nature genetics

Common variants in the *HLA-DQ* region confer susceptibility to idiopathic achalasia

Ines Gockel[1,29], Jessica Becker[2,3,29], Mira M Wouters[4], Stefan Niebisch[1], Henning R Gockel[1], Timo Hess[2,3], David Ramonet[5], Julian Zimmermann[5], Ana González Vigo[6], Gosia Trynka[7], Antonio Ruiz de León[8], Julio Pérez de la Serna[8], Elena Urcelay[6], Vinod Kumar[7], Lude Franke[7], Harm-Jan Westra[7], Daniel Drescher[1], Werner Kneist[1], Jens U Marquardt[9], Peter R Galle[9], Manuel Mattheisen[10,11], Vito Annese[12], Anna Latiano[13], Uberto Fumagalli[14], Luigi Laghi[15], Rosario Cuomo[16], Giovanni Sarnelli[16], Michaela Müller[17], Alexander J Eckardt[17], Jan Tack[4], Per Hoffmann[2,3,18,19], Stefan Herms[2,3,18,19], Elisabeth Mangold[2,3], Stefanie Heilmann[2,3], Ralf Kiesslich[20], Burkhard H A von Rahden[21], Hans-Dieter Allescher[22], Henning G Schulz[23], Cisca Wijmenga[7], Michael T Heneka[5], Hauke Lang[1], Karl-Peter Hopfner[24,25], Markus M Nöthen[2,3], Guy E Boeckxstaens[4], Paul I W de Bakker[26,27], Michael Knapp[28,30] & Johannes Schumacher[2,3,30]

Idiopathic achalasia is characterized by a failure of the lower esophageal sphincter to relax due to a loss of neurons in the myenteric plexus[1,2]. This ultimately leads to massive dilatation and an irreversibly impaired megaesophagus. We performed a genetic association study in 1,068 achalasia cases and 4,242 controls and fine-mapped a strong MHC association signal by imputing classical HLA haplotypes and amino acid polymorphisms. An eight-residue insertion at position 227–234 in the cytoplasmic tail of HLA-DQβ1 (encoded by *HLA-DQB1*05:03* and *HLA-DQB1*06:01*) confers the strongest risk for achalasia ($P = 1.73 \times 10^{-19}$). In addition, two amino acid substitutions in the extracellular domain of HLA-DQα1 at position 41 (lysine encoded by *HLA-DQA1*01:03*; $P = 5.60 \times 10^{-10}$) and of HLA-DQβ1 at position 45 (glutamic acid encoded by *HLA-DQB1*03:01* and *HLA-DQB1*03:04*; $P = 1.20 \times 10^{-9}$) independently confer achalasia risk. Our study implies that immune-mediated processes are involved in the pathophysiology of achalasia.

Idiopathic achalasia has a lifetime prevalence of 1 in 10,000 and is characterized by neuronal degeneration in the myenteric plexus[1,2]. Although the cause is still unknown, an autoimmune mechanism has been proposed because circulating autoantibodies and inflammatory cell infiltrates in the myenteric plexus of affected individuals have been found[2].

We present the first systematic association study in achalasia using a sample of 1,068 cases and 4,242 controls from central Europe (630 cases and 2,652 controls), Spain (273 cases and 327 controls) and Italy (165 cases and 1,263 controls). In cases, achalasia diagnosis was made according to a standard procedure including esophageal manometry and esophagography. Genotyping was performed on the Immunochip, covering 196,524 SNPs at immune-related loci throughout the genome. Of all the markers, 116,672 SNPs passed our quality control filters. Details on the studied samples, genotyping and quality control are provided in the Online Methods.

