nature genetics

Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect

Alienke J Monsuur[1], Paul I W de Bakker[2], Behrooz Z Alizadeh[1], Alexandra Zhernakova[1], Marianna R Bevova[1], Eric Strengman[1], Lude Franke[1], Ruben van't Slot[1], Martine J van Belzen[1,7], Ineke C M Lavrijsen[1], Begoña Diosdado[1], Mark J Daly[2], Chris J J Mulder[3], M Luisa Mearin[4], Jos W R Meijer[5], Gerrit A Meijer[6], Erica van Oort[1], Martin C Wapenaar[1], Bobby P C Koeleman[1] & Cisca Wijmenga[1]

Celiac disease is probably the best-understood immune-related disorder. The disease presents in the small intestine and results from the interplay between multiple genes and gluten, the triggering environmental factor[1]. Although HLA class II genes explain 40% of the heritable risk, non-HLA genes accounting for most of the familial clustering have not yet been identified. Here we report significant and replicable association ($P = 2.1 \times 10^{-6}$) to a common variant located in intron 28 of the gene myosin IXB (*MYO9B*), which encodes an unconventional myosin molecule that has a role in actin remodeling of epithelial enterocytes[2,3]. Individuals homozygous with respect to the at-risk allele have a 2.3-times higher risk of celiac disease ($P = 1.55 \times 10^{-5}$). This result is suggestive of a primary impairment of the intestinal barrier in the etiology of celiac disease, which may explain why immunogenic gluten peptides are able to pass through the epithelial barrier.

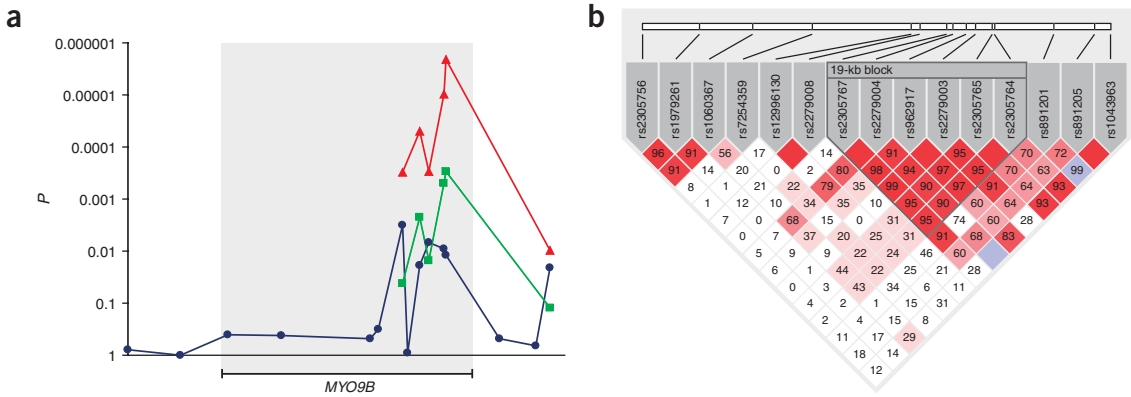

Figure 1 Fine-mapping study in and surrounding *MYO9B*. (**a**) The association data are plotted in blue (circles) on a logarithmic scale for each of the 15 SNPs across a 191-kb region (17,030,936–17,222,116 bp) encompassing *MYO9B* and tested in 216 individuals with Marsh III celiac disease and 216 controls (set 1). Subsequent follow-up studies in a second cohort of 247 affected individuals and 470 controls (set 2) of the six SNPs with $P < 0.05$ are plotted in green (squares). The association data from the combined cohorts (463 affected individuals and 686 controls) are plotted in red (triangles). The spacing between SNPs reflects the distances between them. (**b**) Pairwise LD between the 15 SNPs, given by the D' statistics computed with the genotype data from the 216 controls. The overall LD structure of these data is very similar to that of the data on a European population in the HapMap database (Centre d'Etude du Polymorphisme Humain; Utah residents with ancestry from northern and western Europe), where all associated SNPs fall into a single 22-kb haplotype block (defined as previously described[24]). The darker shade of red indicates the higher D'.

