An integrated taxonomy for monogenic inflammatory bowel disease

Chrissy Bolton[1,2,*], Christopher S. Smillie[3, 4,*], Sumeet Pandey[1*], Rasa Elmentaite[5], Gabrielle Wei[7], Carmen Argmann[7], Dominik Aschenbrenner[1], Kylie R James[5, 6], Dermot P.B McGovern[8], Marina Macchi[1], Judy Cho[7], Dror S. Shouval[9], Jochen Kammermeier[10], Sibylle Koletzko[11,12], Krithika Bagalopal[1], Melania Capitani[1], Athena Cavounidis[1], Elisabete Pires[13], Carl Weidinger[14], James McCullagh[13], Peter D. Arkwright[15-16], Wolfram Haller[16], Britta Siegmund[14], Lauren Peters[6], Luke Jostins[17], Simon P.L. Travis[1,18], Carl A. Anderson[5], Scott Snapper[19], Christoph Klein[11], Eric Schadt[7], Matthias Zilbauer[20-21,], Ramnik Xavier[3], Sarah Teichmann[5,22-23], Aleixo M. Muise[24, 25, 26], Aviv Regev[3, 27, 28], Holm H. Uhlig[1, 17, 29]

Institutions:

1. Translational Gastroenterology Unit, University of Oxford, Oxford, UK.
2. Institute of Child Health, University College London, London, UK.
3. Klarman Cell Observatory, Broad Institute of *MIT* and Harvard, Cambridge, United States.
4. Center for Computational and Integrative Biology, Massachusetts General Hospital, Boston, United States
5. Wellcome Sanger Institute, Wellcome Genome Campus, Hinxton, UK.
6. Garvan Institute of Medical Research, The Kinghorn Cancer Centre, Darlinghurst, Australia
7. Department of Genetics and Genomic Sciences, Icahn School of Medicine at Mount Sinai, New York, NY, USA.
8. F. Widjaja Foundation, Inflammatory Bowel and Immunobiology Research Institute, Cedars-Sinai Medical Center, Los Angeles, CA, USA.
9. Institute of Gastroenterology, Nutrition and Liver Diseases, Schneider Children's Medical Center of Israel, Petah-Tiqva, Israel, affiliated to Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.
10. Gastroenterology Department, *Evelina* London Children's Hospital, London, UK.

11. Dr. von Hauner Children’s Hospital, Department of Pediatrics, University Hospital, LMU Munich, Munich, Germany.
12. Department of Pediatrics, Gastroenterology and Nutrition, School of Medicine Collegium Medicum University of Warmia and Mazury, Olsztyn, Poland.
13. Chemistry Research Laboratory, Department of Chemistry, University of Oxford, Mansfield Road, Oxford OX1 3TA, UK
14. Charité – Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin and Berlin Institute of Health Department of Gastroenterology, Rheumatology and Infectious Disease, Campus Benjamin Franklin, Berlin, Germany
15. Lydia Becker Institute of Immunology and Inflammation, University of Manchester, UK
16. Department of Gastroenterology and Nutrition, Birmingham Children’s Hospital, Birmingham, United Kingdom
17. The Kennedy Institute of Rheumatology, University of Oxford, Oxford, UK.
18. Biomedical Research Center, University of Oxford, Oxford, United Kingdom.
19. Division of Gastroenterology and Nutrition, Boston Children's Hospital, Boston, USA.
20. Department of Paediatric Gastroenterology, Hepatology and Nutrition, Addenbrooke's Hospital, Cambridge, UK.
21. Department of Paediatrics, University of Cambridge, Cambridge, UK.
22. Theory of Condensed Matter, Cavendish Laboratory, Department of Physics, University of Cambridge, Cambridge UK
23. European Molecular Biology Laboratory, European Bioinformatics Institute (EBI), Wellcome Genome Campus, Hinxton UK
24. Gastroenterology Division, The Hospital for Sick Children, Toronto, Canada.
25. SickKids Inflammatory Bowel Disease Center and Cell Biology Program, Research Institute, Hospital for Sick Children, Toronto, ON, Canada.
26. Department of Pediatrics and Biochemistry, University of Toronto, Hospital for Sick Children, Toronto, ON, Canada.

27. Howard Hughes Medical Institute, Department of Biology, Massachusetts Institute of Technology, Cambridge MA, USA
28. Current Address: Genentech, 1 DNA Way, South San Francisco, CA, USA
29. Department of Paediatrics, University of Oxford, Oxford, United Kingdom.

* Authors share co-first authorship.

Grant Support: We acknowledge the contribution of the BRC Gastrointestinal biobank (11/YH/0020, 16/YH/0247), supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre (BRC). This work was supported by the Leona M. and Harry B. Helmsley Charitable Trust (CK, SS, AMM, DMG, JC, GW, CA, ES, and HHU), the Manton Foundation (RX and AR), the Klarman Cell Observatory (AR), HHMI (AR). AMM is funded by a Canada Research Chair (Tier 1) in Pediatric IBD, CIHR Foundation Grant and NIDDK (RC2DK118640) Grant. ST, KJ and HHU are supported by the Wellcome Trust Human Cell Atlas grant. BS and CW are supported by the CRC-TRR 241 of the German Research Foundation (DFG), CW is fellow of the BIH Charité Clinician Scientist Program.

Acknowledgements:
We thank many authors of original papers and the UK Cystic Fibrosis Trust Registry for clarifications and kindly providing additional unpublished information. We thank Leslie Gaffney for her help with figure preparation.

Correspondence:
To whom correspondence should be addressed: holm.uhlig@ndm.ox.ac.uk (HU), Phone: 0044 1865 8 57963, Translational Gastroenterology Unit, John Radcliffe Hospital, University of Oxford, Oxford, OX3 9DU, UK.

Disclosures:
None of the authors have a conflict of interest related to this article. HHU received research support or consultancy fees from Janssen, Eli Lilly, UCB Pharma, Celgene, MiroBio, OMass, and Mestag. SPLT has been adviser to, in receipt of educational or research grants from, or invited lecturer for AbbVie;

Amgen; Asahi; Biogen; Boehringer Ingelheim; BMS; Cosmo; Elan; Enterome; Ferring; FPRT Bio; Genentech/Roche; Genzyme; Glenmark; GW Pharmaceuticals; Immunocore; Immunometabolism; Janssen; Johnson & Johnson; Lilly; Merck; Novartis; Novo Nordisk; Ocera; Pfizer; Shire; Santarus; Sensyne; SigmoidPharma; Synthon; Takeda; Tillotts; Topivert; Trino Therapeutics with Wellcome Trust; UCB Pharma; Vertex; VHsquared; Vifor; Warner Chilcott and Zeria. A.R. is a co-founder and equity holder of Celsius Therapeutics, an equity holder in Immunitas, and was an SAB member of ThermoFisher Scientific, Syros Pharmaceuticals, Neogene Therapeutics and Asimov until July 31, 2020. From August 1, 2020, A.R. is an employee of Genentech. BS has served as consultant for Abbvie, Arena, BMS, Boehringer, Celgene, Falk, Janssen, Lilly, Pfizer, Prometheus and Takeda and received speaker's fees from Abbvie, CED Service GmbH, Falk, Ferring, Janssen, Novartis, Pfizer, Takeda [served as representative of the Charité].

