Published in final edited form as:

Nat Genet. 2015 June ; 47(6): 577–578. doi:10.1038/ng.3268.

Fine-mapping in the MHC region accounts for 18% additional genetic risk for celiac disease

Javier Gutierrez-Achury^{1,10}, Alexandra Zhernakova^{1,10}, Sara L. Pulit², Gosia Trynka³, Karen A. Hunt⁴, Jihane Romanos¹, Soumya Raychaudhuri^{5,6,7,8}, David A. van Heel⁴, Cisca Wijmenga^{1,11,12}, and Paul I.W. de Bakker^{2,9,11}

¹University of Groningen, University Medical Center, Department of Genetics, Groningen, the Netherlands ²Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, The Netherlands ³Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Cambridge, UK ⁴Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, London, UK ⁵Division of Rheumatology, Immunology, and Allergy, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA ⁶Division of Genetics, Brigham and Women's Hospital and Harvard Medical School, Boston, MA 02115, USA ⁷Program in Medical and Population Genetics, Broad Institute, Cambridge, MA 02142, USA ⁸Arthritis Research UK Epidemiology Unit, Centre for Musculoskeletal Research, Institute of Inflammation and Repair, University of Manchester, Manchester M13 9PT, UK ⁹Department of Epidemiology, Julius Center for Health Sciences and Primary Care, University Medical Center Utrecht, Utrecht, The Netherlands

Abstract

Although dietary gluten is the trigger, celiac disease risk is strongly influenced by genetic variation in the major histocompatibility complex (MHC) region. We fine-mapped the MHC association signal to identify additional risk factors independent of the *HLA-DQ* alleles and observed five novel associations that account for 18% of the genetic risk. Together with the 57 known non-MHC loci, genetic variation can now explain up to 48% of celiac disease heritability.

Celiac disease is an autoimmune disease triggered by the consumption of dietary gluten in which the primary organ affected is the small intestine. The major histocompatibility complex (MHC) is the main genetic factor in disease development, with the strongest effects

Users may view, print, copy, and download text and data-mine the content in such documents, for the purposes of academic research, subject always to the full Conditions of use:http://www.nature.com/authors/editorial_policies/license.html#terms

¹²Corresponding author: Professor C Wijmenga, Department of Genetics, University Medical Centre Groningen, PO Box 30001, 9700 RB Groningen, The Netherlands. Tel: + 31 50 361710; Fax: + 31 50 3617230; c.wijmenga@umcg.nl.

¹⁰These authors contribute equally to this study

¹¹These authors jointly directed this study

AUTHOR CONTRIBUTIONS

J.G.-A., A.Z., C.W. and P.I.W.d.B. designed the study and analyzed and interpreted the data. S.P. imputed the data. J.G.-A., G.T., K.H. and J.R. prepared the data for analysis. D.v.H. and C.W. collected and genotyped samples. J.G.-A., A.Z., C.W., and P.I.W.d.B. wrote the manuscript and J.G.-A., A.Z., S.P., D.v.H., S.R., C.W. and P.I.W.d.B. critically discussed the manuscript and received feedback from the other authors.

COMPETING INTERESTS STATEMENT

The authors declare no competing financial interests.

mapped to the classical *HLA-DQA1* and *HLA-DQB1* genes. Specifically, the common haplotypes DQ2.5, DQ2.2 and DQ8 have been shown to increase disease risk by 6-fold on average^{1,2}. Genome-wide association analyses have identified another 57 independent SNPs located in 39 risk loci outside the MHC^{3,4}. In this study, we hypothesized that a substantial fraction of celiac disease heritability might be explained by unknown independent variants across the MHC (but outside the classical HLA-DQ region). These variants might have remained undetected in previous genome-wide association studies due to the complexities of fine-mapping association signals in the extended MHC region. We therefore imputed the region and tested for associated SNPs, amino acid polymorphisms, and classical HLA alleles in 12,016 cases and 11,920 controls of European ancestry (in six cohorts from five countries), all genotyped on the Illumina Immunochip array (Supplementary Table 1, Supplementary Methods).

