



# OPEN Mucosal expression of *PI3*, *ANXA1*, and *VDR* discriminates Crohn's disease from ulcerative colitis

Jaslin Pallikkunnath James<sup>1✉</sup>, Boye Schnack Nielsen<sup>2</sup>, Ib Jarle Christensen<sup>1</sup>, Ebbe Langholz<sup>3,4</sup>, Mikkel Malham<sup>5,6</sup>, Tim Svenstrup Poulsen<sup>1</sup>, Kim Holmstrøm<sup>2</sup>, Lene Buhl Riis<sup>1,4</sup> & Estrid Høgdall<sup>1,4</sup>

Differential diagnosis of inflammatory bowel disease (IBD) to Crohn's disease (CD) or ulcerative colitis (UC) is crucial for treatment decision making. With the aim of generating a clinically applicable molecular-based tool to classify IBD patients, we assessed whole transcriptome analysis on endoscopy samples. A total of 408 patient samples were included covering both internal and external samples cohorts. Whole transcriptome analysis was performed on an internal cohort of FFPE IBD samples (CD,  $n=16$  and UC,  $n=17$ ). The 100 most significantly differentially expressed genes (DEG) were tested in two external cohorts. Ten of the DEG were further processed by functional enrichment analysis from which seven were found to show consistent significant performance in discriminating CD from UC: *PI3*, *ANXA1*, *VDR*, *MTCL1*, *SH3PXD2A-AS1*, *CLCF1*, and *CD180*. Differential expression of *PI3*, *ANXA1*, and *VDR* was reproduced by RT-qPCR, which was performed on an independent sample cohort of 97 patient samples (CD,  $n=44$  and UC,  $n=53$ ). Gene expression levels of the three-gene profile, resulted in an area under the curve of 0.84 ( $P=0.02$ ) in discriminating CD from UC, and therefore appear as an attractive molecular-based diagnostic tool for clinicians to distinguish CD from UC.

Inflammatory bowel disease (IBD) includes a group of chronic conditions that trigger inflammation in the lining of the gastrointestinal (GI) tract and affects more than 5 million people worldwide<sup>1</sup>. The diagnosis of IBD is based on clinical, endoscopic, and radiological characteristics together with histological evidence of structural and inflammatory changes<sup>2</sup>. The exact etiology of IBD remains unknown, however, it is believed to be caused by a mixture of genetic predisposition, abnormal immunity, and environmental exposures<sup>3</sup>. Crohn's disease (CD) and Ulcerative colitis (UC) are the most common forms of IBD. UC is characterized by relapsing and remitting mucosal inflammation stretching from the rectum to proximal segments of the colon, whereas CD may affect the whole GI tract and cause transmural inflammation<sup>4,5</sup>. In CD, one-third of the patients have a pure colonic disease location and at least 60% of them have non-stricturing or non-penetrating behavior at the time of diagnosis. In the absence of granulomas, these cases can be difficult to differentiate from UC<sup>4</sup>, and are termed IBD unclassified (IBDU). Even though 25% of IBDU patients are eventually sub-classified into either CD or UC within 6 months of follow up, the majority remain as IBDU<sup>4</sup> even after subsequent colectomy<sup>6</sup>. When confirming the subtype is crucial before surgery or when choosing an advanced medical treatment, it is necessary to separate IBDU cases into CD and UC at an early stage of the disease or during the disease. This underlines the need for biomarkers to support correct diagnosis of uncertain IBD cases for better disease management of the patients.

Since surgery, modern treatments, and emerging prognostic indices work on disease-specific strategies, it is important to determine the correct diagnosis from start for optimal clinical management in IBD patients. At present, there are several molecular diagnostic studies available for IBD, each offering valuable insights into disease mechanisms and patient management, however there are so far no universal molecular diagnostic biomarkers. Serologic tests and fecal markers have been studied intensively to improve the diagnosis of IBD. Fecal markers such as lactoferrin and calprotectin are indicative for any inflammation in the GI tract, which can discriminate between IBD and functional disorders without inflammation but have proved unsuccessful in

<sup>1</sup>Department of Pathology, Herlev University Hospital, Borgmester Ib Juuls Vej 73, 2730 Herlev, Denmark. <sup>2</sup>Bioneer A/S, Hørsholm, Kogle Allé 2, 2970 Hørsholm, Denmark. <sup>3</sup>Gastroenheden D, Herlev University Hospital, 2730 Herlev, Denmark. <sup>4</sup>Institute for Clinical Medicine, University of Copenhagen, 2200 Copenhagen, Denmark. <sup>5</sup>The Department of Pediatric and Adolescence Medicine, Copenhagen University Hospital-Amager and Hvidovre, 2650 Hvidovre, Denmark. <sup>6</sup>Copenhagen Center for Inflammatory Bowel Disease in Children, Adolescents and Adults, Hvidovre Hospital, University of Copenhagen, 2650 Hvidovre, Denmark. ✉email: jaslin.pallikkunnath.james@regionh.dk; jaslin.pj@gmail.com

distinguishing CD from UC. Serologic markers such as autoantibodies, and neutrophils<sup>7–9</sup> are not alone sufficient to have a distinctive diagnosis in IBD. There have been growing efforts to characterize and explore the pathogenic mechanisms within the inflammatory cascade of IBD subtypes to identify more appropriate treatment regimens on a case-by-case basis. By using differential gene expression of 7 marker genes on frozen biopsies, Stein et al. developed a biomarker tool for differential diagnosis in IBD, thereby demonstrating that CD and UC have distinct expression signatures<sup>10</sup>. Former studies have showed that IBD subtyping using fresh frozen colon biopsies can be performed, but the results have not been substantiated in follow-up studies<sup>11–13</sup>. Formalin-fixed, paraffin-embedded (FFPE) colon tissue biopsies may have significant promise for analyzing IBD, IBDU, and several other diseases because they are representative tissues collected and preserved by pathologists worldwide. FFPE tissues are routinely produced in the clinical routine setting and can be stored at room temperature for decades. One of the concerns when using FFPE tissue samples is the risk that the mRNAs may be of poor quality due to cross-linking or oxidative deamination which hampers subsequent gene expression studies. However, Turnbull et al. showed that comparable gene expression profiles can be obtained using FFPE material from breast tumor samples and FFPE is a reliable substitute to fresh frozen biopsies<sup>14</sup>. Importantly, also using FFPE colon biopsy samples, Christophi et al. quantified gene expression by RT-PCR, which gave insights into the pathogenetic differences between CD and UC<sup>15</sup>.

In this study we aimed at identifying biomarkers for discriminating CD from UC patients using FFPE-derived RNA from archived patient samples.

## Results

### Discovery study using next generation sequencing

For the discovery study, total RNA was extracted from mucosal FFPE tissue sections from 33 IBD and 10 HC cases with a mean RNA concentration of 45.19 ng/μL (range: 7 to 156 ng/μL). Targeted RNA sequencing with the Ion AmpliSeq™ transcriptome human gene expression panel resulted in an average of 13 million (range: 2 to 32 million) mapped reads per sample. The average fraction of valid reads among all the mapped reads per sample was 83% (range, 48 to 95%) and that of the targets detected in the valid reads were with an average of 65% (range, 36 to 75%) per sample. Unsupervised clustering of the samples based on gene expression using the PCA plot, clustered the samples into different disease types as shown in Fig. 1A. Out of more than 20,800 genes in the transcriptome panel, 18,848 genes were selected for further analysis and is uploaded into GEO database (GSE222070). Genes with less than 10 reads in all samples were excluded, which resulted in a list of 14,190 genes. From the logistic regression analysis, we found that a set of 429 genes were significantly differentially expressed between CD and UC with 85 genes being up-regulated and 344 genes being down-regulated in CD relative to UC. Hierarchical clustering of differentially regulated genes separated CD and UC as shown in the heatmap, Fig. 1B.

### Testing in external cohorts

The set of 100 most significantly expressed genes from our cohort (fold change, median = 2.48, IQR 1.95–3.2) was tested in two external cohorts from GEO database having gene expression data from mucosal biopsies of CD and UC patients: GSE16879<sup>16</sup> and GSE117993<sup>17</sup>. Both GSE16879<sup>16</sup> and GSE117993<sup>17</sup> are based on RNA from frozen samples, but obtained the expression data with different sequencing methods, Affymetrix micro-array and Illumina HiSeq 2500 sequencing platforms, respectively. OR and AUC values for all the 100 genes in the three different datasets are shown in supplementary Table 2. Among these, *PI3* showed the highest AUC in all three datasets [0.9044 (OR, 0.2137), 0.8271 (OR, 0.4223), and 0.7939 (OR, 0.5650) in GSE222070 (1: Herlev), GSE117993<sup>17</sup> (2: GSE11799), and GSE16879<sup>16</sup> (3: GSE16879), respectively]. This confirmed that our results are robust and reproducible also in cohorts of different ethnicities, type of sample material, and sequencing methods.