Author Contributions:

CB, CS, SP, RE, GW, CA, MM, DA, KRJ, JK, WH, KB, MC, AC, EP, CW, AR and HUH provided acquisition and analysis of the data. CB, CS, SP, RE, GW, CA, DA, KRJ, KB, MC, AC, EP, CW, DM, JC, DS, JK, SK, LP, ST, LJ, CAA, SS, CK, ES, MZ, RX, JM, PA, BS, ST, AMM, AR and HUH contributed to interpretation of the data. HUH, CB, CS and SP drafted the first version manuscript. CB, CS, SP, RE, CA, DA, KRJ, AR and HUH contributed to data visualisation. All authors read and provided feedback on the final manuscript. HHU is the guarantor.

Data Transparency Statement:

Raw and processed datasets from colon single-cell RNA-seq (scRNA-seq) are available[9]. Data from Ref[5,56] will be available on publication at gutcellatlas.org. Codes used for analysis are available at:
https://www.github.com/cssmillie/ulcerative_colitis

ABSTRACT:

Background and aims: Monogenic forms of inflammatory bowel disease (IBD) illustrate the essential roles of individual genes in pathways and networks safeguarding immune tolerance and gut homeostasis.

Methods: To build a taxonomy model we assessed 165 disorders. Genes were prioritized based on penetrance of IBD and disease phenotypes were integrated with multi-omics datasets. Monogenic IBD genes were classified by: (1) overlapping syndromic features; (2) response to hematopoietic stem cell transplantation; (3) bulk RNA-seq of 32 tissues; (4) single-cell RNA-seq of >50 cell subsets from the intestine of healthy individuals and IBD patients (pediatric and adult), and (5) proteomes of 43 immune subsets. The model was validated by addition of newly identified monogenic IBD defects. As a proof-of-concept, we explore the intersection between immunometabolism and antimicrobial activity for a group of disorders (*G6PC3*/*SLC37A4*).

Results: Our quantitative integrated taxonomy defines the cellular landscape of monogenic IBD gene expression across 102 genes with high and moderate penetrance (81 in the model set and 21 genes in the validation set). We illustrate distinct cellular networks, highlight expression profiles across understudied cell types (*e.g.,* $CD8^+$ T cells, neutrophils, epithelial subsets and endothelial cells) and define genotype-phenotype associations (perianal disease and defective antimicrobial activity). We illustrate processes and pathways shared across cellular compartments and phenotypic groups and highlight cellular immunometabolism with mTOR activation as one of the converging pathways. There is an overlap of genes and enriched cell-specific expression between monogenic and polygenic IBD.

Conclusion: Our taxonomy integrates genetic, clinical and multi-omic data; providing a basis for genomic diagnostics and testable hypotheses for disease functions and treatment responses.

Abbreviations

CD	Crohn's disease
CGD	Chronic granulomatous disease
DC2s	type 2 dendritic cells
FACS	fluorescence-activated cell sorting
GOF	Gain-of-function
IBD	inflammatory bowel diseases
IPEX	"Immune dysregulation, polyendocrinopathy, enteropathy, X-linked" (syndrome)
HSCT	Hematopoietic stem cell transplantation
LOF	Loss-of-function
MDM	Monocyte-derived macrophages
pS6	phosphorylation of S6 ribosomal protein
PBMCs	peripheral blood mononuclear cells
scRNA-seq	single-cell RNA-sequencing
T_{reg}	Regulatory T cells
UC	Ulcerative colitis
WAS	Wiskott-Aldrich syndrome

Key words:

Immunodeficiency; genomics; RNAseq; next-generation sequencing

INTRODUCTION

Inflammatory bowel diseases (IBD) arise from breakdown in mucosal barrier function and immune homeostasis[1,2]. IBD has been conventionally classified as Crohn's disease (CD), ulcerative colitis (UC) and unclassified IBD. Through Montreal and Paris criteria these disease forms are endo-phenotyped[3]; however in routine clinical practice they lack strong predictive power of disease course, have a limited impact on management across the wide spectrum of patients and a restricted ability to delineate molecular etiology. Substantial progress has been made in understanding the genetic and immunological etiology of IBD through genetic association studies and single-cell RNA-seq (scRNA-seq)[4–6]. However whilst over 240 loci have been associated with risk of polygenic IBD[4] only ~10% have been mapped to causal variants[7]. scRNA-seq has started to reveal the heterogeneity and interactions of immune, parenchymal and stromal cells that modify cellular or molecular responses to exert a protective or pathogenic effect[5,8–10].

A small proportion of patients develop IBD due to rare monogenic defects[11]. There is clinical need as well as scientific advantages to systematically investigating monogenic IBD: (**1**) patients present with more severe and distinguishable phenotypes; (**2**) the genetic cause of the disease, often identified from protein-coding variants, and corresponding molecular pathways, are potentially easier to map; and (**3**) many conventional therapies fail in these patients, but pathway-specific therapies can provide functional evidence. Indeed, for some forms of monogenic IBD, a genetic diagnosis will indicate pathway-specific therapies, predict postoperative recurrence, inform screening for malignancies/infections, facilitate genetic counselling and help to avoid an extremely prolonged course of intractable inflammation[11,12]. Nonetheless, the mechanistic model we have for different monogenic IBD disorders is limited by a strong bias towards a narrow subset of disorders, implicating selected groups of immune cells into functional models[12].

Here we tackled this challenge by appraising 165 Mendelian disorders and syndromes and building a comprehensive taxonomy of the 102 monogenic

disorders that are the most strongly associated with IBD, identifying shared cellular and molecular mechanisms (**Fig. 1a**).

MATERIALS AND METHODS

Monogenic and polygenic gene selection

A literature search was performed to identify Mendelian disorders and syndromes associated with IBD (**Supplementary Methods, Table S1**). For inclusion, patients required an IBD/CD/UC diagnosis or histological or endoscopic evidence of IBD-like intestinal inflammation in the absence of causal infection. IBD penetrance was estimated from unselected cohort descriptions available or summated from individual studies (**Table S2-3**). Candidate polygenic IBD genes were selected from three meta-analyses[4,13,14] and one mapping study[7] (**Supplementary Methods, Table S4**).

Cumulative allele frequency of nonsense variants was based on the Exome Aggregation Consortium (ExAC) variant server[15] (**Supplementary Methods**).

Bulk mRNA and protein expression

Bulk RNA-seq data of 32 human tissues were analyzed using the Human Protein Atlas[16]. Quantitative protein levels of 43 cell types from peripheral blood mononuclear cells (PBMCs) were obtained from ImmProt[17].

Single-cell mRNA analysis

scRNA-seq analyses were based on original publications[6,8,18]. Briefly, 366,650 cells were obtained from human colon[8] and 58,900 cells were obtained from pediatric ileum[6]. scRNA-seq data of human blood-derived neutrophils was added to the healthy colon data[18]. Published annotations and expression data were used to analyze mean expression levels. Gene modules and differential expression analyses were performed as described in (**Supplementary Methods**).

Pathway analysis

Monogenic IBD genes represented in pathways from the Reactome database[19] with an FDR ≤ 0.05 were visualized (**Supplementary Methods**).

