

Published in final edited form as:

Nat Genet. 2007 July ; 39(7): 827–829.

A genome-wide association study for celiac disease identifies risk variants in the region harboring *IL2* and *IL21*

David A van Heel¹, Lude Franke^{2,17}, Karen A Hunt^{1,17}, Rhian Gwilliam^{3,17}, Alexandra Zhernakova², Mike Inouye³, Martin C Wapenaar⁴, Martin CNM Barnardo⁵, Graeme Bethel³, Geoffrey KT Holmes⁶, Con Feighery⁷, Derek Jewell⁸, Dermot Kelleher⁷, Parveen Kumar¹, Simon Travis⁹, Julian RF Walters¹⁰, David S Sanders¹¹, Peter Howdle¹², Jill Swift¹³, Raymond J Playford¹, William M McLaren³, M Luisa Mearin^{14,15}, Chris J Mulder¹⁶, Ross McManus⁷, Ralph McGinnis³, Lon R Cardon⁸, Panos Deloukas³, and Cisca Wijmenga^{2,4}

¹Centre for Gastroenterology, Institute of Cell and Molecular Science, Queen Mary University of London, London E1 2AT, UK ²Complex Genetics Section, DBG-Dept Biomedical Genetics, University Medical Center Utrecht, 3584 CG Utrecht, The Netherlands ³Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, UK ⁴Genetics Department, University Medical Center Groningen, 9700 RB Groningen, The Netherlands. ⁵Transplant Immunology, Oxford Transplant Centre, Churchill Hospital, Oxford OX3 7LJ, UK ⁶Department of Gastroenterology, Derbyshire Royal Infirmary, London Road, Derby DE1 2QY, UK. ⁷Departments of Clinical Medicine and Immunology, Institute of Molecular Medicine, Trinity College Dublin, Dublin, Ireland. ⁸Wellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK. ⁹Gastroenterology Unit, University of Oxford, Oxford OX3 7BN, UK. ¹⁰Gastroenterology Section, Imperial College London, Hammersmith Hospital, London W12 0HS, UK ¹¹Department of Gastroenterology and Liver Unit, Royal Hallamshire Hospital, Sheffield S10 2JF, UK. ¹²Department of Gastroenterology, St James's University Hospital, Leeds LS9 7TF, UK ¹³Department of Gastroenterology, Llandough Hospital, Penarth CF64 2XX, UK. ¹⁴Department of Paediatric Gastroenterology, Leiden University Medical Center, 2300 RC Leiden, The Netherlands. ¹⁵Department of Pediatric Gastroenterology, VU Medical Center, 1007 MB Amsterdam, The Netherlands. ¹⁶Department of Gastroenterology, VU Medical Center, 1007 MB Amsterdam, The Netherlands.

Abstract

We tested 310,605 single-nucleotide polymorphisms for association in 778 celiac disease cases and 1422 controls. Outside the HLA, the most significant finding (rs13119723, $P=2.0 \times 10^{-7}$, empirical genome-wide significance $P=0.045$) was in the *KIAA1109/Tenr/IL2/IL21* linkage disequilibrium block. Association was independently confirmed in two further collections (strongest at rs6822844, 24kB 5' of *IL21*, meta-analysis $P=1.3 \times 10^{-14}$, OR 0.63), suggesting genetic variation in this region predisposes to celiac disease.

Celiac disease is a common (1% prevalence) small intestinal inflammatory condition induced by dietary wheat, rye, and barley. However despite high heritability (estimated at 87% from twin studies¹), no non-HLA genetic factors have been identified and convincingly replicated. The majority of celiacs possess HLA-DQ2 (the remainder mostly HLA-DQ8²), and how HLA-DQ2 presents cereal peptides to intestinal T cells is understood³. However HLA-DQ2 is common in healthy individuals, demonstrating it is necessary but not sufficient for disease development.

Correspondence should be addressed to D.A.vH (d.vanheel@qmul.ac.uk)..