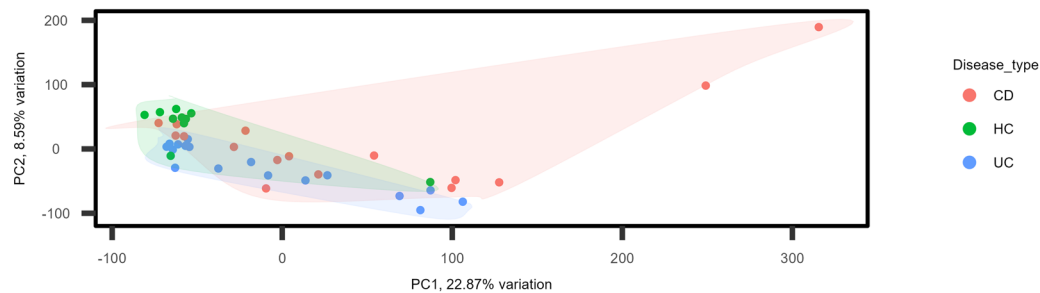
### Functional enrichment analysis

Based on AUC values in three datasets, expression of a set of ten genes, *PI3*, *ANXA1*, *VDR*, *MTCL1*, *SH3PXD2A-AS1*, *CLCF1*, *ME1*, *RPBMS*, *PRSS22* and *CD180*, was selected for enrichment analysis. These differentially expressed genes (DEG) were subjected to functional enrichment analysis and the results are shown in Fig. 2A<sup>18,19</sup>. The DEGs were typically involved in biological processes such as proliferation, immune response, tissue development and ion transport. With the analysis of cellular components, DEGs were mainly enriched in extracellular matrix and plasma membrane. Molecular functions such as protease binding, single stranded RNA binding and serine hydrolase activities were enriched. Receiver operating characteristic statistics discriminating CD from UC in the discovery study cohort for the 10 selected genes are shown in Fig. 3A. Box plots comparing expression levels of the 10 most relevant candidates among CD, UC, and HC are shown in Fig. 3B. A panel of 7 genes (*PI3*, *ANXA1*, *VDR*, *MTCL1*, *SH3PXD2A-AS1*, *CLCF1*, and *CD180*) out of the initially 10 genes were selected for further validation based on their relation to the extra intestinal manifestations (EIM) connected to IBD and involvement in immune response. Other factors considered in the selection process were high expression levels, allowing detection with basic methods such as RT-qPCR or ISH, and previous literature<sup>20–28</sup>. Expression of *PI3*, *ANXA1*, *VDR*, *MTCL1*, *SH3PXD2A-AS1*, *CLCF1*, and *CD180* have been shown to be involved in immune response, lymphocyte proliferation and keratinocyte differentiation. E.g. the protein encoded by *PI3*, Elafin, in serum is known as a marker for psoriasis and inflammation<sup>29</sup>, and *ANXA1* is highly expressed in intestinal tissue of UC patients compared to CD patients and promotes mucosal homeostasis<sup>24</sup>, see also Fig. 2B, and the discussion section below. Complete list of gene ontology analysis results is presented in supplementary file 4.

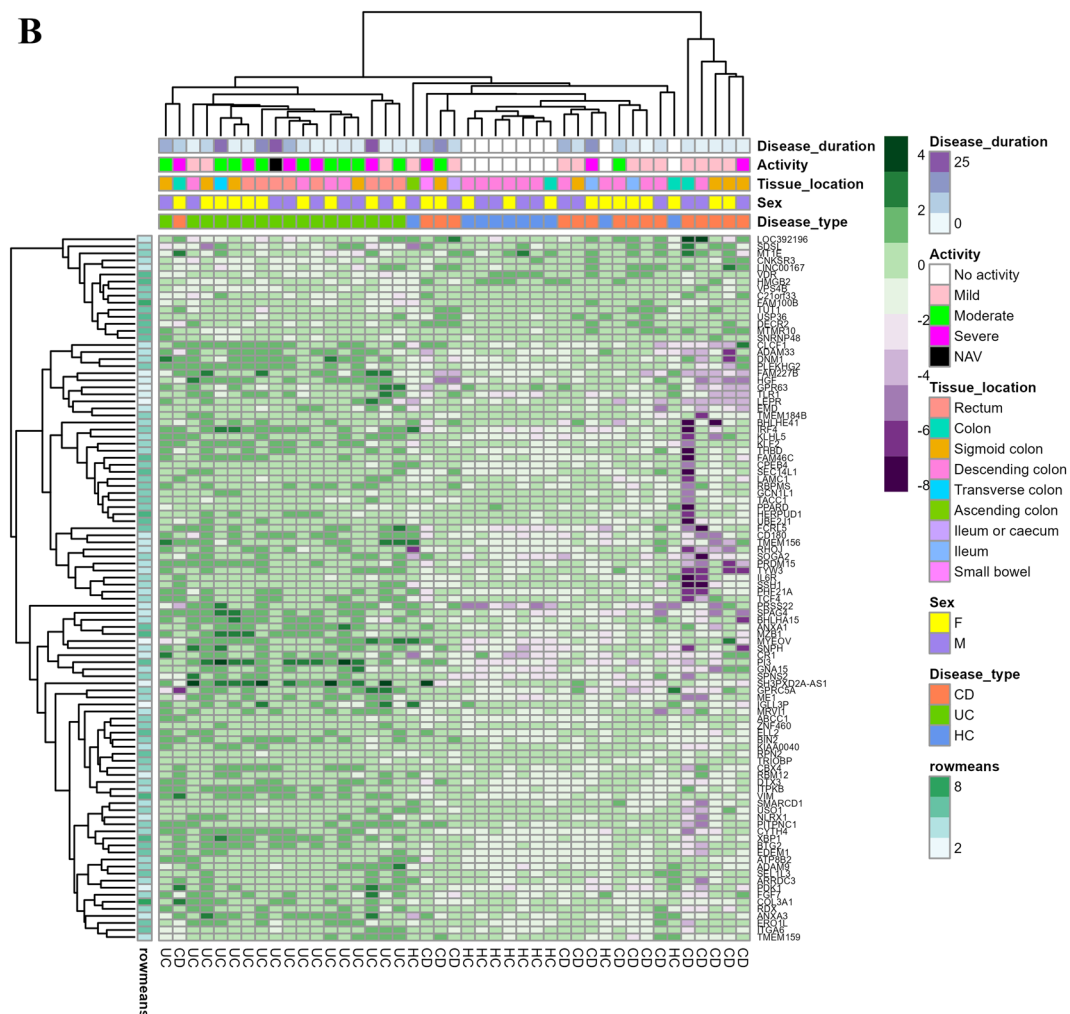
### Validation study using RT-qPCR

The final set of seven selected genes were tested in an independent cohort of 109 FFPE mucosal biopsies including 45 CDs, 53 UCs, and 12 HCs using RT-qPCR method. Expression of β-actin mRNA (*ACTB*) was used as the

**A**

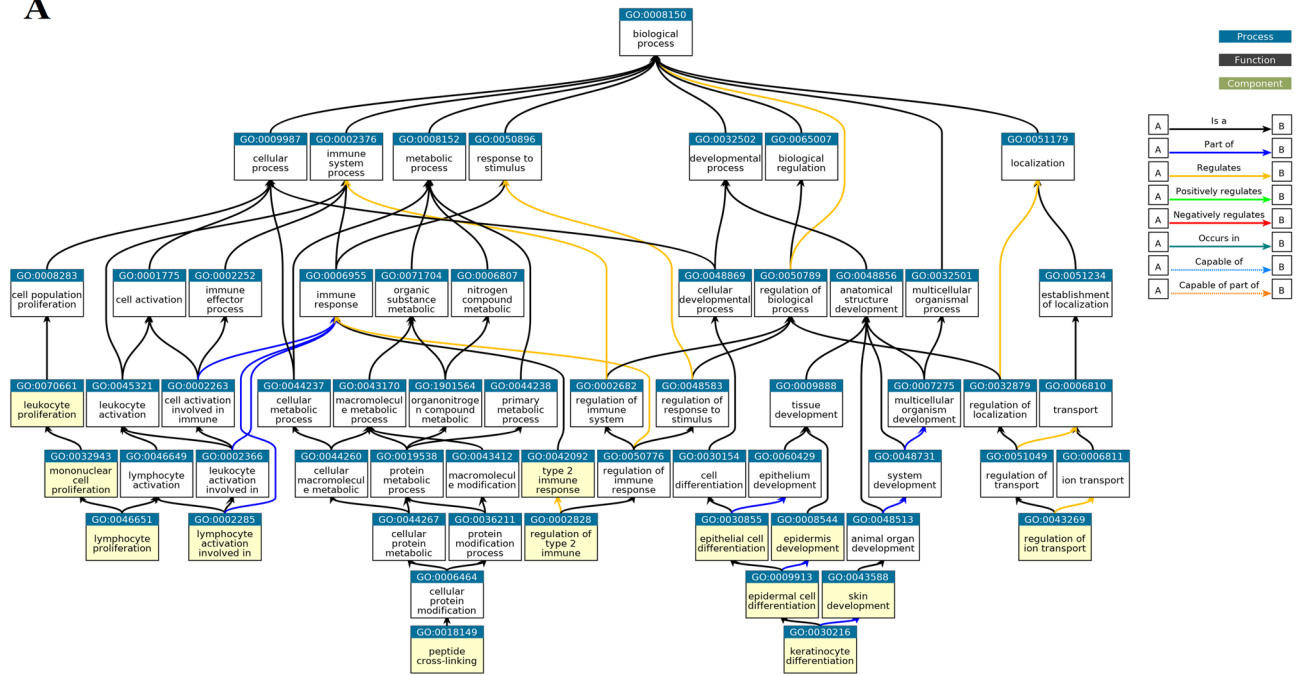


**B**



**Figure 1.** Heatmap and unsupervised hierarchical clustering of differentially expressed genes between CD (16) and UC (17) samples from the discovery study. **(A)** Principal component analysis (PCA) plot showing the unsupervised clustering of the discovery study samples into CDs, UCs, and healthy controls (HC) with the normalized gene expression values. Each point represents one sample with colors indicating which group as indicated in the figure. **(B)** Heatmap of the best 100 differentially expressed genes based on the *P* values from the discovery study. Factors such as disease duration, disease activity, tissue location, sex and disease type are shown with colored bars at the top of the heatmap. Disease duration is presented in years from 0 to 25 years. Purple color represents the highest expression and green color being the lowest expression for corresponding genes in respective samples. *PC* principal component, *CD* Crohn’s disease, *UC* ulcerative colitis, *HC* healthy control, *NAV* not available, *F* female, *M* male.