Gentamicin protection assay

PBMCs were isolated and adherent human monocytes were differentiated into macrophages over 5 days. Monocyte-derived macrophages (MDM) were treated with a *Salmonella enterica* strain expressing green fluorescent protein (GFP; NCTC12023) for 1 hour, followed by gentamicin (100µg/ml) for 2 hours, as described[20] (**Supplementary Methods**). Intracellular bacteria were quantified by lysing MDM and plating on LB agar plates using the track method. The NCTC12023 strain expressing GFP under a pH sensitive promoter was used for a fluorescence-activated cell sorting (FACS)-based gentamicin protection assay.

Myeloid metabolic analysis

Metabolic flux analyses were performed using the Agilent glycolysis stress and mito-stress kits[21] (**Supplementary Methods**). Metabolomics was performed on primary human MDMs differentiated in the presence and absence of glucose-6-phosphate transporter inhibitor S3483[21]. Quantification of phosphorylation of S6 ribosomal protein (pS6) was performed by intracellular staining and FACS (**Supplementary Methods**).

Taxonomy

We integrated clinical and multi-omics data for 81 monogenic IBD disorders (published before 31/09/18), using weighted hierarchical clustering. To validate the robustness of the taxonomy, we repeated the construction of this taxonomy using 29 newly reported disorders (published 2018-2021).

Statistics

Calculations of significance were performed using designated statistical tests. For hierarchical clustering, rows and columns were clustered using correlation distance and average linkage.

RESULTS

Selection of monogenic genes: penetrance stratification and population-based genetics

To build a taxonomy model, we investigated a diverse group of 136 Mendelian disorders and syndromes based on a comprehensive literature review to identify patients with monogenic defects causing intestinal inflammation (**Fig. S1**). To ensure high-confidence gene mapping, we only included genes where patients had been described with histologically proven intestinal inflammation, attributed to a validated monogenic defect.

We classified the gene defects by the penetrance of inflammation, using a model that accounts for effect size and integrates patient numbers (**Fig. 1b-d**, **S2**). Homozygous or compound heterozygous *NOD2* variants possess the strongest known genetic risk for polygenic CD in European ancestry populations (≤4.9% penetrance[22–25]). These variants serve as a benchmark to differentiate monogenic genes of high impact from polygenic IBD loci. Therefore, we identified high penetrance genes with a higher IBD penetrance than *NOD2* variants (>4.9%; **Fig. 1b**). We defined moderate penetrance genes as those with an IBD penetrance between 1-5% (per the confidence interval), given the highest baseline risk of IBD in the population is estimated at 1%[26] (**Fig. 1c**). This model allowed the inclusion of rare defects, where even a single rigorously validated case report could be included by its expected effect size. On the other hand, this model also allowed exclusion of common disorders (such as cystic fibrosis (*CFTR*)), where many IBD patients are reported, but the penetrance is not significantly higher than the population risk (**Fig. 1d**).

We distilled hundreds of heterogeneous case reports down to 55 disorders with high-penetrance IBD, 26 disorders with moderate-penetrance, and 23 disorders with insufficient evidence for an IBD penetrance greater than 1% (**Fig. 1b-d**). The importance of directionality of effect was illustrated by the contrasting functional effects of loss-of-function (LOF) and gain-of-function (GOF) variants in many genes (**Fig. 1b-d**, **S3a**). Having selected monogenic IBD genes, we investigated the overlap with 278 polygenic IBD risk loci confidently mapped from genome-wide association studies[4,7,13,14]. The overlap of 13 genes was significant, suggesting some shared genetic basis between disease forms (although the variants themselves and their functional consequences are different, $P < 3.04*10^{-5}$; hypergeometric test; **Fig. 1e**).

Since Mendelian disorders are very rare within a population, we reasoned that essential LOF defects among the 76 monogenic IBD genes (with LOF consequences) should be rare. Indeed, monogenic IBD genes in the ExAC database[15] displayed low median frequencies of essential LOF alleles ($6.96*10^{-5}$ for autosomal, $1.28*10^{-5}$ for X-linked; **Fig. S3d**). Additionally, the combined frequencies of alleles with high-confidence LOF variants were well below that of *NOD2* (**Fig. S3b-d**). Conversely, genes like *DUOX2* and *NOX1* had higher LOF allele frequencies, confirming that these are unlikely monogenic IBD genes (**Fig. S3b-c, 1d**).

Monogenic IBD genes vary by age-of-onset, intestinal and extra-intestinal phenotypes, and therapeutic responses

We next identified phenotypic features of high- and moderate-penetrance monogenic IBD genes (**Fig. 2a-c**). Focusing on age of onset (*n*=338 records), patients with monogenic IBD defects were significantly younger than polygenic IBD patients ($P < 1*10^{-4}$, one-sided Mann-Whitney test; **Fig. 2a**), while those with high-penetrance defects presented with significantly younger age of onset than those with moderate-penetrance ($P = .0065$, one-sided Mann-Whitney test; **Fig. 2b**). Infantile IBD was associated with variants in 30 genes (37%), but others presented with an older median age of onset, including adulthood. Even among high-penetrance genes, 37% of defects were associated with a median age of onset above 6 years (**Fig. 2b**).

Patients in disease groups were largely diagnosed with CD (73%; n=59/81) or *both* CD and UC (14.8%; n=12/81) (**Fig. 2c**), showing the limited utility of the Montreal classification in distinguishing diverse monogenic IBD mechanisms. Montreal/Paris sub-phenotypes were not consistently reported, preventing a systematic analysis. However, perianal abscess formation was often reported, a distinct feature of CD[3], and was associated with defects in bacterial handling as assessed with assays like the gentamicin protection assay[20] ($P < 1*10^{-5}$, Fisher's exact test). In terms of extra-intestinal features, 67% of genes were curated into 14 syndromic phenotypic categories that were shared between ≥2 genes (**Fig. 2c**).

Sustained resolution of intestinal inflammation after allogenic HSCT suggests an underlying defect in hematopoietic cells, whereas non-response implicates non-hematopoietic cellular compartments[11]. HSCT was efficacious for IL-10 signaling defects; chronic granulomatous disease (CGD) and “Immune dysregulation, Polyendocrinopathy, Enteropathy, X-linked” (IPEX)-like syndrome (**Fig. 2c, Table S2**). There was moderate evidence for an additional 10 gene defects. For *TTC7A* and *IKBKG*, HSCT did not consistently cure intestinal inflammation, although it impacted the immunodeficiency phenotype, suggesting non-hematopoietic drivers of intestinal inflammation (**Fig. 2c**).

Gene and protein expression of monogenic IBD genes reveals enriched cellular compartments

To investigate the tissue compartments affected by monogenic IBD genes, we analyzed bulk expression profiles of 32 tissues[16]. Hierarchical clustering organized the monogenic IBD genes into tissue-specific subgroups, with most genes enriched in intestinal or primary/secondary lymphoid tissue (**Fig. S4**).

To further elucidate the cell types that may be affected by monogenic IBD disorders, we combined scRNA-seq spanning over 50 cell types from the healthy adult colon[8] (12 individuals) and healthy pediatric ileum[6] (8 individuals) (**Fig. 3a-d**). These data span much of the cellular diversity of the intestinal mucosa, with the exception of neutrophils, which due to technical reasons have eluded scRNA-seq studies. To mitigate this weakness, we added scRNA-seq data from blood-derived neutrophils of healthy humans[18] to the healthy colonic data and validated this approach using a separate immune proteomic dataset[17] (**Fig. 3a, 3e**). Overall, the expression of monogenic IBD genes was distinctly enriched in specific lineages across an array of colonic and ileal epithelial, mesenchymal, endothelial, myeloid, T cell, and B cell types (**Fig. 3a, 3c**).

