ARTICLE

https://doi.org/10.1038/s41467-018-07841-3 OPEN

Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response

Yael Haberman1,2, Rebekah Karns1, Phillip J. Dexheimer1, Melanie Schirmer3, Judith Somekh4,5, Ingrid Jurickova1, Tzipi Braun2, Elizabeth Novak6, Laura Bauman1,7, Margaret H. Collins1, Angela Mo8, Michael J. Rosen1, Erin Bonkowski1, Nathan Gotman9, Alison Marquis9, Mason Nistel1, Paul A. Rufo10, Susan S. Baker11, Cary G. Sauer12, James Markowitz13, Mariam D. Pfefferkorn14, Joel R. Rosh15, Brendan M. Boyle16, David R. Mack17, Robert N. Baldassano18, Sapan Shah19, Neal S. Leleiko20, Melvin B. Heyman21, Anne M. Griffiths22, Ashish S. Patel23, Joshua D. Noe24, Bruce J. Aronow1, Subra Kugathasan2, Thomas D. Walters22, Greg Gibson8, Sonia Davis Thomas9,25, Kevin Mollen6, Shai Shen-Orr4, Curtis Huttenhower3,26, Ramnik J. Xavier3,27, Jeffrey S. Hyams28 & Lee A. Denson1

Molecular mechanisms driving disease course and response to therapy in ulcerative colitis (UC) are not well understood. Here, we use RNAseq to define pre-treatment rectal gene expression, and fecal microbiota profiles, in 206 pediatric UC patients receiving standardised therapy. We validate our key findings in adult and paediatric UC cohorts of 408 participants. We observe a marked suppression of mitochondrial genes and function across cohorts in active UC, and that increasing disease severity is notable for enrichment of adenoma/adenocarcinoma and innate immune genes. A subset of severity genes improves prediction of corticosteroid-induced remission in the discovery cohort; this gene signature is also associated with response to anti-TNFα and anti-α4β7 integrin in adults. The severity and therapeutic response gene signatures were in turn associated with shifts in microbes previously implicated in mucosal homeostasis. Our data provide insights into UC pathogenesis, and may prioritise future therapies for nonresponders to current approaches.

1 Cincinnati Children’s Hospital Medical Center, and the University of Cincinnati College of Medicine, 45229 Cincinnati, OH, USA. 2 Sheba Medical Center, Tel Hashomer, affiliated with the Tel Aviv University, Tel Aviv 5265601, Israel. 3 Broad Institute of MIT and Harvard University, Cambridge 02142 MA, USA. 4 Faculty of Medicine, Technion, Haifa 3109601, Israel. 5 Department of Information Systems, University of Haifa, Haifa 3498838, Israel. 6 Children’s Hospital of Pittsburgh, Pittsburgh 15224 PA, USA. 7 Department of Pediatrics, University of California at San Diego, La Jolla 92162 CA, USA. 8 Georgia Institute of Technology, Atlanta 30332 GA, USA. 9 Collaborative Studies Coordinating Center, University of North Carolina, Chapel Hill 27516 NC, USA. 10 Harvard—Children’s Hospital Boston, Boston 02115 MA, USA. 11 Women & Children’s Hospital of Buffalo WCHOB, Buffalo 14222 NY, USA. 12 Emory University, Atlanta 30322 GA, USA. 13 Cohen Children’s Medical Center of New York, 11040 New Hyde Park, NY, USA. 14 Riley Hospital for Children, Indianapolis 46202 IN, USA. 15 Goryeb Children’s Hospital—Atlantic Health, Morristown 07960 NJ, USA. 16 Nationwide Children’s Hospital, Columbus 43205 OH, USA. 17 Children’s Hospital of East Ontario, Ottawa, Ontario K1P 1J1, Canada. 18 The Children’s Hospital of Philadelphia, Philadelphia 19104 PA, USA. 19 Children’s Hospital of Pittsburgh of UPMC, Pittsburgh 15224 PA, USA. 20 Hasbro Children’s Hospital, Providence 02903 RI, USA. 21 University of California at San Francisco, San Francisco 94143 CA, USA. 22 Hospital for Sick Children, Toronto M5G 1X8, Canada. 23 UT Southwestern, Dallas 75390 TX, USA. 24 Medical College of Wisconsin, Milwaukee 53226 WI, USA. 25 RTI International, Research Triangle Park 27709 NC, USA. 26 Harvard School of Public Health, Boston 02115 MA, USA. 27 Massachusetts General Hospital, Harvard Medical School, Boston 02114 MA, USA. 28 Connecticut Children’s Medical Center, Hartford 06106 CT, USA. Correspondence and requests for materials should be addressed to L.A.D. (email: lee.denson@cchmc.org)
Ulcerative colitis (UC) is a chronic relapsing–remitting inflammatory bowel disease (IBD) diagnosed primarily in young individuals. The disease burden has increased with globalization; newly industrialized countries show the greatest increase in incidence and the highest prevalence is recorded in Western countries. Disease severity and treatment response are strikingly heterogeneous with some patients quickly and continuously responding to initial therapies while others experience ongoing inflammation ultimately requiring surgical resection of the affected bowel. Greater understanding of individualized pathways driving clinical and mucosal severity and response to therapy, and the clinical translation of these data, is needed to proactively identify targeted therapeutic approaches. To improve our understanding of UC pathogenesis and its potential clinical personalized translation, we applied a standardized approach to a large, multicenter inception cohort that collected samples before treatment initiation, and included subjects representing the full spectrum of disease severities. The Predicting Response to Standardized Pediatric Colitis Therapy (PROTECT) study included 428 UC patients from 29 pediatric gastroenterology centers in North America. At diagnosis, disease severity was centrally read, and clinical and demographic data were recorded. Patients were assigned a specific standardized initial therapy with mesalamine or corticosteroids, and outcomes were recorded. Rectal biopsies from a representative subcohort of 206 patients underwent high-throughput RNA sequencing (RNAseq) prior to medical therapy, representing the largest UC transcriptomic cohort to date (Supplementary Table 1). We capture robust gene expression and pathways that are linked to UC pathogenesis, severity, response to corticosteroid therapy, and gut microbiota, which provide new insights into molecular mechanisms driving disease course.

Results
A unique treatment-naive UC inception cohort. The PROTECT study systematically examined response of 428 newly diagnosed pediatric UC patients to consensus-defined disease severity-based treatment regimens guided by the Pediatric Ulcerative Colitis Activity Index (PUCAI). mRNAseq defined pretreatment rectal gene expression for a representative discovery group of 206 UC PROTECT patients, a validation group of 50 UC PROTECT patients, and 20 age- and sex-matched non-IBD controls (Table 1). The validation group had similar characteristics to the discovery group, but with a higher frequency of nonwhite participants. More severe endoscopic disease (Grade 3 Mayo endoscopic subscore, Chi squares p < 0.001) and more extensive disease or pancolitis (Chi squares p < 0.001) were noted in moderate-severe cases. Of the patients with mild disease, 53% of 54 received initial therapy with mesalamine, and all moderate-severe patients received initial therapy with corticosteroids. Week 4 remission was defined as PUCAI < 10 without additional therapy or colectomy and was achieved by 105 of 206 (51%) patients in the discovery cohort. One hundred and fifty-six patients also had 16S rRNA sequencing to characterize their gut microbial communities.

The core UC gene signature. We defined a core rectal UC gene expression signature composed of 5296 genes (Fig. 1a) differentially expressed (FDR < 0.001 and fold change (FC) ≥ 1.5) in comparison to controls (Ctl, Fig. 1 and Supplementary Dataset 1).

