SUPPLEMENTARY NOTE

Major histocompatibility complex (MHC)

UC association. For both EAS and EUR, the strongest IBD association in MHC is a UC-specific association, which is also the strongest UC association across the genome. The best UC-associated variant (rs6927022), from the IBD MHC fine-mapping study[10] (conducted in EUR), has consistent genetic effect and variance explained in EAS and EUR, with logOR of -0.399 ± 0.015 and -0.434 ± 0.026 (mean ± s.e.), and variance explained of 2.4% and 2.9% in EUR and EAS respectively. Note that the best UC-associated variant in EAS is a different variant in weak LD with rs6927022 (R2 = 0.2 in EAS), which has a greater genetic effect (-0.639±0.039) and explains more variance (3.9%). This variant is in strong LD with the best UC-associated variant in the latest NFE IBD GWAS in EUR[4] (rs9271176, R2 = 0.86 in EAS).

CD association. Previous studies have identified multiple independent HLA alleles contributing to the disease, with some having shared roles for CD and some showing non-additive effects[10]. While a full-scale comparative analysis for all the HLA alleles associated with IBD is beyond the scope of this study, we investigated the strongest CD association in MHC to look into the relative importance of CD vs UC in EAS and EUR. The strongest CD association from the IBD MHC fine-mapping study is an HLA allele not available in our EAS data. We, therefore, used the strongest CD-associated variant from the latest NFE IBD GWAS[4] in EUR. This variant (rs145568234) has logOR of 0.86 ± 0.06 and variance explained of 0.6% in EUR. It was, however, not analyzed in the EAS study as it is very rare (MAF < 0.001), suggesting that this variant likely explains very little phenotypic variance in EAS. In EAS, we found a different CD association in MHC (rs9270965) with logOR of 0.53 ± 0.03 (P = 1E-76) and variance explained of 3.5%, which had consistent effects across all four EAS sample collections. This variant is not present in the latest NFE IBD GWAS despite being a common variant (EUR MAF = 0.26). The FIN data in our study reported a significantly reduced logOR of 0.18 ± 0.04 and variance explained of 0.5%. The two CD-associated variants (rs145568234 and rs9270965) are not in LD with each other (R2 < 0.03 in either EUR or EAS).

Summary. We found, in MHC, the primary UC association appears to be consistent across EAS and EUR with similar OR and explains similar phenotypic variances. The primary CD association, instead, appears to be heterogeneous with the best CD-associated variant in EUR not present in EAS, and the best CD-associated variant in EAS has a significantly smaller genetic effect in EUR. While the primary UC association is UC specific for both EUR and EAS (Extended Data Figure 9), when combined with the additional CD association, the MHC, in aggregation, contributes to CD and UC in a more balanced manner in EAS than in EUR: the variance explained are 2.4% vs 0.6% for UC and CD in EUR, and 3.9% vs 3.5% in EAS.

This interesting heterogeneity in MHC across EAS and EUR, and the heterogeneity within EAS (Extended Data Figure 6), can be driven by factors such as the long-range LD and gene-gene interactions. MHC associations may also be involved in the adaptive immune response to the colonic environment, which can be an interesting topic to investigate for genetic-environment interactions across populations[10]. A study to delineate the reasons underlying this heterogeneity requires a large-

scale MHC imputation panel, to begin with, which only exists in EUR until very recently. We note that a recent publication[62] made available an HLA reference panel including over 2,000 EAS subjects, among which the majority is from Japan. Given the significant heterogeneity across the three EAS ancestries in the MHC[13,44], it is unclear to what extent this panel will be helpful, which warrants a full investigation in the future.

IBD genes of interest

Tier 1 (*IL23R* not discussed as it has been well studied)

MFSD12 (Major Facilitator Superfamily Domain Containing 12) is a risk gene of skin pigmentation and inflammation bowel disease identified in African and Korean populations[7,63], respectively. It plays a pivotal role in maintaining normal levels of cystine and the oxidized dimer of cysteine in melanosomes. Studies revealed MFSD12 as a key promoter of cell proliferation and a potential therapeutic target in melanoma[64]. Meanwhile, MFSD12 is a necessary component of the cysteine importer for lysosomes[65], which is critical in maintaining the intestinal epithelial barrier integrity[66]. MFSD12 inhibits the proliferation of SH-SY5Y cells, arrests the cell cycle at G1 phase, and sensitizes the SH-SY5Y cells to sodium butyrate treatment[67].