The association analysis yielded 33 markers reaching genome-wide significance ($P < 5.0 \times 10^{-8}$; **Fig. 1**), of which rs3129727 was the most

[1]Department of General, Visceral and Transplant Surgery, University Medical Center, University of Mainz, Mainz, Germany. [2]Institute of Human Genetics, University of Bonn, Bonn, Germany. [3]Department of Genomics, Life & Brain Center, University of Bonn, Bonn, Germany. [4]Translational Research Center for Gastrointestinal Disorders, Catholic University of Leuven, Leuven, Belgium. [5]Department of Neurology, Division of Clinical Neurosciences, University of Bonn, Bonn, Germany. [6]Department of Immunology, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain. [7]Department of Genetics, University of Groningen, University Medical Center Groningen, Groningen, the Netherlands. [8]Department of Gastroenterology, Instituto de Investigación Sanitaria del Hospital Clínico San Carlos (IdISSC), Madrid, Spain. [9]First Department of Internal Medicine, University Medical Center, University of Mainz, Mainz, Germany. [10]Department of Biomedicine, Aarhus University, Aarhus, Denmark. [11]Centre for Integrative Sequencing (iSEQ), Aarhus University, Aarhus, Denmark. [12]Department of Gastroenterology, Careggi Hospital, University of Florence, Florence, Italy. [13]Division of Gastroenterology, Istituto di Ricovero e Cura a Carattere Scientifico, Casa Sollievo della Sofferenza Hospital, San Giovanni Rotondo, Italy. [14]Department of Surgery, Istituto Clinico Humanitas, Istituto di Ricovero e Cura a Carattere Scientifico, Milan, Italy. [15]Department of Gastroenterology, Istituto Clinico Humanitas, Istituto di Ricovero e Cura a Carattere Scientifico, Milan, Italy. [16]Department of Clinical Medicine and Surgery, Division of Gastroenterology, Federico II University Hospital School of Medicine, Naples, Italy. [17]Department of Gastroenterology, German Diagnostic Clinic, Wiesbaden, Germany. [18]Division of Medical Genetics, University Hospital, Basel, Switzerland. [19]Human Genetics Research Group, Department of Biomedicine, University of Basel, Basel, Switzerland. [20]Department of Internal Medicine, Hospital St. Marien, Frankfurt, Germany. [21]Department of General, Visceral, Vascular and Pediatric Surgery, University of Würzburg, Würzburg, Germany. [22]Center of Internal Medicine, Hospital Garmisch-Partenkirchen, Garmisch-Partenkirchen, Germany. [23]Department of General and Abdominal Surgery, Protestant Hospital Castrop-Rauxel, Castrop-Rauxel, Germany. [24]Department of Biochemistry, Gene Center, Ludwig Maximilians University, Munich, Germany. [25]Center for Integrated Protein Sciences, Munich, Germany. [26]Department of Epidemiology, University Medical Center Utrecht, Utrecht, the Netherlands. [27]Department of Medical Genetics, Center for Molecular Medicine, University Medical Center Utrecht, Utrecht, the Netherlands. [28]Institute for Medical Biometry, Informatics and Epidemiology, University of Bonn, Bonn, Germany. [29]These authors contributed equally to this work. [30]These authors jointly directed this work. Correspondence should be addressed to J.S. (johannes.schumacher@uni-bonn.de).

Received 11 February; accepted 11 June; published online 6 July 2014; doi:10.1038/ng.3029

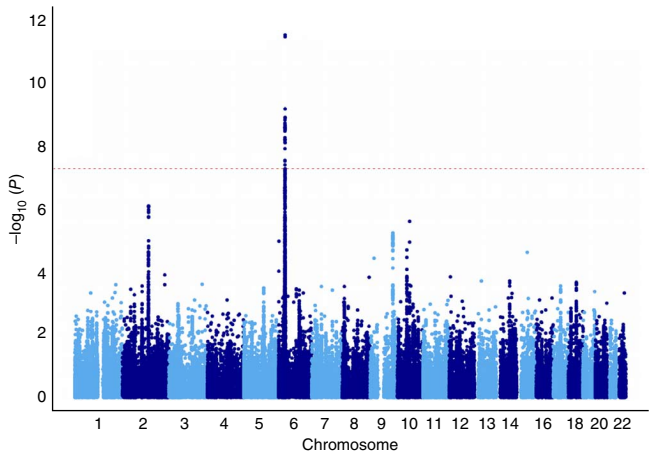

Figure 1 Immunochip association analysis across 1,068 achalasia cases and 4,242 controls. $-\log_{10} P$ values are shown for SNPs that passed quality control. The genome-wide significance threshold ($P = 5 \times 10^{-8}$) is indicated by the dashed line.

