower in CT60 G/G-positive disease-susceptible



Mouse CD45 ACCACTGAATCCACACCCCCAAGCATCTCTGAA Figure 3 Sequence, splicing and expression of Ctla4 in NOD mice and congenic strains. a, Schematic of the four splice forms of Ctla4. b, Autoradiogram of a Southern blot illustrating the splicing of CTLA-4 mRNA. Primers are in exons 1 and 4 of Ctla4, producing bands of 624, 514, 276 and 166 bp corresponding to fICTLA-4, sCTLA-4, liCTLA-4 and the exons 1/4 isoform, respectively. c, Agarose gel of the products generated by RT-PCR using primers specific to each CTLA-4 isoform mRNA. d, Expression of fICTLA-4, sCTLA-4 and liCTLA-4 mRNA by real-time quantitative PCR from purified splenic T cells before and 6 h after anti-CD3 (5 µg per mouse) plus anti-CD28 (2 µg per mouse) antibody stimulation in vivo. Open bars, NOD T cells; filled bars, diabetes-resistant NOD Idd5 T cells. Error bars represent 1 s.d. e, Sequence alignment of CD45 and CTLA4 flanking the SNP at base 77 of exon 4 of CD45 and exon 2 of CTLA4. Residues in bold are conserved between all sequences, and residues in bold and italicized are conserved between all except those in which splicing is not observed (human and rat CTLA4). Position 77 is underlined.

individuals compared with A/A protected individuals (Fig. 2c; P = 0.002). Consistent with the semi-dominant mode of inheritance of the disease locus at *CTLA4*, the heterozygous genotype expressed more sCTLA-4 mRNA than the homozygous susceptible genotype (P = 0.02). Our results suggest that the 6.1-kb region determines the efficiency of the splicing and production of sCTLA-4, with the CT60G disease-susceptibility haplotype producing less sCTLA-4 transcript than the resistant CT60A haplotype.

Mouse CTLA-4

Recent genetic mapping has defined the location of the *Idd5.1* T1D disease locus to a 1.1 cM (2.1 megabase) region of chromosome 1 containing the human chromosome 2q33 orthologues *Ctla4*, *Icos*, *Als2cr19* and part of *Nrp2* (ref. 11) (L.S.W., manuscript in preparation). *Cd28* was excluded as the *Idd5.1* causal locus. In the *Idd5.1* region the C57BL/10 (B10) haplotype is protective for T1D whereas the NOD haplotype confers susceptibility to disease. Comparison of genomic sequence between the NOD and B10 strains of the two functional candidate genes *Ctla4* and *Icos* revealed only two exonic variants: one translationally silent change in exon 2 of *Ctla4*, and one coding variant causing a non-conservative amino acid change in the putative leader sequence encoded in exon 1 of *Icos*. No variants were found in either gene at any of the splice sites or branch points. Nevertheless, splicing of *Ctla4* was investigated in detail (Fig. 3).

We discovered a new splice form of mouse CTLA-4 with a fully intact open reading frame that encodes a transmembrane isoform of CTLA-4 with the CD80/CD86-binding domain missing, referred to here as the ligand-independent isoform (liCTLA-4; Fig. 3a–c). We also observed the transcripts for the previously described full-length and soluble isoforms (Fig. 3a–c). The liCTLA-4 transcript is formed by skipping exon 2, and has a fourfold increase in expression in both activated and resting spleen cells from the diabetes-resistant strain as compared with the NOD strain (Fig. 3d).

