bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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. Lamb[3], Yang Luo[1,4,5],
4 Nicholas A. Kennedy[6,7], Luke Jostins[8,9], Daniel L. Rice[1], Javier Gutierrez-Achury[1], Sun-Gou Ji[1], Graham
5 Heap[6,7], Elaine R. Nimmo[10], Cathryn Edwards[11], Paul Henderson[12,13], Craig Mowat[14], Jeremy
6 Sanderson[15], Jack Satsangi[10], Alison Simmons[16,17], David C. Wilson[18,19], Mark Tremelling[20], Ailsa Hart[21],
7 Christopher G. Mathew[22,23], William G. Newman[24,25], Miles Parkes[2], Charlie W. Lees[10], Holm Uhlig[26],
8 Chris Hawkey[27], Natalie J. Prescott[22], Tariq Ahmad[6,7], John C. Mansfield[28], Carl A. Anderson*[1], Jeffrey C.
9 Barrett*[1]

10 **Genetic association studies have identified 210 risk loci for inflammatory bowel disease[1–7], which**
11 **have revealed fundamental aspects of the molecular biology of the disease, including the roles of**
12 **autophagy and Th17 cell signaling and development. We performed a genome-wide association**
13 **study of 25,305 individuals, and meta-analyzed with published summary statistics, yielding a total**
14 **sample size of 59,957 subjects. We identified 26 new genome-wide significant loci, three of which**
15 **contain integrin genes that encode molecules in pathways identified as important therapeutic**
16 **targets in inflammatory bowel disease. The associated variants are also correlated with**
17 **expression changes in response to immune stimulus at two of these genes (*ITGA4, ITGB8*) and at**
18 **two previously implicated integrin loci (*ITGAL*, *ICAM1*). In all four cases, the stimulus-dependent**
19 **expression increasing allele also increases disease risk. We applied summary statistic fine-**
20 **mapping and identified likely causal missense variants in the primary immune deficiency gene**
21 ***PLCG2* and the negative regulator of inflammation, *SLAMF8*. Our results demonstrate that new**
22 **common variant associations continue to identify genes and pathways of relevance to therapeutic**
23 **target identification and prioritization.**

24 Inflammatory bowel disease (IBD) is a chronic, debilitating, disorder of the gastrointestinal tract that
25 includes two common disease subtypes, Crohn's disease and ulcerative colitis. Disease pathogenesis is
26 poorly understood but is likely driven by a dysregulated immune response to unknown environmental
27 triggers in genetically susceptible individuals. Treatment regimes often use potent immunomodulators to
28 achieve and maintain remission of symptoms. However, patients commonly experience side effects, lose

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

29 response to treatment, or develop complications of IBD, with many ultimately requiring major abdominal
30 surgery. Previous genome-wide association studies (GWAS) and targeted follow-up using the
31 Immunochip have been very successful at identifying genetic risk loci for IBD, but increased biological
32 understanding has not yet had a significant impact on therapy for these disorders.

33 In order to further expand our understanding of the biology of these disorders we carried out a GWAS of
34 12,160 IBD cases and 13,145 population controls of European ancestry that had not been included in any
35 genome-wide meta-analysis of IBD to date (Supplementary Table 1, Online Methods). We imputed
36 genotypes using a reference panel comprising whole genome sequences from 4,686 IBD cases[8] and
37 6,285 publically available population controls[9,10]. Following quality control (Online Methods) we tested 9.7
38 million sites for association. At the 232 IBD associated SNPs in the latest meta-analysis by the
39 International IBD Genetics Consortium[1], 228 had effects in the same direction in our data, 188 showed at
40 least nominal evidence of replication ($P<0.05$) and none showed significant evidence of heterogeneity of
41 effect by Cochrane's Q test. Among these replicated loci was a genome-wide significant association on
42 chromosome 10q25 that was only previously significantly associated with Crohn's disease in individuals
43 of East Asian ancestry[3,7], further supporting near complete sharing of genetic risk loci across
44 populations[1]. We meta-analyzed our new GWAS data with previously published summary statistics from
45 12,882 IBD cases and 21,770 population controls imputed using the 1000 Genomes Project reference
46 panel[1] (Supplementary Figures 1-3, Supplementary Table 2). We observed inflation of the summary
47 statistics (λ_{GC} = 1.23 and 1.29 for Crohn's and ulcerative colitis, respectively), but LD score regression
48 demonstrated that this was due to broad polygenic signal, rather than confounding population
49 substructure (both intercepts = 1.09, Online Methods).

50 We identified 26 new loci at genome-wide significance (**Table 1**). In order to identify causal variants,
51 genes and mechanisms, we performed a summary-statistic fine-mapping analysis on these loci, as well
52 as 39 previously discovered loci where fine-mapping had not yet been attempted[11] (Online Methods,
53 Supplementary Table 3). In order to be confident about fine-mapping inferences, we restricted
54 subsequent analyses to 12 signals where we had high quality imputed data for all relevant variants
55 (Online Methods). At 6 of these 12 loci we identified a single variant with >50% probability of being causal
56 (**Table 2**, Supplementary Figures 4-6). Among these were two loci where a single variant had >99%
57 probability of being causal: a missense variant predicted to affect protein function in *SLAMF8*, (**Figure**

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

58 **1a**), and an intronic variant in the key regulator of Th17 cell differentiation, *RORC*[12]. SLAMF8 is a cell
59 surface receptor that is expressed on activated myeloid cells and has been reported to negatively
60 regulate inflammatory responses by repressing the production of reactive oxygen species (ROS)[13] and
61 inhibiting their migration to sites of inflammation[14]. *RORC* encodes RORγt, the master transcriptional
62 regulator of Th17 cells[12] and group 3 innate lymphoid cells[15]. Both of these cell types play important roles
63 in defence at mucosal surfaces, especially in the intestine, and have been shown to contribute to the
64 homeostasis between the intestinal immune system and gut microbiota[16,17], an equilibrium that is known
65 to be lost in inflammatory bowel disease[18]. Pharmacologic inhibition of RORγt has been shown to offer
66 therapeutic benefit in mouse models of intestinal inflammation, and reduces the frequency of Th17 cells
67 isolated from primary intestinal samples of IBD patients[19].

