

# SCIENTIFIC REPORTS



OPEN

## The non-invasive exfoliated transcriptome (exfoliome) reflects the tissue-level transcriptome in a mouse model of NSAID enteropathy

Canaan M. Whitfield-Cargile<sup>1</sup>, Noah D. Cohen<sup>1,5</sup>, Kejun He<sup>2,7</sup>, Ivan Ivanov<sup>3,5</sup>, Jennifer S. Goldsby<sup>4,5</sup>, Ana Chamoun-Emanuelli<sup>1</sup>, Brad R. Weeks<sup>6</sup>, Laurie A. Davidson<sup>4,5</sup> & Robert S. Chapkin<sup>4,5</sup>

Non-steroidal anti-inflammatory drugs (NSAIDs) are among the most frequently used classes of medications in the world, yet they induce an enteropathy that is associated with high morbidity and mortality. A major limitation to better understanding the pathophysiology and diagnosis of this enteropathy is the difficulty of obtaining information about the primary site of injury, namely the distal small intestine. We investigated the utility of using mRNA from exfoliated cells in stool as a means to surveil the distal small intestine in a murine model of NSAID enteropathy. Specifically, we performed RNA-Seq on exfoliated cells found in feces and compared these data to RNA-Seq from both the small intestinal mucosa and colonic mucosa of healthy control mice or those exhibiting NSAID-induced enteropathy. Global gene expression analysis, data intersection, pathway analysis, and computational approaches including linear discriminant analysis (LDA) and sparse canonical correlation analysis (CCA) were used to assess the inter-relatedness of tissue (invasive) and stool (noninvasive) datasets. These analyses revealed that the exfoliated cell transcriptome closely mirrored the transcriptome of the small intestinal mucosa. Thus, the exfoliome may serve as a non-invasive means of detecting and monitoring NSAID enteropathy (and possibly other gastrointestinal mucosal inflammatory diseases).

Non-steroidal anti-inflammatory drugs (NSAIDs) are among the most frequently consumed pharmaceuticals worldwide because of their anti-inflammatory, anti-neoplastic, and analgesic effects. Their use can result in an enteropathy that has an alarmingly high rate of morbidity and mortality. In the United States alone, NSAID enteropathy results in approximately 100,000 hospitalizations and 16,500 deaths each year<sup>1</sup>. An additional 2/3 of both short- and long-term NSAID users develop subclinical or undiagnosed distal small intestinal lesions<sup>2</sup>. Although detection and management of NSAID-induced lesions of the *proximal* GI tract (*i.e.*, gastropathy) are well documented, diagnosis and treatment of NSAID-induced damage to the GI tract *distal to the duodenum* (also known as NSAID enteropathy, affecting primarily the distal jejunum and ileum) remain elusive<sup>3,4</sup>. This is noteworthy because the incidence of NSAID enteropathy is expected to increase as a result of greater use of NSAIDs to treat rising numbers of inflammatory conditions, to meet the needs of aging populations in North America

<sup>1</sup>Department of Large Animal Clinical Sciences, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, Texas, United States of America. <sup>2</sup>Department of Statistics, College of Science, Texas A&M University, College Station, Texas, USA. <sup>3</sup>Department of Veterinary Physiology and Pharmacology, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, Texas, USA. <sup>4</sup>Program in Integrative Nutrition & Complex Diseases, College of Agriculture and Life Sciences, Texas A&M University, College Station, Texas, USA. <sup>5</sup>Center for Translational Environmental Health Research, Texas A&M University, College Station, Texas, USA. <sup>6</sup>Department of Veterinary Pathobiology, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, College Station, Texas, USA. <sup>7</sup>Present address: Institute of Statistics and Big Data, Renmin University of China, Beijing, 100872, China. Correspondence and requests for materials should be addressed to N.D.C. (email: [Ncohen@cvm.tamu.edu](mailto:Ncohen@cvm.tamu.edu))

and Europe, and for their anti-neoplastic effects<sup>5</sup>. The lower GI tract of multiple mammalian species is affected by NSAIDs in a similar manner in terms of anatomic location, pathological findings, and severity of clinical signs<sup>6–8</sup>.

The pathophysiology of NSAID enteropathy is complex and poorly understood<sup>9</sup>. Deleterious effects of NSAIDs on the intestinal mucosa including enterocyte cell death, increased mucosal permeability, and interaction of the damaged mucosa with luminal contents including bacteria (*i.e.*, GI microbiota) and bacterial products or components such as lipopolysaccharide (LPS)<sup>4,10</sup> has been proposed. The resulting inflammatory cascade is mediated by the innate immune response to LPS and several pro-inflammatory cytokines including tumor necrosis factor (TNF), interleukin (IL)-1, and IL-6<sup>11–13</sup>. Although the GI microbiota has recently been implicated as an important contributor to NSAID enteropathy, the precise mechanisms of host-microbiota interactions remain to be elucidated<sup>14–18</sup>.

An important limitation to understanding the pathogenesis of NSAID enteropathy is the difficulty in obtaining longitudinal (sequential) data from individuals regarding intestinal function and health. A great clinical and investigative need exists to develop non-invasive methods to characterize the health and function of the GI tract distal to the stomach to more effectively identify, study, and manage NSAID enteropathy. A potential strategy to address this limitation is the use of exfoliated intestinal epithelial cells (IECs) and other cell types found in voided stool. Approximately 1/3 of human colonic epithelial cells (up to 10<sup>10</sup> cells in an adult) are exfoliated and shed in feces each day<sup>19</sup>. Isolation and sequencing of the mRNA (host transcriptome) from exfoliated cells has been validated in the context of colon carcinogenesis in rats and humans, and in characterizing human neonatal gastrointestinal development<sup>20–24</sup>. Exfoliated cells, however, have not been used to evaluate a disease affecting the *small intestine*. Thus, the objective of this study was to determine whether exfoliated cells could be used as a non-invasive method for detecting and studying NSAID enteropathy in a murine model. Specifically, we performed RNA-Seq on colonic and small intestinal mucosa and exfoliated host cells in feces. We then applied computational approaches, *e.g.*, linear discriminant analysis (LDA) and sparse canonical correlation analysis (CCA), to analyze the inter-relatedness of these data. The goals of these studies were to provide proof-of-principle that the exfoliated cell transcriptome (*i.e.*, the exfoliome) could be used to gain information about NSAID-induced small intestinal injury. Our specific aims were: 1) to determine whether the transcriptome of exfoliated cells reflected gene expression of the small intestinal mucosa; 2) to demonstrate that the transcriptome of exfoliated cells can be used to differentiate healthy and diseased phenotypes; and, 3) to generate hypotheses regarding key biological pathways and processes involved in the pathogenesis of NSAID enteropathy.

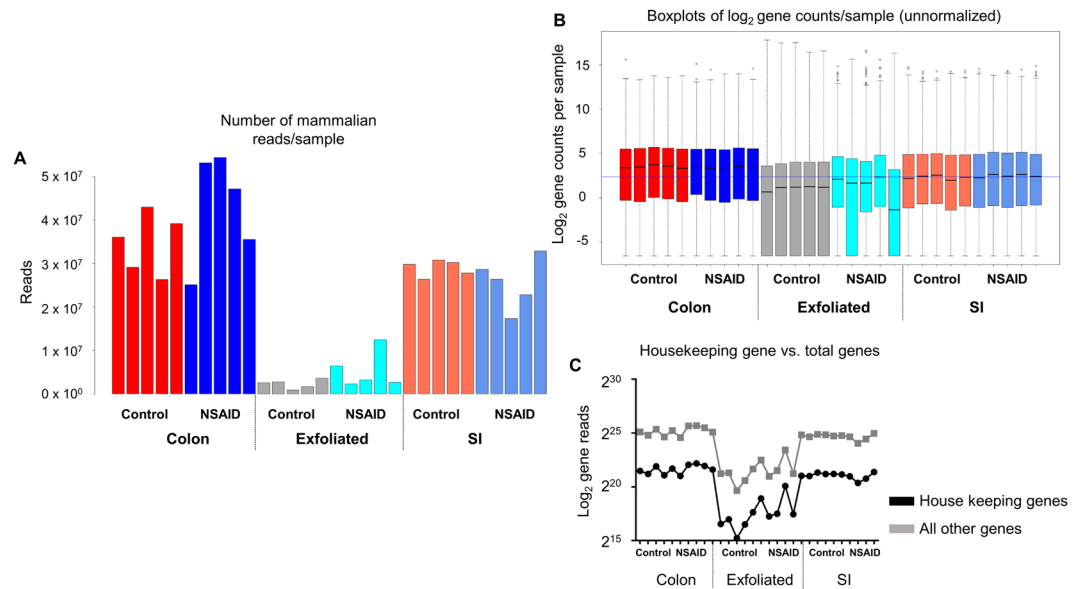
## Results

The lumen of the GI tract is a formidable environment for RNA transcripts, particularly the small intestine because of the abundance of both host and microbial enzymes and the longer transit time in the SI relative to the colon. Thus, after extracting RNA from tissues and exfoliated cells we examined RNA quality owing to the potential for degradation of mRNA from exfoliated SI cells in stool as it passes through the GI tract. As expected, Bioanalyzer results revealed lower quality RNA in the exfoliated cells than in the tissue (Figure S1a–c). Bioanalyzer traces show that the majority of RNA in the exfoliated cell samples is of microbial origin (23S and 16S rRNA subunits) (Figure S1b). However, due to the oligo dT probe used in the first step of library construction, the mouse transcripts were selectively targeted for cDNA production and subsequent library content. Fastq results of the subsequent sequenced transcripts revealed high fidelity and quality with all fecal and tissue samples (Figure S2a,b).