A literature survey and DNA sequence comparison indicated that the mouse CTLA-4 exon 2 SNP is the probable genetic cause of the variation in liCTLA-4 mRNA levels (Fig. 3e). In the human CD45 gene, substitution of the wild-type C allele with a mutant G allele by a SNP at base 77 of exon 4 reduces the function of an exonic splicing silencer (ESS) embedded in the exon 4 DNA sequence, resulting in alteration of the relative levels of functionally distinct isoforms of this important T-cell differentiation molecule²³. Comparison of the human CD45 gene exon 4 sequence and CTLA-4 gene exon 2 sequence, and their mouse and rat orthologues, identified the conserved wild-type sequence CNNNTGNATNNNCACCA/ CNCANNCANCNNTGA (see Fig. 3e). In marked similarity to the human CD45 exon 4 SNP, the mouse Ctla4 exon 2 SNP has an A at position 77 in the diabetes-resistant strains B10 and B6, whereas the NOD strain carries the mutant G allele. The G allele would lead to an increase in the inclusion of exon 2, forming flCTLA-4 mRNA, and therefore a reduction of the exon-2-negative transcript that encodes liCTLA-4. Eight of the conserved 18 nucleotides in this 31-base pair (bp) region differed in human CTLA4, which probably accounts for our inability to detect the liCTLA-4 transcript in human T cells (Fig. 2a). The rat Ctla4 exon 2 sequence differed only by two nucleotides among the conserved 18 compared to the mouse sequence, at positions 75 and 77: both are T nucleotides versus C and A or C in ESS wild-type sequences (Fig. 3e). Rat spleen does not express detectable levels of the liCTLA-4 isoform (data not shown), supporting a critical role for position 77 for the splicing efficiency of liCTLA-4. Mutational studies of an ESS in the late pre-mRNA of bovine papillomavirus type 1 and also sequence comparisons with the few other known ESSs have demonstrated the importance of this central CCCCC sequence²⁴.

Discussion

Our results highlight the power of positional genetics in the identification of key physiological DNA variants, and they provide evidence that certain common autoimmune diseases are caused in part by inherited changes in CTLA-4 expression that presumably increase T-cell self-reactivity. The importance of CTLA-4 in the counter-regulation of CD28 T-cell antigen receptor (TCR) activation of T cells is well established^{4,13-16}; however, the molecular basis of the interactions between CTLA-4 and CD28 is uncertain. CTLA-4, by means of its ligand-binding domain, may compete with CD28 for the receptors (CD80 and CD86) that they share. CTLA-4 might also deliver a negative signal to each T cell by inhibiting the TCR signalling complex, an activity that would not necessarily depend on ligand binding¹⁶. Furthermore, and non-exclusively, CTLA-4 is expressed constitutively by T cells that suppress the effector T-cell response; referred to as T-regulatory cells (T_R)^{4,25,26}. Depletion of T_R cells causes autoimmune disease, and the function of some T_R cells is dependent on CTLA-4²⁷⁻²⁹. T_R cells may exert part of their function by CTLA-4 binding to CD80/CD86 on immature dendritic cells, leading to increases in tryptophan catabolism and downregulation of T effector activity^{14,15,26}. Non-T cells including placental fibroblasts also express CTLA-4 (ref. 30). Thus, the effects of alterations of the splicing of CTLA4 might be wide ranging. Our results provide insights that focus attention on the sCTLA-4 isoform and its little-studied role in the maintenance of immune tolerance and autoimmunity. sCTLA-4 is known to be translated, secreted and present in human serum¹⁹⁻²¹. It can bind the CD80/86 molecules, and recombinant sCTLA-4 inhibits T-cell proliferation *in vitro*²¹. The reduction in the level of sCTLA-4 mRNA produced by the disease-susceptible haplotype of CTLA4 could lead to reduced blocking of CD80/CD86, causing increased activation through CD28, or to less stimulation of CD80/CD86 on immature dendritic cells, resulting in less tryptophan catabolism. The published correlation between in vitro T-cell activation and CTLA-4 genotype³¹⁻³³ could be explained by variation in levels of sCTLA-4.

In the NOD mouse we propose that *Idd5.1* is a SNP in exon 2 of Ctla4 that controls the splicing efficiency of a liCTLA-4 form. We can be certain that this RNA is translated into a functional protein from two other experiments. Recent studies of the human CD28 gene^{34,35} demonstrated that a CD28 isoform (CD28i) equivalent to liCTLA-4 exists. As found for liCTLA-4, CD28i contains the signal peptide, the transmembrane domain and the cytoplasmic tail, and, owing to an alternative splicing mechanism different from the one that we have observed in this study for CTLA-4, is missing the ligand-binding domain, encoded by CD28 exon 2. CD28i, which exists as both a homodimer and as a noncovalent heterodimer with CD28, functions as a co-stimulatory signal amplifier in human T cells³⁵. Second, there is direct evidence that liCTLA-4 protein causes a negative signal in CTLA-4-negative T cells transfected with liCTLA-4 complementary DNA (L. Vijayakrishnan, V. J. Kuchroo, L.S.W. and L.B.P., unpublished data). liCTLA-4 would presumably act by inhibiting signalling in T-cell activation and expansion. The existence of this functional form and our genetic data provide support for a ligand-independent mode of action of CTLA-4.