[1]Complex Genetics Section, DBG-Department of Medical Genetics, University Medical Center Utrecht, Universiteitsweg 100, 3584 CG Utrecht, the Netherlands. [2]Broad Institute of Harvard and MIT, 320 Charles St., Cambridge, Massachusetts 02141, USA; and Center for Human Genetics Research, Massachusetts General Hospital, 185 Cambridge Street, Boston, Massachusetts 02114, USA. [3]Department of Gastroenterology, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. [4]Department of Pediatrics, Leiden University Medical Center, P.O. Box 9600, 2300 RC Leiden, and VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. [5]Department of Pathology, Rijnstate Hospital, P.O. Box 9555, 6800 TA Arnhem, The Netherlands. [6]Department of Pathology, VU University Medical Center, P.O. Box 7057, 1007 MB Amsterdam, The Netherlands. [7]Present address: Laboratory of DNA Diagnostics, Leiden University Medical Center, P.O. Box 9503, 2300 RA Leiden, The Netherlands. Correspondence should be addressed to C.W. (t.n.wijmenga@med.uu.nl).

Received 29 June; accepted 27 September; published online 13 November 2005; doi:10.1038/ng1680

Table 1 Names, chromosomal locations and *P* values for the 15 SNPs tested in and around *MYO9B*

SNP	rs2305756	rs1979261	rs1060367	rs7254359	rs12996130	rs2279008	rs2305767	rs2279004	rs962917	rs2279003	rs2305765	rs2305764	rs891201	rs891205	rs1043963
Position (bp)	17,030,936	17,054,518	17,076,202	17,100,566	17,140,752	17,144,303	17,155,296	17,157,679	17,163,247	17,167,031	17,173,992	17,174,833	17,199,099	17,215,586	17,222,116
P															
Set 1	0.7836	0.9885	0.4085	0.4198	0.4741	0.3114	0.0032*	0.9049	0.0186	0.0069	0.0089	0.0121	0.4832	0.6614	0.0205
Set 2							0.0412		0.0022*	0.0151	0.0005*	0.0003*			0.1233
Set 1 and set 2							0.0003*		0.000049*	0.0003*	0.0000097*	0.0000021*			0.0095

P* values still significant after correcting for multiple testing. *P* values were calculated in Haploview using the χ^2 test. The locations of the SNPs in *MYO9B* are given in **Supplementary Figure 2.

With a prevalence close to 1% (ref. 4), celiac disease is the most common food intolerance in general western populations. In individuals with celiac disease, ingestion of gluten leads to inflammation and tissue remodeling of the intestinal mucosa, resulting in malnutrition and severe complications. Long regarded as a gastrointestinal disorder of childhood, the disease is now considered to be a chronic systemic autoimmune disease[5] and is more often diagnosed in adults than in children[6]. Celiac disease has a strong heritable component, although the inheritance is complex and multifactorial[1]. Gluten is the key environmental risk factor, and HLA class II genes are associated to celiac disease[7]. HLA-DQ2 has a key role in the disease by presenting gluten peptides to $CD4^+$ cells in the lamina propria[8]. Much of the research on celiac disease has focused on the activation and regulation of gluten-reactive T cells. But the etiological steps preceding this T-cell activation are still poorly understood; for example, why are the antigenic gluten peptides resistant to further breakdown in the intestinal lumen, and how do the gluten peptides pass through the epithelial barrier? The integrity of the intestinal barrier is impaired in active celiac disease[9,10], which implies that the epithelial cell barrier has a role in the early pathogenesis of the disease.

Genome-wide screens have been done to identify non-HLA genes involved in celiac disease but, so far, no gene has been positionally cloned. We previously obtained strong evidence for linkage (multiple maximum lod score 4.43, nominal $P = 6.2 \times 10^{-6}$) to chromosome 19p13.1 in celiac disease (*CELIAC4*) in a cohort of affected sibling pairs of European descent from the Netherlands[11]. This chromosomal location was also suggested by meta and pooled analyses of European celiac disease data that did not include the Dutch cohort[12]. In our original study[11], further association analysis was done with five microsatellite markers spanning the 3.5-Mb lod-1 interval in a cohort of 216 case-control pairs, resulting in weak association to *D19S899* ($P = 1.3 \times 10^{-3}$), located in intron 1 of the gene myosin IXB (*MYO9B*).

