AMERICAN JOURNAL OF PHYSIOLOGY
GASTROINTESTINAL AND LIVER PHYSIOLOGY

Am J Physiol Gastrointest Liver Physiol 321: G41–G51, 2021.
First published May 5, 2021; doi:10.1152/ajpgi.00017.2021

RESEARCH ARTICLE

Translational Physiology

Exfoliated epithelial cell transcriptome reflects both small and large intestinal cell signatures in piglets

Grace Yoon,[1] Laurie A. Davidson,[2,3] Jennifer S. Goldsby,[2,3] Destiny A. Mullens,[3,4] Ivan Ivanov,[4] Sharon M. Donovan,[5] and Robert S. Chapkin[2,3]

[1]*Department of Statistics, Texas A&M University, College Station, Texas;* [2]*Department of Nutrition, Texas A&M University, College Station Texas;* [3]*Program in Integrative Nutrition & Complex Diseases, Texas A&M University, College Station, Texas;* [4]*Department of Veterinary Pathobiology, Texas A&M University, College Station, Texas; and* [5]*Department of Food Science and Human Nutrition, University of Illinois, Urbana-Champaign, Illinois*

Abstract

Assessing intestinal development and host-microbe interactions in healthy human infants requires noninvasive approaches. We have shown that the transcriptome of exfoliated epithelial cells in feces can differentiate breast-fed and formula-fed infants and term and preterm infants. However, it is not fully understood which regions of the intestine that the exfoliated cells represent. Herein, the transcriptional profiles of exfoliated cells with that of the ileal and colonic mucosa were compared. We hypothesized that exfoliated cells in the distal colon would reflect mucosal signatures of more proximal regions of the gut. Two-day-old piglets (n = 8) were fed formulas for 20 days. Luminal contents and mucosa were collected from ileum (IL), ascending colon (AC), and descending (DC) colon, and mRNA was extracted and sequenced. On average, ~13,000 genes were mapped in mucosal tissues and ~10,000 in luminal contents. The intersection of detected genes between three mucosa regions and DC exfoliome indicated an approximately 99% overlap. On average, 49% of the genes in IL, AC, and DC mucosa were present in the AC and DC exfoliome. Genes expressed predominantly in specific anatomic sites (stomach, pancreas, small intestine, colon) were detectable in exfoliated cells. In addition, gene markers for all intestinal epithelial cell types were expressed in the exfoliome representing a diverse array of cell types arising from both the small and large intestine. Genes were mapped to nutrient absorption and transport and immune function. Thus, the exfoliome represents a robust reservoir of information in which to assess intestinal development and responses to dietary interventions.

NEW & NOTEWORTHY The transcriptome of exfoliated epithelial cells in stool contain gene signatures from both small and large intestinal mucosa affording a noninvasive approach to assess gut health and function.

colon; ileum; intestine; transcriptome

INTRODUCTION

The early postnatal period represents a pivotal opportunity within the first 1,000 days of life to influence both short- and long-term health outcomes (1). In the past decade, how early nutrition regulates the development of the gut microbiome and the subsequent programming of infant growth (2, 3), and intestinal (4), immune (5, 6), and cognitive (7) development have been a focus of both basic and clinical research. Large-scale studies have shown that receipt of human milk is the most significant factor associated with microbiome structure (8) due to the presence of prebiotics (e.g., human milk oligosaccharides), microbes, and bioactive proteins and lipids (9, 10).

This key role of human milk in early infant development has led to the recent call for "timely investments in research designed to clarify the operations and biological effects of the mother-breastmilk-infant 'triad' and their translation to public health" (11). To gain insight into this triad, it is essential to understand the molecular crosstalk between the host and the microbiota. This requires suitable noninvasive approaches that can be applied longitudinally to healthy infants (12). To address this challenge, we have developed an approach to interrogate the intestinal exfoliome using infant stool (13) to examine the transgenomic crosstalk between the microbial metagenome and cells of the infant intestine (14). Using this approach, we have demonstrated that differences in intestinal gene expression exist between breast and formula-fed infants (13) and term and preterm infants (15) and have identified potential nutritionally- and developmentally regulated biomarker genes. In addition, integrating the host

Correspondence: R. S. Chapkin (r-chapkin@tamu.edu); S. M. Donovan (sdonovan@illinois.edu).
Submitted 15 January 2021 / Revised 9 April 2021 / Accepted 29 April 2021

transcriptome and bacterial metagenome through novel multivariate statistical analyses showed that the gut microbiota metagenome virulence characteristics concurrently varied with immunity-related gene expression in epithelial cells between the formula-fed and the breast-fed infants (14).