SHC1 (SHC adaptor protein 1) encodes three main isoforms that differ in activities and subcellular location. SHC1 is an important regulator of proliferation and tumorigenesis through transcriptional activation of downstream signal cascades[68]. The balance of pro-inflammatory versus anti-inflammatory cytokines generated by dendritic cells upon Toll-like receptors stimulation could be strictly controlled by SHC1, which may act as a switch in response to immunity[69]. Genetic deletion of p66shc-/- increases susceptibility to myocardial injury in response to short-term ischaemia and reperfusion in mice[70].

ZBTB46 (zinc finger and BTB domain containing 46) encodes zinc finger and BTB domain-containing protein that plays a major role in the function of conventional dendritic cells in immunity, inflammation, and tolerance[71,72]. ZBTB46+ group 3 innate lymphoid cells (ILC3) have a vital and non-redundant role in controlling the intestinal inflammation induced by bacterial infection. And targeting the ZBTB46-leukemia inhibitory factor axis may inhibit castration-resistant prostate cancer development and neuroendocrine differentiation after androgen deprivation therapy[73]. Variants in *ZBTB46* are associated with post-thrombolytic parenchymal haematoma and multiple sclerosis[74,75].

ADO (2-aminoethanethiol dioxygenase) encodes cysteamine dioxygenase and is involved in the biosynthesis of taurine[76]. As a low oxygen affinity amino-terminal cysteine dioxygenase, ADO transduces the oxygen-regulated stability of proteins by the N-degron pathway in human cells[77]. *ADO* is associated with leprosy, Ewing sarcoma and psoriasis[78–80].

Tier 2 (*FUT2* not discussed as it has been well studied)

GPR35 (G protein-coupled receptor 35) encodes a rhodopsin-like, 7-transmembrane class A GPCR. Activated by the metabolites tryptophan-derived kynurenic acid (KYNA), the chemokine CXCL17, and

phospholipid derivate lysophosphatidic acid (LPA) species, GPR35 triggers signals that are important for host–microbiome interactions, the host cell function, survival, proliferation, and expansion[81,82]. GPR35 is highly expressed in the small intestine, colon, spleen, and immune cells, including monocytes, macrophages, and dendritic cells[83]. In GWAS, GPR35 has been found to be associated with increased risk of immune-related disorders[84–86].

GTF2I (General Transcription Factor II I) encodes a phosphoprotein containing six characteristic repeat motifs. The encoded protein binds to the initiator element and E-box element in promoters and functions as a regulator of transcription. This locus is deleted in Williams-Beuren syndrome[87], and it is a risk gene of primary Sjögren's syndrome[88,89], thymic epithelial tumors and systemic lupus erythematosus[90]. Genomic and proteomic analysis of GTF2I reveals different interaction patterns of GTF2I at active, inducible, or repressed gene, and implicates GTF2I in the regulation of transcription elongation. Furthermore, GTF2I mediates transcriptional repression of GLI2 through binding the INR, which is shown to be applicable to a subset of tumor supportive genes and identified GTF2I as an important factor in the regulation of RNAPII pausing[91].

Tier 3 (*GPR35* and *FUT2* not discussed as they have been well studied)

ADAP1 (Arf GTPase-Activating Protein with Dual PH Domains 1) enables GTPase activator activity and is highly expressed in brain and colon. *ADAP1* is regulated by a known IBD gene *NKX2-3*[92]*.* Meta-analysis of patients with CD or leprosy has identified *ADAP1* as a shared lead risk gene[93]. ADAP1 promotes latent HIV-1 reactivation by selectively tuning KRAS–ERK–AP-1 T cell signaling transcriptional axis[94]. ADAP1 is also decreased by Salmonella enterica serovar typhimurium to invade nonphagocytic host cells, showing an obvious interplay between Arf guanine nucleotide reaction factors and GAPs that is decreased by Salmonella to establish infection[95].

