

## Supplementary Appendix

This appendix has been provided by the authors to give readers additional information about their work.

Supplement to: Smyth DJ, Plagnol V, Walker NM, et al. Shared and distinct genetic variants in type 1 diabetes and celiac disease. *N Engl J Med* 2008;359:2767-77. DOI: 10.1056/NEJMoa0807917.

# Shared and Distinct Genetic Variants in Type 1 Diabetes and Celiac Disease

## Supplementary Appendix

Deborah J Smyth, Vincent Plagnol, Neil M Walker, Jason D Cooper, Kate Downes, Jennie HM Yang, Joanna MM Howson, Helen Stevens, Ross McManus, Cisca Wijmenga, Graham A. Heap, Patrick C. Dubois, David G. Clayton, Karen A Hunt, David A van Heel and John A Todd.

Names of Genes	Page 3
Glossary	Page 4
Methods	Pages 5-9
Supplementary Table 1	Pages 10-12
Supplementary Table 2	Page 13
Supplementary Table 3	Pages 14-18
Supplementary Table 4	Page 19
Supplementary Table 5	Page 20
Supplementary Table 6	Page 21
References	Pages 22-23

## NAMES OF CANDIDATE GENES (Chromosome Location)

<http://www.tlbase.org>

*BACH2* (6q15) = BTB and CNC homology 1, basic leucine zipper transcription factor 2  
*C1QTNF6* (22q13) = C1q and tumor necrosis factor related protein 6  
*CCL3* (17q12) = chemokine (C-C motif) ligand 3  
*CCL4* (17q12) = chemokine (C-C motif) ligand 4  
*CCL3L1* (17q21) = chemokine (C-C motif) ligand 3-like 1  
*CCR2* (3p21) = chemokine (C-C motif) receptor 2  
*CCR3* (3p21) = chemokine (C-C motif) receptor 3  
*CCR5* (3p21) = chemokine (C-C motif) receptor 5  
*CD226* (18q22) = CD226 molecule  
*CLECI6A* (16p13) = C-type lectin domain family 16, member A  
*CTLA4* (2q33) = cytotoxic T-lymphocyte-associated protein 4  
*CTSH* (15q24) = cathepsin H  
*ERBB3* (12q13) = v-erb-b2 erythroblastic leukemia viral oncogene homolog 3 (avian)  
*HLA-DRB1* (6p21) = major histocompatibility complex, class II, DR beta 1  
*HLA-DQB1* (6p21) = major histocompatibility complex, class II, DQ beta 1  
*IL2* (4q27) = interleukin 2  
*IL2RA* (10p15) = interleukin 2 receptor, alpha  
*IL7R* (5p13) = interleukin 7 receptor  
*IL12A* (3q25) = interleukin 12A  
*IL18RAP* (2q12) = interleukin 18 receptor accessory protein  
*IL21* (4q27) = interleukin 21  
*IFIH1* (2q24) = interferon induced with helicase C domain 1  
*INS* (11p15) = insulin  
*LPP* (3q28) = LIM domain containing preferred translocation partner in lipoma  
*PRKCQ* (10p15) = protein kinase C, theta  
*PTPN2* (18p11) = protein tyrosine phosphatase, non-receptor type 2  
*PTPN22* (1p13) = protein tyrosine phosphatase, non-receptor type 22 (lymphoid)  
*RGS1* (1q31) = regulator of G-protein signaling 1  
*SH2B3* (12q24) = SH2B adaptor protein 3  
*TAGAP* (6q25) = T-cell activation RhoGTPase activating protein  
*UBASH3A* (21q22) = ubiquitin associated and SH3 domain containing, A

## GLOSSARY

**Genome-wide association study (GWA):** Genotyping hundreds of thousands of SNPs on an array, capturing a substantial proportion of common variation in the genome, in a set of samples that are informative for a trait of interest, leading to the mapping of genetic risk factors through the detection of associations between SNP genotype frequencies in cases and controls and trait status.

**Linkage disequilibrium (LD):** The non-random association of alleles at nearby variants on individual chromosomes as a result of recent mutation, genetic drift or selection, manifest as correlations between genotypes at closely linked markers.

**Transmission disequilibrium test (TDT):** A family-based test of genetic association that measures the transmission of an allele from heterozygous parents to affected offspring.

**Pseudo case-control:** A case-control data set can be derived from family data, with the affected offspring as “cases” and between one and seven matched pseudo-controls constructed from parental genotypes. The case and pseudo-control samples, matched by derived family group, were analyzed using conditional logistic regression

**Logistic regression:** A statistical model that is used when the outcome is binary, for example disease status, and relates the log odds of the probability of an event:  $Pr(\text{event})$  to a linear combination of predictor variables. To minimize any confounding due to variations in allele frequencies across Great Britain, all analyses of the case-control collection were performed using a logistic regression model adjusted for 12 broad geographical regions within the country. To provide the most powerful test of whether a SNP is associated with disease, we tested whether genotypic effects model was significantly different from the multiplicative allelic effects model.

**Logistic forward regression:** In a small genetic region with strong LD, it is difficult to distinguish between a disease associated SNP that is the actual causative loci and SNPs that are in strong LD with the causative locus but do not have any effect on disease and that show disease association purely because of LD. The premise for the analysis is that to distinguish the effects of two loci, A and B, the least associated locus, B, is added to a model containing the most associated locus, A. If the B locus is fully explained by the A locus then there will be no improvement in the fit of the model, whereas if locus B adds to the model there will be a significant improvement in the model, suggesting that locus B is causal or in stronger LD with the yet un-genotyped causal variant. The reverse situation is then investigated by taking a model, containing locus B, and adding locus A to see if there is any improvement in fit.

