

University of Groningen

Human Leukocyte Antigen Signatures as Pathophysiological Discriminants of Microscopic Colitis Subtypes

MC-Europe GETECCU GWAS group; Zheng, Tenghao; Roda, Giulia; Zabana, Yamile; Escudero-Hernández, Celia; Liu, Xingrong; Chen, Ye; Camargo Tavares, Leticia; Bonfiglio, Ferdinando; Mellander, Marie-Rose

Published in:
Journal of Crohn's & colitis

DOI:
[10.1093/ecco-jcc/jjad165](https://doi.org/10.1093/ecco-jcc/jjad165)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2024

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

MC-Europe GETECCU GWAS group, Zheng, T., Roda, G., Zabana, Y., Escudero-Hernández, C., Liu, X., Chen, Y., Camargo Tavares, L., Bonfiglio, F., Mellander, M.-R., Janczewska, I., Vigren, L., Sjöberg, K., Ohlsson, B., Almer, S., Halfvarson, J., Miehlke, S., Madisch, A., Lieb, W., ... D'Amato, M. (2024). Human Leukocyte Antigen Signatures as Pathophysiological Discriminants of Microscopic Colitis Subtypes. *Journal of Crohn's & colitis*, 18(3), 349-359. <https://doi.org/10.1093/ecco-jcc/jjad165>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the "Taverne" license. More information can be found on the University of Groningen website: <https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment>.

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Human Leukocyte Antigen Signatures as Pathophysiological Discriminants of Microscopic Colitis Subtypes

Tenghao Zheng,^{a,b,c} Giulia Roda,^{d,*} Yamile Zabana,^{e,f,*} Celia Escudero-Hernández,^{g,h} Xingrong Liu,^{b,c} Ye Chen,ⁱ Leticia Camargo Tavares,^a Ferdinando Bonfiglio,^{a,b} Marie-Rose Mellander,^j Izabella Janczewska,^k Lina Vigren,ⁱ Klas Sjöberg,ⁱ Bodil Ohlsson,^l Sven Almer,^m Jonas Halfvarson,ⁿ Stephan Miehke,^o Ahmed Madisch,^p Wolfgang Lieb,^q Juozas Kupčinskas,^r Rinse K. Weersma,^s Luis Bujanda,^{f,t} Antonio Julià,^u Sara Marsal,^u Maria Esteve,^{e,f} Danila Guagnozzi,^{f,v} Fernando Fernández-Bañares,^{e,f} Carmen Ferrer,^w MC-Europe GETECCU GWAS group, Inga Peter,^x Jonas F. Ludvigsson,^y Darrell Pardi,^z Bas Verhaegh,^{aa} Daisy Jonkers,^{aa} Marieke Pierik,^{aa} Andreas Münch,^{g,bb,cc} Andre Franke,^h Francesca Bresso,^m Hamed Khalili,^{l,dd,ee} Jean-Frederic Colombel,^{ff} Mauro D'Amato^{b,gg,hh,ii}

^aSchool of Biological Sciences, Monash University, Clayton, VIC, Australia

^bClinical Epidemiology Division, Department of Medicine Solna, Karolinska Institutet, Stockholm, Sweden

^cCenter for Molecular Medicine, Karolinska Institutet, Stockholm, Sweden

^dBiostructures and Biosystems National Institute, Rome, Italy

^eGastroenterology Department, Hospital Universitari Mútua de Terrassa, Barcelona, Spain

^fCentro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Barcelona, Spain

^gDepartment of Biomedical and Clinical Sciences, Linköping University, Linköping, Sweden

^hInstitute of Clinical Molecular Biology, Kiel University and University Medical Center Schleswig-Holstein, Kiel, Germany

ⁱDivision of Gastroenterology, Massachusetts General Hospital, Boston, MA, USA

^jDepartment of Clinical Sciences, Lund University, Lund, Sweden

^kMedical Department, Ersta Hospital, Stockholm, Sweden

^lDepartment of Clinical Sciences, Lund University, Skane University Hospital, Malmö, Sweden

^mDivision of Gastroenterology, Department of Gastroenterology, Dermatology and Rheumatology, Karolinska University Hospital, Stockholm, Sweden

ⁿDepartment of Gastroenterology, Faculty of Medicine and Health, Örebro University Hospital, Örebro, Sweden

^oCentre for Digestive Diseases, Internal Medicine Centre Eppendorf, and Centre for Oesophageal Disorders, University Hospital Eppendorf, Hamburg, Germany

^pDepartment of Gastroenterology, CRH Clinic Siloah, Hannover, Germany

^qInstitute of Epidemiology, Christian-Albrechts-University of Kiel, Kiel, Germany

^rDepartment of Gastroenterology, Lithuanian University of Health Sciences, Kaunas, Lithuania

^sDepartment of Gastroenterology and Hepatology, University of Groningen and University Medical Center Groningen, Groningen, The Netherlands

^tDepartment of Liver and Gastrointestinal Diseases, Biodonostia Health Research Institute, Donostia University Hospital, Universidad del País Vasco, San Sebastian, Spain

^uRheumatology Research Group, Vall d'Hebron Research Institute, Barcelona, Spain

^vDepartment of Gastroenterology, Vall d'Hebron University Hospital, Neuro-Immuno-Gastroenterology Group, Digestive System Research Unit, Vall d'Hebron Institut de Recerca, Barcelona, Spain

^wPathology Department, Hospital Universitari Mútua de Terrassa, Barcelona, Spain

^xDepartment of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA

^yDepartment of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden

^zDivision of Gastroenterology and Hepatology, Mayo Clinic, Rochester, MN, USA

^{aa}Division Gastroenterology-Hepatology, Department of Internal Medicine, NUTRIM School of Nutrition and Translational Research in Metabolism, Maastricht University Medical Centre+, Maastricht, The Netherlands

^{bb}Department of Gastroenterology and Hepatology, Linköping University, Linköping, Sweden

^{cc}Department of Health, Medicine, and Caring Sciences, Linköping University, Linköping, Sweden

^{dd}Clinical and Translational Epidemiology Unit, Massachusetts General Hospital, Boston, MA, USA

^{ee}Broad Institute of Massachusetts Institute of Technology and Harvard, Cambridge, MA, USA

^{ff}Dr Henry D. Janowitz Division of Gastroenterology, Icahn School of Medicine at Mount Sinai, New York, NY, USA

^{gg}Gastrointestinal Genetics Lab, CIC bioGUNE - BRTA, Derio, Spain

^{hh}Ikerbasque, Basque Foundation for Science, Bilbao, Spain

ⁱⁱDepartment of Medicine and Surgery, LUM University, Casamassima, Italy

Corresponding author: Mauro D'Amato, Prof, Department of Medicine and Surgery, LUM University, SS 100 Km 18, 70010 Casamassima, Italy. Email: damato@lum.it

*Equal contributions.

Abstract

Background and Aims: Microscopic colitis [MC] is currently regarded as an inflammatory bowel disease that manifests as two subtypes: collagenous colitis [CC] and lymphocytic colitis [LC]. Whether these represent a clinical continuum or distinct entities is, however, an open question. Genetic investigations may contribute important insight into their respective pathophysiologies.

Methods: We conducted a genome-wide association study [GWAS] meta-analysis in 1498 CC, 373 LC patients, and 13 487 controls from Europe and the USA, combined with publicly available MC GWAS data from UK Biobank and FinnGen [2599 MC cases and 552 343 controls in total]. Human leukocyte antigen [HLA] alleles and polymorphic residues were imputed and tested for association, including conditional analyses for the identification of key causative variants and residues. Genetic correlations with other traits and diagnoses were also studied.

Results: We detected strong HLA association with CC, and conditional analyses highlighted the DRB1*03:01 allele and its residues Y26, N77, and R74 as key to this association (best $p = 1.4 \times 10^{-23}$, odds ratio [OR] = 1.96). Nominally significant genetic correlations were detected between CC and pneumonia [$r_g = 0.77$; $p = 0.048$] and oesophageal diseases [$r_g = 0.45$, $p = 0.023$]. An additional locus was identified in MC GWAS analyses near the *CLEC16A* and *RMI2* genes on chromosome 16 [rs35099084, $p = 2.0 \times 10^{-9}$, OR = 1.31]. No significant association was detected for LC.