A



QuickGO - <https://www.ebi.ac.uk/QuickGO>

B

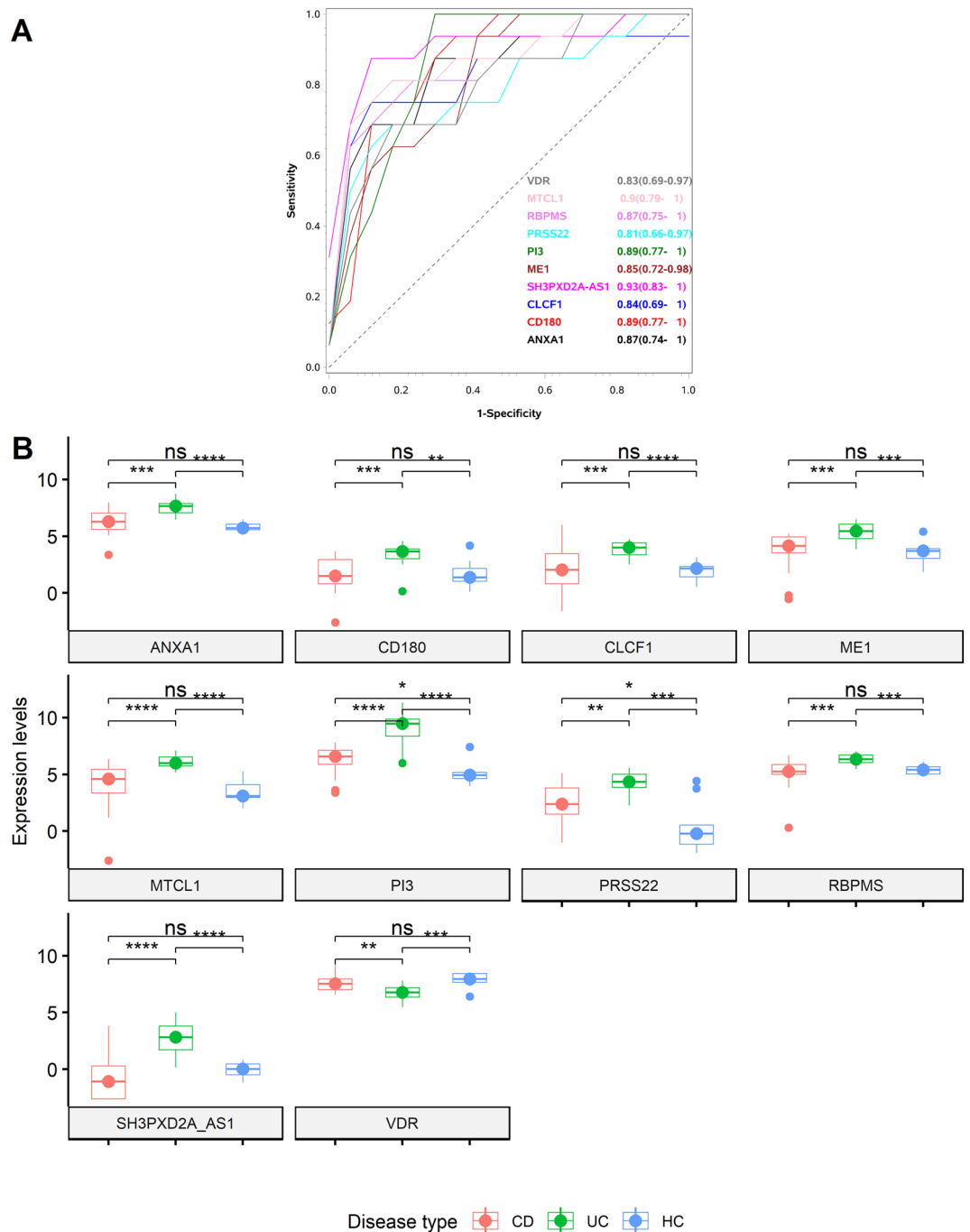
ID	Name	p-value	q-value Bonferroni	q-value FDR B&H	q-value FDR B&Y	Hit Count in Query List	Hit Count in Genome	Hit in Query List
GO:0002828	Regulation of type 2 immune response	9.24E-05	7.33E-02	2.91E-02	2.11E-01	2	33	CLCF1,ANXA1
GO:0002285	Lymphocyte activation involved in immune response	9.51E-05	7.54E-02	2.91E-02	2.11E-01	3	215	CD180,CLCF1,ANXA1
GO:0018149	Peptide cross-linking	1.10E-04	8.74E-02	2.91E-02	2.11E-01	2	36	PI3,ANXA1
GO:0042092	Type 2 immune response	1.50E-04	1.19E-01	2.98E-02	2.16E-01	2	42	CLCF1,ANXA1
GO:0030216	Keratinocyte differentiation	3.05E-04	2.42E-01	4.13E-02	2.99E-01	3	319	PI3,ANXA1,VDR
GO:0046651	Lymphocyte proliferation	3.52E-04	2.79E-01	4.13E-02	2.99E-01	3	335	CD180,CLCF1,ANXA1
GO:0032943	Mononuclear cell proliferation	3.64E-04	2.89E-01	4.13E-02	2.99E-01	3	339	CD180,CLCF1,ANXA1
GO:0070661	Leukocyte proliferation	4.63E-04	3.68E-01	4.59E-02	3.33E-01	3	368	CD180,CLCF1,ANXA1

**Figure 2.** Functional enrichment analysis of the differentially expressed genes. Differentially expressed genes between CDs and UCs identified in the discovery study after testing in two external cohorts and enriched biological processes associated with them: *PI3*, *ANXA1*, *VDR*, *MTCL1*, *SH3PXD2A-AS1*, *CLCF1*, *ME1*, *RPBMS*, *PRSS22*, and *CD180*. **(A)** Flow chart of the enriched biological processes with the 10 selected genes as highlighted in yellow. **(B)** Gene ontology (GO) terms with *P* values and the corresponding genes involved in the respective biological processes. Functional enrichment analysis was performed using ToppGene and QuickGO.

endogenous control for the assay and was found to be a key factor for the analysis, thus, the highest AUC values were obtained when samples with poor *ACTB* Ct values (>28) were removed. RT-qPCR expression values of *PI3*, *ANXA1*, and *VDR* along with the age at diagnosis and sex of the patients resulted in an AUC of 0.84 ( $P=0.02$ ) in discriminating CD from UC (Fig. 4A). Higher expression of *VDR* alone was detected in CD samples with an AUC of 0.77 ( $P=0.01$ ). The Ct values for *ACTB* varied from 28 to 31 among the samples, which indicated that efficiency of the discriminating algorithm was improved with increasing quality of RNA (Fig. 4B). Consequently, the number of samples included in the analysis was also reduced from 61 samples at *ACTB* threshold < 31 to 38 samples at *ACTB* threshold < 28, see also Fig. 4B. With an *ACTB* threshold < 31 the AUC for discriminating CD from UC was 0.67 ( $P=0.61$ ) whereas an *ACTB* threshold < 28, increased AUC to 0.88 ( $P=0.001$ ). Expression of *MTCL1*, *SH3PXD2A-AS1*, *CLCF1*, and *CD180* were found to be below the detection limit of the RT-qPCR method and excluded from the analysis.