strongly associated variant ($P = 2.79 \times 10^{-12}$; relative risk (RR) = 2.15; **Supplementary Table 1**). All 33 markers were located in the major histocompatibility (MHC) region on chromosome 6, which represents one of the most gene-dense regions in the human genome with long-range linkage disequilibrium (LD)[3]. To capture most of the genetic variants within this region, we used a reference panel of 5,225 European individuals[4] to impute classical human leukocyte antigen (HLA) haplotypes, their corresponding amino acid changes and SNPs. In total, we tested 126 classical HLA haplotypes at 2-digit resolution, 298 classical HLA haplotypes at 4-digit resolution, 402 polymorphic amino acid positions and 8,907 SNPs across the MHC region for association. Details on the imputation and statistical tests are provided in the Online Methods.

Of all variants tested, an eight-residue insertion at position 227–234 of HLA-DQβ1 showed the strongest association with achalasia ($P = 1.73 \times 10^{-19}$; **Supplementary Fig. 1a** and **Supplementary Table 2**). Two haplotypes (*HLA-DQB1*05:03* and *HLA-DQB1*06:01*) encoded this insertion, resulting in the introduction of a PQGPPPAG stretch in the cytoplasmic tail of HLA-DQβ1, which is encoded by an alternatively spliced exon 5 of the corresponding gene. The effect on splicing is caused by SNP rs28688207, located in the splice acceptor site[5]. Because rs28688207 was not imputed in our data, we genotyped this marker in a subset of achalasia cases from central Europe ($n = 350$). The genotype at rs28688207 in this subset was in near-perfect agreement with the insertion genotype (concordance rate = 99.7%), reflecting the robustness of our imputation.

The protein products of *HLA-DQB1* and *HLA-DQA1* form the HLA-DQ heterodimer receptor on antigen-presenting cells. To identify other functional amino acids in this receptor associated with achalasia, we adopted the same approach as presented in a recent study on rheumatoid arthritis[6] (Online Methods). We first conditioned on the eight-residue insertion and tested each HLA-DQ amino acid position for association by grouping classical *HLA-DQB1* and *HLA-DQA1* haplotypes according to the specific amino acid at each position. This analysis showed that the lysine at position 41 and the alanine at position 130 of HLA-DQα1 conferred risk of achalasia ($P = 5.60 \times 10^{-10}$; **Supplementary Table 3**). Because *HLA-DQA1*01:03* is the only haplotype encoding a lysine at position 41 and the only haplotype encoding an alanine at position 130, both positions were (statistically) equivalent with regard to achalasia risk. Next, we conditioned on the eight-residue insertion in HLA-DQβ1 and HLA-DQα1 positions 41 and 130, thereby identifying a third association signal for achalasia. The *HLA-DQB1* haplotypes *HLA-DQB1*03:01* and *HLA-DQB1*03:04* encoding a glutamic acid residue at position 45 were significantly more frequent in cases in comparison to controls ($P = 1.20 \times 10^{-9}$; **Supplementary Table 3**). Conditioning on all three risk variants did not identify any residual association at other HLA-DQ amino acid positions ($P > 1 \times 10^{-4}$; **Supplementary Table 3**).

We characterized all identified risk-associated variants in detail. We observed that the effect of each of the three variants was preserved when conditioning on the two other variants (**Table 1**), confirming that all variants contribute independently to achalasia risk. Furthermore, there was no evidence ($P > 0.5$) for epistasis between any of the risk variants and no evidence ($P > 0.1$) for dominance effects of any of the three variants (data not shown). In addition, none of the 121,485 possible combinations of 3 amino acid positions in HLA-DQ outperformed the combination of the 8-residue HLA-DQβ1 insertion, HLA-DQβ1 position 45 and HLA-DQα1 position 41 (**Supplementary Table 4**). We also compared the individual disease risk predicted by a full model that included all four-digit classical *HLA-DQB1* and *HLA-DQA1* haplotypes with that predicted by a simpler model defined by the three identified positions. In this comparison, there was nearly perfect correlation of $r = 0.977$ (**Supplementary Fig. 2**), demonstrating that almost all of the achalasia association for HLA-DQ is captured by the three amino acid polymorphisms. Although the identified risk variants represent the most parsimonious explanation, we cannot completely rule out a more complex HLA-DQ heterodimer model involving *HLA-DQB1*05:03* and *HLA-DQB1*06:01* (both encoding the eight-residue insertion) as an alternative explanation (**Supplementary Tables 5** and **6**).