It is possible that the common functional variants of the CTLA-4 gene have been the subject of evolutionary selection in host resistance to infectious disease. The autoimmune-disease susceptibility alleles could provide enhanced cellular immune responses for particular viral or intracellular infections, whereas the autoimmune-resistance alleles may augment humoral immune responses against a different set of pathogens, such as parasitic worms. Immunoglobulin- ε is a major feature of the anti-worm response and also of atopic illness such as allergy and asthma. The *CTLA4* autoimmune-protective haplotype has been associated recently with susceptibility to atopy³⁶. It has been proposed that the significant increases in the incidence of T1D³⁷ and atopic

illness³⁸ in children over the last 40 years in affluent western populations might be due, in part, to advances in healthcare and in the decrease in frequency of worm and other infections. Taken together with our results we suggest that common allelic variation of *CTLA4* and other genes encoding regulators of the immune system may underpin the molecular basis for such gene–environment interactions. Our results point to a primary role in auto-immune disease for the alternative splice forms of CTLA-4—the soluble isoform in humans and the newly discovered ligand-independent form in mice. These observations highlight the usefulness of the positional genetics approach in discovering insights into disease mechanisms in a complex, multifactorial disorder, even for a gene as well studied as *CTLA4*.

Methods

(See Supplementary Information for detailed Methods)

Subjects

White, UK-born Graves' disease cases (n = 672), Graves' disease family members (n = 210) and AIH cases (n = 228) were recruited from thyroid clinics in Birmingham (Queen Elizabeth and Heartlands Hospitals), Exeter (Royal Devon and Exeter Hospital) and Bournemouth (Royal Bournemouth Hospital), as described previously8. Ethnically matched control subjects (n = 462) with no history of autoimmune disease were recruited at various sites including the Blood Transfusion Service, Birmingham Heartlands Hospital and the Queen Elizabeth Hospital, Birmingham, and 382 control subjects (for which there are cell lines available at the European Collection Animal Cell Cultures) originated from blood donors from Oxford and Birmingham Blood Transfusion Services. T1D families were white European or of white European descent, with two parents and at least one affected child: 472 Diabetes UK Warren 1 multiplex from the United Kingdom³⁹; 322 United States multiplex from the Human Biological Data Interchange40; 80 Yorkshire simplex from the United Kingdom; 32 southwest simplex from the United Kingdom; 216 Bristol simplex from the United Kingdom; 263 Belfast multiplex/simplex⁴¹ from the United Kingdom; 360 Norwegian simplex; 423 Romanian simplex; and 1,503 Finnish multiplex/simplex⁴². The Cambridge Local Research Ethics Committee approved experiments involving human subjects and informed consent was obtained from all subjects. The mouse research was approved by the Home Office and Cambridge University Biomedical Support Services

SNP identification and genotyping

SNPs were identified in PCR products from 32 individuals (Supplementary Information A). SNPs were mainly genotyped using the Invader assay (Third Wave Technologies) in two different formats, uniplex⁴³ and biplex, which was validated for the present study (Supplementary Information A) and by others⁴⁴. The accuracy of genotyping data, its management and quality control are detailed in Supplementary Information A.

Statistical analyses

All statistical analysis, unless otherwise stated, was performed using STATA (http://www.stata.com). Particular use was made of the 'GenAssoc' routine, available from http://www-gene.cimr.cam.ac.uk/clayton/software/stata/ (see also Supplementary Information B).

