HAL
open science

Multiple independent variants in 6q21-22 associated with susceptibility to celiac disease in the Dutch, Finnish and Hungarian populations

Elisabet Einarsdottir, Marianna R Bevova, A Zhernakova, Alienke Monsuur, Lotta L. E. Koskinen, Ruben Vant Slot, Chris J.J. Mulder, M. Luisa Mearin, Ilma Korponay-Szabo, Katri Kaukinen, et al.

To cite this version:

Elisabet Einarsdottir, Marianna R Bevova, A Zhernakova, Alienke Monsuur, Lotta L. E. Koskinen, et al.. Multiple independent variants in 6q21-22 associated with susceptibility to celiac disease in the Dutch, Finnish and Hungarian populations. European Journal of Human Genetics, 2011, 10.1038/ejhg.2011.2 . hal-00614742

HAL Id: hal-00614742

https://hal.science/hal-00614742v1

Submitted on 16 Aug 2011

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Title

Multiple independent variants in 6q21-22 associated with susceptibility to celiac disease in the Dutch, Finnish and Hungarian populations

Running title: Celiac disease susceptibility at 6q21-22 in European populations

Authors:

Elisabet Einarsdottir 1*, Marianna R. Bevova 2,3*, Alexandra Zhernakova 2, 4 , Alienke Monsuur 2 , Lotta L.E. Koskinen 1, Ruben van't Slot 2, Chris Mulder 5, M. Luisa Mearin 6, Ilma R. Korponay-Szabo 7, Katri Kaukinen 8, Kalle Kurppa 9 Juha Kere 10, Markku Mäki 9, Cisca Wijmenga 11, Päivi Saavalainen 1

* Equal author contribution

Affiliations

1. Department of Medical Genetics and Research Program for Molecular Medicine, Biomedicum Helsinki, P.O. Box 63, FIN-00014 University of Helsinki, Finland

2. Department of Medical Genetics, University Medical Centre Utrecht, The Netherlands.

3. Department of Clinical Genetics, VU University Medical Center, Amsterdam, The Netherlands

4. Department of Rheumatology, Leiden University Medical Center, The Netherlands

5. Department of Gastroenterology, VU University Medical Center, Amsterdam, The Netherlands

6. Department of Paediatric Gastroenterology, Leiden University Medical Centre, Leiden, The Netherlands.

7. Heim Pal Children's Hospital, H-1089 Budapest and University of Debrecen, H-4032 Debrecen, Hungary

8. Department of Gastroenterology and Alimentary Tract Surgery, Tampere University Hospital and Medical School, University of Tampere, Tampere, Finland

9. Paediatric Research Centre, University of Tampere Medical School and Tampere University Hospital, FIN-33014 University of Tampere, Finland

10. Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, and Clinical Research Center, Karolinska University Hospital, Huddinge, Sweden

11. Department of Genetics, University Medical Center Groningen, and University of Groningen, P.O. Box 30001, 9700 RB Groningen, the Netherlands

Corresponding author

Elisabet Einarsdottir, Ph.D.
Department of Medical Genetics and
Research Program for Molecular Medicine
Biomedicum Helsinki
P.O. Box 63
FIN-00014 University of Helsinki
Finland
Tel +358-9-19125107
Fax +358-9-19125624
E-mail elisabet.einarsdottir@helsinki.fi

Abstract

Celiac disease is an inflammatory enteropathy caused by intolerance to gluten. Previous linkage studies in the Dutch, Finnish and Hungarian populations have revealed a locus on chromosome 6q21-22 conferring susceptibility to celiac disease. This locus has previously been implicated in susceptibility to other autoimmune diseases such as Crohn's disease and type 1 diabetes. We performed finemapping on 446 independent individuals with celiac disease and 641 controls of Dutch origin, testing 872 tagging SNPs in a 22 Mb region of chromosome 6. The twelve most promising SNPs were followed up in 2071 individuals from 284 Finnish and 357 Hungarian celiac disease families to identify risk variants in this region. Multiple markers in the region were significantly associated with celiac disease in the Dutch material. Two SNPs, rs9391227 and rs4946111, were significantly associated with celiac disease in the Finnish population. The association to rs9391227 represents the strongest association signal found in the Finnish (p=0.0032, OR 0.66) as well as the combined Dutch, Finnish and Hungarian populations (p=3.6*10-5, OR 0.76). Rs9391227 is situated downstream of the HECT domain and ankyrin repeat containing, E3 ubiquitin protein ligase 1 (*HACE1*) gene and is contained within a region of strong linkage disequilibrium enclosing *HACE1*. Two additional, independent, susceptibility variants in the 6q21-22 region were also found in a meta-analysis of the three populations. The 6q21-22 region was confirmed as a celiac disease susceptibility locus and harbours multiple independent associations, some of which may implicate ubiquitin-pathways in celiac disease susceptibility.