To further assess whether there were independent associations outside HLA-DQ, we conditioned on all identified risk variants and tested the remaining variants (HLA haplotypes, amino acid changes and SNPs) within the MHC region. We did not see any independent association ($P > 1 \times 10^{-6}$) after correcting for all three amino acid changes in HLA-DQ (**Supplementary Fig. 1b** and **Supplementary Table 7**).

We genotyped the strongest achalasia risk variant rs28688207 causing the 8-residue HLA-DQβ1 insertion in an independent sample of 412 cases and 1,020 controls from central Europe. Diagnosis in cases was based on the same criteria as used in our initial sample. We replicated our initial finding. The eight-residue insertion in HLA-DQβ1

Table 1 Joint logistic regression analysis of the achalasia-associated amino acid positions in HLA-DQ

Protein	Amino acid position[a]	Amino acid risk/other variant	Frequency (%)[b] Cases	Frequency (%)[b] Controls	*P*	RR (95% CI)	Classical HLA haplotype carrying the risk variant
HLA-DQβ1	227–234	PQGPPPAG/–	9.9	4.7	2.17×10^{-18}	2.44 (2.00–2.99)	*DQB1*05:03*, *DQB1*06:01*
HLA-DQα1[c]	41	K/R	12.5	7.1	2.37×10^{-12}	1.86 (1.56–2.22)	*DQA1*01:03*
HLA-DQα1[c]	130	A/S					
HLA-DQβ1	45	E/G	24.9	22.4	1.20×10^{-9}	1.47 (1.30–1.66)	*DQB1*03:01*, *DQB1*03:04*

RR, relative risk; CI, confidence interval.
[a]According to the IMGT/HLA database (Release 3.9.0)[11]. [b]Frequency of the risk variant. [c]Risk-associated amino acid at positions 41 and 130 in HLA-DQα1 are both encoded only by *HLA-DQA1*01:03*.

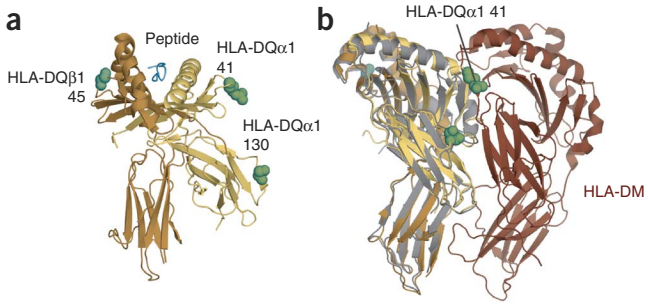

Figure 2 Structural model of the extracellular part of HLA-DQ. The structure for HLA-DQ is based on Protein Data Bank (PDB) entry 4D8P (ref. 10). (**a**) Identified risk variants in HLA-DQβ1 (orange) and HLA-DQα1 (yellow) are highlighted as green spheres. (**b**) Superposition of HLA-DQ (orange and yellow) on HLA-DR (gray; PDB entry 4FQX; ref. 10) bound to HLA-DM (red). HLA-DR and HLA-DQ are structural homologs. In HLA-DR, position 41 directly binds to HLA-DM[7], suggesting that the corresponding position in HLA-DQα1 might also represent the binding site for HLA-DM. Through its interaction, HLA-DM promotes antigen loading onto HLA receptors in endosomes.

showed almost the same frequencies in cases and controls as in our initial sample (replication study: 9.30% in cases versus 3.80% in controls; $P = 7.09 \times 10^{-9}$; RR = 2.59). In the combined sample (1,480 cases and 5,262 controls), the 8-residue HLA-DQβ1 insertion was associated with achalasia with $P_{combined} = 7.72 \times 10^{-26}$ (RR = 2.47).

