# scientific reports

# OPEN



# Identifying metabolic shifts in Crohn's disease using 'omics-driven contextualized computational metabolic network models

Philip Fernandes<sup>1,5</sup>, Yash Sharma<sup>1,5</sup>, Fatima Zulqarnain<sup>1</sup>, Brooklyn McGrew<sup>1</sup>, Aman Shrivastava<sup>1</sup>, Lubaina Ehsan<sup>1</sup>, Dawson Payne<sup>1</sup>, Lillian Dillard<sup>1</sup>, Deborah Powers<sup>1</sup>, Isabelle Aldridge<sup>1</sup>, Jason Matthews<sup>2</sup>, Subra Kugathasan<sup>2</sup>, Facundo M. Fernández<sup>3,4</sup>, David Gaul<sup>3,4</sup>, Jason A. Papin<sup>1</sup> & Sana Syed<sup>1</sup>

Crohn's disease (CD) is a chronic inflammatory disease of the gastrointestinal tract. A clear gap in our existing CD diagnostics and current disease management approaches is the lack of highly specific biomarkers that can be used to streamline or personalize disease management. Comprehensive profiling of metabolites holds promise; however, these high-dimensional profiles need to be reduced to have relevance in the context of CD. Machine learning approaches are optimally suited to bridge this gap in knowledge by contextualizing the metabolic alterations in CD using genome-scale metabolic network reconstructions. Our work presents a framework for studying altered metabolic reactions between patients with CD and controls using publicly available transcriptomic data and existing gene-driven metabolic network reconstructions. Additionally, we apply the same methods to patient-derived ileal enteroids to explore the utility of using this experimental in vitro platform for studying CD. Furthermore, we have piloted an untargeted metabolomics approach as a proof-of-concept validation strategy in human ileal mucosal tissue. These findings suggest that in silico metabolic modeling can potentially identify pathways of clinical relevance in CD, paving the way for the future discovery of novel diagnostic biomarkers and therapeutic targets.

Crohn's disease (CD) is a chronic, relapsing inflammatory disease of the gastrointestinal tract. CD prevalence ranges between 96 and 318 per 100,000 person-years in North America alone, with pediatric patients being the fastest-growing CD incidence group<sup>1-3</sup>. Pediatric-onset CD affects the upper gastrointestinal tract more frequently than adult-onset disease and is more likely to follow a more severe disease course<sup>4,5</sup>. The progressive inflammation resulting from CD disrupts critical developmental periods in children, impeding physical and psychosocial growth<sup>6</sup>. Our current understanding of CD pathogenesis has expanded in recent years to include dietary and lifestyle factors, environmental influences, genetic predilection, and gut microbiome composition<sup>7,8</sup>. The initial clinical diagnostic evaluation of CD includes, but is not limited to, panels of blood and stool biomarkers, such as hemoglobin, C-reactive protein (CRP), erythrocyte sedimentation rate (ESR), albumin, and fecal calprotectin. These biomarkers have sensitivities ranging from 80 to 90% but relatively low pooled specificity to CD in pediatric patients<sup>9,10</sup>. For a definitive diagnosis, clinicians use invasive and often costly investigations such as endoscopy, biopsy, and small-bowel cross-sectional imaging<sup>2</sup>. Thus, there is a clear gap in highly specific biomarkers for CD diagnosis and subsequent prognosis.

<sup>1</sup>Department of Pediatrics, Division of Gastroenterology, Hepatology & Nutrition, University of Virginia School of Medicine & UVA Child Health Research Center, University of Virginia, MR-4 Bldg, 409 Lane Rd., Charlottesville, VA 22908, USA. <sup>2</sup>Division of Pediatric Gastroenterology, Emory Children's Center, Atlanta, GA, USA. <sup>3</sup>Petit Institute for Bioengineering and Bioscience, Georgia Institute of Technology, Atlanta, GA, USA. <sup>4</sup>School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, GA, USA. <sup>5</sup>These authors contributed equally: Philip Fernandes and Yash Sharma. <sup>\Box</sup>email: papin@virginia.edu; sana.syed@virginia.edu

The management protocols for CD hinge on achieving suppression of intestinal inflammation during an induction phase with subsequent maintenance of remission. Strategies for induction include a bottom-up or top-down approach using a combination of steroids, immunomodulators (azathioprine or 6-mercaptopurine), antibiotics, or biologics that inhibit tumor necrosis factor-alpha, such as infliximab and adalimumab<sup>11–13</sup>. Unfortunately, these medications have variable efficacy between patients and may cause severe side effects<sup>14,15</sup>. Therefore, clarifying pathways contributing to CD may reveal novel targets for directed therapy, reducing immuno-suppression-related morbidity.

Constraint-based metabolic modeling has emerged as a useful in silico computational approach to investigate variations in the metabolism under certain biological conditions by analyzing large-scale relationships between genotypes and phenotypes<sup>16</sup>. Through the interrogation of transcriptomic variations in metabolism, we may be able to illuminate not only pathophysiological pathways for targeted therapeutics but also specific metabolic biomarkers of CD. Genome-scale metabolic Network Reconstructions (GENREs) provide a platform to study transcriptomic data in the context of metabolic shifts between various disease states<sup>17</sup>. Recon3D is the most comprehensive human metabolic network reconstruction to date; it includes gene, protein structure, and metabolite data from over 3,500 genes<sup>18</sup>. Genes related to metabolism are annotated and matched to the enzymes they code for and the reactions they catalyze to generate a catalog of gene-protein-reaction (GPR) associations. Flux Balance Analysis (FBA) is one method that can be employed to create a mathematical function of the flux of substrates along a given pathway in a metabolic network reconstruction with the assumption of a steady state<sup>19</sup>. FBA data obtained from constrained-based modeling can subsequently be used as inputs for machine learning models to perform analysis in a sample-specific manner to identify relevant targets for biotechnology and biomedical interventions. FBA-based methodologies effectively offer an approach to pruning unused pathways to identify relevant utilized pathways from the dense reaction-centric metabolic network. Using one pruning approach, parsimonious flux balance analysis (pFBA), metabolic reactions relevant to specific physiological or pathological conditions can be determined by identifying the most efficient reactions in generating biomass<sup>20-23</sup>. Reaction Inclusion by Parsimony and Transcript Distribution (RIPTiDe) provides an additional method for pruning reactions by factoring in transcriptomic abundances to identify the most important energy-efficient metabolic pathways in a given biological state. In CD, this multi-tiered in silico approach potentially allows us to discern the metabolic underpinnings of CD and clarify possible biomarkers and druggable targets<sup>24-26</sup>.

To this end, we have used an FBA-based machine learning (ML) framework for studying altered metabolic reactions between patients with CD and controls, leveraging 'omics data from a public repository from the Risk Stratification and Identification of Immunogenetic and Microbial Markers of Rapid Disease Progression in Children with Crohn's Disease (RISK) study, one of the largest pediatric inception cohorts of children with a new diagnosis of CD, to identify biomarkers found in the blood or stool to help predict which children are at risk of developing complications<sup>1,27</sup>. We aimed to (1) identify a list of metabolic pathways that have altered flux or flow in CD versus controls using in silico metabolic modeling utilizing ileal transcriptomics (2) explore the overlap of metabolic pathways in CD ileal enteroids, and (3) pilot an approach to validate our findings of altered metabolism using untargeted mass-spectrometry-based metabolomics in ileal mucosal biopsies. We hypothesized that the predicted metabolic shifts in the CD ileum will mirror those in the enteroid model, as well as those metabolites identified using metabolomics, and thus this metabolic modeling approach potentially holds power to elucidate novel biomarkers and therapeutic targets, allowing for improved methods for the diagnosis and treatment of Crohn's disease.