Sequencing of these data revealed that the RNA sequencing reads for SI mucosa, colonic mucosa, and exfoliated cells mapped to an average of 19,324 genes, 20,743 genes, and 13,944 genes per sample, respectively (Figure S3a–c). Genes present in low abundance (*i.e.*,  $\leq 4$  animals or  $\leq 50$  times) across all samples were removed from all datasets and the remaining genes subjected to downstream analysis. This reduced the number of transcripts for downstream analysis in the SI to an average of 17,229 genes, the colon data to 17,244 genes, and the exfoliated data to 10,865 genes per sample. Although filtering reduced the total number of genes in the SI and colon (by 11% and 17%, respectively) less than in the exfoliated cells (22%), the total number of reads was negligibly affected in all datasets (SI reduced from 273,065,200 reads across all samples to 273,026,424 – a 0.001% reduction; colon data reduced from 389,222,037 to 389,122,199 – a 0.003% reduction; exfoliated cell data reduced from 38,292,084 to 38,160,589 – a 0.003% reduction) (Figure S3d–f).

After filtering to remove genes present in low abundance, we examined scatter plots of log(2) counts per million (CPM) for each gene in the 2 treatment groups (*i.e.*, NSAID and control), comparing the exfoliome to the SI and colonic transcriptome in a pairwise manner (Figure S4a,b). There was strong and significant correlation of the CPM data for each of the pairwise comparisons (Spearman's correlation coefficient; R value > 0.8 and P < 0.0001 for each pairwise comparison), irrespective of treatment group. We then examined total mammalian gene counts and the number of reads/gene obtained for each sample of the raw, filtered data for all datasets (Fig. 1A,B). There was obvious variation between sources (*i.e.*, exfoliated cells, SI, or colonic RNA) in total mammalian reads and distribution of counts. This difference was thought to be due to microbial RNA contamination of the RNA extracted from exfoliated cells (Figure S1b) resulting in fewer reads mapping to the mouse genome. To confirm this, we extracted total counts of 532 genes that have been previously identified as housekeeping genes from each sample and plotted those relative to total gene counts<sup>25</sup>. Examination of the relative abundance of these genes in each sample revealed that between-sample total count differences were represented by similar magnitudes of differences in abundance of these 532 house-keeping genes, indicating these differences were due to smaller library size attributable to mammalian transcripts and not to sequencing artifact (Fig. 1C).

Given the magnitude of differences in library size attributable to mammalian transcripts between the exfoliome and the tissue transcriptome, it was necessary to perform the remainder of the analyses separately for each source of RNA (*i.e.*, SI, colon, or exfoliated cells). To account for between-sample variation in read-counts, RNA-Seq data for each dataset were normalized with edgeR accounting for group effects using the function

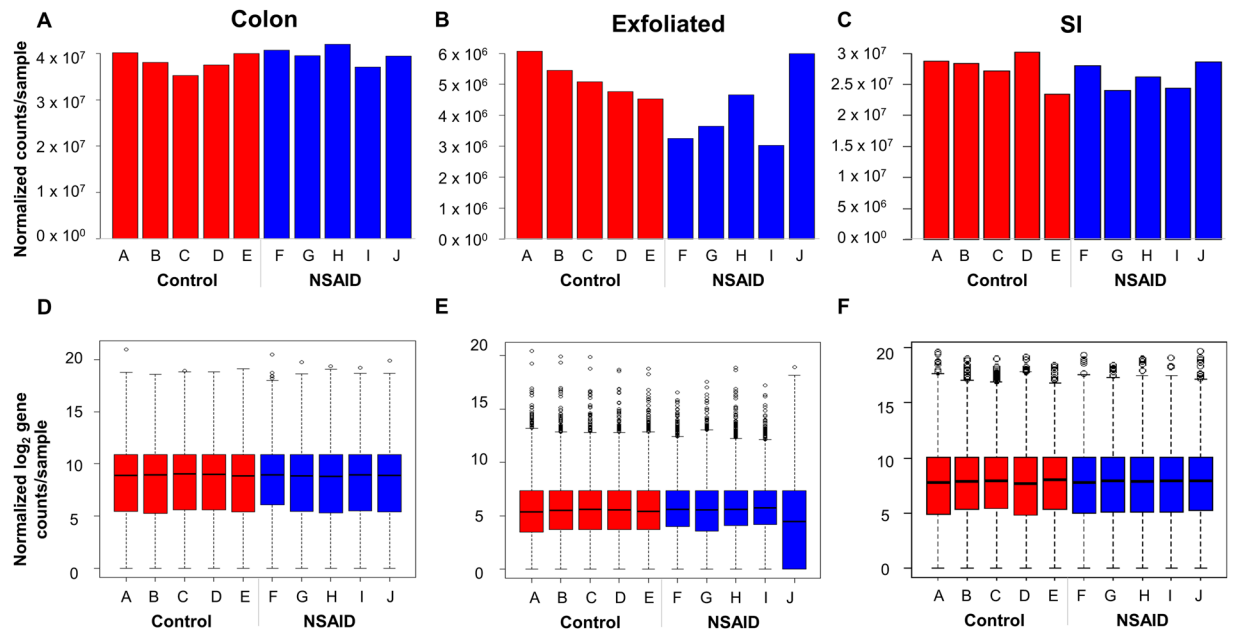


**Figure 1.** The exfoliome contains fewer reads attributable to the mouse genome than the tissue transcriptomes due to bacterial RNA contamination: (A) Number of mammalian reads per sample for each animal and data source colored by treatment group. (B) Log(2) counts per gene per sample across all treatment groups from the sequenced RNA colored by treatment group. (C) Log(2) total gene reads of 532 murine house-keeping genes (black) and all other genes (grey) per sample across all animals and treatment groups from each data source.

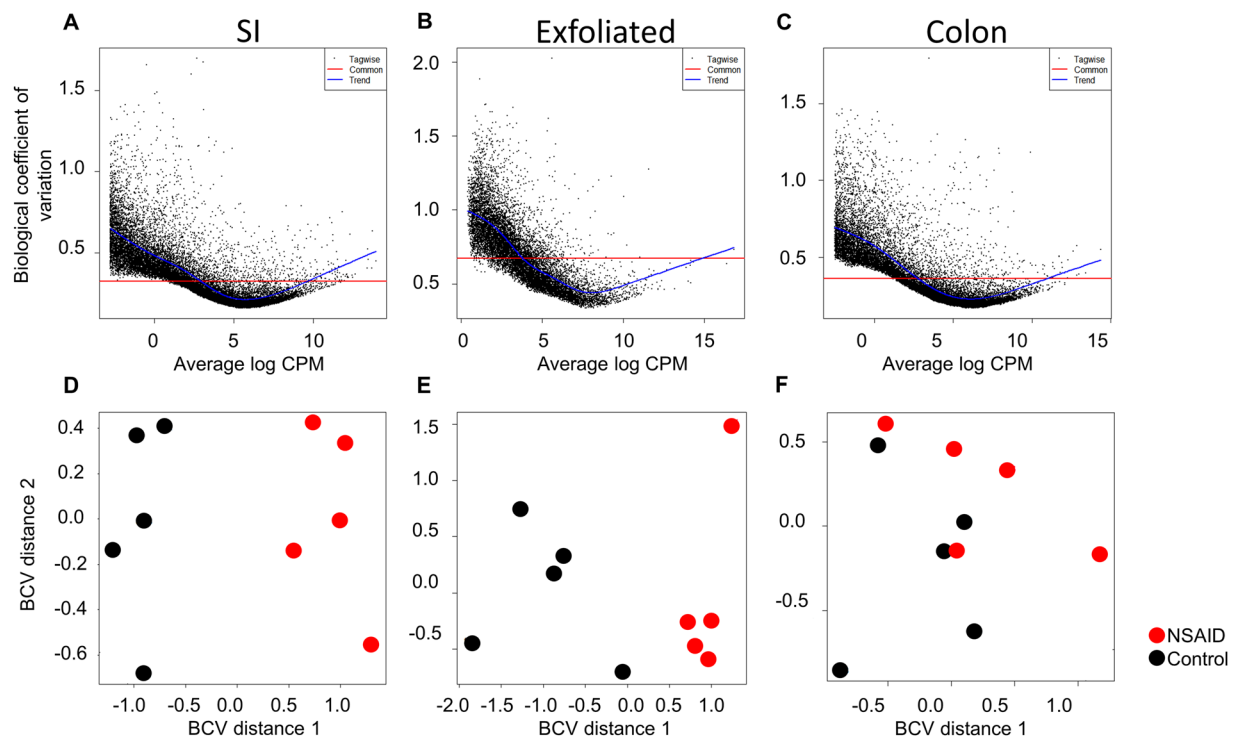
*calcNormFactors* and the upper-quartile method. Total gene-counts and boxplots of the number of reads/gene for each sample of the normalized data for both tissue and exfoliated cell datasets are shown in Fig. 2A–F. Variability in abundance of post-normalization total house-keeping genes was also improved (Figure S5). Prior to identifying differentially expressed (DE) genes, we assessed biological variability in the exfoliome relative to the tissue transcriptomes. Biological coefficient of variation (BCV) versus the mean log counts per million (CPM) and multi-dimensional scaling (MDS) plots were used to visually assess the similarity of the samples within each treatment group (Fig. 3A–F). These results demonstrate a relatively high degree of variation in the exfoliome (common dispersion = 0.592 and BCV = 0.769) (Fig. 3B) as compared to both tissue transcriptomes (Fig. 3A and C), with a common dispersion of 0.143 and BCV of 0.379 for the SI and common dispersion of 0.126 and BCV of 0.329 for the colon. Notably, MDS based on BCV revealed clear separation of the treatment groups in both the SI transcriptome and exfoliome but not the colonic transcriptome (Fig. 3D–F).