Here we focused on the region surrounding *D19S899* and, more specifically, on *MYO9B* using the same 216 cases and 216 controls (set 1; **Supplementary Fig. 1** and **Supplementary Table 1** online). Initially, we typed 15 SNPs with an average spacing of 13 kb across *MYO9B*. Using a conservative Bonferroni correction, a single SNP (rs2305767) located in intron 14 of *MYO9B* showed significant association ($P = 0.0032$; **Fig. 1a**, **Table 1** and **Supplementary Table 2** online), confirming our initial observation. Another five SNPs proximal to rs2305767 also showed significant association without correction ($P < 0.05$; **Table 1**); these SNPs allowed us to refine the region of association, as they were located in the 3′ part of *MYO9B*, with one SNP (rs1043963) located 36.4 kb downstream of the gene. A replication study of the six SNPs with $P < 0.05$ was done in a second, fully independent, cohort of 247 unrelated individuals with Marsh III celiac disease who met the same strict diagnostic criteria[11], and 470 Dutch blood bank donors (set 2). We observed significant association ($P < 0.05$) for all five SNPs in *MYO9B* but not for the SNP downstream of the gene (rs1043963, $P = 0.123$). Three SNPs (rs962917, rs2305765 and rs2305764) were still significant after correcting for six independent tests (nominal Bonferroni $P < 0.05/6 = 8 \times 10^{-3}$; **Fig. 1** and **Table 1**). Combining the cohorts (set 1 and set 2) strengthened the association considerably and showed highly significant association for all five SNPs in *MYO9B* (**Fig. 1** and **Table 1**). We observed the smallest *P* value for rs2305764 ($P = 2.1 \times 10^{-6}$) located in intron 28. Our data showed strong linkage disequilibrium (LD) between the five associated SNPs (**Fig. 1b**), which are located in a single, 19-kb haplotype block spanning exons 15–27 of *MYO9B* (**Supplementary Fig. 2** online).

To rule out the possibility that our initial study (comprising only five microsatellite markers[11]) had overlooked an association signal elsewhere in the *CELIAC4* locus, we carried out a comprehensive fine-mapping study using 359 tag SNPs with minor allele frequencies > 0.02 and $r^2 > 0.7$, and covering the lod-1.5 region (99% confidence interval (c.i.); 15,385,880–21,075,237 bp) on set 2 supplemented with the cases from set 1 (totaling 463 independent cases and 470 controls). After carrying out quality checks on the raw data, we excluded 60 poorly performing or monomorphic tag SNPs (**Supplementary Table 3** online), yielding 299 tag SNPs for further analysis, of which 19 were located in the 191-kb region covered previously. Three SNPs were overlapping in the two screens (rs7254359, rs2279008 and rs2305767; **Supplementary Table 2**). After excluding tag SNPs that showed deviation from Hardy-Weinberg equilibrium, we were able to analyze 291 tag SNPs successfully. These allowed us to capture most of the untyped genetic variation present in the Centre d'Etude du Polymorphisme Humain population of the HapMap Project (**Supplementary Table 3**). We observed a single, strong peak of association in *MYO9B* (**Fig. 2**, **Table 2** and **Supplementary Table 3**). The most-associated tag SNP (rs1457092, $P = 7.8 \times 10^{-5}$), located between rs962917 and rs2279003, remained significant after stringent Bonferroni correction for 291 tag SNPs ($P_c < 0.05/291 = 1.72 \times 10^{-4}$) and is in strong LD with the five associated SNPs identified in the random screen described above (rs2305767, rs962917, rs2279003, rs2305765

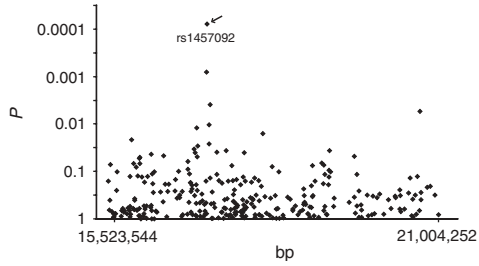

Figure 2 Comprehensive fine-mapping study of the *CELIAC4* region on 19p13.1. *P* values are plotted for all 291 tag SNPs within the 99% c.i. (15,385,880–21,075,237 bp, in chromosomal order) in a cohort of 463 independent individuals with Marsh III celiac disease and 470 controls. *P* values were calculated in Haploview using the χ^2 test. The *P* value cutoff after Bonferroni correction was 1.72×10^{-4}. Only SNP rs1457092, located in intron 20 of *MYO9B* and indicated with an arrow, was significant after correction for multiple testing.