68 In loci where fine-mapping was less clearly resolved, we searched for likely functional variants, observing
69 a missense variant predicted to affect protein function (CADD = 16.45, 50.2% probability of causality) in
70 *PLCG2.* Furthermore, after conditioning on this variant, we discovered a second, independent, likely
71 functional (CADD = 34, 74.6% probability of causality) missense variant in the same gene ($P=2 \times 10^{-8}$).
72 *PLCG2* encodes a phospholipase enzyme that plays a critical role in regulating immune pathway
73 signalling[20], and has previously been implicated in two autosomal dominant immune disorders. Intragenic
74 deletions in its autoinhibitory domain cause antibody deficiency and immune dysregulation (familial cold
75 autoinflammatory syndrome 3, MIM 614468)[21] and heterozygous missense variants (e.g. p.Ser707Tyr)
76 lead to a phenotype that includes intestinal inflammation[22] (**Figure 1b**).

77 A more general overlap between candidate IBD GWAS genes and Mendelian disorders of inflammation
78 and immunity has been previously observed[23]. In addition to *PLCG2* we identified an association between
79 Crohn's disease and an intronic variant in *NCF*4 ($P=1.76 \times 10^{-8}$). This gene encodes p40phox, a
80 component of the NADPH-oxidase system that is responsible for the oxidative burst in innate immune
81 cells and which is a key mechanism of killing phagocytosed bacteria. Rare pathogenic variants in *NCF4*
82 cause autosomal recessive chronic granulomatous disease, characterized by Crohn's disease-like
83 intestinal inflammation and defective ROS production in neutrophils[24]. Our associated variant, rs4821544,
84 had previously been suggestively associated with small bowel Crohn's disease[25,26], and when we
85 stratified patients by disease location we found that the effect was consistently stronger for small bowel
86 compared to large bowel disease (Supplementary Figure 7).

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

87 Among the remaining 22 novel loci we noted three that were within 150kb of integrin genes (*ITGA4,*
88 *ITGAV* and *ITGB8*), while two previously associated loci overlap with a fourth integrin, *ITGAL*, and its
89 binding partner *ICAM1.* Integrins are cell adhesion mediators with bi-directional signalling capabilities that
90 play a crucial role in leukocyte homing and cell differentiation in inflammation and cancer[27]. Given the
91 strong candidacy of these genes, we sought potentially causal molecular mechanisms that would connect
92 the IBD associated SNPs to integrin regulation. Our fine-mapping analysis excluded the possibility that
93 these associations are caused by protein-coding changes, so we next tested for effects of IBD risk SNPs
94 on integrin gene expression in immune cells using nine publicly available eQTL datasets. While many
95 eQTLs and GWAS signals show some degree of correlation, inferences about causality require more
96 robust statistical co-localization of the two signals. Remarkably, we observed three of our five
97 associations had >90% probability of being driven by the same variants as monocyte-specific stimulus
98 response eQTLs (*ITGA4,* P_{LPS_24hr}=0.984; *ITGAL,* P_{LPS_24hr}=0.980; *ICAM1,* P_{LPS_2hr}=0.961; Supplementary
99 Table 4). A fourth association, *ITGB8*, is difficult to map due to extended linkage disequilibrium in the
100 locus, but shows intermediate evidence of co-localization (P_{LPS_24hr}=0.712) in response to the same
101 stimulus (**Figure 2**). These observations suggest upregulation of pro-inflammatory cell surface markers as
102 a potential mechanism of action, as all four of the IBD risk increasing alleles upregulate expression of
103 their respective genes.

104 Integrins and their counter-receptors have recently emerged as important therapeutic targets in IBD. Most
105 promisingly monoclonal antibodies that target the components of the α4β7 dimer, responsible for the gut-
106 homing specificity of leukocyte subsets, have demonstrated efficacy in IBD[28–30]. Additionally, an antisense
107 oligonucleotide targeting *ICAM1* has shown promise in the treatment of ulcerative colitis and pouchitis[31].
108 The importance of gut-selectivity for therapeutic approaches is highlighted by the success of antibodies
109 that bind the αL and α4 integrin subunits. Whilst αL-directed therapy with efalizumab demonstrated
110 potential in Crohn's disease[32], and α4-directed therapy (which binds α4β1 in addition to α4β7 integrin)
111 with natalizumab is licensed in the USA for Crohn's disease[33], both medications have been associated
112 with occurrences of progressive multifocal leukoencephalopathy (PML). This potentially fatal condition is
113 likely mediated by impaired leukocyte migration to the central nervous system leading to JC virus infection
114 of the brain. Owing to the risk of PML, efalizumab has been withdrawn from the market and natalizumab
115 is not licensed for Crohn's disease in Europe.

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

116 Integrins are not only important in cell trafficking, but can also participate in cellular signalling. For
117 example, the αVβ8 heterodimer – both subunits of which are now within confirmed IBD loci – is a potent
118 activator of TGFβ[34], with a range of cell-type specific effects. Furthermore, mice with dendritic-cell specific
119 deletion of this complex had impaired regulatory T cell function and severe colitis[35], whereas deleting the
120 complex in regulatory T cells themselves prevented them from suppressing pathogenic T cell responses
121 during active inflammation[36]. While no current IBD therapeutics target αVβ8 directly, promising early
122 results of an oral antisense oligonucleotide to the inhibitory TGFβ-signalling protein SMAD7[37], itself
123 encoded by a locus identified by genetic association studies[23], demonstrate the therapeutic potential of
124 modifying TGFβ signaling in Crohn's disease.

125 In addition to the connections to anti-integrin and anti-TGFβ therapies described above, IBD GWAS have
126 previously implicated loci containing other therapeutically relevant genes, such as those in pathways
127 targeted by anti-TNF and anti-p40 IBD therapies (**Figure 3**). These discoveries have demonstrated that
128 the importance of the biological pathways underlying associations, and their potential therapeutic
129 relevance, are not necessarily reflected in their GWAS effect sizes. For example, the modest odds ratios
130 of the signals near integrin genes (1.10-1.12) required tens of thousands of samples to detect at genome-
131 wide significance. Furthermore, analyses aimed at understanding the specific cellular contexts in which
132 these genes are active in IBD, as well as the risk-increasing direction of effect (e.g. consistent up-
133 regulation of integrins in response to LPS stimulus), are only beginning to bear fruit.

