Version: Oct 1 2014

High density mapping of the MHC reveals a common role for HLA-DRB1*01:03 in IBD and heterozygous advantage in ulcerative colitis

Philippe Goyette[1] *, Gabrielle Boucher[1]*, Dermot Mallon [2], Eva Ellinghaus[3], Luke Jostins[4] Hailiang Huang[5], Stephan Ripke[5], Elena Gusareva[6], Vito Annese[7], Stephen L. Hauser[8], Jorge R. Oksenberg[8], Ingo Thomsen[3], Stephen Leslie[9,10], International IBD Genetics Consortium, Mark J. Daly[5], Kristel van Steen[6], Richard H. Duerr[11], Jeff Barrett[12], Dermot McGovern[13], Phil Schumm[14], James A. Traherne[15], Mary N. Carrington[16], Vasilis Kosmoliaptsis[2], Tom H. Karlsen[17,18,19]*, Andre Franke[3]*, John D. Rioux[1,20]*

Affiliations:
[1]Montreal Heart Institute; [2]Department of Surgery, University of Cambridge and NIHR Cambridge Biomedical Research Centre; [3]Christian-Albrechts-University Kiel; [4]Wellcome Trust Centre for Human Genetics; [5]Massachusetts General Hospital & the Broad Institute; [6]University of Liège; [7]Florence University Hospital Careggi; [8]University of California San Francisco; [9]Murdoch Childrens Research Institute; [10]University of Melbourne Department of Mathematics and Statistics; [11]University of Pittsburgh School of Medicine and Graduate School of Public Health; [12]Sanger Institute; [13]Cedars Sinai Medical Center; [14]University of Chicago; [15] Department of Pathology, Cambridge Institute for Medical Research, University of Cambridge, United Kingdom; [16] Cancer and Inflammation Program, Laboratory of Experimental Immunology, Leidos Biomedical Research, Inc., Frederick National Laboratory for Cancer Research and Ragon Institute of MGH, MIT and Harvard;[17]Norwegian PSC Research Center, Department of Transplantation Medicine, Division of Cancer, Surgery and Transplantation, Oslo University Hospital, Oslo, Norway; [18]Institute of Clinical Medicine, University of Oslo, Oslo, Norway; [19]K.G. Jebsen Inflammation Research Centre, Research Institute of Internal Medicine, Division of Cancer Medicine, Surgery and Transplantation, Oslo University Hospital, Rikshospitalet, Oslo, Norway; [20]Université de Montréal.

* These authors contributed equally to this work.

Correspondence and requests for materials should be addressed to J.D.R. (John.david.rioux@umontreal.ca).

Recent genome-wide association studies of the highly related chronic inflammatory bowel diseases (IBD) known as Crohn's disease (CD) and ulcerative colitis (UC) have shown strong evidence of association to the major histocompatibility complex (MHC). This region encodes a large number of immunological candidate genes including those that encode the human leukocyte antigens (HLA) that present peptide antigens to T lymphocytes[1]. Previous studies of this region in IBD have indicated that multiple independent associations are likely to exist at HLA genes and non-HLA genes, but lacked the statistical power to define the architecture of association and plausible causal alleles [2, 3]. To address this, we performed high-density SNP typing of the MHC in >32,000 patients with IBD, implicating multiple HLA alleles, with a primary role for HLA-DRB1*01:03 in both CD and UC. Significant differences, however, were observed between these two diseases, including a predominant role of class II HLA variants and heterozygous advantage observed in UC, suggesting an important role of the adaptive immune response to the colonic environment in UC pathogenesis.

Meta-analyses of genome-wide association studies (GWAS) have recently shown that CD (MIM266600) and UC (MIM191390) share the majority of the 163 known genetic risk factors for IBD, with the MHC being one of the notable exceptions [4]. Data from these GWAS, however, have had insufficient variant density to define the association signals within the MHC. Targeted studies of IBD with higher variant density within the MHC region but with modest sample sizes have indicated that multiple independent associations are likely to exist at HLA genes and non-HLA genes, with the most consistent associations being to HLA class II loci, mainly *HLA-DRB1* and *HLA-DQB1*, with some reports of association at the *HLA-C* class I locus and potentially also at non-HLA genes [2, 3, 5-8]. In the current study we generated high quality genotypes for 7,406 SNPs within the MHC region on a total of 18,405 patients with CD, 14,308 patients with UC and 34,241 controls subjects. Using this SNP data, we imputed and benchmarked the genetic variation within the class I (*HLA-B*, *HLA-C*, and *HLA-A*) and class II (*HLA-DRB1*, *HLA-DRB3*, *HLA-DRB4*, *HLA-DRB5*, *HLA-DQA1*, *HLA-DQB1*, *HLA-DPA1* and *HLA-DPB1*) HLA genes at the level of classical HLA alleles and amino acid positions (please refer to the Online Methods).