To interpret the functional consequences of the identified risk variants, we mapped them onto the crystal structure of HLA-DQ (**Fig. 2**). This mapping showed that HLA-DQβ1 position 45 is located near the peptide-binding site, and its alteration might therefore perturb antigen presentation (**Fig. 2a**). Although both amino acid substitutions in HLA-DQα1 had no apparent structural consequences (**Fig. 2a**), there was evidence that position 41 represents the binding site necessary for the interaction between HLA-DQ and the peptide exchange chaperone HLA-DM[7] (**Fig. 2b**). However, it would be premature to conclusively support position 41 as the causal variant over position 130 without additional functional studies. The strongest achalasia risk variant (the eight-residue insertion in HLA-DQβ1) cannot be represented within the crystal structure because of its cytoplasmic localization. One hypothesis is that the insertion might lead to altered intracellular trafficking, as the C-terminal tails of HLA molecules are thought to promote these cellular processes[8]. Functional studies are now required to demonstrate the exact cellular effect of the eight-residue insertion.

Finally, we tested whether genetic risk variants for other immune-mediated diseases were enriched in achalasia cases, as shared genetic risk factors among autoimmune diseases have been observed[9]. We focused on SNPs that had associations reaching genome-wide significance for at least 1 of 15 different immune-mediated disorders. Details on the enrichment analysis are given in the Online Methods. We observed no evidence of shared risk loci in achalasia and any of the 15 autoimmune disorders (**Supplementary Table 8**). Also, at the single-marker level, none of the SNPs selected for the enrichment study appeared to be associated with achalasia after correction for multiple testing (data not shown).

In conclusion, we have identified genetic risk factors for achalasia within the HLA-DQ receptor, thereby confirming a key role for immune-mediated processes in the disease. The strongest signal maps to an eight-residue insertion in HLA-DQβ1, and, to the best of our knowledge, this is the first risk variant in the cytoplasmic tail of an MHC class II receptor that has been identified in a multifactorial disease. In addition, two amino acid substitutions in the extracellular domain of HLA-DQα1 and HLA-DQβ1 confer achalasia risk. All identified variants are located on different HLA haplotypes, which might explain why previous studies of HLA haplotypes in achalasia (**Supplementary Table 9**) failed to find any conclusive association. In this regard, our study exemplifies the value of imputation followed by conditional association analysis as an approach to identify disease-associated variants within the MHC region. Furthermore, given the prevalence of achalasia, our study highlights the role of common genetic variants in rare diseases.

URLs. Genome-wide association study (GWAS) catalog, http://www.genome.gov/26525384/; HLA Explorer, http://www.hlaexplorer.net/; IMGT/HLA database, http://www.ebi.ac.uk/ipd/imgt/hla/; Protein Data Bank (PDB), http://www.rcsb.org/pdb/; R Project for Statistical Computing, http://www.r-project.org/; SNP2HLA, http://www.broadinstitute.org/mpg/snp2hla/; UCSC Genome Browser, http://genome.ucsc.edu/.

METHODS

Methods and any associated references are available in the online version of the paper.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS

We thank all subjects for participating to this study. We acknowledge our collaborating clinical partners, all colleagues who contributed to patient recruitment, and our laboratory technicians and colleagues responsible for database management (for a complete list of all individuals, see the **Supplementary Note**). I.G., M.K. and J.S. received support for this work from the Deutsche Forschungsgemeinschaft (DFG), individual grants GO 1795/1-1, KN 378/2-1 and SCHU 1596/5-1. M.M.N. received support for this work from the Alfried Krupp von Bohlen und Halbach-Stiftung. M.M.N. is a member of the DFG-funded Excellence Cluster ImmunoSensation. The Heinz Nixdorf Recall cohort was established with the generous support of the Heinz Nixdorf Foundation, Germany. We thank B. Pötzsch (University of Bonn) for help with collecting DNA samples from anonymous blood donors. In addition, we thank the Type I Diabetes Genetics Consortium (T1DGC) for data access and J. Sauter (German Bone Marrow Donor Center (DKMS), Tübingen) for help differentiating *HLA-DQA1* alleles. P.I.W.d.B. is the recipient of a Vernieuwingsimpuls VIDI Award from the Netherlands Organization for Scientific Research (NWO project 016.126.354). M.M.W. is supported by a postdoctoral fellowship of the Fund for Scientific Research (FWO) Flanders, Belgium. G.E.B. is supported by a grant from the Research Foundation–Flanders (FWO) (Odysseus program). K.-P.H. is supported by DFG SFB 684 and the Center for Integrated Protein Sciences Munich.