In the colonic epithelial compartment, two gene expression patterns emerged. One subset of genes was expressed in mature enterocytes and goblet cells (*ALPI, SLC26A3, SLC9A3*, and *TTC7A*; **Fig. 3a, 3c**), suggesting impairment of the brush border. Another subset of genes was more broadly expressed in

epithelial (stem, crypt and transit-amplifying cells) and non-epithelial cells (*FERTM1-MVK*; **Fig. 3a, 3c**). Many epithelial-specific genes were not detectable at a protein level in PBMCs, supporting their specific relevance to non-hematopoietic cells (**Fig. 3e**).

Endothelial cells showed distinctly high expression of *TGFBR2* and *SLCO2A1* (**Fig. 3a, c**), particularly in capillary and microvascular cells. A distinct module was less clear for the mesenchymal compartment, where only *COL7A1* and *TTC37* (in inflamed colon) were enriched, highlighting inflammatory fibroblasts, WNT5B/2B+ fibroblasts, and smooth muscle cells (**Fig. 3a, 3c**).

Phagocyte-enriched genes comprised the largest group, but gene expression was not uniform across cell types (**Fig. 3a, 3c**). Peripheral blood neutrophils showed distinctly strong expression of many genes, including Wiskott Aldrich syndrome (WAS)-like genes (*WAS/WIPF1)* and CGD genes (*NCF2*/*NCF4)* (**Fig. 3a, 3e**). In contrast, monocyte, macrophage and type 2 conventional dendritic cell (DC2) subsets expressed high levels of *IL10RA, LACC1, CECR1/ADA2* and *HPS3* (**Fig. 3a, 3c**). We used an independent immune proteomic dataset[17] to confirm the enrichment of these gene products in neutrophils (**Fig. 3e**) and their congruent levels within monocytes (Spearman's $\rho = 0.83$, $P < .0001$; **Fig. 3e-f**). Most monogenic IBD genes had ~10-fold higher protein levels in neutrophils than in activated classical monocytes (**Fig. 3f**).

A set of genes, including *ZAP70, ICOS, CD40LG, CD3G and IL2RG*, were highly expressed across T lymphocytes (**Fig. 3a, 3c**). Regulatory T cells (T_{regs}) showed particularly strong enrichment, with distinctively high expression of *CTLA4, FOXP3* and *IL2RA.* T_{regs} also expressed a broader module of genes associated with IPEX-like syndrome, confirmed by proteomic data (**Fig. 3e**). $CD8^+IL17^+$ T cells, (expanded in ulcerative colitis[27]), showed enrichment of *CD40LG*, *ADA, CD3G* and *IL2RG* in inflamed colonic samples (**Fig. 4a**). Although not exclusive to B cells, *ADA, BACH2*, *NCF1,* and *DCLRE1C* showed highest mRNA expression in B cell subsets (**Fig. 3a, 3c**).

We investigated monogenic IBD gene expression during inflammation, using samples from the colon of 18 adult UC patients and the ileum of 7 pediatric CD patients (**Fig. 4a-d**). Monogenic IBD genes displayed up- and down-regulation during inflammation across a number of cell types; enterocytes, macrophages and $CD8^+$ intra-epithelial lymphocytes displayed the most differentially expressed genes between inflamed and healthy samples (**Fig. 4e**).

The expression of monogenic and polygenic IBD genes was strongly correlated across cell types (**Fig. 4f**; Spearman's $\rho = 0.61$; $P < 1x10^{-5}$). In particular, both monogenic and polygenic IBD genes were highly co-expressed in phagocytes (cycling monocytes, neutrophils, DC2) and activated T cells ($CD8^+IL\text{-}17^+$, cycling T cells) (**Fig. 4f**). Monogenic IBD genes showed greater enrichment in T_{regs}, inflammatory monocytes and $CD4^+PD1^+$ T cells, whereas polygenic IBD genes were particularly enriched in enterocytes and mesenchymal cells (**Fig. 4f**).

Pathway analysis and functional circuits of monogenic IBD

Hierarchical clustering of monogenic IBD genes according to their membership in known pathways revealed clusters of syndromic phenotypes (**Fig. 5a**). For example, CGD genes featured in Rho and RAC1 GTPase cycle pathways. Genes causing WAS-like disease and *BTK* were featured in pathways for phagosome actin dynamics and FcγR phagocytosis. Ectodermal dysplasia genes (*IKBKG, NFKBIA*) and *RELA* were involved in Toll-like-receptor and NF-kB activation pathways. Other clusters were enriched in TGFβ signaling, RUNX1-transcriptional regulation and in controlling development of T_{regs} (IPEX-like syndrome related genes *IL2RA, FOXP3*, *CTLA4*) (**Fig. 5a**). Broadly, this gene clustering infers underlying biology amongst syndromes, but also highlights key mechanistic gaps: 25% of all monogenic IBD genes, particularly epithelial genes, were not represented within Reactome pathways (with similar results for KEGG and GO pathways; **Fig. S5**).

In addition to cell-intrinsic pathways, the scRNA-seq data implicated extracellular communication and cytokine circuits across diverse compartments (**Fig. 5b**). IL-2 produced by T lymphocytes can act on macrophages,

monocytes, and T_{regs} expressing the IL-2 receptor. T_{regs} strongly expressed the CTLA4 checkpoint inhibitor, and together with monocytes, were major producers of IL-10. IL-10 may primarily exert its anti-inflammatory effects via monocytes and macrophages, both of which express TNF and IL-1. TNF acts on many cells including enterocytes, fibroblasts and endothelial cells, where its effects are modulated by *TNFAIP3* (A20). In contrast, IL-1 primarily appears to act on mesenchymal cells (**Fig. 5b**). The involvement of monogenic IBD genes in these cytokine networks, which form amplification cascades and feedback loops, supports a model where dysregulated inflammation across diverse cell types may be involved in the pathogenesis of monogenic IBD.

To determine whether monogenic and polygenic IBD genes may impact similar cell types and pathways, we used non-negative matrix factorization to learn modules of co-expressed genes in the healthy and inflamed bowel, and identified those significantly enriched in monogenic and polygenic IBD genes (**Fig. 5c,** FDR< $1x10^{-4}$; permutation test). Monogenic and polygenic genes were significantly more likely to be co-expressed in the same cell types (*e.g.,* T cells) – and even the same biological pathways within these cell types (*e.g.,* TCR signaling) – than a background set of genes with the same expression statistics, suggesting they may converge onto common pathways, with shared gene networks (**Fig. S6**).

Closing knowledge gaps: myeloid dysfunction in glucose-6-phosphate metabolism defects

For some genes, overlapping biochemical pathways and intestinal phenotypes suggest a shared mechanism, but the multi-omic data and HSCT response implicate seemingly contradictory cellular compartments. For example, defects in *G6PC3* and *SLC37A4* cause congenital neutropenia and monogenic IBD and are considered inborn errors of metabolism[28]. We found divergent transcriptomic and proteomic signals for these two genes that are involved in the metabolism (*G6PC3*) and transport (*SLC27A4*) of glucose-6-phosphate. Both genes were enriched in neutrophils in the proteomics data, but had low scRNA-seq expression (**Fig. 3a, 3e**), suggesting that glycolysis in neutrophils is particularly regulated by translational and post-transcriptional mechanisms[18].