134 Our study has demonstrated that continuing to pursue GWAS, even in a well studied complex disease
135 like IBD, has the potential to complement other powerful approaches, such as targeted genotyping (via
136 the Immunochip) and large-scale genome and exome sequencing. In two cases we have implicated
137 genes in which different variants have previously been shown to cause immune-related Mendelian
138 disorders, echoing a connection made to the very first Crohn's disease risk gene, *NOD2*, in which rare
139 missense mutations cause the autosomal dominant granulomatous disorder Blau syndrome[38]. Finally,
140 while the individual effect sizes of our newly discovered associations are modest, we believe that our
141 results show that GWAS continues to deliver new understanding of disease biology and new therapeutic
142 opportunities.

143

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

144 **Acknowledgements**

145 We would like to thank all individuals who contributed samples to the study. This work was co-
146 funded by the Wellcome Trust [098051] and the Medical Research Council, UK
147 [MR/J00314X/1]. Case collections were supported by Crohn's and Colitis UK. KMdL, LM, CAL,
148 YL, DR, JG-A, NJP, CAA and JCB are supported by the Wellcome Trust [098051;
149 093885/Z/10/Z; 094491/Z/10/Z]. KMdL is supported by a Woolf Fisher Trust scholarship. CAL is
150 a clinical lecturer funded by the NIHR. We thank Anna Stanton for co-ordinating the Guy's and
151 St Thomas' patient recruitment. We acknowledge support from the Department of Health via the
152 NIHR comprehensive Biomedical Research Centre awards to Guy's and St Thomas' NHS
153 Foundation Trust in partnership with King's College London and to Addenbrooke's Hospital,
154 Cambridge in partnership with the University of Cambridge. This research was also supported
155 by the NIHR Newcastle Biomedical Research Centre.

156 **Author contributions:** KMdL, LM, YL, LJ, DLR, CAA, and SGJ performed statistical analysis.
157 KMdL, LM, YL, LJ, JCL, JGA, SGJ, CAL, NAK, and CAA analysed the data. GH, ERN, CE, CM,
158 AS, DCW, MT, AH, CGM, MP, WGN, CWL, HU, CH, NJP, TA, JCM, JackS, JerS, and PH
159 contributed samples/materials. CAA, JCB, KMdL, LM, JCL, CGM, MP, CAL, NAK, YL, and PH
160 wrote the paper. JCB, CAA, JCM, MP, CWL, TA, and NJP conceived & designed experiments.
161 JCB and CAA jointly supervised research. KMdL, LM, and JCL contributed equally to this work.

162 **Competing financial interests:**

163 The authors declare no competing financial interests.

164 **Materials & Correspondence:**

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

167

168

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

169 ## Author affiliations

170 [1] Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, UK
171 [2] Inflammatory Bowel Disease Research Group, Addenbrooke's Hospital, Cambridge, UK
172 [3] Institute of Cellular Medicine, Newcastle University, Newcastle upon Tyne
173 [4] Division of Genetics and Rheumatology, Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA
174 [5] Program in Medical and Population Genetics, Broad Institute of Harvard and MIT, Cambridge, MA, USA
175 [6] Precision Medicine Exeter, University of Exeter, Exeter, UK
176 [7] IBD Pharmacogenetics, Royal Devon and Exeter Foundation Trust, Exeter, UK
177 [8] Wellcome Trust Centre for Human Genetics, University of Oxford, Headington, UK
178 [9] Christ Church, University of Oxford, St Aldates, UK
179 [10] Gastrointestinal Unit, Wester General Hospital University of Edinburgh, Edinburgh, UK
180 [11] Department of Gastroenterology, Torbay Hospital, Torbay, Devon, UK
181 [12] Department of Child Life and Health, University of Edinburgh, Edinburgh, UK
182 [13] Department of Paediatric Gastroenterology and Nutrition, Royal Hospital for Sick Children,Edinburgh, UK
183 [14] Department of Medicine, Ninewells Hospital and Medical School, Dundee, UK
184 [15] Guy's & St Thomas' NHS Foundation Trust, St Thomas' Hospital, Department of Gastroenterology, London, UK
185 [16] Translational Gastroenterology Unit, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DS, UK
186 [17] Human Immunology Unit, Weatherall Institute of Molecular Medicine, University of Oxford, Oxford OX3 9DS, UK
187 [18] Paediatric Gastroenterology and Nutrition, Royal Hospital for Sick Children, Edinburgh, UK
188 [19] Child Life and Health, University of Edinburgh, Edinburgh, Scotland, UK
189 [20] Gastroenterology & General Medicine, Norfolk and Norwich University Hospital, Norwich, UK
190 [21] Department of Medicine, St Mark's Hospital, Harrow, Middlesex, UK
191 [22] Department of Medical and Molecular Genetics, Faculty of Life Science and Medicine, King's College London, Guy's
192 Hospital, London, UK
193 [23] Sydney Brenner Institute for Molecular Bioscience, Faculty of Health Sciences, University of Witwatersrand, South Africa.
194 [24] Genetic Medicine, Manchester Academic Health Science Centre, Manchester, UK
195 [25] The Manchester Centre for Genomic Medicine, University of Manchester, Manchester, UK
196 [26] Translational Gastroenterology Unit and the Department of Paediatrics, University of Oxford, Oxford, United Kingdom
197 [27] Nottingham Digestive Diseases Centre, Queens Medical Centre, Nottingham, UK
198 [28] Institute of Human Genetics, Newcastle University, Newcastle upon Tyne, UK

199
200
201
202
203
204

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

205

206 **Online Methods**
207

208 **New genome-wide genetic data**

209 *GWAS samples and genotyping.* 11,768 British IBD cases, diagnosed using accepted endoscopic,
210 histopathological and radiological criteria, were genotyped on the Human Core Exome v12.1. 10,484
211 population control samples genotyped on the Human Core Exome v12.0 were obtained from the
212 Understanding Society Project. Genotypes were called using optiCall[39].

213 *GWAS quality control.* We removed variants that did not overlap between the two versions of the chip,
214 had missingness > 5%, a significant difference in call rate between cases and controls ($P < 1x10^{-5}$),
215 deviated from Hardy-Weinberg equilibrium (HWE) in controls ($P < 1x10^{-5}$), or that were affected by a
216 genotyping batch effect (significant association [$P < 1x10^{-5}$] between an outlier group of cases discovered
217 using principal component analysis [PC1 < -0.005], and the remainder of the samples). We then removed
218 samples with missingness > 1%, heterozygosity ±3 standard deviations from the mean, mismatch
219 between reported and genotypic sex, first-degree relatives or closer (kinship coefficient > 0.177), and
220 non-European samples identified through principal component analysis with HapMap3 populations. After
221 quality control, data were available for 4,474 Crohn's disease, 4,173 ulcerative colitis, 592 IBD-
222 unclassified cases and 9,500 controls for 296,203 variants.