Gene expression

Peripheral blood mononuclear cells (PBMCs) were isolated from heparinized blood of healthy volunteers by density gradient centrifugation using Lympholyte H (Cedarlane Laboratories). CD4 T cells were purified by negative selection using magnetic beads (Dynal Biotech ASA). As determined by flow cytometry (FACS Scan and CELLQuest Software), cells were routinely >92-94% CD4⁺ and <1.5% CD8⁺. ASTQ²² was carried out on fICTLA-4 and sCTLA-4 cDNA products that had been amplified by PCR, with $[\alpha\text{-}^{32}P]dCTP$ incorporated in the final PCR step (Fig. 2b). cDNA amplification was performed using the primers: flCTLA-4, 5'-ACTTCCTGAAGACCTGAACACCG-3', 5'-GCACGGTTCTGGATCAATTACA-3'; sCTLA-4, 5'-CTGAAGACCTGAACACCGCT-3', $5^{\,\prime}$ -GGCTTCTTTTCTTTAGCAATTACATA-3 $^{\prime}$. Cycling conditions were 95 °C for 25 s, 55 °C (sCTLA-4) or 60 °C (flCTLA-4) for 25 s and 72 °C for 25 s. Final cycle labelling with $[\alpha\text{-}^{32}P]dCTP$ was performed after 30 cycles of PCR. ASTQ PCR products were digested with BbvI and resolved by 8% polyacrylamide gel electrophoresis. Gels were dried and bands quantified after film exposure using the Quantity One software (BIO-RAD). The bands sometimes resolved into doublets, all of which were included in the quantitation. For the flCTLA-4/sCTLA-4 ratio, a real-time quantitative (Taqman) PCR assay was used employing the ß2 microglobulin gene as a control transcript and RNA purified from CD4 T cells (Fig. 2c). Relative levels were calculated from the ratio of the sCTLA-4 cycle threshold (CT) to flCLTA-4 CT values both of which were normalized to β_2 -microglobulin mRNA using the formula, $2^{(\Delta CT(sCTLA-4)-\Delta CT(flCTLA - 4))}$, with ΔCT representing the difference in CT values between the target and control transcripts. Twelve of the volunteers in Fig. 2c had been previously tested in three additional experiments involving a smaller number of subjects in each (n = 7, 4 and 5). Aside from the two previously untested donors present in the fourth experiment shown in Fig. 2c, all samples have been replicated once with an inter-assay replication rate of 92%. A new cDNA synthesis of RNA was performed for each assay. Across all experiments the observed 50%

increase in the sCTLA-4/flCTLA-4 ratio in disease-protected CT60 A/A-genotype-positive individuals compared with susceptible G/G subjects was reproducible. In Fig. 3, primers (Supplementary Table A7) were designed across unique exon boundaries for flCTLA-4 (exons 2/3; 78 bp), sCTLA-4 (exons 2/4; 72 bp) and liCTLA-4 (exons 1/3; 113 bp). All expression differences were evaluated using a two-tail Student's *t*-test distribution.

Received 23 December 2002; accepted 3 April 2003; doi:10.1038/nature01621. Published online 30 April 2003.