Conclusion: Our results suggest CC and LC have distinct pathophysiological underpinnings, characterised by an HLA predisposing role only in CC. This challenges existing classifications, eventually calling for a re-evaluation of the utility of MC umbrella definitions.

Key Words: Microscopic colitis; collagenous colitis; GWAS; HLA; genetics

1. Introduction

Microscopic colitis [MC] is currently classified as an idiopathic form of inflammatory bowel disease [IBD], characterised by chronic non-bloody watery diarrhoea.¹ Because endoscopic exploration of MC usually lacks macroscopic features, MC diagnosis relies on histological examination from colon biopsies. By these means, MC is classically subtyped into collagenous colitis [CC] based on the presence of a thickened [$>10 \mu\text{m}$] subepithelial collagen band, and lymphocytic colitis [LC] where the collagen band is normal [$\sim 3 \mu\text{m}$] but intraepithelial lymphocytes infiltrate the mucosa.² Epidemiological studies from Europe and the USA report an increased prevalence, especially in older women, and a rising MC incidence in recent decades [4.9 per 100 000 person-years for CC and 5.0 for LC], with figures nearing those observed for the two major forms of IBD, Crohn's disease [CD] and ulcerative colitis [UC].^{2,3} The diagnosis of MC and its subtypes requires accurate histological examination, and no disease- or disease subtype-specific markers exist.

The aetiology of MC, CC, and LC and their underlying pathophysiological mechanisms are still poorly understood.^{1,4} Pathogenetic models include impaired epithelial barrier function, bacterial infections, genetics, bile acid malabsorption, and use of certain medications (non-steroidal anti-inflammatory drugs [NSAIDs], and proton pump inhibitors).¹ Given the strong association with autoimmune conditions like coeliac disease, thyroiditis, and type 1 diabetes mellitus, MC is often considered a disease of autoimmune origin,⁵ and recent large-scale epidemiological data on comorbidities support this notion particularly for CC.⁵ A burning question is whether CC and LC represent a clinical continuum caused by common pathogenetic mechanisms, or rather independent entities with overlapping clinical manifestations.

Familial clustering has been reported for MC more than 20 years ago, suggesting the existence of genetic predisposing factors.⁶ However, it was only in 2017 that the first convincing association with the human leukocyte antigens [HLA] region was described in a study of immune-related genes [using

ImmunoChip arrays] in 314 CC patients from Sweden and Germany.⁷ This association was further confirmed in 804 CC patients from the USA,⁸ and in a genome-wide association study [GWAS] of 423 MC diagnoses [lacking CC and LC classification] extracted from the electronic medical records of UK Biobank participants.⁹ As of today, no genetic investigation has been performed separately for CC and LC beyond HLA and immune-related loci; hence the genetic architecture of these conditions, including eventual differences and/or similarities, remains uncharacterised.

In this study we aimed to fill this knowledge gap, by performing GWAS meta-analyses of MC, CC, and LC, using the largest collection of multiple, population-based biobanks and case-control cohorts from Europe and the USA.

2. Materials and Methods

2.1. Study subjects

2.1.1. Case-control cohorts

Patients with MC ($N = 1876$: 1498 with CC; 373 with LC; 5 with incomplete forms of MC [MCi]) were recruited at tertiary gastroenterology clinics in Sweden [Stockholm, Malmö, and Linköping], Germany [Hamburg], The Netherlands [Maastricht], Lithuania [Kaunas], Spain [Barcelona, Girona, and Tomelloso], and the USA [Cleveland, New York, Rochester Minnesota, and Boston]. Most of the individuals have already been described elsewhere,^{7,8,10} and their demographics are reported in [Table 1](#). Ancestry-matched healthy controls [$N = 13\ 487$] from Sweden, Germany, The Netherlands, Lithuania, Spain, and the USA were from previously published studies or general population cohorts.^{11–13} Additional information on MC patients and controls is provided in the [Supplementary Methods](#), and their demographics are reported in [Table 1](#). Ethical approvals were obtained at respective centres from relevant local authorities. The overall study was approved by Stockholm Ethics Review Board [protocol no. 2016/271-31/1].

Table 1. Demographics of study subjects, and their inclusion in different analyses

Cohort	N	Sex, F%	Age, mean [±SD]	MC meta	CC meta	LC meta	HLA assoc
Case-control							
South Europe	1580	42.8	50.8 [±9.0]	●	●	●	●
MC	126	74.6	65.7 [±14.4]				
CC	74	82.4	66.4 [±14.3]				
LC	52	63.5	64.7 [±14.6]				
Controls	1454	40	49.6 [±7.0]				
North Europe	3651	54.1	52.7 [±14.9]	●	●	●	●
MC	783	81.4	64.6 [±12.1]				
CC	542	83.6	64.5 [±11.5]				
LC	241	76.8	64.8 [±13.6]				
Controls	2868	46.7	49.5 [±14.0]				
USA [Boston]	510	95.9	64.8 [±12.7]	●	●	●	●
MC	179	95.5	64.0 [±12.6]				
CC	94	94.7	65.1 [±11.2]				
LC	80	96.3	63.1 [±14.4]				
MCi	5	100	59.6 [±6.4]				
Controls	331	96.1	65.2 [±12.8]				
USA [Cleveland]	2482	60.6	67.6 [±9.2]	●	●		●
MC	215	80	61.6 [±13.6]				
CC	215	80	61.6 [±13.6]				
Controls	2267	58.7	68.2 [±8.5]				
USA [Others] ^a	7140	60.2	-	●	●		●
MC	573	83.6	-				
CC	573	83.6	-				
Controls	6567	58.2	68.2 [±8.6]				
Total case-controls	15363	58.2	57.6 [±14.2]				
MC	1876	82.8	64.1 [±12.7]				
CC	1498	83.7	64.0 [±12.3]				
LC	373	79.1	64.4 [±13.9]				
MCi	5	100	59.6 [±6.4]				
Controls	13487	54.8	57.6 [±14.2]				
Population-based							
UKBB	445655	54.3	58.6 [±9.9]	●			
MC	423	65.6	61.9 [±6.8]				
Controls	445232	54.2	58.6 [±9.9]				
FinnGen	93924	56.2	57.5	●			
MC	300	-	-				
Controls	93624	-	-				
Total population-based	539579	54.6	58.4				
MC	723	-	-				
Controls	538856	-	-				

SD: Standard Deviation; MC: microscopic colitis; CC: collagenous colitis; LC: lymphocytic colitis; MCi: incomplete forms of MC, UKBB: UK Biobank; HLA: Human leukocyte antigen; meta: GWAS meta-analysis; HLA assoc: association tests on HLA alleles and amino acids

^aUSA [Others] includes MC samples recruited from New York or Rochester Minnesota.

—: data not available.

2.1.2. Population cohorts

UK Biobank [UKBB] is a large, population-based cohort with available genotype data and health- and lifestyle-related information from approximately 500 000 individuals.¹⁴ Summary statistics of UKBB MC GWAS results from a previous study were obtained from the GWAS Catalog [<https://www.ebi.ac.uk/gwas/studies/GCST009081>] and included 423 MC cases (identified via hospital admission records using

the International Classification of Diseases [ICD10] code K52.8) and 445 232 controls [the remainder of the cohort].⁹ As previously reported,⁹ K52.8 is an imperfect proxy for MC diagnoses [subtype K52.83] in that it also includes *eosinophilic gastritis or gastroenteritis* [K52.81] and *eosinophilic colitis* [K52.82]. However, these are extremely rare conditions [prevalence is 5.1/100 000 and 2.1/100 000 for K52.81 and K52.82, respectively],¹⁵ and hence the eventual bias due

to such diagnoses potentially contaminating the MC group is negligible. The publicly available Data Freeze 2 from the FinnGen study [https://www.finnngen.fi/en/access_results] includes GWAS results in the form of summary statistics from testing several ICD10-defined conditions in 93 924 participants over 18 years from Finland.¹⁶ Among all FinnGen data releases [up to Data Freeze 8, released in Dec 1, 2022], Data Freeze 2 is the only release that includes both phenotypic and GWAS data for MC, which was defined based on code K52.88 of the Finnish version of ICD10. Despite its endpoint name [‘COLITCOLLAG’ in FinnGen] this corresponds to ‘Other non-infectious gastroenteritis and/or colitis’ [verified at source], which is the same diagnosis as K52.8 in UK Biobank. In total, 300 MC cases were included in this FinnGen GWAS release, compared with the remainder of the cohort [$N = 93\ 624$].

