Supplementary Information A.

Genomics

The RPCI-1, 3, 4, and 5 human P1 artificial chromosome (PAC), the RPCI-11(Ref)[1] human, the RPCI-23(Ref)[2] mouse (B6) and the DIL-NOD bacterial artificial chromosome (BAC) libraries were obtained (Table A1). For all RPCI library information see http://www.chori.org/bacpac/. The DIL-NOD library was prepared from the NOD/MrkTac mouse at the Institute of Physical and Chemical Research (RIKEN Ibaraki, Japan) and is available for distribution from the JDRF/WT DIL. A number of probes were used to screen the libraries (data not shown). From the resulting positive clones six were sequenced (Table A1).

Table A1: Library screening and sequencing information.

Library	Species/strain	Probe	Sequenced clone name	Accession number
RPCI-1	Human	$(AT)_n$ 3'UTR	61E2	AF225900
RPCI-11	Human	$(AT)_n$ 3'UTR	278L16	AJ535716, AJ535717, AJ535718
RPCI-1	Human	CD28	219D7	AF225899
RPCI-23	Mouse B6	Cd28	157N11	AL646054
RPCI-23	Mouse B6	Ctla4	146J17	AL663047
DIL-NOD	Mouse NOD	Ctla4/Icos	257N2	AL596283

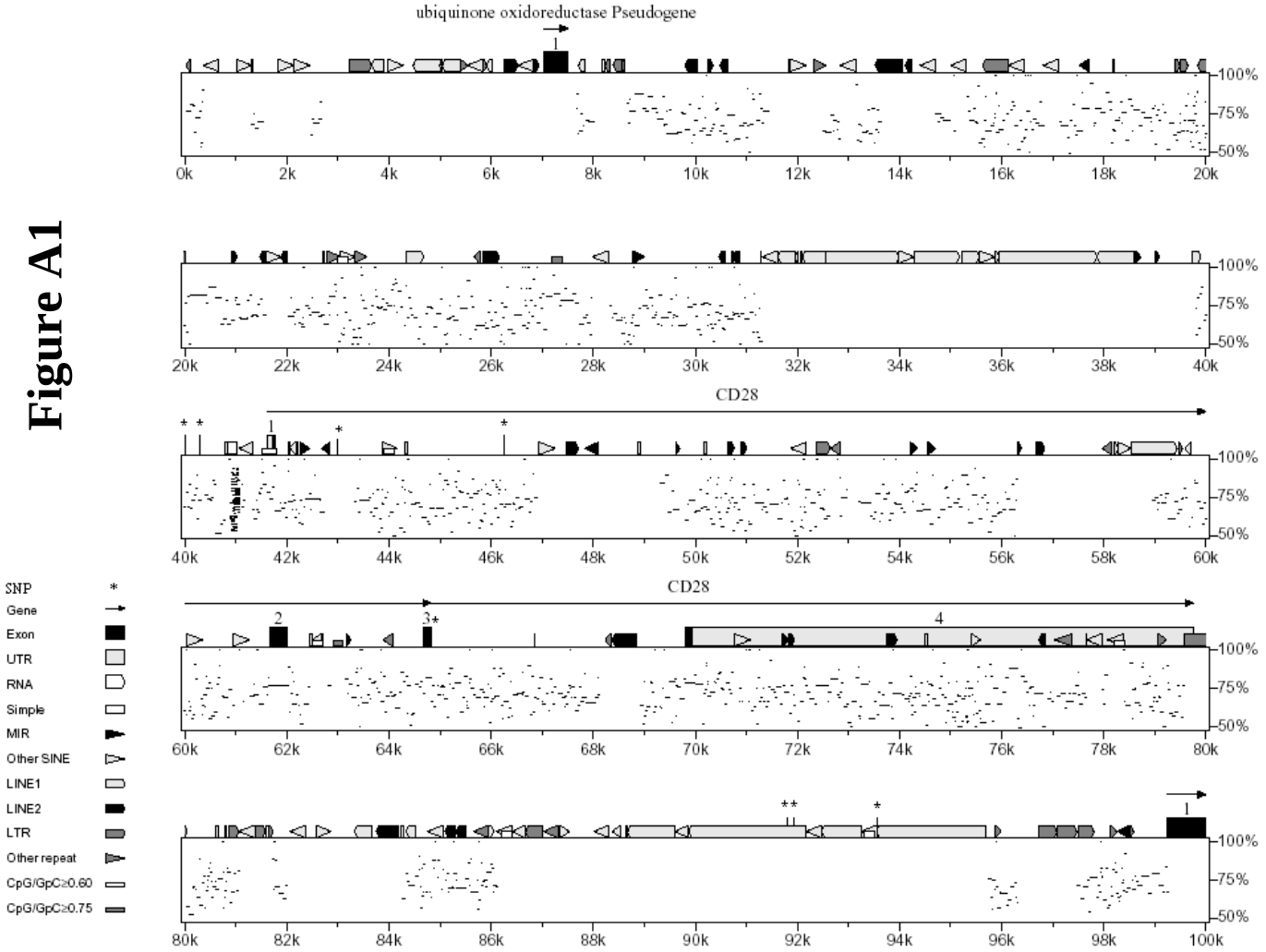
The human and mouse sequences were assembled and used in database searches. The homology information was integrated into T1Dace, our in-house

adaptation of ACeDB[3]with the sequence, repeat information (RepeatMasker A.F.A. Smit and P. Green, unpublished, http://www.genome.washington.edu/UWGC/analysistools/repeatmask.htm), automatically annotated genes from EnsEMBL (http://www.ensembl.org), primers used for SNP harvesting, the identified SNPs, and SNPs from the public domain (dbSNP, TSC SNPs and HGBASE). The mouse and human genomic sequences were compared by constructing percent-identity plots (PIP)[4] (Figure A1). A view of this region using Generic Genome Browser (Gbrowse) (http://www.gmod.org/ggb/) is provided (Figure. A2). Only three functional genes exist in the region, *CD28*, *CTLA4* and *ICOS*, and three pseudogenes (a sequence with similarity to the ubiquinone-oxidoreductase MLRQ subunit gene, keratin, and nucleolar phosphoprotein). A sequence alignment of mouse and human sequence around the CT60 SNP revealed 66.1% homology from the poly-A signal to CT60 (ClustalW, 1.74, http://bimas.dcrt.nih.gov/clustalw/clustalw.htmp/), (Figure A3).