Table 1 Characteristics of controls and PROTECT ulcerative colitis discovery and validation cohorts

|                | Ctl (n = 20) RNAseq | UC (n = 428) Full PROTECT Cohort | UC (n = 206) RNAseq | UC mild (n = 54) RNAseq | UC mod-sev Discovery (n = 152) RNAseq | UC mod-sev Validation (n = 50) RNAseq |
|----------------|---------------------|----------------------------------|--------------------|------------------------|----------------------------------------|---------------------------------------|
| Age (Mean ± SD) | 13.9 ± 3.3          | 12.7 ± 3.3                       | 12.9 ± 3.2         | 13.1 ± 3.5             | 12.8 ± 3.1                             | 12.4 ± 3.4                            |
| Sex M (%)       | 9 (45%)             | 216 (50%)                        | 112 (54%)          | 32 (59%)               | 80 (53%)                               | 23 (46%)                              |
| BMI z score (Mean ± SD) | 0.3 ± 1.6 | −0.2 ± 1.3                       | −0.26 ± 1.32       | −0.08 ± 1.19           | −0.33 ± 1.36                          | −0.28 ± 1.27                          |
| White           | 17/20 (85%)         | 351/420 (84%)                    | 204/206 (99%)      | 52/54 (96%)            | 152/152 (100%)                        | 28/50 (56%)                           |
| Mayo endoscopy subscore (range 0–3) | | | | | | |
| Grade 1 Mild    | 102 (24%)           | 54 (26%)                         | 54 (100%)          | —                      | —                                     |                                       |
| Grade 2 Moderate| 185 (43%)           | 84 (41%)                         | 83 (55%)           | 69 (45%)               | 21 (42%)                              |                                       |
| Grade 3 Severe  | 141 (33%)           | 68 (33%)                         | 71 (34%)           | 66 (43%)               | 26 (52%)                              |                                       |
| Disease location |                        |                                  |                    |                        |                                       |                                       |
| Proctosigmoiditis| 29 (7%)             | 14 (7%)                          | 11 (20%)           | 3 (2%)                 | 0 (0%)                                |                                       |
| Left-sided colitis| 44 (10%)         | 25 (12%)                         | 14 (26%)           | 11 (7%)                | 1 (2%)                                |                                       |
| Extensive/Pancolitis | 355 (83%) | 167 (81%)                        | 138 (91%)          | 49 (98%)               |                                       |                                       |
| Unassessable    | 136 (32%)           | 53 (26%)                         | 53 (98%)           | —                      |                                       |                                       |
| Initial treatment |                                |                                  |                    |                        |                                       |                                       |
| Mesalamine      | 136 (32%)           | 53 (26%)                         | 53 (98%)           | —                      |                                       |                                       |
| Oral or IV steroids | 292 (68%) | 153 (74%)                        | 152 (100%)         | 50 (100%)              |                                       |                                       |
| Oral steroids   | 144 (34%)           | 82 (40%)                         | 81 (53%)           | 20 (40%)               |                                       |                                       |
| IV steroids     | 148 (34%)           | 71 (34%)                         | 71 (47%)           | 30 (60%)               |                                       |                                       |
| Week 4 remission (PUCAI < 10) | 211/422 (50%) | 105 (51%)                        | 75 (49%)           | 21 (42%)               |                                       |                                       |
| Week 4 fecal calpro < 250 | 56/282 (20%) | 39/150 (26%)                     | 14/42 (33%)        | 9/28 (32%)             |                                       |                                       |

PUCAI: Pediatric Ulcerative Colitis Activity Index
*aUnassessable: severe/fulminant disease at presentation and the clinician performed a flexible sigmoidoscopy for safety concerns. Data are mean ± SD, n (%) unless noted otherwise. n/N values show missing data
Fig. 1 The core genes and pathways of newly diagnosed treatment-naive pediatric Ulcerative colitis emphasize lymphocyte activation and mitochondrial dysfunction. a Volcano plot of the 5296 differentially expressed genes between 206 UC and 20 Ctrl samples (FC ≥ 1.5 and FDR < 0.001). Functional annotation enrichment analyses of the b 3600 upregulated and c 1696 downregulated UC core genes using CluGO charts. Detailed functional annotation enrichment analyses of the d 3600 upregulated and e 1696 downregulated UC core genes using ToppGene, ToppCluster, and Cytoscape are shown. GO Biological Process, Cellular Component, and Molecular Function (pink), pathways (light blue), mouse phenotype (blue), gene family (yellow), and cells (IEC13) and functions of the above studies. Box and whisker plot with central line indicating median, box ends representing upper and lower quartile, and whisker represent 10–90 percentile. UC ulcerative colitis.

3600 up-regulated genes

1696 down-regulated genes
Functional annotation enrichment analyses using ToppGene<sup>6</sup>, ToppCluster<sup>2</sup>, and CluGO<sup>8</sup> mapped groups of related genes to biological processes<sup>9</sup>. Overview CluGo pie charts (Fig. 1b, c) showed highest enrichment for increased lymphocyte activation and associated cytokine signaling, and a robust decrease in mitochondrial, aerobic tricarboxylic acid (TCA) cycle, and metabolic functions. P values for the top specific biological processes were obtained as an output from ToppGene (Supplementary Dataset 1) and more detailed ToppCluster pathways analysis output is shown in Fig. 1d for the 3600 upregulated and Fig. 1e for the 1686 downregulated genes. Uregulated gene signatures were enriched for integrin signaling (<i>P</i> < 1.08E-12), JAK-STAT cascade, and TNF production (<i>P</i> < 9.9E-93), pathways that are already associated with therapeutic advances in UC<sup>10,11</sup>. The downregulated UC signature showed a robust decrease of mitochondrial-encoded and nuclear-encoded mitochondrial genes (<i>P</i> < 2.76E-35). Applying a computational gene expression deconvolution approach to estimate the relative composition of immune cell subsets, epithelia, and other stromal cell types in each sample (Methods) showed a significant increase in the estimated proportion of several immune cells including T and B cells, dendritic cells (DC), and monocytes (Fig. 1f). Using RISK cohort rectal biopsies mRNAseq data for treatment-naïve pediatric UC patients (Fig. 1g, Supplementary Dataset 1) and colonic biopsies microarray data of adults with active UC (Fig. 1i, Supplementary Dataset 1, GSE59071<sup>12</sup>), we demonstrated that 87% of the differentially expressed genes in RISK UC, and 80% of the adult UC genes, were within our core PROTECT signature. Comparing the differentially expressed genes from isolated intestinal epithelial cells (IEC) from another pediatric UC inception cohort<sup>13</sup> showed an overlap of 94% of the genes identified in that study (Fig. 1h) with the PROTECT genes, validating the majority of the core PROTECT UC signature in whole biopsies and in isolated epithelia. Functional annotation enrichment analyses of the shared genes further confirmed many of the common enriched pathways (Supplementary Dataset 1). Comparing the shared downregulated genes and pathways between PROTECT, RISK, adult UC cohort GSE59071<sup>12</sup>, and the IEC UC cohort<sup>13</sup> using ToppGene/ToppCluster confirmed the reduction of mitochondrial metabolic-associated genes and pathways, genes associated with lipid metabolism, and genes associated with formation of adenoma and adenocarcinoma (Fig. 1j).

**Robust colonic mitochondrialopathy in UC.** Notably, the mitochondrial genome encodes 13 genes regulating ATP production and all 13 were significantly reduced in UC (Fig. 2a). Real-time analysis of cellular respiration<sup>14</sup> was subsequently evaluated in colonic biopsies from UC and control patients. Mitochondrial electron transport chain Complex I activity, the rate-limiting step in oxidative phosphorylation<sup>15,16</sup> was reduced in active UC rectal biopsies compared to those from control patients (Fig. 2b). There was also a nonsignificant trend toward a decrease in Complex II activity (Fig. 2c). The mitochondrial membrane potential (MMP) that provides an integrated measure of the cellular capacity for ATP production was measured using JC-1 staining and FACS analysis of freshly isolated EpCAM+ colon epithelial cells (Fig. 2d) and CD45+ leukocytes (Fig. 2e). This showed a specific reduction in epithelial cells of active UC, with recovery in inactive UC (Supplementary Fig. 1). In addition, PPARGCA (PGC-1α), the master regulator of mitochondrial biogenesis, was profoundly reduced in UC patients in comparison to controls in PROTECT, RISK, and adult UC (Fig. 2f, h, j), and the IEC UC cohort (Supplementary Dataset 1)<sup>13</sup>. Principal coordinates analysis (PCA) principal components 1 (PC1) to summarize the Krebs cycle (TCA) genes variations between patients showed reduction of genes regulating mitochondrial energy production in the UC groups (Fig. 2g, i, k). The RISK dataset revealed a spectrum of mitochondrial gene expression downregulation in inflamed whole rectal biopsies, ranging from no significant suppression in mucosal biopsies obtained from inflamed rectum of ileo-colonic CD (L3 iCD) patients, to moderate suppression in samples from inflamed rectal biopsies of colon-only CD (L2 cCD) patients, and profound suppression in samples from pediatric UC samples with inflamed rectum (Fig. 2h, i). The spectrum between UC and CD was validated in the adult IBD cohort (GSE59071<sup>12</sup>, Fig. 2j, k), and we noted a recovery of this pathway in inactive adult UC. However, the larger PROTECT mRNAseq cohort permitted identification of an additional 3106 differentially expressed genes, which primarily demonstrated more robustly the suppression of mitochondrial pathways (Supplementary Dataset 1). Immuno-histochemistry confirmed reduced epithelial abundance of both mitochondrial-encoded MT-CO1 and nuclear-encoded COX5A genes which comprise complex IV in active UC (Fig. 2l, m).

