





Original research

Host transcriptome signatures in human faecal-washes predict histological remission in patients with IBD

Bella Ungar ^{1,2}, Miri Yavzori,¹ Ella Fudim,¹ Orit Picard,¹ Uri Kopylov,¹ Rami Eliakim,¹ Dror Shouval,³ Yishai Levin,⁴ Alon Savidor,⁴ Shani Ben-Moshe,² Rita Manco,² Stav Dan ², Adi Egozi,² Keren Bahar Halpern,² Chen Mayer,⁵ Iris Barshack,⁵ Shomron Ben-Horin ¹, Shalev Itzkovitz ²

► Additional supplemental material is published online only. To view, please visit the journal online (<http://dx.doi.org/10.1136/gutjnl-2021-325516>).

For numbered affiliations see end of article.

Correspondence to

Shalev Itzkovitz, Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel; shalev.itzkovitz@weizmann.ac.il

Received 29 June 2021

Accepted 8 December 2021

ABSTRACT

Background Colonoscopy is the gold standard for evaluation of inflammation in inflammatory bowel diseases (IBDs), yet entails cumbersome preparations and risks of injury. Existing non-invasive prognostic tools are limited in their diagnostic power. Moreover, transcriptomics of colonic biopsies have been inconclusive in their association with clinical features.

Aims To assess the utility of host transcriptomics of faecal wash samples of patients with IBD compared with controls.

Methods In this prospective cohort study, we obtained biopsies and faecal-wash samples from patients with IBD and controls undergoing lower endoscopy. We performed RNAseq of biopsies and matching faecal-washes, and associated them with endoscopic and histological inflammation status. We also performed faecal mass-spectrometry proteomics on a subset of samples. We inferred cell compositions using computational deconvolution and used classification algorithms to identify informative genes.

Results We analysed biopsies and faecal washes from 39 patients (20 IBD, 19 controls). Host faecal-transcriptome carried information that was distinct from biopsy RNAseq and faecal proteomics. Transcriptomics of faecal washes, yet not of biopsies, from patients with histological inflammation were significantly correlated to one another ($p=5.3 \times 10^{-12}$). Faecal-transcriptome had significantly higher statistical power in identifying histological inflammation compared with transcriptome of intestinal biopsies (150 genes with area under the curve >0.9 in faecal samples vs 10 genes in biopsy RNAseq). These results were replicated in a validation cohort of 22 patients (10 IBD, 12 controls). Faecal samples were enriched in inflammatory monocytes, regulatory T cells, natural killer-cells and innate lymphoid cells.

Conclusions Faecal wash host transcriptome is a statistically powerful biomarker reflecting histological inflammation. Furthermore, it opens the way to identifying important correlates and therapeutic targets that may be obscured using biopsy transcriptomics.

INTRODUCTION

Mucosal healing, assessed via colonoscopy, is the desired outcome in therapy of inflammatory bowel disease (IBD) patients.¹ Histological healing (as per

Significance of this study

What is already known on this subject?

- Colonoscopy is the gold standard for evaluation of tissue inflammation and treatment selection in inflammatory bowel diseases (IBDs), yet entails cumbersome preparations and risks of injury.
- Transcriptomics of colonic biopsies have been inconclusive in their association with clinical features.

What are the new findings?

- Transcriptomics of faecal washes, yet not of biopsies, from patients with histological inflammation were significantly correlated to one another ($p=5.3 \times 10^{-12}$).
- Faecal wash transcriptome was significantly more associated with histological inflammation compared with transcriptome of intestinal biopsies (150 genes with area under the curve >0.9 in faecal wash samples vs 10 genes in biopsy RNAseq).
- Faecal wash samples were enriched in inflammatory monocytes, regulatory T cells, natural killer-cells and innate lymphoid cells.

How might it impact on clinical practice in the foreseeable future?

- Faecal wash host transcriptome is a statistically powerful biomarker reflecting histological inflammation.
- Faecal wash host transcriptome will assist us in diagnosis, disease monitoring and treatment selection in patients with IBD.

mucosal biopsy obtained during colonoscopy and analysed by a pathologist) is considered a more stringent therapeutic goal.² In a recent publication on 101 patients with ileal Crohn's disease (CD) in clinical remission, histological healing, but not endoscopic healing, was associated with decreased risk of clinical relapse, medication escalation or corticosteroid use.³ Biological therapies have revolutionised therapy for moderate to severe IBD.⁴ While 50%–60% of patients significantly improve



© Author(s) (or their employer(s)) 2022. No commercial re-use. See rights and permissions. Published by BMJ.

To cite: Ungar B, Yavzori M, Fudim E, *et al.* *Gut* Epub ahead of print: [please include Day Month Year]. doi:10.1136/gutjnl-2021-325516

with biologics and experience less hospitalisations and surgeries, many patients are either primary non-responders or experience loss of response over time.⁵ Non-invasive markers that may provide information on histological inflammation, and therefore predict patient prognosis or response to therapies, are critically needed.

Several studies performed RNA sequencing of colonic biopsies obtained during lower endoscopies, with the aim of staging the disease and predicting therapeutic outcomes.^{6,7} Furthermore, certain mucosal micro-RNA and long noncoding RNA have been associated with IBD natural history.^{8,9} Recent studies used single cell RNA sequencing (scRNAseq) and single cell mass-cytometry of IBD biopsy samples to reveal distinct populations and genes that are altered in specific disease states.^{10–12} In addition to transcriptomics, unique DNA methylation patterns have been identified in biopsies of patients with IBD compared with controls.¹³ Data from RNA bulk sequencing of intestinal biopsies has also been integrated with genome-wide-associations to identify genes most associated with regulatory pathways in IBD.¹⁴ Nevertheless, an outstanding challenge of the analysis of biopsies is that they provide localised information and may miss out on inflammatory processes, especially in cases where endoscopic inflammation is not apparent.

A complementary method to assess intestinal inflammation is the use of faecal samples. A recent study demonstrated that patients with active CD had a distinct microRNA profile measured in their stool.¹⁵ Faecal proteomics can also inform on intestinal inflammation status. Indeed, calprotectin, a leucocyte protein, is a widely applied biomarker of intestinal inflammation. Nevertheless, the calprotectin assay is limited in sensitivity and specificity and only few additional proteins have been shown to be both resistant to proteolysis and associated with inflammation.^{16,17} An advantage of faecal samples is that they may provide broad sampling of processes that occur throughout the gastrointestinal tract. Recent works demonstrated that faecal host transcriptomes may carry prognostic information related to colorectal cancer,¹⁸ however, the utility of this approach to staging and prognosis in IBD has not been explored.

Here, we perform transcriptomics of host mRNA of faecal washes (faecal samples obtained by suction of colonic faecal fluid at the beginning of a colonoscopy, before any through-the-scope washing is applied) and demonstrate that this approach provides markers for identifying histological inflammation. Faecal wash host transcriptomics enables broad sampling of the shed cells throughout the colonic axis, preferentially capturing the immune cells involved in intestinal inflammation. It can, therefore, constitute an alternative, potentially less invasive, method for predicting patient outcome and tailoring personalised medicine.

METHODS

Patient population

In this cross-sectional study, patients arriving for a lower endoscopy at Sheba Medical Centre Gastroenterology institute were recruited. For this pilot study, we deliberately focused on patients with IBD with inflammation of the colon, so that the study groups included patients with ulcerative colitis (UC) or Crohn's colitis, or healthy controls. All control patients performed lower endoscopy for colorectal screening purposes and patients with IBD underwent the procedure due to clinical indications (screening for dysplasia/assessment of disease status). Clinical and demographic parameters were obtained from patients' computerised files.