GIT2 (GIT ArfGAP 2) encodes a member of the GIT protein family, which interacts with G protein-coupled receptor kinases and possesses ADP-ribosylation factor GTPase-activating protein activity. GIT2 is an essential terminator of TLR signaling, the loss of which leads to uncontrolled inflammation[96]. GIT2 may also act as one of these network coordinators in controlling numerous aspects of the complex aging process[97]. It has shown that GIT2 is necessary for directional chemotaxis and for the suppression of superoxide production in G protein–coupled receptor–stimulated neutrophils[98]. Git2-/- mice are more susceptible to dextran sodium sulfate induced colitis, endotoxin-shock challenge, or Escherichia coli, and likely by the TLR signaling[96].

CELA3B (Chymotrypsin Like Elastase 3B) is one of the six human elastase genes that encode structurally similar proteins. Unlike other elastases, CELA3B possesses little elastolytic activity. A rare missense mutation in the gene encoding CELA3B is cosegregated with familial pancreatitis with diabetes and pancreatic adenocarcinoma[99]. *CELA3B* is a risk gene for chronic pancreatitis and the loss-of-function of *CELA3B* is associated with a higher potential of chronic pancreatitis[100,101].

Cross-ancestry meta-analysis

IL21R (interleukin 21 receptor) belongs to the type I cytokine receptors and has been shown to form a heterodimeric receptor complex with the common gamma-chain. *IL21R* is mainly expressed in T cells, B cells, and natural killer cells. Autonomous IL21R-dependent signaling by CD8+ T cells is required for sustained cell proliferation and cytokine production during chronic infection[102]. Il21r deletion in transgenic EAE mouse model (T and B cells overexpressing receptors for myelin oligodendrocyte glycoprotein) reduces the incidence and severity of spontaneous EAE, which is associated with the defect in Th17 cell generation[103]. IL21/IL21R signaling contributes to protection against DSS-induced acute colitis through suppression of Th1 and activation of Th2, Th17 and Treg responses in mice[104]. Neutralization of IL-21 in experimental CD4 T cell-driven colitis is associated with reduction in clinical and pathological findings[105]. *IL21R* has higher expression in the inflamed gut of CD patients compared with controls[105–107] and has been reported to be a risk locus for primary biliary cholangitis, systemic lupus erythematosus, multiple sclerosis and ischemic stroke[108–111].

RUNX3 (Runt-related transcription factor 3) is one of key regulators of human autoimmunity[112]. RUNX3 is predominantly expressed in hematopoietic lineages and regulates several aspects of immune function, including T-cell differentiation, dendritic cell (DC) maturation and natural killer (NK) cell activation[113]. Dysregulation of RUNX3 links to hematopoietic and gastrointestinal pathologies, including leukemia, spontaneous colitis, and gastric carcinogenesis[114,115]. In Runx3 knockout mice, loss of leukocytic cell-autonomous function of Runx3 develops IBD and gastric lesion[115]. In cancer studies, RUNX3 is a tumor suppressor in gastric and colorectal cancer[116,117]. RUNX3 inhibits epithelial-to-mesenchymal transition(EMT), which promotes metastasis, by regulating the Wnt signaling pathway[118,119].

ABO (alpha 1-3-N-acetylgalactosaminyltransferase and alpha 1-3-galactosyltransferase) variation is the basis of the ABO blood group. This gene encodes a protein related to the first discovered blood group system and is highly expressed in colon and small intestine. Genetic variation in this gene is associated with duodenal ulcer, epithelial ovarian cancer, pancreatic cancer, Behçet's disease and depression[79,120–123]. *ABO* is also a genetic susceptibility locus in patients with Covid-19 with respiratory failure[124,125]. Moreover, *ABO* also plays a vital role in gut microbiome, indicating the potential for implementing microbiota in disease treatment and management in European population[126–129].

HORMAD1 (HORMA domain containing 1) encodes a HORMA domain-containing protein. HORMA domains are involved in chromatin binding and play a role in cell cycle regulation[130,131]. SKP1-Cullin-F-box ubiquitin E3 ligase regulates DNA double-strand breaks through proteasome-mediated degradation of HORMAD1 in early meiotic recombination[132]. Overexpression of *HORMAD1* contributes to homologous recombination deficiency in triple-negative breast cancers[133].