As a first step to defining the nature of the association to CD and UC within the MHC, we performed univariate analyses of the SNPs, classical HLA alleles, and HLA amino acids. These analyses revealed a very large number of variants across the MHC region with significant association to these phenotypes (**Fig. 1**), with major peaks of association centered in and around the classical HLA genes, suggesting a role for classical HLA alleles in CD and UC risk. This observation is consistent with gene-based analyses, which show strong association at the HLA genes for both UC and CD (e.g. $P<1\times10^{-300}$ for *HLA-DRB1* in UC) (**Supplementary Table 1**). In particular, these analyses demonstrated a role of *HLA-DRB1* that cannot be attributed to other HLA genes, with evidence of residual association in class I and class II regions (**Supplementary Table 1**). In order to be more quantitative, we calculated the variance explained by the class I and class II alleles. Whereas the contribution of class I and class II alleles are relatively equivalent in CD, not only is the overall impact of HLA on disease risk greater in UC, but the alleles in the class II region have nearly three-fold greater impact than class I alleles (**Fig. 2).** Moreover, these analyses have revealed that classical HLA alleles explain three- to ten-fold more of the disease variance than that explained by the index SNPs that were previously identified (~3% vs ~0.3% in CD; ~6% vs ~2% in UC) (**Fig. 2).**

Specifically, in our univariate analyses, the most significant association in CD is to HLA-DRB1*01:03 ($P<4\times10^{-62}$, OR= 2.53), with a p-value over 10 orders of magnitude more significant than the next best associated variants in the region. Importantly, HLA-DRB1*01:03 has an effect

in CD which is statistically independent from the other most associated variants in the MHC, as shown by reciprocal conditional logistic regression (**Supplementary Fig. 1**). While the single most significant variant in the UC univariate analysis is a non-coding SNP (rs6927022, $P<5\times10^{-153}$, OR=1.49) near *HLA-DQA1,* previously identified in the recent meta-analysis of GWAS [4], strikingly the next strongest independent association is also to HLA-DRB1*01:03, with a much greater OR ($P<1\times10^{-120}$, OR=3.63; $P_{cond}<2\times10^{-89}$, OR=3.06) (**Supplementary Fig.2**). Reciprocal conditioning on HLA-DRB1*01:03 did not abolish the effect seen at rs6927022 ($P<9\times10^{-123}$, OR=1.43), indicating that these have mostly statistically independent effects in UC. Taken together, our analyses point to HLA-DRB1*01:03 as likely being causal in both diseases, with additional causal alleles in the class II and class I regions. Given this observation, it is probable that additional alleles within *HLA-DRB1* contribute to IBD risk.

We thus examined an *HLA-DRB1*-centric model and identified seven *HLA-DRB1* alleles with independent effects on CD risk (study-wide significance threshold of 5×10^{-6}) (**Supplementary Table 2**). Moreover, when controlling for these seven *HLA-DRB1* alleles, we identified only a single additional class II allele (HLA-DPA1*01:03) independently associated with CD. Using the same conditional logistic regression framework for the analysis of the class I locus, we identified seven class I HLA alleles that are significantly associated with CD, after conditioning on the eight class II alleles (**Fig. 3** and **Supplementary Table 2**). This *HLA-DRB1*-centric model explains about 2% of disease variance **(Fig. 2).** In UC, we identified a total of 12 *HLA-DRB1* alleles, 1 *HLA-DPB1* allele and 3 class I alleles (**Supplementary Table 3**) that can explain the association to the MHC and which account for about 5% of disease variance (**Fig. 2**).

As can be seen in **Figure 3**, for many of the alleles identified in the *HLA-DRB1*-centric model, a few other candidate alleles in class I or class II can be considered. In particular, multiple HLA-DRB1 alleles have equivalent associations at *HLA-DQA1* and *HLA-DQB1* (e.g. HLA-DQA1*03:01 is equivalent to HLA-DRB1*04 and HLA-DRB1*09 alleles in UC) equally supporting a role for genetic variation within *HLA-DQA1* and/or *HLA-DQB1* in disease susceptibility, particularly for UC (**Fig. 3**). However, several of the alleles in these models, including HLA-DRB1*01:03, do not have any such proxies and thus are strong candidates for being causal (**Fig. 3**). Further dissection of these class II correlated signals for identifying potential causal alleles may only be feasible in admixed or ethnically diverse populations [9]. Further refinement may also be possible by examining the impact of clinical sub-phenotype and associated autoimmune co-morbidities on observed associations, although functional studies will be needed to infer causality. For the present analysis we were able to assess the impact of colonic vs. non-colonic inflammation, and found that HLA-DRB1*01:03 is associated with colonic CD and that HLA-DRB1*07:01 is associated with the absence of colon involvement (**Supplementary Fig. 3),** in line with previous suggestions [10]. This explains the shared associations for CD and UC at HLA-DRB1*01:03 and strongly suggests that this allele is critically involved in determining the colonic immune response to local flora.