Because HLA class II haplotypes are established risk factors for celiac disease, we first focused on the region encompassing *HLA-DRB1*, *HLA-DQA1* and *HLA-DQB1*. Consistent with functional studies, we confirmed strong associations for *HLA-DQ* ($p = 10^{-177}$ for a direct comparison between all classical 4-digit alleles of *HLA-DQA1/DQB1* and of *HLA-DRB1*; Supplementary Methods, Supplementary Table 2). We also observed remarkable allele frequency differences across European populations (Supplementary Table 3, Supplementary Fig. 1). In a step-wise conditional analysis within the HLA-DQ heterodimer molecule, we identified amino acid positions 25 and 47 in HLA-DQ α 1, and positions 57 and 74 in HLA-DQ β 1, as the strongest independent associations with celiac disease risk (Supplementary Fig. 2, Supplementary Table 4). The haplotype model based on these four associated amino acid positions, of which β 57 and β 74 are thought to make direct contact with the epitope in the HLA-DQ binding groove, almost perfectly recapitulates the associations with the classical *HLA-DQA1/DQB1* alleles (99.6% of the explained variance). For example, the haplotype defined by HLA-DQA1-Tyr25/Cys47 and HLA-DQB1-Ala57/Ala74 completely mirrors the risk conferred by HLA-DQ2.5 (amino acid haplotype OR=19.13, $P = 10^{-1.754}$; HLA-DQ2.5 OR=21.56, $P = 10^{-1.801}$) (Supplementary Tables 2 and 4).

In order to identify additional associations beyond HLA-DQ, we tested all the other variants across the MHC region after correcting for the collective effects of all classical *HLA-DQA1* and *DQB1* 4-digit alleles. We identified five further independent variants with genome-wide significance: amino acid position 9 in HLA-DP β 1 (omnibus $p = 10^{-41}$), classical alleles *HLA-B*08:01* ($p = 10^{-22}$) and *HLA-B*39:06* ($p = 10^{-8}$), and two SNPs, rs1611710 ($p = 10^{-9}$) and rs2301226 ($p = 10^{-9}$) (Fig. 1, Supplementary Fig. 3, Supplementary Table 5). eQTL analysis of publicly available datasets⁵ demonstrated a *cis*-eQTL effect of rs1611710 on *HLA-F* expression (eQTL $p = 1.22 \times 10^{-56}$) and of rs2301226 on *B3GALT4* (eQTL $p = 1.67 \times 10^{-39}$) and on *HLA-DPBI* expression (eQTL $p = 3.87 \times 10^{-13}$). When we did not adjust *a priori* for the effects of *HLA-DQ* alleles, we found the same five independent effects, in addition to several strong effects in the class II region (Supplementary Fig. 4 and Supplementary Table 5), providing support for the robustness of these novel findings.

Based on all the genetic associations for celiac disease risk, we estimated the genetic variance explained (h^2) with a liability threshold model, assuming a disease prevalence of

1%^{6–10} (Table 1 and Supplementary Methods). The classical *HLA-DQ* loci alone explain $h^2 = 0.232$ on the liability scale, whereas the five novel variants reported here explain an additional $h^2 = 0.180$, which is more than twice that explained by the 57 independent SNPs outside the MHC³ ($h^2 = 0.065$). In total, all the variants identified to date as being robustly associated with celiac disease risk explain approximately $h^2 = 0.477$. These liability estimates are more modest than those previously published (~87% explained genetic risk¹¹). This discrepancy illustrates the difficulty of comparing explained variance estimates between studies if the model assumptions are not the same¹². When we changed the assumed prevalence to 2% and 3%, as reported in some populations (Supplementary Methods), we observed substantially increased heritability estimates, which illustrates not only how sensitive the estimates are to this parameter, but also how estimates may change in populations with an empirically higher disease prevalence (Supplementary Table 6).

Although the importance of the MHC region in celiac disease and other autoimmune diseases is widely recognized, the particular properties of the region¹³ have made it difficult to identify independent loci and pinpoint functional variants that play a role in disease pathology. Taking advantage of high-density imputation¹⁴, we implicated a set of five variants (Fig. 1, Supplementary Figs. 2 and 3) that act independently of HLA-DQ and which can explain an additional 18% of celiac disease heritability. On the basis of these results, it seems likely that fine-mapping the MHC region in other diseases with classical HLA associations will reveal some of the “missing” heritability.