The distribution of cohorts and study individuals in relation to their inclusion in various analyses is reported in [Table 1](#).

2.2. GWAS and meta-analyses

According to the samples’ origin and their phenotype integrity, samples were subgrouped into five separate batches [[Table 1](#)]. DNA samples from MC cases were genotyped using Global Screening Arrays from Illumina, followed by rigorous quality control and imputation protocols [detailed pipeline in the [Supplementary Methods](#)]. GWAS were performed in individual MC, CC, and LC case-control batches, using a logistic regression model with Plink2 including age, sex, and top 10 principal components as covariates. GWAS summary statistics were then harmonised using the R package ‘EasyQC’ [v9.2], and finally combined into inverse-variance weighted fixed effect meta-analyses [for MC, CC, and LC] adopting the method implemented in METAL.¹⁷ Annotation of GWAS results including gene mapping was performed using the web-based platform FUnctional Mapping and Annotation of Genome-Wide Association Studies [FUMA, see [Supplementary Methods](#)].

2.3. Phenome-wide association studies [PheWAS]

For the C-Type Lectin Domain Containing 16A [*CLEC16A*]/RecQ Mediated Genome Instability 2 [*RIM2*] locus, we screened the lead SNP [single nucleotide polymorphism] and its LD [linkage disequilibrium] proxies [$r^2 > 0.8$] for previously reported GWAS associations extracted from the GWAS Catalog [<https://www.ebi.ac.uk/gwas/>],¹⁸ GWAS ATLAS [<https://atlas.ctglab.nl/>],¹⁹ and using Phenoscanner v2 [<http://www.phenoscanner.medschl.cam.ac.uk/>].²⁰ Only genome-wide significant associations [$p < 5 \times 10^{-8}$] were considered, and redundancy was avoided by only reporting associations from one database if multiple records were retrieved from different sources.

2.4. SNP heritability and genetic correlation

The SNP heritability [h^2_{SNP}] of MC, CC, and LC and genetic correlations were estimated using LDSC [linkage disequilibrium score regression] v1.0.1.²¹ European ancestry GWAS summary statistics for IBD, UC, and CD were obtained from the International IBD Genetics Consortium [IIBDGC] [<https://www.ibdgenetics.org/>] and correspond to the latest IBD GWAS meta-analysis,²² which includes data from the IIBDGC²³ and additional British IBD cases.²⁴ Correlations between CC and other traits were obtained using the online

platform Complex-Traits Genetics Virtual Lab [CTG-VL],²⁵ which integrates summary statistics of 1376 traits from multiple sources including UKBB GWAS round 2 [<http://www.nealelab.is/uk-biobank/>], the Psychiatric Genomics Consortium [PGC], and the Genetic Investigation of ANthropometric Traits [GIANT] consortium.

2.5. HLA imputation and association test

Array-genotyped SNPs within the HLA region [chromosome 6 at 29-34 Mb, hg19 assembly] were extracted from each individual for HLA imputation. We imputed classical alleles [HLA-A, -C, -B, -DRB1, -DQA1, -DQB1, -DPA1, and -DPB1, at the four-digits level] and corresponding amino acid residues using HIBAG [HLA Genotype Imputation with Attribute Bagging] with its pre-fit, European-specific model generated from the HLARES reference panel.²⁶ Only imputed alleles and amino acids with frequencies > 0.01 were included in the downstream analyses. Associations with CC and LC were tested using the HLA Analysis Tool Kit [HATK].²⁷ In brief, we first harmonised HLA nomenclature based on the IMGT/HLA database,²⁸ then generated binary variables representing the presence or absence of an HLA allele or amino acid residue. Finally, the association of alleles and amino acid residues with CC and LC was tested in individual cohorts via logistic regression under a dominant genetic model, adjusting for sex, age [wherever available], and the top five principal components [PCs]. HLA associations were tested under a dominant genetic model based on their known biological function [allele-specific antigen presentation properties].²⁹ Association results from each cohort were then meta-analysed using the inverse-variance weighted model using the R package ‘meta’. Bonferroni correction was applied to take into account multiple testing. For selected HLA alleles and amino acids, conditional analyses were also performed [adopting the same logistic regression model and covariates] to identify independent risk effects from the HLA locus. Three-dimensional structural models of the HLA-DR molecule was downloaded from the Protein Data Bank [PDB IDs ‘3PDO’]³⁰ and visualised using PyMOL v.2.5.2 [Schrödinger].

3. Results

3.1. GWAS meta-analysis of microscopic colitis

We conducted five independent GWAS of MC in respective cohorts from tertiary centres in Europe [909 cases and 4322 controls] and the USA [967 cases and 9165 controls] [[Table 1](#)], using a common pipeline for QC, imputation, and association testing [[Supplementary Methods](#)]. The results were included in a GWAS meta-analysis together with publicly available summary statistics from similar analyses in UK Biobank [UKBB]⁹ and FinnGen [using ICD10 codes from electronic medical records as proxies for MC diagnoses]. In total, data for 8 662 703 high-quality SNP markers and 554 942 individuals [2599 MC cases and 552 343 controls] were included in the MC GWAS meta-analysis. This showed no population stratification [$\lambda = 1.04$, LD score regression, LDSC intercept = 1.03, [Supplementary Figure 1](#)], and identified 3714 genome-wide-significant [$p \leq 5.0 \times 10^{-8}$] associations from two independent loci [[Figure 1](#) and [Table 2](#)]: one on chromosome 6 within the HLA region (lead SNP rs9267445 with $p = 4.3 \times 10^{-39}$ and odds ratio [OR] = 1.83), near the *HLA-B* and MHC Class I Polypeptide-Related Sequence B