Table 2 Names, chromosomal locations and *P* values for the 11 tag SNPs from the comprehensive screen located in *MYO9B* (17,072,928–17,185,703 bp)

SNP	rs7246865	rs4808571	rs7254359	rs1870068	rs10409461	rs11673417	rs2279009	rs2279008	rs7259292	rs2305767	rs1457092
Position (bp)	17,080,105	17,087,229	17,100,566	17,119,828	17,129,784	17,131,499	17,143,899	17,144,303	17,153,014	17,155,296	17,165,236
P	0.3022	0.4125	0.9009	0.4977	0.7905	0.6813	0.4592	0.1617	0.7726	0.0008	0.000078*

P* value was still significant after correcting for multiple testing. Details are shown in **Supplementary Table 3.

and rs2305764; **Supplementary Fig. 3** online). One of the three overlapping SNPs (rs2305767 located in intron 14) was already significant in the initial cohort ($P = 3.2 \times 10^{-3}$) and showed even stronger association in the combined cohorts (set 1 and set 2, comprising 463 cases and 686 controls; $P = 3 \times 10^{-4}$; **Table 1**).

Haplotype analysis with these six SNPs showed that in this 19-kb region, only four haplotypes account for >97% of all the observed haplotypes; these can be captured by typing only three SNPs (rs2305767, rs1457092 and rs2305764). The AAACTA haplotype was present in 39% of cases and 30% of controls (odds ratio (OR) = 1.56; 95% c.i. = 1.27–1.93, $P = 1.8 \times 10^{-5}$; **Table 3**). The GGCTCG haplotype is the most low-risk haplotype and is very unlikely to encompass the unknown causal disease variant. We were able to narrow the haplotype to a single SNP, as we observed that the presence of the A allele of SNP rs2305764 constantly predisposes to a higher risk of celiac disease, independent of the variation in the other five SNPs. This observation was confirmed by logistic regression analysis using a conditional forward selection procedure showing that only SNP rs2305764 was left in the regression model (data not shown) and, similarly, that none of the haplotypes showed stronger association. Therefore, the rs2305764 SNP located in intron 28 of *MYO9B* can alone completely explain the association observed at *MYO9B* and can be considered a marker for celiac disease risk. Sequence analysis of the entire *MYO9B* coding region in 16 affected individuals homozygous with respect to the at-risk haplotype identified 22 variants, 2 of which were already included in our study. Of the 20 new SNP variants, 2 were considered relevant (*i.e.*, nonsynonymous SNPs) and were typed in our entire cohort of 463 independent cases and 686 independent controls. Unfortunately, neither of these two SNPs turned out to be more strongly associated than rs2305764, which agrees with the observation that these two SNPs are not in LD with the at-risk haplotype (**Supplementary Table 4** and **Supplementary Fig. 4** online).

The genotype frequency data for rs2305764 showed that the risk of celiac disease rose in proportion to the number of copies of the A allele (frequency of 46.5% in affected individuals compared with 37.9% in controls), implying a codominant allele-dose effect. Individuals heterozygous with respect to the A allele have a modest but significantly higher risk of developing celiac disease (OR = 1.66; 95% c.i. = 1.23–2.13; $P = 5.3 \times 10^{-4}$), whereas individuals homozygous with respect to the A allele have a risk of developing celiac disease that increases to 2.27 (95% c.i. = 1.56–3.30; $P = 1.55 \times 10^{-5}$), with population-attributable risks of 25% and 23%, respectively. These results suggest that the 3′ part of *MYO9B* is associated with celiac disease.