#### Results

**Demographics of clinical cohorts RISK and enteroids.** The RISK prospective pediatric inception study included n = 243 patients with a new diagnosis of Crohn's disease and n = 43 age-matched control patients. Our analysis included n = 163 patients (59% male) diagnosed with CD localized to the ileum with a mean (±SD) age of 11.9 (±3.1) years and n = 42 controls (62% male) with a mean (±SD) age of 11.1 (±3.1) years. In addition, publicly available RNA sequencing data from enteroids that were derived from a previously established cohort of n = 16 patients with CD with a mean age (±SD) of 15.8 (±3.2) years, 88% of whom were male, and n = 12 non-IBD controls with a mean (±SD) age of 11.4 (±3.5), 42% of whom were male was also used for this analysis<sup>28</sup>. A descriptive summary of both cohorts is outlined in Table 1.

	A. RISK Cohort		B. Enteroid Cohort	
	Crohn's disease n = 163	Controls n = 42	Crohn's disease n = 16	Controls n = 12
Age* (yr), mean (SD)	11.9 (3.1)	11.1 (3.1)	15.8 (3.2)	11.4 (3.5)
Male (%)	59% (n=96)	62% (n=26)	88% (n=14)	42% (n=5)

**Table 1.** Demographic data of cohorts used for metabolic modeling: A. RISK cohort. Publicly available Risk Stratification and Identification of Immunogenetic and Microbial Markers of Rapid Disease Progression in Children with Crohn's Disease (RISK) cohort data was accessed using GEO accession series GSE57945. RISK is a prospective inception cohort study that enrolled 1,276 pediatric patients with IBD diagnosed at 28 sites in North America between 2008 and 2012. It includes n = 163 patients with ileal Crohn's disease and n = 42 healthy controls. B. Enteroid cohort. Gene expression data for all mucosal terminal ileal biopsies used to generate enteroids included in this study have been accessed from the Sequence Read Archive (SRA) using the SRA series accession SUB7687325. \*Age at diagnosis; SD: standard deviation.

To identify peripheral biomarkers and metabolic pathways indicative of CD, we used a computational metabolic modeling approach to generate metabolic profiles of the disease (Fig. 1). We independently constructed the metabolic profiles and lists of relevant reactions for CD and control groups from (1) the open access RISK transcriptomic dataset with Recon3D and (2) the enteroid transcriptomic dataset with Recon3D, as shown in Fig. 1.

Differentially utilized reactions in the RISK metabolic model. Next, we sought to develop a metabolic model of ileal tissue from patients with CD. To this end, we applied RIPTiDe to publicly available transcriptomic data from the previously completed RISK study<sup>1</sup>. We identified context-specific metabolic pathway activity in both CD and control groups (n = 166, n = 42, respectively). This metabolic profiling revealed that the activity of approximately 200 reactions varied between control and CD populations (Supplemental Table 1), demonstrating that CD is associated with significant changes in ileal metabolism. Subsequently, we sought to determine whether this metabolic modeling approach could potentially identify CD using the metabolic signature generated from a transcriptomic validation dataset. Using random forest classification to identify discriminating reactions, our models determined whether a sample was from diseased versus control tissue, achieving 80% accuracy after 100 train-test splits in classifying CD versus control patients. Given these high accuracy measures, we can confidently use these methodologies to identify the most relevant metabolic alterations specific to CD. The average number of reactions pruned from the RISK dataset was 314 reactions (Supplemental Table 3). We next identified the top metabolic reactions that differentiated disease from controls. The reaction IDs, as documented in the Recon3D human network model were: AMP/ATP Transporter, endoplasmic reticulum (ATP2ter), hydroxymethylglutaryl coenzyme A reversible mitochondrial transport (HMGCOAtm), transport of (R)-mevalonate (MEV\_Rt), (R)-mevalonate: NADP + oxidoreductase (CoA Acylating) (r0488), exchange of (R)-mevalonate (EX\_mev\_R[e]), alpha-linolenoyl-CoA metabolism (sink\_lnlncacoa[c]), exchange of uridine (EX\_uri[e]), long-chain-acyl Coenzyme A dehydrogenase (r1466), linoleic acid metabolism (sink\_lnlc[c]), and uridine facilitated transport in the cytosol (URIt) (p value < 0.001; Supplemental Table 1). These top reactions were further grouped based on the metabolites involved in the reactions and the overall biological process. Upon literature review, the reactions were grouped under the gross biochemical pathways to which they belonged. The reactions found to be significantly altered were the mevalonate pathway, fatty acid oxidation, and uridine transport (p value < 0.001; Fig. 2A–D). Our model identified several alterations in the mevalonate pathway, including the conversion of mevalonate to HMG-CoA and HMG-CoA transport and exchange (Fig. 2B), and in the fatty acid oxidation pathway, such as linoleic acid transport, alpha-linolenoyl CoA exchange, and reactions involving long-chain-acyl coenzyme A dehydrogenase (Fig. 2C). Finally, alterations noted in uridine transport included



Figure 1. Schematic Overview of the performed metabolic modeling experiments: (A) On publicly-available RISK transcriptomic dataset and (B) Enteroid cultures that were generated for CD pediatric patients and controls.

![](_page_3_Figure_1.jpeg)

Figure 2. Metabolic pathways with an altered flow in patients with Crohn's disease (CD) versus control patients. The Flux Balance Analysis/Random Forest framework outlined in Fig. 1 were applied to the RISK dataset. After obtaining the list of top metabolic reactions that were altered between CD and control groups, these reactions were grouped into "families" (A-D) based on the biological processes they were involved in: (A) ATP transport processes were found to have significant rates of flux. (B) Mevalonate pathway. Several enzymes within the mevalonate pathway were altered in patients with CD, including the exchange of R- mevalonate, hydroxymethylglutaryl coenzyme A reversible mitochondrial transport, transport of R- mevalonate, and Rmevalonate NADP + oxidoreductase (CoA acylating), (C) Fatty acid oxidation. Long-chain-acyl Coenzyme A dehydrogenase, linoleic acid transport, and alpha-linolenoyl-CoA exchange (D) Uridine transport and exchange. For all graphs, the x-axis describes the reaction that is altered between controls (orange) and diseased states (blue). The y-axis shows the flux values generated by RIPTiDe by analyzing the flow of metabolites through an ileal-specific metabolic network reconstruction. The scale of flux values (y-axis) varies with the reactions as the efficiency of different metabolic pathways in generating biomass varies in a given biological system. A Mann-Whitney U test was done to compare reactions that varied between patients with Crohn's disease and controls. The stars (\*) are used to flag levels of significance. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001. AMP: adenosine monophosphate; ATP: adenosine triphosphate; ER: endoplasmic reticulum; NADP+: nicotinamide adenine dinucleotide phosphate CoA: Coenzyme A.

- - .

the exchange of uridine and uridine-facilitated transport in the cytosol (Fig. 2D). Together, these findings highlight the significant metabolic differences between controls and patients with CD.