223 *Whole-genome sequenced samples*. We generated low-coverage whole genome sequences for 4,686
224 IBD cases and 3,781 population controls from the UK IBD Genetics Consortium (UKIBDGC) and UK10K
225 Consortium, respectively. Detailed information on sequencing, genotype refinement and quality control
226 are described elsewhere[8].

227 *Imputation.* These sequences were combined with 2,504 samples from the Phase 3 v5 release of the
228 1000 Genomes project (2013-05-02 sequence freeze) to create a phased imputation reference panel
229 enriched in IBD-associated variants. We used PBWT[40] to impute from this reference panel (114.2 million
230 total variants) into our new GWAS described above.

231

232

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

234 *Association testing.* Prior to association testing, we removed all samples that were included in previous
235 IBD GWAS meta-analyses (Supplementary Table 1). We then tested for association to ulcerative colitis,
236 Crohn's disease and IBD separately within the sequenced samples and new GWAS using SNPTEST
237 v2.5, performing an additive frequentist association test conditioned on the first ten principal components
238 for each cohort. We filtered out variants with minor allele frequency (MAF) < 0.1%, INFO < 0.4, or strong
239 evidence for deviations from HWE in controls ($p_{HWE}<1x10^{-7}$).

240 *Meta-analysis.* We used METAL (release 2011-03-05) to perform a standard error weighted meta-
241 analysis of our sequencing and GWAS cohorts with the publicly available International Inflammatory
242 Bowel Disease Genetics Consortium (IIBDGC) meta-analysis summary statistics[1], after applying the
243 additional MAF ≥0.1%, and INFO ≥0.4 filters to the IIBDGC data.

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

248 *LD score regression.* We performed LD score regression using LDSC v1.0.0 and European linkage
249 disequilibrium (LD) scores from the 1000 Genomes Project (downloaded from
250 https://data.broadinstitute.org/alkesgroup/LDSCORE/eur_w_ld_chr.tar.bz2) on our filtered meta-analysis
251 summary statistics for all sites with INFO > 0.95. This INFO threshold is to avoid confounding due to poor
252 imputation, as recommended by the authors[41].

253 **Locus definition**

254 *Computing LD windows.* An LD window was calculated for every genome-wide significant variant in any
255 of the three traits (Crohn's disease, ulcerative colitis, IBD), defined by the left-most and right-most
256 variants that are correlated with the main variant with an r^2 of 0.6 or more. The LD was calculated in the
257 GBR and CEU samples from the 1000 Genomes Phase 3, release v5 (based on 20130502 sequence
258 freeze and alignments). Loci with overlapping LD windows, as well as loci whose lead variants were
259 separated by 500kb or less, were subsequently merged, and the variant with the strongest evidence of
260 being associated was kept as the lead variant for each merged locus.

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

269 **Fine-mapping**

270 Approximate Bayes factors were calculated from the meta-analysis effect sizes and standard errors
271 described above by applying equation (2) of Wakefield[42], assuming a prior variance on the log odds ratios
272 of 0.04 (the default prior used by the software SNPTest, and used by Maller *et al*[43]). We then performed
273 fine-mapping using these Bayes factors as described in Maller et al to calculate the posterior that each
274 variant is causal, and the 95% credible set for each association (the smallest set of variants with
275 posteriors that sum to at least 95%). For each association we use the meta-analysis results for the
276 phenotype (Crohn's disease, ulcerative colitis or IBD) specified in Table 1. We only consider a locus to be
277 confidently fine-mapped if there are no variants in the Phase 3 v5 release of the 1000 Genomes project
278 (2013-05-02 sequence freeze) in high LD ($r^2 \geq 0.6$) with our hit SNP, but missing from our dataset, and no
279 variants in our data within high LD ($r^2 > 0.8$) that fail during our QC procedure.

280 **eQTL overlap**

281 *Identifying eQTL overlaps.* Nine eQTL datasets were searched to identify variants within the 26 newly
282 identified IBD risk loci that are associated with variation in gene expression (Supplementary Table 5).
283 Splice-QTLs based on exon-ratio[44] and transcript-ratio[45–47] were also included in the search where
284 available (Supplementary Table 5). The most significant variant-gene associations were extracted from
285 each eQTL/splice-QTL dataset and were reported as candidates if that variant had r2 > 0.8 with any of
286 the lead SNPs in the 26 IBD risk loci.

287 *Testing for co-localization*. We tested for co-localization between IBD association signals and eQTLs
288 using the coloc2 method[48], implemented in the R package coloc. We used a window size of 250kb on

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

289 either side of the IBD association, and implemented the default settings as recommended. Each test was
290 repeated using two different values for the prior probability of co-localization, p_{12}: $1x10^{-5}$ and $1x10^{-6}$.