¹⁷these authors contributed equally

We therefore designed a genome wide association study to identify predisposing genetic factors in celiac disease. We genotyped samples with Illumina BeadChips (Supplementary Methods). After quality control, association analysis was performed on 310,605 SNPs with minor allele frequency >1% genotyped in 778 UK celiac cases and 1422 UK population controls (Supplementary Table 1). Overall SNP call rate was 99.87%. Single SNP association statistics are presented in Supplementary Figure 1.

Highly significant association was seen around the HLA locus, as expected. Association was strongest at rs2187668, which maps to the first intron of HLA-DQA1 ($\chi^2=769.1$, $P<10^{-19}$, frequency of A allele in controls 13.8%, cases 53.1%, odds ratio (OR) 7.04 [95% CI 6.08 - 8.15]). When compared with classical HLA typing (Supplementary Methods), the rs2187668-A allele efficiently tagged HLA-DQ2.5*cis* ($r^2=0.97$, Supplementary Table 2). HLA-DQ2.5*cis* is the most common HLA-DQ2 haplotype associated with celiac disease, where the two chains of the DQ2 heterodimer are encoded on the same chromosome. One or two copies of HLA-DQ2.5*cis* (inferred by rs2187668 genotype) were present in 89.2% of UK celiac patients versus 25.5% of population controls. In order to identify other HLA predisposing variants occurring in the presence, or absence, of HLA-DQ2.5*cis*, we performed further analyses stratified by rs2187668 genotype. In cases (n=558) and controls (n=331) of rs2187668-AG genotype, peak association was seen at rs9357152 ($P=5.2 \times 10^{-14}$); and in cases (n=83) and controls (n=1059) of rs2187668-GG genotype peak association was seen at rs9275141 ($P=3.9 \times 10^{-16}$). Numbers of rs2187668-AA cases (n=31) were too small for analysis. The finding that rs2187668, rs9275141 and rs9357152 map within or adjacent to HLA-DQA1 and -DQB1 underpins the critical role of HLA-DQ2/8 in antigen presentation in celiac disease.

Outside the HLA region, we observed more significant SNPs than would be expected by chance, with 56 SNPs showing association at $P<10^{-4}$ (Supplementary Table 3). Many of these SNPs are in close proximity, suggesting that some of the excess in low p-value SNPs might be due to true disease associations among multiple SNPs in linkage disequilibrium with nearby disease variants. We therefore prioritised these findings for rapid replication (interim results in Supplementary Table 3), whilst designing a more extensive SNP replication study. We noted weak evidence for association in the previously reported region CD28-CTLA4-ICOS⁴ (rs4675374 $P=0.007$, rs11681040 $P=0.008$) but not the MYO9B⁵ region.

The most significant (non-HLA) finding was rs13119723 ($P=2.0 \times 10^{-7}$, frequency of G allele in controls 15.8%, cases 10.1%). Permutation of affection status labels demonstrated genome-wide significance (in 9 of 200 ($P=0.045$) permutations the most significant permuted P value was $\leq 2.0 \times 10^{-7}$). The location of rs13119723 close to *IL2* and *IL21* made it a highly plausible celiac disease candidate gene. We did not observe any evidence for statistical interaction between rs13119723 genotype and inferred HLA-DQ2.5*cis* genotype ($P>0.20$). We then confirmed association of rs13119723 with celiac disease in two separate collections (Table 1). The G allele of rs13119723 was more common in controls in each collection, and meta-analysis (all 4,680 samples) established highly significant disease association at rs13119723 ($P=4.8 \times 10^{-11}$).

The rs13119723 SNP maps to a region of strong linkage disequilibrium (Supplementary Fig. 2). We had genotyped 27 SNPs in this 4q27 region, extending ~480kb from rs6835946 to rs6840978. In addition to rs13119723, four other SNPs showed association with celiac disease at $P<10^{-4}$ in the UK dataset (Fig. 1, Table 1). We further genotyped rs6822844, rs13151961 and rs6840978 (all strongly correlated with rs13119723, Fig 1.) in the Dutch and Irish collections and replicated the UK dataset associations (Table 1, Supplementary Table 4). The strongest association overall was seen at rs6822844, approximately 24kb 5' of *IL21* (meta-analysis $P=1.3 \times 10^{-14}$, OR 0.63 [0.57 - 0.71]).