AUTHOR CONTRIBUTIONS

H.L., I.G., J.B., J.S., M.M.N. and M.T.H. initiated the study. J.B., J.S., K.-P.H., M.K. and P.I.W.d.B. analyzed and interpreted the data. D.R., J.B., J.T., J.Z., M.M.W. and T.H. prepared the DNA and performed the molecular genetic experiments. A.G.V., A.J.E., A.L., A.R.d.L., B.H.A.v.R., D.D., E.M., E.U., G.E.B., G.S., H.-D.A., H.R.G., H.G.S., I.G., J.P.d.l.S., J.U.M., L.L., M. Müller, P.R.G., R.C., R.K., S. Heilmann, S.N., U.F., V.A. and W.K. clinically characterized the achalasia cases and/or collected blood samples from cases and/or controls. C.W., G.T., H.-J.W., L.F. and V.K. provided genotype data from the Dutch, Italian and Spanish controls. M. Mattheisen, P.H. and S. Herms provided database management. I.G., J.B., J.S., M.K. and M.M.N. prepared the manuscript, with feedback from the other authors.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

1. Gockel, H.R. *et al.* Achalasia: will genetic studies provide insights? *Hum. Genet.* **128**, 353–364 (2010).
2. Park, W. & Vaezi, M.F. Etiology and pathogenesis of achalasia: the current understanding. *Am. J. Gastroenterol.* **100**, 1404–1414 (2005).

3. de Bakker, P.I. *et al.* A high-resolution HLA and SNP haplotype map for disease association studies in the extended human MHC. *Nat. Genet.* **38**, 1166–1172 (2006).
4. Jia, X. *et al.* Imputing amino acid polymorphisms in human leukocyte antigens. *PLoS ONE* **8**, e64683 (2013).
5. Senju, S. *et al.* Allele-specific expression of the cytoplasmic exon of *HLA-DQB1* gene. *Immunogenetics* **36**, 319–325 (1992).
6. Raychaudhuri, S. *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat. Genet.* **44**, 291–296 (2012).
7. Pos, W. *et al.* Crystal structure of the HLA-DM–HLA-DR1 complex defines mechanisms for rapid peptide selection. *Cell* **151**, 1557–1568 (2012).
8. Khalil, H. *et al.* The MHC class II β chain cytoplasmic tail overcomes the invariant chain p35-encoded endoplasmic reticulum retention signal. *Int. Immunol.* **15**, 1249–1263 (2003).
9. Parkes, M., Cortes, A., van Heel, D.A. & Brown, M.A. Genetic insights into common pathways and complex relationships among immune-mediated diseases. *Nat. Rev. Genet.* **14**, 661–673 (2013).
10. Tollefsen, S. *et al.* Structural and functional studies of *trans*-encoded HLA-DQ2.3 (DQA1*03:01/DQB1*02:01) protein molecule. *J. Biol. Chem.* **287**, 13611–13619 (2012).
11. Robinson, J. *et al.* The IMGT/HLA database. *Nucleic Acids Res.* **39**, D1171–D1176 (2011).

ONLINE METHODS

Sample description. All 1,480 achalasia cases were diagnosed by a board-certified gastroenterologist. The basic diagnostic procedure included esophageal manometry and/or esophagography. In cases where a secondary achalasia could not be excluded, esophagogastroscopy and/or computed tomography or nuclear magnetic resonance (CT/NMR) imaging was conducted. All clinical centers involved in the recruitment of cases and controls received approval from the appropriate institutional review board. All participants signed informed consent. The ancestry of cases and controls was assessed by identifying the origin of the grandparents.