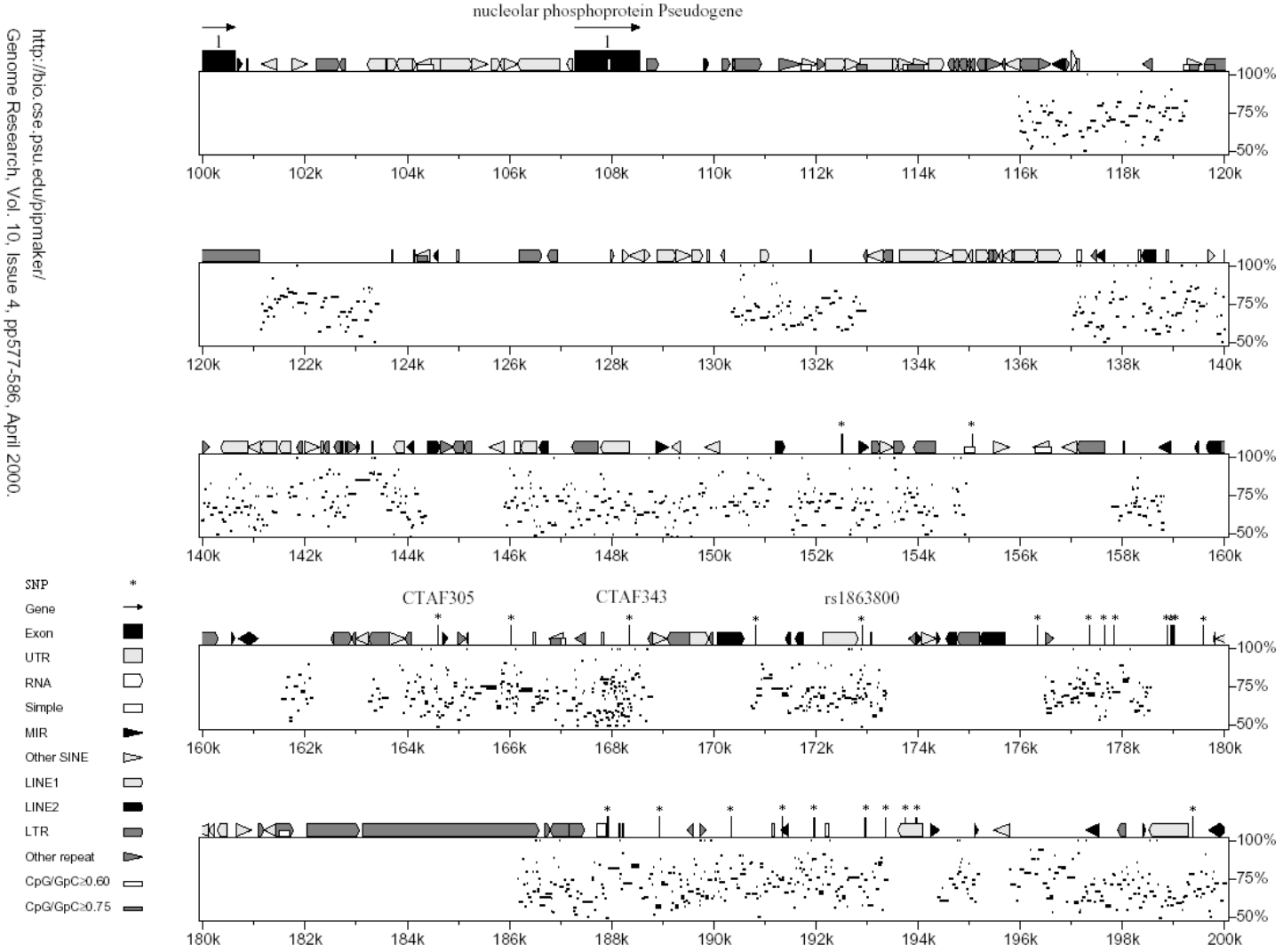
Fig. A1

Figure A1. Percentage Identity Plot (PIP).

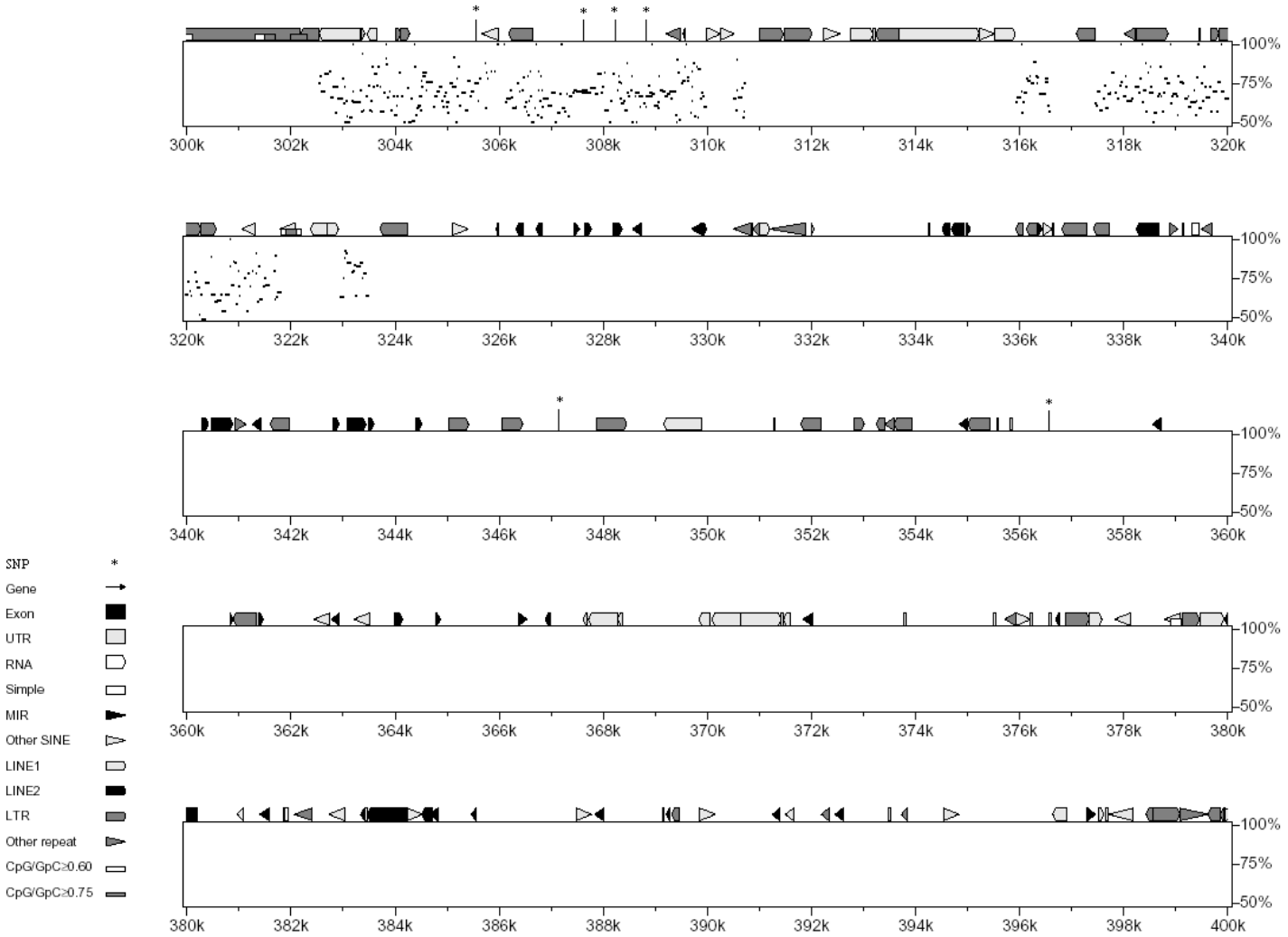
Figure A1



http://bio.cse.psu.edu/pipmaker/
Genome Research, Vol. 10, Issue 4, pp577-586, April 2000.



http://bio.cse.psu.edu/pipmaker/
Genome Research, Vol. 10, Issue 4, pp577-586, April 2000.