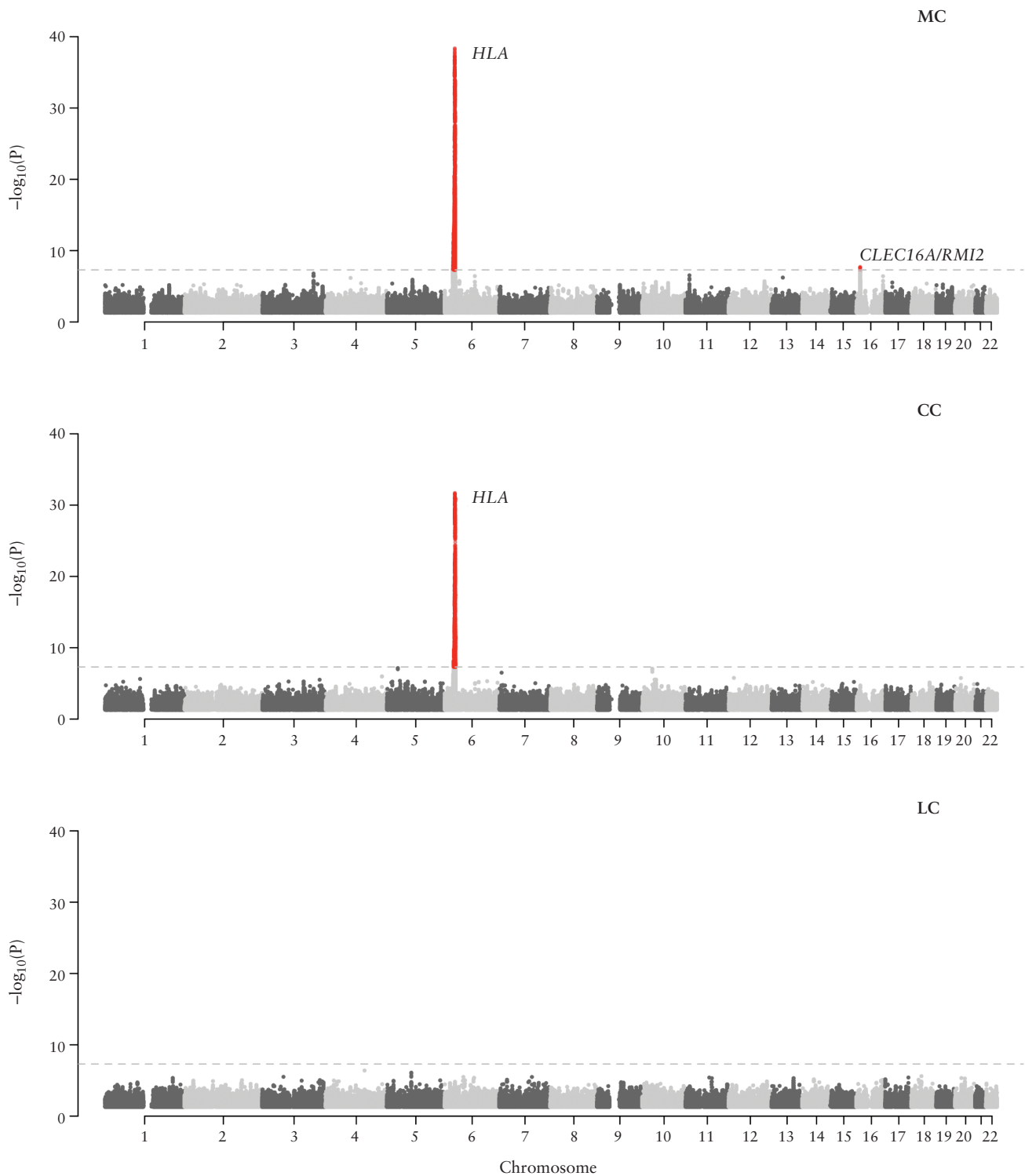


Figure 1. Manhattan plots of MC, CC, and LC GWAS meta-analysis results. Genome-wide association signals [$-\log_{10}(P)$ on the y-axis] for MC, CC, and LC are reported for all markers included in the analyses. Genome-wide significant threshold [$p \leq 5.0 \times 10^{-8}$] is indicated by grey horizontal lines. Markers from the genome-wide significant loci are highlighted in red and are annotated with the nearest gene. Abbreviations: CC = collagenous colitis; HLA = human leukocyte antigen; LC = lymphocytic colitis; MC = microscopic colitis.

[*MICB*] genes, and the other on chromosome 16 [lead SNP rs35099084 with $p = 2.0 \times 10^{-8}$ and OR = 1.31] near the gene *CLEC16A*. Based on LDSC analyses, SNP-based heritability of MC was estimated around 10% [$h^2_{\text{SNP}} = 0.10$; $p = 2.0 \times 10^{-3}$] [Table 3]. Sex-stratified analyses did not reveal any additional association signal [not shown].

3.2. rs35099084 locus-specific analyses

The *CLEC16A* locus was investigated further at different levels. Similar association patterns were observed when this locus was inspected in individual GWAS, with consistent MC genetic risk effects across cohorts [Supplementary Figure 2]. Of interest, based on PheWAS analyses of publicly available

Table 2. MC and CC GWAS meta-analysis risk loci

Lead SNP[s] rsID	CHR	EA	OA	EAF	OR [95% CI]	P	Het.P	Gene content*
MC meta-analysis								
rs9267445*	6	C	G	0.120	1.8 [1.7-2.0]	4.3×10^{-39}	0.55	HLA [253]**
rs35099084	16	C	T	0.808	1.3 [1.2-1.4]	2.0×10^{-8}	0.66	CLEC16A ^{pe} , RMI2 ^{pe} , SOCS1 ^{pe} , TNF2 ^p , PRM3 ^p , PRM2 ^p , PRM1 ^p , RSL1D1 ^c
CC meta-analysis								
rs2844531*	6	G	A	0.119	2.0 [1.8-2.2]	2.0×10^{-32}	0.95	HLA [250]**

MC: microscopic colitis; CC: collagenous colitis; SNP: Single-nucleotide polymorphism; CHR: chromosome; EA: effect allele; OA: other allele; EAF: effect allele frequency; OR: odds ratio; CI: confidence interval; Het.P: p-value from Cochran's Q test for heterogeneity in meta-analyses; HLA: Human leukocyte antigen; CLEC16A: C-Type Lectin Domain Containing 16A, RMI2: RecQ Mediated Genome Instability 2, SOCS1: Suppressor Of Cytokine Signaling 1, TNF2: Transition Protein 2, PRM1: Protamine 1, PRM2: Protamine 2, PRM3: Protamine 3, RSL1D1: Ribosomal L1 Domain Containing 1.

*GWAS signals from markers rs9267445 [MC] and rs2844531 [CC] come from the same HLA locus.

**Mapped genes at the identified locus [total number in brackets for HLA, which are too many to display], based on FUMA positional ['p'] and eQTL ['e'] mapping.

Table 3. Genetic correlations between MC, CC and IBD

Traits ^a	IBD [SNP _{h2} : 0.29; P = 6.7×10^{-29}]			UC [SNP _{h2} : 0.24, P = 7.7×10^{-26}]			CD [SNP _{h2} : 0.42, P = 2.6×10^{-21}]		
	r _g	SE	p	r _g	SE	p	r _g	SE	p
MC [SNP _{h2} : 0.10; p = 0.002]	0.32	0.11	0.0035	0.30	0.12	0.012	0.25	0.11	0.024
CC [SNP _{h2} : 0.10, p = 0.005]	0.30	0.11	0.0089	0.29	0.12	0.019	0.21	0.11	0.054

^aSNP heritability [SNP_{h2}] and its significance [as assessed via LDSC analyses] are reported for all traits.

Abbreviations: CC = collagenous colitis; CD = Crohn's disease; IBD = inflammatory bowel disease; MC = microscopic colitis; SE = standard error; SNP = single-nucleotide polymorphism; UC = ulcerative colitis.

data from previous GWA studies, this locus appears to be involved in a number of other immune-related diseases from the immunological and respiratory domains, including neutrophil and eosinophil counts, asthma, eczema, and others [Supplementary Table 1]. This is in line with gene content at this locus, as most genes mapped to the region [using FUMA, Supplementary Methods] are known to be involved in immune-related activities [Table 2, Supplementary Table 2]. Fine-mapping analysis [Supplementary Methods] identified two independent credible sets of causative variants [posterior probability = 0.92, Supplementary Figure 3], likely deriving from two independent association signals from the same locus. These credible sets contain a total of 45 variants, though none with a posterior probability greater than 0.15, possibly due to high LD in the region [Supplementary Figure 3]; hence the identity of the causal variant from each signal remains elusive. However, at least when looking at the lead SNPs from each credible set [rs35099084 and rs171471], it is of note that both associate with mRNA expression quantitative trait loci [eQTL] from multiple tissues and the gastrointestinal tract, for the genes *CLEC16A* and *RMI2* [Supplementary Figure 4].

3.3. Subtype-specific GWAS meta-analyses of CC and LC

We carried out separate CC and LC GWAS meta-analyses, using the same analytical pipeline on individual datasets with available subtype information [Table 1]. This included 7 486 967 SNP markers tested in 1498 cases and 13 487 controls for CC, and 7 460 065 SNP markers tested in 373 cases and 4653 controls for LC. Individual GWAS showed no genomic inflation or population stratification [$\lambda = 1.03$

and LDSC intercept = 1.01 for CC; $\lambda = 0.98$ and LDSC intercept = 0.96 for LC].

CC GWAS meta-analysis highlighted strong association with the HLA region, where 2851 markers gave rise to genome-wide association signals. These and the corresponding genetic risk effects were even stronger than those observed for MC, despite the reduction in sample size [lead SNP rs2844531, $p = 2.0 \times 10^{-32}$, OR = 2.0; Table 2]. A sensitivity analysis performed only in CC cases with documented exclusion of a coeliac disease diagnosis [negative anti-tissue transglutaminase serology and/or duodenal endoscopy] also returned identical results with a GWAS signal detected in correspondence of the HLA region [Supplementary Figure 5]. On the contrary, no GWAS-significant signal was detected for LC [Figure 1]. This altogether suggests that the HLA association observed in MC may be entirely driven by CC. The *CLEC16A* locus only showed nominal significance for the lead SNP rs35099084, both in CC [$p = 1.9 \times 10^{-4}$, OR = 0.75] and LC [$p = 0.026$, OR = 0.77], with consistent genetic risk effects across individual cohorts. Similar results [and no additional association signals] were obtained when stratifying CC and LC cohorts into male and female groups [not shown].