MYO9B is a good candidate gene for celiac disease because of the function of its encoded protein, and it may be one of the long-sought factors involved in the early mucosal events preceding the well-understood inflammatory response. *MYO9B* encodes a single motor protein[2] belonging to the class IX myosin molecules, which are unique in comparison with other classes because they contain a Rho-GTPase-activating domain within their tails. This GTPase activity converts active Rho-GTP into inactive Rho-GDP, thereby downregulating Rho-dependent signaling pathways[3]. Rho-family GTPases are involved in remodeling of the cytoskeleton and tight junction assembly, both of which result in enhanced epithelial paracellular permeability[13,14]. It is therefore tempting to speculate that a genetic variant in the 3′ part of *MYO9B* leads to an impaired interaction with RhoA, thereby perturbing tight junction gate and fence function. Hence, a subtle, underlying intestinal barrier abnormality may be involved in the etiology of celiac disease, which is in line with the recent observation of intestinal permeability in individuals with celiac disease with normal histology[15,16]. As a consequence, immunogenic gluten peptides can enter the deeper mucosal layer more easily. Notably, this is the site at which the HLA-DQ2–mediated antigen presentation to the $CD4^+$ cells initiates the inflammatory response. So far, *MYO9B* is the only non-HLA gene identified for celiac disease by positional cloning. It will be interesting to determine the effect of this gene in other populations with celiac disease. The identification of *MYO9B* as a susceptibility gene in celiac disease is a notable finding that may open new avenues for studying the early events of celiac disease pathogenesis, a process that has not yet received much attention, but which might prove important in developing alternative treatments to the strict gluten-free diet currently used.

METHODS

Subjects and controls. DNA, isolated from whole blood, was available from two independent cohorts of Dutch individuals with celiac disease (set 1, 216; set

Table 3 The prevalence of *MYO9B* haplotypes reconstructed from selected SNPs and their association to celiac disease

MYO9B haplotypes										
rs2305767	rs962917	rs1457092	rs2279003	rs2305765	rs2305764	Cases (%)[a]	Controls (%)[a]	OR[b]	95% c.i.[b]	*P*[c]
G	G	C	T	C	G	308 (37.3)	565 (45.5)	1.00[d]	–	–
A	G	C	C	C	G	128 (15.5)	216 (17.4)	1.09	0.84–1.41	0.53
A	A	A	C	T	A	319 (38.7)	375 (30.2)	1.56	1.27–1.93	0.000018
A	G	C	C	T	A	46 (5.6)	67 (5.4)	1.26	0.84–1.88	0.26
Rare haplotypes						24 (2.9)	20 (1.6)	2.21	1.24–4.05	0.009

[a]Number represents the frequency of haplotypes estimated using an expectation maximization algorithm embedded log linear model[21,22]. [b]OR represents maximum likelihood estimate of odds ratio, and the corresponding 95% c.i. was approximated using Woolf's method. [c]*P* values were calculated using a χ^2 test. [d]This haplotype was taken as the reference.

2, 247). All the affected individuals were diagnosed in accordance with the revised ESPGHAN criteria[17]. More than 90% of the affected individuals were HLA-DQ2-positive (**Supplementary Table 1**). The initial biopsy specimens of the individuals were retrieved; all showed a Marsh III lesion upon reevaluation by one of two experienced pathologists (G.A.M. and J.W.R.M.). Both cohorts included children and adults. There were also two cohorts of controls available. Set 1 controls ($n = 216$) comprised random hospital controls. Set 2 controls ($n = 470$) were random blood bank donors. All cases and controls were from The Netherlands and of European descent, and at least three of their four grandparents were also born in The Netherlands. The comprehensive screen included all individuals in set 2 supplemented with the cases from set 1. **Supplementary Table 1** lists the baseline demographic parameters of the cohorts. All individuals gave their informed consent. This study was approved by the Medical Ethical Committee of the University Medical Center Utrecht.

Random screen: SNP selection and genotyping. We selected 15 random SNPs from Applied Biosystems, covering *MYO9B* and its surroundings. These SNPs were obtained as Assays on Demand or Assays by Design (Applied Biosystems) and initially typed on set 1 (216 individuals with Marsh III celiac disease and 216 controls) using a 7900 Taqman (Applied Biosystems). In the follow-up study, six SNPs were selected for further typing on set 2 (247 individuals with Marsh III celiac disease and 470 controls).