CTSS (cathepsin S) encodes a lysosomal cysteine proteinase that participates in the degradation of antigenic proteins to peptides for presentation on MHC class II molecules. CTSS is highly expressed in inflamed areas of osteoarthritis synovial membrane[134]. Besides, overexpression of *CTSS* induces chronic atopic dermatitis in mice[135]. CTSS alterations induce a tumor-promoting immune microenvironment in follicular lymphoma[136,137]. Symbionts of the intestinal microbiota regulate host *CTSS* activity and prevent T-cell mediated induction of colonic inflammation[138]. *CTSS* is a risk locus of obesity and congenital cardiovascular left-sided lesions[139,140].

REFERENCES

62. Luo, Y. *et al.* A high-resolution HLA reference panel capturing global population diversity enables multi-ancestry fine-mapping in HIV host response. *Nat. Genet.* **53**, 1504–1516 (2021).
63. Crawford, N. G. *et al.* Loci associated with skin pigmentation identified in African populations. *Science* **358**, (2017).
64. Wei, C.-Y. *et al.* Bioinformatics-based analysis reveals elevated MFSD12 as a key promoter of cell proliferation and a potential therapeutic target in melanoma. *Oncogene* **38**, 1876–1891 (2019).
65. Adelmann, C. H. *et al.* MFSD12 mediates the import of cysteine into melanosomes and lysosomes. *Nature* **588**, 699–704 (2020).
66. Lassen, K. G. & Xavier, R. J. Mechanisms and function of autophagy in intestinal disease. *Autophagy* **14**, 216–220 (2018).
67. Wang, Y., Ma, C., Zhang, H. & Wu, J. Novel protein pp3501 mediates the inhibitory effect of sodium butyrate on SH-SY5Y cell proliferation. *J. Cell. Biochem.* **113**, 2696–2703 (2012).
68. Lai, C.-H. *et al.* DEPDC1B is a tumor promotor in development of bladder cancer through targeting SHC1. *Cell Death Dis.* **11**, 986 (2020).
69. Chen, K.-D. *et al.* Adaptor protein Shc acts as an immune-regulator for the LPS-stimulated maturation of bone marrow-derived dendritic cells. *BMC Immunol.* **12**, 32 (2011).
70. Akhmedov, A. *et al.* Genetic deletion of the adaptor protein p66Shc increases susceptibility to short-term ischaemic myocardial injury via intracellular salvage pathways. *Eur. Heart J.* **36**, 516–26a (2015).
71. Satpathy, A. T. *et al.* Notch2-dependent classical dendritic cells orchestrate intestinal immunity to attaching-and-effacing bacterial pathogens. *Nat. Immunol.* **14**, 937–948 (2013).
72. Cabeza-Cabrerizo, M., Cardoso, A., Minutti, C. M., Pereira da Costa, M. & Reis e Sousa, C. Dendritic Cells Revisited. *Annu. Rev. Immunol.* **39**, 131–166 (2021).
73. Liu, Y.-N. *et al.* Leukemia Inhibitory Factor Promotes Castration-resistant Prostate Cancer and Neuroendocrine Differentiation by Activated ZBTB46. *Clin. Cancer Res.* **25**, 4128–4140 (2019).