SNP
Gene
Exon
UTR
RNA
Simple
MIR
Other SINE
LINE1
LINE2
LTR
Other repeat
CpG/GpC≥0.60
CpG/GpC≥0.75
100%
75%
50%
300k
302k
304k
306k
308k
310k
312k
314k
316k
318k
320k
322k
324k
326k
328k
330k
332k
334k
336k
338k
340k
342k
344k
346k
348k
350k
352k
354k
356k
358k
360k
362k
364k
366k
368k
370k
372k
374k
376k
378k
380k
382k
384k
386k
388k
390k
392k
394k
396k
398k
400k

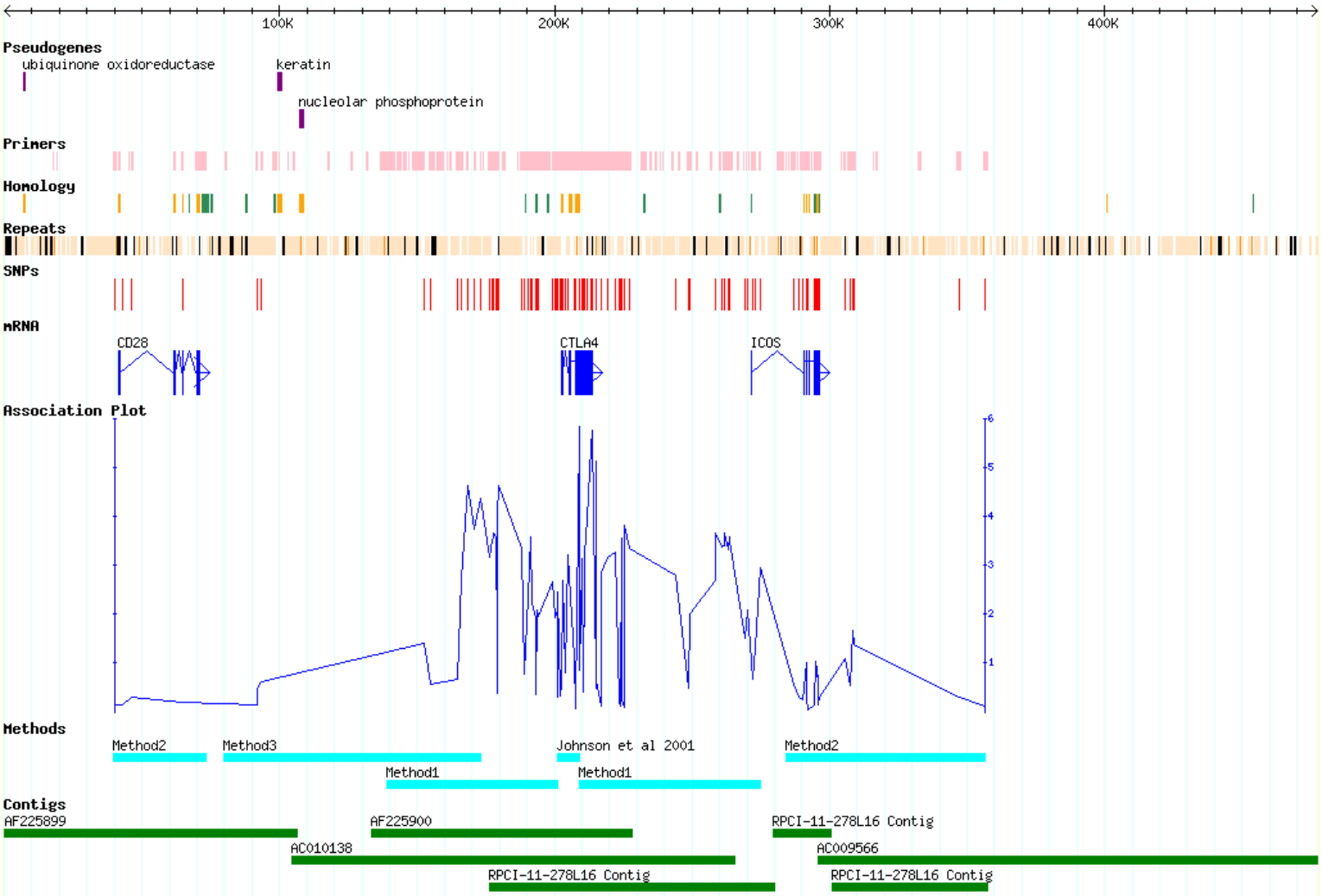


Figure A2. Overview of chromosome 2q33. The relative positions of genes and pseudogenes are shown, along with primers used for sequencing, homologies to dbEST and vertebrate mRNA sequences from EMBL, repeats detected by RepeatMasker, and SNPs confirmed by genotyping. An association curve (scale -log P) for the markers studied (see Supplementary Information B) and contigs used in the analysis are also shown. The Methods track indicates the methods used in SNP detection.

Figure A3. Pairwise alignment of the human and mouse CT60 region. Asterisks show identities between the two sequences and dashes denote gaps in the alignment. A polyadenylation site in the human sequence is shown at position +1,772 with respect to the stop codon of *CTLA4*. The Δ symbol marks the position at which a B3A SINE element was deleted from the mouse sequence to enable a better alignment between the two species. There is no corresponding element in the human sequence. The conservation in the 3ʼUTR of *CTLA4* may suggest a functional role for this region.

SNP identification

The scanned region and the three SNP identification methods used, are shown in Figure A2. PCR products from 32 individuals (the parents of eight probands from UK T1D families, and eight male and eight female control subjects) were scanned, using denaturing HPLC (Transgenomic Inc., Santa Clara, CA). Primers sets (136 in total) were designed for PCR products (500 - 600 bp) using Primer3 (http://www-genome.wi.mit.edu/cgi-bin/primer/primer3_www.cgi), and amplified in each of the 32 individuals and the products analysed according to Transgenomic Application Note 101. Samples with heteroduplex or different homoduplex patterns were then directly sequenced using the same PCR primers (Method 1). Direct sequencing of PCR products from 32 individuals was performed using an Applied Biosystems (ABI) 3700 capillary sequencer (Foster City, CA)(Method 2). For methods 1 and 2, the sequences from AF225900 and RPCI-11-278L16 were used for SNP harvesting. SNPs from the public database (dbSNP) were extracted (Method 3) (Table A2).