**Disease severity gene signatures.** More severe disease is linked in our data and others to higher rates of therapy escalation and colectomy<sup>3,17</sup>, whereas mild disease is associated with remission by 12 weeks. Unsupervised hierarchical clustering analysis using the core 5296 genes grouped 204 of 206 UC cases in the dendrogram cluster A while all 20 non-IBD controls were in cluster B (Supplementary Fig. 2). Most mild cases grouped in A(i), while severe cases tended to be enriched in cluster A(ii) (Supplementary Fig. 2B, <i>P</i> < 0.001). The core UC 5296 gene principle component 1 (PC1) values separated Ctl from UC across both clinical and endoscopic severity (Supplementary Fig. 2C, D), while PC2 contributed to separation within UC severity. In all, 106 genes were significantly differentially expressed between severe vs. moderate and between moderate vs. mild UC clinical disease defined by PUCAI, showing stepwise alteration across cases (Supplementary Fig. 2E). We identified 916 genes that differed between UC with severe vs. mild clinical disease and 1038 genes that differed between severe vs. mild endoscopic subscore (FDR < 0.001 and FC ≥ 1.5, Supplementary Dataset 2). The Venn diagram (Fig. 3a) shows the overlap of the resulting 712 genes (292 down- and 420 upregulated genes) and the core UC signature, referred to hereafter as the UC severity signature. Functional annotation enrichment analyses of the UC severity signature (Supplementary Dataset 2, Fig. 3b) emphasized genes that are down- (<i>P</i> < 4.5E-46) and upregulated (<i>P</i> < 7.62E-51) in colorectal adenoma. Immunohistochemistry (Supplementary Fig. 3) confirmed increased epithelial abundance of REG1A gene, known to be upregulated in both UC and in colitis-associated colorectal cancer (CAC)<sup>18</sup> in active UC. In addition, upregulated severity genes were also enriched for innate immunity (<i>P</i> < 7.07E-19), neutrophil degranulation (<i>P</i> < 1.51E-16), and CXCR1 interactions (<i>P</i> < 9.08E-8). Relative composition of immune cell subsets using a computational gene expression deconvolution approach showed an increase in activated DC, plasma cells, and monocytes in patients with severe vs. mild disease (Fig. 3c). An alternative analytic approach using the Immunological Genome Project data series as a reference through ToppGene<sup>6</sup> also identified an increased proportion of myeloid cells with increased severity (Supplementary Fig. 4).

**Rectal genes correlated with histologic features.** Rectal biopsy histology was evaluated centrally. Surface villiform architectural abnormality (Fig. 3d) was linked to escalation therapy or colectomy in our recent report<sup>32</sup>. We identified 187 genes (69 up- and 118 downregulated, Supplementary Dataset 3) that differed
(FDR < 0.001 and FC ≥ 1.5) between UC patients with and without surface villiform changes. Most of these genes overlapped with the 712 UC severity genes (Fig. 3e) suggesting a molecular link between this histologic feature and UC severity (Supplementary Dataset 3). In contrast, higher eosinophil infiltrate (>32 rectal eosinophils/hpf, Fig. 3f) was associated with a favorable week 12 outcome3,5. Three genes differed significantly (FDR < 0.001 and FC ≥ 1.5) between UC patients with and without higher infiltrating eosinophils (Fig. 3f, g). This included Arachidonate 15-Lipoxygenase (ALOX15) involved in production of lipid
mediators which resolve inflammation. A Histologic Severity Score (HSS) for chronic and active acute neutrophil inflammation was defined3 (Supplementary Table 2). While we noted a higher frequency of patients with moderate-severe disease showing marked acute inflammation withcrypt abscesses (grade 3) histology than the frequency noted within patients with mild disease (Fig. 3h), no such difference was noted within moderate-severe patients that did or did not achieve week 4 (WK4) remission (Fig. 3i).

Corticosteroid response gene signature and microbial shifts. In the full cohort, the strongest predictor of corticosteroid-free remission by week 12 was clinical remission at week 4 (WK4), irrespective of initial corticosteroid status3. When considering WK4 remission, clinical factors associated with this outcome included disease severity and rectal biopsy eosinophil count. Based on these results, we focused our analysis on the WK4 outcome of moderate-severe patients that received corticosteroids. A corticosteroid response gene signature composed of 115 differentially expressed genes (FDR < 0.05 and FC ≥ 1.5) in baseline rectal biopsies between moderate-severe UC patients who did or did not achieve WK4 remission was defined (Fig. 4, Supplementary Dataset 4, and Supplementary Fig. 5). PCA PC1 values summarized variation in the corticosteroid response gene signature which was differentially expressed based on Week 4 clinical remission (R vs. NoR, Fig. 4a), and week 4 mucosal healing defined as fecal calprotectin <250 mcg/gm (Fig. 4b) in the Illumina discovery cohort. Healthy controls showing lower scores, implying that patients destined to respond to CS have a more healthy profile with respect to this gene signature at baseline. The corticosteroid response gene signature PC1 was replicated using the Lexogen platform19 in the subset of 134 UC patients with Illumina data, as well an independent subcohort of 50 UC patients that were not included in the original analysis (Fig. 4c, d).

As there are no mucosal transcriptomic studies that examined response to standardized initial corticosteroid induction therapy, as we tested previous transcriptomic studies that examined anti-TNF (GSE1687920, used by refs. 21,22) or anti-integrin α4β7 (GSE7366123) response. We noted a similar difference with anti-TNF or anti-integrin α4β7 response in adult UC, in this case defined by mucosal healing at colonoscopy (Fig. 4c, f).

Interestingly, Oncostatin M (OSM)21 and TREM122 previously associated with anti-TNF response were within our corticosteroid response gene signature (Fig. 4g), and this signature PC1 showed a high correlation with OSM and TREM1 (0.79 and 0.89, P < 0.0001). We also noted a substantial overlap between the genes from the PROTECT corticosteroid response gene signature and previously described anti-TNF response genes21 (Fig. 4g). Functional annotation enrichment analyses of the corticosteroid response gene signature were performed and the full output from ToppGene (Supplementary Dataset 4) with more detailed ToppCluster output is shown in Fig. 4g. Those analyses indicated that this signature is highly associated with chemokines including CXCR (P < 7.12E-12), innate myeloid immune signatures (P < 1.62E-15), and response to bacteria (P < 2.16E-13). Aberrant immune responses to shifts in commensal microbes likely play a role in UC pathogenesis and treatment responses. In all, 152 of the 206 UC patients in our cohort also had fecal 16S RNA microbial profiles24. By applying hierarchical all-against-all association testing (HAILA: http://huttenhower.sph.harvard.edu/haila), we identified genes and pathways associated with specific microbial operational taxonomic units (OTUs), including associations between disease severity-associated taxa such as Campylobacter, Veillonella, and Enterococcus with genes and pathways linked to a more severe disease form, and refractory disease in connection with initial corticosteroid induction therapy. In contrast, we identified decreased taxa from the Clostridiales order that are considered beneficial, which show a negative correlation with gene signatures associated with disease severity and unfavorable treatment responses (Fig. 4h and Supplementary Dataset 5).