- Lohmueller, K. E., Pearce, C. L., Pike, M., Lander, E. S. & Hirschhorn, J. N. Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nature Genet.* 33, 177–182 (2003).
- Dahlman, I. et al. Parameters for reliable results in genetic association studies in common disease. Nature Genet. 30, 149–150 (2002).
- Lesage, S. & Goodnow, C. C. Organ-specific autoimmune disease: a deficiency of tolerogenic stimulation. J. Exp. Med. 194, F31–F36 (2001).
- Finger, E. & Bluestone, J. When ligand becomes receptor-tolerance via B7 signaling on DCs. Nature Immunol. 3, 1056–1057 (2002).
- Egen, J. G., Kuhns, M. S. & Allison, J. P. CTLA-4: new insights into its biological function and use in tumour immunotherapy. *Nature Immunol.* 3, 611–618 (2002).
- Nistico, L. et al. The CTLA-4 gene region of chromosome 2q33 is linked to, and associated with, type 1 diabetes. Hum. Mol. Gen. 5, 1075–1080 (1996).
- Marron, M. P. et al. Genetic and physical mapping of a type 1 diabetes susceptibility gene (IDDM12) to a 100-kb phagemid artificial chromosome clone containing D2S72-CTLA4-D2S105 on chromosome 2q33. Diabetes 49, 492–499 (2000).
- Allahabadia, A. et al. The MHC class II region, CTLA-4 gene and ophthalmopathy in patients with Graves' disease. Lancet 358, 984–985 (2001).
- Kristiansen, O. P., Larsen, Z. M. & Pociot, F. CTLA-4 in autoimmune diseases—a general susceptibility gene to autoimmunity? *Genes Immun.* 1, 170–184 (2000).
- Colucci, F., Bergman, M., Penha-Goncalves, C., Cilio, C. M. & Holmberg, D. Apoptosis resistance of nonobese diabetic peripheral lymphocytes linked to the *Idd5* diabetes susceptibility region. *Proc. Natl Acad. Sci. USA* 94, 8670–8674 (1997).
- Hill, N. J. et al. The NOD Idd5 locus controls insulitis and diabetes and overlaps the orthologous CTLA4/IDDM12 and NRAMP1 loci in humans. Diabetes 49, 1744–1747 (2000).
- Lamhamedi-Cherradi, S. E. et al. Further mapping of the ldd5.1 locus for autoimmune diabetes in NOD mice. Diabetes 50, 2874–2878 (2001).
- 13. Sharpe, A. H. & Freeman, G. J. The B7-CD28 superfamily. Nature Rev. Immunol. 2, 116–126 (2002).
- Grohmann, U. et al. CTLA-4-Ig regulates tryptophan catabolism in vivo. Nature Immunol. 3, 1097–1101 (2002).
- Lin, H. *et al.* Cytotoxic T lymphocyte antigen 4 (CTLA4) blockade accelerates the acute rejection of cardiac allografts in CD28-deficient mice: CTLA4 can function independently of CD28. *J. Exp. Med.* 188, 199–204 (1998).
- Chikuma, S., Imboden, J. & Bluestone, J. A. Negative regulation of T cell receptor-lipid raft interaction by cytotoxic T lymphocyte-associated antigen 4. J. Exp. Med. 197, 129–135 (2003).
- Gabriel, S. B. et al. The structure of haplotype blocks in the human genome. Science 296, 2225–2229 (2002).
- Cordell, H. J. & Clayton, D. G. A unified stepwise regression procedure for evaluating the relative effects of polymorphisms within a gene using case/control or family data: application to HLA in type 1 diabetes. *Am. J. Hum. Genet.* **70**, 124–141 (2002).
- Magistrelli, G. et al. A soluble form of CTLA-4 generated by alternative splicing is expressed by nonstimulated human T cells. Eur. J. Immunol. 29, 3596–3602 (1999).
- Oaks, M. K. & Hallett, K. M. Cutting edge: a soluble form of CTLA-4 in patients with autoimmune thyroid disease. J. Immunol. 164, 5015–5018 (2000).
- 21. Oaks, M. K. et al. A native soluble form of CTLA-4. Cell Immunol. 201, 144-153 (2000).
- Kaijzel, E. L. et al. Allele-specific quantification of tumour necrosis factor alpha (TNF) transcription and the role of promoter polymorphisms in rheumatoid arthritis patients and healthy individuals. *Genes Immun.* 2, 135–144 (2001).
- Lynch, K. W. & Weiss, A. A CD45 polymorphism associated with multiple sclerosis disrupts an exonic splicing silencer. J. Biol. Chem. 276, 24341–24347 (2001).
- Zheng, Z.-M., He, P.-J. & Baker, C. Function of a bovine papillomavirus type 1 exonic splicing suppressor requires a suboptimal upstream 3' splice site. J. Virol. 73, 29–36 (1999).
- Asano, M., Toda, M., Sakguchi, N. & Sakaguchi, S. Autoimmune disease as a consequence of developmental abnormality of a T cell population. J. Exp. Med. 184, 387–396 (1996).
- Manzotti, C. et al. Inhibition of human T cell proliferation by CTLA-4 utilizes CD80 and requires CD25⁺ regulatory cells. Eur. J. Immunol. 32, 2888–2896 (2002).