Keywords:

Celiac, linkage, association, genome-wide, 6q21-22

Introduction

Celiac disease is a chronic autoimmune disorder of the small intestine with autoimmune features. It is triggered in genetically susceptible individuals by dietary exposure to gluten, a protein present in wheat, barley and rye. Ingestion of gluten by celiac disease patients leads to flattening of the duodenal mucosa, a gradual process classified according to Marsh [(1)]. With a prevalence above 1% [(2)], celiac disease is the most common food intolerance in general western populations. Long regarded as a gastrointestinal disorder of childhood, the disease is now considered to be a chronic systemic autoimmune disease and is more and more often diagnosed in adults [(3)]. Celiac disease is a multifactorial disorder and both environmental factors, such as gluten, and a multigenic predisposition are needed to explain the aetiology of the disease. Concordance rates for celiac disease have been shown to be higher in monozygous twins (86%) than in dizygous twins (20%), indicating a strong genetic component in the development of celiac disease [(4)]. Celiac disease aggregates in families with an approximate 10% recurrence risk for siblings of celiac disease patients. There is, however, no clear Mendelian inheritance of the disease in families, suggesting that the genetic predisposition is derived from the presence and interaction of more than one susceptibility gene.

Over 90% of celiac disease patients express the HLA-DQ2 heterodimer, encoded by *DQA1*0501* and *DQB1*02*. A causal role for HLA-DQ2 has been demonstrated as this molecule can present the gluten-derived peptides to T-cells [(5)]. A further 5% of celiac disease patients carry HLA-DQ8. Although HLA-DQ2 or DQ8 seem to be necessary for celiac disease development, they are not sufficient, as HLA-DQ2 or

DQ8 alone can only explain some 40% of the genetic variation underlying celiac disease. This implies that other, non-*HLA* genes, are also important determinants of disease aetiology.

A recent whole-genome association scan identified a number of genetic variants involved in celiac disease susceptibility [6]. However, all the known risk variants explain not more than 50% of the genetic variance of celiac disease, indicating that novel genes and novel risk variants remain to be identified. This was supported by the results of the second whole-genome association study in celiac disease [7]. This study confirmed a number of the previously identified risk variants, as well as identifying numerous novel variants.

A study in the Dutch population obtained evidence for linkage of celiac disease to the 6q21-22 region, approximately 70 cM downstream of the *HLA* region [8]. The region of interest spans approximately 102 to 122 Mb and is clearly distinct from the *HLA* region as well as from the previously identified *TNFAIP3* and *TAGAP* celiac susceptibility genes [9, 10]. An independent study by Einarsdottir *et al.* (unpublished) of susceptibility genes for celiac disease in Finnish and Hungarian families also showed linkage to 6q16-22, to a region overlapping with the linkage found in the Dutch population. Taken together, these results indicate the possibility of a genetic factor within the chromosome 6q21-22 region that is involved in susceptibility to celiac disease in multiple populations. We therefore performed a fine-mapping scan in Dutch unrelated individuals with celiac disease in order to identify the possible candidate gene. This study was further followed up in family datasets from the Finnish and Hungarian populations.

Material and methods

Patients and controls

Dutch cohort:

DNA from whole blood was available from a cohort of Dutch individuals with celiac disease. All the affected individuals were diagnosed in accordance with the revised ESPGHAN criteria [(11)]. The control cohort comprising random blood bank donors and has been described previously [(6)]. All 446 Dutch cases and 641 controls were from the Netherlands and of European descent, and at least three of their four grandparents were also born in The Netherlands. The sample set has been described previously [(12)]. This study was approved by the Medical Ethical Committee of the University Medical Centre Utrecht.