A gene signature improves prediction of week 4 remission. We next asked whether gene expression data would improve a multivariable regression WK4 prediction model based on clinical factors alone (Table 2 and Supplementary Table 3). A model that included (Table 2, model 1) sex, disease severity (total Mayo clinical and endoscopic severity score), and histologic characterization of rectal eosinophils agreed with the model for the full cohort, adding sex with borderline significance. The corticosteroid response gene signature PC1 was negatively associated with week 4 outcome (model 2, OR 0.36, 95% CI 0.18–0.71; P = 0.003). When this gene signature was included, the AUC improved to 0.774 (Likelihood ratio P value < 0.002), indicating superiority to the model which included clinical factors alone. In model 3, we replaced the eosinophil count with the eosinophil-associated gene ALOX15 without harming the model accuracy with some improvement of the discriminant power (AUC of 0.777, 0.692–0.848), sensitivity of 62.7%, (95% CI 52.8–72.5%), specificity of 76.6% (95% CI 0.68–84.4%), positive predictive value of 72.3%, and negative predictive value of 67.8% (AUC cutoff at ≥0.5). Bootstrapping and multiple imputation were used for internal validation and were generally supportive of the final selected moderate/severe model (Supplementary Fig. 6). The HSS showed moderate correlation with the corticosteroid response gene signature PC1 (Spearman r = 0.31, P < 0.001), but not with WK4 outcome. Moreover, the gene signature was still significant in the model even after adjusting for the HSS (Supplementary Table 3). Similarly, while the monocyte deconvolution score showed high correlation with the corticosteroid response gene signature PC1 (Pearson r = 0.72, P < 0.001) and was different between WK4 responders and nonresponders (Supplementary
Fig. 5), it was not significant when added to the model in place of the gene signature, while the gene signature remained significant in the model after adjusting for the monocyte score (Supplementary Table 3).

Discussion

PROTECT is the largest prospective inception cohort study to examine factors associated with early responses to standardized first-line therapy in pediatric UC. We provide evidence for core
host gene expression profiles driving lymphocyte activation and cytokine signaling which are targeted by current therapies. However, our data also suggest a robust reduction in epithelial mitochondrial genes and associated energy production pathways in UC, not directly addressed by current approaches. This reduction of mitochondrial genes was validated in treatment-naïve pediatric UC, adults with active UC with longstanding disease, and more specifically in viable isolated epithelia of treatment-naïve pediatric UC. We capture genes and pathways that are linked to UC severity and prioritize those regulating epithelial transformation and innate CXCR-mediated leukocyte recruitment. We identified a gene signature linked to corticosteroid response, which was validated in an independent subset of UC patients, and showed substantial overlap with genes previously associated with anti-TNF response. A multivariable analysis combining the corticosteroid response gene signature with clinical variables better predicted corticosteroid responsiveness than clinical factors alone. These findings are summarized in Fig. 4i.

Decreased mitochondrial activity\textsuperscript{25,26} was previously described in UC, but understanding of the molecular mechanism was lacking\textsuperscript{27–30}. Dysfunctional mitochondria exacerbate barrier dysfunction and inflammation, while pro-\textsuperscript{29} and anti-\textsuperscript{30} inflammatory stimuli affect mitochondrial metabolic functions. \textit{PPARGC1A} (PGC1α), the master regulator of mitochondrial biogenesis, ameliorated experimental colitis, whereby intestinal epithelial depletion of PGC1α suppressed mitochondrial function and the intestinal barrier\textsuperscript{31}. Mitochondrial loss also preceded the development of colonic dysplasia in UC\textsuperscript{32}, and high mitochondrial activity reflecting electron transport in the ileum was also associated with protection against CD progression in RISK\textsuperscript{33}. Our findings of a substantial suppression of all 13 electron transport mitochondrial-encoded genes (Complex I, III, IV, and V), \textit{PPARGC1A} (PGC1α), and epithelial MMP further support the robustness of the colonic mitochondrial pathway in UC. Moreover, we demonstrate specificity of mitochondrial gene expression downregulation in colon-only forms of IBD rather than in CD patients with both ileal and colonic inflammation. Interestingly, previous studies in infectious colitis\textsuperscript{34} or diverticulitis\textsuperscript{35} demonstrated an induction of immune and wound-healing genes, with considerable overlap with the immune and wound-healing genes identified in pediatric UC for the current report. However, these studies did not demonstrate a similar reduction in mitochondrial genes, suggesting specificity of this response in UC. Functionally, we observe a decrease in the activity of Complex I of the electron transport chain in the inflamed rectums of patients with UC, and a reduction of mitochondrial depolarization more specifically in epithelia. Although a defect in respiration has been observed in the colons of UC patients previously, to our knowledge mitochondrial function from intestinal biopsies has never before been evaluated via high-resolution respirometry.

With real-time analysis of intact human tissue, this technique offers precise evaluation of mitochondrial membrane integrity and oxidative capacity. In conjunction with our expression data, these results suggest a downregulation and dysfunction of mitochondrial respiration, characterized by a defect at Complex I, the rate-limiting step in oxidative phosphorylation. Supplementing the mitochondrial electron transport axis via medical, environmental, or nutritional approaches can be potential targets for future therapies.

Inflammation has a substantial cumulative role in colitis-associated colorectal cancer (CA-CRC) development and is closely linked to the extent\textsuperscript{36}, duration\textsuperscript{37} and severity\textsuperscript{38}. Studies in the noncancerous IBD mucosa indicated that colorectal cancer development in IBD begins many years before the development of neoplasia as part of the occult evolution within the inflamed bowel\textsuperscript{39}. Here, we detect a profound dysregulation of gene sets associated with disease severity previously implicated in adenocarcinoma. Our results therefore show that not only at the genomic and epigenetic level\textsuperscript{39,40}, but also at the transcriptomic level, already at diagnosis, genes and pathways that are associated with UC severity show associations with epithelial transformation.

Microbial organisms and products affect host immune education, development and response, and aberrant immune responses to commensal microbes likely contribute to gut inflammation which is the hallmark of UC\textsuperscript{41}. We identified positive associations between genes and pathways associated with UC severity and response to treatment and disease-linked microbial taxa. Negative associations involved more beneficial commensal taxa with pathways and genes that were linked to resolution of inflammation or upregulated in non-IBD controls. Those included oral pathobionts \textit{Veillonella dispar}, and \textit{Campylobacter}, and depletion of several commensal organisms such as \textit{Lachnospiraceae}, \textit{Bifidobacterium}, and \textit{Ruminococcaceae} suggesting a substantial depletion of SCFA-producing bacteria that may affect epithelial barrier function\textsuperscript{42}. Those associations will need to be validated in an independent cohort.

In our study\textsuperscript{3} and in previous studies in children and adults\textsuperscript{4,43,44}, higher baseline disease severity identified patients less likely to achieve remission with corticosteroids. We supplement and improve those models by adding baseline gene expression data. We identified a gene signature linked to corticosteroid response and validated it in an independent subset of UC patients. The corticosteroid response gene signature is enriched for cytokines and chemokines (CXCR1/2 and CXCL6/8/10/11/17), which promote activation of the innate immune system and recruitment of neutrophils, and to response to external stimuli and bacteria. Notably, the corticosteroid response gene signature showed a substantial overlap with genes previously associated with anti-TNF response, and exhibited a similar difference between responders and nonresponders to anti-TNF or anti-integrin α\textsubscript{4}β\textsubscript{2} therapies. These similarities support an emerging concept in the field that the mucosal inflammatory state as
A rectal gene signature is associated with response to UC induction therapy and microbial shifts. Samples loading PC1 (Z score) values of the corticosteroid response gene signature are shown for controls and the discovery cohort of 152 moderate-severe UC patients stratified by corticosteroid response. The full list of enriched functions and the top 50 genes that were differentially expressed in pretreatment colon biopsies of anti-TNF refractory vs. responsive UC patients. The full list of enriched functions and P values are in Supplementary Dataset 4. Genes are in blue and biologic functions in pink; connections to each signature are as shown. Heat map summarizing Spearman similarity measures between microbial abundances and gene expression using hierarchical all-against-all association. *False discovery rate < 0.2. Blue and red indicate negative and positive associations, respectively. Graphical summary of the cohort and main findings. Box and whisker plot with central line indicating median, box ends representing upper and lower quartile, and whisker represent 10–90 percentile. Mann–Whitney and ANOVA FDR as applicable, *P < 0.05, **P < 0.01, ***P < 0.001, ****P < 0.0001.
measured by gene expression may better define the likelihood of response to current treatment approaches than conventional clinical measures of severity. By comparison, higher ALOX15 expression was linked to a higher likelihood for remission. Increasing evidence suggests a role for ALOX15 in UC and for later disease outcomes. A validated corticosteroid response gene signature and higher anti-inflammatory ALOX15 expression are associated with higher odds of achieving early clinical remission, with remarkable overlap with genes implicated in response to biologics. A shift to personalized approaches targeting specific mechanisms in individual patients will be key to reducing the increasing disease burden of UC worldwide.