## METHODS

### Genotyping

SNPs were genotyped using TaqMan (Applied Biosystems). All type 1 diabetes (T1D) case-control genotyping data were scored blind to case-control status and were double scored by a second operator to minimize error. The 32 base pair insertion/deletion in *CCR5* was genotyped using fluorescently labeled primers (forward primer AATCTTCTTCATCATCCTCCTGAC, reverse primer GCCCAGAAGAGAAAATAACAATC) and sized on an Applied Biosystems 3730xl and analyzed in GeneMapper. The *PTPN2* SNP rs1893217, which until the present study, was the SNP most associated with T1D the *PTPN2* region, was genotyped in celiac disease patients and controls as part of a SNP set using the Goldengate assay (Illumina) as described previously.<sup>1</sup>

### Statistical Analyses

Statistical analyses on the data presented here were performed in Stata (<http://www.stata.com/>), R statistical systems and PLINK.<sup>2</sup> Logistic regression models were used for T1D and celiac case-control association tests. In all analyses, we stratified by place of collection for the cases and controls for 12 geographical regions of England, Scotland and Wales (Southwestern, Southeastern, London, Eastern, Wales, Midlands, North Midlands, Northwestern, Yorkshire, East and West Ridings, Northern, and Scotland) to exclude the possibility of confounding by geography with little loss of power, given how well the cases and controls were matched geographically.<sup>3,4</sup> In the logistic regression analysis of a SNP, we performed a one-degree of freedom (1-d.f.) likelihood ratio test to determine whether a 1-d.f. multiplicative allelic effects model or a 2-d.f. genotype effects model (no specific mode of inheritance assumed) was more appropriate.<sup>5</sup> We used the Transmission/disequilibrium test (1-d.f. multiplicative allelic effects model), and conditional

logistic regression models (2-d.f. genotypic effects model) using pseudo cases and controls<sup>5</sup> for the T1D family association tests. Control samples for all SNPs were in Hardy-Weinberg equilibrium,  $P > 0.01$ . Celiac results taken from Hunt *et al*<sup>1</sup> were calculated by allele count  $\chi^2$  Cochran-Mantel-Haenszel tests.

### **Combined Test**

A Cochran Armitage 1-df trend test was used, extended to pool information across multiple strata within a single data set or across multiple data sets, and a Wald test was used to combine 2-df  $P$  values.

### **Potential Biases**

The T1D cases were recruited solely on the basis on their T1D status from diabetes clinics, not from celiac disease clinics; in the latter case this would lead to bias and invalidate our results. Similarly, celiac disease cases were not recruited from diabetes clinics. Therefore, the genotype distribution in cases can only be affected by T1D status (or celiac disease status for the celiac disease collection). Any association between genotype and T1D status has to be causal for T1D and shared associations between T1D and celiac disease must be the consequence of shared pathways between both diseases. That the alleles of the 6q25/*TAGAP* and *IL18RAP* loci were associated in the opposite direction in T1D; the 4q27/*IL2-IL21* region associations were distinct (Supplementary Tables 1 and 3); the 3q25/*IL12A* and 3q28 regions did not show association with T1D (Supplementary Table 1); and *PTPN22*, *IL2RA* and *INS*, which are very strongly predisposing to T1D, were not associated with celiac disease in this sample set (Supplementary Table 3), show that we do not have this bias in ascertainment.

We note that for the T1D and celiac disease studies we used the same controls (Supplementary Table 3). This could potentially introduce a problem, if the control frequencies vary by chance and are responsible for a disease association, or for missing one (a false negative result), rather than variation in the case allele frequencies. We know that this is not a problem in T1D because our

family results (Supplementary Table 1) <sup>6-8</sup> are consistent with the case-control data and validate the control group, and provides reassurance for the presented celiac case-control disease data.

### **Study Design and Statistical Power Considerations**

In order to test the hypothesis that celiac loci are associated with T1D, very large collections of carefully-matched cases and controls with high-quality DNA samples and robust genotyping technology are required in order to avoid hidden biases that could lead to false-positive results, especially for associations with odds ratios less than 1.2, and to ensure that the statistical power of the study is maximized.<sup>3,6</sup> To this end, we have collected DNA samples from over 8,000 cases of T1D, diagnosed under age 17 years, from England, Scotland and Wales, and matched these geographically to over 9,000 controls from the sub-regions of Great Britain, resulting in little or no bias due to variation in allele frequencies across Great Britain, the major known covariate in T1D association analyses.<sup>3,4,6</sup> In an idealized situation, we have 90% power for disease-associated allele at MAF >0.10 and false-positive rate  $\alpha = 10^{-4}$  to detect effects at odds ratio 1.2 and above. A collection of over 2,500 celiac disease case samples, also from England, Scotland and Wales, has been assembled and matched with the same 9,000 controls by subregions of Great Britain in order to evaluate directly effects of T1D loci in celiac disease with 30% power to detect odds ratio 1.2 at  $\alpha = 10^{-4}$ .

There has been much recent discussion concerning appropriate significance thresholds for association studies, although most of this has been in the context of genome-wide studies;  $P < 5 \times 10^{-7}$  has emerged as a reliable threshold.<sup>4,6</sup> The choice of an appropriate threshold to control the false positive rate is somewhat arbitrary, depending on an estimate of the prior odds against there being a true association at any locus tested.

Thomas and Clayton <sup>9</sup> suggested that, even in candidate gene studies, these odds could

be as low as 1000:1 against and that, using the arguments advanced by Wacholder et al.,<sup>10</sup> this could suggest thresholds as stringent as  $10^{-4}$ . In our study we would estimate the prior odds against true association as being rather more favorable than those considered by Thomas and Clayton;<sup>9</sup> but, nevertheless, we elected to be conservative and claim "significance" only when  $P < 10^{-4}$ .

### **MHC Associations**

The striking commonality in the genetic bases of celiac disease and T1D prompts comparison with the T1D NOD mouse model in which exchange of disease susceptibility gene regions, such as the MHC or different combination of the non-MHC *Idd* genes produces variant strains of NOD with increased predisposition to other autoimmune diseases, such as autoimmune thyroid disease<sup>11</sup> and liver autoimmune disease.<sup>12</sup> NOD mice carry the H2<sup>g7</sup> MHC haplotype, but when the H2<sup>h</sup> haplotype is introduced the new strain develops autoimmune thyroid disease in response to dietary iodine (AITD).<sup>11</sup> Hence, the same non-MHC genetic background can cause T1D and AITD. MHC class II and class I molecules alter autoantigen, tissue specific targeting, switching the underlying organ-specific autoimmunity from one organ to another. Altering the likelihood of disease outcomes can also be achieved by swapping non-MHC combinations,<sup>12</sup> and by introducing certain transgenes.<sup>13</sup> This comparison is particularly relevant for T1D versus celiac disease comparison in human subjects. The main T1D genetic effects map to the HLA class II loci, *HLA-DQB1* and *HLA-DRB1* on chromosome 6p21,<sup>14-17</sup> with alleles DQB1\*0302, DQB1\*0201 and DRB1\*0401 being most associated, as well as involvement of the class I locus, *HLA-B*.<sup>16</sup> These same HLA Class II alleles predispose to celiac disease. However, the heterozygous genotype, DR3/4, encoding the transdimer combination of DQ alleles, DQA1\*0501 and DQB1\*0302, has a particularly strong association with T1D susceptibility, but not in celiac disease, even though this genotype has both the major celiac disease *DQB1*