In total, more than 60% percent of the 317 kb sequence was scanned in 32 individuals for SNPs (Figure A2), excluding repeats sequences. All 28.6 kb of the CTLA-4 gene and its flanking DNA from 3.5 kb upstream of 5' UTR to 19 kb downstream of the published 3'UTR was contiguously scanned/re-sequenced. For *CD28* 7.7 kb was sequenced. This corresponds to 2.1 kb upstream of 5' UTR, all four exons, and 2.6 kb downstream of 3' UTR. For *ICOS*, SNPs were harvested from 13.9 kb corresponding to 2.7 kb upstream of the 5' UTR, the 5 exons, and all 3' UTR sequence. All the identified SNPs were mapped using the golden path sequence (NCBI build 28) and the information stored in our database. SNP coverage, and SNP positions are shown (Figure A2). Attempts were made to produce a denser SNP map between positions 100 kb and 150 kb using the public databases. No additional SNPs

were found. This may be due to the repetitive nature of this region. In total 108 SNPs were identified and/or confirmed.

All the supporting information, including SNP sequences, shown in Table A3 is also available from our ftp site (ftp://ftp-gene.cimr.cam.ac.uk/CTLA4).

Table A2: The names of re-sequenced SNPs in dbSNP

dbSNP identifiers	confirmed/genotyped
rs1879877	Yes
rs1181390	Yes
rs1181389	-
rs1181388	Yes
rs1980421	-
rs1980422	-
rs1181425	Yes
rs1181426	-
rs875020	-
rs179584	-
rs1974996	-
rs231387	-
rs749933	-
rs890637	-
rs1035846	-
rs231813	-
rs231815	-
rs231817	-
rs231736	-
rs1863800	Yes

Table A3: Sequences of 108 SNPs

SNP names	SNP sequences
1CD28p5_1Rb	AATATTAGCAGTGTTTATATTAGGGTG[T/G]TAAGAATGTGGATGAATCTTGTTTTAAT
2rs1879877	CTGGTGGAATAACCCTCTCT[T/G]CAAAGGGCCTGGGAGTTGAAGAAGGGTG
3rs1181390	TGGAGTAAGATGCTCATTCAAC[T/G]TGTTCTCAGAGTTAGAAAAAGGCCC
4rs1181388	AGGTAACCCAGGCCCCACAGA[G/A]AGGTGGTTGCATGCAAGTGTCTA
5CD28ex3F	TTTATTATTTTCTGGGTAAGAGAAGCAGCAC[T/C]GCTTTTATGTAACTTTTCCACTG
6rs1181425	CATGGTTCAATCTTGGTAGGTTGT[G/A]TGTATCTAGGAATTTATTCATTTATTCTAGTTTT
7rs1181425V	TTTTTCATCTCTGATTATATTTATTTGG[A/G]TCTTCTCTCTTTTCTTCTTA
8rs1181426V	GGCGTGAGCCACCACGCCTGGC[A/C]TGGATTTTTAAAAATTATTTCAATCTCCTTGTT
9CTAF185	TAGTAGGTTTCCTGGGGGATGAG[T/G]TCAAGGAATAAGAGGGGTCACTGACA
10CTAF212	CTCCGCCATCCACACTCCCAGGGTCTG[T/C]GGTGAAATCGAATTTATTCTCATTCTC
11CTAF305	AGGAGCTATTTTAATTTTCATGCAATGT[C/T]TCTGTAGGGTTCAGTTGGCGTTGT

12	CTAF322	AGTATCTACCGTGTCTTAATAT[T/C]GTATCCAATTGGAATTTGATATAAAACT
13	CTAF343	TGGGAGTATTTTACTGTGCTAAAA[T/C]ACATTTAGCATGGGCTGTTATATCTTATGAC
14	CTAF371_1	CAATACTAACACAGATAGG[T/G]GTATCGTAGCAGAGTCCTGATTGTGCGCC
15	rs1863800	GATAAAAAGGAACTGTTTAAA[C/T]TGATAGTAAAGAAAAGCCTTAAATTTTTGG
16	CTAF422	GTTCTATGGTTAAGAATCCGTAAGCATCA[G/A]GTCCTTTATTTTGCATCTCTGGGG
17	CTAF434_2	TCATTTAATTCATTACCCATATACTAC[G/A]TATGGGCATATAAATACCTGCTTT
18	CTAF439-1	CAGGACCAAACAAAACTGAA[T/C]AGGAACTCAAGTCACTCAAGTAACTT
19	CTAF439-2	CTTTGCTTCAAGACTCTCTTCT[C/G]TTTTATAGATGTAATTCCACTCACCCTCA
20	CTAF450_1	GTTTAACCTTTTAGGAGGGTTTTGTTG[T/C]TGCCTTATAGAAGCTAACATAAG
21	CTAF450_2	GCCAAAGAAAAGAAATGCAAAG[A/C]GGGCAGACAAAGAAAAAAAA
22	CTAF450_3	GAAAAAAAAAAAAAACAC[G/A]TGAATGTAAAGCTATGCCACTTTTT
23	CTAF450_4	AATGTAAAGCTATGCCACTTTTTT[T/A]CCCCTCTCATAGGTAATTCTGGAA
24	MH30	AATAAAAACAGAATAAAACAAT[G/C]AGAAAATTTTCACCTTTATTTAATTAGCAGA
25	MH26	CCCACTGACTCTGTCTCAAGAT[G/T]GTAATTAGTAACTGACAATGATTACGC
26	MH23	TATCTAAGACTTGGATTCAGTG[C/A]ATTTCACATCGCTGAGCTCTCACAT
27	MH20	TAGTCTGCCTGACATGCTT[G/A]GTGTATCTTGTATGATTTCTGTAAAGTTAGAA
28	MH18	AAAATTGTTGAGAGTTGAAAGTAAGAGA[T/C]CCTTCCTAATAGTGCCTTCTTATCTCT
29	MH17	CAATTTTATTTGCTAAATTTAGTAC[T/C]AGAGTGACATTATCTGTACATT
30	MH15	TGCATACAGAAACTGAGCTCTATGC[G/T]TGTGCCAGACAAAAACCAAAGAGCTT
31	MH14	CTGGCTGGTGATGGTTCAGTGT[G/A]ATGTGGGATCAGAATTGCAGATTAT
32	MH13_2	ACCCGAAAAGGAAATCTCATAC[T/C]GCCAAGCAGTCATTCCCCATTTTCC
33	MH13_1	CCTTCATTCCTTTTTATGGCT[G/A]AATAATATTCCGTGGTGTAGATAG
34	MH3	AATTAAGGAATCTGGTAAGA[T/C]GTCTTTGAAAGTTCTAATTACAAC
35	MH2	ATTTCCTTCTTTATAAAATGGGAAAAATG[G/A]TAACTCTTGTCTTGTAGGGTTGTTA
36	MH1	AGTAGTGGCAACAGAGACCC[C/T]ACCGTTTGCAAATCATAACATATTTACT
37	CT50	GCTCCTCTACATAATACTTCAA[T/C]TCCAGCATTGATCTCACTCTATCATGATCA
38	CT51	CTCTATCATGATCATGGGTTTAGCTG[T/C]CTGTCCCTGCCACTGCTGTGTGT
39	CT52	CAGGAACATTTGTTTTTCACTTTTT[A/G]AAAAACCTCTGTTGCCCAGTCT
40	CT53	GTTGGCGCTTGAGCTGGGGCTTGAAG[G/A]TTTCTATAATGTGTAGCAGTGTAT
41	CT41	AAAACCAGAGGCAGCTTCTTTTC[C/T]GCCTATTTTCAGTTTATTTCTTGTGAT
42	CT44 (-319C>T)	TCCACTTAGTTATCCAGATCCT[C/T]AAAGTGAACATGAAGCTTCAGT
43	+49G>A	GGCACAAGGCTCAGCTGAACCTGGCT[G/A]CCAGGACCTGGCCCTGCACT
44	CT43	TATCTCTCTCTAGACCTTCTTGG[C/T]TAAGAAACCATGTAGTTTGTATGA
45	CT55	GCAGCCACTATTTTTGAGTTGATGCAAG[T/C]CTCTCTGTATGGAGAGCTGGTCTCC
46	CT57	GACAGTCCCTCTCAGACA[C/A]CTCTGCCTAAGGCCAGCTTTGCCATTGCA
47	CT59	GTTTATAATTCTGTATGCTGTGA[A/G]CATTCATTTTTAACCAGCTAGGGACC
48	CT60	TTTTGATTTCTTCACCACTATTTGGGATATAAC[G/A]TGGGTTAACACAGACATA
49	CT61	TGGGTTAACACAGACATA[G/A]CAGTCCTTTATAAATCAATTGGCATGCTGT
50	JO37_3	TGGTAGCCATGAAGAAAAACACCAATC[G/A]GGAGCCTCAGTGGATA
51	JO37_2	GTATATCATTTCCACTCCTCTAAAC[G/A]TCTTTAGAGAGATTACTCTTTTTCATAGTT
52	JO37_1	GGGAGTGCATCTGGGCACCA[A/C]ATGAAAGCCTCTTCTAATTCCCATGTCA
53	JO36	GTCAGATTTGCTGACACTTTAAGCTC[A/G]TGGATTTCTCCTCTTTTGTTTCATAG
54	JO35	TAGGACATCCAGGACCGTTTT[T/C]CATACAGAACCCATCTGTG
55	JO34	CTACATGCTGGGATAGGGGCTCAT[A/G]GTAAGTTTGCCAGATTCAACCAA
56	JO31	AACAGTCTGTCAGCAAAGCC[G/T]GCAGTACACTGAGAAAGCTCCTATT
57	JO30	CGGACCTCTTGAGGTCAGGAGTTC[G/A]AGACCAGCCTGGCCAACATGGTGA
58	JO27_2	CAGGGCCTGAAGATGCCCC[C/G]AAATGGCAAATTAAGTGAACTGG
59	JO27_1	CCAGAAGTTGAAGTGTAGGAA[T/C]ATCTGGGGTCAAAGCAAAAAAAGACTTT
60	JO26_2	GTGTGTGGTGTGTGTGTGGGCAC[C/T]AAGCAGGGGTGAAAGCTGATTATGGAGG
61	JO26_1	GGTGAAAGCTGATTATGGAGGGG[T/C]GGTGTCAGGAAACATCACAAAGG
62	JO23	AATGAGAGAGTGCTTACCAT[A/G]TTTGTCTACCCAAACCTTTAAAGA
63	JO22	AATGTTGTACCCTTCAGCACA[T/C]TGGCATGAGCTCAGGTGGCAGGAGAACAA