**Differentially utilized reactions in the enteroid metabolic model.** Enteroids are a valuable tool in modern research, facilitating the ex vivo manipulation and investigation of human tissues. RNA sequencing (RNA-seq) from recent studies has demonstrated that enteroids derived from patients with CD have a 90% similarity in protein-coding genes compared to intestinal crypts isolated from these patients<sup>28</sup>. Furthermore, isolated enteroids were found to have similar cytokine and growth factor profiles as inflamed epithelium in patients, suggesting that enteroids exhibit similar characteristics to the diseased epithelium<sup>28</sup>. Therefore, we hypothesized that enteroid transcriptomes could be used to create a generalized in vitro metabolic model for gastrointestinal disease to distinguish quantifiable and targetable metabolic pathways. To test this hypothesis, we used our analytical pipeline designed for the RISK dataset outlined above on enteroid RNA-seq data from a CD cohort<sup>28</sup>. Enteroids were generated as previously reported, and RNA extracted for downstream analysis was accessed from the National Library of Medicine<sup>28</sup>. RNA sequencing data was RPKM normalized and used for downstream metabolic modeling to validate our in silico approach in a tissue CD model. Subsequently, we applied RIPTiDe, followed by random forest, to this enteroid transcriptomic dataset and extracted metabolic pathways most energetically productive to the CD model, as detailed in the methods section. Of note, our model had an average accuracy of 56% in classifying whether transcriptomic data originated from either CD or control patients within

this validation cohort of enteroids after 100 train-test validation splits. Even though our accuracies were on average low when classifying patients with CD versus controls, we hypothesized that these data could potentially reveal significant alterations in the metabolic profiles of CD patients compared to control patients. Thus, as detailed in our methods section, we studied the top relevant reactions from only those train-validation splits where our detection accuracy was high enough (>70%) to ensure the discriminability between CD and control patients to demonstrate the promise of enteroid models to serve as in vitro models of disease. The number of reactions in pruned in the enteroid dataset was 651 (Supplemental Table 3).Together, these data reveal significant alterations in the metabolic profiles of CD patients compared to control patients.

Following metabolic modeling of these enteroids derived from CD patients, we used the Recon3D database to categorize the reaction identification (IDs) of the most energy-efficient reactions in the small bowel. These reaction IDs were superoxide dismutase (SPODMm), catalase (CATm), RE3121R, a reaction involved in fatty acid oxidation, choline phosphatase (PCHOLP\_hs), transport of phosphatidylserine (PSt3), passive diffusion of sphinganine into extracellular space (SPHGNte), palmitoyl coenzyme A hydrolase (RE0577C), exchange of coproporphyrinogen (EX\_HC01610[e]), and HMR\_0692, and a biomass reaction (biomass\_reaction), two reactions involved in energy exchange (Fig. 3, Supplemental Table 1).

The metabolic reactions in the enteroids were classified using the subsystems given on the Virtual Metabolic Network and literature review as belonging to six metabolic subsystems: reactive oxygen species (ROS)

![](_page_4_Figure_4.jpeg)

**Figure 3**. Metabolic pathways with an altered flow in enteroids generated from patients with Crohn's disease (CD) versus control patients: Our Flux Balance Analysis/Random Forest framework was applied to RNA sequencing data from enteroid models generated from n = 16 patients with Crohn's disease and n = 12 healthy controls. After obtaining the list of top metabolic reactions that were altered between the CD and control groups, these reactions were grouped into "families" (A–D) based on the biological processes they were involved in. (A) ROS detoxification: superoxide dismutase and catalase, (B) Glycerophospholipid metabolism: choline phosphatase and the transport of phosphatidylserine. (C) Fatty acid oxidation: RE3121R is involved in pentaenoyl coenzyme A metabolism and palmitoyl coenzyme A hydrolase (D) Sphinganine transport. For all graphs, the x-axis describes the reaction that is altered between controls (orange) and diseased states (blue). The y-axis shows the flux values generated by RIPTiDe by analyzing the flow of metabolites through an ileal-specific metabolic network reconstruction. The scale of flux values (y-axis) varies with the reactions as the efficiency of different metabolic pathways in generating biomass varies in a given biological system. A Mann–Whitney U test was done to compare reactions that were varied between patients with Crohn's disease and controls. The stars (\*) are used to flag levels of significance. \*p < 0.05, \*\*p < 0.01, \*\*\*p < 0.001, ns: not significant. ROS: reactive oxygen species.

detoxification, fatty acid oxidation, glycerophospholipid metabolism, fatty acid synthesis, transport, and exchange/demand<sup>18,29,30</sup>. These were further grouped based on the metabolites involved in the reactions and overall biological process, revealing four subgroups: ROS detoxification, fatty acid metabolism, glycerophospholipid metabolism, and sphinganine transport (Fig. 3A–D). Alterations to the ROS detoxification reactions included superoxide dismutase and catalase. These enzymes occur in series to detoxify superoxide and then catalyze the decomposition of hydrogen peroxide into oxygen and water. Alterations in glycerophospholipid metabolism included phosphatidylserine transport and phospholipase D metabolism. Phosphatidylserine is a lipid component of cellular membranes and has non-structural roles in cell signaling and apoptosis. Phospholipase D serves as a key component of multiple signaling and metabolic pathways. These findings suggest that CD is associated with alterations in several diverse metabolic pathways.

**Comparative analyses of RISK and enteroid metabolic models.** Using the computational approaches detailed above, we generated metabolic models of CD in both RISK-derived tissue and enteroids. When these models were compared, we noted that fatty acid metabolism (Fig. 4A) was a common biological process highlighted between RISK and enteroid models, with additional reactions identified in each model. Furthermore, we noted a significant overlap in gene expression between ileal organoids and tissue biopsies (Fig. 4B).

These data indicated that the fatty acid metabolic profile in the tissue sufficiently maps to epithelial cell function. In addition, the gene count data shown in Fig. 4B suggests transcriptomic changes associated with CD may differ based on different specific cell types being analyzed. The difference in gene count between RISK and enteroid data prior to overlaying on Recon3D suggests that the total gene count data in the publicly available transcriptomic datasets included both protein-coding and noncoding mRNA sequencing. By subtracting the enteroid epithelial gene count data (n = 2445 genes), after overlaying it on Recon3D from the gene count data from RISK (n = 3193 genes), we inferred that 748 genes are likely to be related to functions outside of the epithelium that are not represented in enteroids but would be present in an ileal biopsy.

Contextualization of in silico metabolic modeling approach with metabolomics. As noted above, inflamed pediatric CD patient ileal tissue displays a distinct metabolic signature compared with noninflamed tissue. Subsequently, we sought to experimentally confirm whether the CD ileum contained metabolic changes versus controls using a small external cohort. To this end, we used mass spectrometry-based lipidomics in an experiment complementary to our in silico metabolic modeling to characterize whether the changes in metabolic pathways correlate to specific changes in metabolite composition. A heat map depicting relative metabolite quantification of the most variable compounds between inflamed (red) and non-inflamed tissue (green) shows clusters of metabolites that demonstrate a distinct metabolic signature of CD compared to control samples (Fig. 5A). Detailed LC-MS compound information can be found in Supplemental Table 2. This table includes the feature numbers for the compounds that make up the y-axis labels of Fig. 6A and the elemental formulas, molecular weights, and chromatographic retention time of the identified lipid species. Significant variations in metabolic alterations within individuals were observed (Fig. 6A), suggesting significant genetic heterogeneity within CD pathogenesis. In addition, we performed pathway analysis using the mummichog analytic approach for untargeted metabolomics, which leverages the metabolic networks KEGG, UCSD Recon1, and the Edinburgh human metabolic network, to predict biological activity without the need for the identification of specific metabolites beforehand. Using this approach, we identified vitamin B3 metabolism, pyrimidine metabolism, nitrogen metabolism, and other lipid and amino acid pathways to be enriched in CD compared to control tissue, in alignment with our in silico predictions and tissue LC-MS measurements (Fig. 5B)<sup>31</sup>.