alleles, DQB1\*0201 and, secondarily, DQB1\*0302.<sup>18-21</sup> There is biochemical evidence that the DQA1\*0501-DQB1\*0302 transdimer has unique peptide-binding properties,<sup>22</sup> presumably facilitating immune presentation of key diabetogenic peptides. Hence, it is possible that individuals with the DR3/4 genotype and a sufficient dose of non-MHC susceptibility alleles that do not have major effects in celiac disease, such as *INS*, *IL2RA*, and *PTPN22*, develop anti-islet autoimmunity and T1D, instead of celiac disease, on top of a background of shared pathways. Additionally, DR3/7 genotypes predispose to celiac disease, owing to the DQB1\*0201- DQA1\*0501 molecule, but DR3/7 is not predisposing to T1D.

		Type 1 Diabetes Case-Control Results				Type 1 Diabetes Family Results			
SNP region gene	Minor-allele Genotypes	Genotype		P-values		frequency		P-values	
		Frequency n (%) cases	Frequency n (%) controls	OR (95% CI)	1-df 2-df	n (%) cases	n (%) parents	RR (95% CI)	1-df 2-df
rs2816316 <b>1q31</b> <i>RGS1</i>	C	2,540 (16.58)	3,424 (18.18)	0.89 (0.84-0.95)	1.23 x 10 <sup>-4</sup>	1,120 (16.10)	1,674 (16.24)	0.91 (0.82-1.00)	0.0436
	A/A	5,343 (69.74)	6,331 (67.22)	1.00 (reference)		2202 (71.31)	3,607 (70.03)	1.00 (reference)	
	A/C	2,095 (27.35)	2,752 (29.22)	0.90 (0.84-0.96)	6.24 x 10 <sup>-4</sup>	805 (26.07)	1,414 (27.45)	0.87 (0.79-0.97)	0.0439
	C/C	223 (2.91)	336 (3.57)	0.79 (0.66-0.95)		81 (2.62)	130 (2.52)	0.96 (0.72-1.27)	
<i>P</i> <sub>HWE</sub> 0.0834									
rs917997 <b>2q12</b> <i>IL18RAP</i>	T	3,524 (22.09)	4,304 (22.21)	0.98 (0.93-1.03)	0.416	1,045 (20.73)	1,548 (21.48)	0.87 (0.78-0.96)	8.35 x 10 <sup>-3</sup>
	C/C	4,806 (60.24)	5,887 (60.75)	1.00 (reference)		1,441 (64.33)	2,210 (61.35)	1.00 (reference)	
	C/T	2,819 (35.33)	3,302 (34.07)	1.04 (0.97-1.10)	0.0122	697 (31.12)	1,236 (34.31)	0.83 (0.74-0.93)	6.50 x 10 <sup>-3</sup>
	T/T	353 (4.42)	502 (5.18)	0.83 (0.72-0.96)		102 (4.55)	156 (4.33)	0.89 (0.69-1.15)	
<i>P</i> <sub>HWE</sub> 0.214									
rs6441961 <b>3p21</b> <i>CCR3</i>	A	5,116 (32.09)	5,836 (30.13)	1.09 (1.04-1.14)	3.40 x 10 <sup>-4</sup>	1,734 (30.42)	2,433 (30.65)	1.04 (0.95-1.13)	0.386
	G/G	3,691 (46.30)	4,731 (48.85)	1.00 (reference)		1,247 (49.52)	1,920 (48.39)	1.00 (reference)	
	A/G	3,445 (43.21)	4,072 (42.04)	1.08 (1.01-1.15)	0.00151	1,031 (40.95)	1,663 (41.91)	1.05 (0.95-1.16)	0.677
	A/A	836 (10.49)	882 (9.11)	1.20 (1.08-1.33)		240 (9.53)	385 (9.70)	1.07 (0.88-1.29)	
<i>P</i> <sub>HWE</sub> 0.924									
rs17810546 <b>3q25</b> <i>IL12A</i>	G	1,918(12.33)	1,802 (12.31)	1.00 (0.93-1.07)	0.96	N/A			
	A/A	5,987 (76.95)	5,633 (76.89)	1.00 (reference)					
	A/G	1,667 (21.43)	1,582 (21.59)	1.00 (0.92-1.08)	0.926				
	G/G	126 (1.62)	111 (1.52)	1.04 (0.80-1.34)					
<i>P</i> <sub>HWE</sub> 0.867									