Patients with *G6PC3* defects present with perianal abscess formation and resolution of colitis following HSCT (**Fig. 2c**), implicating hematopoietic cells. However, *G6PC3* and *SLC37A4* were enriched in non-hematopoietic cells in scRNA-seq and bulk RNA-seq data (epithelial and endothelial, respectively; **Fig. 3a, 3c, Fig. S4a**). PBMC protein levels were highest in neutrophils (*G6PC3*) and lymphocytes (*SLC37A4)* (**Fig. 3e**). This contradictory evidence illustrates the difficulties in classifying defects without knowledge of how cellular metabolism links with functional mechanisms of intestinal inflammation.

We investigated the role of *G6PC3* and *SLC37A4* defects in macrophages, given the association between perianal abscess formation and deficiencies in phagocytosis (*e.g.,* CGD, **Fig. 2c**). MDMs from patients with validated defects showed impaired bacterial clearance following infection with *Salmonella* in a gentamicin protection assay (**Fig. 6a**). We independently validated this finding *in vitro* by treating healthy MDMs with S3483, a specific inhibitor of the G6PT encoded by the *SLC37A4* gene, across varying concentrations and time scales (**Fig. 6b-c, S7**). Since both *SLC37A4* and *G6PC3* control glucose metabolism we tested whether S3483-treated macrophages have impaired metabolic flux. Indeed, S3483-treated MDMs showed elevated glycolysis (**Fig. 6d**) and reduced oxidative phosphorylation (**Fig. 6e**).

We performed metabolomics on S3483-treated macrophages to characterize their metabolic profiles. We found significantly reduced levels of adenosine monophosphate (**Fig. 6f**), which regulates AMPK and subsequent mTORC1 activity[21]. We next used pS6, a surrogate marker of mTORC1 signaling, to confirm significantly higher levels of mTORC1 signaling in S3483-treated macrophages *vs.* controls (**Fig. 6g**). Given the link between mTOR signaling and autophagy[29], we tested whether rapamycin (an mTORC1 inhibitor) could rescue bacterial handling defects, and strikingly found it reverted the defect (**Fig. 6h**). We confirmed this effect in primary macrophages from monogenic IBD patients with *SLC37A4* and *G6PC3* defects (**Fig. 6h**). These findings implicate defective immune-metabolism and subsequent impaired bacterial killing in phagocytes as driving factors in these inflammatory disorders. Thus,

targeting metabolism via the mTOR pathway may compensate for some genetically driven antimicrobial defects (**Fig. 6h-i**). This supports the concept that the overlapping syndromic phenotypes are caused by shared metabolic perturbations and defective antimicrobial activity via mTOR activation, leading to elevated inflammatory responses. Defective mTOR signaling also provides a plausible link between *G6PC3*, *SLC37A4* and other gene defects (such as IL10 signaling defects and *LACC1)*[29,30].

An integrated data-driven taxonomy of monogenic IBD

Taken together, our results suggested the classical polygenic IBD phenotype classification is largely not informative for monogenic IBD: age of onset marks a spectrum and CD or UC phenotypes span disparate underlying mechanisms. Monogenic IBD genes cluster into subgroups according to distinct disease syndromes, responses to therapy, gene or protein expression and involvement in biological pathways. We therefore integrated these diverse sources of information into a systematic taxonomy of monogenic IBD using weighted hierarchical clustering (**Fig. S8,** visualized in **Fig. S9**).

Across the taxonomy, not only were syndromic phenotypes associated with shared membership in Reactome pathways, but monogenic IBD genes were also more likely to be co-expressed in the same cell types and gene modules ($P < .001$, permutation test; **Fig. S10b**). Although preliminary due to small patient numbers and variable outcome measures, associations emerge between gene sub-clusters and response to biologic treatments. Treatment data was applied from an independent systematic review of monogenic IBD[31], supporting the potential relevance of the sub-clusters to predicting treatment responses (**Fig. S9**).

To update the classification and to confirm the robustness and scalability of the taxonomy, as well as the relationship between multi-omic and phenotypic data, we performed a search for monogenic disorders associated with IBD that were newly identified since 31/09/18 (**Fig. S1**). In the replicated analysis, we applied the same methods to generate an updated taxonomy (**Fig. 7a, S11-12, Table S5**). 21 genes demonstrated sufficient penetrance for inclusion (**Fig. S1,**

S11a). Despite the addition of 26% more disorders (21/81), only 5% of disorders (5/102) changed sub-cluster (**Fig. S13**). The new genes clustered plausibly by syndromic phenotypes, gene expression, and Reactome pathway involvement (*e.g., PRKCD, FMNL2, AGR2, PTPN2* and *IL2RB;* **Fig. 7a, S11b**).

In total, 81% (22/27) of monogenic IBD genes with evidence of response to HSCT showed enrichment in T lymphocyte and phagocytic groups (**Fig. 7a**). This association was not significant given the paucity of HSCT treatment data for non-immunodeficiency-related genes. Across the detailed criteria included, the taxonomy provides plausible evidence that intestinal inflammation associated with other gene defects may also respond to HSCT, such as those related to CGD, WAS-like and IPEX-like syndromes (**Fig. 7a**).

The taxonomy of 102 genes suggests overlapping pathway involvement and disease features across syndromic phenotypes. Among the phagocyte-predominant subgroups, sub-clusters contain genes particularly associated with immunometabolic functions; defective pathways involved in phagocytosis and impaired reactive oxygen species generation (**Fig. 7a-b**). Phagocytic-enriched genes were associated with impaired antimicrobial activity (P < .00001, Fisher's exact test) and perianal abscess formation (P = .0008, Fisher's exact test, **Fig. 7a, Table S5**). Many of these genes were strongly expressed in myeloid cells and displayed impaired bacterial clearance. Reactome pathway involvement would suggest defective autophagy via an impact on oxidative burst, actin dynamics involved in phagocytic cup formation or RHO GTPase regulation in genes of this group. In the immune-metabolic group, we have shown that *G6PC3* and *SLC37A4* defects cause impaired bacterial clearance due to increased mTOR signaling. This represents a central immune checkpoint that is also activated in metabolic defects of LACC1- and NPC1-deficient patients[29,32] (also in this sub-cluster, **Fig. 7a**). Defects in autophagy are associated with increased inflammasome activation and IL-1β production, during attempts to clear phagocytosed bacteria[33,34]. Although data is only available for a small number of syndromes, the suggested non-response to α4β7-inhibition (involved in T cell migration) and

efficacy of total IL-1 blockade would support the importance of these mechanisms in disease (**Fig. S9**).

In the T lymphocyte-enriched group, the IPEX-like genes showed specific enrichment in T_{regs} and were represented in Reactome pathways involving IL-2 signaling and T_{reg} development, reflecting impaired tolerance mechanisms in these defects (**Fig. 5a, S11b**). Directionality of gene defects has implications for treatment considerations, as *STAT1* and *STAT3* GOF defects have been successfully treated with JAK inhibitors, which corrects their dysregulated signal transduction[35]. Defects in *PTPN2* were recently associated with an IPEX-like phenotype[36] (**Table S5-6**) and closely clustered with *STAT3* GOF variants (**Fig. 7a**). Both genes were represented in IL-1 signaling pathways and IFN-y signaling regulation in pathway analysis (**Fig. 7a, S11b, Table S6**).