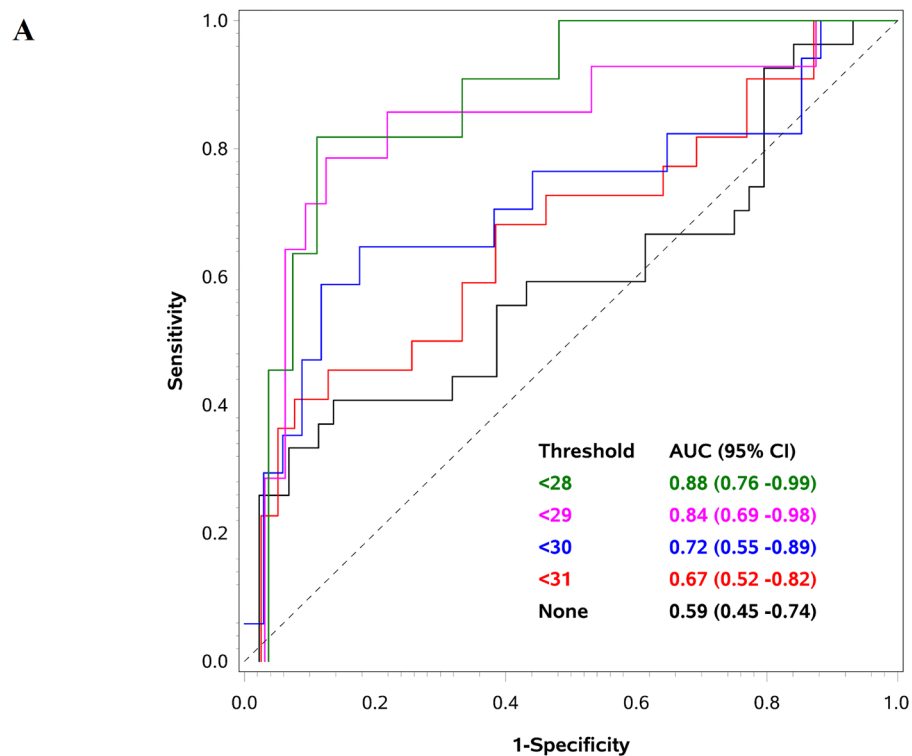
### Localization of *PI3*, *ANXA1* and *VDR*

A subset of 18 FFPE samples from the discovery study were selected for localization study based on the amount of remaining tissue and quality of the material. The samples were stained with RNAscope probes to *PI3*, *ANXA1*, and *VDR* mRNAs in parallel with *PIIB* mRNA and bacterial *dapB* mRNA probes as positive and negative controls, respectively (Fig. 5). UC samples showed relatively strong expression of *PI3* and *ANXA1* in the epithelial cells, whereas CD samples had low expression of *PI3* and *ANXA1*. *VDR* mRNA was intensely expressed in



**Figure 3.** Expression levels of the differentially expressed genes *PI3*, *ANXA1*, *VDR*, *MTCL1*, *SH3PXD2A-AS1*, *CLCF1*, *ME1*, *RBPMS*, *PRSS22*, and *CD180* in CD, UC and HC based on sequencing assay on the GSE222070 (discovery cohort). **(A)** ROC plot with AUC values (confidence intervals) for the corresponding genes in discriminating CD (16) from UC (17) samples among the discovery study samples. **(B)** Box plot of the 10 gene candidates and their expression levels in CD, UC, and HC samples from the discovery cohort. Symbols "\*\*\*\*", "\*\*\*", "\*\*", "\*" and "ns" corresponds to *P* values < 0.001, 0.001, 0.01, 0.05 and  $\geq 0.05$  respectively.

the epithelium of CD samples compared to UC samples. All three genes' relative staining intensities were in concordance with the sequencing and RT-qPCR results. *PI3* mRNA and *ANXA1* mRNA were detected mainly in the mucosal epithelium, and with the strongest staining being located towards the lumen. There was no *PI3* mRNA signal in the lymphoid follicles (LF) or lamina propria, whereas *ANXA1* mRNA was also detected in lamina propria and in the LF. *VDR* mRNA was highly expressed in the epithelium, but only sporadic staining was noted in the LF.



**B**

	AUC	Lower 95% CI	Upper 95% CI	P value	Number of samples
<i>No_threshold</i>	0.59	0.45	0.74	0.400	71
<i>ACTB_CT&lt;31</i>	0.67	0.52	0.82	0.129	61
<i>ACTB_CT&lt;30</i>	0.72	0.55	0.89	0.100	51
<i>ACTB_CT&lt;29</i>	0.84	0.69	0.98	0.010	46
<i>ACTB_CT&lt;28</i>	0.88	0.76	0.99	0.004	38

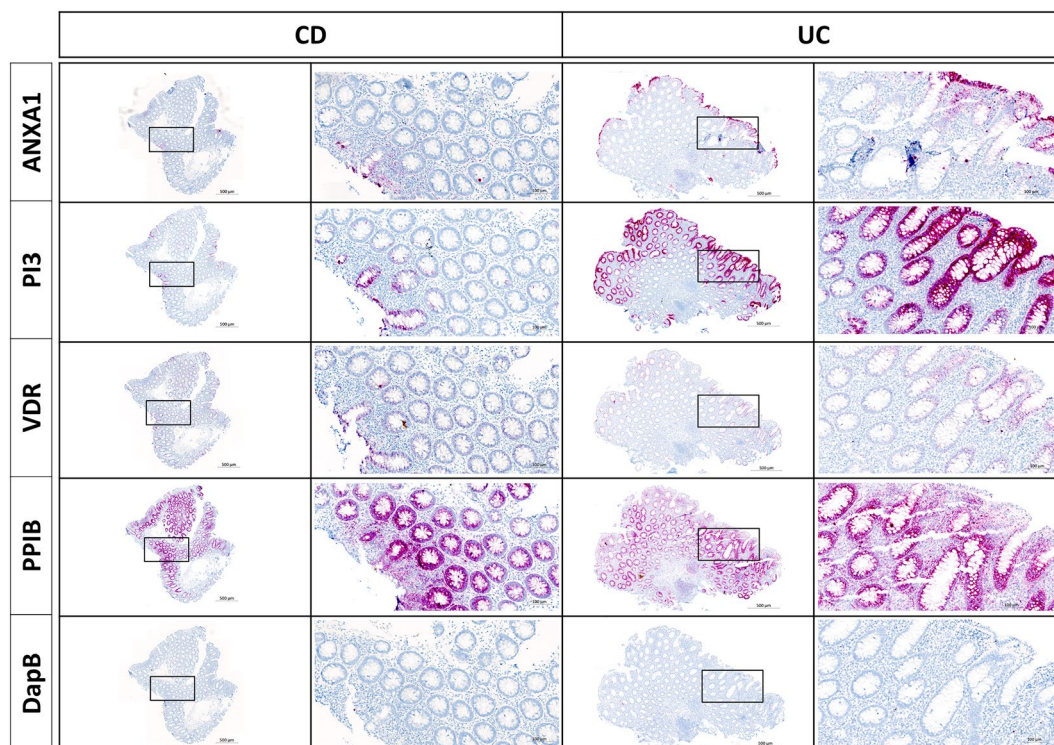
**Figure 4.** ROC plot showing the predictive power of the algorithm in the validation cohort. **(A)** ROC plot with AUC values for discriminating Crohn's disease (CD) and ulcerative colitis (UC) based on the algorithm using RT-qPCR expression values of PI3, ANXA1, and VDR mRNAs in combination with the age and sex of the patient at different endogenous control thresholds. ACTB mRNA was used as the endogenous control for the RT-qPCR assay. This data is generated based on the gene expression values from the validation cohort consisting of 44 CD and 53 UC patient samples. **(B)** Table with AUC values, *P* values, confidence intervals (CI) and number of samples included in the analysis at different ACTB mRNA threshold values.

## Discussion

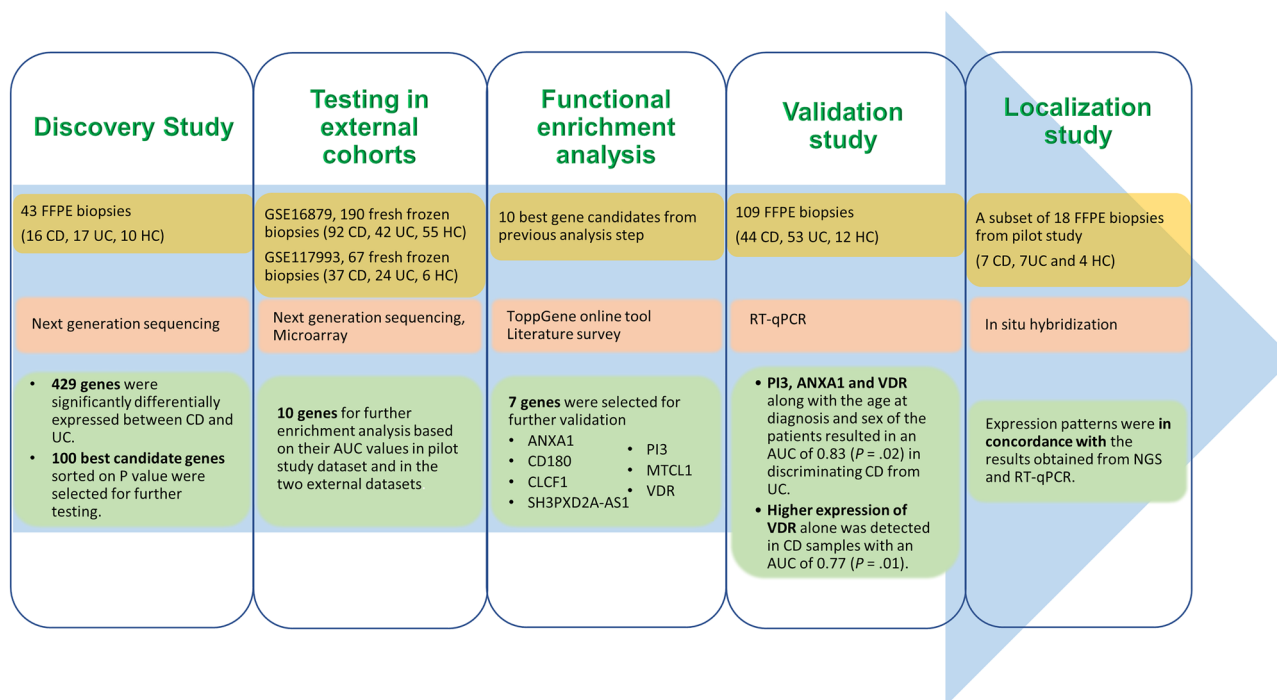
Using whole transcriptome sequencing on archived mucosal biopsies of IBD patients, we identified 429 genes differentially expressed between CD and UC. After testing in two external IBD cohorts, a list of the ten best candidate genes were selected for further validation. Functional enrichment analysis and validation using RT-qPCR in a third independent cohort, paved the way for an algorithm based on three genes (ANXA1, PI3 and VDR), which, in combination with patient's age at diagnosis and sex, can discriminate CD from UC with an average precision of 84%. The workflow of the whole study and the key results are summarized in Fig. 6.