rs9811792	G	6,992 (45.13)	6,496 (44.34)	1.04 (0.99-1.08)	0.147	N/A			
<b>3q25</b>	A/A	2,320 (29.95)	2,270 (30.99)	1.00 (reference)					
<i>IL12A</i>	A/G	3,859 (49.82)	3,615 (49.34)	1.04 (0.97-1.13)	0.335				
	G/G	1,567 (20.23)	1,441 (19.67)	1.07 (0.97-1.17)					
$P_{\text{HWE}} 0.964$									
rs1464510	T	7,280 (45.07)	8,556 (45.16)	1.00 (0.95-1.04)	0.820	N/A			
<b>3q28</b>	G/G	2,462 (30.49)	2,814 (29.67)	1.00 (reference)					
<i>LPP</i>	G/T	3,948 (48.89)	4,776 (50.35)	0.95 (0.89-1.02)	0.241				
	T/T	1,666 (20.63)	1,895 (19.98)	1.00 (0.92-1.09)					
$P_{\text{HWE}} 0.0980$									
rs6822844	T	1,361 (16.49)	1,293 (17.56)	0.95 (0.89-1.00)	0.0559	N/A			
<b>4q27</b>	G/G	5,669 (68.69)	6,672 (67.63)	1.00 (reference)					
<i>IL2-IL21</i>	T/G	2,364 (28.64)	2,922 (29.62)	0.94 (0.85-0.98)	0.0854				
	T/T	220 (2.67)	272 (2.76)	0.98 (0.76-1.73)					
$P_{\text{HWE}} 0.364$									
rs1738074	T	3,340 (41.42)	4,268 (43.65)	0.92 (0.88-0.96)	$7.90 \times 10^{-5}$	2,757 (40.36)	4,261 (41.98)	0.86 (0.80-0.92)	$2.71 \times 10^{-5}$
<b>6q25</b>	C/C	2,743 (34.02)	3,126 (31.97)	1.00 (reference)		1,048 (35.26)	1,721 (33.91)	1.00 (reference)	
<i>TAGAP</i>	C/T	3,962 (49.13)	4,767 (48.76)	0.95 (0.89-1.02)	$1.62 \times 10^{-4}$	1,439 (48.42)	2,447 (48.22)	0.88 (0.80-0.97)	$1.06 \times 10^{-4}$
	T/T	1,359 (16.85)	1,884 (19.27)	0.83 (0.76-0.91)		485 (16.32)	907 (17.87)	0.72 (0.62-0.84)	
$P_{\text{HWE}} 0.377$									
rs3184504	A	7,311 (54.99)	7,208 (48.51)	1.28 (1.22-1.35)	$2.72 \times 10^{-24}$	N/A			
<b>12q24</b>	G/G	1,362 (20.49)	1,951 (26.26)	1.00 (reference)					
<i>SH2B3</i>	A/G	3,259 (49.03)	3,750 (50.47)	1.35 (1.25-1.47)	$8.04 \times 10^{-24}$				
	A/A	2,026 (30.48)	1,729 (23.27)	1.63 (1.52-1.81)					

$P_{\text{HWE}} 0.537$				
------------------------	--	--	--	--

**Supplementary Table 1:** Type 1 Diabetes Case-Control and Family Association Results for SNPs followed up from Celiac Disease.

OR = odds ratio for the minor allele, 95% CI = 95% confidence interval, RR = relative risk. We report a 1-df test  $P$ -value for allelic effects and a 2-df test  $P$ -value for genotype effects.  $P_{\text{HWE}}$  = Hardy Weinberg equilibrium in controls.

SNP region gene	Minor-allele Genotypes	Type 1 Diabetes Case Control Results				Type 1 Diabetes Family Results				Combined <i>P</i> 1-df 2-df
		frequency (%) cases	frequency (%) controls	OR (95% CI)	<i>P</i> -values 1-df 2-df	frequency (%) cases	frequency (%) parents	RR (95% CI)	<i>P</i> -values 1-df 2-df	
rs333 <b>3p21</b> <i>CCR5</i>	del	1625 (10.33)	2309 (11.89)	0.85 (0.80-0.92)	5.72 x 10 <sup>-6</sup>	309 (10.07)	503 (10.68)	0.90 (0.75-1.09)	0.12	1.49 x 10 <sup>-6</sup>
	ins/ins	6,320 (80.23)	7,554 (77.83)	1.00 (reference)		2,496 (80.65)	1,887 (80.13)	1.00 (reference)		
	ins/del	1,487 (18.88)	1,994 (20.54)	0.89 (0.82-0.96)	1.88 x 10 <sup>-6</sup>	572 (18.48)	433 (18.39)	0.97 (0.86-1.10)	9.10 x 10 <sup>-3</sup>	
	del/del	70 (0.89)	158 (1.63)	0.54 (0.40-0.72)		27 (0.87)	35 (1.49)	0.53 (0.34-0.82)		
<i>P</i> <sub>HWE</sub> 0.0514										
rs1799864 <b>3p21</b> <i>CCR2</i>	A	1159 (7.47)	564 (7.68)	0.97 (0.89-1.06)	0.51	N/A				
	G/G	6,638 (85.62)	6,122 (85.31)	1.00 (reference)						
	A/G	1,071 (13.81)	1,006 (14.02)	0.99 (0.90-1.08)	0.56					
	A/A	44 (0.57)	48 (0.67)	0.80 (0.53-1.21)						
<i>P</i> <sub>HWE</sub> 0.457										

**Supplementary Table 2:** Type 1 Diabetes Association Results for *CCR5* and *CCR2*.

OR = odds ratio for the minor allele, 95% CI = 95% confidence intervals, RR = relative risk. N/A not attempted

We report a 1-df test *P*-value for allelic effects and a 2-df test *P*-value for genotype effects. *P*<sub>HWE</sub> = Hardy Weinberg equilibrium in controls.

