



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease

Citation for published version:

de Lange, KM, Moutsianas, L, Lee, JC, Lamb, CA, Luo, Y, Kennedy, NA, Jostins, L, Rice, DL, Gutierrez-Achury, J, Ji, S-G, Heap, G, Nimmo, ER, Edwards, C, Henderson, P, Mowatt, C, Sanderson, J, Satsangi, J, Simmons, A, Wilson, DC, Tremelling, M, Hart, A, Matthew, CG, Newman, WG, Parkes, M, Lees, CW, Uhlig, H, Hawkey, C, Prescott, NJ, Ahmad, T, Mansfield, JC, Anderson, CA & Barrett, JC 2017, 'Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease', *Nature Genetics*, vol. 49, no. 2, pp. 256–261. <https://doi.org/10.1038/ng.3760>

Digital Object Identifier (DOI):

[10.1038/ng.3760](https://doi.org/10.1038/ng.3760)

Link:

[Link to publication record in Edinburgh Research Explorer](#)

Document Version:

Peer reviewed version

Published In:

Nature Genetics

General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.



1 **Genome-wide association study implicates immune activation of** 2 **multiple integrin genes in inflammatory bowel disease**

3 Katrina M. de Lange*¹, Loukas Moutsianas*¹, James C. Lee*², Christopher A. Lamb³, Yang Luo^{1,4,5},
4 Nicholas A. Kennedy^{6,7}, Luke Jostins^{8,9}, Daniel L. Rice¹, Javier Gutierrez-Achury¹, Sun-Gou Ji¹,
5 Graham Heap^{6,7}, Elaine R. Nimmo¹⁰, Cathryn Edwards¹¹, Paul Henderson^{12,13}, Craig Mowat¹⁴,
6 Jeremy Sanderson¹⁵, Jack Satsangi¹⁰, Alison Simmons^{16,17}, David C. Wilson^{18,19}, Mark Tremelling²⁰,
7 Ailsa Hart²¹, Christopher G. Mathew^{22,23}, William G. Newman^{24,25}, Miles Parkes², Charlie W. Lees¹⁰,
8 Holm Uhlig²⁶, Chris Hawkey²⁷, Natalie J. Prescott²², Tariq Ahmad^{6,7}, John C. Mansfield²⁸, Carl A.
9 Anderson^{†1}, Jeffrey C. Barrett^{†1}

- 10 [1] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK
11 [2] Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, Cambridge, UK
12 [3] Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne
13 [4] Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
14 [5] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA
15 [6] Precision Medicine Exeter, University of Exeter, Exeter, UK
16 [7] IBD Pharmacogenetics, Royal Devon and Exeter Foundation Trust, Exeter, UK
17 [8] Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, UK
18 [9] Christ Church, University of Oxford, St Aldates, UK
19 [10] Gastrointestinal Unit, Wester General Hospital University of Edinburgh, Edinburgh, UK
20 [11] Department of Gastroenterology, Torbay Hospital, Torbay, Devon, UK
21 [12] Department of Child Life and Health, University of Edinburgh, Edinburgh, UK
22 [13] Department of Paediatric Gastroenterology and Nutrition, Royal Hospital for Sick Children, Edinburgh, UK
23 [14] Department of Medicine, Ninewells Hospital and Medical School, Dundee, UK
24 [15] Guy's & St Thomas' NHS Foundation Trust, St Thomas' Hospital, Department of Gastroenterology, London, UK
25 [16] Translational Gastroenterology Unit, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK
26 [17] Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK
27 [18] Paediatric Gastroenterology and Nutrition, Royal Hospital for Sick Children, Edinburgh, UK
28 [19] Child Life and Health, University of Edinburgh, Edinburgh, Scotland, UK
29 [20] Gastroenterology & General Medicine, Norfolk and Norwich University Hospital, Norwich, UK
30 [21] Department of Medicine, St Mark's Hospital, Harrow, Middlesex, UK
31 [22] Department of Medical and Molecular Genetics, Faculty of Life Science and Medicine, King's College London, Guy's
32 Hospital, London, UK
33 [23] Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of Witwatersrand, South Africa.
34 [24] Genetic Medicine, Manchester Academic Health Science Centre, Manchester, UK
35 [25] The Manchester Centre for Genomic Medicine, University of Manchester, Manchester, UK
36 [26] Translational Gastroenterology Unit and the Department of Paediatrics, University of Oxford, Oxford, United Kingdom
37 [27] Nottingham Digestive Diseases Centre, Queens Medical Centre, Nottingham, UK
38 [28] Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK

39 * These authors contributed equally to this work

40 † These authors jointly supervised this work

41 Correspondence should be addressed to Jeffrey C. Barrett (jb26@sanger.ac.uk) and Carl A. Anderson (ca3@sanger.ac.uk)

42 **Genetic association studies have identified 215 risk loci for inflammatory bowel disease¹⁻⁸,**
43 **which have revealed fundamental aspects of its molecular biology. We performed a genome-**
44 **wide association study of 25,305 individuals, and meta-analyzed with published summary**
45 **statistics, yielding a total sample size of 59,957 subjects. We identified 25 new loci, three of**
46 **which contain integrin genes that encode proteins in pathways identified as important**
47 **therapeutic targets in inflammatory bowel disease. The associated variants are correlated**
48 **with expression changes in response to immune stimulus at two of these genes (*ITGA4*,**
49 ***ITGB8*) and at previously implicated loci (*ITGAL*, *ICAM1*). In all four cases, the expression**
50 **increasing allele also increases disease risk. We also identified likely causal missense**
51 **variants in the primary immune deficiency gene *PLCG2* and a negative regulator of**
52 **inflammation, *SLAMF8*. Our results demonstrate that new common variant associations**
53 **continue to identify genes relevant to therapeutic target identification and prioritization.**

54 Inflammatory bowel disease (IBD) is a chronic, debilitating, disorder of the gastrointestinal tract that
55 includes two common disease subtypes, Crohn's disease and ulcerative colitis. Disease
56 pathogenesis is poorly understood but is likely driven by a dysregulated immune response to
57 unknown environmental triggers in genetically susceptible individuals. Treatment regimes often use
58 potent immunomodulators to achieve and maintain remission of symptoms. However, patients
59 commonly experience side effects, lose response to treatment, or develop complications of IBD, with
60 many ultimately requiring major abdominal surgery. Previous genome-wide association studies
61 (GWAS) and targeted follow-up using the ImmunoChip have been very successful at identifying
62 genetic risk loci for IBD, but increased biological understanding has not yet had a significant impact
63 on therapy for these disorders.