Based on the molecular profile, the presented algorithm has the potential to support IBD subtyping at disease presentation and thereby to plan the optimal treatment course. Earlier studies have reported that IBD molecular subtyping is possible with differential gene expression from fresh frozen colon biopsies and FFPE tissue biopsies<sup>10,15,30,31</sup>. Other studies have also used differential expression of microRNAs, lamina propria T-helper subsets, and matrix-assisted laser desorption/ionization (MALDI) imaging techniques aiming at discriminating CD and UC<sup>32,33</sup>. However, the earlier findings have not been extensively supported in follow-up studies<sup>11-13</sup>.

In this study, we obtained a sample cohort of 43 archived FFPE colon biopsies for the discovery study and another 109 archived FFPE colon biopsies for the RT-qPCR validation. The discovery study sample cohort consisted of a mix of samples with different clinical characteristics such as disease subtype, sex, level of inflammation,



**Figure 5.** In situ hybridization (ISH) for ANXA1, PI3 and VDR mRNAs in CD and UC. RNAScope probes for the genes ANXA1, PI3 and VDR were applied to mucosal biopsies from a CD and a UC sample. The two samples are presented with overview images having framed areas shown in the zoomed images. UC samples in general showed relatively strong expression of *PI3* and *ANXA1* in the epithelial cells, compared to CD samples which in general had low expression of *PI3* and *ANXA1*. Conversely, *VDR* was intensely expressed in CD samples compared to UC samples. *PIIB* and *DapB* mRNA probes were used as positive and negative controls, respectively. The *PIIB* mRNA ISH signal is equally intense in the 2 samples. No ISH signal is seen with *DapB* probe.



**Figure 6.** Schematic illustration of the study design and strategy. Summary of samples used, methods applied, and key results obtained are shown as a flow chart.

disease duration, treatment status, and site of the disease. To test the gene expression differences based on the age at diagnosis, one-third of our validation cohort were pediatric patients. Our proposed 3-gene algorithm was validated in external data sets comprising 134 and 61 fresh frozen rectal or colon biopsies including both adults and children<sup>34,35</sup>, thus, emphasizing the advantage of covering diversity in the selection criteria terms. We used total RNA extracted from the archived FFPE blocks with relatively low RNA integrity numbers (RIN)  $\leq 2$ . Despite of this, our molecular profile was demonstrated to be robust since it was reproduced in external data-sets that were based on different sample types and sequencing platforms, as well as by being verified in our own independent sample cohort using RT-qPCR and partly by ISH. We believe that RT-qPCR is a feasible and cost-efficient approach to perform the molecular-based diagnosis in the daily routine. Although the sample size of this retrospective cohort is limited, there is sufficient power to detect a number of highly significant differences between disease groups.

IBD is believed to be caused by reduced barrier function, and the mucous epithelial barriers play a vital role in regulating tissue homeostasis<sup>36</sup>. In normal conditions, the inflammatory response and tissue repair are precisely balanced to guarantee restoration of mucosal homeostasis<sup>36</sup>. In IBD, such resolution of inflammation is compromised and thereby leads to sustained inflammation and impaired healing<sup>37</sup>. Even though UC and CD both are associated with reduced barrier function, their etiology and treatment regimens are different. UC is characterized by acute and chronic inflammation in the colonic mucosa, whereas in CD the inflammation is transmural, and may be characterized by occurrence of complications such as stenosis and fistulas<sup>6,38</sup>. In our functional enrichment analysis, we found that the expression of *ANXA1*, *PI3*, and *VDR* are involved in tissue homeostasis and immune related biological processes. Whereas *ANXA1* and *PI3* were high in UC, *VDR* expression was high in CD. All three genes are well-known parameters in IBD and in the maintenance of epithelial barrier integrity.

In line with our findings, previous studies have also reported elevated expression levels of *ANXA1* in UC patient samples and loss of *ANXA1* expression in patient samples with progressive CD compared to healthy controls<sup>25,39</sup>. By ISH analysis we found *ANXA1* mRNA in both epithelial cells and a subset of the cells located in lamina propria and submucosa. Similarly, *ANXA1* protein (annexin-A1) has been observed in monocytes, macrophages, neutrophils, and epithelial cells<sup>40</sup>. Annexin-A1 is a phospholipid-binding protein that acts in a variety of inflammatory pathways<sup>41</sup>. In their recent review, Leoni and Nusrat<sup>42</sup> discuss the repair-promoting functions of annexin-A1 within the inflamed mucosa<sup>42</sup>. Indeed, *ANXA1* deficient mice exhibited increased susceptibility to DSS-induced colitis<sup>43</sup>. Together, our findings support the hypothesis that annexin-A1 promotes the clearance of inflammation and restoration of compromised mucosal integrity and suggest that the high expression of *ANXA1* in the epithelium of UC samples serves as a defense mechanism.

Peptidase Inhibitor 3 (*PI3*) encodes the protein originally named SKALP or elafin, where elafin is formed by processing of the precursor protein trappin-2 by transglutaminases<sup>44–46</sup>. *PI3* is induced in keratinocytes by inflammatory mediators like tumor necrosis factor  $\alpha$ <sup>47</sup>. Elafin functions as a protease inhibitor and inhibits elastase 2a (ELA2A), recently discovered to be highly expressed in IBD and suggested to drive inflammation and loss of barrier function<sup>48</sup>. Elafin is well-known in anti-inflammatory processes as a defense against luminal microbes by its ability to disrupting bacterial cell membranes and by suppressing inflammation-associated transcription factors like NF- $\kappa$ B and AP-1<sup>49,50</sup>. In addition to being expressed in epithelium, elafin has also been found in neutrophils, the main function being anti-inflammatory, anti-microbial and pro-resolution of inflammation<sup>22</sup>. In the murine DSS model, mice with transgenic over-expression of elafin, elafin protected against the development of colitis<sup>51</sup>. In the IBD cases included in our study, as well as in those studied in Arijs et al<sup>16</sup>, Haberman<sup>17</sup>, Flach et al<sup>52</sup>, the *PI3* mRNA levels were higher in UC and CD patient mucosa than in HC. This finding was supported by our RT-qPCR and in situ hybridization analyses, showing abundant *PI3* mRNA in the mucosal epithelium of UC patients. In addition, immunohistochemical analysis found elafin strongly enhanced in inflamed versus non-inflamed UC<sup>53</sup> and elafin protein levels measured in serum were elevated in UC patients<sup>23,47</sup>, suggesting that the increased *PI3* mRNA levels is reflected in increased elafin protein levels. In the UC tissue, the increased level of *PI3*/elafin expression however appears insufficient to impede the hostile inflammation in the mucosa, and it is therefore encouraging that experimental therapeutic approaches using elafin have proven promising<sup>54</sup>.

In contrast to *ANXA1* and *PI3*, we found elevated *VDR* expression in CD compared to UC. Vitamin D/*VDR* signaling directs homeostasis, stimulates functionality of tight junction proteins and is important for bacteria recognition, pathogen clearance and dysbiosis prevention<sup>55</sup>. Cellular response to vitamin D was among the top hits of negatively enriched GO terms in non-inflamed CD patients compared to healthy controls in a study with treatment-naive IBD patients and their mucosal gene expression, thus, strengthening the body of evidence regarding vitamin D as a significant actor in the pathogenesis of CD<sup>56,57</sup>. The active form of vitamin D, calcitriol, achieves its effects through *VDR*, which functions as a transcription factor in the cell nucleus and controls gene expression. *VDR* regulation of e.g. Claudin-15 is important in maintaining epithelial integrity and for protecting against colitis<sup>38,58</sup>. *VDR* appears to have several functions as transcription factor and the low *VDR* expression in IBD has been discussed to correlate with dysfunctional vitamin D/*VDR* signaling in IBD patients<sup>59</sup>.

Autoimmune inflammatory conditions are common for IBD and psoriasis, and it may therefore not seem surprising that “keratinocyte differentiation” showed up in the functional enrichment analysis. Interaction between keratinocytes, immune cells, and other skin-resident cells is involved in the pathogenesis of psoriasis<sup>60</sup>. E.g. studies have shown that elafin is upregulated in psoriasis<sup>29,61,62</sup>, similar to our finding of *PI3* upregulation in UC. From the functional enrichment analysis, we found that also *ANXA1* and *VDR* are involved in keratinocyte differentiation and tissue development. Previous studies have shown that CD and psoriasis are related within individuals, leading to relatively more CD patients also suffer from psoriasis, and also that the level of association of psoriasis to CD is higher than to UC<sup>63</sup>.