SNP region gene	Minor-allele Genotypes	Type 1 Diabetes Case Control Results				Celiac Disease Case Control Results			
		Genotype Frequency % cases	Genotype Frequency % controls	OR (95% CI)	<i>P</i> -values 1-df 2-df	Genotype Frequency % cases	Genotype Frequency % controls	OR (95% CI)	<i>P</i> -values 1-df 2-df
rs2476601 <b>1p13</b> <i>PTPN22</i>	T	2,647(17.78)	1,335 (9.53)	2.05 (1.90-2.20)	1.13 x 10 <sup>-88</sup>	543 (10.64)	1,63 (9.56)	1.09 (0.98-1.22)	0.130
	C/C	5,038 (67.60)	5,722 (81.74)	1.00 (reference)		2,037 (79.85)	5,397 (81.70)	1.00 (reference)	
	C/T	2,161 (29.04)	1,221 (17.44)	2.01 (1.85-2.18)	1.73 x 10 <sup>-87</sup>	485 (19.01)	1,155 (17.48)	1.07 (0.95-1.21)	0.254
	T/T	243 (3.27)	57 (0.81)	4.76 (3.55-6.40)		29 (1.14)	54 (0.82)	1.38 (0.85-2.21)	
<i>P</i> <sub>HWE</sub> 0.357									
rs1990760 <b>2q24</b> <i>IFIH1</i>	G	4,997 (35.20)	6,140 (38.94)	0.86 (0.82-0.90)	2.13 x 10 <sup>-10</sup>	1,992 (39.67)	5,725 (38.99)	1.02 (0.95-1.09)	0.547
	A/A	3,002 (42.29)	2,964 (37.60)	1.00 (reference)		895 (35.64)	2,756 (37.54)	1.00 (reference)	
	A/G	3,197 (45.03)	3,698 (46.91)	0.85 (0.79-0.92)	1.72 x 10 <sup>-9</sup>	1,240 (49.38)	3,447 (46.95)	1.09 (0.99-1.22)	0.168
	G/G	900 (12.68)	1,221 (15.49)	0.73 (0.66-0.81)		376 (14.97)	1,139 (15.51)	1.00 (0.87-1.16)	
<i>P</i> <sub>HWE</sub> 0.229									
rs3087243 <b>2q33</b> <i>CTLA4</i>	A	4,331 (40.45)	8,276 (45.17)	0.82 (0.78-0.86)	1.27 x 10 <sup>-14</sup>	2,086 (41.08)	7,798 (45.14)	0.85 (0.80-0.90)	1.26 x 10 <sup>-6</sup>
	G/G	1,876 (35.05)	2,773 (30.27)	1.00 (reference)		884 (34.82)	2,616 (30.29)	1.00 (reference)	
	A/G	2,623 (49.00)	4,500 (49.12)	0.86 (0.79-0.93)	4.42 x 10 <sup>-14</sup>	1,224 (48.21)	4,244 (49.14)	0.85 (0.77-0.94)	7.93 x 10 <sup>-6</sup>
	A/A	854 (15.95)	1,888 (20.61)	0.67 (0.60-0.74)		431 (16.98)	1,777 (20.57)	0.72 (0.63-0.83)	
<i>P</i> <sub>HWE</sub> 0.458									
rs333 <b>3p21</b> <i>CCR5</i>	del	1,625 (10.33)	2,309 (11.89)	0.85 (0.80-0.91)	5.87 x 10 <sup>-6</sup>	479 (9.54)	2,173 (11.80)	0.79 (0.71-0.88)	9.18 x 10 <sup>-6</sup>
	ins/ins	6,320 (80.23)	7,554 (77.84)	1.00 (reference)		2,051 (81.71)	7,176 (77.97)	1.00 (reference)	
	ins/del	1,487 (18.88)	1,994 (20.54)	0.89 (0.82-0.96)	1.93 x 10 <sup>-6</sup>	439 (17.49)	1,883 (20.46)	0.81 (0.72-0.92)	2.50 x 10 <sup>-5</sup>
	del/del	70 (0.89)	158 (1.63)	0.54 (0.40-0.72)		20 (0.80)	145 (1.58)	0.48 (0.30-0.78)	
<i>P</i> <sub>HWE</sub> 0.0514									

rs2069763 <b>4q27</b> <i>IL2-IL21</i>	T	5,706 (35.77)	6,274 (32.91)	1.13 (1.08-1.18)	1.28 x 10 <sup>-7</sup>	1,761 (34.58)	6,274 (32.91)	1.08 (1.01-1.16)	0.0180
	G/G	3,274 (41.05)	4,278 (44.89)	1.00 (reference)		1,102 (43.28)	4,278 (44.89)	1.00 (reference)	
	G/T	3,698 (46.36)	4,232 (44.40)	1.13 (1.06-1.21)	8.67 x 10 <sup>-7</sup>	1,127 (44.27)	4,232 (44.40)	1.03 (0.93-1.13)	
	T/T	1,004 (12.59)	1,021 (10.71)	1.27 (1.15-1.41)		317 (12.45)	1,021 (10.71)	1.24 (4.60-7.64)	0.02
<i>P</i> <sub>HWE</sub> 0.593									
rs6897932 <b>5p13</b> <i>IL7R</i>	A	3,023 (25.07)	3,955 (27.40)	0.89 (0.84-0.94)	4.13 x 10 <sup>-4</sup>	1,297 (25.37)	3,955 (27.40)	0.91 (0.84-0.97)	7.23 x 10 <sup>-3</sup>
	G/G	3,382 (56.10)	3,805 (52.72)	1.00 (reference)		1,415 (55.30)	3,805 (52.72)	1.00 (reference)	
	A/G	2,269 (37.64)	2,869 (39.75)	0.89 (0.83-0.96)	8.07 x 10 <sup>-4</sup>	989 (38.65)	2,869 (39.75)	0.94 (0.85-1.03)	
	A/A	377 (6.25)	543 (7.52)	0.78 (0.68-0.91)		155 (6.06)	543 (7.52)	0.77 (0.64-0.93)	0.0162
<i>P</i> <sub>HWE</sub> 0.232									
rs11755527 <b>6q15</b> <i>BACH2</i>	G	7,866 (49.54)	8,835 (46.45)	1.13 (1.09-1.18)	8.57 x 10 <sup>-9</sup>	2,504 (49.06)	8,528 (46.53)	1.10 (1.03-1.18)	2.78 x 10 <sup>-3</sup>
	C/C	2,088 (26.30)	2,658 (27.95)	1.00 (reference)		660 (25.86)	2,556 (27.89)	1.00 (reference)	
	C/G	3,836 (48.32)	4,869 (51.20)	1.01 (0.94-1.09)	4.37 x 10 <sup>-11</sup>	1,280 (50.16)	4,688 (51.16)	1.08 (0.96-1.20)	
	G/G	2,015 (25.38)	1,983 (20.85)	1.30 (1.19-1.42)		612 (23.98)	1,920 (20.95)	1.22 (1.07-1.39)	9.64 x 10 <sup>-3</sup>
<i>P</i> <sub>HWE</sub> 0.0143									
rs947474 <b>10p15</b> <i>PRKCQ</i>	G	2,654 (17.11)	3,549 (18.74)	0.88 (0.83-0.93)	1.48 x 10 <sup>-5</sup>	882 (17.29)	3,415 (18.77)	0.90 (0.83-0.98)	0.0178
	A/A	5,327 (68.68)	6,257 (66.07)	1.00 (reference)		1,747 (68.48)	6,004 (66.00)	1.00 (reference)	
	A/G	2,204 (28.42)	2,877 (30.38)	0.89 (0.83-0.95)	7.90 x 10 <sup>-5</sup>	726 (28.46)	2,771 (30.46)	0.91 (0.81-1.00)	
	G/G	225 (2.90)	336 (3.55)	0.76 (0.64-0.91)		78 (3.06)	322 (3.54)	0.82 (0.63-1.07)	0.0603
<i>P</i> <sub>HWE</sub> 0.917									
rs12722495 <b>10p15</b> <i>IL2RA</i>	G	1,120 (7.21)	1,624 (11.27)	0.62 (0.57-0.68)	1.74 x 10 <sup>-30</sup>	595 (12.02)	1,542 (11.29)	1.06 (0.95-1.17)	0.316
	A/A	6,684 (86.11)	5,667 (78.62)	1.00 (reference)		1,919 (77.50)	5,362 (78.56)	1.00 (reference)	
	A/G	1,036 (13.35)	1,458 (20.23)	0.61 (0.56-0.67)	1.82 x 10 <sup>-29</sup>	519 (20.96)	1,384 (20.28)	1.03 (0.92-1.16)	
	G/G	42 (0.54)	83 (1.15)	0.45 (0.31-0.65)		38 (1.53)	79 (1.16)	1.30 (0.86-1.95)	0.425
<i>P</i> <sub>HWE</sub> 0.318									