3.4. HLA association as a distinctive feature of CC

We investigated further the HLA signal, which appears to be a distinctive feature of CC. We first considered the possibility that association with HLA was not detected in LC because of the much smaller sample size compared with CC [Table 1]. However, although at least 600 cases would be needed to detect a genome-wide significant signal, the LC cohort

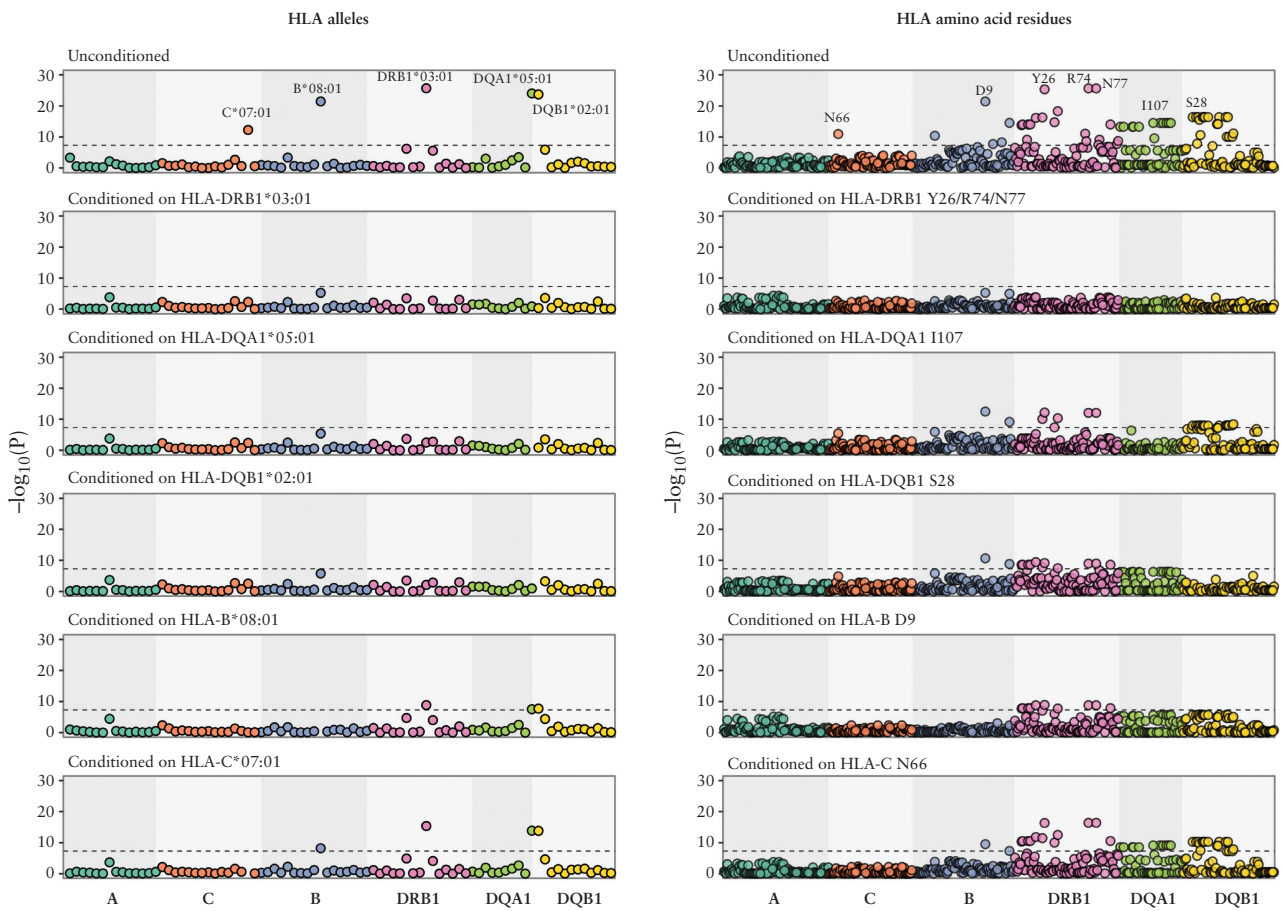


Figure 2. HLA associations in CC, alleles, and amino acids. The plots show HLA association signals observed in CC [$-\log_{10}(P)$ on the y-axis], relative to tested HLA classical alleles [left] and amino acid residues [right] [reported in full in Supplementary Table S4]. Association results after conditioning on selected HLA alleles and amino acids are also reported as indicated [Y26/R74/N77 corresponds to results for R74 for illustrative purposes, and identical or nearly identical results were obtained respectively for Y26 and N77, both also showing no residual association signal]. The grey horizontal line indicates the genome-wide association threshold 5×10^{-8} . Abbreviations: CC = collagenous colitis; HLA = human leukocyte antigen.

[373 cases and 4653 controls] still had >80% statistical power to detect association $<2 \times 10^{-6}$ for an SNP with the same frequency and risk effect as the HLA lead marker in CC [rs2844531, MAF = 0.12, OR = 2.0]. Moreover, a simulation including 10 000 rounds of random CC case sampling and association testing to reproduce LC sample size-based analyses [Supplementary Methods] returned results more significant than those observed for LC 100% of the time [p -value range 2.5×10^{-4} - 4.2×10^{-17} vs 1.8×10^{-2} as detected for LC]. This confirms the specificity of the HLA signal in relation to CC risk, and suggests this region is not relevant to LC.

3.5. HLA variants and residues most relevant to CC risk

In order to refine the CC risk signal coming from the HLA region, and to identify associations with individual HLA alleles and specific amino acid residues, we carried out HLA imputation using HIBAG. We detected a total of 94 HLA alleles and 534 amino acid residues present in at least 1% of individuals for the loci *HLA-DQA1*, *HLA-DQB1*, *HLA-DRB1*, *HLA-DPB1*, *HLA-A*, *HLA-B*, and *HLA-C*. HLA testing in CC returned strong association signals for multiple alleles from the ancestral haplotype 8.1, including DRB1*03:01 [OR = 1.96, Bonferroni-corrected p -value [P_c] = 1.3×10^{-23}], B*08:01 [OR = 1.84, $P_c = 2.3 \times 10^{-19}$], DQA1*05:01 [OR = 1.92,

$P_c = 6.0 \times 10^{-22}$], DQB1*02:01 [OR = 1.90, $P_c = 1.3 \times 10^{-21}$], C*07:01 [OR = 1.58, $P_c = 3.5 \times 10^{-10}$] [Supplementary Table 3]. DRB1*03:01 appeared to be most likely responsible for the association signal, since it showed strongest association from the 8.1 haplotype, and no residual association was observed when conditioning on this allele in further analyses [Figure 2 and Supplementary Table 3].

When HLA associations were tested at the amino acid level, strongest associations were detected for tyrosine [Y] at amino acid position 26, asparagine [N] at position 77, and arginine [R] at position 74, all from the HLA-DRB1 molecule [best $P_c = 1.4 \times 10^{-23}$ for R74 and N77, OR = 2; Supplementary Table 3]. Of note, these are all exclusively found in the HLA-DRB1*03:01 allele (other alleles, DRB1*03:02 and HLA-DRB1*09:01, also carry these residues but are extremely rare [respective allele frequencies 0.00015 and 0.0033 in our study] and were not tested here), and are located in the peptide-binding groove where interaction with the antigen to be presented to T cells occurs [Figure 3]. Conditioning on any of these alleles in additional analyses abrogated all association signals at the amino acid level [whereas association was still detected for DRB1 residues when conditioning on the most significant residues from other 8.1 alleles] [Figure 2]. This again suggests that HLA-DRB1*03:01 is the major driver of the association with CC.