64 In order to further expand our understanding of the biology of these disorders we carried out a
65 GWAS of 12,160 IBD cases and 13,145 population controls of European ancestry that had not been
66 included in any genome-wide meta-analysis of IBD to date (Supplementary Table 1, Online
67 Methods). We imputed genotypes using a reference panel comprising whole genome sequences
68 from 4,686 IBD cases⁹ and 6,285 publically available population controls^{10,11}. Following quality
69 control (Online Methods) we tested 9.7 million sites for association. At the 232 IBD associated SNPs
70 in the latest meta-analysis by the International IBD Genetics Consortium¹, 228 had effects in the

71 same direction in our data, 188 showed at least nominal evidence of replication ($P < 0.05$) and none
72 showed significant evidence of heterogeneity of effect by Cochran's Q test. Among these replicated
73 loci was a genome-wide significant association on chromosome 10q25 that was only previously
74 significantly associated with Crohn's disease in individuals of East Asian ancestry^{3,7}, further
75 supporting near complete sharing of genetic risk loci across populations¹. We meta-analyzed our
76 new GWAS data with previously published summary statistics from 12,882 IBD cases and 21,770
77 population controls imputed using the 1000 Genomes Project reference panel¹ (Supplementary
78 Figures 1-3, Supplementary Table 2). We observed inflation of the summary statistics ($\lambda_{GC} = 1.23$
79 and 1.29 for Crohn's and ulcerative colitis, respectively), but LD score regression demonstrated that
80 this was due to broad polygenic signal, rather than confounding population substructure (both
81 intercepts = 1.09, Online Methods).

82 We identified 25 new loci at genome-wide significance (**Table 1**). In order to identify causal variants,
83 genes and mechanisms, we performed a summary-statistic fine-mapping analysis on these loci, as
84 well as 40 previously discovered loci that were genome-wide significant in our data but where fine-
85 mapping had not yet been attempted¹² (Online Methods, Supplementary Table 3). In order to be
86 confident about fine-mapping inferences, we restricted subsequent analyses to 12 signals where we
87 had high quality imputed data for all relevant variants (Online Methods). At 6 of these 12 loci we
88 identified a single variant with >50% probability of being causal (**Table 2**, Supplementary Figures 4-
89 6). Among these were two loci where a single variant had >99% probability of being causal: a
90 missense variant predicted to affect protein function in *SLAMF8*, (rs34687326, p.Gly99Ser, **Figure**
91 **1a**), and an intronic variant in the key regulator of Th17 cell differentiation, *RORC*¹³. *SLAMF8* is a
92 cell surface receptor that is expressed on activated myeloid cells and has been reported to
93 negatively regulate inflammatory responses by inhibiting their migration to sites of inflammation¹⁴
94 and repressing their production of reactive oxygen species (ROS)¹⁵. **This, together with the**
95 **observation that the risk-decreasing allele (MAF=0.1) is predicted to affect protein function**
96 **(CADD=32.0, 92nd percentile of missense variants)¹⁶, suggests further experiments evaluating a**
97 **possible gain-of-function mechanism may be worthwhile.** *RORC* encodes ROR γ t, the master
98 transcriptional regulator of Th17 cells¹³ and group 3 innate lymphoid cells¹⁷. Both of these cell types
99 play important roles in defence at mucosal surfaces, especially in the intestine, and have been
100 shown to contribute to the homeostasis between the intestinal immune system and gut

101 microbiota^{18,19}, an equilibrium that is known to be lost in inflammatory bowel disease²⁰.
102 Pharmacologic inhibition of ROR γ t has been shown to offer therapeutic benefit in mouse models of
103 intestinal inflammation, and reduces the frequency of Th17 cells isolated from primary intestinal
104 samples of IBD patients²¹.

105 In loci where fine-mapping was less clearly resolved, we searched for likely functional variants,
106 observing a missense variant (CADD=16.5, 50.2% probability of causality) in *PLCG2*. Furthermore,
107 after conditioning on this variant, we discovered a second, independent, likely functional
108 (CADD=34.0, 74.6% probability of causality) missense variant in the same gene ($P=2 \times 10^{-8}$). *PLCG2*
109 encodes a phospholipase enzyme that plays a critical role in regulating immune pathway
110 signalling²², and has previously been implicated in two autosomal dominant immune disorders.
111 Intragenic deletions in its autoinhibitory domain cause antibody deficiency and immune dysregulation
112 (familial cold autoinflammatory syndrome 3, MIM 614468)²³ and heterozygous missense variants
113 (e.g. p.Ser707Tyr) lead to a phenotype that includes intestinal inflammation²⁴ (**Figure 1b**).

114 A more general overlap between candidate IBD GWAS genes and Mendelian disorders of
115 inflammation and immunity has been previously observed in 163 loci discovered at that time²⁵. We
116 replicated this finding in our list of 241 loci ($p < 10^{-6}$, Supplementary Table 4), and observed that this
117 enrichment is even stronger when considering just the 26 loci where a gene can be confidently
118 implicated by fine-mapping to a coding variant or colocalisation with an eQTL (27% vs 3%, $p=2 \times 10^{-5}$).
119 In addition to *PLCG2* we identified an association between Crohn's disease and an intronic
120 variant in *NCF4* ($P=1.76 \times 10^{-8}$). This gene encodes p40phox, a component of the NADPH-oxidase
121 system that is responsible for the oxidative burst in innate immune cells and which is a key
122 mechanism of killing phagocytosed bacteria. Rare pathogenic variants in *NCF4* cause autosomal
123 recessive chronic granulomatous disease, characterized by Crohn's disease-like intestinal
124 inflammation and defective ROS production in neutrophils²⁶. Our associated variant, rs4821544, had
125 previously been suggestively associated with small bowel Crohn's disease^{27,28}, and when we
126 stratified patients by disease location we found that the effect was consistently stronger for small
127 bowel compared to large bowel disease (Supplementary Figure 7).