rs11594656	A	3,104 (22.17)	4,335 (24.56)	0.87 (0.83-0.93)	$2.03 \times 10^{-6}$	1,188 (23.39)	4,054 (24.47)	0.94 (0.87-1.01)	0.0921
<b>10p15</b>	T/T	4,218 (60.25)	5,014 (56.82)	1.00 (reference)		1,501 (59.12)	4,715 (56.93)	1.00 (reference)	
<i>IL2RA</i>	A/T	2,462 (35.17)	3,285 (37.23)	0.89 (0.83-0.95)		888 (34.97)	3,080 (37.19)	0.90 (0.81-0.99)	
	A/A	321 (4.59)	525 (5.95)	0.74 (0.64-0.86)	$9.16 \times 10^{-6}$	150 (5.91)	487 (5.88)	0.96 (0.78-1.18)	0.0864
$P_{HWE}$ 0.670									
rs689	A	2,085 (15.13)	5,735 (29.25)	0.42 (0.41-0.46)	$8.93 \times 10^{-195}$	1,456 (28.58)	5,735 (29.25)	0.95 (0.89-1.03)	0.201
<b>11p15</b>	T/T	5,038 (73.10)	4,877 (49.75)	1.00 (reference)		1,305 (51.24)	4,877 (49.75)	1.00 (reference)	
<i>INS</i>	A/T	1,623 (23.55)	4,117 (42.00)	0.38 (0.35-0.40)		1,028 (40.36)	4,117 (42.00)	0.91 (0.83-1.00)	
	A/A	231 (3.35)	809 (8.25)	0.27 (0.23-0.32)	$1.86 \times 10^{-202}$	214 (8.40)	809 (8.25)	0.97 (0.82-1.15)	0.166
$P_{HWE}$ 0.146									
rs2292239	A	5,675 (40.66)	5,408 (35.23)	1.31 (1.22-1.41)	$5.79 \times 10^{-22}$	1,804 (35.94)	5,408 (35.23)	1.02 (0.96-1.10)	0.498
<b>12q13</b>	C/C	2,433 (34.86)	3,230 (42.08)	1.00 (reference)		1,033 (41.16)	3,230 (42.08)	1.00 (reference)	
<i>ERBB3</i>	A/C	3,417 (48.96)	3,484 (45.39)	1.27 (1.21-1.33)		1,150 (45.82)	3,484 (45.39)	1.03 (0.93-1.14)	
	A/A	1,129 (16.18)	962 (12.53)	1.64 (1.39-1.92)	$3.49 \times 10^{-21}$	327 (13.03)	962 (12.53)	1.04 (0.90-1.21)	0.785
$P_{HWE}$ 0.635									
rs3184504	A	7,311 (54.99)	7,208 (48.51)	1.28 (1.22-1.35)	$2.72 \times 10^{-24}$	2,658 (52.34)	6,710 (48.67)	1.15 (1.08-1.23)	$2.85 \times 10^{-5}$
<b>12q24</b>	G/G	1,362 (20.49)	1,951 (26.26)	1.00 (reference)		708 (27.88)	1,618 (23.47)	1.00 (reference)	
<i>SH2B3</i>	A/G	3,259 (49.03)	3,750 (50.47)	1.35 (1.25-1.47)	$8.04 \times 10^{-24}$	1,242 (48.92)	3,474 (50.40)	1.22 (1.09-1.37)	$7.62 \times 10^{-5}$
	A/A	2,026 (30.48)	1,729 (23.27)	1.63 (1.52-1.81)		589 (23.20)	1,801 (26.13)	1.33 (1.15-1.52)	
$P_{HWE}$ 0.537									
rs3825932	C	4,555 (28.68)	6,050 (31.77)	0.86 (0.82-0.90)	$4.62 \times 10^{-10}$	1,705 (33.37)	5,820 (31.82)	1.07 (1.00-1.14)	0.0559
<b>15q24</b>	T/T	4,009 (50.49)	4,429 (46.52)	1.00 (reference)		1,149 (44.97)	4,248 (46.45)	1.00 (reference)	
<i>CTSH</i>	C/T	3,307 (41.65)	4,134 (43.42)	0.88 (0.83-0.94)		1,107 (43.33)	3,976 (43.47)	1.03 (0.94-1.14)	
	C/C	624 (7.86)	958 (10.06)	0.72 (0.64-0.80)	$2.07 \times 10^{-9}$	299 (11.70)	922 (10.08)	1.18 (1.02-1.38)	0.0946