Given that classical HLA alleles consist of combinations of specific amino acids at multiple positions, we tested whether the association to disease could be better explained by single amino acid positions. Indeed we observed very strong association signals at many single amino acid variants in CD (e.g. five amino acids of HLA-DRβ at positions 67, 70 and 71) and in UC (e.g. 4 amino acid variants of HLA-DQα at positions 50 and 53 and 215 and 4 amino acid variants in HLA-DRβ at positions 98 and 104) and also performed per position omnibus analyses that confirm the predominant association to HLA-DRβ position 11 in UC, as previously reported [5], and to HLA-DRβ position 70 in CD (**Supplementary Tables 4-5** and **Supplementary Fig. 4).** While the hypothesis of a positional effect is appealing, the interpretation of these position-based tests is not straightforward in the context of likely multiple causal alleles **(Supplementary**

Note on amino acids, Supplementary Table 6 and **Supplementary Fig. 5)**. Furthermore in this study the amino acid-based models did not capture the association at *HLA-DRB*1 in a more parsimonious way than the HLA allele-based models (**Supplementary Note on amino acids**). To further explore the basis for the observed HLA associations, we performed three-dimensional protein structure modeling followed by analysis of the electrostatic properties of the binding groove of associated ($P<10^{-4}$) and common (frequency>1%) *HLA-DRB1* alleles. These analyses suggest that HLA-DR alleles associated with increased risk of UC and CD, share common structural and electrostatic properties within or near their peptide binding groove that are largely distinct from those of HLA-DR alleles associated with decreased risk of UC and CD (**Fig. 4**).

While we performed the primary analyses based on a dose effect model, our sample size allowed us to investigate the effects further, by testing for non-additive effects. In fact we found significant departure from additive effects in UC, but not for CD (**Fig 5a-c)**. Specifically, we found evidence of decreased heterozygosity in UC patients for genotyped and imputed variants across the MHC and at HLA genes, mostly in class II (**Supplementary Tables 7-8**). This heterozygote advantage could be explained by an enrichment of dominant protective and recessive risk alleles[11], that is absent or much less important in CD (**Fig. 5** and **Supplementary Fig. 6**). Notably, we also detected multiple overdominant effects in UC, the strongest of which is captured by HLA-DRB1*03:01 (**Fig. 5, Supplementary Fig. 6-7,** and **Supplementary Table 9)**. This allele is mostly found on the ancestral haplotype 8.1, a relatively common (~5-10%) haplotype that is conserved in European populations and that is implicated in other immune diseases [12-14]. The overdominance effect of this haplotype in UC is possibly due to the presence of both dominant protective and recessive risk alleles, which would be consistent with the reported recessive risk of this haplotype in the UC-related biliary disease primary sclerosing cholangitis (**Supplementary Fig. 8-9**) [15, 16]. Analogous with an infectious paradigm [11], these data may suggest that decreased HLA class II heterozygosity may impair the ability to appropriately control colonic microbiota in UC.

Although there is a significant challenge in defining the causal alleles for CD and UC in the MHC given the LD structure in the region, a number of conclusions can be drawn regardless of the models tested. First, the high density mapping of this region in a large cohort revealed the significant contribution of the MHC to disease risk, a contribution that is not apparent in the previous GWAS. Second, for both CD and UC it would appear that variation within HLA genes as opposed to variation in other genes within the MHC plays a predominant role in disease susceptibility. Third, while the contribution of class I and class II HLA variants to disease risk is relatively equivalent in CD, HLA class II variation plays a more important role in UC. Fourth, in contrast to the majority of non-MHC susceptibility loci being shared between CD and UC, most associated HLA alleles have a predominant role in either CD or UC, with very few having shared IBD risk **Fig. 6**). Finally, the decreased heterozygosity in UC suggests that the ability to recognize a broader set of antigens, potentially of colonic microbial origin, is important to mount protective immunity.

Figure Legends

Figure 1. Primary univariate association analyses of CD and UC. Univariate association analysis results for 8,939 SNPs (dark grey) **(Supplementary Table 10)**, 90 2-digit and 138 4-digit resolution HLA alleles (yellow) (**Supplementary Table 11**), as well as 741 single amino acid variants (red) (**Supplementary Table 4**) in the MHC region are shown for 18,405 CD cases and 14,308 UC cases (with 34,241 common control subjects). Given that previous genetic analyses have identified distinct effects in the MHC for CD and UC, with different non-correlated alleles identified in each disease, we opted to perform the finemapping analyses for CD and UC separately. (**a**) The primary univariate association analysis in CD reveals over 1,789 markers showing study-wide significant association ($P<5\times10^{-6}$) across the MHC, including 32 4-digit resolution classical HLA alleles (**Fig. 3** and **Supplemental Table 2**). The single most significant variant for CD is HLA-DRB1*01:03 ($P=3x10^{-62}$, OR= 2.51). (**b**) The primary univariate association analysis in UC reveals over 2,762 markers showing study-wide significant association across the MHC, including 50 4-digit resolution classical HLA alleles (**Fig. 3** and **Supplemental Table 3**). The single most significant variant for UC is rs6927022 ($P=8\times10^{-154}$, OR= 1.49) while the best HLA allele is HLA-DRB1*01:03 ($P=3\times10^{-119}$, OR=3.59).

Figure 2. Variance explained by 4-digit HLA alleles in CD and UC. Proportion of variance explained on a logit scale (McKelvey and Zavoina's Pseudo R^2, see **Online Methods**) for different models in CD (left) and UC (right). The top boxes show the variance explained by previously identified GWAS index SNPs within the MHC[4]. The middle boxes illustrate the variance explained by HLA models including all 4-digit alleles of frequency > 0.5% (126 alleles in CD and UC) and models restricted to 4-digit alleles within either class I (63 alleles) or class II regions (63 alleles), respectively. The Venn diagram illustrates the proportion of variance explained that is unique to class I, class II or shared. The bottom boxes indicate the variance explained by the proposed HLA models (15 and 16 alleles in CD and UC, respectively). To be noted, these estimations of variance explained were performed on the logit scale for practical reasons, and should not be directly compared to heritability estimates computed on the (Gaussian) liability scale.