128 Among the remaining 21 novel loci we noted three that were within 150kb of integrin genes (*ITGA4*,
129 *ITGAV* and *ITGB8*), while a previously associated locus overlaps with a fourth integrin, *ITGAL*.

130 Furthermore, a recent study demonstrated that there is an IBD specific association that affects
131 expression of *ICAM1*, which encodes the binding partner of *ITGAL*²⁹. Integrins are cell adhesion
132 mediators with bi-directional signalling capabilities that play a crucial role in leukocyte homing and
133 cell differentiation in inflammation and cancer³⁰. Given the strong candidacy of these genes, we
134 sought potentially causal molecular mechanisms that would connect the IBD associated SNPs to
135 integrin regulation. Our fine-mapping analysis excluded the possibility that these associations are
136 caused by protein-coding changes, so we next tested for effects of IBD risk SNPs on integrin gene
137 expression in immune cells using twelve publicly available eQTL datasets. While many eQTLs and
138 GWAS signals show some degree of correlation, inferences about causality require more robust
139 statistical co-localization of the two signals. Remarkably, we observed three of our five associations
140 had >90% probability of being driven by the same variants as monocyte-specific stimulus response
141 eQTLs (*ITGA4*, $P_{LPS_24hr}=0.984$; *ITGAL*, $P_{LPS_24hr}=0.980$; *ICAM1*, $P_{LPS_2hr}=0.961$; Supplementary
142 Table 5). A fourth association, *ITGB8*, is difficult to map due to extended linkage disequilibrium in the
143 locus, but shows intermediate evidence of co-localization ($P_{LPS_24hr}=0.712$) in response to the same
144 stimulus (**Figure 2**). All four of the IBD risk increasing alleles upregulate expression of their
145 respective genes, suggesting that increased levels of pro-inflammatory cell surface markers in
146 response to stimulus may be a consistent mechanism of action. Proving this hypothesis would
147 require showing that IBD risk alleles causally change stimulus-response expression (e.g. by targeted
148 editing of each allele in cell lines homozygous for the low risk haplotype), and moreover that such
149 changes have physiological relevance to disease processes.

150 One line of evidence that supports such disease relevance for integrins and their counter-receptors
151 is their recent emergence as important therapeutic targets in IBD. Most promisingly, the monoclonal
152 antibodies vedolizumab and etrolizumab, which target the components of the $\alpha4\beta7$ dimer (encoded
153 by *ITGA4* and *ITGB7*, and responsible for the gut-homing specificity of certain leukocytes), have
154 demonstrated efficacy in IBD³¹⁻³³. Additionally, an antisense oligonucleotide targeting ICAM1 has
155 shown promise in the treatment of ulcerative colitis and pouchitis³⁴. The importance of gut-selectivity
156 for therapeutic approaches is highlighted by the antibodies that bind the αL and $\alpha4$ integrin subunits
157 (encoded by *ITGAL* and *ITGA4*, respectively). Therapies targeting αL (efalizumab) and $\alpha4$
158 (natalizumab) demonstrated potential in Crohn's disease^{35,36}, but both medications have been
159 associated with progressive multifocal leukoencephalopathy (PML)³⁷. This potentially fatal condition

160 is likely mediated by binding to integrin dimers that are not gut-specific, leading to impaired
161 leukocyte migration to the central nervous system and JC virus infection of the brain. Owing to the
162 risk of PML, efalizumab has been withdrawn from the market and natalizumab is not licensed for
163 Crohn's disease in Europe.

164 Integrins are not only important in cell trafficking, but can also participate in cellular signalling. For
165 example, the $\alpha V\beta 8$ heterodimer – both subunits of which are encoded by genes which are now
166 within confirmed IBD loci (*ITGAV* and *ITGB8*, respectively) – is a potent activator of TGF β ³⁸, with a
167 range of cell-type specific effects. Furthermore, mice with dendritic-cell specific deletion of this
168 complex had impaired regulatory T cell function and severe colitis³⁹, whereas deleting the complex in
169 regulatory T cells themselves prevented them from suppressing pathogenic T cell responses during
170 active inflammation⁴⁰. While no current IBD therapeutics target $\alpha V\beta 8$ directly, promising early results
171 of an oral antisense oligonucleotide to the inhibitory TGF β -signalling protein SMAD7⁴¹, itself
172 encoded by a locus identified by genetic association studies²⁵, demonstrate the therapeutic potential
173 of modifying TGF β signaling in Crohn's disease.

174 In addition to the connections to anti-integrin and anti-TGF β therapies described above, IBD GWAS
175 have previously implicated loci containing other therapeutically relevant genes, such as those in
176 signalling pathways relevant to the targets of anti-TNF and anti-p40 IBD therapies (**Figure 3**,
177 Supplementary Table 6). These discoveries have demonstrated that the importance of the biological
178 pathways underlying associations, and their potential therapeutic relevance, are not necessarily
179 reflected in their GWAS effect sizes. For example, the modest odds ratios of the signals near integrin
180 genes (1.10-1.12) required tens of thousands of samples to detect at genome-wide significance.
181 Furthermore, analyses aimed at understanding the specific cellular contexts in which these genes
182 are active in IBD, as well as the risk-increasing direction of effect (e.g. consistent up-regulation of
183 integrins in response to LPS stimulus), are only beginning to bear fruit.