Finnish and Hungarian cohorts:

The Finnish families and cases were collected at the University of Tampere and have been described previously [(13)]. The collection of the Hungarian samples has also been described previously [(14)]. A total of 1001 individuals from 284 Finnish families (19 unknown phenotype, 359 healthy, 623 affected) and 1070 individuals from 357 Hungarian families (7 unknown phenotype, 471 healthy, 592 affected) were included for genotyping. All sample collections were approved by the ethical committees of the Tampere and Helsinki University Hospitals, Finnish National Public Health Institute, Heim Pal Children's Hospital, Budapest, and the University of Debrecen. All

147 participants were informed about the study according to the study protocol and gave
148 their written informed consent.
149

150 ***Marker selection and genotyping***

151 For a comprehensive follow-up screen in an independent Dutch dataset, we took the
152 lod-1 region (95% confidence interval; ranging from rs9390679 (101,261,700 bp) to
153 rs11154152 (123,694,693 bp) (National Center for Biotechnology Information
154 NCBI34/hg16) of the previously identified linkage peak on chromosome 6 [(15)]. SNPs
155 for fine-mapping were selected by downloading all the SNPs typed in the CEPH
156 (Centre d'Etude du Polymorphisme Humain) population (Utah residents with ancestry
157 from northern and western Europe) that were located in the genomic sequence of the
158 region of interest from the HapMap database (November 2004, Phase I,
159 http://www.hapmap.org/). The program Tagger (available at
160 http://www.broad.mit.edu/mpg/tagger/) was subsequently used to select tag SNPs to
161 cover the region. We used $r^2 \geq 0.7$ to perform pairwise tagging. Additionally we set up
162 a minor allele frequency (MAF) cut off $\geq 5\%$ and took Illumina quality design scores
163 into account for preferential picking of SNPs that are more likely to work on the
164 Illumina platform. The obtained tag SNPs were genotyped using the GoldenGate
165 assay on an Illumina BeadStation 500 GX (Illumina Inc., San Diego, USA). After
166 genotyping all tag SNPs from the initial scan in the Dutch cohort were examined for
167 their resulting quality; all those that were not polymorphic, had a low signal or had
168 clusters that were too wide were excluded (data not shown). In addition, tag SNP that
169 deviated from Hardy-Weinberg equilibrium (HWE) in the controls ($p < 0.001$) were
170 excluded from further analysis, yielding a final set of 872 tag SNPs obtained for
171 genotype analysis.

172

173 For the replication in Finnish and Hungarian cohorts, we chose the 18 markers
174 showing the highest allelic association in the Dutch cohort ($p< 0.01$). 14 of the SNPs
175 showed the highest association in the unadjusted model and 4 additional SNP were
176 chosen from the top markers in a sex adjusted model ($p<0.01$). As the differences in
177 the analysis were not very large we have not proceeded further with the sex adjusted
178 model (data not shown). The markers were genotyped by MALDI-TOF analysis on
179 the Massarray Analyzer platform (Sequenom, San Diego, USA). Genotyping was
180 performed at the MAF core facility in Karolinska Institute, Stockholm, Sweden.
181 Genotyping results and sample information were collected in a BC|GENE LIMS
182 database (Biocomputing Platforms, Espoo, Finland). As five markers in the
183 replication study failed (assay design failed or call rate < 90%, data not shown),
184 genotype information for twelve SNP markers was available for analysis.
185 Deviations from Mendelian inheritance in the family materials were determined with
186 PEDCHECK 1.1 [(16)]. Mendelian errors were detected in 0.21% of genotypes in the
187 Finnish samples, and in 0.17% of the genotypes in the Hungarian samples. All
188 genotypes of a marker were discarded in a family if inheritance errors were detected.
189

190 To ensure that the annotation of the rs9391227 SNP on both the Illumina and
191 Sequenom platforms was performed on the same strand, we performed an additional
192 genotyping of rs9391227 on the Taqman platform using 12 controls from Finland
193 (previously genotyped by Sequenom) and 54 Dutch celiac disease cases and controls
194 (previously genotyped by Illumina). All the samples showed 100% concordance of
195 the genotypes obtained on different platforms (data not shown).

Association analysis

For the initial association analysis in the Dutch sample set, we performed a single SNP association test using the Haploview program [17].

Unphased (version 3.0) [18] was used to look for association in families by calculating allele counts in affected offspring and comparing these with counts from untransmitted alleles from parents and unaffected offspring. All affected individuals in a family were used in the association analysis. As each affected individual within a family is not independent, this constitutes in effect, a test of linkage in the presence of possible association.

Unphased was also used to perform a meta-analysis of the Finnish, Hungarian and Dutch sample sets. Population differences were assessed by a global Wald test. All reported p-values two-tailed and uncorrected unless noted. Only single-marker association analysis was performed as the genetic markers in the follow-up study were too distant from each other. Permutation testing (1000 permutations) was used to estimate empirical p-values, allowing for multiple testing corrections over all tests performed in a run. PLINK (version 1.06) was used to perform quantitative trait analysis in order to explore if the associated markers were possible eQTLs for the genes within a 200 kb window around them.