Prior to analyzing these data in order to derive biological meaning, we first wished to determine the anatomic origin of the cellular signature derived from the exfoliome. In order to determine the source of this signature we extracted the counts of genes previously identified and expressed predominantly in specific anatomic locations (*i.e.* stomach, pancreas, small intestine, and colon). Interestingly, we found that the exfoliome contained virtually no reads from genes representing the stomach or pancreas. In contrast, there was a clear signature arising from both the colon and small intestine (Fig. 4A). As expected genes representing the colon and small intestine were heavily represented in the transcriptomes arising from those locations with some overlap. Similarly, in addition to anatomic origin we also wished to determine the cell types represented in the exfoliome. Clearly, the intestinal mucosa is comprised not only of IECs but also stem cells, crypt cells, goblet cells, Paneth cells (SI), as well as a host of infiltrating immune cells depending on depth of the sample (*i.e.*, lamina propria) and disease state of the GI tract (*e.g.*, inflammation vs. homeostasis). To try to determine the cell types present in these data, we reviewed the literature for marker genes expressed either solely by a specific cell type or at least highly enriched in a specific cell type<sup>26–38</sup>. In particular, we extracted the numbers of reads in each sample across all 3 datasets for the following cell types: intestinal stem cells, IECs, crypt base columnar cells, Paneth cells, tuft cells, goblet cells, macrophages, lymphocytes, neutrophils, and smooth muscle. A list of the genes used as biomarkers for each of these cell types is shown in Table S1. Interestingly, we found that all cell types were present in all datasets as identified by the presence of at least 2 marker genes per cell type (Fig. 4B). Visually there were minimal differences among the 3 datasets with the expected exception of fewer reads in the exfoliated cell data and absence of a few marker genes of intestinal stem cells in the exfoliome with concurrent low expression in the tissue transcriptomes. These data suggest that the mucosal transcriptome and exfoliome represent signatures from similar cell types that comprise not only IECs but reads from the diverse array of cell types expected to be found in the intestinal mucosa.

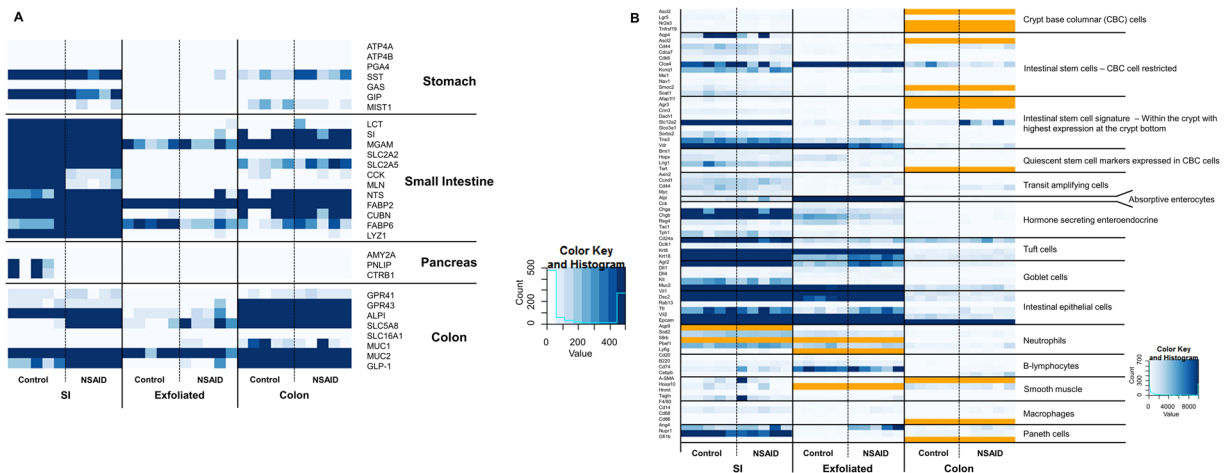
After characterizing the cellular source of the signatures derived from these datasets, we examined each dataset for alterations induced by NSAIDs by comparing the transcriptome (or exfoliome) between the control group and NSAID group. In human subjects and preclinical models, NSAID-induced lower GI damage primarily occurs in the distal jejunum and ileum<sup>3,4</sup>. As expected, we observed marked pathological findings in the distal SI but no notable microscopic abnormalities in the colon in NSAID-treated mice (Fig. 5A). Therefore, RNA extracted from



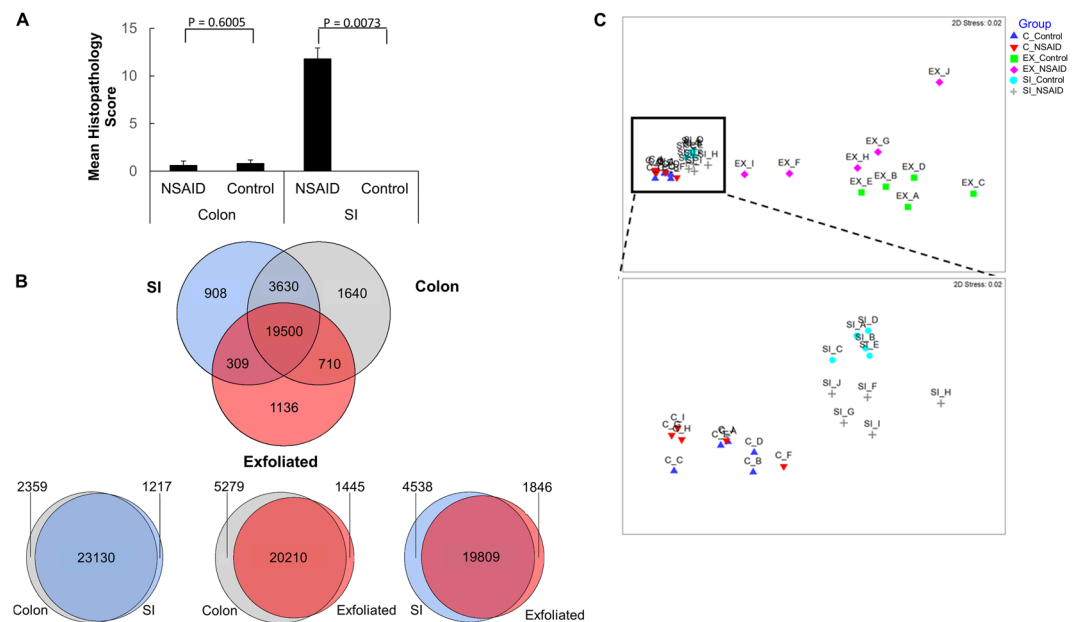
**Figure 2.** Raw data after filtering and normalization show that the between sample variation in exfoliated cell reads is improved and similar to tissue reads. Total gene counts after normalization for each sample across all treatment groups from the sequenced RNA extracted from (A) colonic mucosa, (B) exfoliated cells and (C) SI mucosa. Normalized log<sub>2</sub> counts per gene per sample across both treatment groups from sequenced RNA extracted from (D) colonic mucosa, (E) exfoliated cells and (F) SI mucosa.



**Figure 3.** SI transcriptome and exfoliome cluster by treatment group in contrast to colonic transcriptome and biological variability is higher in the exfoliome than the tissue transcriptomes. Biological coefficient of variation (BCV) versus the mean log counts per million (CPM) of the SI transcriptome (A), exfoliome (B) and colonic transcriptome (C). (D) Treatment-based multi-dimensional scaling (MDS) plots of the SI transcriptome and that of the exfoliome (E) and colon (F).



**Figure 4.** (A) The exfoliome signature arises from cells sloughed from both the small intestine and colon and comprises reads from the diverse array of cell types expected to be found in the intestinal mucosa (A) Heatmap showing counts of genes that are reported to be primarily expressed at specific anatomic locations (stomach, pancreas, small intestine, colon). All genes with counts greater than 400 are colored dark blue. (B) Heatmap showing counts of biomarker genes from each sample and each data source (orange = gene not expressed).

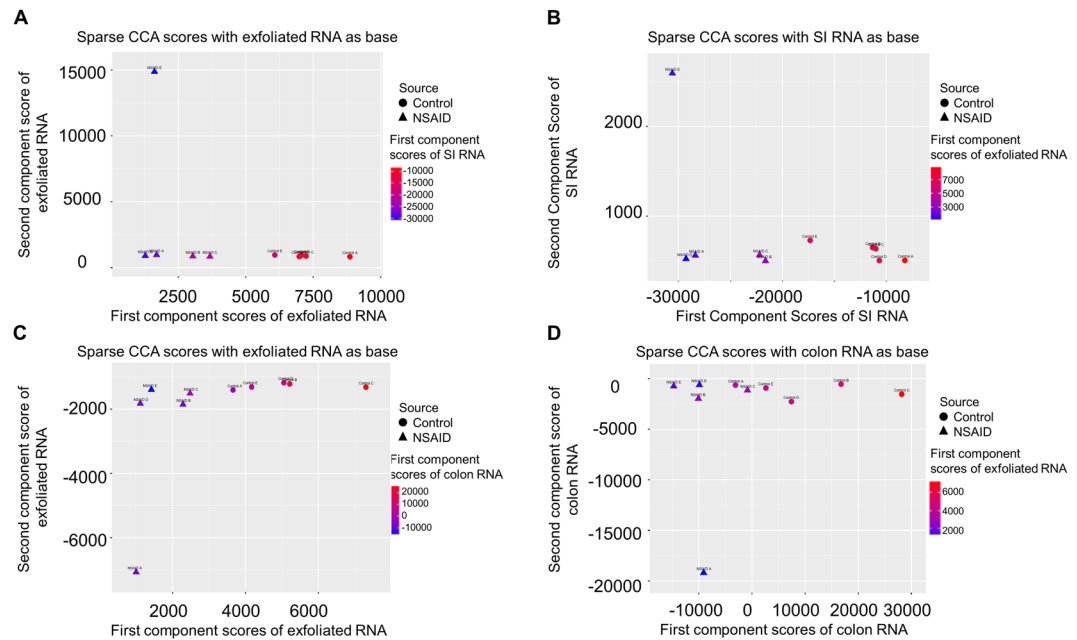


**Figure 5.** Microscopic pathology reveals NSAID injury is confined to the SI. Despite great overlap between the three RNA-Seq datasets, the exfoliome distinguishes between NSAID and control animals similar to the SI transcriptome, whereas the colonic transcriptome does not. (A) Microscopic pathology scores from colon and small intestinal mucosa in control mice and NSAID-treated mice. (B) Venn diagram showing intersection of gene lists among the datasets. (C) Multi-dimensional scaling plot of each sample color-coded by source and treatment group. Inset of panel C enlarged to show degree of separation of groups in the SI transcriptome and lack of clear separation in colonic transcriptome.