$P_{HWE}$ 0.884									
rs12708716	G	3,522 (30.63)	4,393 (35.07)	0.81 (0.77-0.86)	$3.19 \times 10^{-13}$	1,830 (36.45)	3,990 (34.87)	1.06 (0.99-1.14)	0.120
<b>16p13</b>	A/A	2,789 (48.51)	2,642 (42.19)	1.00 (reference)		1,033 (41.16)	2,433 (42.52)	1.00 (reference)	
<i>CLEC16A</i>	A/G	2,398 (41.71)	2,847 (45.46)	0.79 (0.73-0.86)	$1.96 \times 10^{-12}$	1,124 (44.78)	2,588 (45.23)	1.00 (0.90-1.11)	
	G/G	562 (9.78)	773 (12.34)	0.68 (0.60-0.77)		353 (14.06)	701 (12.25)	1.17 (1.00-1.36)	0.116
$P_{HWE}$ 0.887									
rs478582	C	4,592 (40.82)	5,484 (44.93)	0.83 (0.79-0.88)	$8.83 \times 10^{-12}$	2,149 (43.24)	4,991 (44.88)	0.93 (0.87-1.00)	0.0408
<b>18p11</b>	T/T	1,959 (34.83)	1,841 (30.17)	1.00 (reference)		788 (31.71)	1,686 (30.32)	1.00 (reference)	
<i>PTPN2</i>	C/T	2,740 (48.71)	3,040 (49.81)	0.83 (0.76-0.90)		1,245 (50.10)	2,757 (49.59)	0.96 (0.85-1.07)	
	C/C	926 (16.46)	1,222 (20.02)	0.69 (0.62-0.77)	$7.68 \times 10^{-11}$	452 (18.19)	1,117 (20.09)	0.86 (0.74-0.99)	0.102
$P_{HWE}$ 0.606									
rs45450798	G	3,102 (20.17)	2,538 (16.59)	1.28 (1.21-1.36)	$1.15 \times 10^{-16}$	946 (19.00)	2,417 (16.62)	1.18 (1.08-1.30)	$2.61 \times 10^{-4}$
<b>18p11</b>	C/C	4,894 (63.64)	5,329 (69.68)	1.00 (reference)		1638 (65.60)	5,064 (69.66)	1.00 (reference)	
<i>PTPN2</i>	C/G	2,490 (32.38)	2,100 (27.46)	1.30 (1.21-1.39)		772 (30.92)	1,995 (27.44)	1.20 (1.07-1.33)	
	G/G	306 (3.98)	219 (2.86)	1.60 (1.33-1.92)	$1.00 \times 10^{-15}$	87 (3.48)	211 (2.90)	1.35 (1.03-1.78)	$1.2 \times 10^{-3}$
$P_{HWE}$ 0.486									
rs763361	A	8,141 (49.74)	8,818 (47.14)	1.14 (1.07-1.19)	$1.56 \times 10^{-8}$	2,491 (49.09)	8,341 (47.16)	1.09 (1.01-1.16)	0.0137
<b>18q22</b>	G/G	2,050 (25.31)	2,603 (27.83)	1.00 (reference)		639 (25.19)	2,463 (27.85)	1.00 (reference)	
<i>CD226</i>	A/G	4,041 (49.90)	4,682 (50.06)	1.14 (1.06-1.23)		1,305 (51.44)	4,421 (49.99)	1.4 (1.02-1.27)	
	A/A	2,008 (24.79)	2,068 (22.11)	1.28 (1.18-1.39)	$1.09 \times 10^{-7}$	593 (23.37)	1,960 (22.16)	1.17 (1.03-1.34)	0.0275
$P_{HWE}$ 0.666									
rs3788013	A	7,447 (46.54)	8,281 (43.27)	1.13 (1.08-1.18)	$3.09 \times 10^{-8}$	2,464 (45.42)	8,281 (43.27)	1.09 (1.02-1.16)	$8.88 \times 10^{-3}$
<b>21q22</b>	C/C	2,282 (28.52)	3,046 (31.84)	1.00 (reference)		749 (29.68)	3,046 (31.86)	1.00 (reference)	
<i>UBASH3A</i>	A/C	3,989 (49.86)	4,763 (49.78)	1.12 (1.04-1.20)		1,260 (49.92)	4,763 (49.78)	1.08 (0.97-1.19)	
	A/A	1,729 (21.61)	1,759 (18.38)	1.28 (1.17-1.40)	$2.07 \times 10^{-7}$	515(20.40)	1,759 (18.38)	1.19 (1.05-1.35)	0.0457

$P_{\text{HWE}}$ 0.172									
rs229541	C	7,209 (45.53)	8,125 (42.77)	1.12 (1.07-1.17)	$6.96 \times 10^{-7}$	2,168 (42.41)	7,815 (42.79)	0.97 (0.91-1.04)	0.423
<b>22q13</b>	T/T	2,340 (29.56)	3,118 (32.82)	1.00 (reference)		850 (33.26)	2,989 (32.73)	1.00 (reference)	
<i>CIQTNF6</i>	C/T	3,943 (49.81)	4,637 (48.82)	1.12 (1.05-1.20)	$4.40 \times 10^{-6}$	1,244 (48.67)	4,471 (48.96)	0.97 (0.88-1.08)	0.725
	C/C	1,633 (20.63)	1,744 (18.36)	1.24 (1.14-1.36)		462 (18.08)	1,672 (18.31)	0.95 (0.83-1.08)	
$P_{\text{HWE}}$ 0.783									

**Supplementary Table 3:** Celiac Disease Association Results for Type 1 Diabetes Associated SNPs.

OR = odds ratio for the minor allele, 95% CI = 95% confidence intervals.

We report a 1-df test  $P$ -value for allelic effects and a 2-df test  $P$ -value for genotype effects.  $P_{\text{HWE}}$  = Hardy Weinberg equilibrium in controls.

SNP	<i>PTPN2</i> -T1D Association			add SNP to rs45450798	add rs45450798 to SNP
	OR	95% CI	1df <i>P</i>	1df <i>P</i>	1df <i>P</i>
rs45450798	1.316	1.22 - 1.42	2.29x 10 <sup>-12</sup>	-	-
rs1893217	1.301	1.21 - 1.41	1.09x 10 <sup>-11</sup>	0.570	0.073

**Supplementary Table 4:** *PTPN2* Association Results in Type 1 Diabetes Case-Control Collection

Single locus test results and regression analysis *P* values for a completely typed data set of 4,996 case and 4,359 control samples for the *PTPN2* SNPs rs45450798 and rs1893217.

OR = odds ratio for the minor allele, 95% CI = 95% confidence intervals.