Power analysis of follow-up study

The Genetic Power Calculator program (http://pngu.mgh.harvard.edu/~purcell/gpc/) [19] was used to estimate the power of the Finnish and Hungarian family-based association analysis to replicate the association in the Dutch case-control samples. Parameters for power estimation in the Finnish and Hungarian follow-up materials

were estimated from the initial Dutch results and marker rs1998166 as it showed the highest association in the discovery phase. We estimated the power to replicate the association to rs1998166 (protective allele frequency 0.123, odds-ratio 0.616, and celiac disease prevalence 0.01) in the 284 Finnish families to be 64%, and in the 357 Hungarian families to be 74%. When combining the Finish and Hungarian samples we had an estimated power of 93%.

Results

Dutch association

In this study, we performed finemapping of the 6q linkage region by genotyping 872 SNP markers in 446 cases and 641 controls. The association of each SNP to celiac disease is shown in figure 1 and supplementary table 1. From 872 SNPs, 18 showed association at $p<0.01$ and were included for replication in CD families from Finland and Hungary.

Because population sub-structure is a possible cause for false positive results, this issue has been investigated previously for all the Dutch cases and a majority of the controls as a part of another study (Carolien de Kovel, unpublished data). No significant population structure was detected using the STRUCTURE software (http://pritch.bsd.uchicago.edu/structure.html). The likelihood of increasing numbers of subpopulations decreased continuously, indicating that it is more likely that the individuals were drawn from a single population than from two or more. Also, the assignment of the individuals to the different sub-populations, when the software was forced to create more than one, was never significantly different between cases and controls (Caroline de Kovel, unpublished data). In view of these results, it is unlikely

that the results of the association analysis in the Dutch dataset were due to population sub-structure.

Follow-up association in Finnish and Hungarian families

Table 1 shows the markers studied in the follow-up association study. The empirical 5% quantile of the best p-values for the Finnish dataset was 0.0029, 0.0033 for the Hungarian dataset, and 0.0027 for the combined Finnish, Hungarian and Dutch dataset (the markers with p-values lower than these are denoted by * in table 1). Three SNPs showed consistent replication in either the Finnish or Hungarian families, to the same allele as observed in the Dutch population. Association analysis revealed that the T-allele of rs9391227 was associated with protection from celiac disease in Finnish families (p 0.0032, OR 0.66), as well as allele T of rs4946111 (p 0.006479, OR 0.66). Allele T of rs9391227 also showed a trend of association with celiac disease in Hungarian families. No other markers were associated with celiac disease in the Hungarian family material.

Association to SNP rs9391227 located downstream of HACE1 and within a region of strong LD surrounding the HACE1 gene was confirmed in a meta-analysis of the Finnish and Hungarian family datasets and the Dutch case-control dataset (p 3.625*10-5, protective allele T OR 0.76). Even if we correct for multiple testing of the 872 SNPs originally tested for, this signal remains significant (Pc = 0.032). This represents the strongest association in the combined materials, and it is stronger than in the Dutch case-control dataset alone. Two additional markers showed a strong association in the meta-analysis, these were rs1998166 (p 0.0002, OR 1.56), located in the RWDD1 gene and rs4946578 (p 0.0018, OR 1.25), located upstream from the GJA1 gene. The association signal of these two markers to celiac disease appears to

be contributed to by all three populations, as the meta-p value is lower than that in the Dutch dataset only, and the allele frequency differences in patients and controls show the same trend in each of the populations.

Tests of heterogeneity between the datasets yielded significant differences ($p < 0.05$) between populations in the markers rs2216084, rs1146229, rs9487060 and rs4946111. It is expected that there will be significant differences between allele frequencies even in well sampled datasets from different populations. The markers rs2216084, rs1146229, and rs9487060, while showing some association in the Dutch dataset, show very little indication of association in the Finnish and Hungarian datasets. This is not likely to be only due to different allele frequencies as the Dutch and Hungarian datasets show similar frequencies in these markers. Rather, it may be due to different haplotype backgrounds in the different populations. Marker rs4946111 shows some association in the Finnish dataset, but less in the Dutch populations. Furthermore, the effect is even reversed in the Hungarian population. This marker may thus constitute a locus on a highly heterogeneous haplotype background between the studied populations. The association in the Finnish samples may be due to an otherwise rare mutation enriched in the Finnish population through genetic drift.

An analysis (data not shown) in 40 Finnish control samples (Dukes *et al.* 2010, unpublished) identified no evidence of eQTLs regarding markers rs9391227 and HACE1, rs1998166 and RWDD1, rs4946578 and GJA1, nor of rs9391227 and PREP expression. *PREP* is not in LD with rs9391227 (data not shown), further arguing against genetic factors within PREP itself explaining the

association with rs9391227. Variation at rs9391227 may still, however, be in LD with and have effects on *PREP* regulatory regions outside of the gene itself.