Figure 3. Correlated association signals at HLA alleles support potential alternate association models for both CD and UC. Equivalence of effect at the different study-wide significant associated 4-digit HLA alleles is shown for **(a)** CD and **(b)** UC. The structures illustrated in the figure are not classically defined haplotype structures, but were identified entirely based on the correlation of signal defined through pairwise reciprocal conditional logistic regression analyses (see **Supplementary tables 2** and **3**); although such correlations are clearly dependent on the underlying haplotypic structure of the region. Alleles identified as primary tags for independent association signals in our *HLA-DRB1* focused models are shown in light blue boxes, while alternate alleles with equivalent effects are shown in grey boxes. Alleles in white boxes show study-wide significant secondary effects that can be explained entirely by the selected HLA alleles. Alleles at the *HLA-DRB3, -DRB4 and -DRB5* genes were omitted in order to simplify the display; many of the alleles at these genes show high frequency and as such are correlated to many different alleles (both risk and protective) at the other class II genes. Of note, the HLA-DRB4*null allele is the second strongest associated allele in UC **(see Supplementary table 3).**

Figure 4. HLA-DR peptide binding groove electrostatic properties and risk of IBD
The electrostatic potential of all HLA-DR alleles associated with UC or CD, and of all common HLA-DR alleles (frequency >1%), was calculated. HLA-DR alleles associated with increased or

decreased risk of IBD at study wide-significance level ($P< 5\times10^{-6}$) are shown in dark red or dark blue, respectively. Respective risk associations at suggestive level ($1\times10^{-4}<P<5\times10^{-6}$) are shown in pale red and pale blue. Electrostatic potential comparisons among HLA-DR molecules were performed in a pairwise, all-versus-all, fashion (see **Online Methods**) to produce distance matrices that are displayed as symmetrical heatmaps (scale ranges from 0 [identical] to 1 [maximum difference]). (**a**) The electrostatic potential in seven regions within the peptide binding groove (see **Online Methods** and **Supplementary Fig. 10**), which interact with the presented peptide, were compared among the HLA-DR alleles and pooled onto a single Euclidian distance matrix. The distance-based clustering identifies four clusters, with an enrichment of risk alleles in two of these. Comparison of the electrostatic potential at individual peptide binding groove regions is shown in **Supplementary Fig. 11**. (**b**) Heatmap representing electrostatic potential differences among the HLA-DR alleles at a spherical region that encompasses amino acid residues 67, 70 and 71 of the HLA-DRβ chain (associated with risk for UC and CD; **Supplementary Table 11**). The distance-based clustering identifies two clusters that correlate with directionality of effect in IBD.

Figure 5. Non-additive effect models in CD and UC. Evidence for non-additive effect of common variants (frequency >5%) across the MHC tested under a general model of additive and dominance effects (**Online Methods**) in CD (**a,b**) and UC (**c,d**). Low frequency variants, such as HLA-DRB1*01:03, were excluded from the analyses of non-additive model due to the low number of homozygotes. The p-values and directionality for departure from additive effect (dominance term) are represented on the y-axis (**a,c**). Variants showing either higher or lower risk than expected for heterozygotes under an additive model are shown on the upper and lower part of the plot respectively. HLA alleles and amino acids variants are in yellow and red respectively, while SNPs are represented in dark grey. The dotted lines represent the study wide significance threshold of 5×10^{-6}, with non-significant variants plotted in less pronounced colors. A clear enrichment for lower risk in heterozygotes is observed in UC (**c**) as suggested by the large number of significant negative dominance term (lower part of the plot). This effect is absent in CD (**a**), or much less important. The dominance term OR for common variants is illustrated (y-axis) versus the additive term (x-axis), giving an overview of the association model (**b,d**). Variants were plotted using the same color code as in (**a,c**). Protective and risk minor alleles are shown on the left and right sides of the plot respectively. Strictly recessive or dominant variants are expected to fall on the diagonals, while strictly additive variants lay on or close to the x-axis. The y-axis is the expected position for pure over/under dominance. Variants showing evidence of overdominance fall in the lower triangle (light blue). In UC, many alleles with non-additive effects fall into the bottom left (protective dominant) and bottom right (risk recessive) quadrants of the plots, with some being close to the vertical axis in the lower triangle (overdominance) (see **Supplementary Table 9** for pairwise comparison of HLA-DRB1 alleles in UC). As an example, HLA-DRB1*03:01 and HLA-DQB1*02:01 are located between pure overdominance and pure recessive risk (**Supplementary Fig. 6**). These non-additive effects are observed for many variants in UC, but are mostly absent in CD (**a,b**); notable exception being the HLA-B*08 allele (**Supplementary Fig. 6**).