SI mucosal scrapings of NSAID-treated subjects should demonstrate marked mucosal pathology whereas the RNA from colonic mucosal scrapings should reflect minimal pathology.

In order to determine whether the exfoliome more closely resembled the colonic transcriptome or the SI transcriptome, we first crudely examined the intersection of gene lists from each source. This analysis revealed >90% overlap among the 3 datasets in terms of presence and absence of genes (Fig. 5C). Despite this overlap of genes, non-metric multidimensional scaling (NMDS) plots demonstrated differences among these 3 datasets (Fig. 5C and inset). Although the exfoliated cell data differed from the SI and colon data because of smaller library size and fewer mammalian reads potentially resulting from degradation of RNA in the GI tract, evidence of clustering of the treatment groups was observed in the SI and exfoliated cell data but was absent in the colonic data (Fig. 5C



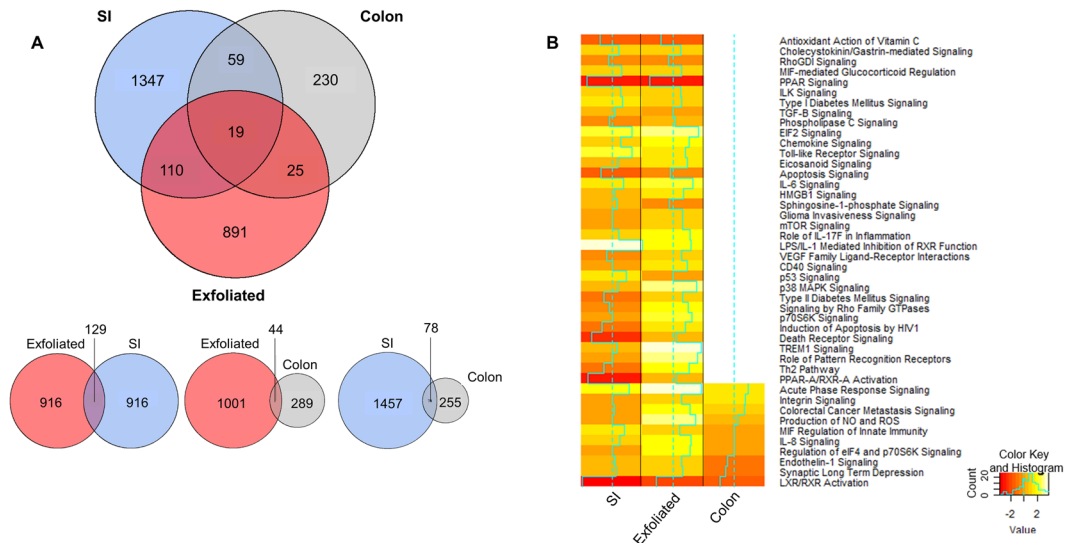


**Figure 6.** Sparse canonical correlation analysis (CCA) reveals that the global transcriptome profiles from exfoliated cells correlates well with the transcriptome profile of the SI. In contrast, the colonic transcriptome data do not discriminate well between treatment groups. Sparse CCA plots positioned by 1<sup>st</sup> and 2<sup>nd</sup> component scores from (exfoliated cells and colored by the 1<sup>st</sup> component SI scores, (B) SI and colored by 1<sup>st</sup> component scores from exfoliated cells, (C) exfoliated cells and colored by the 1<sup>st</sup> component colon scores and (D) colonic mucosa and colored by the 1<sup>st</sup> component exfoliated scores.

and inset). Analysis of similarity (ANOSIM) based on the Bray-Curtis dissimilarity metric quantitatively demonstrated differences between NSAID and control groups for the SI (R value = 0.760;  $P = 0.008$ ) and exfoliated cells (R value 0.280;  $P = 0.024$ ) but not for the colonic data (R value 0.194;  $P = 0.090$ ).

To further examine the interdependent relationship between these 3 transcriptomic profiles, we utilized sparse CCA, a novel multivariate statistical analysis approach. Sparse CCA is a dimensionality reduction technique that identifies the fewest numbers of genes that show the greatest amount of correlation between datasets according to specific optimality criteria. Although the sparse CCA plots should not be assigned any particular biological interpretation, they can be considered a stringent method for determining correlation of large datasets<sup>39</sup>. As revealed by NMDS plots, sparse CCA plots demonstrated that the transcriptome from exfoliated cells correlated well with the SI transcriptome in these mice, and that the SI and exfoliated cell datasets discriminated NSAID-treated from control groups, whereas the colonic mucosal transcriptome did not (Fig. 6A–D).

We next examined the similarities and differences in the gene expression profiles from exfoliated cells compared with those from the scraped intestinal mucosa. We identified DE genes between control and NSAID-treated mice for each dataset. Interestingly, both the exfoliome and SI transcriptome had >1000 DE genes (FDR  $P < 0.05$ ) between control and NSAID-treated mice, whereas there were relatively few DE genes ( $n = 333$  genes) in the colonic transcriptome (Fig. 7A). Venn diagrams revealed sparse overlap (12%) of the DE expressed genes between the SI transcriptome and the exfoliome and even less overlap between the colonic transcriptome and the exfoliome (6%; Fig. 7A). Despite this sparse overlap, the pathways enriched in the exfoliome and SI transcriptome were similar whereas there was much less similarity between the exfoliome and colonic transcriptome. Specifically, IPA Ingenuity Knowledgebase ([www.ingenuity.com](http://www.ingenuity.com)) pathway analysis revealed both SI and exfoliated cell datasets exhibited similar occupancy and predicted directionality (Z-score) in the canonical pathways represented (Fig. 7B). In contrast, few pathways were represented in the colonic data (Fig. 7B). For example, Toll-like receptor signaling (which is known to play a crucial role in the pathogenesis of NSAID enteropathy)<sup>9,10,13,18,40,41</sup> was upregulated by NSAID administration in both the exfoliome and small intestinal transcriptome but genes related to this pathway were not altered in the colonic transcriptome resulting in no occupancy of this canonical pathway. Indeed, the proportion of pathways represented in the colon (31%; 21/67) was significantly less than that of the exfoliome (84%; 56/67;  $P < 0.0001$ ) and the SI (93%; 62/67;  $P < 0.0001$ ). Moreover, Pearson's correlation revealed significant correlation of the Z scores from the exfoliome canonical pathways with those of the SI transcriptome ( $R = 0.3468$ ;  $P = 0.0110$ ). Further, McNemar's test revealed significant ( $P < 0.0001$ ) discordance in pathways represented between colon and SI and colon and exfoliated cells, but not between SI and exfoliated cells ( $P = 0.149$ ). A similar finding was observed when examining upstream regulators of these pathways (Figure S6). Specifically, there were significantly fewer upstream regulators expressed in the colon (41%; 780/1888) than in either the SI (73%; 1380/1888;  $P < 0.0001$ ) or exfoliome (78%; 1469/1888;  $P < 0.0001$ ). When comparing the upstream regulators expressed in both SI and exfoliated cells, there was significant correlation between the fold-changes in SI and exfoliome gene expression ( $R = 0.616$ ;  $P < 0.001$ ). In order to determine the correlation between expression



**Figure 7.** The exfoliated cell transcriptome is similar to the tissue transcriptome as shown by overlapping gene lists and pathways. **(A)** Venn diagram of the intersection of differentially expressed genes found in the exfoliome and tissue transcriptomes. **(B)** Heat map of the Z-scores of the canonical pathways to which the differentially expressed genes between control and NSAID-treated animals were mapped from the SI transcriptome (left column), exfoliome (middle column) and the colonic transcriptome (right column).

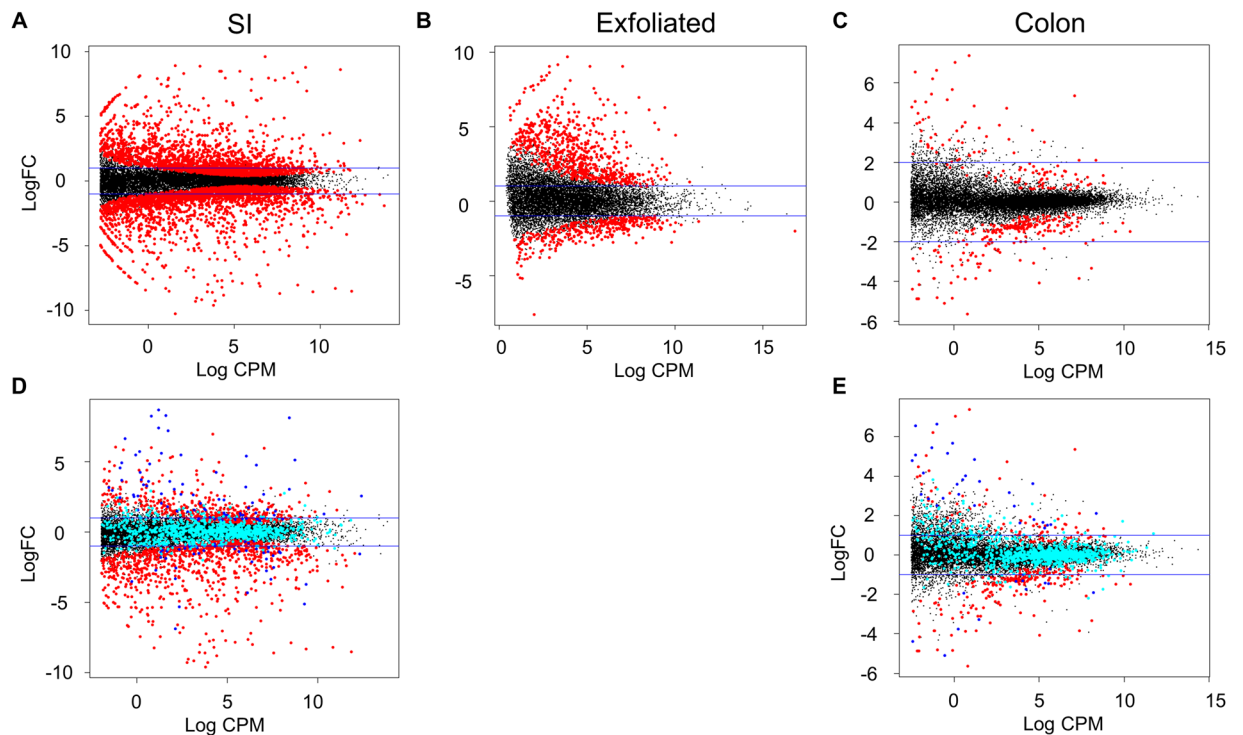
levels in exfoliated cells and the tissues and to demonstrate effect sizes of the differences in gene expression, MA plots for each dataset were constructed and colored to show where the DE genes in the exfoliome resided in both tissues (Fig. 8A–E)<sup>42</sup>. These plots confirmed lack of overlap between DE genes within each dataset and show that the DE genes within the exfoliome had a greater effect-size than those within the tissue.