Comprehensive screen: tag SNP selection and genotyping. For the comprehensive screen we took the 99% c.i. (1.5-lod) of the *CELIAC4* linkage peak on chromosome 19, ranging from 15,385,880 to 21,075,237 bp (National Center for Biotechnology Information build 34). SNPs were selected by downloading all the SNPs typed in the Centre d'Etude du Polymorphisme Humain (Utah residents with ancestry from northern and western Europe) population in this region from the HapMap database[18]. From these SNPs, the program Tagger[19] was used to select tag SNPs so that all SNPs with a minor allele frequency $\geq 2\%$ were captured with $r^2 \geq 0.7$ (excluding SNPs with low Illumina quality design scores). A final set of 359 tag SNPs was obtained for genotype analysis.

We genotyped SNPs using the GoldenGate assay on an Illumina BeadStation 500 GX (Illumina). All tag SNPs were examined for their resulting quality; all those that were not polymorphic in our population, had a low signal or had clusters that were too wide were excluded (**Supplementary Table 3**). We also carried out an evaluation of the tagging efficiency for two reasons. First, several tag SNPs were dropped because of low-quality or bad clusters, because they were not in Hardy-Weinberg equilibrium in the controls, or because they were not polymorphic in our population. Second, in the meantime, HapMap had expanded the number of SNPs typed in the region. We evaluated the tagging efficiency of the 291 working tag SNPs using the following parameters (**Supplementary Table 3**): aggressive tagging was used to get the most information; all the SNPs in the region were used; lod score threshold was set to 1; number of iterations was set to 1,000; maximum number of tries was set to 100,000.

Statistical analysis. We calculated association χ^2 and two-tailed P values using the Haploview program[20], for each stage of the study (set 1, set 2, set 1 and set 2, and the comprehensive screen). SNPs that were not in Hardy-Weinberg equilibrium ($P < 0.001$) in the controls were excluded from further analysis. We used multiple logistic regression analysis to estimate allelic and genotypic OR and the corresponding 95% c.i. for SNP rs2306764. To obtain genotype information for the combined cohorts on the six SNPs comprising this 19-kb region and showing $P < 10^{-3}$ (rs2305767, rs962917, rs1457092, rs2279003, rs2305765 and rs2305764), the associated tag SNP (rs1457092) was also typed for the 216 controls from set 1. As parental information was missing, an expectation maximization algorithm was used to estimate haplotype frequencies in a multiple locus system[21,22]. This function recast the regression model as a generalized linear mixed model with random effects, which is fitted as part of the maximization step and allows testing for LD and disease association[21,22]. The OR and the corresponding 95% c.i. were calculated for all haplotypes. Population attributable risk was calculated for rs2306764 using the corresponding genotype frequencies of this marker as explained previously[23]. All the analyses were done using STATA statistical software, version 8.0 for MS Windows.

URLs. The website for Applied Biosystems is http://myscience.appliedbiosystems.com/. The program Tagger is available at http://www.broad.mit.edu/mpg/tagger/. The HapMap database and the Haploview program are available at http://www.hapmap.org/.

Accession codes. GenBank: *Homo sapiens* chromosome 19 complete sequence, NC_000019; *H. sapiens MYO9B* mRNA, NM_004145. Ensembl: transcript *MYO9B*, ENST00000319396.

Note: Supplementary information is available on the Nature Genetics website.