Figure 6. Comparison of odds ratio in CD and UC for HLA alleles identified from HLA-focused models. Odds ratio (OR) from the primary univariate association analyses in CD and UC for all alleles identified in the HLA-focused models of CD and/or UC are presented (**a**). Odds ratio for CD and UC are in blue and red respectively; darker colors indicate study-wide significant effect ($P< 5\times10^{-6}$), lighter colors indicate nominal significance level ($0.05>P\geq5\times10^{-6}$) and white indicates non-significance ($P\geq0.05$) (for specific effect and significance values refer to **Fig. 3** and **Supplementary Tables 2** and **3**). Allele HLA-B*52:01 is indicated for UC in place of

the equivalent HLA-C*12:02 to simplify the display of this shared signal. For the same HLA alleles, odds ratio for an IBD analysis are plotted against the odds ratio for the CD versus UC analysis with the IBD risk allele as the reference (**b**). Alleles identified as significant in CD or UC only are plotted in blue and red, respectively. Variants identified as significant in both are shown in purple, with empty circles representing opposite direction of effect. Shared association signals are expected to fall in the upper triangle of the plot. Most variants fall outside of this region, highlighting the difference between CD and UC in the MHC (the same representation with all MHC variants is available in **Supplementary Figure 12**).

METHODS

Please refer to attached files

ACKNOWLEDGEMENTS

This project has been funded in whole or in part with federal funds from the Frederick National Laboratory for Cancer Research, under Contract No. HHSN261200800001E. The content of this publication does not necessarily reflect the views or policies of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government. This Research was supported in part by the Intramural Research Program of the NIH, Frederick National Lab, Center for Cancer Research (M.N.C). S.L. wishes to acknowledge the support from the Australian National Health and Medical Research Council (RD Wright Career Development Fellowship, APP1053756). J.B. was supported by a Wellcome Trust grant (#WT098051). D.H.M. and V.K. are supported by the NIHR Cambridge Biomedical Research Centre. P.S. is supported by the NIDDK IBD Genetics Consortium Data Coordinating Center grant (7 U01 DK062429-14). J.T. is supported by the Medical Research Council. DPBM is supported by The Leona M. and Harry B. Helmsley Charitable Trust, The European Union (305479), and grants DK062413, DK046763-19, AI067068, HS021747 and U54DE023789-01. R.H.D is supported by the U.S. National Institute of Diabetes and Digestive and Kidney Diseases (DK062420). S.L.H. and J.R.O. would like to acknowledge the support of National Institutes of Health (U19AI067152; ARRA administrative supplement). Grant support for THK was received from EU 7th Framework Programme (FP7/2007-2013, grant number 262055, ESGI). J.D.R. holds a Canada Research Chair and his current work was supported by grants from the U.S. National Institute of Diabetes and Digestive and Kidney Diseases (DK064869; DK062432). We are grateful to Benedicte A. Lie and Kristian Holm for helpful discussions.

Members of the International Inflammatory Bowel Disease Genetics Consortium:

Clara Abraham[12], Jean-Paul Achkar[25,26], Tariq Ahmad[27], Leila Amininejad [28], Ashwin Ananthakrishnan[29], Vibeke Andersen[30], Carl A Anderson[1], Jane M Andrews[31], Vito Annese[52,79], Guy Aumais[86], Leonard Baidoo[5], Tobias Balschun[32], Peter A Bampton[33], Murray Barclay[44], Jeffrey C Barrett[1], Theodore M Bayless[88], Joshua C Bis[89], Alain Bitton[34], Gabrielle Boucher[23], Stephan Brand[35], Steven R Brant[82] , Carsten Büning[36], Mathias Chamaillard[90], Angela Chew[54], Judy H Cho[9,12], Sven Cichon[38], Isabelle Cleynen[16], Ariella Cohain[37], Jean Frederic Colombel[91], Anthony Croft[69], Mauro D'Amato[39], Mark J Daly[2,3], Silvio Danese[87], Dirk De Jong[4], Goda Denapiene[92], Olivier Dewit[93], Kathy L Devaney[29], Renata D'Inca[94], Marla Dubinsky[40], Richard H Duerr[5,6], Cathryn Edwards[41], David Ellinghaus[32], Jonah Essers[13], Lynnette R Ferguson[42], Eleonora A Festen[4], Philip Fleshner[7], Tim Florin[95], Denis Franchimont[28], Andre Franke[32], Karin Fransen[5,43], Richard Gearry[44,45], Michel Georges[17], Christian Gieger[46], Jürgen Glas[34], Philippe Goyette[23], Todd Green[13], Katherine Hanigan[69], Nicholas K Hayward[101], Hakon Hakonarson[80,81], Talin Haritunians[7], Ailsa Hart[47], Chris Hawkey[48], Matija Hedl[12], Xinli Hu[20], Hailiang Huang[3], Andrew Ippoliti[7], Laimas Jonaitis[50], Luke Jostins[1], Tom H Karlsen[49],Gediminas Kiudelis[50], Limas Kupcinskas[50], Subra Kugathasan[51], Anna Latiano[52], Debby Laukens[53], Ian C Lawrance[54], James C Lee[10], Charlie W Lees[55], Paolo Lionetti[96], Jimmy Z Liu[1], Edouard Louis[18], Gillian Mahy[56], John Mansfield[57], Christopher G Mathew[19], Dermot P McGovern[7,8,] Mitja Mitrovic[14,43], Raquel Milgrom[78], Grant Montgomery[101], Craig Mowat[58], William Newman[59], Kaida Ning[12], Orazio Palmieri[52], Miles Parkes[84], Cyriel Y Ponsioen[60], Uros Potocnik[14,61], Natalie J Prescott[62], Graham