Biallelic IL10 signaling defects confer the strongest susceptibility to monogenic IBD with onset in early infancy (**Fig. 1b, 2a**). A high gene expression of IL10 and IL10RA is found in T_{regs} and the phagocytic compartment (**Fig. 5c**). The importance of this pathway is suggested by the association of classical IBD with common polymorphisms in the *IL10RB* and *IL10* locus, rare protein-coding variants in *IL10RA*[37,38] and the finding that functional IL10-resistance may be induced in monocytes following exposure to whole bacteria[39].

A group of disorders with causal genes predominantly expressed in intestinal epithelial cells show impaired epithelial barrier integrity or disrupted intestinal milieu from altered brush border enzymes (**Fig. 3a, 3c, 7a**). This included gene defects impacting structural integrity associated with distorted crypt architecture or epithelial polarization (*TTC7A, PI4KA*), and defective epithelial adhesion or impaired mucus barrier formation (*AGR2*) due to disturbances in goblet cell function and development (**Table S5-6**). A number of genes encoding solute carriers and associated with a phenotype of secretory congenital diarrhea were also included in this group. Intestinal inflammation in this group of disorders is likely a consequence of dysbiosis and Th17 differentiation defects secondary to disrupted luminal sodium and chloride electrolyte milieu[40]. Multiple gene products within this compartment were not

detected as proteins in PBMCs (**Fig. 3e**), suggesting their selective role in non-hematopoietic cells. Additionally, this group illustrated a trend suggesting a lack of response of TNF-α inhibition (**Fig. S9**).

A complex group of disorders with genes linked to NF-Kβ signaling defects showed both hematopoietic and non-hematopoietic cell expression (**Fig. 7a**). Patients with LOF *RIPK1* and *CASP8* variants display ulcerations on endoscopy, were suggested to respond to TNFα inhibition and these genes were represented in regulation of TNF-induced cell death pathways (**Fig. S11b**, **S9**; **Table S2, S5**). Notably, variants causing complete LOF in *RIPK1* were associated with combined immunodeficiency, whereas patients with missense mutations in the RIPK1 death domain presented with more of an IBD-like phenotype[41], suggesting dysregulation in signaling, as oppose to complete LOF may account for the efficacy of anti-TNFα therapy. Similarly hypomorphic defects in *IKBKG* are associated with IBD and ectodermal dysplasia, as oppose to complete LOF defects[42]. The variable function of hypomorphic *IKBKG* variants may account for the inconsistent response to HSCT[42].

Among genes enriched in endothelial cells, *SLCO2A1* is associated with multiple chronic, nonspecific ulcers and strictures of the small intestine, while *TGFBR2* is linked to Loeys-Dietz syndrome and defective TGF-β signaling. This is consistent with the key vascular manifestations of aortic aneurisms and dissection in Loeys-Dietz connective tissue disorder. However multiple compartments likely play a role TGF-β receptor signaling defects, for example *TGFB1* was strongly upregulated under inflamed conditions in T lymphocyte subsets and NK cells (**Fig. 4e**). Defects in epithelial-dominant *FERMT1* or mesenchymal-dominant *COL7A1* disrupt intestinal epithelial adhesion (**Fig. 7a**). *COL7A1* is a proof-of-concept defect that mesenchymal cells can selectively contribute to intestinal inflammation, although most genes with a mesenchymal expression peak showed expression across multiple cell types.

The final sub-cluster comprised a group of genes that lacked representation in Reactome pathways or distinct phenotypic features and displayed mixed

expression patterns, suggesting they require further investigation (**Fig. 7a**). Several genes in this group have been associated with autoinflammation (*NLRC4, POLA1, MVK, TRNT1*)[43], macrophage activation syndrome (*MVK, ITCH* and *NLRC4,* **Table S2**) and response to total IL-1 blockade (**Fig. S9**). Many demonstrated strong endothelial expression.

Discussion

This taxonomy provides data-driven insights into monogenic IBD, which facilitates a rational genetic panel design for clinical genomics, extended beyond expert opinion[44]. Our approach has several implications for data-driven classifications of monogenic primary immunodeficiencies or classical IBD. It highlights the limits of pure expression analyses, the value of integrating directionality of the signals and multiple datasets to understand cellular communication. Benchmarking arises from incorporating well-studied disease groups with plausible mechanisms (for example, IL10 signaling defects, defective antimicrobial activity in CGD or defects in T_{reg} activity in IPEX syndrome). Our taxonomy highlights a role for genes in understudied cell types (*e.g.,* $CD8^+$ T cells, neutrophils, endothelial cells), identifies genotype-phenotype associations (perianal disease and defective antimicrobial activity), the emerging role of cellular immunometabolism converging at the mTOR signaling checkpoint and knowledge gaps (*e.g.,* mesenchymal and epithelial pathway biology).

Whilst some gene modules clearly implicate specific cellular compartments, interaction networks or distinct processes, other genes may be expressed in multiple pathways across cellular compartments but impact certain cell types disproportionately. For example, goblet cells, Paneth cells and glial cells are specifically implicated in polygenic IBD pathogenesis[6,45]. This selective impact may be due to increased sensitivity to perturbation in these cells or less redundancy across processes. In our initial gene set, there was a noticeable absence of monogenetic evidence highlighting distinct roles for these cells, however the recent identification of monogenic defects impacting goblet cells[46], cells[46], enteric neurons[47] and Paneth cells[48] continues to suggest an overlap between disease forms.

We demonstrate that the taxonomy is scalable for the increasing number of monogenic conditions and genes being discovered and will benefit from systematically recorded phenotypes and treatment outcomes. It will be strengthened by combining further single-cell omics data that allows confidence in the gene expression and spatial omics technologies that resolves the architectural micro-niches in which the dysfunctional cells act[49]. Consideration is also warranted for the emerging role of the intestinal microbiota in the development and maintenance of gut homeostasis in primary immunodeficiencies and monogenic IBD. Given the very early onset in many patients with monogenic IBD, developmental aspects of the immune system likely contribute to the pathogenesis and could be incorporated in future classifications[50].

References:

1. Uhlig HH, Powrie F. Translating Immunology into Therapeutic Concepts for Inflammatory Bowel Disease. Annu Rev Immunol 2018;36:755–781.
2. Graham DB, Xavier RJ. Pathway paradigms revealed from the genetics of inflammatory bowel disease. Nature 2020;578:527–539.
3. Levine A, Griffiths A, Markowitz J, et al. Pediatric modification of the Montreal classification for inflammatory bowel disease: The Paris classification. Inflamm Bowel Dis 2011;17:1314–1321.
4. Lange KM de, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. Nat Genet 2017;49:256–261.
5. Martin JC, Chang C, Boschetti G, et al. Single-Cell Analysis of Crohn's Disease Lesions Identifies a Pathogenic Cellular Module Associated with Resistance to Anti-TNF Therapy. Cell 2019;178:1493-1508.e20.
6. Elmentaite R, Ross ADB, Roberts K, et al. Single-Cell Sequencing of Developing Human Gut Reveals Transcriptional Links to Childhood Crohn's Disease. Dev Cell 2020;55:771-783.e5.
7. Huang H, Fang M, Jostins L, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. Nature 2017;547:173–178.
8. Smillie CS, Biton M, Ordovas-Montanes J, et al. Intra- and Inter-cellular

Rewiring of the Human Colon during Ulcerative Colitis. Cell 2019;178:714-730.e22.