Sample collection

On endoscopy, biopsies (two consecutive biopsies per patient—'double bite') from the sigmoid colon were obtained and faecal fluid was suctioned from the sigmoid colon at the beginning of the procedure, before any through-the-scope washing was applied. In patients with endoscopic inflammation of the sigmoid colon, the biopsies were obtained from the inflamed area, adjacent to ulcers, if detected. Samples were snap-frozen in liquid nitrogen and stored in -80°C until further analysis. In addition, stool samples were obtained from four patients (two patients with IBD and two controls) for concomitant proteomics analysis and stool calprotectin measurements. Stool calprotectin was also measured in additional 22 patients' faecal washes, comprising the validation cohort.

Study outcomes

The primary outcome was to map the transcriptomic profile of faecal washes in different patient groups (control, IBD with or without endoscopic and histological inflammation) and to identify biomarkers for classifying these groups. Secondary outcomes included a comparison of faecal washes with colonic biopsies and inference of the cellular composition of the faecal washes using computational deconvolution based on scRNAseq data.

Exclusion criteria

- ▶ Patients younger than 18.
- ▶ Undetermined diagnosis of UC or CD (IBD unclassified).
- ▶ Missing clinical/demographic data.
- ▶ Patients with active endoscopic inflammation in the right colon only.

Biomarker measurements

Stool calprotectin was measured using a commercially available ELISA assay (Quantum Blue Calprotectin Quantitative Lateral Flow Assay, LF-CAL25, Buhlmann laboratories, Switzerland).

Definition of clinical remission

Clinical status was determined by HBI (Harvey-Bradshaw index) for CD and by SCCAI (Simple Clinical Colitis Activity Index) for UC patients. Clinical remission was defined as HBI <5 for CD patients and SCCAI ≤ 3 for UC patients.^{19,20}

Definition of mucosal healing and histological healing

Endoscopic and histological inflammation were graded according to standardised indices and by blinded gastroenterologists and pathologists respectively. Endoscopic scores were determined prospectively during lower endoscopy. Mucosal healing was defined as absence of ulcers and lack of inflammation on endoscopic examination, for CD and UC, respectively.²¹ Histological inflammation was determined by a certified pathologist at the Sheba medical centre based on biopsies from the same sigmoid colon region used for the biopsy transcriptomics. Histological healing was retrospectively defined as grade 0 on the Nancy histological index. Histologically active disease was graded between 1 and 4 by certified blinded pathologist.²²

RNA extraction

For colonic biopsies—snap frozen tissues (2×2 mm) were thawed in 300 μL Tri-reagent and mechanically homogenised with bead beating, followed by a short centrifugation step to pull down beads and any tissue left-overs. For colonic washes—Tri-reagent was added at a ratio of 3:1, samples were allowed to thaw on ice followed by thorough mixing. A first centrifugation step was

used (1 min, 18 000 rpm) to eliminate faecal solids. Following this, ethanol was added in a ratio of 1:1 to the supernatant from the previous step and continued according to the manufacturer instructions of Direct-zol mini and micro prep kit (ZYMO research, R2052).²³

RNA sequencing of samples

RNA was processed by the mcSCRbseq protocol²⁴ with minor modifications. RT reaction was applied on 10 ng of total RNA with a final volume of 10 μ L (1 \times Maxima RT Buffer, 1 mM dNTPs, 2 μ M TSO* E5V6NEXT, 7.5% PEG8000, 20U Maxima H enzyme, 1 μ L barcoded RT primer). Subsequent steps were applied as mentioned in the protocol. Library preparation was done using Nextera XT kit (Illumina) on 0.6 ng amplified cDNA. Library final concentration of 2 nM was loaded on NextSeq 500/550 (Illumina) sequencing machine aiming at 20 M reads per sample²³ with the following setting: Read1—16 bp, Index1—8 bp, Read2—66 bp.

A total of 8/78 samples in the original cohort failed due to insufficient RNA for extraction / insufficient UMI counts for either wash or biopsy. Only samples which had a UMI count of above 10 000 were included in the analysis. In total 7/39 washes were not included in the analysis due to insufficient RNA for sequencing (2 controls) or due to insufficient UMI counts after sequencing (two UC patients and three controls). One biopsy was also removed due to insufficient UMI count after sequencing (colonic CD patient).

Proteomic analysis

Faecal samples were lysed in lysis buffer containing 5% SDS, proteins were extracted, digested with trypsin, and tryptic peptides were subjected to LC-MS/MS analysis.²⁵ Acquired raw data were analysed using the MaxQuant software while searching against the human protein database, and downstream quantitative comparisons were calculated using the Perseus software.²⁶

Bioinformatics and computational analysis

Illumina output sequencing raw files were converted to FASTQ files using bcl2fastq package. To obtain the UMI counts, FASTQ files were aligned to the human reference genome (GRCh38.91) using zUMI package.²⁷ Statistical analyses were performed with MATLAB R2018b. Mitochondrial genes and non-protein coding genes were removed from the analysis. Protein coding genes were extracted using the annotation in the Ensembl database (BioMart) for reference genome GRCh38 V91, using the R package 'biomaRt' (V2.44.4). Gene expression for each sample was consequently normalised by the sum of the UMIs of the remaining genes. Samples with less than 10 000 UMIs over the remaining genes were removed from the analyses. Clustering and principle component analysis (PCA) were performed in MATLAB using the Zscore-transformed expression matrix. Clustering was done with the matlab function clustergram, using Spearman distances. Differential gene expression was performed using Wilcoxon rank-sum tests and Benjamini-Hochberg FDR corrections. Computational deconvolution was performed using CIBERSORTx²⁸ using signature tables obtained from a single cell atlas of control and UC patients.²⁹ Original cell type annotations were used, but subsequently coarse-grained into small number of cell types (online supplemental table 1). M cells were removed from the analysis due to their low abundance. Receiver Operating Curve analyses were performed using the MATLAB function perfcurve. Gene Set Enrichment Analysis (GSEA)³⁰ was performed over the Hallmark and Kegg gene sets.

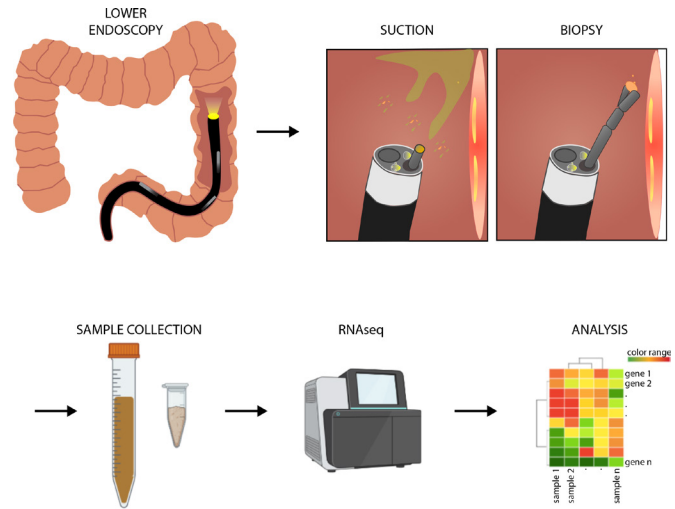


Figure 1 Illustration of experimental layout.

Pathway analysis for the top-classifying faecal wash genes was performed using EnrichR.³¹ N-way analysis of variance was applied to analyse expression profiles of different genes per sample according to Nancy score of histological inflammation.