64	JO18	CTTTTACCTTTTCTTTATTTTCCCGTCAG[T/C]AAATATTTGTTAGGCAAATAAGAGCC
65	JO13	TGGCAAATCTGGTCACCACAA[T/C]ACTCTTTAAAAAACACGCTCATGTTATT
66	JO10	CTCCCAGAGTTTGCCCTGTCCAAG[G/T]AATTCAGGGGAAAACGATGGTGGGCTC
67	JO9	CATGTAATTAGCCTACTGGCTCTA[C/G]AATAGTCATGATATATCCATTAAAA
68	JO8_2	AGAAAGACCCAAACCCACTTTTATACCAAACCCAC[T/C]CTTGTGATAACAAA
69	JO8_1	TTGTGATAACAAAC[T/C]CACTCCTGTTCTAATGACAGAGCCCTCATGACCT
70	JO6_2	CTAACAATCTGAAATGTGAAA[T/C]TATCTCRCTTTCACACAAGAAG
71	JO6_1	CAATCTGAAATGTGAAAYTATCTC[A/G]CTTTCACACAAGAAGAGATGTAGAA
72	JO3	AATATTAACCCTCAACAAC[C/A]GTAATGGATATAAACGCATGTGCTCATCA
73	CTBC358	GAAGGTGGTGCTCAGAGCAG[G/T]CTGCTTCCCTGCCAGGGTTATAGTCC
74	CTBC313	CTCAGTTTCAGAAATGTCCTCT[C/T]ATAGAAGCCCCTAGGAAAACAAGCCT
75	CTBC305	TCTGGTTTGCGTAATTTATGC[C/T]TTTGAAAGGCTGATATTATTGCATT
76	CTBC217_2	ATTTCTTTCTGGATATTGAAAA[A/C]CAATAGGGGTAGTGYCTTCCTTC
77	CTBC217_1	GAAAAMCAATAGGGGTAGTG[C/T]CTTCCTTCTAACCTCAGGGGATCCCTTCAG
78	CTBC190	GCAGCTATGTTTCAGATGTAATT[T/C]TGTTTCCACCAATCAGATGTGCCGAT
79	CTBC182_2	ACACTTCAGTGCATCTCTAATTT[G/T]GGAAATGATCAATGGTCAACAGAG
80	CTBC182_1	TAGTAGTTTTTACAATTGTATT[C/T]ATTAGGCACAGGAAAATGTAAACTCAG
81	CTBC165_3	AAGGGCTGGAGAAGCAGATT[A/C]TTGGGCTTAGCTTCTAGAAATA
82	CTBC165_2	AGTTCCAAAAGCAAGTGAGGTATTGGCA[C/T]ATCTGATTGGTGGAAACAAAATTA
83	CTBC165_1	ACAAAATTACATCTGAATTTTGTTT[G/T]CTGAATTGCTACATGGGCATAGG
84	CTBC106	TTCTTGACCAAGATTTTTGAGAAGG[C/T]GAGGCATGGGAAGGATCATTG
85	CTBC099	GGGGAAGCTTCAGCAGCTAGG[G/A]GGATCCCTGCAGGATCAGCCT
86	CTBC078	CATTTGGACTGGGAATTCTT[T/C]CTGTGGTAGAAAATAAAATG
87	CTBC073	TTAAGTGCAGTTAAGAGGA[T/C]TGAGGAATGTAGAATGAGCACAAG
88	CTBC053	GAAGTCTGTGAAGATGACTGACC[T/C]TGGAGCCCCTCTTAGCCTCTCCAC
89	CIIC065	AAAGCAACAGTGGGGCAACATATTT[T/C]ACTGTCAGACAAATAAAGCCC
90	IC082R	ATGAAAATTTGTCATATTAAAAAG[G/A]AAAATAAAACAAAAAGAA
91	CTIC098	TGCCTCCAATATGAAAGCTAC[C/T]AAAGGACAGAGGCCATACCAGTTT
92	CTIC114_1	GACCAACAGGGAGACAATT[C/T]CTTCCCCCSAAGACATACCTACTAAT
93	CTIC114_2	CCAACAGGGAGACAATTYCTTCCCCC[C/G]AAGACATACCTACTAATTAAAC
94	CTIC114_3	ATTTTTAAAAAATTATTATGC[A/G]TGGGCTGCGAATTTGAATCCAAGGTTCCTGA
95	CTIC142_1	TGTAGGGAACTGGCACATGGAGAGCATTTAGA[T/G]AATTTATGCTGAATTTTTGTTAC
96	CTIC142_2	GATGTGACCCTATAAT[A/C]TGGAACTCTGGCACCCAGGCATG
97	CTIC142_3	AAACAAACACATTTACAA[G/A]AAAAATGTTTTAAAGATGCCAGGGGTACT
98	CTIC148	AGTTAAAATGGTTTACTTTGTTCA[A/T]GTTAGTGGTAGGAAACATTGCCC
99	CTIC154_1	TCTTTAATGGGCCAGCATTCT[C/T]ATGGGGTAGAGCAGAATATTC
100	CTIC154_2	GAATCACAGTGGTCTACCTGC[A/G]TTCATAATTCCAGGATCTGTGAA
101	CTIC154_3	TGGCTAGAAAGATTCTTAAATAT[G/A]TGGAATATGATTATTCTTAGCTGGAATATTT
102	CTIC159	TGTGTGTGGGGTGGGGTATGGGGAGGA[G/C]AACCTTCATGGTGGCCCACCTGG
103	JC034sR	CATTCAAAGAAATGATGGTCC[G/A]TCAGAAATACATTTTAACTGAG
104	JC058sR	ATGGTATATGTTAAATGCT[G/A]TGGTGTCTGAATTTATTTCCTAGCAACA
105	JC068sFa	ATCCCTAAATATTCATTCTCCT[C/G]TCTTATCTCCCATCTTCCTCTA
106	JC068sRb	CTGCAACTCCAGCAGGTAAAATCTG[C/G]TTACCTAACCACCAATTGTGGTATGG
107	JC473sR_4	GTTTTCCTAGAATTGATTC[A/C]ATTAACAACTTACTGGGGTTTTAAAAG
108	JC569sF	GTAACAAGGAAAACTCTACAGGGTGG[G/A]ATTGAGATATTGGATCTTGATATTA