9. Kinchen J, Chen HH, Parikh K, et al. Structural Remodeling of the Human Colonic Mesenchyme in Inflammatory Bowel Disease. Cell 2018;175:372-386.e17.
10. Corridoni D, Chapman T, Antanaviciute A, et al. Inflammatory Bowel Disease Through the Lens of Single-cell RNA-seq Technologies. (Online ahead of print). Inflamm Bowel Dis 2020.
11. Uhlig HH, Schwerd T, Koletzko S, et al. The diagnostic approach to monogenic very early onset inflammatory bowel disease. Gastroenterology 2014;147:990-1007.e3.
12. Uhlig HH. Monogenic diseases associated with intestinal inflammation: implications for the understanding of inflammatory bowel disease. Gut 2013;62:1795–805.
13. Jostins L, Ripke S, Weersma RK, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. Nature 2012;491:119–24.
14. Liu JZ, Sommeren S van, Huang H, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. Nat Genet 2015;47:979–986.
15. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature 2020;581:434–443.
16. Uhlen M, Fagerberg L, Hallstrom BM, et al. Tissue-based map of the human proteome. Science (80-) 2015;347:1260419–1260419.
17. Rieckmann JC, Geiger R, Hornburg D, et al. Social network architecture of human immune cells unveiled by quantitative proteomics. Nat Immunol 2017;18:583–593.
18. Xie X, Shi Q, Wu P, et al. Single-cell transcriptome profiling reveals neutrophil heterogeneity in homeostasis and infection. Nat Immunol 2020;21:1119–1133.
19. Fabregat A, Sidiropoulos K, Viteri G, et al. Reactome pathway analysis: a high-performance in-memory approach. BMC Bioinformatics 2017;18:142.

20. Schwerd T, Pandey S, Yang H-T, et al. Impaired antibacterial autophagy links granulomatous intestinal inflammation in Niemann–Pick disease type C1 and XIAP deficiency with NOD2 variants in Crohn's disease. Gut 2017;66:1060–1073.
21. Schulthess J, Pandey S, Capitani M. The Short Chain Fatty Acid Butyrate Imprints an Antimicrobial Program in Macrophages. Immunity 2019;50:432-445.e7.
22. Brant SR, Wang M-H, Rawsthorne P, et al. A Population-Based Case-Control Study of CARD15 and Other Risk Factors in Crohn's Disease and Ulcerative Colitis. Am J Gastroenterol 2007;102:313–323.
23. Silver J. The Importance of Penetration. Inflamm Bowel Dis 2003;9:341.
24. Zhou Z, Lin X, Akolkar PN, et al. Variation at NOD2/CARD15 in familial and sporadic cases of Crohn's disease in the Ashkenazi Jewish population. Am J Gastroenterol 2002;97:3095–3101.
25. Yazdanyar S, Kamstrup PR, Tybjærg-Hansen A, et al. Penetrance of NOD2/CARD15 genetic variants in the general population. 2010;182:661–665.
26. Ng SC, Shi HY, Hamidi N, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 2017;390:2769–2778.
27. Tom MR, Li J, Ueno A, et al. Novel CD8+ T-Cell Subsets Demonstrating Plasticity in Patients with Inflammatory Bowel Disease. Inflamm Bowel Dis 2016;22:1596–1608.
28. Ghosh A, Schlecht H, Heptinstall LE, et al. Diagnosing childhood-onset inborn errors of metabolism by next-generation sequencing. Arch Dis Child 2017;102:1019–1029.
29. Omarjee O, Mathieu A-L, Quiniou G, et al. LACC1 deficiency links juvenile arthritis with autophagy and metabolism in macrophages. J Exp Med 2021;218.
30. Ip WKE, Hoshi N, Shouval DS, et al. Anti-inflammatory effect of IL-10 mediated by metabolic reprogramming of macrophages. Science (80-) 2017;356:513–519.
31. Nambu R, Warner N, Mulder DJ, et al. A Systematic Review of Monogenic Inflammatory Bowel Disease. Clin Gastroenterol Hepatol

2021:1212.

32. Davis OB, Shin HR, Lim CY, et al. NPC1-mTORC1 Signaling Couples Cholesterol Sensing to Organelle Homeostasis and Is a Targetable Pathway in Niemann-Pick Type C. Dev Cell 2021;56:260-276.e7.
33. Lee PP, Lobato-Márquez D, Pramanik N, et al. Wiskott-Aldrich syndrome protein regulates autophagy and inflammasome activity in innate immune cells. Nat Commun 2017;8.
34. Luca A De, Smeekens SP, Casagrande A, et al. IL-1 receptor blockade restores autophagy and reduces inflammation in chronic granulomatous disease in mice and in humans. Proc Natl Acad Sci U S A 2014;111:3526–3531.
35. Forbes LR, Vogel TP, Cooper MA, et al. Jakinibs for the treatment of immune dysregulation in patients with gain-of-function signal transducer and activator of transcription 1 (STAT1) or STAT3 mutations. J Allergy Clin Immunol 2018;142:1665–1669.
36. Parlato M, Nian Q, Charbit-Henrion F, et al. Loss-of-Function Mutation in PTPN2 Causes Aberrant Activation of JAK Signaling Via STAT and Very Early Onset Intestinal Inflammation. Gastroenterology 2020;159:1968-1971.e4.
37. Gettler K, Levantovsky R, Moscati A, et al. Common and Rare Variant Prediction and Penetrance of IBD in a Large, Multi-ethnic, Health System-based Biobank Cohort. Gastroenterology 2021;160:1546–1557.
38. Sazonovs A, Stevens CR, Venkataraman GR, et al. Sequencing of over 100,000 individuals identifies multiple genes and rare variants associated with Crohns disease susceptibility. medRxiv 2021:2021.06.15.21258641.
39. Aschenbrenner D, Quaranta M, Banerjee S, et al. Deconvolution of monocyte responses in inflammatory bowel disease reveals an IL-1 cytokine network that regulates IL-23 in genetic and acquired IL-10 resistance. Gut 2021;70:1023–1036.
40. Wilck N, Matus MG, Kearney SM, et al. Salt-responsive gut commensal modulates TH17 axis and disease. Nature 2017;551:585–589.
41. Li Y, Führer M, Bahrami E, et al. Human RIPK1 deficiency causes combined immunodeficiency and inflammatory bowel diseases. Proc Natl Acad Sci 2019;116:970–975.