DISCUSSION

In this study we followed up previous findings of linkage to 6q21-22 in celiac disease by performing a fine-mapping study in the Dutch population, and a replication study in the Finnish and Hungarian populations.

HACE1 & rs9391227

In the combined analysis we identified association of rs9391227 to celiac disease in three populations using both case-control and family datasets. Rs9391227 is an intergenic SNP downstream of and within a region of strong LD surrounding the *HACE1* gene (suppl fig 1). HACE1 (HECT domain and ankyrin repeat containing, E3 ubiquitin protein ligase 1) is an ubiquitination-protein ligase. As celiac disease is believed to be due, at least to some extent, to dysregulation of the immune system and antigen presentation, the importance of ubiquitination in celiac disease is plausible. Whether rs9391227 is involved in this mechanism is unknown. .

Rs1998166 and rs4946578

The marker that is second most strongly independently associated with celiac disease in the meta-analysis is rs1998166. This marker is within an intron of the RWD domain containing 1 (*RWDD1*) gene. Little is known about this gene, but the RWD domain is found in RING finger and WD repeat containing proteins and in the DEXDc-like helicase subfamily, related to the ubiquitin-conjugating enzymes domain.

The third, independent, association found in the meta-analysis was to rs4946578. This marker is approx. 124kb away from, but contained within an LD block that contains the Gap junction protein, alpha 1, 43kDa (*GJA1*, *CX43*) gene. This gene is a member of the connexin gene family and encodes for a component of gap junctions. Connexin-43 is the major protein of gap junctions in the heart, and gap junctions are thought to have a crucial role in the synchronized contraction of the heart and in embryonic development. Mutations in this gene have also been found in deafness [(20)] and in palmoplantar keratoderma [(21)].

PREP

The genetic region we studied on 6q also harbours the gene encoding prolyl endopeptidase (PREP), an enzyme able to cleave proline-rich gluten peptides. An altered PREP activity in the intestinal mucosa could be responsible for the inefficient breakdown of gluten peptides, thus contributing to the onset of celiac disease. *PREP* has previously been identified as a potential celiac disease candidate; however, a previous study indicated no association of this gene with celiac disease in the Dutch population [(22)]. Although this gene is located within the region studied in the current Dutch, Finnish and Hungarian samples, no indication was found that this gene could explain the linkage and/or association to 6q.

Multiple associations and implications

Linkage to 6q21-22 was found in two independent whole-genome linkage studies, and following up the results, we found association to multiple independent SNPs in the

339 Dutch, Finnish, and Hungarian populations. This replication in several independent
340 datasets makes the association signals in the region credible and likely to be true
341 celiac disease risk factors. We did not find one single gene with a major effect on
342 celiac disease susceptibility, suggesting that there may be several susceptibility genes
343 at this locus, each with small effects. The previously conducted GWAS on celiac
344 disease failed to identify strong associations to celiac disease in the studied region on
345 chromosome 6. There are 8 SNPs among the top 1000 signals from the GWAS of
346 Dubois *et al.* [7], which are located in this region, These association signals, however,
347 are in distinct LD blocks (D' ≤ 0.2) from the two most highly associated markers in
348 our study (data not shown). Furthermore, the current study is estimated to adequately
349 tag the eight GWAS signals (data not shown). Of the markers in our study, only three
350 (rs1146229, rs2282854, and rs4946578) were also included in the Dubois GWAS
351 study, allowing for a direct comparison of association signals. These three markers
352 were significantly associated with celiac disease neither in the meta-analysis within
353 the Dubois GWAS study nor in the Finnish or Dutch populations (data not shown),
354 these markers were not part of the follow-up study. In addition to this, the arrays used
355 in the GWAS do not contain perfect proxies for the most highly associated markers in
356 our study (data not shown), arguing that the association we see in the current study
357 may have been missed in the GWAS. We thus conclude that our results likely indicate
358 true association that is independent of the GWAS signals, and that it is unlikely that
359 we are missing major associations at this locus due to insufficient marker coverage.
360
361 The current study cannot fully explain the underlying cause of the multiple
362 independent association signals in the region. The multiple association signals show
363 the complexity of disease susceptibility in the region, but the results are consistent in

the different study populations and the observed risk effect of the associated variants is relatively high. This argues that each of the identified variants (or nearby untyped variants in LD with the associated variants) is a true, novel, susceptibility factor for celiac disease.