Immunochip study. We recruited 630 achalasia cases of central European descent at the university hospitals of Mainz (Germany; n = 350), Leuven (Belgium; n = 135) and Amsterdam (the Netherlands; n = 145). We also recruited 273 achalasia cases of Spanish origin at the Hospital Clínico San Carlos in Madrid. The 165 Italian achalasia cases were recruited at the university hospitals of San Giovanni Rotondo (n = 76), Milan (n = 38) and Naples (n = 51). A total of 503 cases were females, and 565 cases were males. The population-based controls for the central European sample (n = 2,652) included 1,489 German individuals from the Heinz Nixdorf Recall (HNR) study, a population-based cohort to study risk factors for cardiovascular diseases[12], and 1,163 individuals from the Netherlands who have been described previously[13]. The Spanish (n = 327) and Italian (n = 1,263) controls were unselected healthy individuals and hospital employees and have also been described previously[13]. A total of 2,423 controls were females, and 1,819 controls were males. Multidimensional scaling (MDS) analysis (**Supplementary Fig. 3**) showed the homogeneity in ancestry of the European sample and found no evidence for population outliers.

Replication study of rs28688207 (causing the eight-residue insertion in HLA-DQβ1). We recruited 412 cases (201 females and 211 males) of central European descent at the university hospitals of Mainz (Germany; n = 350), Leuven (Belgium; n = 27) and Amsterdam (the Netherlands; n = 35). The 1,020 controls were of central European descent (523 females and 497 males) and were recruited at the University Hospital Bonn, Germany.

Genotyping and quality control using the Illumina Immunochip. Whereas all achalasia cases (n = 1,068) and the central European controls from Germany (n = 1,489) were genotyped within this study at the Life & Brain Center (Bonn, Germany), genotypes for the remaining controls from the Netherlands (part of the central European controls) and Spain and Italy (n = 2,753) were obtained from a study performed previously[13]. Immunochip markers were removed from this study if, in any sample, a marker was monomorphic, the minor allele frequency (MAF) was <1%, the call rate was <95% or the Hardy-Weinberg equilibrium P value was $<1.0 \times 10^{-4}$ in controls or $<1.0 \times 10^{-6}$ in cases. Unrelatedness for individuals was confirmed by the program KING[14] and an identical-by-state (IBS)-based in-house program. The genomic control inflation factor inferred from 2,629 SNPs associated with reading and writing (P.I.W.d.B. and J.C. Barrett, unpublished data) was $\lambda_{GC} = 1.071$.

Genotyping and quality control of rs28688207. To assess the quality of the imputed eight-residue insertion in HLA-DQβ1, we genotyped the corresponding SNP marker rs28688207. Genotyping was performed in a sample of 350 German cases who were included in the Immunochip study, using a custom TaqMan genotyping assay (Applied Biosystems). Non-missing genotypes were obtained for 345 of the 350 samples studied, resulting in a SNP call rate of >98.5%. The same custom TaqMan genotyping assay was used for the replication study of rs28688207 in our 412 independent achalasia cases and 1,020 independent controls of central European descent.

Imputing HLA genotypes and validation. As previously described[6], we imputed classical HLA haplotypes, their corresponding amino acid changes and SNPs using reference data collected by the Type 1 Diabetes Genetics Consortium (T1DGC). This reference data set contains genotype data for 5,868 SNPs and classical haplotypes for *HLA-A*, *HLA-B*, *HLA-C*, *HLA-DRB1*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1* at 4-digit resolution in 5,225 unrelated individuals of European descent[4]. In total, 4,586 SNPs overlapped in our Immunochip data and the T1DGC reference panel. We encoded all variants in the reference panel as biallelic markers, which facilitated the application of Beagle for imputation (using default parameters)[15]. For each of the three data sets (central European, Italian and Spanish), we imputed cases and controls together. The frequencies, RRs and patient/control (P/C) ratios of all imputed HLA haplotypes in the combined sample as well as in the central European, Italian and Spanish subsets are shown in **Supplementary Table 10**.