It is noted that the quality of RNA from FFPE samples can be a challenge. Increasing the quality of RNA could give higher discriminating power for the RT-qPCR method, see also Fig. 4B. If implemented in routine practice, the quality of RNA is likely to be higher, as the FFPE samples will not be stored for long time. To



avoid gender-related bias or prejudice based on other clinical characteristics including the level of inflammation, treatment history, or illness location, samples were chosen using a strict and unbiased random sampling procedure. Despite our best efforts to create a completely balanced gender distribution, for instance, a slight variation of 3–9% more female patient samples—which we estimate is within the range of normal statistical fluctuation—occurred.

Since the samples we utilized were drawn at random, they came from both the time of the diagnosis and later in the course of the treatment. Another limitation may be that our samples were selected with a clear CD or UC phenotype and not cases with an uncertain phenotype. This precludes the possibility of testing the algorithm in IBDU cases.

In conclusion, we have developed an algorithm based on RT-qPCR expression values of 3 genes in combination with the patient's age and sex, to support clinicians in distinguishing CD from UC. This method could be applied in a molecular pathology routine clinical setup and potentially to categorize IBDU cases into CD and UC patients at early presentation of disease or during disease course when reconfirmation of the subtype is of utmost importance prior to surgery or choice of advanced medical treatment.

## Methods

### Samples

FFPE mucosal tissue biopsies were used for the discovery study [33 IBD patients and 10 healthy controls (HC)] and the validation study (97 IBD patients and 12 HCs). The samples were stored in clinical biobanks at hospitals in the capital region of Denmark as part of the initial diagnostic procedure. Patients were selected according to IBD subtype and had diverse disease characteristics regarding age at diagnosis, degree of inflammation, treatment history, and disease location. Patient characteristics are presented in Table 1. According to the amount of tissue remaining in the FFPE block, 18 samples (7 CDs, 7UCs and 4 HCs) were selected from the discovery study cohort with varying sample characteristics for localization study using in situ hybridization (ISH).

### RNA extraction

Total RNA was isolated from two 10 µm thick sections of FFPE tissue samples using MagMAX™ FFPE DNA/RNA Ultra Kit from Applied Biosystems™ according to the manufacturer's instructions. RNA quantity and purity were assessed using the NanoDrop spectrophotometer (ThermoFisher Scientific) and Qubit RNA HS Assay kit (Life

	Discovery Study cohort (N = 43)			Validation study cohort (N = 109)		
	CD	UC	HC	CD	UC	HC
Number of samples	16	17	10	44	53	12
Age at diagnosis						
Median (IQR)	27 (24–46)	34 (29–43)	NAP	25 (14–45)	24 (15–37)	NAP
Sex						
Female, N (%)	9 (56)	9 (53)	5 (50)	26 (59)	28 (53)	5 (42)
Age of the FFPE sample blocks in years						
Median (range)	4 (0–11)	1 (0–9)	0 (0–9)	6 (1–20)	5 (1–15)	1 (1–9)
Rate of inflammation						
Mild/moderate/severe (N)	10/2/4	3/9/4*	1**/0/0	20/10/14	21/25/7	NAP
Disease duration in years						
Median (range)	4 (0–19)	1 (0–27)	NAP	2 (0–30)	1 (0–20)	NAP
Medical treatment (non-biological) at time of biopsy						
Yes/no/NA(N)	10/6/0	13/3/1	1/0/0	42/0/2	46/0/7	NAP
Medical treatment (biological) at time of biopsy						
Yes/no/NA (N)	2/14/0	3/13/1	NAP	17/25/2	14/32/7	NAP
Number of samples and sample location						
Small bowel	1					
Colon (unspecified)	2		2	10	9	1
Ileum/cecum	3			14		
Ascending colon			1	4	4	1
Transverse colon		1		2		
Descending colon	5	4	7	2	10	5
Sigmoid colon	5	4		7	19	5
Rectum		8		5	11	

**Table 1.** Patient characteristics of the discovery study and validation study cohort. *N* number of samples, *CD* Crohn's disease, *UC* ulcerative colitis, *HC* healthy controls, *NAP* not applicable, *IQR* inter quartile range, *FFPE* formalin fixed paraffin embedded, *NA* not available. \*One sample information is not available. \*\*Slight inflammation due to bowel preparation prior to endoscopy.

Technologies). RNA quality and integrity were evaluated using Agilent RNA 6000 Nano Kit on Agilent 2100 Bioanalyzer System. RNA samples were kept at  $-80^{\circ}\text{C}$  until further use. All RNA samples used for analysis had an RNA integrity number value of less than 2.

### Next generation sequencing

Total RNAs were diluted to 10 ng using nuclease free water and reverse transcribed using SuperScript™ VILO™ cDNA Synthesis Kit. Transcriptome libraries were generated using automated library preparation with the Ion AmpliSeq™ Transcriptome Human Gene Expression Panel containing 20,800 primer pairs and the Chef-Ready Kit (A31446). Transcriptome libraries were barcoded, templated, and sequenced on Ion GeneStudio™ S5 System using the Ion 550 chip. Initial analysis for AmpliSeq sequencing data was performed using the ampliSeqRNA plugin (V5.12) available with Ion Torrent™ sequencing platforms using Torrent Mapping Alignment Program (TMAP). Raw sequencing reads were aligned against the reference library hg19 Ampliseq transcriptome V1.1. Raw read counts obtained from Ampliseq were used for further statistical analysis.

### Statistical analysis of NGS data

Raw read counts obtained from ampliSeqRNA plugin were filtered, normalized using the trimmed mean method<sup>64</sup> and log transformed in R using the packages limma and edgeR<sup>65</sup>. Statistical calculations were based on logistic regression comparing the expression results of samples obtained from patients with CD and UC diagnoses. The probability of CD was modeled. Multivariate analysis on the factors such as selected genes, age, sex, and disease duration were performed. Model assessment has been done using goodness of fit (Hosmer–Lemeshow) statistics. The best 100 genes were selected based on the p-values conditional on the numerical value of the regression coefficient be greater than one i.e., the odds ratio was greater than 2.71. Results are presented by odds ratios (OR) with 95% confidence limits, the AUC [area under the receiver operating characteristic curve (ROC)] as a measure of discrimination, and the p-value. Expression levels for HC samples were compared to CD and UC using T-test. P-values less than 5% were considered significant. Analysis was performed using the R<sup>66</sup> packages limma<sup>67</sup>, edgeR<sup>68</sup> and SAS (v9.4, SAS Institute, Cary, NC., USA).

### Testing in external cohorts

Gene expression omnibus (GEO) were searched for datasets containing references to IBD and related studies. Datasets GSE16879<sup>16</sup> and GSE117993<sup>17</sup> were chosen based on the sample size of the cohort, type of samples and type of techniques used. Fresh frozen samples were used in both GEO datasets. Both external datasets have used other platforms (Illumina HiSeq 2500 and Affymetrix Human Genome U133 Plus 2.0 Array) for the gene expression and different method for normalization analysis (moderated T-statistics). The 100 best performing genes from our cohort were assessed in the external datasets by logistic regression. Comparisons between CD and UC were only performed, not with HC. Due to differences in platforms used for RNA sequencing, some of the genes in Ion AmpliSeq™ Transcriptome Human Gene Expression Panel was having different nomenclatures in the external datasets. Two of the genes were missing in both external datasets and one gene was present only in one of the datasets.

### Functional enrichment analysis

Genes having an AUC of greater than 0.70 in all three datasets were selected for further functional enrichment analysis. Functional annotation enrichment analyses of the ten most significant genes were performed using ToppFun function in ToppGene online tool (<https://toppgene.cchmc.org/>). Related genes were grouped according to features such as gene ontology (GO): molecular function, biological processes, and cellular component, and disease. FDR B&H cutoff was set to 0.05. Depending on the molecular functions, previous literature in relation to IBD and level of expression in colonic mucosa, 7 genes were selected for further validation using RT-qPCR.

### RT-qPCR

Expression of the 7 genes panel were tested in a validation cohort of 109 FFPE samples having 44 CDs, 53 UCs and 12 HCs using RT-qPCR. Total RNA from the FFPE mucosal tissue sections were extracted using MagMAX™ FFPE DNA/RNA Ultra Kit from Applied Biosystems™ and were quantified using Quant-iT™ RiboGreen RNA reagent and kit from Invitrogen. RT-qPCR was performed using QuantiNova SYBR® Green RT-qPCR Kit (Qiagen). QuantiNova LNA PCR Custom Panel (GeneGlobe ID: SBGA09807) was designed with the 7 selected genes. The primers used in the custom panel are listed in supplementary file 1. HS\_PPC\_2467741, HS\_HGDC\_2467744 and HS\_QIC\_2467742 were used as reference assays according to custom panel manufacturer's recommendation. Real-time PCR amplifications were performed using the Light Cycler 480 system (Roche). RT-qPCR data was analyzed by the  $2^{-\Delta\Delta\text{Ct}}$  method<sup>69</sup> using ACTB (Beta actin) as reference gene to yield relative fold expression levels of the seven different genes. Statistical analysis was done using univariate and multivariate logistic regression, modeling the probability of CD. These analyses were stratified according to the CT values of ACTB from 28 to 31. Relative fold expression values for MTCL1, SH3PXD2A-AS1, CLCF1, and CD180 genes were below detection limit and hence they were excluded from further analysis.