![](_page_5_Figure_5.jpeg)

**Figure 4.** Overlap between gene count data and pathways with an altered flow in the RISK dataset and enteroids: (**A**) Overlap between the reaction "families" that were found to be altered in the RISK dataset and enteroids. After conducting in silico RIPTiDe studies and performing random forest analysis, reactions involved in fatty acid oxidation were found to be altered in both the RISK dataset and the enteroids as detailed in Table 1. (**B**). Gene overlap between the RISK cohort and the ileal organoids- "enteroids." Gene count analysis showed that 99% of the genes present in enteroids were also present in the RISK dataset. Only 96 genes were found to be unique to the enteroids.

![](_page_6_Figure_1.jpeg)

![](_page_6_Figure_2.jpeg)

Figure 5. Ileal tissue from inflamed pediatric CD patients vs. non-inflamed have distinct metabolic signatures. (A) A dendrogram depicting relative abundances of the 50 most significant compounds (top to bottom) compared between inflamed ileal tissue (red, left) and non-inflamed ileal tissue (green, right) patients (left to right). Data was obtained with reverse phase liquid chromatography mass spectrometry (LC-MS) in positive ion mode. Detailed LC-MS compound information is detailed in Supplemental Table 2. Supplemental Table 2 clarifies the y-axis labels, which are the feature numbers, the compounds, their elemental formulas, their molecular weight, as their chromatographic retention time. (B) Pathway analysis of pilot data using the mummichog analytic approach showed the enrichment of multiple metabolites including vitamin B3 metabolism, pyrimidine metabolism, nitrogen metabolism and other lipid and amino acid pathways in CD compared to control tissue groups, in alignment with our in silico findings.

![](_page_7_Figure_1.jpeg)

**Figure 6.** Overview of Methods Framework: (**A**) Risk Stratification and Identification of Immunogenetic and Microbial Markers of Rapid Disease Progression in Children with Crohn's Disease (RISK) is a prospective inception cohort study, which enrolled 1,276 pediatric patients with IBD at diagnosis at 28 sites in North America between 2008 and 2012. It includes n = 163 patients with ileal Crohn's disease and n = 42 healthy controls. These transcriptomic abundances were overlaid onto Recon3D. Recon 3D is a comprehensive human metabolic network model consisting of three-dimensional (3D) metabolite and protein structure data. (**B**) Flux Balance Analysis workflow. Reaction Inclusion by Parsimony and Transcript Distribution (RIPTiDe) Flux Balance Analysis algorithm was used to generate flux values for the flow of metabolites through our ileal-specific metabolic network reconstruction. RIPTiDe generates stoichiometric equations for reactions and utilizes the transcriptomic expression of critical enzymes and metabolites to quantify flux within a specific disease state. A growth constraint is built into this algorithm to identify the reactions most efficient in generating biomass. (**C**) Random Forest Classifier Overview. Discriminating reactions between CD patient subtypes were determined using Random Forests (RF). The RF classifier used multiple classification trees to sample a given dataset and asked a series of T/F questions related to flux values to "learn" which metabolic reaction fluxes distinguished Crohn's disease versus normal tissues.

Together, these findings demonstrate the utility of computational metabolic modeling to specify the characteristic, potentially functional profiles of CD compared to control tissue. Furthermore, the individual metabolic pathways identified by our in silico model displayed significant overlap with the findings from our in vitro enteroid model and our method using untargeted metabolomics, demonstrating the potential value of this metabolic modeling approach. Our work lays the groundwork for future studies to continue probing alterations in the mucosal transcriptome and tissue metabolome for further insights into CD.

#### Discussion

This work presents a framework for the in silico metabolic modeling of Crohn's disease (CD) to identify metabolic pathway alterations using archival transcriptomic data and in vitro enteroid models. Cell-to-cell metabolic variability occurs within cell populations as these populations execute a diverse range of functions, which change depending upon environmental impacts on transcriptomic and epigenetic regulation. Metabolic modeling provides a functional snapshot of the ongoing biological processes in a given cell or organism based on complex genotype-phenotype-biochemical relationships. Identifying specific transcriptional signatures of disease through bulk RNA sequencing poses a significant challenge given the heterogeneity in gene expression between patients with the same disease. To distill large amounts of transcriptomic data into a contextualized disease model capable of elucidating alterations in metabolic pathways, which may serve as either disease biomarkers or targets for therapeutic intervention, we have employed metabolic modeling to distinguish alterations in metabolism in diseased versus control tissue across two different datasets. Alterations in metabolic flux can be validated and subsequently quantified externally through in vitro models such as enteroids, untargeted metabolomics, and lipidomics to clarify actionable biomarkers or therapeutic targets. To develop an in silico model, we applied our metabolic modeling approach to the archival RISK dataset, which showed that mevalonate metabolism, fatty acid oxidation, and uridine metabolism have variable flux when comparing patients with CD to controls (Fig. 2), suggesting the relevance of these three classes of biochemical pathways in CD pathogenesis. Our in vitro model comprised organoids generated from patients with CD and controls ileal tissue samples. These organoids were previously found to have a similar transcriptomic signature as in vivo epithelium and retain disease-specific gene expression patterns<sup>28</sup>. In enteroids, we found glycerophospholipid, linoleic acid, and sphingolipid metabolism to be altered (Fig. 3). Lastly, we piloted an additional validation step by using untargeted metabolomics to reveal individual metabolic pathways associated with CD, in which we found vitamin B3, pyrimidine, and glutamate metabolism to be significantly enriched in CD.

The metabolic modeling pipeline detailed in this study showed that reactions involved in mevalonate metabolism were among the top reactions that differentiated CD from controls in the RISK dataset (Fig. 2B). Alterations in mevalonate metabolism in T-lymphocytes have been implicated in decreased inflammatory suppressive activity<sup>31</sup>. In addition, derivatives of mevalonate are known to be directly involved in limiting the cytotoxic effector response of T-lymphocytes in inflammation and the production of immunomodulatory precursors<sup>32</sup>. Alterations of this pathway, such as mevalonate kinase deficiency (MKD), lead to inflammatory bowel disease, or IBD-like intestinal inflammation, possibly due to decreased immunosuppressive isoprenoid intermediates formed through the mevalonate kinase pathway<sup>33</sup>. Therefore, the changes in mevalonate metabolism between CD and controls tissue identified here may contribute to establishing the inflammatory environment associated with CD.

Fatty acid metabolism is a highly regulated process in which dysregulation can lead to an imbalance of proand anti-inflammatory mediators. Of note, the dysregulation of fatty acid metabolism contributes to both the type and degree of inflammatory responses in the intestine during IBD<sup>34–36</sup>. Recently, in-depth gene array analysis has also revealed that various genes involved in fatty acid uptake and synthesis are differentially expressed in the ileum of IBD patients<sup>37</sup>. One such pathway differently expressed in IBD patients involves uridine transport and exchange. Uridine is a critical regulator of lipid metabolism and a pathway identified in our current study (Fig. 2D)<sup>38</sup>. As such, pharmacologic agents targeting uridine metabolism are promising candidates for future treatment of inflammatory bowel disease (IBD). Remarkably, in studies utilizing uridine to treat mice with Dextran Sulfate Sodium (DSS)-Induced Colitis, levels of pro-inflammatory cytokines IL-6, IL-1β, and TNF in the serum, and mRNA expression in the colon were significantly decreased in the uridine-treated groups<sup>39</sup>, further demonstrating the critical role for fatty acid metabolism in regulating gastrointestinal inflammation.