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

291 **References**

292 1. Liu, J. Z. *et al.* Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight
293 shared genetic risk across populations. *Nat. Genet.* **47,** 979–989 (2015).
294 2. Parkes, M. *et al.* Sequence variants in the autophagy gene IRGM and multiple other replicating loci contribute to
295 Crohn's disease susceptibility. *Nat. Genet.* **39,** 830–832 (2007).
296 3. Yamazaki, K. *et al.* A Genome-Wide Association Study Identifies 2 Susceptibility Loci for Crohn's Disease in a
297 Japanese Population. *Gastroenterology* **144,** 781–788 (2013).
298 4. Anderson, C. A. *et al.* Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of
299 confirmed associations to 47. *Nat. Genet.* **43,** 246–252 (2011).
300 5. Kenny, E. E. *et al.* A genome-wide scan of Ashkenazi Jewish Crohn's disease suggests novel susceptibility loci.
301 *PLoS Genet.* **8,** (2012).
302 6. Julià, A. *et al.* A genome-wide association study identifies a novel locus at 6q22.1 associated with ulcerative
303 colitis. *Hum. Mol. Genet.* **23,** 6927–6934 (2014).
304 7. Yang, S.-K. *et al.* Genome-wide association study of Crohn's disease in Koreans revealed three new
305 susceptibility loci and common attributes of genetic susceptibility across ethnic populations. *Gut* **63,** 80–87
306 (2014).
307 8. Luo, Y. *et al.* Exploring the genetic architecture of inflammatory bowel disease by whole genome sequencing
308 identifies association at ADCY7. *Nature Genetics, submitted*
309 9. Walter, K. *et al.* The UK10K project identifies rare variants in health and disease. *Nature* **526,** 82–90 (2015).
310 10. 1000 Genomes Project Consortium *et al.* A global reference for human genetic variation. *Nature* **526,** 68–74
311 (2015).
312 11. Huang, H. *et al.* Association mapping of inflammatory bowel disease loci to single variant resolution. *bioRxiv*
313 028688 (2015). doi:10.1101/028688
314 12. Ivanov, I. I. *et al.* The Orphan Nuclear Receptor RORγt Directs the Differentiation Program of Proinflammatory
315 IL-17+ T Helper Cells. *Cell* **126,** 1121–1133 (2006).
316 13. Wang, G. *et al.* Cutting edge: Slamf8 is a negative regulator of Nox2 activity in macrophages. *J. Immunol.* **188,**
317 5829–5832 (2012).
318 14. Wang, G. *et al.* Migration of myeloid cells during inflammation is differentially regulated by the cell surface
319 receptors Slamf1 and Slamf8. *PLoS One* **10,** e0121968 (2015).
320 15. Luci, C. *et al.* Influence of the transcription factor RORgammat on the development of NKp46+ cell populations in
321 gut and skin. *Nat. Immunol.* **10,** 75–82 (2009).
322 16. Yang, Y. *et al.* Focused specificity of intestinal TH17 cells towards commensal bacterial antigens. *Nature* **510,**
323 152–156 (2014).

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

324 17. Sawa, S. *et al.* RORγt+ innate lymphoid cells regulate intestinal homeostasis by integrating negative signals from
325 the symbiotic microbiota. *Nat. Immunol.* **12,** 320–326 (2011).
326 18. Gevers, D. *et al.* The treatment-naive microbiome in new-onset Crohn's disease. *Cell Host Microbe* **15,** 382–392
327 (2014).
328 19. Withers, D. R. *et al.* Transient inhibition of ROR-[gamma]t therapeutically limits intestinal inflammation by
329 reducing TH17 cells and preserving group 3 innate lymphoid cells. *Nat. Med.* **22,** 319–323 (2016).
330 20. Fu, G., Chen, Y., Schuman, J., Wang, D. & Wen, R. Phospholipase Cγ2 plays a role in TCR signal transduction
331 and T cell selection. *J. Immunol.* **189,** 2326–2332 (2012).
332 21. Ombrello, M. J. *et al.* Cold urticaria, immunodeficiency, and autoimmunity related to PLCG2 deletions. *N. Engl. J.*
333 *Med.* **366,** 330–338 (2012).
334 22. Zhou, Q. *et al.* A hypermorphic missense mutation in PLCG2, encoding phospholipase Cγ2, causes a dominantly
335 inherited autoinflammatory disease with immunodeficiency. *Am. J. Hum. Genet.* **91,** 713–720 (2012).
336 23. Jostins, L. *et al.* Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease.
337 *Nature* **491,** 119–124 (2012).
338 24. Matute, J. D. *et al.* A new genetic subgroup of chronic granulomatous disease with autosomal recessive
339 mutations in p40 phox and selective defects in neutrophil NADPH oxidase activity. *Blood* **114,** 3309–3315
340 (2009).
341 25. Rioux, J. D. *et al.* Genome-wide association study identifies new susceptibility loci for Crohn disease and
342 implicates autophagy in disease pathogenesis. *Nat. Genet.* **39,** 596–604 (2007).
343 26. Roberts, R. L. *et al.* Confirmation of association of IRGM and NCF4 with ileal Crohn's disease in a population-
344 based cohort. *Genes Immun.* **9,** 561–565 (2008).
345 27. Hynes, R. O. Integrins: bidirectional, allosteric signaling machines. *Cell* **110,** 673–687 (2002).
346 28. Sandborn, W. J. *et al.* Vedolizumab as induction and maintenance therapy for Crohn's disease. *N. Engl. J. Med.*
347 **369,** 711–721 (2013).
348 29. Feagan, B. G. *et al.* Vedolizumab as induction and maintenance therapy for ulcerative colitis. *N. Engl. J. Med.*
349 **369,** 699–710 (2013).
350 30. Vermeire, S. *et al.* Etrolizumab as induction therapy for ulcerative colitis: a randomised, controlled, phase 2 trial.
351 *Lancet* **384,** 309–318 (2014).
352 31. Hosten, T. A., Zhao, K., Han, H. Q., Liu, G. & He, X. H. Alicaforsen: An Emerging Therapeutic Agent for
353 Ulcerative Colitis and Refractory Pouchitis. *Gastroenterol. Res. Pract.* **7,** 51–55 (2014).
354 32. James, D. G., Seo, D. H., Chen, J., Vemulapalli, C. & Stone, C. D. Efalizumab, a human monoclonal anti-CD11a
355 antibody, in the treatment of moderate to severe Crohn's disease: An open-label pilot study. *Dig. Dis. Sci.* **56,**
356 1806–1810 (2011).
357 33. Sandborn, W. J. *et al.* Natalizumab induction and maintenance therapy for Crohn's disease. *N. Engl. J. Med.*