RESULTS

Cohort characteristics

In total, 39 biopsies and 39 matching faecal wash samples were obtained from 16 patients with UC, 4 patients with Crohn's colitis and 19 control subjects undergoing colonoscopy at Sheba Medical centre. Pairs of biopsies and matching washes were obtained concomitantly (figure 1, table 1). Control patients were those undergoing lower endoscopy for screening purposes, recommended over the age of 50, and, therefore, they were significantly older than the IBD group ($p < 0.0008$), with more comorbidities, other than IBD ($p = 0.0015$, table 1). Eleven (58%) of all patients with IBD were treated with immunomodulator/biological therapy and five (26%) were on concomitant steroids at time of enrolment. Nine (47%) of the patients were in clinical remission, twelve (63%) were in endoscopic remission and seven (37%) achieved histological remission as determined on the day of the lower endoscopy. Five faecal wash samples were excluded from the analysis due to technical drop-outs. Validation cohort consisted of 10 patients with IBD and 12 controls (see online supplemental table 2 for demographic and clinical parameters).

Faecal wash host transcriptome is more informative than biopsy transcriptome in classifying patient disease status

We performed bulk RNA sequencing of all samples using the UMI-based mcSCRbseq protocol (see Methods) and mapped the reads to the human genome (online supplemental table 3). We found that gene expression signatures of colonic biopsies were different from those of colonic washes (figure 2A,B). Biopsy samples with histological inflammation were not distinct from biopsy samples of patients without histological inflammation in the PCA or clustering analysis (figure 2B,C). In contrast, colonic faecal wash samples showed a clear separation between samples with and without histological inflammation (figure 2D).

We next sought to quantify the comparative ability of biopsy and faecal wash transcriptomics to inform on histological inflammation (evaluated by the Nancy score). To this end, we examined correlations between gene expression profiles of pairs of samples

Table 1 Patients' demographic and clinical characteristics of IBD versus controls patients

	IBD	Controls	P value
N (%)	20 (51)	19 (49)	
Age, years (median, IQR)	49 (36–56)	67 (58–73)	0.0008
Female gender (%)	11 (55)	11 (58)	0.85
Smoking at induction, n (%)	0 (0)	3 (15)	0.12
Weight, kg - median (IQR)	80 (69–87)	68 (63.6–79)	0.9
Concomitant medical condition, n (%)	14 (70)	16 (84)	0.0015
Disease duration, years—median (IQR)	14.5 (4–31)		
Previous surgery, n (%)	1 (5)		
Concomitant 5-aminosalicylic acid (5-ASA) therapy, n (%)	3 (14)		
Concomitant immunomodulator therapy, n (%)	1 (5)		
Concomitant steroids, n (%)	5 (24)		
Concomitant biological therapy, n (%)	10 (48)		
Disease location			
CD, ileocolitis n (%)	4 (100)*		
UC, pancolitis n (%)	8 (50)†		
UC, left-sided colitis n (%)	8 (50)†		
Clinical remission at time of endoscopy (median, IQR)‡	9 (47)		
Endoscopic remission at time of endoscopy (median, IQR)	12 (63)		
Histologic remission at time of endoscopy (median, IQR)	7 (37)		

*Out of total CD patients (n=4).

†Out of total UC patients (n=16).

‡Clinical remission was defined using the HBI and SCCAI scores for CD and UC, respectively.

§Bold indicates statistically significant P values.

CD, Crohn's disease; HBI, Harvey-Bradshaw index; IBD, inflammatory bowel disease; SCCAI, Simple Clinical Colitis Activity Index; UC, ulcerative colitis.

that both have histological inflammation compared with mixed pairs (one with and one without histological inflammation). There was no significant difference between transcriptomic profiles obtained from biopsies with histological inflammation compared with correlations between mixed biopsies (with or without histological inflammation) ($p=0.98$). However, faecal washes with histological inflammation were significantly more correlated to each other than mixed washes (figure 2E, $p=5.3 \times 10^{-12}$). Our analysis, therefore, demonstrates that faecal wash transcriptomics may provide signatures for classifying patients with or without histological inflammation.

When assessing concordance of faecal washes and biopsies with endoscopic, rather than histological inflammation, similarly, faecal washes, rather than biopsies, were associated with endoscopic remission ($p=0.004$ vs $p=0.6$, respectively, online supplemental figure 1A,B). Furthermore, we observed statistically higher concordance of faecal wash transcriptomics with histological inflammation status, compared with biopsy transcriptomics when stratifying according to patients' age or biological therapy (online supplemental figure 1C–F). The expression signatures of faecal washes were generally more similar to their matching biopsies than to other biopsies (online supplemental figure 2)

Gene expression patterns are significantly different between faecal washes of patients with and without histological inflammation

We next asked which genes differed in expression between the faecal washes from patients with and without histological

inflammation (figure 3A). We found that 1168 genes out of 3999 highly expressed genes (normalised expression above 5×10^{-5}) were significantly different between the groups (q -value <0.1 , Wilcoxon rank sum tests with Benjamini-Hochberg false discovery rate correction). Genes that were upregulated in inflamed sample washes included S100A8 and S100A9, encoding the subunits of the calprotectin protein, as well as other immune-related genes such as NFKBIA, TNF, TNFRSF1B, CCR1, STAT1 and IFIT3. Using GSEA³⁰ we found that washes from inflamed patients were enriched in genes associated with TNF α signalling, IL6 signalling, chemokine signalling pathway and the JAK STAT pathway, and depleted in pathways such as glycolysis and glutathione metabolism (figure 3B).

To examine whether faecal wash host transcriptomics further carries information indicative of the grade of histological inflammation, we used N-way analysis of variance among the samples of patients with histological inflammation to identify genes that exhibit significant differences among the Nancy score classes (graded as 0–4 according to increasing grade of histological inflammation). We found that 703 genes were significantly variable in the faecal wash measurements, vs only 95 genes in the biopsy samples (q -value <0.25 , online supplemental figure 4A,B). Moreover, expression levels of the genes that were most variable across the Nancy score classes increased more sharply with severity in the faecal wash transcriptomics compared with the biopsy transcriptomics (online supplemental figure 4C).

Inflamed faecal washes exhibit distinct cellular composition

We inferred cell compositions among inflamed vs non inflamed faecal washes and biopsies using CIBERSORTx²⁸ (Methods/Bioinformatics and computational analysis), using gene expression signatures of human colonic cell types that we parsed based on a recent scRNAseq study²⁹ (Methods, online supplemental figure 3). We found elevated representation of distinct immune cell subtypes in the washes of patients with histological inflammation (figure 4). Cell types that were elevated in faecal washes from patients with histological inflammation included regulatory T cells ($p=3.6 \times 10^{-4}$), natural killer (NK) cells ($p=5.2 \times 10^{-3}$), inflammatory monocytes (1.4×10^{-8}) and innate lymphoid cells (ILCs, $p=2.6 \times 10^{-6}$). The increased differential representation of these immune subsets was higher in faecal washes when compared with biopsies (figure 4B, online supplemental table 1). Non-inflamed washes had a significantly higher representation of enterocytes ($p=1.8 \times 10^{-3}$) and goblet cells (5.4×10^{-8}) compared with inflamed washes.