The ability of both the SI transcriptome and exfoliome to differentiate between control and NSAID-treated animals was further substantiated by LDA of the list of differentially expressed genes between treatment groups in the 3 datasets to identify the most biologically-relevant genes. LDA was performed on the counts of DE genes identified from the exfoliome, SI transcriptome, and colonic transcriptome. The 1-feature (*i.e.*, single gene) LDA identified 34 and 49 genes from the SI and exfoliated datasets, respectively, that differentiated control from NSAID-treated animals with less than 5% error, but only 17 genes from the colonic data (Table S2). These genes were subsequently uploaded into IPA along with their respective P-values and fold-changes in order to determine which pathways identified by LDA could best classify NSAID-treated versus control mice with respect to all differentially expressed genes. The top (lowest P-value) upstream regulators and the networks to which these genes belong are shown in Fig. 9.

## Discussion

Here we have shown that the transcriptome of exfoliated cells closely resembles that at the affected tissue level in a murine model of NSAID enteropathy. Using RNA-Seq, we compared the exfoliome to the transcriptome of mucosa from the SI and colon. Although gene expression of exfoliated cells has been used to examine GI development in neonates and for biomarkers of colorectal cancer, we show for the first time the application of this approach to examine a disease that primarily affects the SI<sup>2,9,13,14,43</sup>. The use of exfoliated cell mRNA presents several important advantages relative to mRNA obtained from the standard approach of tissue biopsy. First, use of fecal samples is non-invasive and results in no physical pain to the patient. Second, greater requirement of time, effort, and cost is required for collection of biopsy specimens than for collection of feces. Third, although the primary area of injury with NSAID enteropathy is the distal small intestine, NSAIDs can affect both the large and small intestine of humans and other animals<sup>44,45</sup>. Exfoliated cells arise from both the colon and SI and thus provide an overview of the gene expression profiles of the entire GI tract, whereas a tissue specimen only provides a site-specific gene expression profile<sup>21</sup>. Fourth, intestinal biopsies generally can be collected once (typically at a terminal stage of the study) in mice, whereas feces can be collected sequentially, thereby allowing the temporality of gene expression profiles from exfoliated cells to be used to investigate and monitor disease progression. If the same findings hold true in humans, fecal samples could be used as a non-invasive tool for detecting and monitoring patients with NSAID enteropathy.

Concerns regarding the use of mammalian RNA found in voided stool are valid given the likelihood of RNA degradation as it passes through the GI tract. Indeed, we observed differences in the quality of RNA found in stool versus mucosa. It appears that appropriate selection of polyA tails and utilization of purification steps outlined in this manuscript results in RNA that is suitable for sequencing. Based on both the results of the quality control measures utilized here and the similarity of the signature in the exfoliome versus the tissue transcriptomes, results obtained by sequencing the exfoliome can be accurate and informative. An additional aspect that might have improved quality scores is the fact that we performed 50-bp, single-end sequencing because longer transcripts



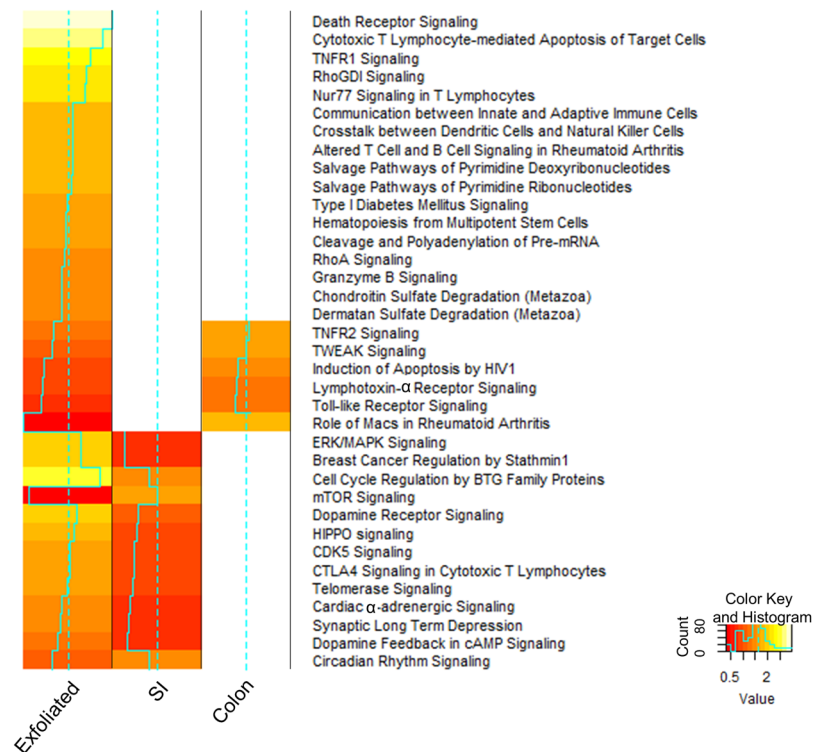
**Figure 8.** MA plots demonstrate the expression of genes identified as differentially expressed (DE) in the exfoliome and in the tissue transcriptomes. MA plot of genes from the (A) SI transcriptome, (B) exfoliome and (C) colon transcriptome with DE genes (FDR < 0.05) colored red. (D) MA plot of genes from the exfoliome AND SI transcriptome colored blue. Genes only differentially expressed in the exfoliome are colored cyan. (E) MA plot of genes from the colonic transcriptome with genes differentially expressed in the exfoliome AND colonic transcriptome colored blue. Genes only differentially expressed in the exfoliome are colored cyan.

likely underwent further degradation as they passed through the GI tract, which could have resulted in lower quality scores for the bases located beyond the 50-bp segments sequenced in our study.

In general, the exfoliome correlated well with the tissue transcriptome as evidenced by the intersection of genes present in all datasets. From a clinical perspective, the exfoliome was similar to the SI mucosal transcriptome, and discriminated NSAID-treated from control mice similarly to mucosa from the ileum, the region of the GI tract most severely affected in NSAID enteropathy in both mice and human subjects<sup>40,46–49</sup>. In contrast, the transcriptome of the colonic mucosa, which sustained less injury due to NSAID treatment, failed to discriminate between treatment groups. These results are interesting because much of the RNA extracted from exfoliome undoubtedly originates from the colon. Although we did not attempt to identify the precise source of these cells, marker genes representing various anatomic locations within the GI tract (*i.e.*, stomach, small intestine, pancreas, and colon) revealed that the signature in the exfoliome was derived from both the small intestine and the colon with virtually no signature coming from the stomach or pancreas. Additionally, the many cell types that reside in the mucosa and lamina propria were present in the exfoliome (and tissue), indicating that these data represent many cell types. It is impossible to determine the exact source of each transcript found in the exfoliome and this may explain, in part, the lack of overlap between DE genes between the exfoliome and SI transcriptomes. For example, inflammatory cytokines are expressed following NSAID-induced impairment of barrier function and resultant influx of LPS and other components of the microbiome<sup>16,50,51</sup>. The many cell types found in the intestinal mucosa are then exposed to these cytokines and each cell type may behave differently following exposure to these cytokines. One example of such differing behavior in response to similar stimulation is the dichotomy of the effects of macrophages and neutrophils in response to intestinal inflammation whereby macrophages dampen the effects of neutrophilic inflammation in the gut<sup>52</sup>. This could explain similar pathway analysis results arising from lists of DE genes that do not have substantial overlap. Nonetheless, the ability of the exfoliome to discriminate between control and NSAID-treated mice in a manner similar to the SI transcriptome indicates that studying the exfoliome can yield informative results about the distal small intestine irrespective of knowledge of the precise source of the exfoliated cells. Further investigation of the impact of concomitant colonic disease and other potential confounding factors on use of the exfoliated cell transcriptome is merited.