ACKNOWLEDGEMENTS

All the study subjects are thanked for participating in the study. We thank the Dutch Celiac Disease Foundation for participating in this study, H. van Someren and F. Mulder for clinical database management, and Ruben van‘t Slot at UMC Utrecht for technical assistance. We thank Wieke HM Verbeek for help in sample collection and Jackie Senior for critical reading of the manuscript. We thank Hanne Ahola and Anne Heimonen for excellent technical assistance and Erzsébet Szathmári, Judit B.Kovács, Margit Lörincz and Anikó Nagy for their work with the Hungarian families. We thank Carolien de Kovel at UMC Utrecht for providing unpublished data on the population structure of the Dutch cohort. We acknowledge funding from the Coeliac Disease Consortium (an innovative cluster approved by the Netherlands Genomics Initiative and partly funded by the Dutch Government, grant BSIK03009 to CW); the Netherlands Organization for Scientific Research (VICI grant 918.66.620 to CW) and the Dutch Digestive Disease Foundation. This work has been funded by the EU Commission through a Marie Curie Excellence Grant (FP6 contract MEXT-CT-2005-025270 to PS), by the Academy of Finland, the Hungarian Scientific Research Fund (contract OTKA 61868), the University of Helsinki Funds, Biocentrum Helsinki, the Competitive Research Funding of the Tampere University Hospital, the Foundation of Pediatric Research and the Sigrid Juselius Foundation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Marsh M. N. Gluten, major histocompatibility complex, and the small intestine. A molecular and immunobiologic approach to the spectrum of gluten sensitivity ('celiac sprue'). *Gastroenterology* 1992;**102:**330-354.
2. Lohi S., Mustalahti K., Kaukinen K., *et al*. Increasing prevalence of coeliac disease over time. *Aliment.Pharmacol.Ther.* 2007;**26:**1217-1225.
3. Ravikumara M., Tuthill D. P. and Jenkins H. R. The changing clinical presentation of coeliac disease. *Arch.Dis.Child.* 2006;**91:**969-971.
4. Greco L., Romino R., Coto I., *et al*. The first large population based twin study of coeliac disease. *Gut* 2002;**50:**624-628.
5. Molberg O., Mcadam S. N., Korner R., *et al*. Tissue transglutaminase selectively modifies gliadin peptides that are recognized by gut-derived T cells in celiac disease. *Nat.Med.* 1998;**4:**713-717.
6. van Heel D. A., Franke L., Hunt K. A., *et al*. A genome-wide association study for celiac disease identifies risk variants in the region harboring IL2 and IL21. *Nat.Genet.* 2007;**39:**827-829.
7. Dubois P. C., Trynka G., Franke L., *et al*. Multiple common variants for celiac disease influencing immune gene expression. *Nat.Genet.* 2010;**42:**295-302.
8. Van Belzen M. J., Meijer J. W., Sandkuijl L. A., *et al*. A major non-HLA locus in celiac disease maps to chromosome 19. *Gastroenterology* 2003;**125:**1032-1041.
9. Hunt K. A., Zhernakova A., Turner G., *et al*. Newly identified genetic risk variants for celiac disease related to the immune response. *Nat.Genet.* 2008;**40:**395-402.
10. Trynka G., Zhernakova A., Romanos J., *et al*. Coeliac disease associated risk variants in TNFAIP3 and REL implicate altered NF-{kappa}B signalling. *Gut* 2009;
11. Anonymous Revised criteria for diagnosis of coeliac disease. Report of Working Group of European Society of Paediatric Gastroenterology and Nutrition. *Arch.Dis.Child.* 1990;**65:**909-911.
12. Monsuur A. J., de Bakker P. I., Alizadeh B. Z., *et al*. Myosin IXB variant increases the risk of celiac disease and points toward a primary intestinal barrier defect. *Nat.Genet.* 2005;**37:**1341-1344.
13. Koskinen L. L., Einarsdottir E., Dukes E., *et al*. Association study of the IL18RAP locus in three European populations with coeliac disease. *Hum.Mol.Genet.* 2009;**18:**1148-1155.
14. Koskinen L. L., Korponay-Szabo I. R., Viiri K., *et al*. Myosin IXB gene region and gluten intolerance: linkage to coeliac disease and a putative dermatitis herpetiformis association. *J.Med.Genet.* 2008;**45:**222-227.