More genes have expression levels that are highly predictive of histological inflammation in faecal washes compared with biopsies

We next sought to assess whether expression levels of individual genes can classify samples as belonging to patients with or without histological inflammation. We performed receiver operating characteristic curve (ROC) analysis for all genes in the biopsies and faecal washes and examined the area under the curve (AUC, figure 5A). We found that in the washes, the expression levels of multiple individual genes were significantly more predictive of histological inflammation compared with the biopsies. This was evident by the significantly higher AUC of the 5% genes with highest AUC levels in both groups ($p=1.5 \times 10^{-78}$, figure 5B). Faecal washes included 150 genes with AUC >0.9 , whereas biopsies had only 10 such genes (figure 5C–E). Pathway analysis demonstrated that the 5% genes with the highest AUC in faecal washes were enriched for TNF α signalling via NF- κ B, and

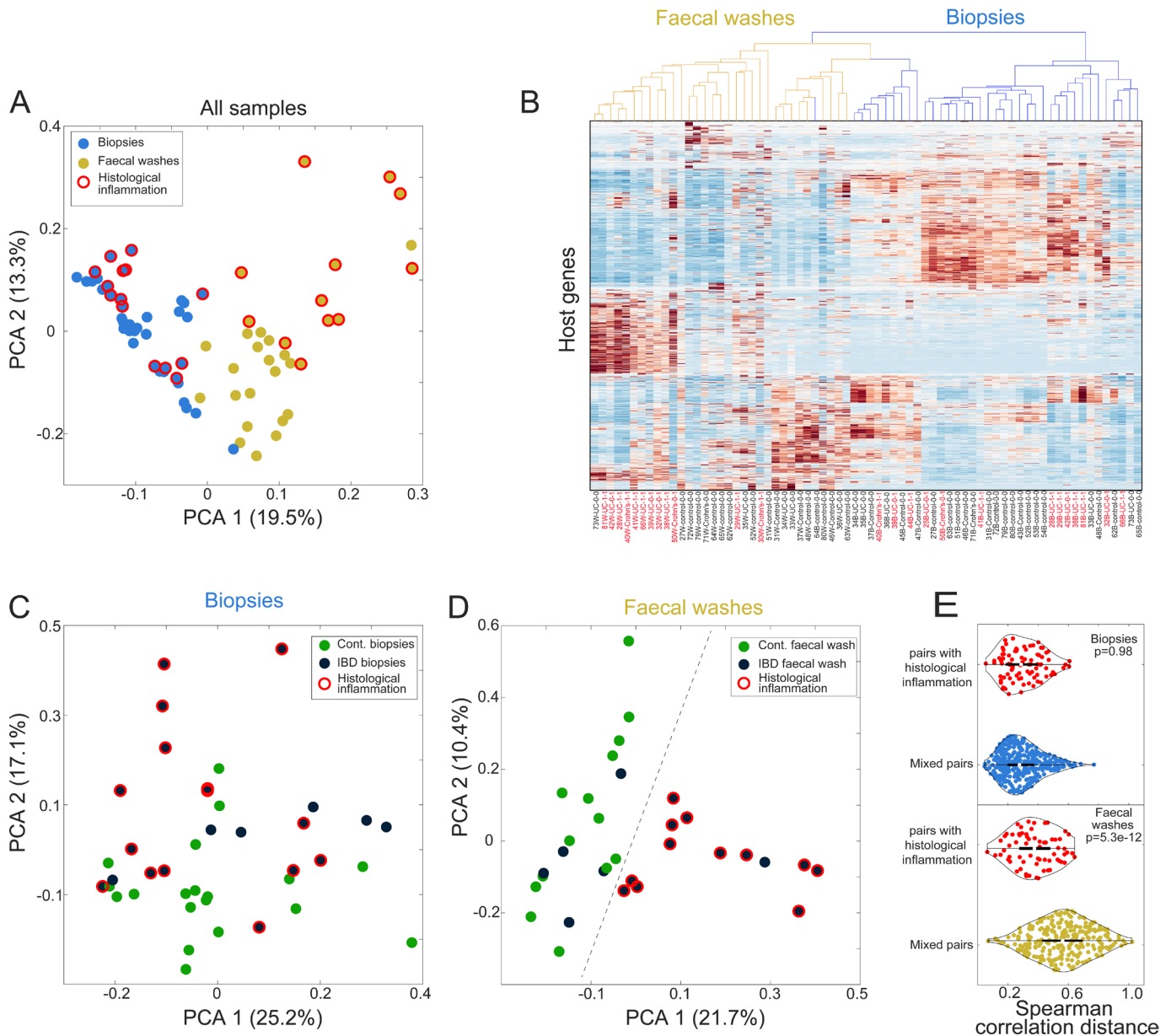


Figure 2 Faecal wash gene expression patterns are more indicative of histological inflammation compared with those of biopsies. (A) Principal component analysis (PCA) plot showing biopsies (blue circles) and faecal washes (brown circles). Red outer circles denote samples that correspond to patients with histological inflammation determined based on pathology examination of the colonic biopsies. (B) Hierarchical clustering of faecal wash samples (brown branches) and colonic biopsies (blue branches). Samples corresponding to patients with active histological inflammation are marked in red. Naming nomenclature: sample name-condition-endoscopic inflammation (0/1)—histological inflammation (0/1). Genes included had maximal expression above 5×10^{-4} of the samples' UMIs. (C, D) PCA plots of biopsies (C), and faecal wash samples (D). Red outer circles denote patients with IBD with corresponding histological inflammation. (E) Violin plots demonstrating that the correlation distances between transcriptomic signatures of pairs of samples that both have histological inflammation (red dots) are significantly smaller than the distances between mixed samples with and without histological inflammation when examining faecal washes (brown dots, bottom) but not when examining biopsies (blue dots, top). White circles are medians, black boxes denote the 25th–75th percentiles. IBD, inflammatory bowel disease. UMI, unique molecular identifier.

inflammatory response, interferon α and γ signalling pathways, and IL-6/JAK/STAT signalling (online supplemental table 4).

Validation cohort demonstrates high overlap between inflammation-predicting faecal wash genes

To examine the robustness of our findings we replicated the study on an independent validation cohort of 22 additional patients (10 IBD and 12 controls, UMI-based reads listed in online supplemental table 3). In the validation cohort, gene expression values in the faecal washes were more indicative of

histological inflammation than in biopsies (online supplemental figure 5A–C). A total of 155 genes had AUC higher than 0.9 in the faecal washes vs only 5 genes in the biopsies. Moreover, we found that the overlap in the two cohorts between the inflammation-predicting faecal wash genes was highly significant (47%, 219/467, of the genes that were in the top 10% AUC values overlapped, hypergeometric $p < 1e-15$). Online supplemental figure 5G–I shows examples of ROC curves over the validation cohort for three of the top classifying faecal wash genes detected in the original cohort. Online supplemental table