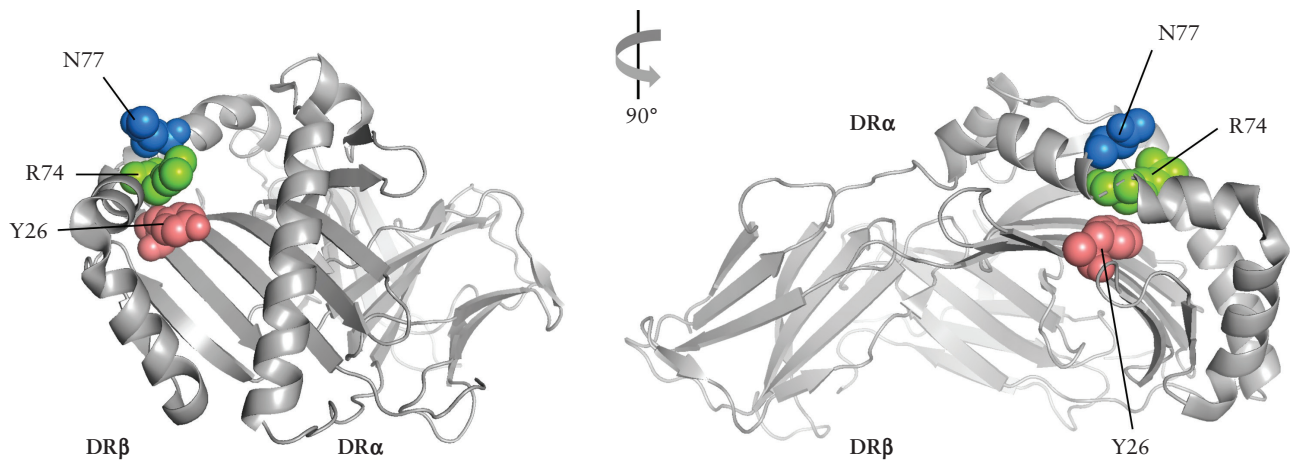


Figure 3. Functional relevance of key HLA amino acids associated with CC risk. A three-dimensional ribbon model of the peptide-binding portion of the HLA DR molecule is represented as top view [left] and side view [right]. The structure obtained from Protein Data Bank [PDB ID code 3PDO] displays the key CC-predisposing residues tyrosine at position 26 [Y26], asparagine at position 77 [N77], and arginine at position 74 [R74], all contributing to the specificity of the peptide-binding groove of the HLA-DRB1*03:01 molecule. Abbreviations: CC = collagenous colitis; HLA = human leukocyte antigen.

3.6. Genetic correlation between CC and other traits

Finally, we sought to further characterise the genetic architecture of CC by seeking similarities with other traits and conditions. LDSC analysis indicated that CC is a [partially] heritable trait, with an estimated SNP-based heritability of 9.6% [$p = 4.9 \times 10^{-3}$], whereas no evidence of heritability was detected for LC [not shown]. Using European-ancestry summary statistics from the most recent IBD GWAS meta-analysis²² [see Methods], we compared CC with IBD, UC, and CD [also via LDSC analyses] and detected significant overlap with IBD [$r_g = 0.30$, $p = 0.0089$] and UC [$r_g = 0.29$, $p = 0.019$] [Table 3].

Broader analyses of genetic overlap with other traits and conditions were conducted using the CTG-VL platform [Supplementary Methods]. We found nominally significant genetic commonalities between CC and 48 other diseases or traits, with strongest correlation observed for ‘ICD10: J18 Pneumonia, organism unspecified’ [$r_g = 0.77$, $p = 0.048$] [Supplementary Table 4]. However, none of these correlations remained significant after correction for multiple testing.

4. Discussion

We report here the first investigation of MC subtypes CC and LC across the entire genome, based on the analysis of multiple cohorts from Europe and the USA. We confirm HLA as the strongest genetic driver of CC risk, whereas no other unequivocal associations were observed in the rest of the genome. We also demonstrate that HLA association is a genetic determinant of CC but not LC, thus suggesting these subtypes may be associated with different pathophysiological mechanisms. This expands our knowledge of these conditions, challenging current views and classifications and eventually calling for a re-evaluation of the appropriateness of using MC as an umbrella term.

GWAS meta-analyses of MC subtypes CC and LC provided most insightful results as to the nature of the potential mechanisms beyond predisposition to these conditions. SNP heritability could be detected for CC [but not LC], indicating that

the former is a [partially] inherited condition. In LDSC correlation analyses compared with IBDs, the genetic architecture of CC appeared to be more similar to that of UC than CD, confirming and expanding previous observations obtained from ImmunoChip data.¹⁰ Broader analyses of genetic similarities between CC and other diseases highlighted nominally significant correlations with pneumonia and diseases of the oesophagus, including conditions of known or suspected infectious origin such as achalasia and ulcer of the oesophagus. For the latter, a role for *Helicobacter pylori* [*H. pylori*] infection has been proposed, although often complicated by the co-occurrence of gastrointestinal reflux disease [GERD] with oesophageal ulcers.^{31,32} At the same time, although these genetic correlations are detected independent of the HLA signal [see Supplementary Methods for LDSC details],²¹ HLA-only associations like the one observed here for CC appear to be typically seen in infectious diseases.³³ These findings may be worth considering in view of recent large-scale epidemiological surveys reporting higher risk of CC following gastroenteritis due to *Campylobacter concisus*, *Clostridium difficile*, *Norovirus*, and other infections,^{34,35} as well as increased risk of severe COVID-19 infection in CC.³⁶ Indeed, a decrease in microbiome diversity and dysbiosis has also been detected in CC patients.³⁷ Altogether, this warrants additional research to better understand the potential contribution of infections to the pathogenesis of CC.

HLA associations and their in-depth analysis provided valuable insight into the pathophysiology of CC and LC. These associations were exclusively detected in the former: although LC sample size was considerably smaller than in CC [respectively 373 vs 1498 cases], reduced statistical power was not the reason for the lack of HLA associations in this group, as our simulation analyses clearly demonstrated [LC sample size was adequately powered to detect associations of the magnitude observed in CC]. Thus, our results suggest that HLA association is a distinctive feature of CC and not LC, which bears implications on the current view of these conditions as similar clinical entities jointly classified under the unifying MC umbrella term. This has been a matter of debate for a long time,³⁸ and the results presented here challenge this view and warrant additional thoughts on the utility and eventual

appropriateness of using ‘microscopic colitis’ to interchangeably refer to CC or LC. At the same time, HLA genotype may help to better define disease forms that do not entirely fulfil histological criteria for a differential diagnosis into CC or LC, though still exhibiting characteristic features typical of MC [including response to MC treatments] and are therefore categorised as incomplete MC [MCi].^{1,2,39} Additional evidence has been recently accumulating also at the molecular level, that CC and LC are more diverse than initially thought, especially when looking at immune-related activities and cell types, as well as RNA-seq gene expression data.^{40–42} In this context, the direction of HLA associations is also intriguing, as strong and exclusive HLA associations have been reported in GWAS of other conditions characterised by fibrotic processes, involving the same alleles [HLA-DRB1*03:01] and amino acid residues identified here as strongest risk factors for CC.⁴³ This includes conditions like fibrotic idiopathic interstitial pneumonia, frontal fibrosing alopecia, and idiopathic retroperitoneal fibrosis that, like CC, are characterised by collagen deposition.^{43–45} The knowledge that HLA-DRB1*03:01, or one or more of its residues 26, 74, and 77, are likely to mediate the CC-predisposing role, might help in the identification of eventual self or microbial pathogenic epitopes, as the antigen-binding properties of many HLA molecules have been thoroughly characterised. HLA-DR molecular modelling shows that modifying neutral amino acids to a positively charged hydrophilic arginine at position 74 can significantly affect the properties of pocket 4 in the peptide-binding groove,⁴⁶ one of the regions key to determining overall peptide binding affinity and avidity. Tyrosine 26 and asparagine 77, strongly linked to arginine 74, also contribute to the peptide-binding specificity of the HLA-DR molecule.^{43,47} These amino acid positions have already been highlighted in previous studies of rheumatoid arthritis [RA] and primary biliary cholangitis [another disease characterised by fibrosis],^{48,49} in fact earlier studies reported that residue 74 [and 71] in the DR4 rheumatoid arthritis-associated molecule is key to the preferential binding of a fragment derived from human collagen II [residues 256–271].⁵⁰ Hence, these amino acids may determine specific HLA-DR pocket signatures that accommodate CC-triggering peptides of as yet unknown characteristics.