### In situ hybridization (ISH)

ISH for mRNAs PI3, ANXA1 and VDR were performed on 5  $\mu\text{m}$  thick FFPE sections using the RNAscope® 2.5 HD (Advanced Cell Diagnostics, CA)<sup>70</sup> as described by Anderson et al.<sup>71</sup> on automated Ventana Discovery Ultra instrument. Probes used for PI3, ANXA1 and VDR mRNA were RNAscope® 2.5 VS Probe- Hs-PI3 (Target region:2–564, 12 zz pairs), RNAscope® 2.5 VS Probe- Hs-ANXA1 (Target region:142–1095, 20 zz pairs),

RNAscope<sup>®</sup> 2.5 VS Probe-Hs-VDR (Target region:161–1191, 20 zz pairs), respectively. Probes to housekeeping gene PPIB mRNA (Target region: 139–989, 16 zz pairs) and negative control bacteria dapB mRNA (Target region: 414–862, 10 zz pairs) were used as positive and negative control, respectively. An alkaline phosphatase-based signal amplification system was applied and detected using the chromogenic substrate fast red. Positive mRNA expression was visible as red signal dots. Digital whole slides were obtained using a 20X objective in a bright-field slide scanner (Axio Scan.Z1; Zeiss). The probes for housekeeping gene PPIB and the bacterial dapB were applied in a subset of cases.

### Ethics approval and consent to participate

The study has been approved by the National Ethical Committee (j. nr. H-20032221) and the Danish Data Protection Agency (P-2020-737). The need for consent was waived by Danish National Ethical committee as tissue used for the project is formalin fixed paraffin embedded tissue previously collected and stored in the clinical biobank (tissue archive) at the department of Pathology at Herlev Hospital. The study was conducted in accordance with the ethical standards, according to the declaration of Helsinki, and according to national and international guidelines.

### Data availability

Transcript profiling: The National Center for Biotechnology Information Gene Expression Omnibus database accession number for data from the discovery study reported in this paper is GSE222070. Data transparency statement: The raw data generated during and/or analysed during the current study are not publicly available, which is in accordance with the rules concerning processing of personal data set out in the EU General Data Protection Regulation (GDPR) and the Danish Data Protection Act. Nonetheless, may a researcher have an interest in our data they are welcome to contact us and collaborate. The data that support the findings of this study can be requested from The National Secretariat for Bio- and Genome Bank Denmark, RBGB.sekretariat.herlev-og-gentofte-hospital@regionh.dk, Herlev Hospital, Borgmester Ib Juuls Vej 73, 2730 Herlev, Denmark. Normalized data of the gene expression from the discovery study is uploaded to gene expression omnibus with the reference number: GSE222070.

Received: 27 May 2023; Accepted: 20 October 2023

Published online: 27 October 2023

### References

- Bernstein, C. N. Review article: Changes in the epidemiology of inflammatory bowel disease-clues for aetiology. *Aliment. Pharmacol. Ther.* **46**, 911–919 (2017).
- Olsen, J. *et al.* Diagnosis of ulcerative colitis before onset of inflammation by multivariate modeling of genome-wide gene expression data. *Inflamm. Bowel Dis.* **15**, 1032–1038 (2009).
- Bouma, G. & Strober, W. The immunological and genetic basis of inflammatory bowel disease. *Nat. Rev. Immunol.* **3**, 521–533 (2003).
- Tontini, G. E. Differential diagnosis in inflammatory bowel disease colitis: State of the art and future perspectives. *World J. Gastroenterol.* **21**, 21 (2015).
- Chang, M. D. & Liu, X. Overview of histopathology of ulcerative colitis and Crohn's disease. In *Interventional Inflammatory Bowel Disease: Endoscopic Management and Treatment of Complications*. 49–68 (Elsevier, 2018). <https://doi.org/10.1016/B978-0-12-811388-2.00005-1>.
- Kellermann, L. & Riis, L. B. A close view on histopathological changes in inflammatory bowel disease, A narrative review. *Dig. Med. Res.* **4**, 3–3 (2021).
- Ferrante, M. *et al.* New serological markers in inflammatory bowel disease are associated with complicated disease behaviour. *Gut* **56**, 1394–1403 (2007).
- Bossuyt, X. Serologic markers in inflammatory bowel disease. *Clin. Chem.* **52**, 171–181 (2006).
- Bruining, D. H. & Loftus, E. V. Evolving diagnostic strategies for inflammatory bowel disease. *Curr. Gastroenterol. Rep.* **8**, 478–485 (2006).
- von Stein, P. *et al.* Multigene analysis can discriminate between ulcerative colitis, Crohn's disease, and irritable bowel syndrome. *Gastroenterology* **134**, 1869–1881 (2008).
- Von Stein, P. Inflammatory bowel disease classification through multigene analysis: Fact or fiction?. *Expert Rev. Mol. Diagn.* **9**, 7–10. <https://doi.org/10.1586/14737159.9.1.7> (2009).
- Janczewska, I. *et al.* Clinical application of the multigene analysis test in discriminating between ulcerative colitis and Crohn's disease: A retrospective study. *Scand. J. Gastroenterol.* **47**, 162–169 (2012).
- Bjerrum, J. T., Nyberg, C., Olsen, J. & Nielsen, O. H. Assessment of the validity of a multigene analysis in the diagnostics of inflammatory bowel disease. *J. Intern. Med.* **275**, 484–493 (2014).
- Turnbull, A. K. *et al.* Unlocking the transcriptomic potential of formalin-fixed paraffin embedded clinical tissues: Comparison of gene expression profiling approaches. *BMC Bioinform.* **21**, 30 (2020).
- Christophi, G. P., Rong, R., Holtzapfel, P. G., Massa, P. T. & Landas, S. K. Immune markers and differential signaling networks in ulcerative colitis and Crohn's disease. *Inflamm. Bowel Dis.* **18**, 2342–2356 (2012).
- Arijs, I. *et al.* *Mucosal Expression Profiling in Patients with Inflammatory Bowel Disease Before and After First Infliximab Treatment*. GSE16879 (2009).
- Karns R, Haberman Y, D. T. *Transcription Profiles of Rectal Biopsies Obtained During Diagnostic Colonoscopy for Pediatric Inflammatory Bowel Diseases*. GSE117993 (2018).
- Huntley, R. P. *et al.* The GOA database: Gene Ontology annotation updates for 2015. *Nucleic Acids Res.* **43**, D1057 (2015).
- Carbon, S. *et al.* The Gene Ontology resource: Enriching a GOLD mine. *Nucleic Acids Res.* **49**, D325–D334 (2021).
- Low, E. N., Den Mokhtar, N. M., Wong, Z. & Raja Ali, R. A. Colonic mucosal transcriptomic changes in patients with long-duration ulcerative colitis revealed colitis-associated cancer pathways. *J. Crohn's Colitis* **13**, 755–763 (2019).
- Greenhill, C. Elafin—A potential IBD therapy. *Nat. Rev. Gastroenterol. Hepatol.* **9**, 686–686 (2012).
- Zhang, W., Teng, G., Wu, T., Tian, Y. & Wang, H. Expression and clinical significance of Elafin in inflammatory bowel disease. *Inflamm. Bowel Dis.* **23**, 2134–2141 (2017).