42. Miot C, Imai K, Imai C, et al. Hematopoietic stem cell transplantation in 29 patients hemizygous for hypomorphic IKBKG /NEMO mutations. Blood 2017;130:1456–1467.
43. Papa R, Picco P, Gattorno M. The expanding pathways of autoinflammation: a lesson from the first 100 genes related to autoinflammatory manifestations. Adv Protein Chem Struct Biol 2020;120:1–44.
44. Uhlig HH, Charbit-Henrion F, Kotlarz D, et al. Clinical Genomics for the Diagnosis of Monogenic Forms of Inflammatory Bowel Disease: A Position Paper From the Paediatric IBD Porto Group of European Society of Paediatric Gastroenterology, Hepatology and Nutrition. J Pediatr Gastroenterol Nutr 2021;72:456–473.
45. Parikh K, Antanaviciute A, Fawkner-Corbett D, et al. Colonic epithelial cell diversity in health and inflammatory bowel disease. Nature 2019;567:49–55.
46. AA A-S, UM A-M, SZ H, et al. Human AGR2 Deficiency Causes Mucus Barrier Dysfunction and Infantile Inflammatory Bowel Disease. Cell Mol Gastroenterol Hepatol 2021.
47. Sifuentes-Dominguez L, Li H, Llano E, et al. SCGN deficiency results in colitis susceptibility. Elife 2019;8.
48. Yu J, He X, Wei A, et al. HPS1 Regulates the Maturation of Large Dense Core Vesicles and Lysozyme Secretion in Paneth Cells. Front Immunol 2020;11:560110.
49. Fawkner-Corbett D, Antanaviciute A, Parikh K, et al. Spatiotemporal analysis of human intestinal development at single-cell resolution ll Spatiotemporal analysis of human intestinal development at single-cell resolution. Cell 2021;184:810–826.
50. Elmentaite R, Kumasaka N, King HW, et al. Cells of the human intestinal tract mapped across space and time. bioRxiv 2021:2021.04.07.438755.

FIGURE LEGENDS

Figure 1: Approach to selecting and investigating monogenic IBD genes

(**a**) Schematic representation of classification. (**b-d**) Monogenic gene defects were classified by penetrance. Black dots: number of patients with validated defect. Bars: 90% confidence interval.

(**e**) Network diagram showing the genes belonging to different polygenic and monogenic IBD gene classes, highlighting the significant overlap between monogenic and polygenic IBD (labelled red nodes; n=13, $P < 3x10^{-5}$, hypergeometric test).

Figure 2: Phenotypic characteristics of people with monogenic IBD gene defects

(**a**) Left: age of onset of intestinal inflammation for patients (n=338) with validated monogenic IBD defects. Right: age of onset distribution for polygenic IBD patients (n=1608). Red dashed line denotes the median.

(**b**) Age of onset for high- and moderate-penetrance IBD genes (visualized with log scale, * $P < .01$). Red line: median.

(**c**) Phenotype associations for patients with monogenic gene defects.

Figure 3: Monogenic IBD gene expression in samples from healthy participants

(**a-d**) Monogenic IBD gene expression was enriched in specific cell subsets from healthy samples from the adult colon (n=12) and pediatric ileum (n=8). (**a, c**) Scaled mean expression of monogenic IBD genes across cell subsets. Black outlines: q <.05.

(**b, d**) PCA of mean expression levels across all cell types showing cell lineage specificity.

(**e**) Scaled mean protein copy numbers of monogenic IBD genes across PBMCs (n=3-4 donors), sorted by FACS into steady state (ss) or activated (a) subsets. Grey: no data available (n=12 genes).

(**f**) Mean protein copy numbers of activated classical monocytes vs. peripheral blood neutrophils in healthy subjects (n=3-4) showing strongly correlated gene expression.

Figure 4: Monogenic IBD gene expression in inflamed ileum and colon

(**a-d**) Monogenic IBD gene expression was enriched in specific cell subsets from inflamed samples from the adult colon (*n*=18) and pediatric ileum (*n*=7). (**a,c**) Scaled mean expression (color) of monogenic IBD genes across cell subsets. Black outlines: $q < .05$. (**b,d**) Principal components of mean gene expression levels of monogenic IBD genes across cell types showing cell lineage specificity.

(**e**) Differentially expressed (DE) genes in inflamed *vs*. healthy samples across cell subsets, from adult colon. Dot color: DE model coefficients ($q < .05$; discrete coefficient from MAST).

(**f**) Monogenic and polygenic IBD genes were enriched in overlapping cell types: mean expression of genes for each cell type from the healthy human colon.

Figure 5: Monogenic IBD gene signalling pathways and gene module co-expression

(**a**) Monogenic IBD genes with the same syndromic phenotypes are enriched in the same signalling pathways. Clustering of the presence/absence profiles of monogenic IBD genes across Reactome pathways ($q < .05$).

(**b**) Mean expression levels of cytokines and their cognate receptors across selected cell subsets from the inflamed adult colon. Asterisks denote monogenic IBD genes.

(**c**) Monogenic and polygenic IBD genes co-occur in gene expression modules in adult colon (left) or pediatric ileum (right), the normalized usage of each module across each cell subset (bars), colored by lineage. Black text: monogenic, Grey text: polygenic.

Figure 6: Myeloid dysfunction in congenital neutropenia caused by *G6PC3* and *SLC37A4* gene defects

(**a-b**) Defective bacterial clearance was observed from the gentamicin protection assay on (**a**) *G6PC3*-/- (red) and *SLC37A4*-/- (blue) patients and (**b**) S4383-treated MDMs at baseline compared to controls. Each dot represents an individual donor sample.

(**c**) Left: Macrophages were infected with *Salmonella* expressing GFP under acid sensitive promoter, treated with gentamicin and sorted. Right: representative FACS plot shows impaired *Salmonella* killing in S3483-treated MDMs (red) vs controls (black).

(**d-e**) S3483-treated MDMs showed disrupted metabolism, with increased glycolysis (**d**) and reduced oxidative phosphorylation (**e**). Data represents the mean of 6 biological replicates from 2 independent experiments. Green dashed line shows time of injection as indicated.

(**d**) Real-time changes in the Extracellular acidification rate (ECAR) of MDMs as assessed by Seahorse glycolysis stress assay. (**e**) Real-time changes in the oxygen consumption rate (OCR) of MDMs as assessed by Seahorse mito-stress assay.

(**f**) Metabolomic changes in S3483-treated macrophages. Left: the abundances of all metabolites that differ significantly between samples from control *vs.* S3483-treated macrophages (*n*=5 for each group). Right: their corresponding fold-changes (positive: enriched in control samples). (**g**) Quantification of pS6 from control and S3483 treated macrophages via FACS.

(**h**) Defective bacterial clearance was restored by rapamycin in gentamycin protection assay. Colony-forming units (CFU) for S3483-treated and patient macrophages +/-rapamycin (Rapa), normalized to the control..

(**i**) Schematic summary: glucose is converted to glucose-6-phosphate which may enter: glycolysis, the pentose phosphate pathway (R5P), uptake into the endoplasmic reticulum (ER) or glycogen storage. Genetic mutations in *SLC37A4* or *G6PC3* lead to increased glycolysis, mTOR activation and impaired antimicrobial activity. mTOR1 inhibitor rapamycin can restore antimicrobial activity. GSD1b = glycogen storage disease 1b, R5P = pentose phosphate pathway

Statistical significance was determined using paired t-test **P* < .05, ***P* < .01, ***P* < .001 and ****P* < .0001.

Figure 7: An integrated taxonomy of monogenic IBD genes (Oxford classification of monogenic IBD)

(**a**) Weighted hierarchical clustering provides a scalable classification, which is shown schematically (**b**). (**a**) For each gene cluster, putative mechanisms integrating multiple lines of evidence are provided (right).