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

358 **353,** 1912–1925 (2005).
359 34. Travis, M. A. & Sheppard, D. TGF-β activation and function in immunity. *Annu. Rev. Immunol.* **32,** 51–82 (2014).
360 35. Travis, M. A. *et al.* Loss of integrin alpha(v)beta8 on dendritic cells causes autoimmunity and colitis in mice.
361 *Nature* **449,** 361–365 (2007).
362 36. Worthington, J. J. *et al.* Integrin αvβ8-Mediated TGF-β Activation by Effector Regulatory T Cells Is Essential for
363 Suppression of T-Cell-Mediated Inflammation. *Immunity* **42,** 903–915 (2015).
364 37. Monteleone, G. *et al.* Mongersen, an oral SMAD7 antisense oligonucleotide, and Crohn's disease. *N. Engl. J.*
365 *Med.* **372,** 1104–1113 (2015).
366 38. Miceli-Richard, C. *et al.* CARD15 mutations in Blau syndrome. *Nat. Genet.* **29,** 19–20 (2001).
367 39. Shah, T. S. *et al.* optiCall: a robust genotype-calling algorithm for rare, low-frequency and common variants.
368 *Bioinformatics* **28,** 1598–1603 (2012).
369 40. Durbin, R. Efficient haplotype matching and storage using the positional Burrows-Wheeler transform (PBWT).
370 *Bioinformatics* **30,** 1266–1272 (2014).
371 41. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide
372 association studies. *Nat. Genet.* **47,** 291–295 (2015).
373 42. Wakefield, J. Bayes factors for genome-wide association studies: comparison with P-values. *Genet. Epidemiol.*
374 **33,** 79–86 (2009).
375 43. Wellcome Trust Case Control Consortium *et al.* Bayesian refinement of association signals for 14 loci in 3
376 common diseases. *Nat. Genet.* **44,** 1294–1301 (2012).
377 44. Zhernakova, D. *et al.* Hypothesis-free identification of modulators of genetic risk factors. *bioRxiv* 033217 (2015).
378 doi:10.1101/033217
379 45. Battle, A. *et al.* Characterizing the genetic basis of transcriptome diversity through RNA-sequencing of 922
380 individuals. *Genome Res.* **24,** 14–24 (2014).
381 46. Monlong, J., Calvo, M., Ferreira, P. G. & Guigó, R. Identification of genetic variants associated with alternative
382 splicing using sQTLseekeR. *Nat. Commun.* **5,** 4698 (2014).
383 47. Ongen, H., Buil, A., Brown, A. A., Dermitzakis, E. T. & Delaneau, O. Fast and efficient QTL mapper for
384 thousands of molecular phenotypes. *Bioinformatics* **32,** 1479–1485 (2016).
385 48. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using
386 summary statistics. *PLoS Genet.* **10,** e1004383 (2014).

387

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

388

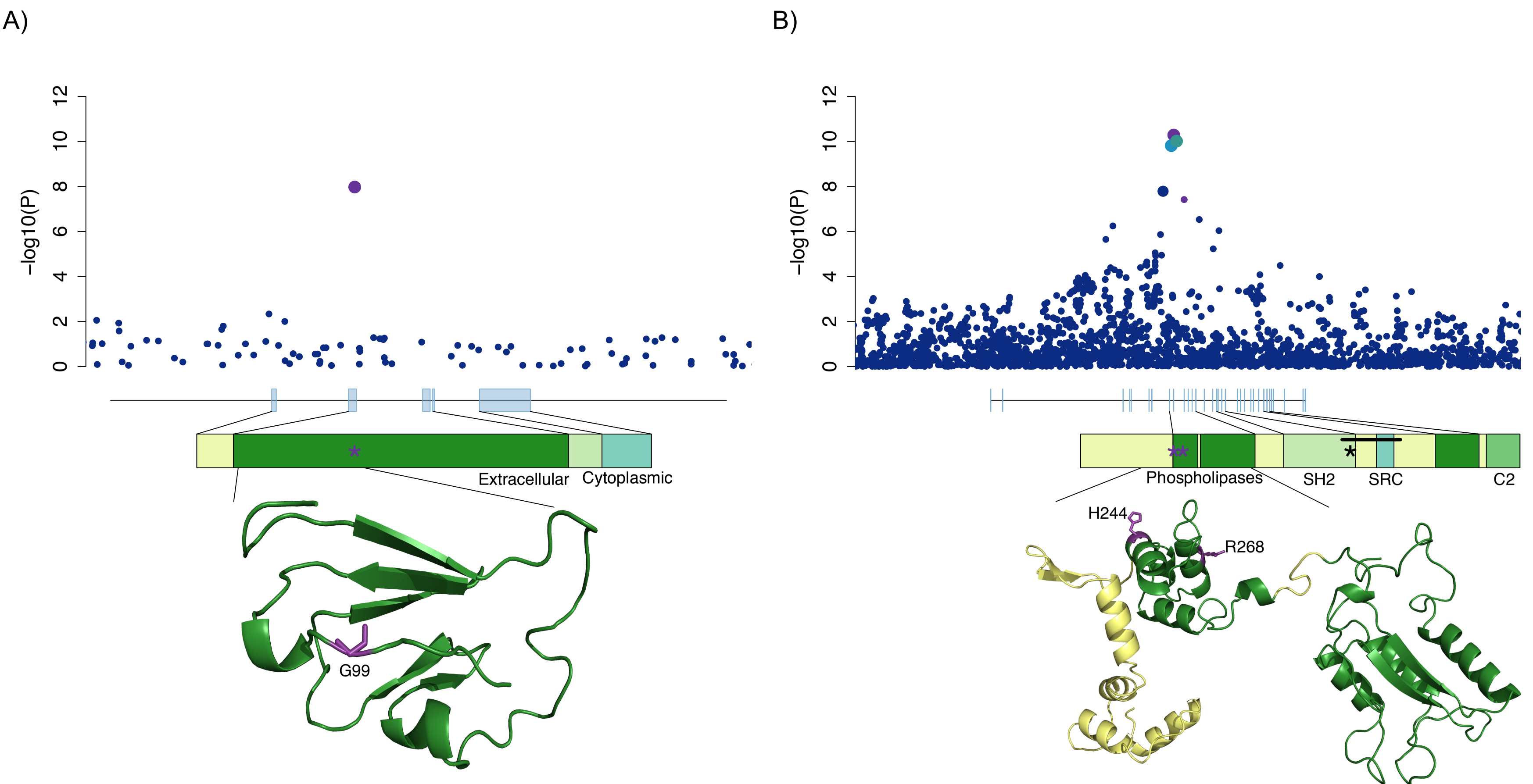
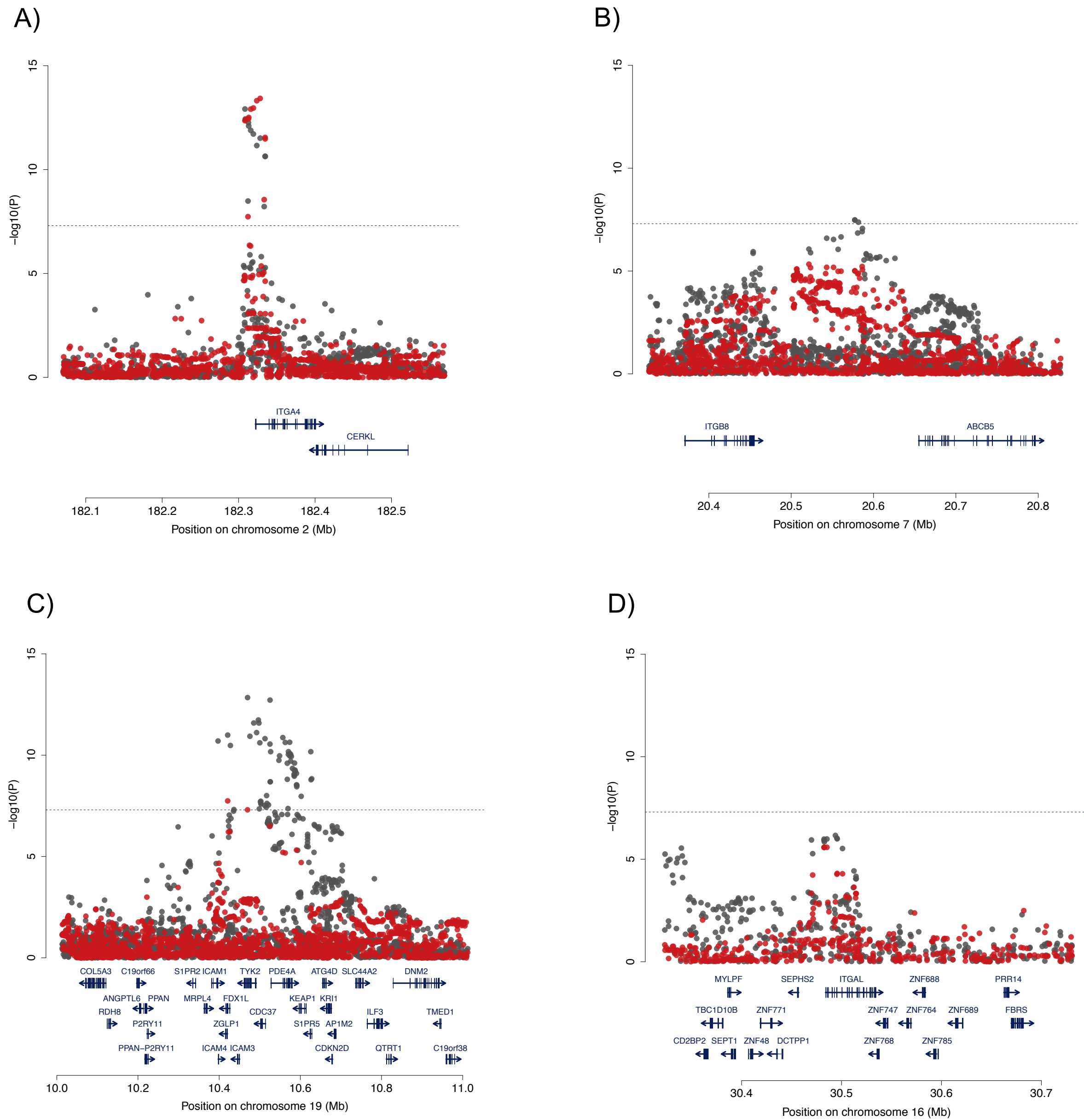