Having demonstrated that the exfoliome reflected the SI transcriptome, we used gene expression data to examine which pathways and processes were implicated in NSAID enteropathy. Our results further substantiate the importance of innate immune responses in NSAID enteropathy<sup>53,54</sup>. Specifically, and as reported previously, we identified several pro-inflammatory pathways and upstream regulators including IL-8 (mouse homologs), IL-17, NF-KB, IL-1, and MYD88 as being involved in NSAID enteropathy<sup>13,40,53–55</sup>. Toll-like receptor (TLR) signaling





**Figure 9.** 1-Feature linear discriminate analysis (LDA) substantiates the ability of both the SI transcriptome and exfoliome to differentiate between control and NSAID-treated animals. Heatmap showing Z-scores of canonical pathways identified by 1-feature LDA of genes identified as being DE (FDR < 0.05) in each dataset.

and specifically TLR-4 is known to be a critical component of the pathophysiology of NSAID enteropathy<sup>40,41</sup>. As expected, this pathway was highly up-regulated in both the SI transcriptome and the exfoliome but expression of the genes belonging to this pathway were not altered in the colonic mucosa and therefore this pathway showed no occupancy in the data arising from the colonic mucosa. The lack of occupancy of this and other pathways crucial to the pathogenesis of NSAID enteropathy in the colonic data, including apoptosis, and IL-1 and IL-6 pathways, provide further evidence that the exfoliome signature reflects the SI mucosa in the presence of SI disease. The consistency of our exfoliome results with those of prior studies implicating innate immunity as a critical player in NSAID enteropathy adds evidence to the validity of using this novel approach for the study of NSAID enteropathy (and potentially numerous other small intestinal disorders).

There were several limitations to this study. First, the evaluated sample size was small and consisted of a single time-point (after 7 days of NSAID treatment). Despite the small sample size, we clearly demonstrated that the exfoliated cell transcriptome reflects that of the tissue at the level of the ileum. In addition, our approach revealed the importance of several pathways involved in both NSAID enteropathy. Therefore, even with a small sample size, we were able to recapitulate known biologically relevant findings with this technique. In the latter or more severe stages of NSAID enteropathy, the major pathways involved are the host response to increased mucosal permeability. Therefore, to better understand the temporal events contributing to NSAID enteropathy, a more appropriate approach might be to examine earlier time points.

In summary, we have demonstrated that the exfoliated cell transcriptome closely mirrors the transcriptome of the tissue at the level of the ileum. Although use of the exfoliome has been described previously, this is the first time these findings have been correlated to the tissue-level transcriptome. In addition, this technique has not been utilized to examine a disease that primarily affects the SI. We have shown that the exfoliated cell transcriptome can be a useful tool to examine NSAID enteropathy and potentially extend to a number of other GI diseases, including those affecting the SI. Additionally, the application of noninvasive exfoliated cell-based techniques might be useful for the early identification and monitoring of therapeutic interventions targeting NSAID enteropathy and other intestinal disorders.

## Methods

**Animals and sample collection.** Animal protocols were approved by the Texas A&M Institutional Animal Care and Use Committee in accordance with appropriate institutional and regulatory bodies' guidelines. Animals were handled and treated as previously described<sup>43</sup>. Briefly, 8- to 10-week-old, specific-pathogen-free female C57BL/6J mice were purchased and allowed to acclimate for 2 weeks. Mice were fed standardized laboratory rodent diet (Teklad rodent diet, Envigo product number 8604) and sterile water *ad libitum*. Mice were randomly divided into 2 groups (n = 5 mice/group): NSAID (indomethacin) or control (vehicle). Mice were then rehoused on the basis of treatment group assignment, with 5 animals/group-cage. To induce NSAID enteropathy, mice

in the NSAID group were gavaged once daily with indomethacin (5 mg/kg; for 7 days; Sigma Aldrich, St. Louis, MO) dissolved in DMSO (Sigma Aldrich, St. Louis, MO) and further diluted in phosphate buffered saline (PBS). Mice in the control group were gavaged with equal volumes (200  $\mu$ L) of PBS including an equal concentration of DMSO (0.001%). Treatments occurred between 9 AM and 11 AM CST once daily for 7 days. Feces were collected daily by placing individual animals in sterile plastic cups that were RNase- and DNase-free until they passed feces. The mice were returned to their home cages and the feces immediately placed in 500  $\mu$ L denaturation solution (ThermoFisher Scientific), homogenized and flash frozen at  $-80^{\circ}\text{C}$ . On day 7, the mice were euthanized via  $\text{CO}_2$  asphyxiation. The small intestine and colon were removed, flushed with ice-cold PBS, fixed in 4% paraformaldehyde, Swiss-rolled, paraffin-embedded, and stained with hematoxylin and eosin. Stained sections of the small intestine were scored for intestinal inflammation as previously described<sup>56</sup>.

**Extraction of mRNA from exfoliated cells and library preparation.** PolyA<sup>+</sup> RNA was isolated from stool samples from mice as previously described<sup>24,57</sup>. RNA was quantified using a Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA) and the quality was assessed using the Nano6000 chip on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Each mouse sample was processed with the NuGEN Ovation 3'-DGE kit (San Carlos, CA) to convert RNA into cDNA followed by NuGEN Encore Rapid Library kit to create Illumina libraries, as per manufacturer's instructions. Briefly, 30 ng of each sample were used to synthesize first and second strand cDNA, followed by purification using Agencourt RNAClean XP beads (Brea, CA) included in the kit. The cDNA was linearly amplified using the NuGEN SPIA primer, and cDNA quantity was determined using Nanodrop spectrophotometer (Thermo Fisher Scientific). Two micrograms of cDNA were fragmented using a Covaris S2 sonicator (Woburn, MA) with the following settings: duty cycle 10%, intensity 5, cycles/burst 100, and time 5 min. Fragmented samples were concentrated using the MinElute Reaction Cleanup kit (Qiagen, Germantown, MD) as per manufacturer's instructions. Samples were subsequently quantified using a Nanodrop spectrophotometer. Following cDNA fragment repair and purification, Illumina adaptors were ligated onto fragment ends and amplified to create the final library. Libraries were quantified using the Kapa Library Quantification kit (Kapa Biosystems-Sigma Aldrich, St. Louis, MO) and run on an Agilent DNA High Sensitivity Chip to confirm appropriate sizing and the exclusion of adapter dimers.

**Tissue mRNA extraction and library preparation.** RNA was extracted from ileal and colonic mucosal scrapings using an RNeasy mini kit (QIAGEN, Redwood City, CA) following the manufacturer's instructions and including on-column DNase treatment. RNA quantity was determined using a Nanodrop spectrophotometer (Thermo Fisher Scientific) and the quality was assessed using the Nano6000 chip on a Bioanalyzer 2100 (Agilent Technologies). The samples were randomized prior to RNA-Seq library preparation. Sequencing libraries were made using 250 ng of RNA and the TruSeq RNA Sample Preparation kit (Illumina) following the manufacturer's instructions.

**RNA sequencing and downstream processing.** Libraries were pooled and sequenced on an Illumina HiSeq. 2500 at the Texas AgriLife Genomics and Bioinformatics Services Core Facility (College Station, TX). Sequence data were uploaded into NCBI small reads archive (Accession number PRJNA290483). Sequencing data were provided in a de-multiplexed format and aligned using Spliced Transcripts Alignment to a Reference (STAR) software with default parameters and referenced against the genome of *Mus musculus* (Ensembl version GRCm38)<sup>58</sup>. Differentially expressed genes were determined using EdgeR based on the matrix of gene counts<sup>42</sup>. The open source package edgeR was selected based on two considerations: first, it uses the state-of-the-art general linear model framework for detection of differentially expressed genes and second, there is evidence that it performs well relative to other tools for differential gene expression<sup>59-61</sup>. Gene pathway involvement and intersections were analyzed using QIAGEN's Ingenuity<sup>®</sup> Pathway Analysis (IPA, QIAGEN, Redwood City, CA) by uploading appropriate gene lists with fold-change and false discovery rate (FDR) P-values.

**Sparse canonical correlation analysis (CCA).** In order to take into account the multivariate structure when assessing and ranking genes, the multivariate relationships between the exfoliated cell and tissue transcriptomic datasets were analytically quantified. Sparse CCA provides measures of the strength of multivariate association between variable sets as well as a means to interpret the role of variables in terms of the underlying multivariate relationship<sup>23</sup>. Briefly, for the 2 random vectors  $x = (x_1, x_2, \dots, x_p)^T$  and  $y = (y_1, y_2, \dots, y_q)^T$  CCA aims to find 2 projection directions  $u_1 \in \mathbb{R}^p$  and  $v_1 \in \mathbb{R}^q$  so that;

$$(u_1, v_1) = \underset{u,v}{\operatorname{argmax}} \operatorname{Corr}(u^T x, v^T y) = \underset{u,v}{\operatorname{argmax}} \frac{u^T \Sigma_{xy} v}{\sqrt{(u^T \Sigma_{xx} u)(v^T \Sigma_{yy} v)}}, \quad (1)$$

where,  $\Sigma_{xy}$ ,  $\Sigma_{xx}$ ,  $\Sigma_{yy}$  are covariance and variance matrices. Let  $X$  and  $Y$  denote  $n \times p$  and  $n \times q$  data matrices as the observations from random vectors  $x$  and  $y$ , respectively. Thus, the empirical version of the above objective function can be written as;

$$\begin{aligned} & \max_{u,v} u^T X^T Y v \\ & \text{subject to } u^T \frac{X^T X}{n} u = 1, \quad v^T \frac{Y^T Y}{n} v = 1. \end{aligned} \quad (2)$$

This objective function, however, does not have a closed-form solution when the sample size  $n$  is less than  $\min(p, q)$ . In addition, the result does not perform variable selection and hence there could be problems with the biological interpretability. To overcome these limitations, sparse CCA adds regularization conditions and obtains a sparse solution with a reduced computational cost<sup>39,62,63</sup>. The main idea of this approach is to utilize;

$$\begin{aligned} & \max_{u,v} && u^T X^T Y v \\ \text{subject to} &&& u^T X^T X u \leq 1, \quad v^T Y^T Y v \leq 1 \\ &&& \|u\|_1 \leq c_1, \quad \|v\|_1 \leq c_2, \end{aligned} \quad (3)$$

where the tuning parameters  $c_1, c_2$ , are positive. It has been shown that in high-dimensional problems, treating the covariance matrix as a diagonal one yields satisfactory results<sup>39,62</sup>. Thus;

$$\begin{aligned} & \max_{u,v} && u^T X^T Y v \\ \text{subject to} &&& \|v\|_2 \leq 1, \quad \|u\|_2 \leq 1, \\ &&& \|u\|_1 \leq c_1, \quad \|v\|_1 \leq c_2, \end{aligned} \quad (4)$$

yields the first pair of sparse canonical correlation loadings. The algorithm to solve this optimization problem uses soft-thresholding, binary search, and finds the second pair of sparse canonical correlation components via the deflation method<sup>62,64</sup>. In our implementation of the sparse CCA method, a pair of tuning parameters ( $c_1, c_2$ ) with values between 0 and 1 were utilized. To select the tuning parameters, the leave-one-out cross-validation was used. Specifically, a grid of dense values from  $(0, 1) \times (0, 1)$  were used as potential values of ( $c_1, c_2$ ) and the absolute value of correlation without penalty was calculated as the criterion function in a cross validation method.