### Methods

#### Study design and participants.

PROTECT was a multicenter inception cohort study based at 29 centers in the USA and Canada. Children aged 4–17 years with a diagnosis of UC based on accepted clinical, endoscopic, and histological parameters, disease extent beyond the rectum, a baseline Pediatric Ulcerative Colitis Activity Index (PUCAI) score of at least 10, no previous therapy for colitis, and stool culture negative for enteric bacterial pathogens and *Clostridium difficile* toxin were included. Informed consent or assent was obtained in all cases and the study was approved by the local investigational review board at all investigative sites. Detailed protocol and study description can be found in Hyams et al. Disease extent was classified as proctosigmoiditis, left-sided colitis (to the splenic flexure), extensive colitis (to the hepatic flexure), or pancolitis (beyond the hepatic flexure) by visual evidence. Patients with severe or fulminant disease at presentation who refused flexible sigmoidoscopy because of safety concerns were assigned to the flexible sigmoidoscopy group (unassessable). Clinical activity at diagnosis was established with the PUCAI (range 0–120). PUCAI less than 10 denoted inactive disease or first remission, 10–30 denoted mild disease, 30–60 denoted moderate disease, and 60–120 denoted severe disease. A linear mixed-effects model was used to compare disease activity of patients at diagnosis with the presence of ALOX15 expression. The results of the model are presented in Table 2.

### Table 2 Multivariable models of baseline characteristics and gene expression associated with week 4 remission in 147 patients with moderate-severe disease that received corticosteroids

| Model # | Model variables | OR (95% CI) | Variable P | Model AIC | Model AUC | Model ChiSq | Model P |
|---------|----------------|-------------|------------|-----------|-----------|-------------|---------|
| 1       | Total Mayo score (range 0–12) | 0.68 (0.54–0.85) | 0.0007 | 186.03 | 73.7 | 25.75 | <0.0001 |
|         | Rectal eosinophil level (count > 32 /hpf) | 2.27 (1.11–4.63) | 0.039 | | | | |
|         | Sex (M vs. F) | 0.47 (0.23–0.96) | | | | | |
| 2       | Total Mayo score (range 0–12) | 0.77 (0.61–0.98) | 0.32 | 178.51 | 77.4 | 35.27 | <0.0001 |
|         | Rectal eosinophil level (count > 32 /hpf) | 1.81 | 0.122 | | | | |
|         | Sex (M vs. F) | 0.85–3.84 | 0.003 | | | | |
|         | Corticosteroid response gene signature (PC1 z-score values) | 0.47 (0.22–0.99) | | | | | |
|         | | 0.36 (0.18–0.71) | | | | | |
| 3       | Total Mayo score (range 0–12) | 0.79 | 0.055 | 192.98 | 77.7 | 40.80 | <0.0001 |
|         | ALOX15 gene exp. (TPM) | 0.63–1.00 | 0.014 | | | | |
|         | Sex (M vs. F) | 2.59 | 0.038 | | | | |
|         | Corticosteroid response gene signature (PC1 z-score values) | 0.12–5.52 | 0.009 | | | | |
|         | | 0.45 (0.21–0.96) | | | | | |
|         | | 0.40 (0.2–0.79) | | | | | |

LR = 9.519 and LR P value = 0.002 when comparing model 2 to model 1. OR odds ratio, AIC Akaike's information criterion, AUC area under the ROC curve, LR likelihood ratio, ROC receiver operator characteristic
Hospit‌al Medical Center patients were enrolled under another IBB approved protocol, and were included in the current analyses as non-IBD controls after clinical endoscopic, and biopsies evaluation demonstrated no histologic and endoscopic inflammation. Rectal mucosal biopsies from a representative subcohort of 206 PROTECT UC patients and 20 age and gender matched non-IBD controls underwent high coverage transcriptomic profiling using Illumina RNAseq (Table 1). These constituted the Discovery cohort for the current study.

The representative subcohort for RNAseq was defined by having a baseline rectal biopsy available to be included in the RNAseq analysis, and must also have the following data available in order to be assigned to the appropriate clinical subgroup: baseline PUCAI, medication data including the need for rescue or colectomy, a baseline rectal biopsy and a week 4 PUCAI. Four patient PUCAI required a rescue or a colectomy during the first 4 weeks. The following PROTECT participants were not eligible for the RNAseq analysis: patients with a diagnosis other than UC after enrollment, patients with significant baseline violations, patients who took rescue medications for a non-UC reason within the first 4 weeks, baseline rectal biopsy not available, race or ethnicity not identified as “Black”, “Native American” or “Unknown”, baseline PUCAI < 35 but did not start on mesalamine as first therapy, baseline PUCAI ≥ 35 but did not start on corticosteroids as first therapy. A total of 219 were selected, and data for 206 were ultimately available, after excluding 5 subjects based on the RNAseq data as described below, and 8 with insufficient RNA.

Rectal RNA extraction and RNAseq analysis. RNA was isolated from rectal biopsies obtained during diagnostic colonoscopy using the Qiagen AllPrep RNA/DNA Mini Kit. PolyA RNA selection, fragmentation, cDNA synthesis, adapter ligation, TruSeq RNA sample library preparation (Illumina, San Diego, CA), and paired-end 75 bp sequencing was performed. An additional validation of the baseline rectal gene expression at diagnosis utilized the independent RISK cohort of treatment-naive pediatric patients (55 non-IBD controls, 43 UC patients, and 92 CD patients with rectal inflammation) and single-end 75 bp mRNA sequencing. Reads were quantified by kallisto23, using Genode v24 as the reference genome and transcripts per million (TPM) as an output. We included 14,085 protein-coding mRNA genes with TPM above 1 in 20% of the samples in our downstream analysis. Only samples for which the gene expression (Y encoded genes and XIST)-determined gender matched the clinical-reported gender were included in the analyses (we excluded one sample with unmatched gender). Four other PROTECT samples were excluded due to poor read quality. A total of 226 RNAseq samples with mean read depth of ~47 M (14 M std. deviation) were stratified into specific clinical subgroups including CI (n = 20), and UC (n = 206), and were stratified based on disease severity, and on histologic findings. Differentially expressed genes were determined in GeneSpring24 software with fold change differences (FC) ≥ 1.5 and using the Benjamini–Hochberg false discovery rate correction (FDR, 0.001) for all analyses except for the corticosteroid response genes that was calculated out of the 712 severity genes with FDR < 0.05. Unsupervised hierarchical clustering using Euclidean distance metric and Ward’s linkage rule was used to test for groups of rectal biopsies with similar patterns of gene expressions. ToppGene25 and ToppCluster26 software were used to test for functional annotation enrichment analyses of immune cell types, pathways, phenotype, immune cell-type enrichments, and biologic functions. Visualization of the network was obtained using CytoScape.27 For validation of the association between baseline gene expression and outcome, we also generated independent Lexogen QuantSeq 3′ mRNAseq libraries15 and single-end 100 bp sequencing for 134 participants who also had Illumina mRNAseq data (the Discovery Cohort) and for 50 participants who did not have Illumina mRNAseq data (the independent Validation cohort, Table 1). PCA was performed to summarize variation in gene expression between patients, and principal components (PC) values were extracted for downstream analyses. We considered several central gene expression pathways CI preidentified by the previous differential expression analyses and functional annotation enrichment analyses of the core 9269 UC genes, the 712 severity genes, and the 15 corticosteroid response signature for building and associations with the microbial composition as described below. PROTECT (GSE109142) and RISK (GSE117993) rectal mRNAseq data sets were deposited into GEO.