184 Our study has demonstrated that continuing to pursue GWAS, even in a well studied complex
185 disease like IBD, has the potential to complement other powerful approaches, such as targeted
186 genotyping (via the ImmunoChip) and large-scale genome and exome sequencing. In two cases we
187 have implicated genes in which different variants have previously been shown to cause immune-
188 related Mendelian disorders, echoing a connection made to the very first Crohn's disease risk gene,

189 *NOD2*, in which rare missense mutations cause the autosomal dominant granulomatous disorder
190 Blau syndrome⁴². Finally, while the individual effect sizes of our newly discovered associations are
191 modest, our results show that GWAS continues to deliver new loci, which help understand many
192 aspects of disease biology, including possible mechanisms of known therapies. For example, four
193 IBD associations that plausibly co-localize with changes in integrin expression underscore the value
194 of comprehensive catalogs of the regulatory consequences of GWAS variants in specific cells and
195 contexts. Even when specific genes are implicated, cellular assays with relevance to disease
196 physiology (for example, protein response to bacterial stimulus in colonic organoids) will be needed
197 to achieve the ultimate payoff from prospectively mining these signals for promising targets for new
198 therapeutics.

199 **Data availability**

200 Genotype data that supports this study has been deposited in the European Genome-phenome
201 Archive (EGA) under the accession code [EGAS00001000924](https://ega-archive.org/studies/EGAS00001000924). Association summary statistics are
202 available from ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2016-11-07/.

203 **Acknowledgements**

204 We would like to thank all individuals who contributed samples to the study. This work was co-
205 funded by the Wellcome Trust [098051] and the Medical Research Council, UK [MR/J00314X/1].
206 Case collections were supported by Crohn's and Colitis UK. KMdL, LM, CAL, YL, DR, JG-A, NJP,
207 CAA and JCB are supported by the Wellcome Trust [098051; 093885/Z/10/Z; 094491/Z/10/Z]. KMdL
208 is supported by a Woolf Fisher Trust scholarship. CAL is a clinical lecturer funded by the NIHR. We
209 thank Anna Stanton for co-ordinating the Guy's and St Thomas' patient recruitment. We
210 acknowledge support from the Department of Health via the NIHR comprehensive Biomedical
211 Research Centre awards to Guy's and St Thomas' NHS Foundation Trust in partnership with King's
212 College London and to Addenbrooke's Hospital, Cambridge in partnership with the University of
213 Cambridge. This research was also supported by the NIHR Newcastle Biomedical Research Centre.
214 The UK Household Longitudinal Study is led by the Institute for Social and Economic Research at
215 the University of Essex and funded by the Economic and Social Research Council. The survey was
216 conducted by NatCen and the genome-wide scan data were analysed and deposited by the
217 Wellcome Trust Sanger Institute. Information on how to access the data can be found on the
218 Understanding Society website <https://www.understandingsociety.ac.uk/>.

219 **Author contributions**

220 KMdL, LM, YL, LJ, DLR, CAA, and SGJ performed statistical analysis. KMdL, LM, YL, LJ, JCL, JGA,
221 SGJ, CAL, NAK, and CAA analysed the data. GH, ERN, CE, CM, AS, DCW, MT, AH, CGM, MP,
222 WGM, CWL, HU, CH, NJP, TA, JCM, JackS, JerS, and PH contributed samples/materials. CAA,
223 JCB, KMdL, LM, JCL, CGM, MP, CAL, NAK, YL, and PH wrote the paper. JCB, CAA, JCM, MP,
224 CWL, TA, and NJP conceived & designed experiments.

225 **Competing financial interests**

226 The authors declare no competing financial interests.

227 References

- 228 1. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and
229 highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–989 (2015).
- 230 2. Parkes, M. *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci
231 contribute to Crohn's disease susceptibility. *Nat. Genet.* **39**, 830–832 (2007).
- 232 3. Yamazaki, K. *et al.* A Genome-Wide Association Study Identifies 2 Susceptibility Loci for Crohn's Disease
233 in a Japanese Population. *Gastroenterology* **144**, 781–788 (2013).
- 234 4. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the
235 number of confirmed associations to 47. *Nat. Genet.* **43**, 246–252 (2011).
- 236 5. Kenny, E. E. *et al.* A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility
237 loci. *PLoS Genet.* **8**, (2012).
- 238 6. Julià, A. *et al.* A genome-wide association study identifies a novel locus at 6q22.1 associated with
239 ulcerative colitis. *Hum. Mol. Genet.* **23**, 6927–6934 (2014).
- 240 7. Yang, S.-K. *et al.* Genome-wide association study of Crohn's disease in Koreans revealed three new
241 susceptibility loci and common attributes of genetic susceptibility across ethnic populations. *Gut* **63**, 80–87
242 (2014).
- 243 8. Ellinghaus, D. *et al.* Analysis of five chronic inflammatory diseases identifies 27 new associations and
244 highlights disease-specific patterns at shared loci. *Nat. Genet.* **48**, 510–518 (2016).
- 245 9. Luo, Y. *et al.* Exploring the genetic architecture of inflammatory bowel disease by whole genome
246 sequencing identifies association at ADCY7. *Nat. Genet.* (In Press)
- 247 10. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526**, 82–90
248 (2015).
- 249 11. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526**, 68–
250 74 (2015).
- 251 12. Huang, H. *et al.* Association mapping of inflammatory bowel disease loci to single variant resolution.
252 *bioRxiv* 028688 (2015). doi:10.1101/028688
- 253 13. Ivanov, I. I. *et al.* The Orphan Nuclear Receptor ROR γ t Directs the Differentiation Program of
254 Proinflammatory IL-17+ T Helper Cells. *Cell* **126**, 1121–1133 (2006).
- 255 14. Wang, G. *et al.* Migration of myeloid cells during inflammation is differentially regulated by the cell surface
256 receptors Slamf1 and Slamf8. *PLoS One* **10**, e0121968 (2015).
- 257 15. Wang, G. *et al.* Cutting edge: Slamf8 is a negative regulator of Nox2 activity in macrophages. *J. Immunol.*
258 **188**, 5829–5832 (2012).
- 259 16. Kircher, M. *et al.* A general framework for estimating the relative pathogenicity of human genetic variants.