DNA preparation

For cell-line samples, DNA was extracted from Epstein-Barr virus (EBV)-transformed peripheral blood lymphocytes. One hundred ml of confluent EBV cells were pelleted at 1000 rpm for 5 min. The supernatant was discarded and cells were re-suspended in 3.5 ml of 6 M guanidine hydrochloride (Sigma, Dorset, UK), 250 µl of 7.5 M ammonium acetate (VWR international Ltd, Lutterworth, UK), 50 µl of Proteinase K (VWR international Ltd, Lutterworth, UK) and 250 µl of 20% sodium N-laurylsarcosinate solution (Fluka, Dorset, UK). The solution was incubated overnight at 37 °C. The tubes were cooled to RT and then 2 ml of chloroform added. The samples were then mixed and spun at 2500 rpm for 3 min. The upper layer was removed and added to 10 ml of absolute ethanol. The tubes were chilled to -20°C overnight. The solution was gently inverted and the precipitated DNA pelleted at 3000 rpm for 15 min in an Allegra 6R micro-centrifuge (Beckman, High Wycombe,UK). Pellets were washed with 70% ethanol and re-suspended in 300 .l TE (10 mM Tris, 0.1 mM EDTA; pH 7.5). DNA was quantified using PicoGreen dsDNA quantitation reagent (Molecular Probes Europe B.V., Leiden, The Netherlands) and diluted to 4 ng/ul in TE (pH 7.5) before use. For blood samples, whole blood was lysed using 30 ml of lysis buffer (0.3 M Sucrose, 1.0 M Tris-HCl, 4.9 M $MgCl_2$, 1% Triton-X (Sigma, Dorset, UK)). DNA was extracted from the resultant white blood cell pellet using the method described above for cell-line samples.

Genotyping PCR

PCR conditions were optimised by varying $MgCl_2$ concentrations between 1 and 5 mM and annealing temperature between 50 °C and 65 °C. PCRs were performed in 384 well polypropylene microtitre plates (ABgene, Epsom, Surrey, UK) in 6 µl final

reaction volume. 3 µl of 4 ng/µl stock of genomic DNA was dispensed into each well. 3 µl of PCR reaction mix containing 0.4 mM dNTP, 2-10 mM $MgCl_2$, 3.8 ng/µl forward and reverse primer, and 0.25 units of TaqGold (Applied Biosystems, Foster City, CA) was added and the plate was covered with Adhesive Sealing Sheet (ABgene, Epsom, Surrey, UK). Reactions were incubated at 95 ^{o}C for 15 min and then cycled for 35 cycles of 95 ^{o}C for 30 s, 50-65 ^{o}C for 30 s, 72 ^{o}C for 30 s, and finally incubated for 15 min at 72 ^{o}C on PCT225 thermocyclers (MJ Research, Watertown, MA) using heated lids. All pipetting steps for PCR preparation were performed with on a Beckman Multimek™ 96 robot (Beckman, High Wycombe, UK), dedicated to pre-PCR work.

Genotyping methods

Apart from early RFLPs[5] all the genotyping on this project was performed using the Invader® genotyping method described previously[6]. Probe sets for each SNP were designed and synthesised by Third Wave Technologies, Inc (Madison, WI). For a uniplex dry-down plate Invader® assay, assays were prepared for each allele separately. In a 384-well format, PCR products were diluted 1:4 in distilled water. Aliquots of 3 µl were dispensed into two pre-prepared 384-well format Invader® Assay FRET detection plates (Third Wave Technologies, Inc. Madison, WI). Six microlitres of probe mix containing 1.3 µM allele specific probe, 0.13 µM Invader™ probe, and 7.5 mM $MgCl_2$ were added. All pipetting steps for this preparation were performed with a Beckman Multimek™ 96 robot. Plates were covered with Adhesive Sealing Sheet (ABgene, Epsom, Surrey, UK). Reactions were incubated at 95 ^{o}C for 5 min, and 65 ^{o}C for 10-80 min on MJ PCT225 thermocyclers. Fluorescence was measured directly at the end of incubation using a Cytofluor 4400 fluorescence

microtiter plate reader (Perkin Elmer Applied Biosystems, Foster City, CA), excitation 485/20, emission 530/25, and gain 50.

For a biplex dry-down plate Invader® assay, assays for both SNP alleles were performed at the same time. Final reaction concentrations of the allele-specific probes were 0.7 μM and the Invader™ probe was 0.07 μM. The probes were labelled either FAM or RED. Fluorescence was measured directly at the end of incubation using a Cytofluor 4400, excitation 485/20, emission 530/25 for FAM and excitation 560/20, emission 620/40 for RED. All pipetting steps were performed with a Beckman Multimek™ 96 robot.

The $(AT)_n$-3' UTR microsatellite marker was genotyped using fluorescently-labelled primers as described elsewhere[7].