Radford-Smith[69,83], Jean-Francois Rahier[97], Soumya Raychaudhuri[20,21,22], Miguel Regueiro[5], Florian Rieder[25], John D Rioux[23], Stephan Ripke[2,3], Rebecca Roberts[44], Richard K Russell[63], Jeremy D Sanderson[64], Miquel Sans[65,66], Jack Satsangi[55], Eric E Schadt[37], Stefan Schreiber[67,68], L Philip Schumm[11], Mark Seielstad[98,99], Yashoda Sharma[12], Mark S Silverberg[78], Lisa A Simms[69], Jurrgita Skieceviciene[50], Sarah L Spain[19], Hillary A Steinhart[78], Joanne M Stempak[78], Jurgita Sventoraityte[50], Stephan R Targan[7], Kent D Taylor[7,8], Anje ter Velde[60], Emilie Theatre[17,18], Mark Tremelling[70], Andrea van der Meulen[71], Suzanne van Sommeren[4], Eric Vasiliauskas[7], Severine Vermeire[85], Hein W Verspaget[71], Martine De Vos[53], Ming-His Wang[82], Rinse K Weersma[4], Zhi Wei[24], David Whiteman[100], Cisca Wijmenga[43], David C Wilson[63,72], Juliane Winkelmann[73], Ramnik J Xavier[29,74], Sebastian Zeissig[75], Clarence K Zhang[76], Zhen Zhao[101], Hongyu Zhao[76]

Affiliations:
[1]Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, UK. [2]Analytic and Translational Genetics Unit, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [3]Broad Institute of MIT and Harvard, Cambridge, Massachusetts, USA. [4]Department of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands. [5]Division of Gastroenterology, Hepatology and Nutrition, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, Pennsylvania, USA. [6]Department of Human Genetics, University of Pittsburgh Graduate School of Public Health, Pittsburgh, Pennsylvania, USA. [7]Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [8]Medical Genetics Institute, Cedars-Sinai Medical Center, Los Angeles, California, USA. [9]Department of Genetics, Yale School of Medicine, New Haven, Connecticut, USA. [10]Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK. [11]Department of Health Studies, University of Chicago, Chicago, Illinois, USA. [12]Department of Internal Medicine, Section of Digestive Diseases, Yale School of Medicine, New Haven, Connecticut, USA. [13]Center for Human Genetic Research, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts, USA. [14]University of Maribor, Faculty of Medicine, Center for Human Molecular Genetics and Pharmacogenomics, Maribor, Slovenia. [15]University Medical Center Groningen, Department of Genetics, Groningen, The Netherlands. [16]Department of Pathophysiology, Gastroenterology section, KU Leuven, Leuven, Belgium. [17]Unit of Animal Genomics, Groupe Interdisciplinaire de Genoproteomique Appliquee (GIGA-R) and Faculty of Veterinary Medicine, University of Liege, Liege, Belgium. [18]Division of Gastroenterology, Centre Hospitalier Universitaire, Universite de Liege, Liege, Belgium. [19]Department of Medical and Molecular Genetics, King's College London School of Medicine, Guy's Hospital, London, UK. [20]Division of Rheumatology Immunology and Allergy, Brigham and Women's Hospital, Boston, Massachusetts, USA. [21]Program in Medical and Population Genetics, Broad Institute, Cambridge, Massachusetts, USA. [22]Division of Genetics, Brigham and Women's Hospital, Boston, Massachusetts, USA. [23]Université de Montréal and the Montreal Heart Institute, Research Center, Montréal, Québec, Canada. [24]Department of Computer Science, New Jersey Institute of Technology, Newark, NJ 07102, USA. [25]Department of Gastroenterology & Hepatology, Digestive Disease Institute, Cleveland Clinic, Cleveland, Ohio. [26]Department of Pathobiology, Lerner Research Institute, Cleveland Clinic, Cleveland, Ohio, USA. [27]Peninsula College of Medicine and Dentistry, Exeter, UK. [28]Erasmus Hospital, Free University of Brussels, Department of Gastroenterology, Brussels, Belgium. [29]Massachusetts General Hospital, Harvard Medical School, Gastroenterology Unit, Boston, Massachusetts, USA. [30]Viborg Regional Hospital, Medical Department, Viborg, Denmark. [31]Inflammatory Bowel Disease Service, Department of Gastroenterology and Hepatology, Royal Adelaide Hospital, and School of Medicine, University of Adelaide, Adelaide, Australia. [32]Institute of Clinical Molecular Biology, Christian-Albrechts-University, Kiel, Germany.