Analyses of microarrays. We obtained colon biopsy gene expression data and patient clinical data from published studies available in Gene Expression Omnibus (GEO) as summarized in Suppl. Table 1. The Affymetrix raw gene array data (.CEL files) were processed to obtain a log2 expression value for each gene probe set using the robust multichip average (RMA) method implemented in R; the Affymetrix GeneChip Human Genome U133 Plus 2.0 Arrays were processed in R with the affy package, and the Affymetrix Human U133 2.0 and the Human U133 Plus 2.0 ST arrays were processed with the oligo package (v1.42.0). For comparative analysis, the LIMMA package28 was used to identify the filtered gene probe sets that showed significant differential expression between the studied groups, based on moderated t-statistics with Benjamin–Hochberg FDR correction for multiple testing. Gene probe sets were then further biologically significant by testing for an overall change (FC) ≥ 1.5. When genes in microarray data were represented by multiple probes, the probe with the greatest interquartile range was selected for analysis. PCA was performed on the normalized log2 microarray data of control and UC samples and PC1 values were calculated.

Microbiome analyses. DNA was extracted from PROTECT UC stool samples and subjected to 16S rRNA amplicon sequencing. OTU clustering and taxonomic assignment was performed24 (NCBI SRA Bioproject: PRJNA436359). Briefly, for the OTU analysis the 16S bioBakery workflow built with AnaDAMA24 was applied and microbial taxonomy was based on the GreenGenes 16S rDNA database (version 13.8). Samples were subsequently filtered (min 3000 reads and OTU prevalence threshold of 20 samples). Statistical significance was established using hierarchical all-against-all association testing (HALLA: http://buttenhoffer.sph.harvard.edu/halla) in all-against-all mode using Spearman as the similarity measure and a cutoff of 0.2 for the FDR. Overall, 156 PROTECT stools at baseline were available that also had mRNAseq data. In total, 149 OTUs were significantly associated with 9 genes, and 15 pathways, with 36 below FDR 0.1 (Supplementary Dataset 5). Overall, only 28 RISK CD cases and 21 PROTECT Lexogen UC validation cohort cases had both fecal microbial profiling and rectal mRNAseq data, providing insufficient power for validation of these results.

Computational deconvolution. To estimate cell subset proportions, we performed a cell-type deconvolution. We utilized xCell2, a computational method that is able to infer 64 variable cell types (e.g., immune cell types, epithelial, and stroma cells) using gene signatures. To ensure robustness of our downstream analyses, we considered only cell types that had significant enrichment scores (FDR corrected P values < 0.1 in at least 80% of the samples). We calculated the significance using two-tailed t-tests, and took into account cell types that were filtered into at least one of them. The first includes randomization of the genes in the signatures used for generating the enrichment scores and the second includes using simulations where the tested cell type is not included in the mixture35. Cell signatures were considered but did not vary significantly between samples. We identified the following cell types: gut-associated lymphoid tissue (GALT), conventional DC, DC, memory B cells, plasma cells, Th1 cells, and monocytes. The scores of active DC and DC as well as B cells and “Memory B cells” across samples were positively and highly correlated (Supplementary Fig. 7) and we consider the more specific and biologically relevant activated DC and memory B cells. Astrocyte cell type was removed from the calculation.

High-resolution respirometry. The Oxxygraph-2k (O2k, Oroboros Instruments, Innsbruck, Austria) was used for measurements of respiration. Each chamber was air-calibrated in Mir05 respiration medium (0.5 mM Ethylenediaminetetraacetic acid (EDTA), 3 mM MgCl2, 60 mM k-lactobionic acid, 20 mM taunine, 10 mM KH2PO4, 20 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), 110 mM Na-glucose, 0.1% Bovine serum albumin (BSA) essentially fatty acid free) before each experiment. All experiments were performed at 37 °C. Oxygen concentrations in each chamber never dropped below 80 μM during any experiment. Patient biopsies were taken from the cecum and rectum in both control patients (N = 5) and patients with UC (N = 9). Cecal and rectal biopsies were homogenized in 1 ml of respiration medium, and 100 μl of the tissue homogenate was added to each chamber. Once baseline oxygen levels in each chamber became stable, cytochrome c (10 μM), malate (2 mM), pyruvate (5 mM), Adenosine diphosphate (ADP) (5 mM), and glutamate (10 mM) were added to stimulate respiration through Complex I. Once the oxygen consumption rate plateaued, succinate (110 mM) was added to assess the respiratory capacity of Complex II. 1 mM rotenone (1 mM) was added to inhibit Complex I activity, and additional succinate was added to analyze maximal Complex II activity. Carboxyl cyanide p-trifluoromethoxyphenylhydrazone (0.5 μM) was then added to uncouple the mitochondrial membrane and induce maximal respiratory. Respiration rates were normalized to the amount of protein added for each sample. Complex I respiration was defined as the rate of respiration of malate/ADP/pyruvate/glutamate (first succinate–rotenone). Complex II respiration was defined as respiration after adding the second dose of succinate minus Complex I respiration. Average rates of oxygen consumption [pmol/(s×ml)/μg protein] + standard error of the mean (SEM) were graphed.

Cold enzyme biopsy prep to generate single cells. Colon biopsies were minced in a Petri dish on ice in the presence of Native Bacillus Licheniformis psychrophilic proteases at 1 mg/ml (Creative Enzymes, Shirley, NY), transferred to an Eppendorf tube, intermittently vortexed for 30–60 s, placed on ice, and gently pipetted over 15 min. The suspension was centrifuged at 90 × g and the supernatant filtered over a 40 μm filter. Additional enzyme was added to reach tissue slurry and the procedure repeated for an additional 15 min. Cells were counted with trypan blue and 85–99% viability was noted.

JCl mitochondrial membrane potential measurement. JCl staining was performed on the above single-cell isolations with flow cytometry using the JC-1 (5,5′,6,6′-tetrachloro-1,1′,3,3′-tetrathiethylbenzimidazolylcarbocyanine iodide, Molecular Probes, Inc. Eugene, OR) and Meloy EDR C staining kit (CytoViva, Microvision, USA) as instructed. In brief, J-1 dye was added at 1 mcM to washed cells, and incubated for 20 min at 37 °C. 5% CO2. Cells were washed and CD45 APC-Cy7 (BD
Bioscience, Franklin Lakes, NJ) and EpCAM APC (BioLegend, San Diego, CA) antibodies were added for an additional 30 min at room temperature. Cells were washed, acquired on a Canto flow cytometer, and data were analyzed using DeNovo software. The MMP was calculated as the ratio of PE-MFI/FTC-MFI in EpCAM+ and CD45+ cells. Representative data are shown in Supplemental Fig. 1. As a positive control for the specificity of the assay we used 50 nM of CCCP (carbonyl cyanide 3-chlorophenylhydrazone) to depolarize the MMP measured using the JC-1 dye.