74. Carrera, C. *et al.* Single nucleotide variations in ZBTB46 are associated with post-thrombolytic parenchymal haematoma. *Brain* **144**, 2416–2426 (2021).
75. International Multiple Sclerosis Genetics Consortium *et al.* MANBA, CXCR5, SOX8, RPS6KB1 and ZBTB46 are genetic risk loci for multiple sclerosis. *Brain* **136**, 1778–1782 (2013).
76. Dominy, J. E., Jr *et al.* Discovery and characterization of a second mammalian thiol dioxygenase, cysteamine dioxygenase. *J. Biol. Chem.* **282**, 25189–25198 (2007).
77. Masson, N. *et al.* Conserved N-terminal cysteine dioxygenases transduce responses to hypoxia in animals and plants. *Science* **365**, 65–69 (2019).
78. Sahlén, P. *et al.* Chromatin interactions in differentiating keratinocytes reveal novel atopic dermatitis- and psoriasis-associated genes. *J. Allergy Clin. Immunol.* **147**, 1742–1752 (2021).
79. Takeuchi, M. *et al.* Dense genotyping of immune-related loci implicates host responses to microbial exposure in Behçet's disease susceptibility. *Nat. Genet.* **49**, 438–443 (2017).
80. Hou, S. *et al.* Genome-wide association analysis of Vogt-Koyanagi-Harada syndrome identifies two new susceptibility loci at 1p31.2 and 10q21.3. *Nat. Genet.* **46**, 1007–1011 (2014).
81. Kaya, B., Melhem, H. & Niess, J. H. GPR35 in Intestinal Diseases: From Risk Gene to Function. *Front. Immunol.* **12**, 717392 (2021).
82. Boleij, A. *et al.* G-protein coupled receptor 35 (GPR35) regulates the colonic epithelial cell response to enterotoxigenic Bacteroides fragilis. *Commun Biol* **4**, 585 (2021).
83. O'Dowd, B. F. *et al.* Discovery of three novel G-protein-coupled receptor genes. *Genomics* **47**, 310–313 (1998).
84. Mackenzie, A. E., Lappin, J. E., Taylor, D. L., Nicklin, S. A. & Milligan, G. GPR35 as a Novel Therapeutic Target. *Front. Endocrinol.* **2**, 68 (2011).
85. Imielinski, M. *et al.* Common variants at five new loci associated with early-onset inflammatory bowel disease. *Nat. Genet.* **41**, 1335–1340 (2009).
86. Horikawa, Y. *et al.* Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat. Genet.* **26**, 163–175 (2000).

87. Kozel, B. A. *et al.* Williams syndrome. *Nat Rev Dis Primers* **7**, 42 (2021).
88. Li, Y. *et al.* A genome-wide association study in Han Chinese identifies a susceptibility locus for primary Sjögren's syndrome at 7q11.23. *Nat. Genet.* **45**, 1361–1365 (2013).
89. Song, I.-W. *et al.* Identification of susceptibility gene associated with female primary Sjögren's syndrome in Han Chinese by genome-wide association study. *Hum. Genet.* **135**, 1287–1294 (2016).
90. Sun, C. *et al.* High-density genotyping of immune-related loci identifies new SLE risk variants in individuals with Asian ancestry. *Nat. Genet.* **48**, 323–330 (2016).
91. Fan, A. X. *et al.* Genomic and proteomic analysis of transcription factor TFII-I reveals insight into the response to cellular stress. *Nucleic Acids Res.* **42**, 7625–7641 (2014).
92. Yu, W. *et al.* NKX2-3 transcriptional regulation of endothelin-1 and VEGF signaling in human intestinal microvascular endothelial cells. *PLoS One* **6**, e20454 (2011).
93. Jung, S. *et al.* Identification of shared loci associated with both Crohn's disease and leprosy in East Asians. *Hum. Mol. Genet.* **31**, 3934–3944 (2022).
94. Ramirez, N.-G. P. *et al.* ADAP1 promotes latent HIV-1 reactivation by selectively tuning KRAS-ERK-AP-1 T cell signaling-transcriptional axis. *Nat. Commun.* **13**, 1109 (2022).
95. Davidson, A. C., Humphreys, D., Brooks, A. B. E., Hume, P. J. & Koronakis, V. The Arf GTPase-activating protein family is exploited by Salmonella enterica serovar Typhimurium to invade nonphagocytic host cells. *MBio* **6**, (2015).
96. Wei, J. *et al.* The GTPase-activating protein GIT2 protects against colitis by negatively regulating Toll-like receptor signaling. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 8883–8888 (2014).
97. van Gastel, J. *et al.* GIT2-A keystone in ageing and age-related disease. *Ageing Res. Rev.* **43**, 46–63 (2018).
98. Mazaki, Y. *et al.* Neutrophil direction sensing and superoxide production linked by the GTPase-activating protein GIT2. *Nat. Immunol.* **7**, 724–731 (2006).
99. Moore, P. C. *et al.* Elastase 3B mutation links to familial pancreatitis with diabetes and pancreatic

adenocarcinoma. *J. Clin. Invest.* **129**, 4676–4681 (2019).