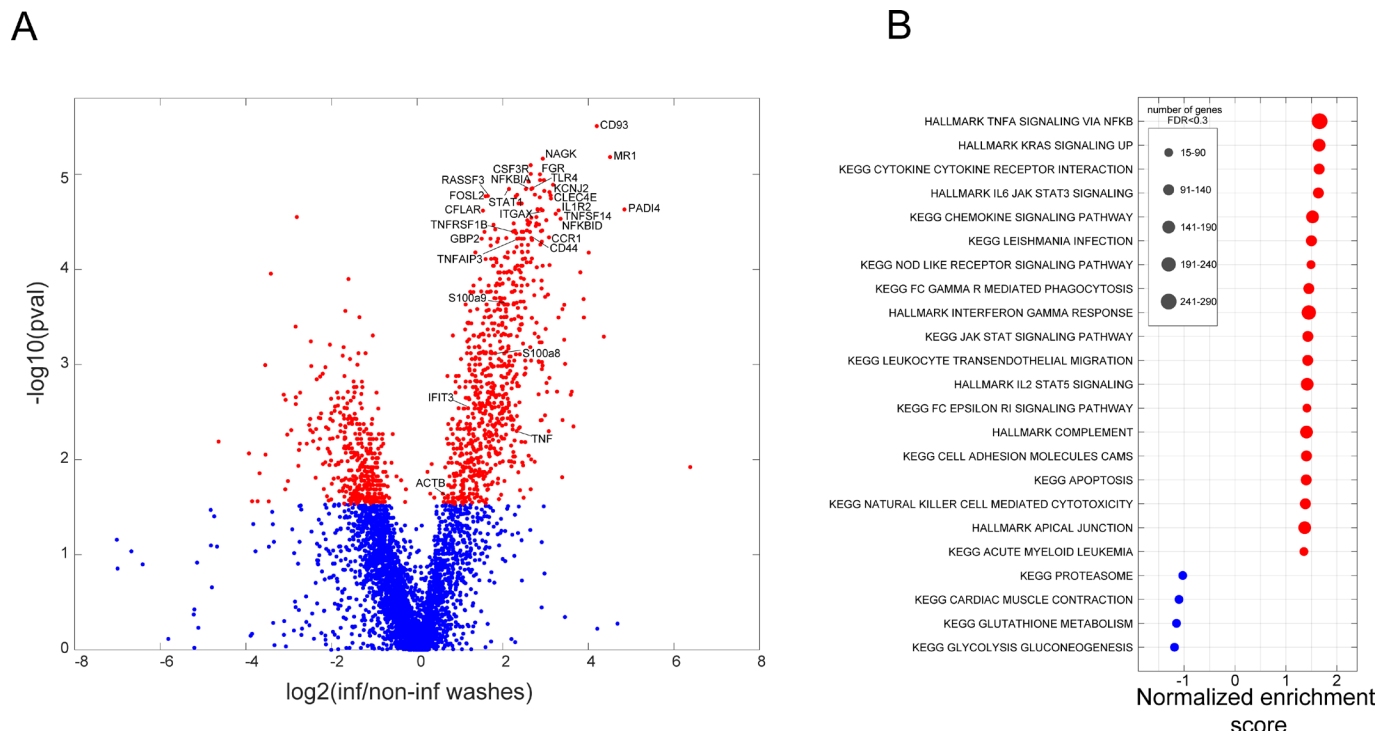


Figure 3 Differentially expressed genes between inflamed and non-inflamed faecal washes. (A) Volcano plot, each dot is a gene, x-axis is the \log_2 -ratio of expression between samples with and without histological inflammation, y axis is $-\log_{10}$ (p value), where p value is computed using Wilcoxon rank-sum tests. Genes with corresponding q-values below 0.1 are marked in red (q-values computed using Benjamini-Hochberg FDR correction). Names of representative up-regulated genes are shown. (B) GSEA analysis over the Hallmark and Kegg sets. Shown are all gene sets with q-value < 0.3 . Inflamed washes (red circles) were associated with immune cell pathways, while non-inflamed washes (blue circles) expressed more epithelial cell related pathways. Naming nomenclature: sample name-condition-endoscopic inflammation (0/1)—histological inflammation (0/1). FDR, false discovery rate; GSEA, gene set enrichment analysis.

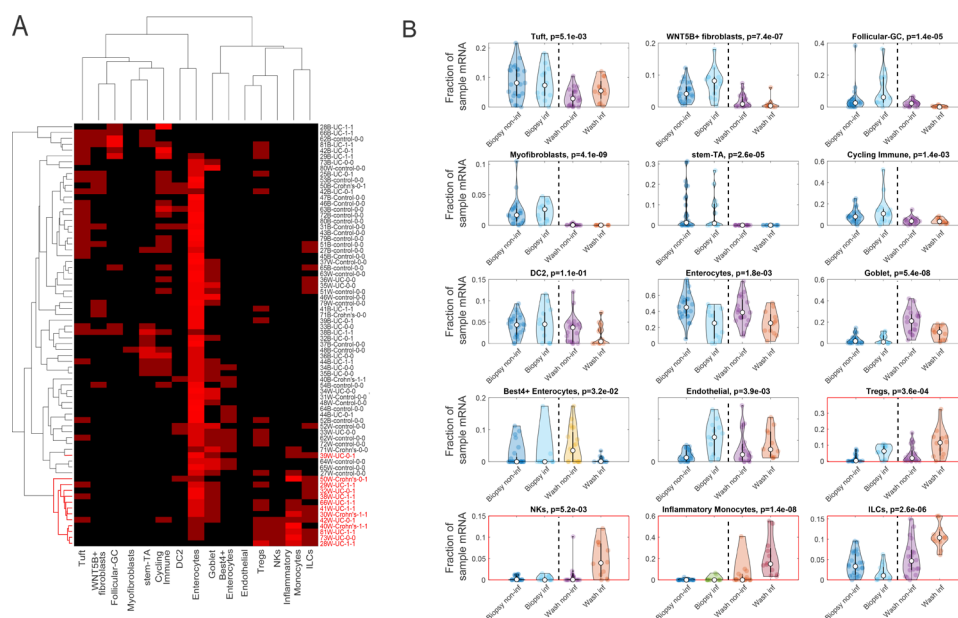


Figure 4 Cell compositions of inflamed versus non inflamed faecal washes and biopsies, inferred by computational deconvolution. (A) hierarchical clustering of cell type representation in faecal wash samples and colonic biopsies. Faecal washes from patients with histological inflammation are marked in red. (B) Inferred relative representation of genes associated with different cell types in histologically inflamed and non-inflamed colonic biopsies and faecal washes. Immune-related cell types, more abundant in the faecal washes of patients with histological inflammation, are marked with a red box. Naming nomenclature: sample name-condition-endoscopic inflammation (0/1)—histological inflammation (0/1). White circles are medians, grey boxes denote the 25th–75th percentiles. P values computed using Kruskal-Wallis tests.