- Maloy, K. J. et al. CD4 + CD25 + T(R) cells suppress innate immune pathology through cytokinedependent mechanisms. J. Exp. Med. 197, 111–119 (2003).
- Sakaguchi, S. Control of immune responses by naturally arising CD4(+) regulatory T cells that express Toll-like receptors. J. Exp. Med. 197, 397–401 (2003).
- Pasare, C. & Medzhitov, R. Toll pathway-dependent blockade of CD4⁺CD25⁺ T cell-mediated suppression by dendritic cells. *Science* 299, 1033–1036 (2003).
- Kaufman, K. A. et al. The CTLA-4 gene is expressed in placental fibroblasts. Mol. Hum. Reprod. 5, 84–87 (1999).
- Huang, D., Giscombe, R., Zhou, Y., Pirskanen, R. & Lefvert, A. Dinucleotide repeat expansion in the CTLA-4 gene leads to T cell hyper-reactivity via the CD28 pathway in myasthenia gravis. *J. Neuroimmunol.* 105, 69–77 (2000).
- Kouki, T. et al. CTLA-4 gene polymorphism at position 49 in exon 1 reduces the inhibitory function of CTLA-4 and contributes to the pathogenesis of Graves' disease. J. Immunol. 165, 6606–6611 (2000).
- Maurer, M. et al. A polymorphism in the human cytotoxic T-lymphocyte antigen 4 (CTLA4) gene (exon 1 +49) alters T-cell activation. Immunogenetics 54, 1–8 (2002).
- Magistrelli, G. et al. Identification of three alternatively spliced variants of human CD28 mRNA. Biochem. Biophys. Res. Commun. 259, 34–37 (1999).
- Hanawa, H. et al. A novel costimulatory signalling in human T lymphocytes by a splice variant of CD28. Blood 99, 2138–2145 (2002).
- Howard, T. et al. Fine mapping of an IgE-controlling gene on chromosome 2q: analysis of CTLA4 and CD28. J. Allergy Clin. Immunol. 110, 743–751 (2002).
- Todd, J. A. A protective role of the environment in the development of type 1 diabetes? *Diabetic Med.* 8, 906–910 (1991).
- 38. Strachan, D. P. Hay fever, hygiene, and household size. Br. Med. J. 299, 1259-1260 (1989).
- Bain, S. C., Todd, J. A. & Barnett, A. H. The British Diabetic Association Warren Repository. Autoimmunity 7, 83–85 (1990).
- 40. Lernmark, A. et al. Family cell lines available for research. Am. J. Hum. Genet. 47, 1028–1030 (1990).
- Patterson, C. C., Carson, D. J. & Hadden, D. R. Epidemiology of childhood IDDM in Northern Ireland 1989–1994: low incidence in areas with highest population density and most household crowding. Northern Ireland Diabetes Study Group. *Diabetologia* 39, 1063–1069 (1996).
- Tuomilehto, J. et al. Epidemiology of childhood diabetes mellitus in Finland—background of a nationwide study of type 1 (insulin-dependent) diabetes mellitus. The Childhood Diabetes in Finland (DiMe) Study Group. Diabetologia 35, 70–76 (1992).
- Mein, C. A. et al. Evaluation of single nucleotide polymorphism typing with invader on PCR amplicons and its automation. Genome Res. 10, 330–343 (2000).
- Olivier, M. et al. High-throughput genotyping of single nucleotide polymorphisms using new biplex invader technology. Nucleic Acids Res. 30, e53 (2002).
- Clayton, D. in Handbook of Statistical Genetics (eds Balding, D., Bishop, M. & Cannings, C.) 519–540 (Wiley, Chichester, 2001).
- Spielman, R., McGinnis, R. & Ewens, W. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). Am. J. Hum. Genet. 52, 506–516 (1993).

Supplementary Information accompanies the paper on www.nature.com/nature.

Acknowledgements We thank the Juvenile Diabetes Research Foundation (JDRF), the Wellcome Trust, NovoNordisk, Novo Nordisk Foundation, the Academy of Finland, The Sigrid Juselius Foundation, Diabetes UK and the Medical Research Council for financial support. H.U. was a Wellcome Trust Travelling Fellow and R.N. is a West Midlands Regional Health Authority Sheldon Medical Research Fellow. The availability of NOD congenic mice through the Taconic Farms Emerging Models Program has been made possible and is supported by grants from the Merck Genome Research Institute, NIAID and the JDRF. We gratefully acknowledge the participation of all patients, controls and family members, including provision of samples from T1D families from the Human Biological Data Interchange and Diabetes UK repositories, and sample collections by The Norwegian Study Group for Childhood, preparation of manuscript materials by J. Brown and T. Thorn, and review of the manuscript by C. Rudd, C. Nathan, M. Bobrow, P. Lyons, R. Glynne, C. Goodnow, J. Trowsdale, N. Proudfoot, T. Merriman, N. Hastie and D. Sansom.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.A.T. (john.todd@cimr.cam.ac.uk) or L.S.W. (linda.wicker@cimr.cam.ac.uk).

articles