Markers on the HumanHap300 BeadChip (Illumina) are haplotype tag SNPs. We found that the 27 SNPs genotyped in the UK collection highly efficiently captured the common genetic variation in the ~480kb region (161 of 165 common phase I+II HapMap SNPs pairwise tagged at $r^2 > 0.8$ in CEU population⁶, Supplementary Methods). Genotyping of further markers in the UK collection was therefore unlikely to contribute substantial additional information.

Finer analysis of haplotype structure in the ~480kb region in the UK collection showed subdivision into two closely correlated ~439kb and ~40kb haplotype blocks (using strict criteria⁷). The rs13119723-G allele was found on a single strongly associated haplotype in both blocks (Supplementary Fig. 2), with haplotype frequencies in the 439kb block of 10.1% in cases and 15.3% in controls ($P = 2.1 \times 10^{-6}$), and in the 40kb block of 16.3% in cases and 21.5% in controls ($P = 4.3 \times 10^{-5}$). We genotyped 10 additional SNPs to tag >5% frequency haplotypes (in addition to the 4 SNPs already tested) in the Dutch and Irish collections, and found similar haplotype structure and association across all three populations (Supplementary Table 4). Due to extensive linkage disequilibrium, these analyses did not enable us to determine the causal variant associated with celiac disease in the 4q27 region. The population specific genetic variance at the associated 4q27 markers (CEU HapMap data) is relatively high, suggesting possible selection in the Northern European population.

The 4q27 celiac disease associated region contains three known protein coding genes NP_640336.1/*Tenr*, *IL2*, *IL21* and a predicted gene of unknown function (*KIAA1109*/Q6ZS70). We manually annotated the human genome sequence in the region (not shown), but did not identify further genes. IL-2, secreted in an autocrine fashion by antigen-stimulated T cells, is a key cytokine for T cell activation and proliferation. Another T cell derived cytokine, IL-21, enhances B, T and NK cell proliferation and interferon- γ production. Both cytokines are implicated in the mechanisms of other intestinal inflammatory diseases^{8,9}. Expression profiles for the four genes across multiple cell/tissue types were examined in the GNF SymAtlas database (Supplementary Methods). *Tenr* is specifically testis expressed, and is an unlikely candidate for the causal celiac disease susceptibility gene. The function of *KIAA1109* is largely unknown¹⁰, although *KIAA1109* is widely expressed as multiple splice variants in multiple tissue types. We specifically looked at gene expression in duodenal tissue from normal and celiac disease individuals (with normal histology, and with villous atrophy). *Tenr* expression levels were mostly undetectable. No differences were seen between normal and treated celiac individuals for *KIAA1109*, *IL2* or *IL21*. In the presence of inflammation (untreated celiac disease), *KIAA1109* and *IL2* levels showed a modest reduction and *IL21* an increase (Supplementary Fig. 3). The syntenic mouse region to human 4q27 (*Idd3*) determines susceptibility to multiple autoimmune diseases in the NOD mouse model, by a mechanism influencing IL-2 mRNA/protein levels and CD4⁺ CD25⁺ regulatory T cell activity¹¹. However further studies are required to determine the human celiac disease susceptibility gene in this region.

Our genome wide association study has identified genetic variation in a linkage disequilibrium block encompassing the *KIAA1109*/*Tenr*/*IL2*/*IL21* genes as a novel susceptibility factor for celiac disease. In addition to further investigation of this 4q27 region, the next steps in dissecting the genetic causes of celiac disease include larger scale replication of other putative associations and additional genome-wide analyses (e.g. of copy number variation¹²).