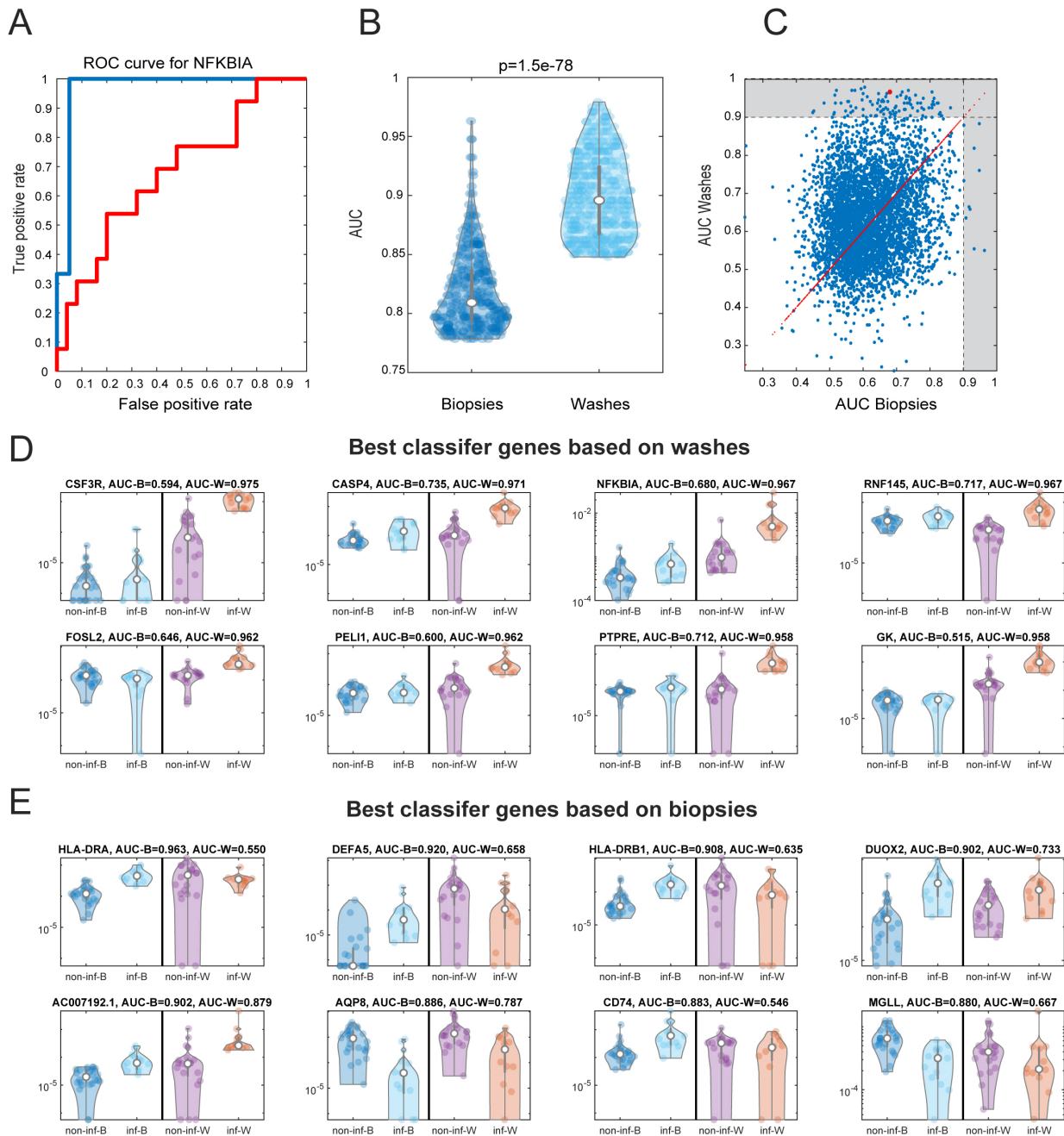


Figure 5 Expression of individual genes in faecal washes has a higher statistical power in classifying histological inflammation compared with biopsy gene expression. (A) ROC curve example for the gene NFKBIA using faecal washes (blue, AUC=0.97) and biopsies (red, AUC=0.67). (B) AUC of 5% genes with the highest AUC for biopsies and washes. The AUC of the top classifier genes is significantly higher for faecal washes compared with biopsies ($p=1.85 \times 10^{-72}$). (C) Comparison of AUC for individual genes based on biopsies (x axis) and faecal washes (y axis). NFKBIA (red dot) is shown as an example. Grey boxes mark the top AUC (>0.9) for both groups. Faecal washes contain 150 genes with AUC >0.9 whereas biopsies contain only 10 such genes. (D, E) Expression levels for the eight genes with the highest AUC levels for washes (D) and biopsies (E). White circles are medians, grey boxes mark the 25th–75th percentiles. AUC, area under the curve; inf-W, inflamed faecal washes; NFKBIA, NF-kappa-B inhibitor alpha; non-inf-B, non-inflamed biopsies, inf-B, inflamed biopsies, non-inf-W, non-inflamed faecal washes; ROC, receiver operating characteristic.

5 presents this set of overlapping genes that are indicative of histological inflammation. Notably, overlap between the most inflammation-predicting biopsy genes was insignificant (10% (51/480), $p=0.27$). The AUC values were significantly correlated between the cohorts in the faecal wash data, yet not in the biopsy data (online supplemental figure 5D,E, $R=0$, $p=0.96$ for biopsies, $R=0.34$, $p=2.3e-130$ for washes, Online supplemental table 6).

We further assessed the predictive power of faecal-wash host transcriptomics in the validation cohort compared with ELISA-based calprotectin levels (online supplemental figure 6A). We found that calprotectin levels had lower statistical power for classification of histological inflammation compared with faecal host transcriptomics (equal error rate of 0.14 for calprotectin vs equal error rate of 0 for 17 of the faecal wash host genes online supplemental figure 6B–D).

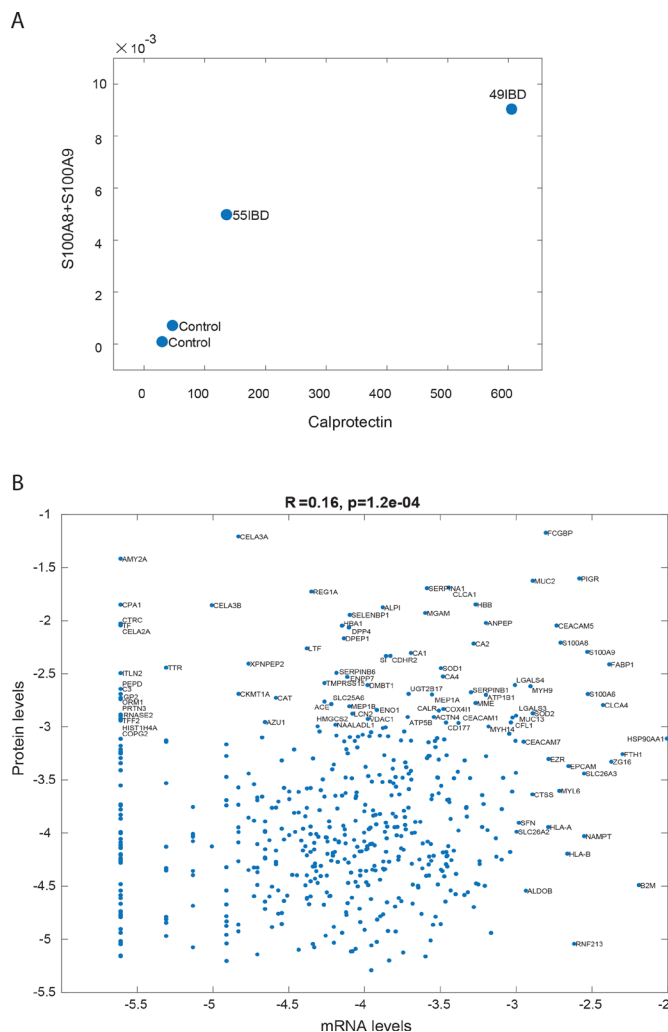


Figure 6 Protein and mRNA levels in faecal washes are only weakly correlated. (A) Faecal calprotectin levels as measured by a commercial ELISA assay are correlated with the mass-spectrometry proteomics levels of the same protein, S100A8, S100A9. Each blue dot denotes a faecal sample. (B) Correlation between protein levels and faecal wash mRNA levels for the same samples. Each dot is the average expression over the four samples. Shown are names of genes with relative expression of either mRNA or protein above $1e^{-3}$.

Faecal wash transcriptomics carries distinct information from faecal proteomics

To assess the information contained by the faecal wash transcriptomics measurements in relation to faecal proteomics we next performed mass spectrometry proteomics of 10 faecal samples (six faecal washes and four stool samples). The six faecal washes had matching faecal wash transcriptomics analyses. Faecal calprotectin levels were measured in the four stool samples. Protein expression of S100A8 and S100A9 were correlated with stool calprotectin levels (figure 6A). Notably, protein and mRNA levels were only weakly correlated ($R=0.16$, $p=1.2 \times 10^{-4}$). Genes with discordant mRNA and protein levels included pancreatic proteins, such as the amylase protein AMY2A, and the elastase proteins CELA2A, CELA3A and CELA3B (figure 6B). These proteins are produced by pancreatic acinar cells and settle on the luminal side of the intestinal epithelium, explaining the lack of mRNAs. Other discordances may represent differential stability of distinct proteins and mRNA species. The faecal host

transcriptomics therefore provides information that is distinct from faecal proteomics.

DISCUSSION

In the current study, we analysed biopsies and faecal washes of 39 IBD and control patients. We found that transcriptomes of faecal washes were significantly more powerful in identifying histological inflammation compared with transcriptomes of intestinal biopsies. This was also demonstrated in a validation cohort of 22 IBD and control patients. The increased statistical power of faecal host transcriptomics over biopsy transcriptomics could be related to two underlying features—the broader sampling of the intestinal wall, and the increased representation of cell populations directly involved in inflammation. Since faecal washes include cells that are shed throughout the gastrointestinal tract, their measurement may detect inflammatory processes that may be missed out when sampling a specific biopsy, therefore providing broad sampling.