**Linear discriminant analysis (LDA).** We used the edgeR output of genes that were found to be differentially expressed between NSAID and control animals with a false discovery rate (FDR) P value of  $< 0.05$  for feature set identification. Subsequently, we used a previously described method to determine, within the differentially expressed genes, which genes (1-feature) accurately discriminated NSAID-treated animals from control animals as given by the classification error rates<sup>21</sup>. Estimation of the classification error is of critical importance when the number of potential feature sets is large. When sample size is limited, an error estimator may have a large variance and, therefore, may often be low, even if it is approximately unbiased. This can produce many feature sets and classifiers with low error estimates. In this case, the problem was mitigated by applying a bolstered error estimation technique<sup>21,65</sup>. The result of this approach is a list of feature sets that are ranked with respect to their bolstered classification error estimates.

**Availability of Data and Materials.** The datasets utilized for the current study are available via the NCBI bioproject (Accession number PRJNA290483) <http://www.ncbi.nlm.nih.gov/bioproject/>.

## References

- Wolfe, M. M., Lichtenstein, D. R. & Singh, G. Gastrointestinal toxicity of nonsteroidal antiinflammatory drugs. *N Engl J Med* **340**, 1888–1899, <https://doi.org/10.1056/nejm199906173402407> (1999).
- Graham, D. Y., Opekun, A. R., Willingham, F. F. & Qureshi, W. A. Visible small-intestinal mucosal injury in chronic NSAID users. *Clinical gastroenterology and hepatology: the official clinical practice journal of the American Gastroenterological Association* **3**, 55–59 (2005).
- Becker, J. C., Domschke, W. & Pohle, T. Current approaches to prevent NSAID-induced gastropathy—COX selectivity and beyond. *British journal of clinical pharmacology* **58**, 587–600, <https://doi.org/10.1111/j.1365-2125.2004.02198.x> (2004).
- Wallace, J. L. Mechanisms, prevention and clinical implications of nonsteroidal anti-inflammatory drug-enteropathy. *World journal of gastroenterology* **19**, 1861–1876, <https://doi.org/10.3748/wjg.v19.i12.1861> (2013).
- Shaheen, N. J., Straus, W. L. & Sandler, R. S. Chemoprevention of gastrointestinal malignancies with nonsteroidal antiinflammatory drugs. *Cancer* **94**, 950–963 (2002).
- Cohen, N. D., Carter, G. K., Mealey, R. H. & Taylor, T. S. Medical management of right dorsal colitis in 5 horses: a retrospective study (1987–1993). *Journal of veterinary internal medicine* **9**, 272–276 (1995).
- McConnico, R. S., Morgan, T. W., Williams, C. C., Hubert, J. D. & Moore, R. M. Pathophysiologic effects of phenylbutazone on the right dorsal colon in horses. *American journal of veterinary research* **69**, 1496–1505, <https://doi.org/10.2460/ajvr.69.11.1496> (2008).
- Marshall, J. F. & Blikslager, A. T. The effect of nonsteroidal anti-inflammatory drugs on the equine intestine. *Equine veterinary journal. Supplement*, 140–144, <https://doi.org/10.1111/j.2042-3306.2011.00398.x> (2011).
- Wallace, J. L. NSAID gastropathy and enteropathy: distinct pathogenesis likely necessitates distinct prevention strategies. *Br J Pharmacol* **165**, 67–74, <https://doi.org/10.1111/j.1476-5381.2011.01509.x> (2012).
- Handa, O., Naito, Y., Fukui, A., Omatsu, T. & Yoshikawa, T. The impact of non-steroidal anti-inflammatory drugs on the small intestinal epithelium. *Journal of clinical biochemistry and nutrition* **54**, 2–6, <https://doi.org/10.3164/jcbs.13-84> (2014).
- Santucci, L. *et al.* Pentoxifylline prevents indomethacin induced acute gastric mucosal damage in rats: role of tumour necrosis factor alpha. *Gut* **35**, 909–915 (1994).
- Appleyard, C. B., McCafferty, D. M., Tigley, A. W., Swain, M. G. & Wallace, J. L. Tumour necrosis factor mediation of NSAID-induced gastric damage: role of leukocyte adherence. *The American journal of physiology* **270**, G42–48 (1996).
- Watanabe, T. *et al.* Non-steroidal anti-inflammatory drug-induced small intestinal damage is Toll-like receptor 4 dependent. *Gut* **57**, 181–187, <https://doi.org/10.1136/gut.2007.125963> (2008).
- Wallace, J. L. *et al.* Proton pump inhibitors exacerbate NSAID-induced small intestinal injury by inducing dysbiosis. *Gastroenterology* **141**, 1314–1322, 1322. e1311–1315, <https://doi.org/10.1053/j.gastro.2011.06.075> (2011).
- Makivuokko, H., Tiitonen, K., Tynkkynen, S., Paulin, L. & Rautonen, N. The effect of age and non-steroidal anti-inflammatory drugs on human intestinal microbiota composition. *The British journal of nutrition* **103**, 227–234, <https://doi.org/10.1017/s0007114509991553> (2010).