- 260 *Nat. Genet.* **46**, 310–315 (2014).
- 261 17. Luci, C. *et al.* Influence of the transcription factor ROR γ on the development of NKp46+ cell
262 populations in gut and skin. *Nat. Immunol.* **10**, 75–82 (2009).
- 263 18. Yang, Y. *et al.* Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature*
264 **510**, 152–156 (2014).
- 265 19. Sawa, S. *et al.* ROR γ t+ innate lymphoid cells regulate intestinal homeostasis by integrating negative
266 signals from the symbiotic microbiota. *Nat. Immunol.* **12**, 320–326 (2011).
- 267 20. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15**,
268 382–392 (2014).
- 269 21. Withers, D. R. *et al.* Transient inhibition of ROR- γ therapeutically limits intestinal inflammation by
270 reducing TH17 cells and preserving group 3 innate lymphoid cells. *Nat. Med.* **22**, 319–323 (2016).
- 271 22. Fu, G., Chen, Y., Schuman, J., Wang, D. & Wen, R. Phospholipase C γ 2 plays a role in TCR signal
272 transduction and T cell selection. *J. Immunol.* **189**, 2326–2332 (2012).
- 273 23. Ombrello, M. J. *et al.* Cold urticaria, immunodeficiency, and autoimmunity related to PLCG2 deletions. *N.*
274 *Engl. J. Med.* **366**, 330–338 (2012).
- 275 24. Zhou, Q. *et al.* A hypermorphic missense mutation in PLCG2, encoding phospholipase C γ 2, causes a
276 dominantly inherited autoinflammatory disease with immunodeficiency. *Am. J. Hum. Genet.* **91**, 713–720
277 (2012).
- 278 25. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel
279 disease. *Nature* **491**, 119–124 (2012).
- 280 26. Matute, J. D. *et al.* A new genetic subgroup of chronic granulomatous disease with autosomal recessive
281 mutations in p40 phox and selective defects in neutrophil NADPH oxidase activity. *Blood* **114**, 3309–3315
282 (2009).
- 283 27. Rioux, J. D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and
284 implicates autophagy in disease pathogenesis. *Nat. Genet.* **39**, 596–604 (2007).
- 285 28. Roberts, R. L. *et al.* Confirmation of association of IRGM and NCF4 with ileal Crohn's disease in a
286 population-based cohort. *Genes Immun.* **9**, 561–565 (2008).
- 287 29. Dendrou, C. A. *et al.* Resolving TYK2 locus genotype-to-phenotype differences in autoimmunity. *Sci.*
288 *Transl. Med.* **8**, 363ra149–363ra149 (2016).
- 289 30. Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**, 673–687 (2002).
- 290 31. Sandborn, W. J. *et al.* Vedolizumab as induction and maintenance therapy for Crohn's disease. *N. Engl. J.*
291 *Med.* **369**, 711–721 (2013).
- 292 32. Feagan, B. G. *et al.* Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N. Engl. J.*
293 *Med.* **369**, 699–710 (2013).

- 294 33. Vermeire, S. *et al.* Etrolizumab as induction therapy for ulcerative colitis: a randomised, controlled, phase 2
295 trial. *Lancet* **384**, 309–318 (2014).
- 296 34. Hosten, T. A., Zhao, K., Han, H. Q., Liu, G. & He, X. H. Alicaforfen: An Emerging Therapeutic Agent for
297 Ulcerative Colitis and Refractory Pouchitis. *Gastroenterol. Res. Pract.* **7**, 51–55 (2014).
- 298 35. James, D. G., Seo, D. H., Chen, J., Vemulapalli, C. & Stone, C. D. Efalizumab, a human monoclonal anti-
299 CD11a antibody, in the treatment of moderate to severe Crohn's disease: An open-label pilot study. *Dig.*
300 *Dis. Sci.* **56**, 1806–1810 (2011).
- 301 36. Sandborn, W. J. *et al.* Natalizumab induction and maintenance therapy for Crohn's disease. *N. Engl. J.*
302 *Med.* **353**, 1912–1925 (2005).
- 303 37. Carson, K. R. *et al.* Monoclonal antibody-associated progressive multifocal leucoencephalopathy in
304 patients treated with rituximab, natalizumab, and efalizumab: a Review from the Research on Adverse
305 Drug Events and Reports (RADAR) Project. *Lancet Oncol.* **10**, 816–824 (2009).
- 306 38. Travis, M. A. & Sheppard, D. TGF- β activation and function in immunity. *Annu. Rev. Immunol.* **32**, 51–82
307 (2014).
- 308 39. Travis, M. A. *et al.* Loss of integrin $\alpha(v)\beta8$ on dendritic cells causes autoimmunity and colitis in mice.
309 *Nature* **449**, 361–365 (2007).
- 310 40. Worthington, J. J. *et al.* Integrin $\alpha v \beta 8$ -Mediated TGF- β Activation by Effector Regulatory T Cells Is
311 Essential for Suppression of T-Cell-Mediated Inflammation. *Immunity* **42**, 903–915 (2015).
- 312 41. Monteleone, G. *et al.* Mongersen, an oral SMAD7 antisense oligonucleotide, and Crohn's disease. *N. Engl.*
313 *J. Med.* **372**, 1104–1113 (2015).
- 314 42. Miceli-Richard, C. *et al.* CARD15 mutations in Blau syndrome. *Nat. Genet.* **29**, 19–20 (2001).
- 315

316 **Figure legends**

317

318 **Figure 1. Likely causal missense variants.** For A) SLAMF8 and B) PLCG2, local association
319 results are plotted with point size corresponding to LD to our lead variant and color to fine-mapping
320 probability (purple > 50%, intermediate blue 10-50%, navy blue <10%). Gene body diagrams and
321 protein domain annotations are taken from ENSEMBL, and partial predicted crystal structures for
322 both proteins are obtained from the SWISS-MODEL repository.