Genotype concordance

Genotyping concordance was assessed by double-typing a small number of key SNPs using either a second batch of biplex Invader® kit, or the Taqman MGB chemistry (Applied Biosystems Inc, Foster City, CA)[8-9].

MH30, CT60 and JO31 were genotyped using both Taqman and Invader® in either UK T1D families, GD cases or controls by two scientists independently. The allelic differences were 0/632(0%) for MH30, 1/688 (0.037%) for CT60, 1/1,898 (0.053%) for JO31, in total, 2/5,218 (0.038%). CTAF371_1 was genotyped using Taqman and biplex Invader® by the same scientist. No allelic difference was observed in 1,990 alleles. No typing difference was observed in 768 alleles for CTIC154_1 comparing two different batches of biplex Invader® kits.

Genotype checks

All the genotyping data was stored in our local MySQL database (http://www.mysql.com). This allows easy double scoring, error checking and data retrieval. Data manipulation errors are reduced to a minimum by employing automated approaches. All Invader® and Taqman data was scored twice independently.

Case-control datasets In the case-control datasets, genotyping data was checked for Hardy-Weinberg equilibrium (HWE) separately for the cases and controls. Markers and datasets showing deviation from HWE ($0.01 < P < 0.05$) are presented in Table A4. Datasets, which showed deviation from HWE ($P < 0.01$) were treated as genotyping errors, and were retyped. No data set used in analysis showed deviation from HWE ($P < 0.01$). Of the 216 datasets (108 SNPs) typed in the Graves' cases and controls, 18 datasets are out of HWE ($0.01 < P < 0.05$). No datasets were out of HWE in the large Graves' case-control dataset or in the AIH cases.

Since regression analyses are sensitive to the number of typed samples, we retyped SNPs if there were > 10% genotyping failures in either the cases or controls. The percentages of genotype failures were as a result less than 10% in each case and control dataset used, except for rs1181426V cases (15.9%).

D'(Ref)[10] and W_n(Ref)[11] were used as the measure of pairwise LD in the LD blocks, measured both in a moving window across the region and against the double-typed key frame markers, MH30 and CT60. If D' or W_n differed substantially between cases and controls between a pair of markers (and D' > 0.8), we retyped the marker.

Table A4: Markers and datasets showing deviation from Hardy-Weinberg equilibrium

(a) Graves' cases and controls

Marker	Dataset	*P* value
MH13_1	cases	0.028
MH2	cases	0.012
MH1	cases	0.034
IC082R	cases	0.033
CTIC098	cases	0.035
CTIC159	cases	0.028
JC068sFa	cases	0.046
rs1879877	controls	0.022
rs1181390	controls	0.044
MH23	controls	0.032
CT43	controls	0.028
CT61	controls	0.041
CTBC106	controls	0.025
CTBC099	controls	0.017
CTBC073	controls	0.011
CTBC053	controls	0.012
JC058sR	controls	0.033
JC473sR_4	controls	0.021

(b) T1D families

CTAF343	USA	0.015

Family collections All typing data on family collections, including the microsatellite marker $(AT)_n$ 3'UTR, were screened for Mendelian inconsistencies using PedCheck[12]. Families that consistently misinherit in > 3% of SNPs were marked in the database, and automatically excluded from analysis. The familial dataset was screened again for Mendelian inconsistencies in STATA (http://www.stata.com). The working hypothesis is that these give a measure of genotyping error, given we have already excluded "real" misinheritances. These apparent misinheritances were eliminated by re-typing or excluding a family from the analysis for that marker. Mendelian inconsistencies in the 3,671 T1D family dataset are tabulated in Table A5. There were

231 inconsistencies out of 41,406 informative transmissions (0.56%). Some families identified by STATA have been recently excluded as newly-identified consistently mis-inheriting families in our lab. The number of apparent offspring inconsistencies arising from genotyping errors is, therefore, less than 0.56%.

D' and W_n values were calculated with two double-typed frame markers (MH30, CT60) and were used for LD pattern checking. If the datasets showed largely different D' or W_n values within each population, we re-typed the marker.

The familial genotyping data was also screened for HWE using parents only by population, grouped as UK (Warren I, Yorkshire, Belfast, South West and Bristol), Norway, Romania, USA and Finland. Only one of 41 datasets showed deviation from HWE in Graves' and T1D families (not shown).

Table A5: Apparent misinheritances in 3,671 T1D families

	Affected sibling			Unaffected			
SNP name	Inconsistencies	Informative	%	Inconsistencies	Informative	%	Total %
CTAF343	18	2321	0.78	5	1211	0.41	0.65%
rs1863800	20	3282	0.61	10	1696	0.59	0.60%
MH30	25	3647	0.69	20	2033	0.98	0.79%
+49G>A	1	3434	0.03	2	1837	0.11	0.06%
CT60	15	3824	0.39	11	2099	0.52	0.44%
JO31	25	3290	0.76	15	1788	0.84	0.79%
JO30	17	2994	0.57	8	1506	0.53	0.56%
JO27_1	20	3177	0.63	9	1609	0.56	0.61%
CTIC154	5	1084	0.46	5	574	0.87	0.60%
Total	146	27053	0.54	85	14353	0.59	0.56%

CTLA-4 gene expression

Primers and probes to detect human CTLA-4 and sCTLA-4 mRNAs were designed using Primer Express software (Applied Biosystems, Foster City, CA), β2-microglobulin by use of Primer 3 program (Table A6). The mouse primer and probe sets are shown in Table A7. All the synthesised probes were labelled FAM and

TAMRA at the 5’ and 3’, respectively. All the sequences and the annealing temperatures used for the 3’UTR mapping are tabulated in Table A8. Figure A4 shows our transcription analysis results of human the CTLA-4 gene.

Table A6: Primers and probes for human Taqman assay

Target Gene	Forward (5’-3’)	Reverse (5’-3’)	Probe (5’-3’)
flCTLA-4	ACCCAGATTTATGTAATTGATCCAGAA’	CCGAACTAACTGCTGCAAGGA’	CGTGCCCAGATTCTGACTTCCTCCTCT
sCTLA-4	CATCTGCAAGGTGGAGCTCAT	GGCTTCTTTTCTTTAGCAATTACATAAATC	ACCGCCATACTACCTGGGCATAGGCA
β2 microglobulin	TGCTCGCGCTACTCTCTCT	TCCATTCTCTGCTGGATGAC	CTGGAGGCTATCCAGCGTACTCCAA

Table A7: Primers and probes for mouse Taqman assay

Target Gene	Forward (5’-3’)	Reverse (5’-3’)	Probe (5’-3’)
flCTLA-4	ACTCATGTACCCACCGCCATA	GGGCATGGTTCTGGATCAAT	CATGGGCAACGGGACGCAGATTTAT
sCTLA-4	ACCCACCGCCATACTTTGTG	AGGACTTCTTTTCTTTAGCAATGACAT	AATCTGCGTCCCGTTGCCCATGC
liCTLA-4	GCCTTTTGTAGCCCTGCTCA	TCAGAATCCGGGCATGGTT	TTCTTTTCATCCCAGTCTTCTCTGAAGATCCA
β2 microglobulin	CCTGCAGAGTTAAGCATGCCA	TGATCACATGTCTCGATCCCA	TATGGCCGAGCCCAAGACCGTCT