16. Hagiwara, M., Kataoka, K., Arimochi, H., Kuwahara, T. & Ohnishi, Y. Role of unbalanced growth of gram-negative bacteria in ileal ulcer formation in rats treated with a nonsteroidal anti-inflammatory drug. *The Journal of medical investigation: JMI* **51**, 43–51 (2004).
17. Uejima, M., Kinouchi, T., Kataoka, K., Hiraoka, I. & Ohnishi, Y. Role of intestinal bacteria in ileal ulcer formation in rats treated with a nonsteroidal anti-inflammatory drug. *Microbiology and immunology* **40**, 553–560 (1996).
18. Koga, H., Aoyagi, K., Matsumoto, T., Iida, M. & Fujishima, M. Experimental enteropathy in athymic and euthymic rats: synergistic role of lipopolysaccharide and indomethacin. *The American journal of physiology* **276**, G576–582 (1999).
19. Potten, C. S., Schofield, R. & Lajtha, L. G. A comparison of cell replacement in bone marrow, testis and three regions of surface epithelium. *Biochimica et biophysica acta* **560**, 281–299 (1979).
20. Chapkin, R. S. *et al.* Dietary fiber differentially alters cellular fatty acid-binding protein expression in exfoliated colonocytes during tumor development. *Nutrition and cancer* **32**, 107–112, <https://doi.org/10.1080/01635589809514727> (1998).
21. Chapkin, R. S. *et al.* Noninvasive stool-based detection of infant gastrointestinal development using gene expression profiles from exfoliated epithelial cells. *American journal of physiology. Gastrointestinal and liver physiology* **298**, G582–589, <https://doi.org/10.1152/ajpgi.00004.2010> (2010).
22. Davidson, L. A. *et al.* Non-invasive detection of fecal protein kinase C betaII and zeta messenger RNA: putative biomarkers for colon cancer. *Carcinogenesis* **19**, 253–257 (1998).
23. Schwartz, S. *et al.* A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in immune response. *Genome biology* **13**, r32, <https://doi.org/10.1186/gb-2012-13-4-r32> (2012).
24. Knight, J. M. *et al.* Non-invasive analysis of intestinal development in preterm and term infants using RNA-Sequencing. *Scientific reports* **4**, 5453, <https://doi.org/10.1038/srep05453> (2014).
25. Eisenberg, E. & Levanon, E. Y. Human housekeeping genes are compact. *Trends in Genetics* **19**, 362–365, [https://doi.org/10.1016/S0168-9525\(03\)00140-9](https://doi.org/10.1016/S0168-9525(03)00140-9).
26. Palmer, C., Diehn, M., Alizadeh, A. A. & Brown, P. O. Cell-type specific gene expression profiles of leukocytes in human peripheral blood. *BMC Genomics* **7**, 115, <https://doi.org/10.1186/1471-2164-7-115> (2006).
27. Munoz, J. *et al.* The Lgr5 intestinal stem cell signature: robust expression of proposed quiescent ‘+4’ cell markers. *The EMBO journal* **31**, 3079–3091, <https://doi.org/10.1038/emboj.2012.166> (2012).
28. Li, N. *et al.* Single-Cell Analysis of Proxy Reporter Allele-Marked Epithelial Cells Establishes Intestinal Stem Cell Hierarchy. *Stem Cell Reports* **3**, 876–891, <https://doi.org/10.1016/j.stemcr.2014.09.011> (2014).
29. Fevr, T., Robine, S., Louvard, D. & Huelsken, J. Wnt/beta-catenin is essential for intestinal homeostasis and maintenance of intestinal stem cells. *Molecular and Cellular Biology* **27**, 7551–7559, <https://doi.org/10.1128/mcb.01034-07> (2007).
30. Grun, D. *et al.* Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* **525**, 251–255, <https://doi.org/10.1038/nature14966> (2015).
31. Rothenberg, M. E. *et al.* Identification of a cKit(+) colonic crypt base secretory cell that supports Lgr5(+) stem cells in mice. *Gastroenterology* **142**, 1195–1205.e1196, <https://doi.org/10.1053/j.gastro.2012.02.006> (2012).
32. Dalerba, P. *et al.* Single-cell dissection of transcriptional heterogeneity in human colon tumors. *Nature biotechnology* **29**, 1120–1127, <https://doi.org/10.1038/nbt.2038> (2011).
33. Trzpis, M., McLaughlin, P. M. J., de Leij, L. & Harmsen, M. C. Epithelial Cell Adhesion Molecule: More than a Carcinoma Marker and Adhesion Molecule. *The American journal of pathology* **171**, 386–395, <https://doi.org/10.2353/ajpath.2007.070152> (2007).
34. el Marjou, F. *et al.* Tissue-specific and inducible Cre-mediated recombination in the gut epithelium. *Genesis (New York, N. Y.: 2000)* **39**, 186–193, <https://doi.org/10.1002/gene.20042> (2004).
35. Chi, J.-T. *et al.* Gene Expression Programs of Human Smooth Muscle Cells: Tissue-Specific Differentiation and Prognostic Significance in Breast Cancers. *PLOS Genetics* **3**, e164, <https://doi.org/10.1371/journal.pgen.0030164> (2007).
36. Mariadason, J. M. *et al.* Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. *Gastroenterology* **128**, 1081–1088, <https://doi.org/10.1053/j.gastro.2005.01.054> (2005).
37. Georgijevic, S. *et al.* Spatiotemporal expression of smooth muscle markers in developing zebrafish gut. *Developmental dynamics: an official publication of the American Association of Anatomists* **236**, 1623–1632, <https://doi.org/10.1002/dvdy.21165> (2007).
38. Khazen, W. *et al.* Expression of macrophage-selective markers in human and rodent adipocytes. *FEBS Letters* **579**, 5631–5634, <https://doi.org/10.1016/j.febslet.2005.09.032> (2005).
39. Witten, D. M. & Tibshirani, R. J. Extensions of sparse canonical correlation analysis with applications to genomic data. *Statistical applications in genetics and molecular biology* **8**, Article 28, <https://doi.org/10.2202/1544-6115.1470> (2009).
40. Narimatsu, K. *et al.* Toll-like receptor (TLR) 2 agonists ameliorate indomethacin-induced murine ileitis by suppressing the TLR4 signaling. *Journal of gastroenterology and hepatology* **30**, 1610–1617, <https://doi.org/10.1111/jgh.12980> (2015).
41. Watanabe, T. *et al.* Anti-tumour necrosis factor agents reduce non-steroidal anti-inflammatory drug-induced small bowel injury in rheumatoid arthritis patients. *Gut* **63**, 409–414, <https://doi.org/10.1136/gutjnl-2013-304713> (2014).
42. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics (Oxford, England)* **26**, 139–140, <https://doi.org/10.1093/bioinformatics/btp616> (2010).
43. Whitfield-Cargile, C. M. *et al.* The microbiota-derived metabolite indole decreases mucosal inflammation and injury in a murine model of NSAID enteropathy. *Gut microbes* **7**, 246–261, <https://doi.org/10.1080/19490976.2016.1156827> (2016).
44. Ananthakrishnan, A. N. *et al.* Aspirin, nonsteroidal anti-inflammatory drug use, and risk for Crohn disease and ulcerative colitis: a cohort study. *Annals of internal medicine* **156**, 350–359, <https://doi.org/10.7326/0003-4819-156-5-201203060-00007> (2012).
45. Imaeda, H. *et al.* Terminal-restriction fragment length polymorphism (T-RFLP) analysis for changes in the gut microbiota profiles of indomethacin- and rebamipide-treated mice. *Digestion* **86**, 250–257, <https://doi.org/10.1159/000341508> (2012).
46. Fukumoto, K. *et al.* Role of tumor necrosis factor-alpha in the pathogenesis of indomethacin-induced small intestinal injury in mice. *International journal of molecular medicine* **27**, 353–359, <https://doi.org/10.3892/ijmm.2011.602> (2011).
47. DiLauro, S. & Crum-Cianflone, N. F. Ileitis: When It Is Not Crohn’s Disease. *Current gastroenterology reports* **12**, 249–258, <https://doi.org/10.1007/s11894-010-0112-5> (2010).
48. Kwo, P. Y. & Tremaine, W. J. Nonsteroidal anti-inflammatory drug-induced enteropathy: case discussion and review of the literature. *Mayo Clinic proceedings* **70**, 55–61, [https://doi.org/10.1016/s0025-6196\(11\)64666-1](https://doi.org/10.1016/s0025-6196(11)64666-1) (1995).
49. Bjarnason, I. *et al.* Nonsteroidal anti-inflammatory drug-induced intestinal inflammation in humans. *Gastroenterology* **93**, 480–489 (1987).
50. Bertrand, V. *et al.* Increase in tumor necrosis factor-alpha production linked to the toxicity of indomethacin for the rat small intestine. *British journal of pharmacology* **124**, 1385–1394, <https://doi.org/10.1038/sj.bjp.0701968> (1998).
51. Singh, D. P., Borse, S. P. & Nivsarkar, M. Clinical importance of nonsteroidal anti-inflammatory drug enteropathy: the relevance of tumor necrosis factor as a promising target. *Translational Research* **175**, 76–91, <https://doi.org/10.1016/j.trsl.2016.03.014> (2016).
52. Qualls, J. E., Kaplan, A. M., van Rooijen, N. & Cohen, D. A. Suppression of experimental colitis by intestinal mononuclear phagocytes. *Journal of leukocyte biology* **80**, 802–815, <https://doi.org/10.1189/jlb.1205734> (2006).
53. Nadatani, Y. *et al.* High mobility group box 1 promotes small intestinal damage induced by nonsteroidal anti-inflammatory drugs through Toll-like receptor 4. *The American journal of pathology* **181**, 98–110, <https://doi.org/10.1016/j.ajpath.2012.03.039> (2012).
54. Higashimori, A. *et al.* Mechanisms of NLRP3 inflammasome activation and its role in NSAID-induced enteropathy. *Mucosal immunology* **9**, 659–668, <https://doi.org/10.1038/mi.2015.89> (2016).



55. Zeino, Z., Sisson, G. & Bjarnason, I. Adverse effects of drugs on small intestine and colon. Best practice & research. *Clinical gastroenterology* **24**, 133–141, <https://doi.org/10.1016/j.bpg.2010.02.008> (2010).
56. Jia, Q. *et al.* Reduced colitis-associated colon cancer in Fat-1 (n-3 fatty acid desaturase) transgenic mice. *Cancer research* **68**, 3985–3991, <https://doi.org/10.1158/0008-5472.can-07-6251> (2008).
57. Davidson, L. A., Lupton, J. R., Miskovsky, E., Fields, A. P. & Chapkin, R. S. Quantification of human intestinal gene expression profiles using exfoliated colonocytes: a pilot study. *Biomarkers: biochemical indicators of exposure, response, and susceptibility to chemicals* **8**, 51–61, <https://doi.org/10.1080/1354750021000042268> (2003).
58. Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* **29**, 15–21, <https://doi.org/10.1093/bioinformatics/bts635> (2013).
59. Zhang, Z. H. *et al.* A Comparative Study of Techniques for Differential Expression Analysis on RNA-Seq Data. *PLoS one* **9**, e103207, <https://doi.org/10.1371/journal.pone.0103207> (2014).
60. McQueen, C. M. *et al.* TRPM2 SNP genotype previously associated with susceptibility to *Rhodococcus equi* pneumonia in Quarter Horse foals displays differential gene expression identified using RNA-Seq. *BMC Genomics* **17**, <https://doi.org/10.1186/s12864-016-3345-3> (2016).
61. McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor RNA-Seq experiments with respect to biological variation. *Nucleic acids research* **40**, 4288–4297, <https://doi.org/10.1093/nar/gks042> (2012).
62. Witten, D. M., Tibshirani, R. & Hastie, T. A penalized matrix decomposition, with applications to sparse principal components and canonical correlation analysis. *Biostatistics (Oxford, England)* **10**, 515–534, <https://doi.org/10.1093/biostatistics/kxp008> (2009).
63. Rouse, M., Singh, N. P., Nagarkatti, P. S. & Nagarkatti, M. Indoles mitigate the development of experimental autoimmune encephalomyelitis by induction of reciprocal differentiation of regulatory T cells and Th17 cells. *British journal of pharmacology* **169**, 1305–1321, <https://doi.org/10.1111/bph.12205> (2013).
64. Warde-Farley, D. *et al.* The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic acids research* **38**, W214–220, <https://doi.org/10.1093/nar/gkq537> (2010).
65. Braga-Neto, U. & Dougherty, E. Bolstered error estimation. *Pattern Recognition* **37**, 1267–1281, <https://doi.org/10.1016/j.patcog.2003.08.017> (2004).

## Acknowledgements

The authors acknowledge Texas A&M Institute for Genome Sciences and Society (TIGSS) for providing computational resources and systems administration support for the TIGSS HPC Cluster. This work was funded in part by grants from the National Institutes of Health R35CA197707 and P30ES023512, by the Link Equine Research Endowment, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, and by a grant to Dr. Whitfield from the College of Veterinary Medicine & Biomedical Sciences, Texas A&M University. Dr. Whitfield-Cargile was supported by a post-residency doctoral training fellowship from the Department of Large Animal Clinical Sciences, Texas A&M University. Kejun He was supported by an Interdisciplinary Seed Grant for Strategic Initiatives, Texas A&M University.

## Author Contributions

C.W.C., N.D.C., I.I., B.R.W., L.A.D., R.S.C. contributed to the conception, design, analysis, and interpretation of data. J.S.G., B.R.W., and A.C.-E. contributed to data acquisition and analysis. K.H. contributed to data analysis and interpretation of data. C.W.C., N.D.C., K.H., A.C.-E., L.A.D., and R.S.C. were integral in writing the manuscript. All authors have read and approved the final manuscript.

## Additional Information

**Supplementary information** accompanies this paper at <https://doi.org/10.1038/s41598-017-13999-5>.

**Competing Interests:** The authors declare that they have no competing interests.

**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 license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license 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 license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2017