15. van Belzen M. J., Vrolijk M. M., Meijer J. W., *et al*. A genomewide screen in a four-generation Dutch family with celiac disease: evidence for linkage to chromosomes 6 and 9. *Am.J.Gastroenterol.* 2004;**99:**466-471.
16. O'Connell J. R. and Weeks D. E. PedCheck: a program for identification of genotype incompatibilities in linkage analysis. *Am.J.Hum.Genet.* 1998;**63:**259-266.
17. Barrett J. C., Fry B., Maller J. and Daly M. J. Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* 2005;**21:**263-265.
18. Dudbridge F. Pedigree disequilibrium tests for multilocus haplotypes. *Genet.Epidemiol.* 2003;**25:**115-121.
19. Purcell S., Cherny S. S. and Sham P. C. Genetic Power Calculator: design of linkage and association genetic mapping studies of complex traits. *Bioinformatics* 2003;**19:**149-150.
20. Paznekas W. A., Karczeski B., Vermeer S., *et al*. GJA1 mutations, variants, and connexin 43 dysfunction as it relates to the oculodentodigital dysplasia phenotype. *Hum.Mutat.* 2009;**30:**724-733.
21. van Steensel M. A., Spruijt L., van der Burgt I., *et al*. A 2-bp deletion in the GJA1 gene is associated with oculo-dento-digital dysplasia with palmoplantar keratoderma. *Am.J.Med.Genet.A.* 2005;**132A:**171-174.
22. Diosdado B., Stepniak D. T., Monsuur A. J., *et al*. No genetic association of the human prolyl endopeptidase gene in the Dutch celiac disease population. *Am.J.Physiol.Gastrointest.Liver Physiol.* 2005;**289:**G495-500.

TITLES AND LEGENDS TO FIGURES

Figure 1. Association of SNPs in the chromosome 6 region to celiac disease in the Dutch cohort. SNP positions are based on NCBI Build 35 and are given in Mb. Diamonds represent individual SNPs.

Table 1. SNP markers in the study and their association to celiac disease in Dutch cases and controls (A) as well as Finnish (B) and Hungarian (C) families, and a meta-analysis of all three populations (D). Case frequencies, in the family materials, refer to the allele frequencies of the affected offspring ("cases") and the control frequencies refer to both unaffected offspring ("healthy") and the untransmitted alleles in parents. P-values lower than 0.01 are considered significant (marked in bold) and odds ratios (OR) and 95% OR confidence intervals for significant markers are given. Emp permutation p-values are based on 10000 permutations.

Supplementary table 1. Allelic and genotypic association of 860 SNP markers to celiac disease in Dutch cases and controls. The 12 markers chosen for follow-up analysis are shown in bold and underlined. Genotype counts (11,12,22), Hardy-Weinberg (HW) p-values and allelic and genotypic p-values are shown.

Figure 1

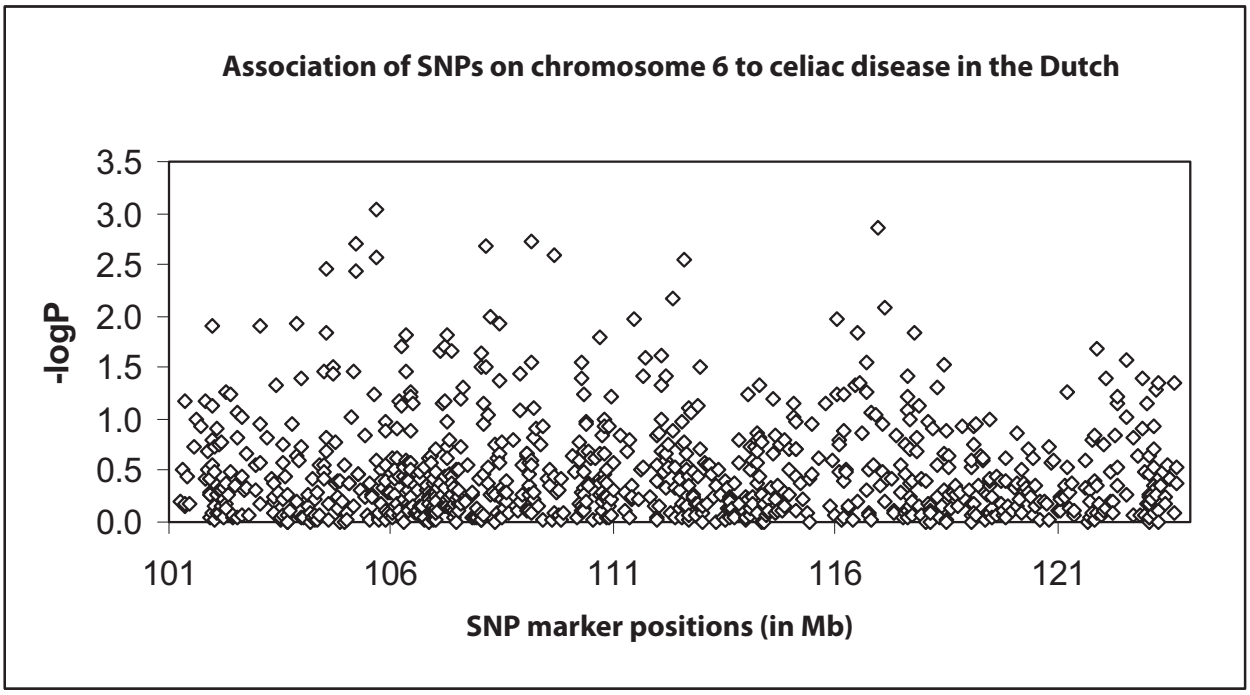
Association of SNPs on chromosome 6 to celiac disease in the Dutch
-logP
3.5
3.0
2.5
2.0
1.5
1.0
0.5
0.0
101
106
111
116
121
SNP marker positions (in Mb)