SNP	<i>IL2RA</i> -T1D Association			add SNP to rs12722495	add rs12722495 to SNP
	OR	95%CI	1df <i>P</i>	1df <i>P</i>	1df <i>P</i>
rs12722495	0.626	0.57-0.68	$9.13 \times 10^{-27}$	-	-
rs41295061	0.641	0.59-0.70	$8.86 \times 10^{-23}$	0.395	$1.15 \times 10^{-5}$

**Supplementary Table 5:** *IL2RA* Association Results in Type 1 Diabetes Case-Control Collection

Single locus test results and regression analysis *P* values for a completely typed data set of 6,658 T1D case and 6,983 control samples for the *IL2RA* SNPs rs41295061 and rs12722495. We reported previously two independent T1D associations within the *IL2RA* region marked by the SNPs rs41295061 and rs11594656.(ref<sup>23</sup>) Genotyping of additional T1D and control samples for *IL2RA* SNPs identified rs12722495, a SNP in high linkage disequilibrium with rs41295061 ( $r^2 = 0.79$ ) as the most associated SNP in the region. Logistic regression analysis revealed that rs12722495 added to the association observed at rs41295061. In the reverse analysis rs41295061 did not add to the association at rs12722495. OR = odds ratio for the minor allele, 95% CI = 95% confidence intervals.

rs1893217 <i>PTPN2</i>		Frequency % cases	Frequency % controls	OR 95% CI	<i>P</i> values 1-df/2-df
Dutch	G	189 (18.64)	296 (16.67)	1.15 (0.94-1.40)	0.188
	A/A	334 (65.88)	619 (69.71)	1.00 (reference)	0.324
	A/G	157 (30.97)	242 (27.25)	1.20 (0.94-1.53)	
	G/G	16 (3.16)	27 (3.04)	1.10 (0.58-2.06)	
Irish	G	180 (21.63)	365 (19.07)	1.16 (0.96-1.43)	0.129
	A/A	261 (62.74)	628 (65.62)	1.00 (reference)	0.170
	A/G	130 (31.25)	293 (30.62)	1.07 (0.82-1.37)	
	G/G	25 (6.01)	36 (3.76)	1.67 (0.98-2.83)	
Combined					0.0452 0.127

**Supplementary Table 6:** Association Results for *PTPN2* rs1893217 in the Dutch and Irish Celiac Case-Control Collections.

OR = odds ratio for the minor allele, 95% CI = 95% confidence intervals.

We report a 1-df test *P*-value for allelic effects and a 2-df test *P*-value for genotype effects.

The 416 Irish cases and 957 Irish controls; and 507 Dutch cases and 888 Dutch controls were genotyped using Illumina GoldenGate.(ref <sup>1</sup>) for rs1893217, previously reported as the most associated SNP in T1D(ref <sup>6</sup>).

## References

1. Hunt KA, Zhernakova A, Turner G, et al. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat Genet* 2008.
2. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 2007;81:559-75.
3. Clayton DG, Walker NM, Smyth DJ, et al. Population structure, differential bias and genomic control in a large-scale, case-control association study. *Nat Genet* 2005;37:1243-6.
4. Wellcome Trust Case Control Consortium. Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls. *Nature* 2007;447:661-87.
5. Cordell HJ, Clayton DG. 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* 2002;70:124-41.
6. Todd JA, Walker NM, Cooper JD, et al. Robust associations of four new chromosome regions from genome-wide analyses of type 1 diabetes. *Nat Genet* 2007;39:857-64.
7. Cooper JD, Smyth DJ, Smiles AM, et al. Meta-analysis of genome-wide association study data identifies additional type 1 diabetes loci. *Nat Genet* 2008:in press.
8. Concannon P, Onengut-Gumuscu S, Todd JA, et al. A human type 1 diabetes susceptibility locus maps to chromosome 21q22.3. *Diabetes* 2008.
9. Thomas DC, Clayton DG. Betting odds and genetic associations. *J Natl Cancer Inst* 2004;96:421-3.
10. Wacholder S, Chanock S, Garcia-Closas M, El Ghormli L, Rothman N. Assessing the probability that a positive report is false: an approach for molecular epidemiology studies. *J Natl Cancer Inst* 2004;96:434-42.
11. Sharma RB, Alegria JD, Talor MV, Rose NR, Caturegli P, Burek CL. Iodine and IFN-gamma synergistically enhance intercellular adhesion molecule 1 expression on NOD.H2h4 mouse thymocytes. *J Immunol* 2005;174:7740-5.
12. Irie J, Wu Y, Wicker LS, et al. NOD.c3c4 congenic mice develop autoimmune biliary disease that serologically and pathogenetically models human primary biliary cirrhosis. *J Exp Med* 2006;203:1209-19.
13. Bour-Jordan H, Thompson HL, Bluestone JA. Distinct effector mechanisms in the development of autoimmune neuropathy versus diabetes in nonobese diabetic mice. *J Immunol* 2005;175:5649-55.
14. Cucca F, Lampis R, Congia M, et al. A correlation between the relative predisposition of MHC class II alleles to type 1 diabetes and the structure of their proteins. *Hum Mol Genet* 2001;10:2025-37.
15. Koeleman BP, Lie BA, Undlien DE, et al. Genotype effects and epistasis in type 1 diabetes and HLA-DQ trans dimer associations with disease. *Genes Immun* 2004;5:381-8.
16. Nejentsev S, Howson JM, Walker NM, et al. Localization of type 1 diabetes susceptibility to the MHC class I genes *HLA-B* and *HLA-A*. *Nature* 2007;450:887-92.
17. Erlich H, Valdes AM, Noble J, et al. HLA DR-DQ Haplotypes and Genotypes and Type 1 Diabetes Risk: Analysis of the Type 1 Diabetes Genetics Consortium Families. *Diabetes* 2008.
18. Mazzilli MC, Ferrante P, Mariani P, et al. A study of Italian pediatric celiac disease patients confirms that the primary HLA association is to the DQ(alpha 1\*0501, beta 1\*0201) heterodimer. *Hum Immunol* 1992;33:133-9.

19. van Belzen MJ, Koeleman BP, Crusius JB, et al. Defining the contribution of the HLA region to cis DQ2-positive coeliac disease patients. *Genes Immun* 2004;5:215-20.
20. Karell K, Louka AS, Moodie SJ, et al. HLA types in celiac disease patients not carrying the DQA1\*05-DQB1\*02 (DQ2) heterodimer: results from the European Genetics Cluster on Celiac Disease. *Hum Immunol* 2003;64:469-77.
21. Maki M, Mustalahti K, Kokkonen J, et al. Prevalence of Celiac disease among children in Finland. *N Engl J Med* 2003;348:2517-24.
22. Tollefsen S, Arentz-Hansen H, Fleckenstein B, et al. HLA-DQ2 and -DQ8 signatures of gluten T cell epitopes in celiac disease. *J Clin Invest* 2006;116:2226-36.
23. Lowe CE, Cooper JD, Brusko T, et al. Large-scale genetic fine mapping and genotype-phenotype associations implicate polymorphism in the IL2RA region in type 1 diabetes. *Nat Genet* 2007;39:1074-82.