Logistic regression model for association testing. We used the previously described[6] logistic regression model

$$\ln\frac{p_i}{1-p_i} = \sum_{a=1}^{m-1}\beta_a g_{ai} + \sum_{j=1}^{3}\delta_{ij}\left(\theta_j + \sum_{k=1}^{5}\pi_{jk}e_{ik}\right)$$

to test (Immunochip-genotyped and MHC-imputed) markers for their disease association in the entire sample. Here g_{ai} is the dosage of allele a in individual i. β_a represents the additive effect per allele. For testing a multiallelic locus with m alleles, we included $m-1$ β parameters and selected the most frequent allele in the controls as the reference allele. δ_{ij} is an indicator variable that was equal to 1 only if individual i belonged to sample j. The θ_j parameters are sample-specific intercept parameters. To account for population stratification, we included as covariates the first five components e_{ik} (which were obtained from MDS analysis[16]) and allowed for sample-specific effects π_{jk} for these covariates.

Analysis of HLA-DQ amino acid polymorphisms. To test the effects of a polymorphic amino acid position in HLA-DQ, we first identified the m residues occurring at that position and then partitioned the *HLA-DQB1* and *HLA-DQA1* alleles into m groups of alleles with identical residues at that position. We used SAS[17] to perform the Wald test for the null hypothesis $\beta_1 = \ldots = \beta_{m-1}$ in the logistic regression model described above.

For the conditional analyses, we note that the addition of another position might result in a refinement of allele groups. We assumed that T of the previous allele groups was refined by the addition of another position and let, for $1 \le t \le T$, $a_{s(t)}, \ldots, a_{e(t)}$ denote the subgroups of the previous allele group a_t generated by the additional position. In case there was no effect of the additional position, all parameters representing the effect of different subgroups descending from the same previous allele group should be identical. Therefore, we assessed the significance of the additional position by performing the Wald test for the null hypothesis $\beta_{as(1)} = \ldots = \beta_{ae(1)}, \ldots, \beta_{as(T)} = \ldots = \beta_{ae(T)}$.

Enrichment analysis. To test for shared genetic factors between other immune-mediated diseases and achalasia, we extracted all markers from the GWAS catalog (accessed 7 November 2013) that were associated at genome-wide significance ($P < 5 \times 10^{-8}$) with a total of 15 immune-mediated disorders (**Supplementary Table 8**). Markers were excluded from the study if they met one of the following criteria: (i) they were located within the MHC region; (ii) they were missing data for the risk allele; (iii) they were AT or CG markers; (iv) they were only reported in GWAS samples with African or Asian ancestry; and (v) they were not present in our Immunochip data set after quality control. This filtering resulted in a total of 346 markers, which were further filtered through LD pruning for each of the 15 immune-mediated disorders separately and for all immune-mediated disorders combined. For LD pruning, we used the controls of our Immunochip study to remove markers with $r^2 > 0.5$ to other markers that were disease associated with stronger odds ratios. The numbers of SNP markers before and after LD pruning are listed in **Supplementary Table 8**. For each condition, an LD-pruned marker set ($r^2 < 0.5$) was used for the enrichment analysis. The score for each individual in our Immunochip sample was calculated as the number of risk alleles weighted by the log of the odds ratio taken from the GWAS catalog. The effect of the score was tested by Wald test in the logistic regression analysis of disease state on score (with the first five MDS components again included as additional covariates).

12. Schmermund, A. *et al.* Assessment of clinically silent atherosclerotic disease and established and novel risk factors for predicting myocardial infarction and cardiac death in healthy middle-aged subjects: rationale and design of the Heinz Nixdorf RECALL Study. Risk Factors, Evaluation of Coronary Calcium and Lifestyle. *Am. Heart J.* **144**, 212–218 (2002).
13. Trynka, G. *et al.* Dense genotyping identifies and localizes multiple common and rare variant association signals in celiac disease. *Nat. Genet.* **43**, 1193–1201 (2011).
14. Manichaikul, A. *et al.* Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873 (2010).
15. Browning, B.L. & Browning, S.R. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *Am. J. Hum. Genet.* **84**, 210–223 (2009).
16. Purcell, S. *et al.* PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
17. SAS Institute, Inc. *SAS/STAT 9.2 User's Guide* (SAS Institute, Inc., Cary, NC, 2008).

doi:10.1038/ng.3029