Finally, in our initial GWAS meta-analysis of MC, we also detected a genome-wide, significant association for a risk locus tagged by SNP rs35099084 on chromosome 16p13.13. In particular, this locus appears to harbour two independent signals linked to quantitative changes [eQTL] at the level of mRNA expression for the genes *CLEC16A* and *RMI2*. The product of the *CLEC16A* gene is a membrane-associated endosomal protein ubiquitously expressed and involved in the control of mitophagy,⁵¹ also shown to contribute to the process of peptide loading onto HLA class II molecules in myeloid and B cells.⁵² Of note, genetic variation in this gene is associated with other immune-mediated traits and conditions including asthma and eosinophil counts.^{53,54} *RMI2* codes for a poorly characterised protein that is a component of the BTR complex [comprising Bloom helicase, Topoisomerase 3, and RMI1/2 scaffold proteins], which plays a role in DNA repair and genome stability,⁵⁵ and associations with eosinophil counts and immune diseases, including IBD and primary biliary cirrhosis, have been reported also for this gene [Supplementary Table 1]. The results from the *CLEC16A* locus therefore point to a potential involvement of eosinophils and immune-related activities as mechanisms underlying

the observed associations with MC, as well as CC and LC given the similar genetic risk effects observed, thus eventually reflecting at least some degree of non-HLA, immune-mediated overlap between these conditions. However, this association signal requires confirmation via replication in additional studies.

In conclusion, we report here the first genome-wide analysis of CC and LC predisposition, which strongly suggests that they have distinct pathophysiological underpinnings, characterised by an important predisposing role of HLA-related mechanisms only in CC. This challenges existing paradigms and clinical definitions that see them grouped together under an overarching diagnosis of MC, calling for an eventual re-evaluation of the utility of such classifications. The identification of specific HLA risk alleles and their key residues warrants further studies aimed at the characterisation of putative colitogenic peptides.

Funding

This study was partially supported by grants from the Swedish Research Council to MD'A [VR 2017-02403].

Conflict of Interest

YZ has received support for conference attendance, speaker fees, research support, and consulting fees from AbbVie, Adaclyte, Amgen, Dr Falk Pharma, FAES Pharma, Ferring, Janssen, MSD, Otsuka, Pfizer, Shire, Takeda, Galapagos, Boehringer Ingelheim, and Tillots. LV has received speaker fees from Ferring and Tillots. JFL has co-ordinated an unrelated study on IBD with funding from Janssen. DJ has received funding from EU projects FP7/Nr 305564 and H2020/Nr 848228, and public private partnership grants from Top Knowledge Institute and as part of the NWO-CCC partnership programme outside the submitted work. HK has received funding from Takeda and Pfizer for unrelated projects; HK has also received consulting fees from Takeda. JFC reports receiving research grants from AbbVie, Janssen Pharmaceuticals, and Takeda; receiving payment for lectures from AbbVie, Amgen, Allergan, Ferring Pharmaceuticals, Shire, and Takeda; receiving consulting fees from AbbVie, Amgen, Arena Pharmaceuticals, Boehringer Ingelheim, BMS, Celgene Corporation, Eli Lilly, Ferring Pharmaceuticals, Galmed Research, Genentech, Glaxo Smith Kline, Janssen Pharmaceuticals, Kaleido Biosciences, Imedex, Immunic, Iterative Scopes, Merck, Microbia, Novartis, PBM Capital, Pfizer, Protagonist Therapeutics, Sanofi, Takeda, TiGenix, Vifor; and holds stock options in Intestinal Biotech Development. MD'A has received unrestricted research grants from QOL Medical for unrelated projects.

MC-Europe GETECCU GWAS group members:

Yamile Zabana,^{1,2} Fernando Fernández-Bañares,^{1,2} Maria Esteve,^{1,2} Marta Piqueras,³ David Busquets,⁴ Eva Iglesias,⁵ Alfredo J Lucendo.⁶

¹Gastroenterology Department, Hospital Universitari Mútua de Terrassa, Barcelona, Spain

²Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Barcelona, Spain

³Consorci Sanitari de Terrassa, Terrassa, Spain

⁴Hospital Universitari de Girona Doctor Josep Trueta, Girona, Spain

⁵Hospital Universitario Reina Sofía de Córdoba e Instituto Maimónides de Investigación Biomédica de Córdoba, Córdoba, Spain

⁶Hospital General de Tomelloso and Instituto de Investigación Sanitaria de Castilla-La Mancha, Instituto de Investigación Sanitaria Princesa, Centro de Investigación Biomédica en Red de Enfermedades Hepáticas y Digestivas, Tomelloso, Spain

Data Availability

Genome-wide summary statistics have been submitted to the European Bioinformatics Institute [www.ebi.ac.uk/gwas] under the accession numbers GCST90204144, GCST90204145 and GCST90204146.

Author Contributions

Concept: MD'A [lead], JFC [supporting]. Investigation and data curation: GR [lead], YZ [lead], MRM [supporting], IJ [supporting], LV [supporting], KS [supporting], BO [supporting], SA [supporting], JH [supporting], S Mielhke [supporting], A Madisch [supporting], WL [supporting], JK [supporting], RKW [supporting], LB [supporting], AJ [supporting], S Marsal [supporting], ME [supporting], DG [supporting], FFB [supporting], CF [supporting], IP [supporting], JFL [supporting], DP [supporting], BV [supporting], DJ [supporting], MP [supporting], A Münch [supporting], AF [supporting], F Bresso [supporting], HK [supporting], JFC [supporting], and MD'A [supporting]. Formal analysis: TZ [lead], XL [supporting], YC [supporting], LCT [supporting] and F Bonfiglio [supporting]. Visualisation: TZ [lead] and F Bonfiglio [supporting]. Funding acquisition, project administration, and supervision: MD'A [lead]. Writing original draft: TZ [equal], MD'A [equal], CEH [supporting]. Writing review and editing: all authors [equally].

Supplementary Data

Supplementary data are available at ECCO-JCC online.

References

- Burke KE, D'Amato M, Ng SC, Pardi DS, Ludvigsson JF, Khalili H. Microscopic colitis. *Nat Rev Dis Primers* 2021;7:39.
- Mielhke S, Guagnozzi D, Zabana Y, *et al.* European guidelines on microscopic colitis: United European gastroenterology and European microscopic colitis group statements and recommendations. *United European Gastroenterol J* 2021;9:13–37.
- Ng SC, Shi HY, Hamidi N, *et al.* Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet* 2017;390:2769–78.
- Zabana Y, Tontini G, Hultgren-Hornquist E, *et al.* Pathogenesis of microscopic colitis: a systematic review. *J Crohns Colitis* 2022;16:143–61.
- Wildt S, Munck LK, Winther-Jensen M, Jess T, Nyboe Andersen N. Autoimmune diseases in microscopic colitis: A Danish nationwide case-control study. *Aliment Pharmacol Ther* 2021;54:1454–62.
- Jarnerot G, Hertervig E, Granno C, *et al.* Familial occurrence of microscopic colitis: a report on five families. *Scand J Gastroenterol* 2001;36:959–62.
- Westerlind H, Mellander MR, Bresso F, *et al.* Dense genotyping of immune-related loci identifies HLA variants associated with increased risk of collagenous colitis. *Gut* 2017;66:421–8.