In previous studies using cultured ileal organoids (enteroids), no morphological characteristics were observed that could be used to differentiate between CD and controls. Further, when gene expression profiles from enteroids were compared to profiles from freshly isolated intestinal crypts, a strong positive correlation was observed in the mean expression levels of many genes<sup>28,40</sup>. This result suggests that most genes expressed in vivo in the epithelium are also expressed in organoid cultures, demonstrating the immense value of enteroids as an important tool for IBD research. As with the RISK dataset, we grouped altered metabolic reactions identified in organoids based on broad biochemical pathways. We noted that four of the top ten reactions were involved in the dysregulation of fatty acid and phospholipid metabolism (Fig. 3B, C). These findings agree with previous lipid profiling studies of CD patients compared to controls. In these studies, fatty acid and phospholipid metabolism was significantly altered, with most alterations found in glycerophospholipid, linoleic acid, and sphingolipid metabolism pathways<sup>41</sup>. Various metabolites involved in these pathways play a role in inflammation, intracellular signaling, pain, immune function, reproduction, and appetite, perpetuating colitis when dysregulated<sup>42</sup>. Future work will test whether drugs targeting these pathways can alleviate the characteristic inflammation of IBD in patient-derived enteroids, laying the groundwork for further studies into potential therapeutic use<sup>40</sup>. Additionally, the observed alterations of different metabolites of fatty acid oxidation between tissue and organoids suggest mitochondrial dysfunction<sup>43</sup>. Fatty acid biosynthesis has previously been shown to be part of a mitochondrial-tocytosolic stress response that results when mitochondrial protein synthesis is disrupted<sup>44</sup>. Staining our patientderived organoids with a mitochondrial membrane potential dye may further help clarify this relationship.

Another biological process that was found to be significantly altered in enteroids derived from CD patients was sphinganine transport (Fig. 3D). Sphinganine is biosynthesized from serine and palmitoyl-CoA through decarboxylating condensation, resulting in a keto intermediate, which is reduced by NADPH. It is then further acylated, then dehydrogenated to form ceramide<sup>45</sup>. Ceramide decreases the release of tumor necrosis factor (TNF), most likely via post-translational regulatory mechanisms and the modulation of TNF-converting enzyme activity<sup>46</sup>. Furthermore, ceramides play an essential role in regulating autophagy, a process strongly linked to the pathogenesis of CD<sup>47</sup>. Sphinganine also blocks lysosomal cholesterol transport and has been linked to Niemann-Pick Type C disease, which predisposes to early-onset IBD with CD phenotype and granuloma formation<sup>48</sup>. However, the exact mechanism by which sphinganine contributes to CD remains unclear and will require further study.

In addition, oxidative stress, caused by increased ROS production, is present locally and systemically in patients with CD<sup>49</sup>. Here, we identified two reactions in this pathway: superoxide dismutase and catalase (Fig. 3A). These enzymes play a role in endogenous antioxidant mechanisms and counteract the effects of excess ROS<sup>50,51</sup>. Notably, multiple enzymes catalyze the production of ROS. Among these enzymes, mucosal NOXs and dual oxidase 2 (DUOX2) have been reported as novel risk factors for IBD<sup>52,53</sup>, further demonstrating that an imbalance in redox homeostasis contributes to the pathogenesis of IBD. Interestingly, In the RISK study,

the authors highlight the role of oxidative stress in gut inflammation. Their works detail the enrichment of pro-inflammatory genes, including antimicrobial dual oxidase (DUOX2), and the decreased expression of antiinflammatory and antioxidant genes, such as those resulting in the production of apoprotein A1 (APOA1) in the ileal tissue of patients with CD<sup>27</sup>.

Using mass spectrometry-based lipidomics, we aimed to pilot an additional level of validation complementary to our in silico metabolic modeling approach. The significant variations in metabolic alterations between diseased and not diseased individuals and within different diseased individuals provide a rationale for future metabolomic analyses of large cohorts examining phenotypic variations within patients with CD. Our metabolomics analysis showed overlaps with our in silico and in vitro modeling. For example, vitamin B3 was found to be significantly enriched, iterating the theme of the dysregulation of lipid metabolism present in CD. Interestingly, high-dose vitamin B3 treatment has been shown to ameliorate ulcerative colitis through increased prostaglandin D2 synthesis in mice<sup>54</sup>. Thus, niacin supplementation can be a potential therapeutic target to be investigated in CD as well. In addition, untargeted metabolomics showed the dysregulation of pathways related to pyrimidine, glutamate, and nitrogen metabolism. Similarly, in the RISK dataset, we found uridine transport and metabolism to be altered between patients with CD and controls, further warranting assessment of the potential efficacy of uridine in CD management.

Limitations in this study are shown in the discrepancy in the specific metabolic reactions that have altered flux when comparing the RISK dataset to the enteroid dataset presents a visible limitation to our study. The use of publicly available, pre-normalized data that does not include row counts presents a limitation to using other normalization techniques for inter-sample comparison. We used Recon3D, which is a generic metabolic network reconstruction and not tissue-specific. In future studies, we hope to generate a tissue-specific reconstruction analogous to the heart-specific iCardio<sup>17</sup>. Further, the inclusion of medium information specific to Crohn's disease may improve FBA performance, as well as inclusion of detailed information on macromolecular synthesis rates in Crohn's disease to support or negate assumptions of hypermetabolism. From a biological standpoint, variations in the mean age and ethnicity of patients also have varying effects on the metabolome and lipidome of the enteroids and, thus are confounders in our analysis. In addition, there is an incongruity between the sizes of the two cohorts, as enteroids were generated from a smaller population (Table 1). Furthermore, it is crucial to recognize that enteroids only consist of epithelial tissue, while ileal tissue contains other mucosal elements, including the submucosa, muscularis mucosa, and the immune cells that are present in these layers.CD affects the entire thickness of the epithelium, involving a complicated interaction between genetic factors, growth factors, cytokines, environmental factors such as diet and smoking, and the constitution of the gut microbiome<sup>27</sup>. Thus, not only does the absence of other cell types present a limitation to our analysis, but the absence of an accounting of the gut microbiome signaling in the enteroids is also a shortcoming. In particular, the absent gut microbiome may serve as an explanation for the dissimilarities between altered fatty acid pathways in the RISK dataset and enteroid models<sup>55,56</sup>. In emerging work studying organoids' interaction with other elements of the gastrointestinal system, researchers added microbiome elements to characterize the carcinogenic effects of E. coli on intestinal organoids<sup>57</sup>. This work clearly demonstrated that microbial elements may significantly impact the genetic signature of organoids, revealing a technical limitation in using organoids. To address this limitation, other groups have started developing novel organoid culture systems having added resident innate immune cells and fibroblasts and observed their interactions with tissue to enable further use of enteroid as an in vitro model for changes to the gut epithelium<sup>58</sup>. Future studies with these complex organoid systems will allow us to better probe the relationship of the microbiome to the metabolic pathways underlying CD pathogenesis.