We found that faecal washes of patients with histological inflammation were enriched in inflammatory monocytes, regulatory T cells, NK cells and ILCs (figure 4). Consistently, faecal washes from histologically-inflamed samples were enriched in genes associated with TNF α signalling, IL6 signalling, chemokine signalling pathway and NK cytotoxicity, whereas non-inflamed samples expressed mostly genes associated with pathways such as glycolysis and glutathione metabolism (figure 3). This result is in line with recent findings of increased leukocyte trafficking to the intestine and to the gut-associated lymphatic tissue on inflammation.^{32 33} Whereas epithelial cells are constantly shed off in the intestine, the process of active inflammation may involve increased shedding of immune cells. Unlike bulk measurements of biopsies, consisting mainly of fibroblasts and epithelial cells (figure 4), the faecal washes enable zooming in on the immune cell populations, avoiding ‘masking’ by other stromal cell types that are abundant in biopsies.

Faecal microbial RNA analysis has been shown to differentiate between CD and UC and alternations in microbiota diversity have been linked with disease activity,³⁴ however human faecal mRNA has scarcely been studied so far,^{35–38} and has not been applied in IBD. While our analyses focused on faecal washes, preliminary data suggests that host transcriptomics from stool samples may also provide comparable information (Ungar *et al*, in preparation). Challenges involved in stool samples, including the ability to sample mRNA from recently-shed cells, will be addressed in future studies.

In recent years, faecal calprotectin has been established as a leading biomarker assisting in the diagnosis and non-invasive management of IBD. Notably, the calprotectin assay’s sensitivity and specificity is on the order of 80%.³⁹ Our study found only weak correlation between faecal protein levels and matching mRNA expression levels (figure 6B). This discrepancy could be a result of differences in protein and RNA degradation rates, which are often non-correlated.⁴⁰ The similarity between human and bacterial RNAses, which are both abundant in human stool, probably also affects human faecal mRNA instability.⁴¹ Furthermore, several studies have reported elevated levels of faecal protease activity in patients with certain gastro-intestinal diseases including IBD and irritable bowel syndrome (IBS), potentially compromising proteomic, but not necessarily transcriptomics measurements in these diseases.⁴²

This study analysed 39 patients with 78 faecal wash and biopsy samples and additional 22 validation—cohort patients. While this cohort revealed the power of faecal transcriptomics in

identifying histological inflammation, future studies with larger sample sizes could detect differences in various subgroups, such as patients with differential responses to biological therapy, or certain age-groups. Endoscopic inflammation and histological inflammation were graded according to standardised indices and by blinded gastroenterologists and pathologists, yet, as in other studies, sampling errors and variations in scaling are possible.

In conclusion, we have demonstrated that faecal wash host transcriptome constitutes a statistically powerful biomarker in IBD, specifically for the challenging identification of histological inflammation. Expression levels of mRNA in faecal washes discriminate between histologically inflamed and non-inflamed patients significantly better than mRNA expression of colonic biopsies. Pathways common to genes associated with inflammation in faecal washes were those related to TNF α , JAK-kinase and IL-6, central pathways that are thought to be mechanistically associated with IBD pathology. Hence, it seems that the colonic washes constitute an 'extract' of exfoliated inflammatory cells in the gut, which is highly indicative of histological inflammation in patients with IBD. Although faecal washes require gut cleansing, they obviate the need for full colonoscopy, with its discomfort, need for sedation and risk for perforation. Future studies will assess the ability of faecal host transcriptome to predict disease course in specific IBD patient populations, to predict response to different biologics and hence to tailor patient-specific therapy. In addition, faecal host transcriptomics may be applied to classify other intestinal inflammatory diseases such as immunotherapy-related colitis, microscopic colitis, intestinal manifestations of autoimmune illnesses such as Behcet's disease, and perhaps even IBS and colorectal cancer. Finally, corroborating studies will also demonstrate whether stool samples, without bowel lavage, could be similarly analysed.

Author affiliations

¹Department of Gastroenterology, Sheba Medical Center Tel Hashomer & Sackler School of Medicine, Tel-Aviv University, Ramat Gan, Israel

²Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, Israel

³Institute of Gastroenterology, Nutrition and Liver Diseases, Schneider Children's Medical Center of Israel, & Sackler School of Medicine, Tel-Aviv University, Petah Tikva, Israel

⁴The De Botton Institute for Protein Profiling, The Nancy and Stephen Grand Israel National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot, Israel

⁵Department of Pathology, Sheba Medical Center Tel Hashomer & Sackler School of Medicine, Tel-Aviv University, Ramat Gan, Israel

Contributors SI and BU conceived the study and drafted the manuscript; SI performed the bioinformatics analysis and is the guarantor of the study; BU was involved in study conception, performed experiments, analyses and interpretation of data; MY, EF, OP, YL and AS took part in sample retrieval and analysis. SB-M, RM, AE and KBH took part in study design and data analysis. UK, RE and DS participated in study design and drafting of the manuscript. IB and SB-H participated in data interpretation and in critical revision of the manuscript. SD and CM were involved in final revision of the manuscript. All authors have approved the final draft submitted.

Funding SI is supported by the Helen and Martin Kimmel Institute for Stem Cell Research, the Wolfson Family Charitable Trust & Wolfson Foundation, the Edmond de Rothschild Foundations, the Fannie Sherr Fund, the Dr. Beth Rom-Rymer Stem Cell Research Fund, the Minerva Stiftung grant, the Israel Science Foundation grant No. 1486/16, the Broad Institute-Israel Science Foundation grant No. 2615/18, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program grant No. 768956, the Bert L. and N. Kuggie Vallee Foundation and the Howard Hughes Medical Institute (HHMI) international research scholar award. Bella Ungar is supported by the Talpiot Medical Leadership Program, The Chaim Sheba Medical Center and by the Kitter Scholarship, Sheba Medical Center. Shomron Ben-Horin is supported by the Leona and Harry Helmsley Charitable Trust. The use of faecal host transcriptomics for evaluation of IBD disease pathology and outcome is currently being patented. Figure 1 was created with Biorender.com.

Competing interests SB-H received consulting and advisory board fees and/or research support from AbbVie, MSD, Janssen, Takeda and CellTrion. UK received

speaker and advisory fees from: AbbVie, Janssen, Gilead, MSD, Medtronic, Takeda and research support from: Janssen, Medtronic, Takeda. RE received consultant and speaker fees from Janssen, AbbVie, Takeda and Medtronic. BU received consultation fees from Neopharm, Takeda, Janssen and AbbVie. DS received research support from Takeda and consultation and lecturing fees from AbbVie.

Patient consent for publication Not applicable.

Ethics approval This study was approved by Sheba Medical Center Helsinki committee reference number 743.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement All data relevant to the study are included in the article or uploaded as online supplemental information. All data relevant to the study is included as online supplemental tables.

Supplemental material This content has been supplied by the author(s). It has not been vetted by BMJ Publishing Group Limited (BMJ) and may not have been peer-reviewed. Any opinions or recommendations discussed are solely those of the author(s) and are not endorsed by BMJ. BMJ disclaims all liability and responsibility arising from any reliance placed on the content. Where the content includes any translated material, BMJ does not warrant the accuracy and reliability of the translations (including but not limited to local regulations, clinical guidelines, terminology, drug names and drug dosages), and is not responsible for any error and/or omissions arising from translation and adaptation or otherwise.