ACKNOWLEDGMENTS

We thank the affected individuals, their physicians and the Dutch Celiac Disease Foundation for participating in this study; C. Gouw, A. Bardoel and H. van Someren for practical assistance and J. Senior for critically reading the manuscript. This study was supported by grants from The Netherlands Organization for Scientific Research, the Dutch Digestive Disease Foundation and the Celiac Disease Consortium.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Van Heel, D.A., Hunt, K., Greco, L. & Wijmenga, C. Genetics in coeliac disease. *Best Pract. Res. Clin. Gastroenterol.* **19**, 323–339 (2005).
2. Post, P.L. *et al.* Myosin-IXb is a single-headed and processive motor. *J. Biol. Chem.* **277**, 11679–11683 (2002).
3. Muller, R.T., Honnert, U., Reinhard, J. & Bahler, M. The rat myosin myr 5 is a GTPase-activating protein for Rho *in vivo*: essential role of arginine 1695. *Mol. Biol. Cell* **8**, 2039–2053 (1997).
4. Dube, C. *et al.* The prevalence of celiac disease in average-risk and at-risk Western European populations: a systematic review. *Gastroenterology* **128**, S57–S67 (2005).
5. Rewers, M. Epidemiology of celiac disease: what are the prevalence, incidence, and progression of celiac disease? *Gastroenterology* **128**, S47–S51 (2005).
6. Green, P.H. The many faces of celiac disease: clinical presentation of celiac disease in the adult population. *Gastroenterology* **128**, S74–S78 (2005).
7. Falchuk, Z.M. & Strober, W. HL-A antigens and adult coeliac disease. *Lancet* **2**, 1310 (1972).
8. Arentz-Hansen, H. *et al.* The intestinal T cell response to alpha-gliadin in adult celiac disease is focused on a single deamidated glutamine targeted by tissue transglutaminase. *J. Exp. Med.* **191**, 603–612 (2000).
9. Schulzke, J.D., Schulzke, I., Fromm, M. & Riecken, E.O. Epithelial barrier and ion transport in coeliac sprue: electrical measurements on intestinal aspiration biopsy specimens. *Gut* **37**, 777–782 (1995).
10. Van Elburg, R.M., Uil, J.J., Mulder, C.J. & Heymans, H.S. Intestinal permeability in patients with coeliac disease and relatives of patients with coeliac disease. *Gut* **34**, 354–357 (1993).
11. Van Belzen, M.J. *et al.* A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* **125**, 1032–1041 (2003).
12. Babron, M.C. *et al.* Meta and pooled analysis of European coeliac disease data. *Eur. J. Hum. Genet.* **11**, 828–834 (2003).
13. Bruewer, M., Hopkins, A.M., Hobert, M.E., Nusrat, A. & Madara, J.L. RhoA, Rac1, and Cdc42 exert distinct effects on epithelial barrier via selective structural and biochemical modulation of junctional proteins and F-actin. *Am. J. Physiol. Cell Physiol.* **287**, C327–C335 (2004).
14. Matter, K. & Balda, M.S. Signalling to and from tight junctions. *Nat. Rev. Mol. Cell Biol.* **4**, 225–236 (2003).
15. Smecuol, E. *et al.* Permeability, zonulin production, and enteropathy in dermatitis herpetiformis. *Clin. Gastroenterol. Hepatol.* **3**, 335–341 (2005).
16. Uil, J.J. *et al.* Follow-up of treated coeliac patients: sugar absorption test and intestinal biopsies compared. *Eur. J. Gastroenterol. Hepatol.* **8**, 219–223 (1996).
17. United European Gastroenterology. When is a coeliac a coeliac? Report of a working group of the United European Gastroenterology Week in Amsterdam, 2001. *Eur. J. Gastroenterol. Hepatol.* **13**, 1123–1128 (2001).
18. The International HapMap Consortium. The International HapMap Project. *Nature* **426**, 789–796 (2003).
19. De Bakker, P.I. *et al.* Efficiency and power in genetic association studies. *Nat. Genet.*, advance online publication 23 October 2005 (doi:10.1038/ng1669).
20. Barrett, J.C., Fry, B., Maller, J. & Daly, M.J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263–265 (2005).
21. Chiano, M.N. & Clayton, D.G. Fine genetic mapping using haplotype analysis and the missing data problem. *Ann. Hum. Genet.* **62**, 55–60 (1998).
22. Bitti, P.P. *et al.* Association between the ancestral haplotype HLA A30B18DR3 and multiple sclerosis in central Sardinia. *Genet. Epidemiol.* **20**, 271–283 (2001).
23. Greenland, S. & Rothman, R.K. Measures of effect and measures of association. in *Modern Epidemiology* 2nd edn. 47–67 (Lippincott-Raven, Philadelphia, 1998).
24. Gabriel, S.B. *et al.* The structure of haplotype blocks in the human genome. *Science* **296**, 2225–2229 (2002).