The notable strength of this study is the detailed framework for in silico prediction with in vitro validation, pruning genes and metabolic reactions from transcriptomic data, and lipidomics analysis to provide a high-order understanding of alterations in metabolism for specific disease states. These altered pathways may define measurable biomarkers specific to certain disease states or even represent targetable therapeutic options. In this study, we identified metabolic pathways with altered flow in both archival data and enteroid models of Crohn's disease and revealed overlap in the overall biochemical processes in which they occur. In the archival RISK dataset, we found mevalonate, fatty acid, and uridine metabolism to be altered, while in enteroids, we found glycerophospholipid, linoleic acid, and sphingolipid metabolism to be altered (Figs. 2, 3). Thus, fatty acid metabolic pathways offer promising therapeutic targets in managing CD. Due to the overlap between metabolic alterations observed in both enteroids and the RISK datasets, ileal enteroids can offer us insight into epithelial response to various pharmaceutical interventions. In addition, the other metabolic pathways identified by metabolic modeling offer a wide range of targetable metabolites and reactions that can be investigated further. Metabolomics analyses can also be used to further validate these findings by confirming whether the metabolites of the pathways with altered flow are present and measurable in ileal enteroids and biopsies. Future studies will determine whether the metabolites identified in this study can be used to monitor disease progression and treatment response and whether therapeutic targeting of these pathways may provide an effective approach to treating Crohn's disease.

# Methods

**Publicly available transcriptomic data from the RISK inception cohort.** RNA sequencing data was obtained from the RISK pediatric prospective inception cohort study<sup>1,27</sup>. All CD patients were required to undergo a baseline colonoscopy during which biopsies were taken to confirm chronic active colitis or ileitis on histology prior to intervention and were followed at a time point 22 months after diagnosis to monitor disease progression<sup>1,27</sup>. Transcriptomic data normalized to reads per kilobase per million (RPKM) transcriptomic data from the RISK study was accessed using the GEO accession series GSE57945. Our analysis included n = 163 patients with ileal CD and n = 42 controls who were not found to have CD on histology.

In silico metabolic modeling platform applied in vitro on enteroids. Enteroids were derived from a previously established CD cohort (n=16 CD, n=12 non-IBD controls)<sup>28</sup>. Publicly available enteroid RNA-seq data was accessed using the accession key PRJNA643576. Data was analyzed using the analytical pipeline designed for the RISK dataset outlined below. RNA sequences from the enteroid study were normalized to RPKM and used for downstream *in-silico* metabolic modeling to validate our *in-silico* approach using a tissue model of CD.

**In silico identification of relevant metabolic reactions.** The publicly available RISK transcriptomic dataset and enteroid transcriptomic data were overlaid onto Recon3D, a human metabolic reconstruction<sup>18</sup>. As RNA sequencing was done on ileal biopsies of patients with Crohn's disease and controls, by overlaying the gene count data onto Recon3D we obtain an ileum-specific model for this study. Genes present in both the Recon3D network and transcriptomic data were retained, while the rest were removed. Retained reactions were specifically linked with transcriptomically-abundant RISK and enteroid genes. Constraint-Based Reconstruction and Analysis (COBRA) Toolbox was used to remove reactions determined to be inactive and generate an ileal-specific metabolic network reconstruction to contextualize our downstream analyses<sup>59,60</sup>.

**Identifying reactions with differential utilization of relevant metabolic pathways.** RIPTiDe, a subset of parsimonious Flux Balance Analysis (pFBA), was used to prune the list of genes and reactions obtained from the diseased and control patient sets<sup>24</sup>. The RISK transcriptomic dataset or enteroid transcriptomic data with Recon3D was used to create a contextualized metabolic model for each patient based on the parsimonious usage of reactions defined by their associated transcriptomic analysis. Multiple possible flux values can be generated from the permissible range of flux bounds for all the active reactions in each patient. The RIPTiDe package randomly samples 50 to 500 flux values depending on the permissible flux ranges. The parameters for the implementation of the RIPTiDe package were kept at default, with the exception of the minimum percent of optimal objective value during the flux balance analysis which we increased from 0.8 to 1 to control the variability of flux ranges assuming a hypermetabolic state in  $CD^{61-63}$ . A random flux sampling method was used to generate multiple data points for training the machine-learning random forest model. The scikit learn implementation was used for random forest<sup>20-22,24,25,64,65</sup>. We kept the open\_exchange parameter in the COBRA optimizer as "false" to allow the FBA module to automatically determine the maximum and minimum bound of reactions based on transcriptome contextualization and FBA optimization.

**Extracting discriminative reactions using machine learning.** Many reactions identified by FBA are essential for normal tissue function and, thus, were found to have similar flow between patients with CD and controls. Therefore, we employed RIPTiDe to extract metabolic reactions that had altered flow between patients with CD and controls. A random forest classifier was used to classify CD versus control patients using reactions found to be altered in the diseased state. An 80–20% ratio was used for separating patient data points to create training data and test data for machine learning modeling. All flux points associated with patients in the training set were used to train the model, and the model was evaluated on flux data points associated with validation approach was used to identify the top reactions in an unbiased approach. Multiple splits of train-validation patients were created to train and evaluate the model. This extensive cross-validation allowed us to demonstrate performance gains and takeaways across the dataset. Top reactions were extracted and aggregated from splits where the model had a high detection accuracy. Due to an imbalance between the control and CD patients, a random under-sampling method was used in each split to balance the number of controls and CD patients. A schematic overview of the methods used can be viewed in Fig. 6.

**Grouping reactions to derive functional groups.** The top 20 reactions identified by our metabolic modeling pipeline (detailed above) were extracted from the splits with an F1 score (the harmonic mean of precision and recall) greater than 70% and used for further analysis. This step ensured only reactions which were substantially altered between CD and control patients in the training and validation patients were included in the functional group analysis. Differences between patients with CD and controls were calculated using the Mann–Whitney U test, and a *p* value <0.05 was considered significant. These top 20 reactions were further grouped based on their metabolic subsystems and description of their overall function based on the virtual metabolic human (VMH) reaction database of Recon3D<sup>18,30</sup>.

**Mass spectrometry-based lipidomics.** Ileal tissue samples from pediatric CD patients (n=11) and control non-inflamed ileal samples (n=14) were analyzed via non-targeted LC–MS lipidomics assays to semiquantitatively measure thousands of non-polar compounds, including lipids from the eight main lipid classes (fatty acyls, glycerolipids, glycerophospholipids, sphingolipids, sterol lipids, prenol lipids, saccharolipids, and polyketides) in these complex biosamples. Biopsies were cryopulverized prior to mixing with extraction solvent (IPA) to extract the non-polar lipids while minimizing highly polar compounds and proteins, which would suppress the signals of the desired analytes. The supernatant resulting from the extraction was stored at 4 °C until analysis. A pooled quality control sample was injected at regular intervals (every 5 samples) throughout the sample batch, bracketing the actual samples, which were randomized. Data collection was performed on a ThermoFisher Scientific Q-Exactive HF (QE HF) Hybrid Quadrupole-Orbitrap MS system coupled with a Vanquish Horizon LC system in positive and negative ion modes. Compounds were separated using a ThermoFisher Scientific Accurcore C30 ( $150 \times 2.1 \text{ mm}$ ,  $2.6 \mu m$  particle size). The chromatographic method for sample analysis involved elution with 80:20 water:MeCN with 10 mM ammonium formate and 0.1% formic acid (mobile phase A) and MeCN and 0.1% formic acid (mobile phase B) using the following gradient program: 0 min 5% A; 0.5 min 5% A; 8 min 60% A; 10.4 min 60% A; 10.5 min 5% A; 14 min 5% A. The flow rate was set at 0.4 mL/min. The column temperature was set to 40 °C, and the injection volume was 2  $\mu$ L. Full MS data was acquired with 240,000 resolutions over the 150–2000 m/z range. Data processing steps for LC–MS data were carried out using Compound Discoverer v3.0. Compound annotation was carried out based on accurate mass, retention time, isotopic pattern, and MS<sup>2</sup> fragmentation pattern matching local and public libraries. Annotations were consistent with the current naming convention in the Metabolomics Workbench and reflected the extent of structural information contained in the collected data<sup>66</sup>.