**Immunohistochemistry.** Immunohistochemistry detection of MT-CO1, COX5A, and REG1A was performed32,58, using anti-Complex IV subunit I (Thermo Fisher using the JC-1 dye. As a positive control for the specific

**DeNovo software.** The MMP was calculated as the ratio of PE-MFI/FITC-MFI in

**ARTICLE NATURE COMMUNICATIONS | https://doi.org/10.1038/s41467-018-07841-3 | www.nature.com/naturecommunications**

**1. Kaplan, G. G. & Ng, S. C. Understanding and preventing the global increase of inflammatory bowel disease. Gastroenterology 152, 313–321, e312 (2017).**

**2. Peery, A. F. et al. Burden of gastrointestinal disease in the United States: 2012 update. Gastroenterology 143, 1179–1187, e1171-e1173 (2012).**

**3. Hyams, J. S. et al. Factors associated with early outcomes following standardised therapy in children with ulcerative colitis (PROTECT): a multicentre inception cohort study. Lancet Gastroenterol. Hepatol. 2, 855–868 (2017).**

**4. Hyams, J. S. et al. Clinical outcome of ulcerative colitis in children. J. Pediatr. 129, 81–88 (1996).**

**5. Boyle, B. et al. Histologic correlates of clinical and endoscopic disease in children newly diagnosed with ulcerative colitis. Am. J. Surg. Pathol. 41, 1491–1498 (2017).**

**6. Chen, J., Bardes, E. E., Aronow, B. J. & Jegga, A. G. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. Nucleic Acids Res. 37, W305–W311 (2009).**

**7. Kaimal, V., Bardes, E. E., Tabar, S. C., Jegga, A. G. & Aronow, B. J. ToppCluster: a multiple gene feature analyzer for comparative enrichment clustering and network-based dissection of biological systems. Nucleic Acids Res. 38, W96–W102 (2010).**

**8. Bindea, G. et al. ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks. Bioinformatics 25, 1091–1093 (2009).**

**9. Haberman, Y. et al. Pediatric Crohn disease patients exhibit specific ideal transcriptome and microbiome signature. J. Clin. Invest. 124, 3617–3633 (2014).**

**10. Flamant, M., Rigal, J., Paul, S. & Robin, X. Advances in the development of Janus kinase inhibitors in inflammatory bowel disease: future prospects. Drugs 77, 1057–1068 (2017).**

**11. Abraham, C., Dulai, P. S., Vermeire, S. & Sandborn, W. J. Lessons learned from trials targeting cytokine pathways in patients with inflammatory bowel diseases. Gastroenterology 152, 374–388, e574 (2017).**

**12. Vanhove, W. et al. Strong upregulation of AIM2 and IFI16 inflammasomes in the mucosa of patients with active inflammatory bowel disease. Inflamm. Bowel Dis. 21, 2673–2682 (2015).**

**13. Howell, K. J. et al. DNA methylation and transcription patterns in intestinal epithelial cells from pediatric patients with inflammatory bowel diseases differentiate disease subtypes and associate with outcome. Gastroenterology 154, 585–598 (2018).**

**14. Pesta, D. & Gnaiger, E. High-resolution respirometry: OXPHOS protocols for human cells and permeabilized fibers from small biopsies of human muscle. Methods Mol. Biol. 810, 25–58 (2012).**

**15. Zielinski, L. P., Smith, A. C., Smith, A. G. & Robinson, A. J. Metabolic flexibility of mitochondrial respiratory chain disorders predicted by computer modelling. Mitochondrion 31, 45–55 (2016).**

**16. Hroudova, J. & Fisar, Z. Control mechanisms in mitochondrial oxidative phosphorylation. Neural Regen. Res. 8, 363–375 (2013).**

**17. Turner, D. et al. Severe pediatric ulcerative colitis: a prospective multicenter study of outcomes and predictors of response. Gastroenterology 138, 2228–2231 (2010).**

**18. Zhao, W. et al. Functional comparison between genes dysregulated in ulcerative colitis and colorectal carcinoma. PLoS ONE 8, e71989 (2013).**

**19. Turker, A., Wiktorin, C. & Guler, S. Mixture models reveal multiple positional bias types in RNA-Seq data and lead to accurate transcript concentration estimates. PLoS Comput. Biol. 13, e1005515 (2017).**

**20. Arjis, I. et al. Murusco gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infiximab treatment. PLoS ONE 4, e7004 (2009).**

**21. West, N. R. et al. Oncostatin M drives intestinal inflammation and predicts response to tumor necrosis factor-neutralizing therapy in patients with inflammatory bowel disease. Nat. Med. 23, 579–589 (2017).**

**22. Gaujoux, R. et al. Cell-centred meta-analysis reveals baseline predictors of anti-TNFAlpha non-response in biopsy and blood of patients with IBD. Gut https://doi.org/10.1136/gutjnl-2017-315949 (2018).**

**23. Arjis, I. et al. Effect of vedolizumab (anti-alpha4beta7-integrin) therapy on histological healing and mucosal gene expression in patients with UC. Gut 67, 43–52 (2018).**

**24. Schirmer, M. et al. Compositional and temporal changes in the gut microbiome of pediatric ulcerative colitis patients are linked to disease course. Cell Host Microbe 24, 600–610, e614 (2018).**

**25. Biferno, A. et al. Mitochondrial respiratory chain complex II-CIII-subunit complex in the colonic mucosa of patients with ulcerative colitis. Mol. Cell. Biochem. 342, 111–115 (2010).**

**26. Santhanam, S. et al. Mitochondrial electron transport chain complex II-CIII-subunit complex in the colonic mucosa of patients with ulcerative colitis. Mol. Cell. Biochem. 342, 111–115 (2010).**

**27. Mottawea, W. et al. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn’s disease. Nat. Commun. 7, 13419 (2016).**

**28. Cardinale, C. J. et al. Transcriptome profiling of human ulcerative colitis mucosa reveals altered expression of pathways enriched in genetic susceptibility loci. PLoS ONE 9, e96153 (2014).**

**29. Marson-McDermott, E. W. et al. Inflammatory bowel disease. Gastroenterology 152, 855–868 (2017).**

**30. Santhanam, S. et al. Mitochondrial respiratory chain complex II-CIII-subunit complex in the colonic mucosa of patients with ulcerative colitis. Mol. Cell. Biochem. 342, 111–115 (2010).**

**31. Mottawea, W. et al. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn’s disease. Nat. Commun. 7, 13419 (2016).**

**32. Cardinale, C. J. et al. Transcriptome profiling of human ulcerative colitis mucosa reveals altered expression of pathways enriched in genetic susceptibility loci. PLoS ONE 9, e96153 (2014).**

**33. Marson-McDermott, E. W. et al. Inflammatory bowel disease. Gastroenterology 152, 855–868 (2017).**

**34. Santhanam, S. et al. Mitochondrial respiratory chain complex II-CIII-subunit complex in the colonic mucosa of patients with ulcerative colitis. Mol. Cell. Biochem. 342, 111–115 (2010).**