100. Tóth, A. *et al.* Loss-of-function variant in chymotrypsin like elastase 3B (CELA3B) is associated with non-alcoholic chronic pancreatitis. *Pancreatology* **22**, 713–718 (2022).

101. Zhong, J. *et al.* A Transcriptome-Wide Association Study Identifies Novel Candidate Susceptibility Genes for Pancreatic Cancer. *J. Natl. Cancer Inst.* **112**, 1003–1012 (2020).

102. Fröhlich, A. *et al.* IL-21R on T cells is critical for sustained functionality and control of chronic viral infection. *Science* **324**, 1576–1580 (2009).

103. Lee, Y. *et al.* IL-21R signaling is critical for induction of spontaneous experimental autoimmune encephalomyelitis. *J. Clin. Invest.* **125**, 4011–4020 (2015).

104. Wang, Y. *et al.* IL-21/IL-21R signaling suppresses intestinal inflammation induced by DSS through regulation of Th responses in lamina propria in mice. *Sci. Rep.* **6**, 31881 (2016).

105. Holm, T. L. *et al.* Evaluating IL-21 as a Potential Therapeutic Target in Crohn's Disease. *Gastroenterol. Res. Pract.* **2018**, 5962624 (2018).

106. Monteleone, G. *et al.* Interleukin-21 enhances T-helper cell type I signaling and interferon-gamma production in Crohn's disease. *Gastroenterology* **128**, 687–694 (2005).

107. Liu, Z. *et al.* Il-21 enhances NK cell activation and cytolytic activity and induces Th17 cell differentiation in inflammatory bowel disease. *Inflamm. Bowel Dis.* **15**, 1133–1144 (2009).

108. Webb, R. *et al.* A polymorphism within IL21R confers risk for systemic lupus erythematosus. *Arthritis Rheum.* **60**, 2402–2407 (2009).

109. Qiu, F. *et al.* A genome-wide association study identifies six novel risk loci for primary biliary cholangitis. *Nat. Commun.* **8**, 14828 (2017).

110. Nohra, R. *et al.* RGMA and IL21R show association with experimental inflammation and multiple sclerosis. *Genes Immun.* **11**, 279–293 (2010).

111. Lee, H. K. *et al.* Natural allelic variation of the IL-21 receptor modulates ischemic stroke infarct volume. *J. Clin. Invest.* **126**, 2827–2838 (2016).

112. Alarcón-Riquelme, M. E. Role of RUNX in autoimmune diseases linking rheumatoid arthritis,

psoriasis and lupus. *Arthritis Res. Ther.* **6**, 169–173 (2004).

113. Kurklu, B. *et al.* Lineage-specific RUNX3 hypomethylation marks the preneoplastic immune component of gastric cancer. *Oncogene* **34**, 2856–2866 (2015).

114. Levanon, D., Brenner, O., Otto, F. & Groner, Y. Runx3 knockouts and stomach cancer. *EMBO Rep.* **4**, 560–564 (2003).

115. Brenner, O. *et al.* Loss of Runx3 function in leukocytes is associated with spontaneously developed colitis and gastric mucosal hyperplasia. *Proc. Natl. Acad. Sci. U. S. A.* **101**, 16016–16021 (2004).

116. Ito, K. *et al.* RUNX3, a novel tumor suppressor, is frequently inactivated in gastric cancer by protein mislocalization. *Cancer Res.* **65**, 7743–7750 (2005).

117. Kim, B. R. *et al.* RUNX3 inhibits the metastasis and angiogenesis of colorectal cancer. *Oncol. Rep.* **36**, 2601–2608 (2016).

118. Chen, F., Liu, X., Bai, J., Pei, D. & Zheng, J. The emerging role of RUNX3 in cancer metastasis (Review). *Oncol. Rep.* **35**, 1227–1236 (2016).

119. Voon, D. C.-C. *et al.* Runx3 protects gastric epithelial cells against epithelial-mesenchymal transition-induced cellular plasticity and tumorigenicity. *Stem Cells* **30**, 2088–2099 (2012).