**Ethical approval.** All experiments and methods were performed in accordance with the relevant guidelines and regulations at Emory University, including approval by Emory University under IRB MOD004-IRB00085516. All patient samples were collected with informed consent.

#### Data availability

RNA sequencing data for the RISK dataset was accessed on the NCBI Gene Expression Omnibus using the accession code GSE57945 (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE57945). RNA sequencing data for the enteroid dataset was accessed at the National Library of Medicine using the accession code PRJNA643576 (https://www.ncbi.nlm.nih.gov/bioproject/?term=643576). The methods of this paper can be accessed at https://github.com/GutIntelligenceLab/ContextualizedMetabolicModel.

Received: 30 May 2022; Accepted: 20 December 2022 Published online: 05 January 2023

#### References

- Kugathasan, S. et al. Prediction of complicated disease course for children newly diagnosed with Crohn's disease: A multicentre inception cohort study. Lancet 389(10080), 1710–1718 (2017).
- Benchimol, E. I. et al. Incidence, outcomes, and health services burden of very early onset inflammatory bowel disease. Gastroenterology 147(4), 803-813 e7 (2014).
- Ng, S. C. et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. Lancet 390(10114), 2769–2778 (2017).
- 4. Freeman, H. J. Natural history and long-term clinical course of Crohn's disease. *World J. Gastroenterol.* **20**(1), 31–36 (2014).
- Rosen, M. J., Dhawan, A. & Saeed, S. A. Inflammatory Bowel disease in children and adolescents. JAMA Pediatr. 169(11), 1053–1060 (2015).
- Thayu, M. et al. Determinants of changes in linear growth and body composition in incident pediatric Crohn's disease. Gastroenterology 139(2), 430–438 (2010).
- 7. Linares, R. et al. Bacterial translocation as inflammatory driver in crohn's disease. Front. Cell Dev. Biol. 9, 703310 (2021).
- McGovern, D. P., Kugathasan, S. & Cho, J. H. Genetics of inflammatory bowel diseases. *Gastroenterology* 149(5), 1163-1176 e2 (2015).
- 9. Henderson, P., Anderson, N. H. & Wilson, D. C. The diagnostic accuracy of fecal calprotectin during the investigation of suspected pediatric inflammatory bowel disease: a systematic review and meta-analysis. *Am. J. Gastroenterol.* **109**(5), 637–645 (2014).
- Mosli, M. H. et al. C-Reactive protein, fecal calprotectin, and stool lactoferrin for detection of endoscopic activity in symptomatic inflammatory bowel disease patients: A systematic review and meta-analysis. Am. J. Gastroenterol. 110(6), 802–819 (2015).
- 11. Walters, T. D. *et al.* Increased effectiveness of early therapy with anti-tumor necrosis factor-alpha vs an immunomodulator in children with Crohn's disease. *Gastroenterology* **146**(2), 383–391 (2014).
- 12. Levine, A. *et al.* Crohn's disease exclusion diet plus partial enteral nutrition induces sustained remission in a randomized controlled trial. *Gastroenterology* **157**(2), 440-450 e8 (2019).
- Ruemmele, F. M. et al. Consensus guidelines of ECCO/ESPGHAN on the medical management of pediatric Crohn's disease. J. Crohns Colitis 8(10), 1179–1207 (2014).
- 14. Cushing, K. & Higgins, P. D. R. Management of Crohn disease: A review. JAMA 325(1), 69-80 (2021).
- 15. Lichtenstein, G. R. *et al.* Infliximab for Crohn's disease: More than 13 years of real-world experience. *Inflamm. Bowel Dis.* **24**(3), 490–501 (2018).
- Knecht, C. et al. Distinct metabolic network states manifest in the gene expression profiles of pediatric inflammatory bowel disease patients and controls. Sci. Rep. 6(1), 32584 (2016).
- Dougherty, B. V. et al. Identifying functional metabolic shifts in heart failure with the integration of omics data and a heart-specific, genome-scale model. Cell Rep. 34(10), 108836 (2021).
- Brunk, E. *et al.* Recon3D enables a three-dimensional view of gene variation in human metabolism. *Nat. Biotechnol.* 36(3), 272–281 (2018).
- 19. Reimers, A.-M. & Reimers, A. C. The steady-state assumption in oscillating and growing systems. J. Theor. Biol. 406, 176–186 (2016).
- 20. Sridhara, V. *et al.* Predicting growth conditions from internal metabolic fluxes in an in-silico model of *E. coli. PLoS ONE* **9**(12), e114608 (2014).
- 21. Orth, J. D., Thiele, I. & Palsson, B. Ø. What is flux balance analysis?. Nat. Biotechnol. 28(3), 245-248 (2010).
- 22. Anand, S., Mukherjee, K. & Padmanabhan, P. An insight to flux-balance analysis for biochemical networks. *Biotechnol. Genet. Eng. Rev.* **36**(1), 32–55 (2020).
- 23. Lee, J. M., Gianchandani, E. P. & Papin, J. A. Flux balance analysis in the era of metabolomics. *Brief. Bioinform.* 7(2), 140–150 (2006).
- 24. Jenior, M. L. *et al.* Transcriptome-guided parsimonious flux analysis improves predictions with metabolic networks in complex environments. *PLoS Comput. Biol.* **16**(4), e1007099 (2020).
- Machado, D. & Herrgård, M. J. Correction: Systematic evaluation of methods for integration of transcriptomic data into constraintbased models of metabolism. *PLoS Computat. Biol.* 10, e1003989 (2014).
- 26. Jerby, L. & Ruppin, E. Predicting drug targets and biomarkers of cancer via genome-scale metabolic modeling. *Clin. Cancer Res* 18(20), 5572–5584 (2012).
- Haberman, Y. et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. J. Clin. Investig. 124(8), 3617–3633 (2014).