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We acknowledge funding from Coeliac UK; the Coeliac Disease Consortium (an innovative cluster approved by the Netherlands Genomics Initiative and partly funded by the Dutch Government, grant BSIK03009); the Netherlands Genomics Initiative (grant 050-72-425); the Netherlands Organization for Scientific Research (grant 901-04-219); the Science Foundation Ireland and the Wellcome Trust (GR068094MA Clinician Scientist Fellowship to D.A.vH; New Blood Fellowship to R.M.; support for the work of R.McG and P.D.). The authors acknowledge use of DNA from the British 1958 Birth Cohort collection, funded by the UK Medical Research Council grant G0000934 and the Wellcome Trust grant 068545/Z/02.

We thank C.A. Mein and the Barts and The London Genome Centre for advice and genotyping support; D.Simpkin, T. Dibling and C. Hand for genotyping (Sanger Institute); M.J. Caulfield for advice on study design; D.P. Kelsell for comments on the manuscript; D. Strachan and W.L. McArdle for 1958 Birth Cohort samples; A. Monsuur for patient recruitment, G. Meijer and J. Meijer for histology review, K. Duran for DNA extraction, H. van Someren for clinical database management (the Netherlands); A. Ryan, G. Turner, M. Abuzakouk, N. Kennedy, F. Stevens and C. O'Morain for patient and control recruitment and sample management (Ireland). We thank J. Loveland for EST annotation and checking. We thank the Wellcome Trust Centre for Human Genetics, University of Oxford for provision of computing facilities. We thank all celiac patients and controls for participating in this study.

References

1. Nistico L, et al. *Gut* 2006;55:803–8. [PubMed: 16354797]
2. Karel K, et al. *Hum Immunol* 2003;64:469–77. [PubMed: 12651074]
3. van Heel DA, West J. *Gut* 2006;55:1037–46. [PubMed: 16766754]
4. Hunt KA, et al. *Eur J Hum Genet* 2005;13:440–4. [PubMed: 15657618]
5. Monsuur AJ, et al. *Nat Genet* 2005;12:1341–4. [PubMed: 16282976]
6. Altshuler D, et al. *Nature* 2005;437:1299–320. [PubMed: 16255080]
7. Gabriel SB, et al. *Science* 2002;296:2225–9. [PubMed: 12029063]
8. Sadlack B, et al. *Cell* 1993;75:253–61. [PubMed: 8402910]
9. Monteleone G, et al. *Gastroenterology* 2005;128:687–94. [PubMed: 15765404]
10. He QY, et al. *Bioinformatics* 2006;22:2189–91. [PubMed: 16632497]
11. Yamanouchi J, et al. *Nat Genet.* 2007
12. Redon R, et al. *Nature* 2006;444:444–54. [PubMed: 17122850]

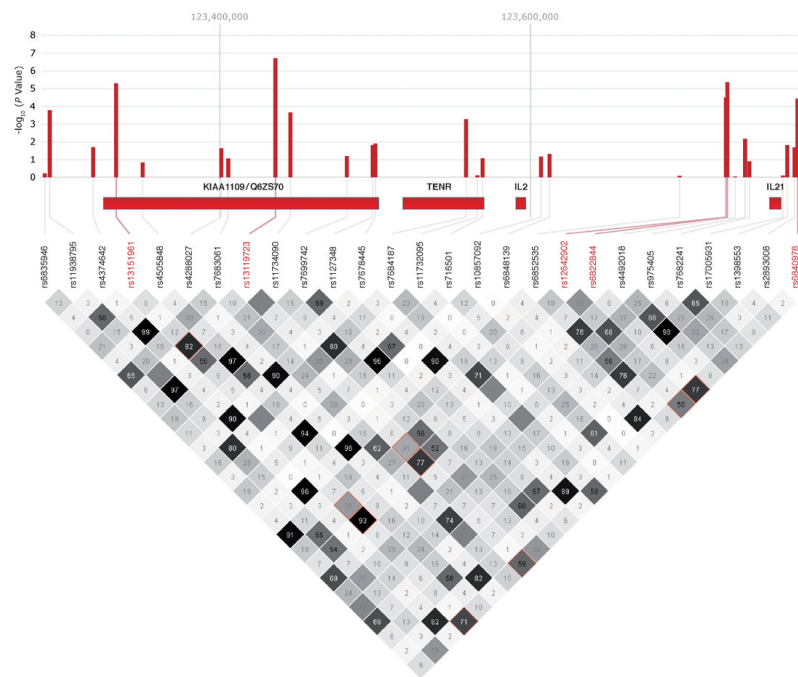


Figure 1. Analysis of chromosome 4q27 region around rs13119723

A ~480kb region between rs6835946 and rs6840978 is shown (build 36 map), with single SNP allelic association test P values, genes, and linkage disequilibrium statistics (r^2) determined from the UK dataset.