Finally, this study provides evidence that the MHC class I region is also an important locus in celiac disease etiology. This is an interesting observation given the presumed role for intraepithelial lymphocytes (IELs) TCR $\alpha\beta^+$ CD8 $\alpha\beta^+$, which have been shown to exacerbate celiac disease and are restricted to MHC class I recognition¹⁵. These cells represent a subtype of antigen-experienced “innate-like” T cells and, under conditions of homeostasis, they are located within the intestinal mucosa, maintaining epithelium integrity and preventing pathogen incursion. The TCR $\alpha\beta^+$ CD8 $\alpha\beta^+$ are most abundant in the intestinal mucosa (70–80% of the IEL subset) and are a hallmark of celiac disease due to their correlation with gluten intake, thereby providing further support for the role of IELs in the etiology of celiac disease.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We thank Jackie Senior for carefully reading the manuscript and Claudia Gonzalez-Arevalo for the graphics. This study was funded by a grant from the Celiac Disease Consortium, an Innovative Cluster approved by the Netherlands Genomics Initiative, to C.W., a European Research Council advanced grant (FP/2007–2013/ERC grant 2012-322698) to C.W., a grant from the Dutch Reumafonds (11-1-101) to A.Z. and a Rosalind Franklin Fellowship from the University of Groningen to A.Z. S.R. and P.I.W.d.B. received support from the National Institutes of Health, USA (1R01AR062886). P.I.W.d.B. holds a VIDI award from the Netherlands Organization for Scientific Research (NWO project 016.126.354). S.R. is also supported by a Doris Duke Clinical Scientist Development Award.

References

1. Liu E, et al. *N Engl J Med*. 2014; 371:42–49. [PubMed: 24988556]
2. Romanos J, et al. *Gastroenterology*. 2009; 137:834–840. 840, e1–e3. [PubMed: 19454285]
3. Trynka G, et al. *Nat Genet*. 2011; 43:1193–1201. [PubMed: 22057235]
4. Hunt KA, et al. *Nature*. 2013; 498:232–235. [PubMed: 23698362]
5. Westra HJ, et al. *Nat Genet*. 2013; 45:1238–1243. [PubMed: 24013639]
6. Cooper GS, Bynum ML, Somers EC. *J Autoimmun*. 2009; 33:197–207. [PubMed: 19819109]
7. Speed D, Hemani G, Johnson MR, Balding DJ. *Am J Hum Genet*. 2012; 91:1011–1021. [PubMed: 23217325]
8. Yang J, Lee SH, Goddard ME, Visscher PM. *Am J Hum Genet*. 2011; 88:76–82. [PubMed: 21167468]
9. Lee SH, Goddard ME, Wray NR, Visscher PM. *Genet Epidemiol*. 2012; 36:214–224. [PubMed: 22714935]
10. Wray NR, Purcell SM, Visscher PM. *PLoS Biol*. 2011; 9:e1000579. [PubMed: 21267061]
11. Nistico L, et al. *Gut*. 2006; 55:803–808. [PubMed: 16354797]
12. Witte JS, Visscher PM, Wray NR. *Nat Rev Genet*. 2014; 15:765–776. [PubMed: 25223781]
13. de Bakker PI, et al. *Nat Genet*. 2006; 38:1166–1172. [PubMed: 16998491]
14. Jia X, et al. *PLoS One*. 2013; 8:e64683. [PubMed: 23762245]
15. Abadie V, Discepolo V, Jabri B. *Semin Immunopathol*. 2012; 34:551–566. [PubMed: 22660791]

Figure 1.

Schematic representation of the stepwise analysis applied in fine-mapping the HLA region. *HLA-DQA1* and *HLA-DQB1* 4-digit alleles were included in the model to control for their effect (red arrow). A green flag represents each of the steps in the analysis and an independent association.

Table 1

Heritability estimates for HLA and non-HLA variants in celiac disease.

Model	h ²	h ²
	Observed scale	Liability scale
Classical HLA-DQ alleles (4-digit)	0.316	0.232
Independent MHC variants outside HLA-DQ	0.357	0.180
Non-MHC variants (57 SNPs)	0.15	0.065
Combined	0.823	0.477