ORCID iDs

Bella Ungar <http://orcid.org/0000-0002-6078-2249>

Stav Dan <http://orcid.org/0000-0003-4469-8247>

Shomron Ben-Horin <http://orcid.org/0000-0002-3984-4580>

Shalev Itzkovitz <http://orcid.org/0000-0003-0685-2522>

REFERENCES

- Neurath MF, Travis SPL. Mucosal healing in inflammatory bowel diseases: a systematic review. *Gut* 2012;61:1619–35.
- Dal Buono A, Roda G, Argollo M, et al. Histological healing: should it be considered as a new outcome for ulcerative colitis? *Expert Opin Biol Ther* 2020;20:407–12.
- Christensen B, Erlich J, Gibson PR, et al. Histologic healing is more strongly associated with clinical outcomes in ileal Crohn's disease than endoscopic healing. *Clin Gastroenterol Hepatol* 2020;18:2518–25.
- Paramsothy S, Rosenstein AK, Mehandru S, et al. The current state of the art for biological therapies and new small molecules in inflammatory bowel disease. *Mucosal Immunol* 2018;11:1558–70.
- Ben-Horin S, Chowers Y. Review article: loss of response to anti-TNF treatments in Crohn's disease. *Aliment Pharmacol Ther* 2011;33:987–95.
- Taman H, Fenton CG, Hensel IV, et al. Transcriptomic landscape of treatment-naïve ulcerative colitis. *J Crohns Colitis* 2018;12:327–36.
- Gaujoux R, Starosvetsky E, Maimon N, et al. Cell-centred meta-analysis reveals baseline predictors of anti-TNF α non-response in biopsy and blood of patients with IBD. *Gut* 2019;68:604–14.
- He C, Shi Y, Wu R, et al. miR-301a promotes intestinal mucosal inflammation through induction of IL-17A and TNF- α in IBD. *Gut* 2016;65:1938–50.
- Rankin CR, Shao L, Elliott J, et al. The IBD-associated long noncoding RNA *IFNG-AS1* regulates the balance between inflammatory and anti-inflammatory cytokine production after T-cell stimulation. *Am J Physiol Gastrointest Liver Physiol* 2020;318:G34–40.
- Mitsialis V, Wall S, Liu P, et al. Single-Cell analyses of colon and blood reveal distinct immune cell signatures of ulcerative colitis and Crohn's disease. *Gastroenterology* 2020;159:591–608.
- Huang B, Chen Z, Geng L, et al. Mucosal profiling of pediatric-onset colitis and IBD reveals common Pathogenics and therapeutic pathways. *Cell* 2019;179:1160–76.
- Kinchen J, Chen HH, Parikh K, et al. Structural remodeling of the human colonic mesenchyme in inflammatory bowel disease. *Cell* 2018;175:372–86.
- Howell KJ, Kraiczky J, Nayak KM, 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* 2018;154:585–98.
- Peters LA, Perrigoue J, Mortha A, et al. A functional genomics predictive network model identifies regulators of inflammatory bowel disease. *Nat Genet* 2017;49:1437–49.
- Verdier J, Breunig IR, Ohse MC, et al. Faecal micro-RNAs in inflammatory bowel diseases. *J Crohns Colitis* 2020;14:110–7.
- Lehmann FS, Burri E, Beglinger C. The role and utility of faecal markers in inflammatory bowel disease. *Therap Adv Gastroenterol* 2015;8:23–36.
- Rokkas T, Portincasa P, Koutroubaki IE. Fecal calprotectin in assessing inflammatory bowel disease endoscopic activity: a diagnostic accuracy meta-analysis. *J Gastrointest Liver Dis* 2018;27:299–306.

- 18 Barnell EK, Kang Y, Campbell KM, *et al.* Prospective clinical study using expressed sequencing variants on stool-derived eukaryotic RNA transcripts (seRNA) for colorectal cancer screening. *JCO* 2019;37:516.
- 19 Walmsley RS, Ayres RC, Pounder RE, *et al.* A simple clinical colitis activity index. *Gut* 1998;43:29–32.
- 20 Vermeire S, Schreiber S, Sandborn WJ, *et al.* Correlation between the Crohn's disease activity and Harvey-Bradshaw indices in assessing Crohn's disease severity. *Clin Gastroenterol Hepatol* 2010;8:357–63.
- 21 Ungar B, Levy I, Yavne Y, *et al.* Optimizing Anti-TNF- α Therapy: Serum Levels of Infliximab and Adalimumab Are Associated With Mucosal Healing in Patients With Inflammatory Bowel Diseases. *Clin Gastroenterol Hepatol* 2016;14:550–7.
- 22 Marchal-Bressenot A, Salleron J, Boulagnon-Rombi C, *et al.* Development and validation of the Nancy histological index for UC. *Gut* 2017;66:43–9.
- 23 Bahar Halpern K, Massalha H, Zwick RK, *et al.* Lgr5+ telocytes are a signaling source at the intestinal villus tip. *Nat Commun* 2020;11:1936.
- 24 Bagnoli JW, Ziegenhain C, Janjic A, *et al.* Sensitive and powerful single-cell RNA sequencing using mSCR-seq. *Nat Commun* 2018;9:2937.
- 25 Elinger D, Gabashvili A, Levin Y. Suspension trapping (S-Trap) is compatible with typical protein extraction buffers and detergents for bottom-up proteomics. *J Proteome Res* 2019;18:1441–5.
- 26 Cox J, Mann M. MaxQuant enables high peptide identification rates, individualized p.p.b.-range mass accuracies and proteome-wide protein quantification. *Nat Biotechnol* 2008;26:1367–72.
- 27 Parekh S, Ziegenhain C, Vieth B, *et al.* zUMIs - A fast and flexible pipeline to process RNA sequencing data with UMIs. *Gigascience* 2018;7:giy059.
- 28 Newman AM, Steen CB, Liu CL, *et al.* Determining cell type abundance and expression from bulk tissues with digital cytometry. *Nat Biotechnol* 2019;37:773–82.
- 29 Smillie CS, Biton M, Ordovas-Montanes J, *et al.* Intra- and Inter-cellular rewiring of the human colon during ulcerative colitis. *Cell* 2019;178:714–30.
- 30 Subramanian A, Tamayo P, Mootha VK, *et al.* Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A* 2005;102:15545–50.
- 31 Kuleshov MV, Jones MR, Rouillard AD, *et al.* Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res* 2016;44:W90–7.
- 32 Habtezion A, Nguyen LP, Hadeiba H, *et al.* Leukocyte trafficking to the small intestine and colon. *Gastroenterology* 2016;150:340–54.
- 33 NEJM. Pathophysiology of inflammatory bowel diseases. Available: <https://www.nejm.org/doi/full/10.1056/NEJMra2002697>
- 34 Pascal V, Pozuelo M, Borruel N, *et al.* A microbial signature for Crohn's disease. *Gut* 2017;66:813–22.
- 35 Chapkin RS, Zhao C, Ivanov I, *et al.* Noninvasive stool-based detection of infant gastrointestinal development using gene expression profiles from exfoliated epithelial cells. *Am J Physiol Gastrointest Liver Physiol* 2010;298:G582–9.
- 36 Bennett WE, González-Rivera R, Shaikh N, *et al.* A method for isolating and analyzing human mRNA from newborn stool. *J Immunol Methods* 2009;349:56–60.
- 37 Ordiz MI, Wold K, Kaimila Y, *et al.* Detection and interpretation of fecal host mRNA in rural Malawian infants aged 6–12 months at risk for environmental enteric dysfunction. *Exp Biol Med* 2018;243:985–9.
- 38 Whitfield-Cargile CM, Cohen ND, He K, *et al.* The non-invasive exfoliated transcriptome (exfoliome) reflects the tissue-level transcriptome in a mouse model of NSAID enteropathy. *Sci Rep* 2017;7:14687.
- 39 Khaki-Khatibi F, Qujeq D, Kashifard M, *et al.* Calprotectin in inflammatory bowel disease. *Clinica Chimica Acta* 2020;510:556–65.
- 40 Schwanhäusser B, Busse D, Li N, *et al.* Global quantification of mammalian gene expression control. *Nature* 2011;473:337–42.
- 41 Houseley J, Tollervey D. The many pathways of RNA degradation. *Cell* 2009;136:763–76.
- 42 Carroll IM, Ringel-Kulka T, Ferrier L, *et al.* Fecal protease activity is associated with compositional alterations in the intestinal microbiota. *PLoS One* 2013;8:e78017.