323

324 **Figure 2. Co-localization of disease association and stimulus response eQTLs in monocytes.**

325 The local pattern of disease association (IBD: (A) *ITGA4*, (B) *ITGB8*, (C) *ICAM1*; (D) UC: *ITGAL*) in
326 grey, and the association of that variant with response to LPS (lipopolysaccharide) stimulation in red.
327 Evidence of co-localization (probability > 70%) is observed for all for signals.

328

329 **Figure 3. IBD-associated loci containing genes in immune pathways related to classes of**

330 **approved therapeutics.** All IBD loci are divided into the studies where they were first identified¹.

331 Loci that contain a gene in one of four signalling pathways related to targets of three classes of
332 **approved** IBD therapeutics (**Online Methods**) are highlighted, with those where the pathway gene
333 has been confidently identified as the causal IBD gene labelled. Despite the general pattern that
334 effect size decreases from left to right, therapeutically relevant associations continue to be found.

335
336
337

Tables

Table 1. Novel IBD-associated loci.

Rsid	Chr	Position bp	Left - right Mb	Risk Allele	Non-risk Allele	Risk Allele Frequency in 1000 Genomes CEU+GBR	P _{Meta}	OR	95% CI	Phenotype	Implicated gene
rs34687326	1	159799910	159.80 - 159.80	G	A	0.900	1.06 x 10 ⁻¹⁰⁰	1.18	1.12 - 1.24	CD	<i>SLAMF8</i>
rs59043219	1	209970610	209.97 - 210.02	A	G	0.379	1.09 x 10 ⁻¹⁰⁵	1.08	1.05 - 1.10	IBD	-
rs6740847	2	182308352	182.31 - 182.33	A	G	0.508	1.22 x 10 ⁻¹³³	1.10	1.07 - 1.12	IBD	<i>ITGA4</i>
rs144344067	2	187576378	187.50 - 187.68	A	AT	0.895	1.29 x 10 ⁻¹⁰⁵	1.12	1.08 - 1.16	IBD	-
rs1811711	2	228670476	228.67 - 228.67	C	G	0.826	6.09 x 10 ⁻¹⁰⁹	1.14	1.10 - 1.18	UC	-
rs76527535	2	242484701	242.47 - 242.49	C	T	0.745	2.87 x 10 ⁻¹⁰⁵	1.09	1.06 - 1.12	IBD	-
rs2581828	3	53133149	53.10 - 53.17	C	G	0.597	6.46 x 10 ⁻¹⁰⁹	1.10	1.07 - 1.13	CD	-
rs2593855	3	71175495	71.16 - 71.19	C	T	0.663	2.54 x 10 ⁻¹⁰⁹	1.09	1.06 - 1.11	IBD	-
rs503734	3	101023748	100.91 - 101.27	A	G	0.513	2.67 x 10 ⁻¹⁰⁰	1.07	1.05 - 1.10	IBD	-
rs56116661	3	188401160	188.40 - 188.40	C	T	0.795	5.67 x 10 ⁻¹¹⁰	1.14	1.10 - 1.18	CD	-
rs11734570	4	38588453	38.58 - 38.59	A	G	0.368	4.80 x 10 ⁻¹⁰⁰	1.07	1.05 - 1.10	IBD	-
rs17656349	5	149605994	149.59 - 149.63	T	C	0.466	1.54 x 10 ⁻¹⁰⁰	1.09	1.06 - 1.13	UC	-
rs113986290	6	19781009	19.72 - 19.83	C	T	0.989	7.59 x 10 ⁻¹⁰⁹	1.36	1.25 - 1.46	UC	-
rs67289879	6	42007403	42.00 - 42.01	T	C	0.179	3.04 x 10 ⁻¹⁰⁰	1.09	1.06 - 1.13	IBD	-
rs11768365	7	6545188	6.50 - 6.55	A	G	0.816	3.88 x 10 ⁻¹⁰⁰	1.09	1.06 - 1.12	IBD	-
rs149169037	7	20577298	20.58 - 20.58	G	A	0.895	3.26 x 10 ⁻¹⁰⁰	1.14	1.10 - 1.19	IBD	<i>ITGB8</i>
rs243505	7	148435339	148.40 - 148.58	A	G	0.624	3.04 x 10 ⁻¹¹⁰	1.08	1.06 - 1.11	IBD	-
rs7911117	10	27179596	27.16 - 27.18	T	G	0.871	1.84 x 10 ⁻¹⁰⁰	1.14	1.10 - 1.19	UC	-
rs111456533	10	126439381	126.32 - 126.55	G	A	0.829	1.18 x 10 ⁻¹⁰⁹	1.11	1.08 - 1.14	IBD	-
rs80244186	13	42917861	42.84 - 42.94	C	T	0.111	3.66 x 10 ⁻¹⁰⁵	1.13	1.09 - 1.18	CD	-
rs11548656	16	81916912	81.91 - 81.92	A	G	0.961	5.18 x 10 ⁻¹¹¹	1.27	1.20 - 1.34	IBD	<i>PLCG2</i>
rs10492862	16	82867456	82.87 - 82.92	A	C	0.308	1.26 x 10 ⁻¹⁰⁹	1.11	1.08 - 1.15	CD	-
rs4256018	20	6093889	6.08 - 6.10	G	T	0.250	1.23 x 10 ⁻¹⁰⁵	1.08	1.05 - 1.11	IBD	-
rs138788	22	35729721	35.72 - 35.74	A	G	0.418	2.95 x 10 ⁻¹⁰⁵	1.09	1.06 - 1.13	UC	-
rs4821544	22	37258503	37.26 - 37.26	C	T	0.321	1.76 x 10 ⁻¹⁰⁰	1.10	1.07 - 1.13	CD	-