- Stahl E, Roda G, Dobbyn A, *et al.* Collagenous colitis is associated with HLA signature and shares genetic risks with other immune-mediated diseases. *Gastroenterology* 2020;159:549–61.e8.
- Green HD, Beaumont RN, Thomas A, *et al.* Genome-wide association study of microscopic colitis in the UK Biobank confirms immune-related pathogenesis. *J Crohns Colitis* 2019;13:1578–82.
- Westerlind H, Bonfiglio F, Mellander MR, *et al.* HLA associations distinguish collagenous from lymphocytic colitis. *Am J Gastroenterol* 2016;111:1211–3.
- Tigchelaar EF, Zhernakova A, Dekens JA, *et al.* Cohort profile: Lifelines deep, a prospective, general population cohort study in the northern Netherlands: Study design and baseline characteristics. *BMJ Open* 2015;5:e006772.
- Krawczak M, Nikolaus S, von Eberstein H, Croucher PJP, El Mokhtari NE, Schreiber S. Popgen: Population-based recruitment of patients and controls for the analysis of complex genotype-phenotype relationships. *Community Genet* 2006;9:55–61.
- Julia A, Domenech E, Ricart E, *et al.* A genome-wide association study on a southern European population identifies a new Crohn's disease susceptibility locus at rbx1-ep300. *Gut* 2013;62:1440–5.
- Sudlow C, Gallacher J, Allen N, *et al.* UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med* 2015;12:e1001779.
- Mansoor E, Saleh MA, Cooper GS. Prevalence of eosinophilic gastroenteritis and colitis in a population-based study, from 2012 to 2017. *Clin Gastroenterol Hepatol* 2017;15:1733–41.
- Mars N, Widen E, Kerminen S, *et al.*; FinnGen. The role of polygenic risk and susceptibility genes in breast cancer over the course of life. *Nat Commun* 2020;11:6383.
- Willer CJ, Li Y, Abecasis GR. Metal: Fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010;26:2190–1.
- Buniello A, MacArthur JAL, Cerezo M, *et al.* The NHGRI-EBI GWAS catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. *Nucleic Acids Res* 2019;47:D1005–12.
- Watanabe K, Stringer S, Frei O, *et al.* A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet* 2019;51:1339–48.
- Kamat MA, Blackshaw JA, Young R, *et al.* Phenoscanner v2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics* 2019;35:4851–3.
- Bulik-Sullivan BK, Loh PR, Finucane HK, *et al.*; Schizophrenia Working Group of the Psychiatric Genomics Consortium. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet* 2015;47:291–5.
- Liu Z, Liu R, Gao H, *et al.*; FinnGen. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat Genet* 2023;55:796–806.
- Liu JZ, van Sommeren S, Huang H, *et al.*; International Multiple Sclerosis Genetics Consortium. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet* 2015;47:979–86.
- de Lange KM, Moutsianas L, Lee JC, *et al.* Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* 2017;49:256–61.
- Cuellar-Partida G, Lundberg M, Kho PF, D'Urso S, Gutierrez-Mondragon LF, Hwang L-D. Complex-traits genetics virtual lab: a community-driven web platform for post-GWAS analyses. *bioRxiv* 2019;518027. <https://doi.org/10.1101/518027>.
- Zheng X, Shen J, Cox C, *et al.* HIBAG-HLA genotype imputation with attribute bagging. *Pharmacogenomics J* 2014;14:192–200.
- Choi W, Luo Y, Raychaudhuri S, Han BH. HLA analysis toolkit. *Bioinformatics* 2021;37:416–8.
- Robinson J, Soormally AR, Hayhurst JD, Marsh SGE. The IPD-IMGT/HLA database: new developments in reporting HLA variation. *Hum Immunol* 2016;77:233–7.

29. Hughes AL, Nei M. Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* 1988;335:167–70.
30. Berman HM, Westbrook J, Feng Z, *et al.* The protein data bank. *Nucleic Acids Res* 2000;28:235–42.
31. Blaser MJ. Helicobacter pylori and esophageal disease: wake-up call? *Gastroenterology* 2010;139:1819–22.
32. Schulz C, Kupcinskas J. Review: helicobacter pylori and non-malignant upper gastro-intestinal diseases. *Helicobacter* 2020;25:e12738.
33. Tian C, Hromatka BS, Kiefer AK, *et al.* Genome-wide association and HLA region fine-mapping studies identify susceptibility loci for multiple common infections. *Nat Commun* 2017;8:599.
34. Nielsen HL, Dalager-Pedersen M, Nielsen H. High risk of microscopic colitis after campylobacter concisus infection: population-based cohort study. *Gut* 2020;69:1952–8.
35. Khalili H, Axelrad JE, Roelstraete B, Olén O, D'Amato M, Ludvigsson JF. Gastrointestinal infection and risk of microscopic colitis: A nationwide case-control study in Sweden. *Gastroenterology* 2021;160:1599–607.e5.
36. Khalili H, Zheng T, Soderling J, *et al.*; COVID-19 and Microscopic Colitis Collaborators. Association between collagenous and lymphocytic colitis and risk of severe coronavirus disease 2019. *Gastroenterology* 2021;160:2585–7.e3.
37. Carstens A, Dicksved J, Nelson R, *et al.* The gut microbiota in collagenous colitis shares characteristics with inflammatory bowel disease-associated dysbiosis. *Clin Transl Gastroenterol* 2019;10:e00065.
38. Fernandez-Banares F, Gisbert JP. Letter: Are lymphocytic colitis and collagenous colitis really the same disease? *Aliment Pharmacol Ther* 2012;36:606.
39. Miehlke S, Verhaegh B, Tontini GE, Madisch A, Langner C, Münch A. Microscopic colitis: pathophysiology and clinical management. *Lancet Gastroenterol Hepatol* 2019;4:305–14.
40. Daferera N, Escudero-Hernandez C, Nystrom S, *et al.* Collagenous colitis mucosa is characterized by an expansion of nonsuppressive foxp3+ T helper cells. *Inflamm Bowel Dis* 2021;27:1482–90.
41. Koch S, Münch A, Escudero-Hernández C. P017. Transcriptional characterisation of lymphocytic colitis and comparison with collagenous colitis. *J Crohns Colitis* 2022;16:i144–5.
42. Carrasco A, Esteve M, Salas A, *et al.* Immunological differences between lymphocytic and collagenous colitis. *J Crohns Colitis* 2016;10:1055–66.
43. Martorana D, Marquez A, Carmona FD, *et al.* A large-scale genetic analysis reveals an autoimmune origin of idiopathic retroperitoneal fibrosis. *J Allergy Clin Immunol* 2018;142:1662–5.
44. Fingerlin TE, Zhang W, Yang IV, *et al.* Genome-wide imputation study identifies novel HLA locus for pulmonary fibrosis and potential role for auto-immunity in fibrotic idiopathic interstitial pneumonia. *BMC Genet* 2016;17:74.
45. Tziotzios C, Petridis C, Dand N, *et al.* Genome-wide association study in frontal fibrosing alopecia identifies four susceptibility loci including HLA-b*07:02. *Nat Commun* 2019;10:1150.
46. Ban Y, Davies TF, Greenberg DA, *et al.* Arginine at position 74 of the HLA-DR beta1 chain is associated with Graves' disease. *Genes Immunol* 2004;5:203–8.
47. Rothwell S, Cooper RG, Lundberg IE, *et al.*; Myositis Genetics Consortium. Dense genotyping of immune-related loci in idiopathic inflammatory myopathies confirms HLA alleles as the strongest genetic risk factor and suggests different genetic background for major clinical subgroups. *Ann Rheum Dis* 2016;75:1558–66.
48. Raychaudhuri S, Sandor C, Stahl EA, *et al.* Five amino acids in three HLA proteins explain most of the association between MHC and seropositive rheumatoid arthritis. *Nat Genet* 2012;44:291–6.
49. Darlay R, Ayers KL, Mellis GF, *et al.* Amino acid residues in five separate HLA genes can explain most of the known associations between the MHC and primary biliary cholangitis. *PLoS Genet* 2018;14:e1007833.
50. Diab BY, Lambert NC, L'Faqihi FE, Loubet-Lescoulié P, de Préval C, Coppin H. Human collagen II peptide 256–271 preferentially binds to HLA-DR molecules associated with susceptibility to rheumatoid arthritis. *Immunogenetics* 1999;49:36–44.
51. Soleimanpour SA, Gupta A, Bakay M, *et al.* The diabetes susceptibility gene clec16a regulates mitophagy. *Cell* 2014;157:1577–90.
52. Rijvers L, Melief MJ, van Langelaar J, *et al.* The role of autoimmunity-related gene clec16a in the B cell receptor-mediated HLA class II pathway. *J Immunol* 2020;205:945–56.
53. Sakaue S, Kanai M, Tanigawa Y, *et al.*; FinnGen. A cross-population atlas of genetic associations for 220 human phenotypes. *Nat Genet* 2021;53:1415–24.
54. Chen MH, Raffield LM, Mousas A, *et al.*; VA Million Veteran Program. Trans-ethnic and ancestry-specific blood-cell genetics in 746,667 individuals from 5 global populations. *Cell* 2020;182:198–1213.e14.
55. Xu D, Guo R, Soback A, *et al.* RMI, a new ob-fold complex essential for bloom syndrome protein to maintain genome stability. *Genes Dev* 2008;22:2843–55.