Table 1 - SNP markers and association in follow-up study

			A DUTCH (446 cases, 641 controls)				B FINNISH (623 cases, 359 healthy)			C HUNGARIAN (592 cases, 471 healthy)			D FINNISH, HUNGARIAN, DUTCH (1661 cases, 1471 controls)		
Markers	**Position (bp)**	**Alleles**	**case,contr freq**	**p value**	**emp p-value**	**OR (95% CI)**	**case,contr freq**	**p value**	**OR (95% CI)**	**case,contr freq**	**p value**	**OR (95% CI)**	**case,contr freq**	**p value**	**OR (95% CI)**
rs1648957	104613397	CT	0.498, 0.433	0.003	0.005	1.30 (1.09-1.55)	0.449, 0.481	0.351		0.545, 0.519	0.440		0.498, 0.460	0.085	
rs9391227	105265327	AT	0.497, 0.559	0.005	0.003	0.78 (0.66-0.93)	0.541, 0.651	**0,003***	0.66 (0.49-0.88)	0.495, 0.555	0.054		0.507, 0.577	**3.07*10-5 ***	0.76 (0.66-0.86)
rs2895615	106292441	CG	0.690, 0.642	0.020	0.018	1.25 (1.03-1.50)	0.668, 0.641	0.398		0.679, 0.678	0.871		0.683, 0.649	0.061	
rs2216084	107660005	CT	0.646, 0.688	0.045	0.058	0.83 (0.69-0.99)	0.809, 0.757	0.048	1.41 (0.99-1.99)	0.706, 0.676	0.603		0.700, 0.699	0.836	
rs1146229	109205720	CT	0.085, 0.052	0.002	0.003	1.72 (1.22-2.43)	0.111, 0.118	0.673		0.070, 0.095	0.082		0.087, 0.075	0.259	
rs9487060	109741149	CG	0.079, 0.048	0.003	0.002	1.73 (1.21-2.47)	0.067, 0.072	0.724		0.060, 0.062	0.763		0.072, 0.054	0.032	1.36 (1.03-1.79)
rs2503770	110373108	AT	0.233, 0.274	0.031	0.028	0.80 (0.66-0.98)	0.324, 0.313	0.884		0.237, 0.217	0.376		0.255, 0.269	0.536	
rs571060	112403213	CT	0.970, 0.987	0.007	0.010	0.44 (0.24-0.81)	0.992, 0.993	1.000		0.982, 0.978	0.841		0.978, 0.986	0.085	
rs2282854	112638353	CT	0.933, 0.962	0.003	0.003	0.55 (0.37-0.82)	0.968, 0.965	0.796		0.948, 0.938	0.452		0.946, 0.958	0.170	
rs4946111	116091112	GT	0.445, 0.502	0.011	0.008	0.80 (0.67-0.95)	0.385, 0.482	**0.007**	0.66 (0.49-0.90)	0.434, 0.390	0.280		0.429, 0.474	0.020	0.86 (0.76-0.98)
rs1998166	117020110	CT	0.921, 0.877	0.001	0.003	1.62 (1.20-2.19)	0.940, 0.917	0.105		0.956, 0.928	0.087		0.934, 0.896	**2.17*10-4 ***	1.56 (1.22-1.98)
rs4946578	121936760	AG	0.584, 0.534	0.023	0.023	1.23 (1.03-1.46)	0.732, 0.695	0.323		0.564, 0.536	0.421		0.606, 0.558	**0.002**	1.25 (1.09-1.44)