338

339 **Table 2. Variants fine-mapped to >50% probability of being causal in their given signal.**

Rsid	Chr	Position (bp)	P _{Causal}	Effect	Credible set size	Phenotype	P _{Meta}	Locus type
rs34687326	1	159799910	1.000	SLAMF8 p.Gly99Ser (missense)	1	CD	1.06 x 10 ⁻⁰⁸	Novel
rs4845604	1	151801680	0.999	RORC (intronic)	1	IBD	7.09 x 10 ⁻¹⁴	Known
rs1811711	2	228670476	0.914		2	UC	6.09 x 10 ⁻⁰⁹	Novel
rs56116661	3	188401160	0.561	LPP (intronic)	11	CD	5.67 x 10 ⁻¹⁰	Novel
rs11548656	16	81916912	0.502	PLCG2 p.His244Arg (missense)	3	IBD	5.18 x 10 ⁻¹¹	Novel
rs1143687	16	81922813	0.746	PLCG2 p.Arg268Trp (missense)	5	IBD	3.83 x 10 ⁻⁰⁸	Novel
rs4821544	22	37258503	0.804	NCF4 (intronic)	2	CD	1.76 x 10 ⁻⁰⁸	Novel

340

341 **Online Methods**
342

343 **New genome-wide genetic data**

344 *GWAS samples and genotyping.* Following ethical approval by Cambridge MREC (reference:
345 03/5/012), 11,768 British IBD cases, diagnosed using accepted endoscopic, histopathological and
346 radiological criteria, were consented into the study and genotyped on the Human Core Exome v12.1.
347 10,484 population control samples genotyped on the Human Core Exome v12.0 were obtained from
348 the Understanding Society Project. Genotypes were called using optiCall⁴³.

349 *GWAS quality control.* We removed variants that did not overlap between the two versions of the
350 chip, had missingness > 5%, a significant difference in call rate between cases and controls ($P <$
351 1×10^{-5}), deviated from Hardy-Weinberg equilibrium (HWE) in controls ($P < 1 \times 10^{-5}$), or that were
352 affected by a genotyping batch effect (significant association [$P < 1 \times 10^{-5}$] between an outlier group of
353 cases discovered using principal component analysis [$PC1 < -0.005$], and the remainder of the
354 samples). We then removed samples with missingness > 1%, heterozygosity ± 3 standard deviations
355 from the mean, mismatch between reported and genotypic sex, first-degree relatives or closer
356 (kinship coefficient > 0.177), and non-European samples identified through principal component
357 analysis with HapMap3 populations. After quality control, data were available for 4,474 Crohn's
358 disease, 4,173 ulcerative colitis, 592 IBD-unclassified cases and 9,500 controls for 296,203 variants.

359 *Whole-genome sequenced samples.* We generated low-coverage whole genome sequences for
360 4,686 IBD cases and 3,781 population controls from the UK IBD Genetics Consortium (UKIBDGC)
361 and UK10K Consortium, respectively. Detailed information on sequencing, genotype refinement and
362 quality control are described elsewhere⁹.

363 *Imputation.* These sequences were combined with 2,504 samples from the Phase 3 v5 release of
364 the 1000 Genomes project (2013-05-02 sequence freeze) to create a phased imputation reference
365 panel enriched in IBD-associated variants. We used PBWT⁴⁴ to impute from this reference panel
366 (114.2 million total variants) into our new GWAS described above.

367

368 **Association testing, meta-analysis, and quality control.**

369 *Association testing.* Prior to association testing, we removed all samples that were included in
370 previous IBD GWAS meta-analyses (Supplementary Table 1). We then tested for association to
371 ulcerative colitis, Crohn's disease and IBD separately within the sequenced samples and new
372 GWAS using SNPTTEST v2.5, performing an additive frequentist association test conditioned on the
373 first ten principal components for each cohort. We filtered out variants with minor allele frequency
374 (MAF) < 0.1%, INFO < 0.4, or strong evidence for deviations from HWE in controls ($p_{\text{HWE}} < 1 \times 10^{-7}$).

375 *Meta-analysis.* We used METAL (release 2011-03-05) to perform a standard error weighted meta-
376 analysis of our sequencing and GWAS cohorts with the publicly available International Inflammatory
377 Bowel Disease Genetics Consortium (IIBDGC) meta-analysis summary statistics¹, after applying the
378 additional MAF \geq 0.1%, and INFO \geq 0.4 filters to the IIBDGC data.

379 *Quality control.* The output of the fixed-effects meta-analysis was further filtered, and sites with high
380 evidence for heterogeneity ($I^2 > 0.90$) were discarded. Only sites for which all cohorts passed our
381 quality control filters were included in our analysis. In addition, we discarded genome-wide
382 significant variants for which the meta-analysis p-value was not lower than all of the cohort-specific
383 p-values.

384 *LD score regression.* We performed LD score regression using LDSC v1.0.0 and European linkage
385 disequilibrium (LD) scores from the 1000 Genomes Project (downloaded from
386 https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2) on our filtered meta-
387 analysis summary statistics for all sites with INFO > 0.95. This INFO threshold is to avoid
388 confounding due to poor imputation, as recommended by the authors⁴⁵.

389 **Locus definition**

390 *Computing LD windows.* An LD window was calculated for every genome-wide significant variant in
391 any of the three traits (Crohn's disease, ulcerative colitis, IBD), defined by the left-most and right-
392 most variants that are correlated with the main variant with an r^2 of 0.6 or more. The LD was
393 calculated in the GBR and CEU samples from the 1000 Genomes Phase 3, release v5 (based on
394 20130502 sequence freeze and alignments). Loci with overlapping LD windows, as well as loci

395 whose lead variants were separated by 500kb or less, were subsequently merged, and the variant
396 with the strongest evidence of being associated was kept as the lead variant for each merged locus.