120. Tanikawa, C. *et al.* A genome-wide association study identifies two susceptibility loci for duodenal ulcer in the Japanese population. *Nat. Genet.* **44**, 430–4, S1-2 (2012).

121. Kuchenbaecker, K. B. *et al.* Identification of six new susceptibility loci for invasive epithelial ovarian cancer. *Nat. Genet.* **47**, 164–171 (2015).

122. Childs, E. J. *et al.* Common variation at 2p13.3, 3q29, 7p13 and 17q25.1 associated with susceptibility to pancreatic cancer. *Nat. Genet.* **47**, 911–916 (2015).

123. Wu, Y. *et al.* GWAS of peptic ulcer disease implicates Helicobacter pylori infection, other gastrointestinal disorders and depression. *Nat. Commun.* **12**, 1146 (2021).

124. Severe Covid-19 GWAS Group *et al.* Genomewide Association Study of Severe Covid-19 with Respiratory Failure. *N. Engl. J. Med.* **383**, 1522–1534 (2020).

125. Ma, Y. *et al.* Integrating single-cell sequencing data with GWAS summary statistics reveals CD16+monocytes and memory CD8+T cells involved in severe COVID-19. *Genome Med.* **14**, 16 (2022).

126. Rühlemann, M. C. *et al.* Genome-wide association study in 8,956 German individuals identifies influence of ABO histo-blood groups on gut microbiome. *Nat. Genet.* **53**, 147–155 (2021).

127. Sanna, S., Kurilshikov, A., van der Graaf, A., Fu, J. & Zhernakova, A. Challenges and future directions for studying effects of host genetics on the gut microbiome. *Nat. Genet.* **54**, 100–106 (2022).

128. Qin, Y. *et al.* Combined effects of host genetics and diet on human gut microbiota and incident disease in a single population cohort. *Nat. Genet.* **54**, 134–142 (2022).

129. Lopera-Maya, E. A. *et al.* Effect of host genetics on the gut microbiome in 7,738 participants of the Dutch Microbiome Project. *Nat. Genet.* **54**, 143–151 (2022).

130. Royo, H. *et al.* ATR acts stage specifically to regulate multiple aspects of mammalian meiotic silencing. *Genes Dev.* **27**, 1484–1494 (2013).

131. Acquaviva, L. *et al.* Ensuring meiotic DNA break formation in the mouse pseudoautosomal region. *Nature* **582**, 426–431 (2020).

132. Guan, Y. *et al.* SCF ubiquitin E3 ligase regulates DNA double-strand breaks in early meiotic recombination. *Nucleic Acids Res.* **50**, 5129–5144 (2022).

133. Watkins, J. *et al.* Genomic Complexity Profiling Reveals That HORMAD1 Overexpression Contributes to Homologous Recombination Deficiency in Triple-Negative Breast Cancers. *Cancer Discov.* **5**, 488–505 (2015).

134. Lambert, C. *et al.* Gene expression pattern of cells from inflamed and normal areas of osteoarthritis synovial membrane. *Arthritis Rheumatol* **66**, 960–968 (2014).

135. Kim, N. *et al.* Overexpression of cathepsin S induces chronic atopic dermatitis in mice. *J. Invest. Dermatol.* **132**, 1169–1176 (2012).

136. Bararia, D. *et al.* Cathepsin S Alterations Induce a Tumor-Promoting Immune Microenvironment in

Follicular Lymphoma. *Cell Rep.* **31**, 107522 (2020).

137. Dheilly, E. *et al.* Cathepsin S Regulates Antigen Processing and T Cell Activity in Non-Hodgkin Lymphoma. *Cancer Cell* **37**, 674-689.e12 (2020).

138. Steimle, A. *et al.* Symbiotic gut commensal bacteria act as host cathepsin S activity regulators. *J. Autoimmun.* **75**, 82–95 (2016).

139. Pei, Y.-F. *et al.* Meta-analysis of genome-wide association data identifies novel susceptibility loci for obesity. *Hum. Mol. Genet.* **23**, 820–830 (2014).

140. Hanchard, N. A. *et al.* A genome-wide association study of congenital cardiovascular left-sided lesions shows association with a locus on chromosome 20. *Hum. Mol. Genet.* **25**, 2331–2341 (2016).