- Niklinska-Schirtz, B. J. et al. Ileal derived organoids from Crohn's disease patients show unique transcriptomic and secretomic signatures. Cell. Mol. Gastroenterol. Hepatol. 12(4), 1267–1280 (2021).
- 29. Yang, Y. et al. Integration of metabolic networks and gene expression in virtual reality. Bioinformatics 21(18), 3645-3650 (2005).
- 30. Virtual metabolic human. Luxembourg Centre for Systems Biomedicine (LCSB) of the University of Luxembourg: Luxembourg.
- 31. Kempkes, R. W. M. et al. Metabolic pathways involved in regulatory T cell functionality. Front. Immunol. 10, 2839 (2019).
- 32. Bianco, A. M. et al. Mevalonate kinase deficiency and IBD: Shared genetic background. Gut 63(8), 1367 (2014).
- Dunn, K. et al. Mevalonate kinase deficiency presenting as recurrent rectal abscesses and perianal fistulae. Ann. Allergy Asthma Immunol. 120(2), 214–215 (2018).
- 34. Parada Venegas, D. *et al.* Corrigendum: short chain fatty acids (SCFAs)-mediated gut epithelial and immune regulation and its relevance for inflammatory bowel diseases. *Front. Immunol.* **10**, 1486 (2019).
- 35. Uchiyama, K. et al. Lipid and bile acid dysmetabolism in Crohn's disease. J. Immunol. Res. 2018, 7270486 (2018).
- 36. De Preter, V. *et al.* Faecal metabolite profiling identifies medium-chain fatty acids as discriminating compounds in IBD. *Gut* **64**(3), 447–458 (2015).
- Heimerl, S. et al. Alterations in intestinal fatty acid metabolism in inflammatory bowel disease. Biochim. Biophys. Acta BBA Mol. Basis Dis. 1762(3), 341–350 (2006).
- Zhang, Y. et al. Uridine metabolism and its role in glucose, lipid, and amino acid homeostasis. Biomed. Res. Int. 2020, 7091718– 7091718 (2020).
- 39. Jeengar, M. K. et al. Uridine ameliorates dextran sulfate sodium (DSS)-induced colitis in mice. Sci. Rep. 7(1), 3924 (2017).
- Yoo, J.-H. & Donowitz, M. Intestinal enteroids/organoids: A novel platform for drug discovery in inflammatory bowel diseases. World J. Gastroenterol. 25(30), 4125–4147 (2019).
- 41. Murgia, A. *et al.* Italian cohort of patients affected by inflammatory bowel disease is characterised by variation in glycerophospholipid, free fatty acids and amino acid levels. *Metabolomics* 14(10), 140 (2018).
- Linggi, B. et al. Meta-analysis of gene expression disease signatures in colonic biopsy tissue from patients with ulcerative colitis. Sci. Rep. 11(1), 18243 (2021).
- Rath, E., Moschetta, A. & Haller, D. Mitochondrial function—gatekeeper of intestinal epithelial cell homeostasis. Nat. Rev. Gastroenterol. Hepatol. 15(8), 497–516 (2018).
- Kim, H.-E. *et al.* Lipid biosynthesis coordinates a mitochondrial-to-cytosolic stress response. *Cell* **166**(6), 1539-1552.e16 (2016).
   Merrill, A. H. & Sweeley, C. C. Chapter 12—sphingolipids: metabolism and cell signaling. In *New comprehensive biochemistry* (eds)
- Vance, D. E. & Vance, J. E.) 309–339 (Elsevier, Amsterdam, 1996).
  46. Lamour, N. F. *et al.* Ceramide kinase regulates the production of tumor necrosis factor α (TNFα) via inhibition of TNFα-converting
- enzyme. J. Biol. Chem. 286(50), 42808–42817 (2011).
  47. Sewell, G. W. et al. Lipidomic profiling in Crohn's disease: abnormalities in phosphatidylinositols, with preservation of ceramide, phosphatidylcholine and phosphatidylserine composition. Int. J. Biochem. Cell Biol. 44(11), 1839–1846 (2012).
- Schwerd, T. *et al.* Impaired antibacterial autophagy links granulomatous intestinal inflammation in Niemann-Pick disease type Cl and XIAP deficiency with NOD2 variants in Crohn's disease. *Gut* 66(6), 1060–1073 (2017).
- Alzoghabia, M. A. Concepts of oxidative stress and antioxidant defense in Crohn's disease. World J. Gastroenterol. 19(39), 6540–6547 (2013).
- Wang, Y. et al. Superoxide dismutases: Dual roles in controlling ROS damage and regulating ROS signaling. J. Cell Biol. 217(6), 1915–1928 (2018).
- DeJong, R. J. et al. Reactive oxygen species detoxification by catalase is a major determinant of fecundity in the mosquito Anopheles gambiae. Proc. Natl. Acad. Sci. U.S.A. 104(7), 2121–2126 (2007).
- Lambeth, J. D. Nox enzymes, ROS, and chronic disease: An example of antagonistic pleiotropy. Free Radical Biol. Med. 43(3), 332-347 (2007).
- Denson, L. A. Epithelial reactive oxygen species and risk for very early onset inflammatory bowel disease. Cell. Mol. Gastroenterol. Hepatol. 1(5), 456–457 (2015).
- Li, J. et al. Niacin ameliorates ulcerative colitis via prostaglandin D2-mediated D prostanoid receptor 1 activation. EMBO Mol. Med. 9(5), 571–588 (2017).
- 55. Khan, I. *et al.* Alteration of gut microbiota in inflammatory bowel disease (IBD): cause or consequence? IBD treatment targeting the gut microbiome. *Pathogens (Basel, Switzerland)* **8**(3), 126 (2019).
- 56. Pascal, V. et al. A microbial signature for Crohn's disease. Gut 66(5), 813-822 (2017).
- 57. Pleguezuelos-Manzano, C. *et al.* Mutational signature in colorectal cancer caused by genotoxic pks+ *E. coli. Nature* **580**(7802), 269–273 (2020).
- Tominaga, K. *et al.* Generation of human intestinal organoids containing tissue-resident immune cells. *Gastroenterology* 162(3), S57 (2022).
- 59. Ebrahim, A. et al. COBRApy: COnstraints-based reconstruction and analysis for python. BMC Syst. Biol. 7, 74 (2013).
- 60. Heirendt, L. *et al.* Creation and analysis of biochemical constraint-based models using the COBRA Toolbox v30. *Nat. Protoc.* **14**(3), 639–702 (2019).
- 61. Al-Jaouni, R. et al. Energy metabolism and substrate oxidation in patients with Crohn's disease. Nutrition 16(3), 173–178 (2000).
- 62. Schneeweiss, B. et al. Energy and substrate metabolism in patients with active Crohn's disease. J Nutr 129(4), 844-848 (1999).
- 63. Zoli, G. et al. Increased energy expenditure in growing adolescents with Crohn's disease. Dig Dis Sci 41(9), 1754–1759 (1996).
- Pedregosa, F. et al. Scikit-learn: Machine learning in Python. J. Mach. Learn. Res. 12, 2825–2830 (2011).
- Schonlau, M. & Zou, R. Y. The random forest algorithm for statistical learning. *Stata J.* 20(1), 3–29 (2020).
- 56. Sumner, L. W. et al. Proposed minimum reporting standards for chemical analysis Chemical Analysis Working Group (CAWG) Metabolomics Standards Initiative (MSI). *Metabolomics* 3(3), 211–221 (2007).

# Acknowledgements

The authors would like to acknowledge Adam Greene, PhD, University of Virginia, for his help proofreading the manuscript.

# Author contributions

P.F., S.S., Y.S., A.S., and L.E. were responsible for the conceptualization and analysis of the experiment. B.M., F.Z., and I.A. were responsible for literature review and manuscript writing. J.P., D.P., L.D., and D.P. provided biomedical engineering support in running the metabolic modeling pipeline. SK was responsible for curating the RISK transcriptomic dataset. FF and DG were responsible for the methods and results of the metabolomics work. JM grew organoids, and helped with data interpretation and manuscript writing.

# Funding

Research reported in this publication was supported by National Institutes of Health (NIH) through the National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK) under award numbers K23DK117061-01A1 (Syed), R01DK132369 (Syed) and R01DK087694 (Kugathasan), as well as the Litwin IBD Pioneers Award of the Crohn's & Colitis Foundation (Syed).

# **Competing interests**

The authors declare no competing interests.

# Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/ 10.1038/s41598-022-26816-5.

Correspondence and requests for materials should be addressed to J.A.P. or S.S.

Reprints and permissions information is available at www.nature.com/reprints.

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

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2022