397 *Identifying novel loci.* A locus was annotated as known if it contained at least one variant previously
398 reported at genome-wide significance (irrespective of the LD between that variant and the most
399 associated variants in the locus). To ensure that putatively novel signals were not due to long-range
400 LD with variants in previously reported loci, we conducted conditional analysis in our new GWAS for
401 all variants in loci which were less than 3Mb away from a known locus. Putatively novel loci already
402 known in a lower order IBD trait (e.g. a previously known Crohn's disease locus coming up as an
403 IBD locus) were also removed from this list. This did not apply where, for example, a known Crohn's
404 disease locus was now associated with ulcerative colitis, or vice versa.

405 **Fine-mapping**

406 Approximate Bayes factors were calculated from the meta-analysis effect sizes and standard errors
407 described above by applying equation (2) of Wakefield⁴⁶, assuming a prior variance on the log odds
408 ratios of 0.04 (the default prior used by the software SNPTest, and used by Maller *et al*⁴⁷). We then
409 performed fine-mapping using these Bayes factors as described in Maller *et al* to calculate the
410 posterior that each variant is causal, and the 95% credible set for each association (the smallest set
411 of variants with posteriors that sum to at least 95%). For each association we use the meta-analysis
412 results for the phenotype (Crohn's disease, ulcerative colitis or IBD) specified in Supplementary
413 Table 2. We only consider a locus to be confidently fine-mapped if there are no variants in the Phase
414 3 v5 release of the 1000 Genomes project (2013-05-02 sequence freeze) in high LD ($r^2 \geq 0.6$) with
415 our hit SNP, but missing from our dataset, and no variants in our data within high LD ($r^2 > 0.8$) that
416 fail during our QC procedure.

417 **eQTL overlap**

418 *Identifying eQTL overlaps.* Twelve eQTL datasets were searched to identify variants within the 25
419 newly identified IBD risk loci that are associated with variation in gene expression (Supplementary
420 Table 7). Splice-QTLs based on exon-ratio⁴⁸ and transcript-ratio⁴⁹⁻⁵¹ were also included in the
421 search where available (Supplementary Table 7). The most significant variant-gene associations

422 were extracted from each eQTL/splice-QTL dataset and were reported as candidates if that variant
423 had $r^2 > 0.8$ with any of the lead SNPs in the 25 IBD risk loci.

424 *Testing for co-localization.* We tested for co-localization between IBD association signals and eQTLs
425 using the coloc2 method⁵², implemented in the R package coloc. We used a window size of 250kb
426 on either side of the IBD association, and implemented the default settings as recommended. Each
427 test was repeated using two different values for the prior probability of co-localization, p_{12} : 1×10^{-5} and
428 1×10^{-6} .

429 **Signalling pathway definitions**

430 We identify the following immune pathways as relevant to classes of approved IBD therapeutics: the
431 IL12 and IL23 signalling pathways (ustekinumab⁵³), the TNFa signalling pathway (infliximab⁵⁴,
432 adalimumab⁵⁵), and the integrin signalling pathway (vedolizumab^{31,32}). Genes involved in these
433 pathways were identified from the Molecular Signatures Database canonical pathways gene sets
434 (C2; <http://software.broadinstitute.org/gsea/msigdb/genesets.jsp?collection=CP>). These gene lists
435 had been previously curated by the Pathway Interaction Database⁵⁶. The integrin signalling gene list
436 was comprised of all unique genes from the following gene sets: integrin beta1 pathway
437 (PID_INTEGRIN1_PATHWAY), integrin beta7 pathway (PID_INTEGRIN5_PATHWAY) and integrin
438 cell surface interactions (PID_INTEGRIN_CS_PATHWAY). The list of TNFa signalling genes was
439 obtained from PID_TNF_PATHWAY and the list of IL-23/IL-12 p40 signalling genes was comprised
440 of all unique genes from the PID_IL12_PATHWAY and PID_IL23_PATHWAY.

441 **References**

- 442 43. Shah, T. S. *et al.* optiCall: a robust genotype-calling algorithm for rare, low-frequency and common
443 variants. *Bioinformatics* **28**, 1598–1603 (2012).
- 444 44. Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform
445 (PBWT). *Bioinformatics* **30**, 1266–1272 (2014).
- 446 45. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide
447 association studies. *Nat. Genet.* **47**, 291–295 (2015).
- 448 46. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet.*
449 *Epidemiol.* **33**, 79–86 (2009).
- 450 47. Wellcome Trust Case Control Consortium *et al.* Bayesian refinement of association signals for 14 loci in 3
451 common diseases. *Nat. Genet.* **44**, 1294–1301 (2012).
- 452 48. Zhernakova, D. *et al.* Hypothesis-free identification of modulators of genetic risk factors. *bioRxiv* 033217
453 (2015). doi:10.1101/033217
- 454 49. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922
455 individuals. *Genome Res.* **24**, 14–24 (2014).
- 456 50. Monlong, J., Calvo, M., Ferreira, P. G. & Guigó, R. Identification of genetic variants associated with
457 alternative splicing using sQTLseeker. *Nat. Commun.* **5**, 4698 (2014).
- 458 51. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for
459 thousands of molecular phenotypes. *Bioinformatics* **32**, 1479–1485 (2016).
- 460 52. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies
461 using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
- 462 53. Sandborn, W. J. *et al.* Ustekinumab induction and maintenance therapy in refractory Crohn's disease. *N.*
463 *Engl. J. Med.* **367**, 1519–1528 (2012).
- 464 54. Hanauer, S. B. *et al.* Maintenance infliximab for Crohn's disease: the ACCENT I randomised trial. *Lancet*
465 **359**, 1541–1549 (2002).
- 466 55. Colombel, J.-F. *et al.* Adalimumab for maintenance of clinical response and remission in patients with
467 Crohn's disease: the CHARM trial. *Gastroenterology* **132**, 52–65 (2007).
- 468 56. Schaefer, C. F. *et al.* PID: the Pathway Interaction Database. *Nucleic Acids Res.* **37**, D674–9 (2009).