[33]Department of Gastroenterology and Hepatology, Flinders Medical Centre and School of Medicine, Flinders University, Adelaide, Australia. [34]Division of Gastroenterology, McGill University Health Centre, Royal Victoria Hospital, Montréal, Québec, Canada. [35]Department of Medicine II, University Hospital Munich-Grosshadern, Ludwig-Maximilians-University, Munich, Germany. [36]Department of Gastroenterology, Charit, Campus Mitte, UniversitŠtsmedizin Berlin, Berlin, Germany. [37]Department of Genetics and Genomic Sciences, Mount Sinai School of Medicine, New York City, New York, USA. [38]Department of Genomics, Life & Brain Center, University Hospital Bonn, Bonn, Germany. [39]Department of Biosciences and Nutrition, Karolinska Institutet, Stockholm, Sweden. [40]Department of Pediatrics, Cedars Sinai Medical Center, Los Angeles, California, USA. [41]Torbay Hospital, Department of Gastroenterology, Torbay, Devon, UK. [42]School of Medical Sciences, Faculty of Medical & Health Sciences, The University of Auckland, Auckland, New Zealand. [43]Department of Genetics, University Medical Center Groningen, Groningen, The Netherlands. [44]Department of Medicine, University of Otago, Christchurch, New Zealand. [45]Department of Gastroenterology, Christchurch Hospital, Christchurch, New Zealand. [46]Institute of Epidemiology, Helmholtz Center Munich, German Research Center for Environmental Health (GmbH), Neuherberg, Germany. [47]St Mark's Hospital, Watford Road, Harrow, Middlesex, HA1 3UJ. [48]Nottingham Digestive Diseases Centre, Queens Medical Centre, Nottingham NG7 1AW, UK. [49]Rikshospitalet University Hospital, Medical Department, Oslo, Norway. [50]Kaunas University of Medicine, Department of Gastroenterology, Kaunas, Lithuania. [51]Department of Pediatrics, Emory University School of Medicine, Atlanta, Georgia, USA. [52]Unit of Gastroenterology, Istituto di Ricovero e Cura a Carattere Scientifico-Casa Sollievo della Sofferenza (IRCCS-CSS) Hospital, San Giovanni Rotondo, Italy. [53]Ghent University Hospital, Department of Gastroenterology and Hepatology, Ghent, Belgium. [54]School of Medicine and Pharmacology, The University of Western Australia, Fremantle, Australia. [55]Gastrointestinal Unit, Molecular Medicine Centre, University of Edinburgh, Western General Hospital, Edinburgh, UK. [56]Department of Gastroenterology, The Townsville Hospital, Townsville, Australia. [57]Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK. [58]Department of Medicine, Ninewells Hospital and Medical School, Dundee, UK. [59]Genetic Medicine, MAHSC, University of Manchester, Manchester, UK. [60]Academic Medical Center, Department of Gastroenterology, Amsterdam, The Netherlands. [61]University of Maribor, Faculty for Chemistry and Chemical Engineering, Maribor, Slovenia. [62]King's College London School of Medicine, Guy's Hospital, Department of Medical and Molecular Genetics, London, UK. [63]Royal Hospital for Sick Children, Paediatric Gastroenterology and Nutrition, Glasgow, UK. [64]Guy's & St. Thomas' NHS Foundation Trust, St. Thomas' Hospital, Department of Gastroenterology, London, UK. [65]Department of Gastroenterology, Hospital Cl'nic/Institut d'Investigaci— Biomdica August Pi i Sunyer (IDIBAPS). [66]Centro de Investigaci—n Biomdica en Red de Enfermedades Hepaticas y Digestivas (CIBER EHD), Barcelona, Spain. [67]Christian-Albrechts-University, Institute of Clinical Molecular Biology, Kiel, Germany. [68]Department for General Internal Medicine, Christian-Albrechts-University, Kiel, Germany. [69]Inflammatory Bowel Diseases, Genetics and Computational Biology, Queensland Institute of Medical Research, Brisbane, Australia. [70]Norfolk and Norwich University Hospital. [71]Department of Gastroenterology, Leiden University Medical Center, Leiden, The Netherlands. [72]Child Life and Health, University of Edinburgh, Edinburgh, Scotland, UK. [73]Clinic for Neurology, TUM, Munich, Germany. [74]Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, Massachusetts, USA. [75]Department for General Internal Medicine, Christian-Albrechts-University, Kiel, Germany. [76]Department of Biostatistics, School of Public Health, Yale University, New Haven, Connecticut, USA. [78]Mount Sinai Hospital Inflammatory Bowel Disease Centre, University of Toronto, Toronto, Ontario, Canada. [79]Azienda Ospedaliero Universitaria (AOU) Careggi, Unit of Gastroenterology SOD2, Florence, Italy. [80]Center for Applied Genomics, The Children's Hospital of Philadelphia, Philadelphia, Pennsylvania, USA. [81]Department of Pediatrics, Center for Pediatric Inflammatory Bowel Disease, The Children's Hospital of