Figure 1. Likely causal missense variants. For A) SLAMF8 and B) PLCG2, local association results are plotted with point size corresponding to LD to our lead variant and color to fine-mapping probability (purple > 50%, intermediate blue 10-50%, navy blue <10%). Gene body diagrams and protein domain annotations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

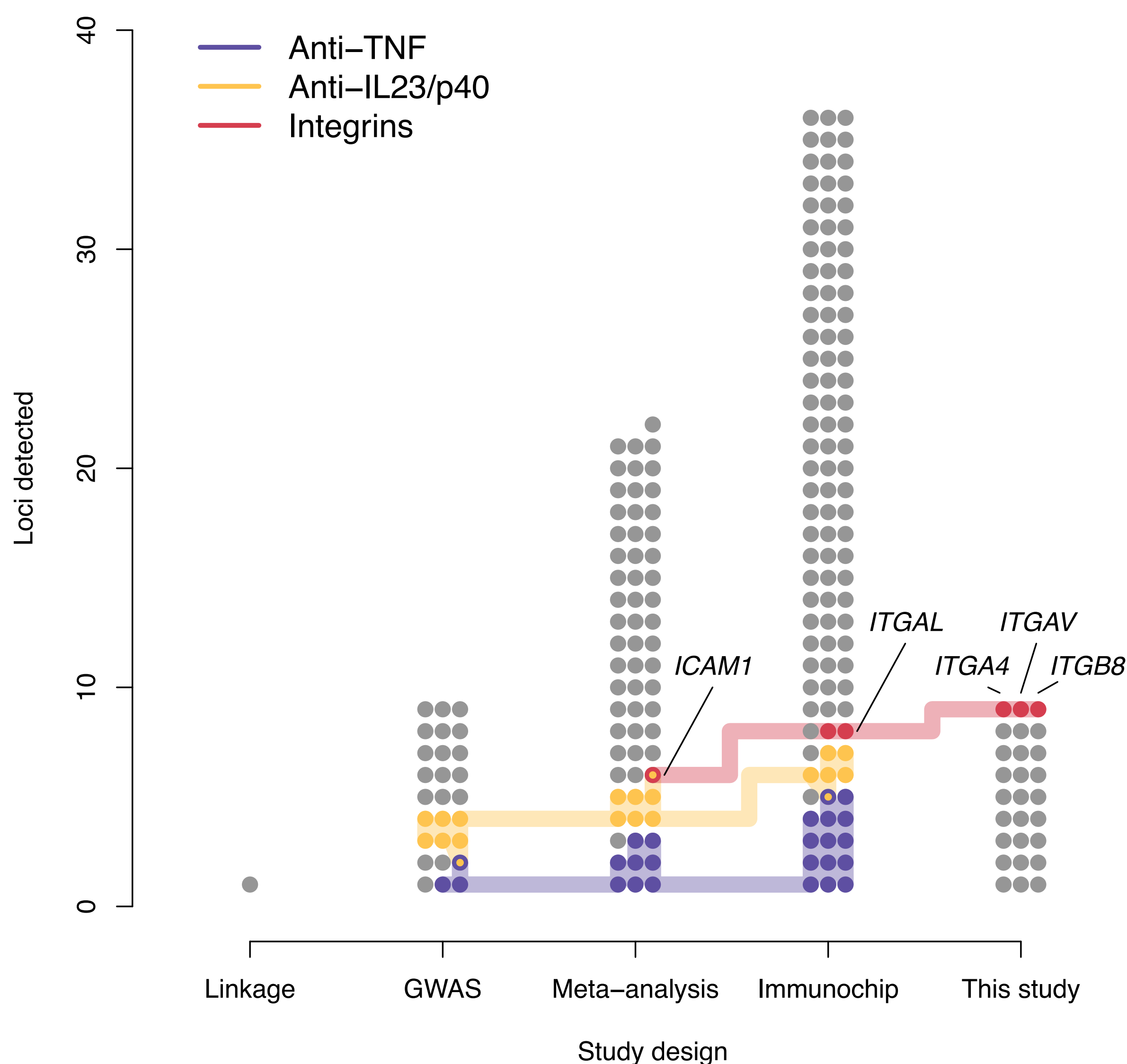
bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