23. Wang, J. *et al.* High circulating elafin levels are associated with Crohn's disease-associated intestinal strictures. *PLoS One* **15**, e0231796 (2020).
24. da Rocha, G. H. O. *et al.* Pioglitazone-mediated attenuation of experimental colitis relies on cleaving of annexin A1 released by macrophages. *Front. Pharmacol.* **11**, 591561 (2020).
25. Vong, L. *et al.* Up-regulation of annexin-A1 and lipoxin A4 in individuals with ulcerative colitis may promote mucosal homeostasis. *PLoS One* **7**, e39244 (2012).
26. Kellermann, L. *et al.* Mucosal vitamin D signaling in inflammatory bowel disease. *Autoimmun. Rev.* **19**, 102672 (2020).
27. Kim, J. *et al.* Implication of intestinal VDR deficiency in inflammatory bowel disease. *Biochim. Biophys. Acta-Gen. Subj.* **1830**, 2118–2128 (2013).
28. Hou, P. *et al.* Long noncoding RNA SH3PXD2A-AS1 promotes colorectal cancer progression by regulating p53-mediated gene transcription. *Int. J. Biol. Sci.* **17**, 1979–1994 (2021).
29. Elgharib, I., Khashaba, S. A., Elsaid, H. H. & Sharaf, M. M. Serum elafin as a potential inflammatory marker in psoriasis. *Int. J. Dermatol.* **58**, 205–209 (2019).
30. Wu, F. *et al.* Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: Insights into distinctive pathogenesis. *Inflamm. Bowel Dis.* **13**, 807–821 (2007).
31. Boyd, M. *et al.* Characterization of the enhancer and promoter landscape of inflammatory bowel disease from human colon biopsies. *Nat. Commun.* **9**, 1661 (2018).
32. Klein, O. *et al.* Classification of inflammatory bowel disease from formalin-fixed, paraffin-embedded tissue biopsies via imaging mass spectrometry. *Proteomics-Clin. Appl.* **14**, 1900131 (2020).
33. James, J. P. *et al.* MicroRNA biomarkers in IBD-differential diagnosis and prediction of colitis-associated cancer. *Int. J. Mol. Sci.* **21**, 1–19 (2020).
34. Haberman, Y. *et al.* Ulcerative colitis mucosal transcriptomes reveal mitochondriopathy and personalized mechanisms underlying disease severity and treatment response. *Nat. Commun.* **10**, 38 (2019).
35. Arijis, I. *et al.* Mucosal gene expression of antimicrobial peptides in inflammatory bowel disease before and after first infliximab treatment. *PLoS One* **4**, e7984 (2009).
36. Quirós, M. & Nusrat, A. Contribution of wound-associated cells and mediators in orchestrating gastrointestinal mucosal wound repair. *Annu. Rev. Physiol.* **81**, 189–209 (2019).
37. Ho, G., Cartwright, J. A., Thompson, E. J., Bain, C. C. & Rossi, A. G. Resolution of inflammation and gut repair in IBD: Translational steps towards complete mucosal healing. *Inflamm. Bowel Dis.* **26**, 1131–1143 (2020).
38. Turpin, W. *et al.* Increased intestinal permeability is associated with later development of Crohn's disease. *Gastroenterology* **159**, 2092–2100.e5 (2020).
39. Sena, A. *et al.* Dysregulation of anti-inflammatory annexin A1 expression in progressive Crohns disease. *PLoS One* **8**, e76969 (2013).
40. Kamal, A. M., Flower, R. J. & Perretti, M. An overview of the effects of annexin 1 on cells involved in the inflammatory process. *Mem. Inst. Oswaldo Cruz* **100**(Suppl 1), 39–48 (2005).
41. Lim, L. H. K. & Pervaiz, S. Annexin 1: The new face of an old molecule. *FASEB J.* **21**, 968–975 (2007).
42. Leoni, G. & Nusrat, A. Annexin A1: Shifting the balance towards resolution and repair. *Biol. Chem.* **397**, 971–979 (2016).
43. Babbitt, B. A. *et al.* Annexin A1 regulates intestinal mucosal injury, inflammation, and repair. *J. Immunol.* **181**, 5035–5044 (2008).
44. Wiedow, O., Schroder, J. M., Gregory, H., Young, J. A. & Christophers, E. Elafin: an elastase-specific inhibitor of human skin. Purification, characterization, and complete amino acid sequence. *J. Biol. Chem.* **265**, 14791–14795 (1990).
45. Alkemade, J. A. C. *et al.* SKALP/elafin is an inducible proteinase inhibitor in human epidermal keratinocytes. *J. Cell Sci.* **107**(Pt 8), 2335–2342 (1994).
46. Kuijpers, A. L. A. *et al.* Skin-derived antileukoproteinase (SKALP) and epidermal fatty acid-binding protein (E-FABP): Two novel markers of the psoriatic phenotype that respond differentially to topical steroid. *Acta Derm. Venereol.* **77**, 14–19 (1997).
47. Krawiec, P. & Pac-Kożuchowska, E. Clinical significance of serum Elafin in children with inflammatory bowel disease. *Biomedicines* **10**, 3267 (2022).
48. Motta, J. P. *et al.* Epithelial production of elastase is increased in inflammatory bowel disease and causes mucosal inflammation. *Mucosal Immunol.* **14**, 667–678 (2021).
49. Deraison, C., Bonnart, C., Langella, P., Roget, K. & Vergnolle, N. Elafin and its precursor Trappin-2: What therapeutic potential for intestinal diseases?. *Br. J. Pharmacol.* <https://doi.org/10.1111/BPH.15985> (2022).
50. Butler, M. W. *et al.* Elafin prevents lipopolysaccharide-induced AP-1 and NF-kappaB activation via an effect on the ubiquitin-proteasome pathway. *J. Biol. Chem.* **281**, 34730–34735 (2006).
51. Motta, J. *et al.* Modifying the protease, antiprotease pattern by Elafin overexpression protects mice from colitis. *Gastroenterology* **140**, 1272–1282 (2011).
52. Flach, C.-F. *et al.* Detection of elafin as a candidate biomarker for ulcerative colitis by whole-genome microarray screening. *Inflamm. Bowel Dis.* **12**, 837–842 (2006).
53. Schmid, M. *et al.* Attenuated induction of epithelial and leukocyte serine antiproteases elafin and secretory leukocyte protease inhibitor in Crohn's disease. *J. Leukoc. Biol.* **81**, 907–915 (2007).
54. Bermúdez-Humarán, L. G. *et al.* Serine protease inhibitors protect better than IL-10 and TGF-β anti-inflammatory cytokines against mouse colitis when delivered by recombinant lactococci. *Microb. Cell Fact.* **14**, 26 (2015).
55. Fakhoury, H. M. A. *et al.* Vitamin D and intestinal homeostasis: Barrier, microbiota, and immune modulation. *J. Steroid Biochem. Mol. Biol.* **200**, 105663 (2020).
56. Vatn, S. S. *et al.* Mucosal gene transcript signatures in treatment naïve inflammatory bowel disease: A comparative analysis of disease to symptomatic and healthy controls in the European IBD-character cohort. *Clin. Exp. Gastroenterol.* **15**, 5 (2022).
57. Mouli, V. P. & Ananthkrishnan, A. N. Review article: Vitamin D and inflammatory bowel diseases. *Aliment. Pharmacol. Ther.* **39**, 125–136 (2014).
58. Chatterjee, I. *et al.* Overexpression of vitamin D receptor in intestinal epithelia protects against colitis via upregulating tight junction protein claudin 15. *J. Crohns Colitis* **15**, 1720 (2021).
59. Bakke, D. & Sun, J. Ancient nuclear receptor VDR with new functions: Microbiome and inflammation. *Inflamm. Bowel Dis.* **24**, 1149–1154 (2018).
60. Zhou, X., Chen, Y., Cui, L., Shi, Y. & Guo, C. Advances in the pathogenesis of psoriasis: from keratinocyte perspective. *Cell Death Dis.* **13**, 81 (2022).
61. Nonomura, K., Yamanishi, K., Yasuno, H., Nara, K. & Hirose, S. Up-regulation of elafin/SKALP gene expression in psoriatic epidermis. *J. Invest. Dermatol.* **103**, 88–91 (1994).
62. Iizuka, H., Takahashi, H., Honma, M. & Ishida-Yamamoto, A. Unique keratinization process in psoriasis: Late differentiation markers are abolished because of the premature cell death. *J. Dermatol.* **31**, 271–276 (2004).
63. Cohen, A. D., Dreiherr, J. & Birkenfeld, S. Psoriasis associated with ulcerative colitis and Crohn's disease. *J. Eur. Acad. Dermatol. Venereol.* **23**, 561–565 (2009).
64. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, 1–9 (2010).
65. Smyth, G. K. *et al.* RNA-seq analysis is easy as 1–2–3 with limma, Glimma and edgeR. *F1000Research* **5**, 1–30 (2018).

66. R Development Core Team, R. R: *A Language and Environment for Statistical Computing*. (2021).
67. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
68. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
69. Schmittgen, T. D. & Livak, K. J. Analyzing real-time PCR data by the comparative CT method. *Nat. Protoc.* **3**, 1101–1108 (2008).
70. Wang, F. *et al.* RNAscope. *J. Mol. Diagn.* **14**, 22–29 (2012).
71. Anderson, C. M. *et al.* Fully automated RNAscope in situ hybridization assays for formalin-fixed paraffin-embedded cells and tissues. *J. Cell. Biochem.* **117**, 2201–2208 (2016).

## Acknowledgements

We would like to thank laboratory technicians Christina Grønhøj, Department of pathology, Herlev hospital and Trine Møller, Bioneer A/S for their valuable technical assistance.

## Author contributions

J.P.J.: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing—Original Draft, Writing—Review & Editing, Visualization, Project administration, Funding acquisition, Resources. BSN: Conceptualization, Methodology, Validation, Formal analysis, Resources, Writing—Review & Editing, Supervision. I.J.C.: Methodology, Software, Validation, Formal analysis, Investigation, Data Curation, Writing—Review & Editing, Visualization. E.L.: Conceptualization, Resources, Data Curation, Writing—Review & Editing, Supervision. M.M.: Conceptualization, Resources, Data Curation, Writing—Review & Editing, Supervision. T.S.P.: Methodology, Software, Validation, Resources, Data Curation, Visualization, Writing—Review & Editing. K.H.: Methodology, Validation, Formal analysis, Resources, Writing—Review & Editing, Visualization. L.B.R.: Conceptualization, Methodology, Formal analysis, Resources, Data Curation, Writing—Review & Editing, Supervision, Project administration, Funding acquisition. E.H.: Conceptualization, Methodology, Formal analysis, Writing—Review & Editing, Supervision, Project administration. All authors have read and approved the final version of the manuscript.

## Funding

This work was supported by Aage Louis Hansen fonden [J.nr.20-2B-5995]; Herlev internal research funding [after year, 2019]; Colitis-Crohn Foreningen [2021]; Torben og Alice Frimodts Fond [FK1900]; Aase and Ejnar Danielsen's Foundation [20-10-0411].

## 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/s41598-023-45569-3>.

**Correspondence** and requests for materials should be addressed to J.P.J.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://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) 2023