Philadelphia, Philadelphia, Pennsylvania, USA. [82]Inflammatory Bowel Disease Center, Department of Medicine, Johns Hopkins University School of Medicine, Baltimore, Maryland, USA. [83]Department of Gastroenterology, Royal Brisbane and Womens Hospital, and School of Medicine, University of Queensland, Brisbane, Australia. [84]Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, University of Cambridge, Cambridge, UK. [85]Division of Gastroenterology, University Hospital Gasthuisberg, Leuven, Belgium. [86] Université de Montréal, Medicine, Montréal, Québec, Canada, Hôpital Maisonneuve-Rosemont, Dept of Gastroenterology, Montréal, Québec, Canada, [87] IBD Center, Department of Gastroenterology, Istituto Clinico Humanitas, Milan, Italy, [88] Johns Hopkins University School of Medicine, Meyeroff Inflammatory Bowel Disease Center, Dept of Medicine, Baltimore, Maryland, USA, [89] Cardiovascular Health Research Unit, Department of Medicine, University of Washington, Seattle, Washington, USA, [90]Inserm, U1019, Lille, France, [91]Université de Lille Department of Hepato-Gastroenterology, Lille, France. [92] LVilnius University, Vilnius, Lithuania. [93]Department of Gastroenterology, UCL, Clinique Universitaire Saint-Luc, Brussels, Belgium. [94]Division of Gastroenterology, University Hospital Padua, Padua, Italy. [95]Mater Health Services, Department of Gastroenterology, Brisbane, Australia., [96] Dipartimento di Neuroscienze, Psicologia, Area del Farmaco e Salute del Bambino (NEUROFARBA), Università di Firenze SOD Gastroenterologia e Nutrizione Ospedale pediatrico Meyer, Firenze, Italy. [97] Department of Gastroenterology, UCL - Mont Godinne, Belgium. [98] Genome Institute of Singapore, Human Genetics, Singapore. [99] Institute for Human Genetics, University of California San Fransisco, San Francisco, California, USA. [100] Molecular Epidemiology, Queensland Institute of Medical Research, Brisbane, Australia. [101]Queensland Institute of Medical Research, Genetic Epidemiology, Brisbane, Australia

AUTHOR CONTRIBUTIONS

Jointly supervised research: J.D.R., M.J.D., K.V.S., V.K., T.H.K., A.F.
Conceived and designed the experiments: J.D.R., P.G., G.B., R.H.D., J.B., D.P.B.M., J.A.T., M.C., V.K., A.F.
Performed statistical analysis: P.G., G.B., D.M., L.J.
Analysed the data: P.G., G.B., L.J., E.G.
Contributed reagents / materials / analysis tools: V.A., S.L.H., J.R.O., I.T., S.L., P.S.
Wrote the paper: J.D.R., P.G., G.B., D.M., D.P.B.M., V.K., T.H.K., A.F.
Data QC and imputation: P.G., G.B., E.E., H.H., S.R.
All authors read and approved the final manuscript before submission

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

References:

1. Horton, R., et al., *Gene map of the extended human MHC.* Nat Rev Genet, 2004. **5**(12): p. 889-99.
2. Rioux, J.D., et al., *Mapping of multiple susceptibility variants within the MHC region for 7 immune-mediated diseases.* Proc Natl Acad Sci U S A, 2009. **106**(44): p. 18680-5.
3. Stokkers, P.C., et al., *HLA-DR and -DQ phenotypes in inflammatory bowel disease: a meta-analysis.* Gut, 1999. **45**(3): p. 395-401.
4. Jostins, L., et al., *Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.* Nature, 2012. **491**(7422): p. 119-24.
5. Achkar, J.P., et al., *Amino acid position 11 of HLA-DRbeta1 is a major determinant of chromosome 6p association with ulcerative colitis.* Genes Immun, 2012. **13**(3): p. 245-52.
6. Jones, D.C., et al., *Killer Ig-like receptor (KIR) genotype and HLA ligand combinations in ulcerative colitis susceptibility.* Genes Immun, 2006. **7**(7): p. 576-82.
7. Kulkarni, S., et al., *Genetic interplay between HLA-C and MIR148A in HIV control and Crohn disease.* Proc Natl Acad Sci U S A, 2013. **110**(51): p. 20705-10.
8. Satsangi, J., et al., *Contribution of genes of the major histocompatibility complex to susceptibility and disease phenotype in inflammatory bowel disease.* Lancet, 1996. **347**(9010): p. 1212-7.
9. Oksenberg, J.R., et al., *Mapping multiple sclerosis susceptibility to the HLA-DR locus in African Americans.* Am J Hum Genet, 2004. **74**(1): p. 160-7.
10. Newman, B., et al., *CARD15 and HLA DRB1 alleles influence susceptibility and disease localization in Crohn's disease.* Am J Gastroenterol, 2004. **99**(2): p. 306-15.
11. Lipsitch, M., C.T. Bergstrom, and R. Antia, *Effect of human leukocyte antigen heterozygosity on infectious disease outcome: the need for allele-specific measures.* BMC Med Genet, 2003. **4**: p. 2.
12. Alper, C.A., et al., *Extended major histocompatibility complex haplotypes in patients with gluten-sensitive enteropathy.* J Clin Invest, 1987. **79**(1): p. 251-6.
13. Aly, T.A., et al., *Multi-SNP analysis of MHC region: remarkable conservation of HLA-A1-B8-DR3 haplotype.* Diabetes, 2006. **55**(5): p. 1265-9.
14. Wiencke, K., et al., *Primary sclerosing cholangitis is associated to an extended B8-DR3 haplotype including particular MICA and MICB alleles.* Hepatology, 2001. **34**(4 Pt 1): p. 625-30.
15. Donaldson, P.T. and S. Norris, *Evaluation of the role of MHC class II alleles, haplotypes and selected amino acid sequences in primary sclerosing cholangitis.* Autoimmunity, 2002. **35**(8): p. 555-64.
16. Liu, J.Z., et al., *Dense genotyping of immune-related disease regions identifies nine new risk loci for primary sclerosing cholangitis.*Nat Genet, 2013. **45**(6): p. 670-5.