389

Figure 2. Co-localization of disease association and stimulus response eQTLs in monocytes. The local pattern of disease association (IBD: (A) *ITGA4,* (B) *ITGB8*, (C) *ICAM1*; (D) UC: *ITGAL*) in grey, and the association of that variant with response to LPS (lipopolysaccharide) stimulation in red. Evidence of co-localization (probability > 70%) is observed for all for signals.

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.



390

Figure 3. IBD-associated loci containing genes related to known drug targets. All IBD loci are divided into the studies where they were first identified[1]. The direct targets and closely related genes of three classes of IBD therapeutics are highlighted, with particular focus on the newly discovered integrin associations. Despite the general pattern that effect size decreases from left to right, therapeutically relevant associations continue to be found.

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

391

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
rs34687326	1	159799910	159.8-159.8	G	A	0.9	1.06×10^{-08}	1.18 (1.12-1.24)	CD
rs59043219	1	209970610	209.97-210.02	A	G	0.379	1.09×10^{-08}	1.08 (1.05-1.10)	IBD
rs6740847	2	182308352	182.31-182.33	A	G	0.508	1.22×10^{-13}	1.10 (1.07-1.12)	IBD
rs144344067	2	187576378	187.5-187.68	A	AT	0.895	1.29×10^{-08}	1.12 (1.08-1.16)	IBD
rs1811711	2	228670476	228.67-228.67	C	G	0.826	6.09×10^{-09}	1.14 (1.1-1.18)	UC
rs76527535	2	242484701	242.47-242.49	C	T	0.745	2.87×10^{-08}	1.09 (1.06-1.12)	IBD
rs2581828	3	53133149	53.1-53.17	C	G	0.597	6.46×10^{-09}	1.10 (1.07-1.13)	CD
rs2593855	3	71175495	71.16-71.19	C	T	0.663	2.54×10^{-09}	1.09 (1.06-1.11)	IBD
rs503734	3	101023748	100.91-101.27	A	G	0.513	2.67×10^{-08}	1.07 (1.05-1.10)	IBD
rs56116661	3	188401160	188.4-188.4	C	T	0.795	5.67×10^{-10}	1.14 (1.10-1.18)	CD
rs11734570	4	38588453	38.58-38.59	A	G	0.368	4.80×10^{-08}	1.07 (1.05-1.10)	IBD
rs17656349	5	149605994	149.59-149.63	T	C	0.466	1.54×10^{-08}	1.09 (1.06-1.13)	UC
rs113986290	6	19781009	19.72-19.83	C	T	0.989	7.59×10^{-09}	1.36 (1.25-1.46)	UC
rs67289879	6	42007403	42-42.01	T	C	0.179	3.04×10^{-08}	1.09 (1.06-1.13)	IBD
rs11768365	7	6545188	6.5-6.55	A	G	0.816	3.88×10^{-08}	1.09 (1.06-1.12)	IBD
rs149169037	7	20577298	20.58-20.58	G	A	0.895	3.26×10^{-08}	1.14 (1.10-1.19)	IBD
rs243505	7	148435339	148.4-148.58	A	G	0.624	3.04×10^{-10}	1.08 (1.06-1.11)	IBD
rs7911117	10	27179596	27.16-27.18	T	G	0.871	1.84×10^{-08}	1.14 (1.10-1.19)	UC
rs111456533	10	126439381	126.32-126.55	G	A	0.829	1.18×10^{-09}	1.11 (1.08-1.14)	IBD
rs11221335	11	128385906	128.38-128.4	C	T	0.224	2.44×10^{-08}	1.09 (1.06-1.12)	IBD
rs80244186	13	42917861	42.84-42.94	C	T	0.111	3.66×10^{-08}	1.13 (1.09-1.18)	CD
rs11548656	16	81916912	81.91-81.92	A	G	0.961	5.18×10^{-11}	1.27 (1.20-1.34)	IBD
rs10492862	16	82867456	82.87-82.92	A	C	0.308	1.26×10^{-09}	1.11 (1.08-1.15)	CD
rs4256018	20	6093889	6.08-6.1	G	T	0.25	1.23×10^{-08}	1.08 (1.05-1.11)	IBD
rs138788	22	35729721	35.72-35.74	A	G	0.418	2.95×10^{-08}	1.09 (1.06-1.13)	UC
rs4821544	22	37258503	37.26-37.26	C	T	0.321	1.76×10^{-08}	1.10 (1.07-1.13)	CD

bioRxiv preprint doi: https://doi.org/10.1101/058255; this version posted June 10, 2016. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

392

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

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	1.06×10^{-08}	Novel
rs4845604	1	151801680	0.999	RORC (intronic)	1	IBD	7.09×10^{-14}	Known
rs1811711	2	228670476	0.914		2	UC	6.09×10^{-09}	Novel
rs56116661	3	188401160	0.561	LPP (intronic)	11	CD	5.67×10^{-10}	Novel
rs11548656	16	81916912	0.502	PLCG2 p.His244Arg (missense)	3	IBD	5.18×10^{-11}	Novel
rs1143687	16	81922813	0.746	PLCG2 p.Arg268Trp (missense)	5	IBD	3.83×10^{-08}	Novel
rs4821544	22	37258503	0.804	NCF4 (intronic)	2	CD	1.76×10^{-08}	Novel