Table A8: Primers for the 3’ UTR mapping

Forward (5’-3’)	Reverse (5’-3’)	PCR condition (annealing)
GAGCTGGGATGTTTCTGTCA	CTTAATTCCTTGACCCACATCA	55°C.
ATGAGTCAGCTTTGCACCAGC CATTAC	GAGGTGAAGAACCTGTGTTAAACAG CATG	TD 70°C →60°C
GAACCACTTGTCACAGGCAAG ACTG	GTCAGCAAATCTGACGGTGGAGGC	TD 70°C →60°C
TGTGATTATCACCACTGTTGCT GCT	CTGTTAGTTGCTATAGTAACTTGACT G	55°C
CTGCTTCCACCTTGTCATGTTA GCCG	GTAGAGACAGGGTTTCACCATGTTGG C	TD 70°C →65°C
CTGCTTCCACCTTGTCATGTTA GCCG	TCTCCTGCCTCAGCCACCCGAGTAGC G	TD 70°C →65°C
CATAGTCTCTACTCTCCTGTGA GC	ATGGAGATGATGGCAGAACCTC	55°C.

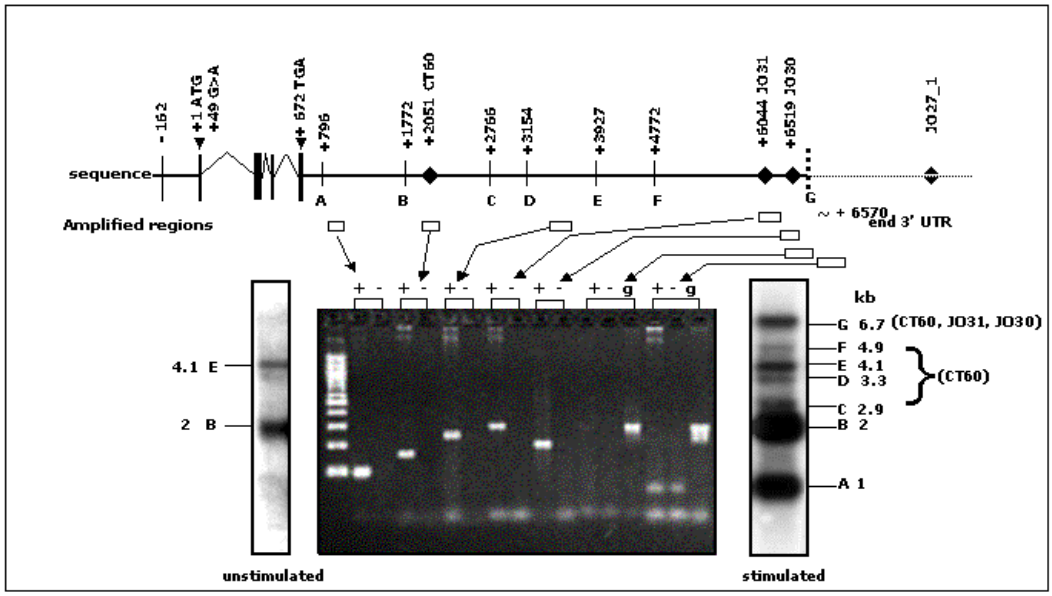


Figure A4**:** Transcription of the human CTLA-4 gene. Seven different sets of primers, indicated by open boxes, were designed on the basis of the genomic sequence downstream the open reading frame of *CTLA4*. For each primer set both +RT (reverse transcriptase)(+) and –RT(-) PCR products are displayed. RT-PCR mapping placed the end of the 3'UTR 6.57 kb 3' the last codon of CTLA-4 and CT60, JO31 and JO30 inside the 3' UTR. Northern Blot of poly(A)$^+$ mRNA derived from unstimulated and 24 h stimulated PBMC show the presence of multiple transcripts of CTLA-4 due to the existence of different poly(A) signal (A-G) in the 3' UTR. JO31 and JO30 are only present in the long 6.7 kb transcript. All five of the 2.9-6.7 kb transcripts include CT60. Primer sequences are given in Table A8. With respect to the pattern of splicing since the disease-associated 6.1 kb region is only 279 bp 3' of the major polyadenylation poly(A) mRNA termination site of *CTLA4*, we propose that its allelic variation might affect termination of the transcript. Sequences 3' of the poly(A) site bind the protein complex that mediates transcription termination[13-15]. The termination

protein complex in turn binds to the proteins mediating splicing of the final exon in the highly orchestrated production of a mature transcript[15]. The pre-mRNA complex destined to form mature sCTLA-4 mRNA may be more stable when the protective allele contributes to the secondary structure or, alternatively, the splicing needed to produce the sCTLA-4 mRNA from the protective allele may be more efficient.

References

1. Osoegawa, K. et al. A bacterial artificial chromosome library for sequencing the complete human genome. *Genome Res* **11**, 483-496 (2001).
2. Osoegawa, K. et al. Bacterial artificial chromosome libraries for mouse sequencing and functional analysis. *Genome Res* **10**, 116-128 (2000).
3. Durbin, R. & Thierry-Mieg, J. AC.elegans database. Documentation, code and data available from anonymous FTP servers at lirmm.limrmm.fr, cele.mrc-lbm.cam.ac.uk and ncbi.nlm.nih.gov. (1991).
4. Schwartz, S. et al. PipMaker--a web server for aligning two genomic DNA sequences. *Genome Res* **10**, 577-586 (2000).
5. Johnson, G. C. et al. Haplotype tagging for the identification of common disease genes. *Nat Genet* **29**, 233-237 (2001).
6. 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).
7. 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).
8. User Bulletin: Primer Express Version 1.5 and TaqMan MGB Probes for Allelic Discrimination (http://www.appliedbiosystem.com)

9. Ranade, K. et al. High-Throughput Genotyping with Single Nucleotide Polymorphisms. *Genome Res.***11**,1262-1268 (2001)

10. Devlin, B. & Risch, N. A comparison of linkage disequilibrium measures for fine-scale mapping. *Genomics* **29**, 311-312 (1995).

11. Klitz, W., Stephens, J. C., Grote, M. & Carrington, M. Discordant patterns of linkage disequilibrium of the peptide- transporter loci within the HLA class II region. *Am J Hum Genet* **57**, 1436-1444 (1995).

12. O'Connell, J. R. & Weeks, D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am J Hum Genet* **63**, 259-266 (1998).

13. Proudfoot, N. J., Furger, A. & Dye, M. J. Integrating mRNA processing with transcription. *Cell* **108**, 501-512 (2002).

14. Dye, M. J. & Proudfoot, N. J. Multiple transcript cleavage precedes polymerase release in termination by RNA polymerase II. *Cell* **105**, 669-681 (2001).

15. Cartegni, L., Chew, S. L. & Krainer, A. R. Listening to silence and understanding nonsense: exonic mutations that affect splicing. *Nat Rev Genet* **3**, 285-298 (2002).