**35. Mottawea, W. et al. Altered intestinal microbiota-host mitochondria crosstalk in new onset Crohn’s disease. Nat. Commun. 7, 13419 (2016).**
34. Peterson, K. M. et al. The expression of REG 1A and REG 1B is increased
35. Schieffer, K. M. et al. RNA-seq implicates deregulation of the immune system
in the pathogenesis of diverticulitis. Am. J. Physiol. Gastrointest. Liver Physiol. 313, G277–G284 (2017).
36. Ekborn, A., Helmač, C., Zack, M. & Adami, H. O. Ulcerative colitis and colorectal cancer. A population-based study. N. Engl. J. Med. 323, 1228–1233 (1990).
37. Rutter, M. D. et al. Thirty-year analysis of a colonoscopic surveillance program for neoplasia in ulcerative colitis. Gastroenterology 130, 1030–1038 (2006).
38. Choi, C. R., Bakır, I. A., Hart, A. L. & Graham, T. A. Clonal evolution of colorectal cancer in IBD. Nature. Rev. Gastroenterol. Hepatol. 14, 218–229 (2017).
39. Leedham, S. J. et al. Clonality, founder mutations, and field cancerization in human ulcerative colitis-associated neoplasia. Gastroenterology 136, 542–550, e546 (2009).
40. Sartor, R. B. & Wu, G. D. Roles for intestinal bacteria, viruses, and fungi in pathogenesis of inflammatory bowel diseases and therapeutic approaches. Gastroenterology 147, 1055–1070, 1070–1081 (2014).
41. Kelly, C. J. et al. Crosstalk between microbiota-derived short-chain fatty acids and intestinal epithelial HIF augments tissue barrier function. Cell Host Microbe 17, 662–671 (2015).
42. Romberg-Camps, M. J. et al. Influence of phenotype at diagnosis and of other potential prognostic factors on the course of inflammatory bowel disease. Am. J. Gastroenterol. 104, 371–383 (2009).
43. Moore, J. C. et al. Clinical variables as prognostic tools in pediatric-onset ulcerative colitis: a retrospective cohort study. Inflamm. Bowel Dis. 17, 15–21 (2011).
44. Ackermann, J. A., Hofheinz, K., Zaiss, M. M. & Kronke, G. The double-edged role of 12/15-lipoxygenase during inflammation and immunity. Biochim. Biophys. Acta 1862, 371–381 (2017).
45. Chan, M. M. & Moore, A. R. Resolution of inflammation in murine autoimmune arthritis is disrupted by cyclooxygenase-2 inhibition and restored by prostaglandin E2-mediated lipoxin A4 production. J. Immunol. 184, 6418–6426 (2010).
46. Storkes, K. et al. A role for 12/15-lipoxygenase-derived proresolving mediators in postoperative ileus: protect DN-regulated neutrophil extravasation. J. Leukoc. Biol. 99, 231–239 (2016).
47. Yamada, T. et al. Eosinophils promote resolution of acute peritonitis by producing proresolving mediators in mice. Faseb J. 25, 561–568 (2011).
48. Planell, N. et al. Transcriptional analysis of the intestinal mucosa of patients with ulcerative colitis in remission reveals lasting epithelial cell alterations. Gut 62, 967–976 (2013).
49. Wu, F. et al. Genome-wide gene expression differences in Crohn’s disease and ulcerative colitis from endoscopic punch biopsies: insights into distinctive pathogenesis. Inflamm. Bowel Dis. 13, 807–821 (2007).
50. Bray, N. L., Pimentel, H., Mehi1, P. & Pachter, L. Near-optimal probabilistic RNA-seq quantification. Nat. Biotechnol. 34, 525–527 (2016).
51. Saito, R. & Yata, T. A general framework for the development of microarray data. Stat. Appl. Genet. Mol. Biol. 3, Article3 (2004).
52. McClure, J. L. et al. bioBakery: a meta’omic analysis environment. Bioinformatics 23, 1295–1297 (2007).
53. McDonald, D. et al. An improved Greengenes taxonomy with explicit ranks for ecological and evolutionary analyses of bacteria and archaea. ISME J. 6, 610–618 (2012).
54. Aran, D., Hu, Z. & Butte, A. J. Cellx: digitally portraying the tissue cellular heterogeneity landscape. Genome Biol. 18, 220 (2017).
55. Adam, M., Potter, A. S. & Potter, S. S. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. Development 144, 3625–3632 (2017).
56. Carey, R. et al. Activation of an IL-6STAT3-dependent transcriptome in pediatric-onset inflammatory bowel disease. Inflamm. Bowel Dis. 14, 446–457 (2008).

Acknowledgements

Support for this study was provided by NIDDK: DK103100, DK59075, and P30 DK078392, for Genomics Analysis and Integrative Morphology Core, NIDDK: DK101753 and DK14446, and NIDDK: DK043351 and AT009708, CBISID: DK043351, Center for Microbiome Informatics and Therapeutics at Massachusetts Institute of Technology. This work was also supported in part by the Israel Science Foundation (Y.H., grant no 908/15), the I-CORE program (Y.H., grants no. 41/11, and the ERC starting grant (Y.H., grant no. 758313). The authors thank Frank Hamilton, MD, Dana Anderson, MD, Joins Everhart, PhD, Jose Serrano, MD, PhD, and Stephen James, MD from NIDDK for their guidance, and William Faubion, MD for his role as safety monitor. The authors thank PROTECT site investigators Maria Oliva-Hemker, MD, Jennifer Strople, MD, David Ziring, MD, Stephen L. Guthery, MD, Boris Sudel, MD, Keith Benkov, MD, Prateek Wali, MD, Dedrick Moulton, MD, Jonathan Evans, MD, Michael D. Kappelman, MD, and Marla Dubinsky, MD, MD for patient recruitment and data gathering. The study investigators are deeply indebted to Shire Pharmaceuticals for providing Pentasa® for this study, to the research coordinators at the sites for their tireless attention, and to the patients and families who agreed to participate in this important study.

Author contributions

Y.H. and L.A.D. analyzed the data, and wrote the first draft of the manuscript. The steering committee (J.S.H., S.D.T., S.K., T.D.W., L.A.D.) conceived and designed the study and participated in drafting the manuscript. R.K., P.J.D., M.S., J.S., T. T.B., E.N., L.B., M.H.C., A. Mo, M.R.K., E.B., N.G., A. Marquesi, M.N., R.J.A., G.O., K.M., S.S.O., C.H., P.J.K. generated and analyzed the data and participated in drafting the manuscript. P.A.R., S.S.B., C.G. S., J.M., M.D.P., J.R.R., B.M.R., D.R.M., R.N.B., S.S., N.S.L., M.R.H., A.M.G., A.S.P., J.D.N. recruited patients, gathered data and participated in manuscript revision. All authors had access to study data and approved the decision to submit the manuscript.

Additional information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41467-018-07841-3.

Competing interests: The authors declare no competing interest associated specifically with this work. However, the following authors report the following disclosures: Sonia Davis Thomas, independent data monitoring committee, Lycera Corporation; Margaret H. Collins, research support, Shire, Receptors, Pflug, Alkook; Michael J. Rosen, advisory board, Celgene; David R. Mack, Shares in Biogenetics; Robert N. Baldassano, Advisory Board, Janssen, Consultant, Abbvie, Lilly, Receptors, Pfizer; Neal S. Leleiko, Consultant, Abbvie; Ashish S. Patel, Speakers Bureau Abbvie, Janssen; James Markowitz, Consultant, Janssen, UCB, Lilly; Anne M. Griffiths: Research support Abbvie, Consultant Abbvie, Janssen, Merck, Takeda, Speaker Abbvie, Janssen; Joel R. Rosh: Consultant, Abbvie, Janssen, UCB, Grant Funding Janssen, Abbvie; Paul A. Rufe: Consultant, Shire, Lycera, Consultant, Shire, Lypseud, Speaker, Abbvie, Research support, TechLab, Cary G. Sauer, Consultant, Abbvie; Subra Kugathasan, Consultant, Janssen, UCB; Jennifer Strople, Consultant and speaker, Abbvie; Melanie B. Heyman, Research grants Genentech, Abbvie, Sucamco, Janssen; Jeffrey S. Hyams Advisory Board, Janssen, Consultant, Abbvie, Takeda, Lilly, Boerhinger-Ingelheim, Allergen, Astra Zeneca, Receptors, Pfizer, Lee A. Denston; Grant Support, Janssen. The following authors have nothing to disclose: Yale Haberman, Rebekah Karam, Philip J. D. Jexheuser, Melanie Schirmer, Judith Somick, Ingred Jurikova, Tizpi Braun, Elizabeth Novak, Angela Mo, Bruce J. Arowo, Thomas D. Walters, Greg Gibson, Laura Bauman, CORE program (Y.H., grants no. 41/11), and the ERC starting grant (Y.H., grant no. 758313). The authors thank Frank Hamilton, MD, Dana Anderson, MD, Joins Everhart, MD, Jose Serrano, MD, PhD, and Stephen James, MD from NIDDK for their guidance, and William Faubion, MD for his role as safety monitor. The authors thank PROTECT site investigators Maria Oliva-Hemker, MD, Jennifer Strople, MD, David Ziring, MD, Stephen L. Guthery, MD, Boris Sudel, MD, Keith Benkov, MD, Prateek Wali, MD, Dedrick Moulton, MD, Jonathan Evans, MD, Michael D. Kappelman, MD, and Marla Dubinsky, MD, MD for patient recruitment and data gathering. The study investigators are deeply indebted to Shire Pharmaceuticals for providing Pentasa® for this study, to the research coordinators at the sites for their tireless attention, and to the patients and families who agreed to participate in this important study.

Publisher’s note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/