Chromosome 4q27 markers in UK genome wide association scan, and replication collections.

	UK genome wide association scan collection				Dutch collection		Irish collection		Meta-analysis	
	Allele Freq. Cases	Allele Freq. Controls	P value ¹	Allele Freq. Cases	Allele Freq. Controls	P value ¹	Allele Freq. Cases	Allele Freq. Controls	Odds Ratio [95% CI]	P value ¹
	n=778	n=1422		n=508	n=929		n=483	n=560		
rs6835946	30.3%	29.6%	0.63	25.1%	25.5%	0.81	31.4%	33.5%		0.32
rs11938795	21.2%	26.3%	0.00017							
rs4374642	10.5%	8.4%	0.020							
rs13151961	12.6%	17.9%	5.2 × 10 ⁻⁶	12.1%	8.4%	0.0015	8.3%	6.7%		0.17
rs4505848	34.4%	32.3%	0.15	12.2%	19.1%	2.2 × 10 ⁻⁶	14.8%	19.4%	0.65 [0.58 - 0.73]	1.3 × 10 ⁻¹²
rs4288027	10.4%	8.4%	0.02	37.6%	34.9%	0.15	31.3%	27.8%		0.084
rs7683061	40.6%	38.0%	0.090							
rs13119723	10.1%	15.8%	2.0 × 10 ⁻⁷	11.5%	16.4%	0.00042	12.8%	16.2%	0.66 [0.58 - 0.74]	4.8 × 10 ⁻¹¹
rs11734090	21.5%	26.5%	0.00023							
rs7699742	34.9%	32.2%	0.066							
rs1127348	24.8%	21.6%	0.016	25.7%	24.0%	0.34	21.7%	16.9%		0.0053
rs7678445	9.9%	7.7%	0.013							
rs7684187	25.2%	30.1%	0.00056							
rs11732095	8.8%	8.6%	0.79	9.9%	9.1%	0.45	8.5%	7.7%		0.50
rs716501	34.8%	32.2%	0.087	7.8%	7.3%	0.60	7.0%	7.7%		0.53
rs10857092	5.2%	6.6%	0.069							
rs6848139	10.2%	8.5%	0.050							
rs6852535	29.7%	29.4%	0.84							
rs12642902	28.5%	34.6%	3.3 × 10 ⁻⁵							
rs6822844	12.6%	17.9%	4.6 × 10 ⁻⁶	12.4%	18.5%	2.1 × 10 ⁻⁵	14.2%	19.7%	0.63 [0.57 - 0.71]	1.3 × 10 ⁻¹⁴
rs4492018	26.6%	26.5%	0.93	22.7%	23.1%	0.80	29.1%	30.2%		0.60
rs975405	38.6%	42.8%	0.0071	41.5%	43.3%	0.34	39.9%	43.9%		0.070
rs7682241	36.6%	34.4%	0.13							
rs17005931	26.5%	26.2%	0.82							
rs1398553	34.1%	30.6%	0.016	35.2%	32.6%	0.18	30.6%	25.7%		0.015
rs2893008	10.0%	8.0%	0.021	11.0%	7.4%	0.0011	7.3%	6.5%		0.50
rs6840978	16.3%	21.5%	3.8 × 10 ⁻⁵	15.2%	21.9%	2.0 × 10 ⁻⁵	19.2%	24.0%	0.70 [0.63 - 0.78]	1.1 × 10 ⁻¹⁰

¹ P values from χ^2 squared test of allele counts. All tests are two-tailed.