Exfoliated epithelial cell gene signatures can be derived from both the small and large intestine (13), however, definitive sources of these cells have not been defined to date. A recent study conducted by our laboratory assessed the interrelatedness of tissue (invasive) and stool (noninvasive) datasets from both the small intestinal mucosa and colonic mucosa of healthy control mice or those exhibiting nonsteroidal anti-inflammatory drug (NSAID)-induced enteropathy (16). The exfoliated cell transcriptome closely mirrored the transcriptome of the small intestinal mucosa, suggesting that the exfoliome may serve as a noninvasive means of detecting and monitoring mucosal inflammatory diseases that are specific to the small intestine (16). Similar data are not available for healthy neonates, thus, the goal of this study was to investigate the transcriptional profile of exfoliated cells in the fecal stream with that of the underlying mucosa collected from the ileum and colon. We hypothesized that exfoliated cells in the distal colon would reflect mucosal signatures of more proximal regions of the gut. These studies were conducted in the piglet, as it is considered the best model for human infant intestinal, immune, and cognitive development (3, 17, 18).

MATERIALS AND METHODS

Animal Experiment

All animal care and experimental procedures were in accordance with the National Research Council Guide for Care and Use of Laboratory Animals and approved by the University of Illinois at Urbana-Champaign Institutional Animal Care and Use Committee. Intact male pigs (1050 Cambro genetics; $n = 8$) were naturally farrowed and allowed colostrum consumption for up to 48 h before transport to the animal facilities on campus. Upon arrival, pigs were administered a subcutaneous injection of *Clostridium perfingens* antitoxin C + D (5.0 mL) (Colorado Serum Company, Denver, CO). Pigs were artificially reared from *postnatal day* (*PND*) *2* until *PND 22* and housed in caging units that contained six individual stainless-steel cages (length × width × height of 87.6 cm × 88.9 cm × 50.8 cm) with clear, polycarbonate facades on three sides of the cage and vinyl-coated, expanded-metal flooring (Tenderfoot, Minneapolis, MN). All pigs were reared in the same room with ambient temperature maintained between 27 and 29°C and a 12-h light/dark cycle maintained from 0600 to 1800 h. Pigs were observed twice daily at ~0800 and 1600 h and given health scores to track any weight loss, vomiting, diarrhea, or lethargic behavior.

Pigs were fed a nutritionally complete milk replacer reconstituted at 183 g of dry powder in 1 L of water (Nutrastart Liqui-Wean, Milk Specialties Global, Eden Prairie, MN). Pigs received ~0.5 L of formula on the day of arrival to allow for adjustment to the milk replacer and were subsequently fed at a rate of 285 mL and 325 mL/kg body wt from *PND 3–6* and *PND 7–22*, respectively. Individual pig body weights were recorded daily to determine the volume of milk to be dispensed to individual animals throughout the day. Formula was administered 10 times/day, between 1000 and 0400 h using an automated feeding system.

Sample Collection

At study conclusion (*PND 22*), piglets were initially anesthetized with an intramuscular injection of Telazol (Tiletamine HCl and Zolazepam HCl, 3.5 mg/kg body wt each, Pfizer Animal Health, Fort Dodge, IA), after which they were euthanized by an intracardiac injection of 72 mg/kg body wt pentobarbital sodium (Fatal Plus, Vortech Pharmaceuticals, Dearborn, MI). The small intestine was excised from the pyloric sphincter to the ileocecal junction, and its length measured. The intestine was divided into three regions, based on length: duodenum (0%–10%), jejunum (11%–74%), and ileum (IL; final 75%–100%). The large intestine was separated into cecum and colon at the cecocolic junction. The colon was further divided into the ascending (AC) and transverse and descending colon (DC) (19). Contents (~1.0–2.0 g) were collected from the IL, AC, and DC and immediately homogenized in 2× volume of Ambion Denaturation Solution (Millipore Sigma, St Louis, MO). Segments (~10 cm) of IL, AC, and DC were opened longitudinally and mucosa (0.5–1.0 g) scraped with a sterile microscope slide and flash frozen in liquid nitrogen. Samples were stored at −80°C and then shipped on dry ice to Texas A&M University for RNA isolation, library preparation, and for RNA sequencing.

RNA Isolation from Mucosa and Intestinal Contents and Library Construction

For intestinal mucosa samples, RNA was extracted from the scraped mucosa of the IL, AC, and DC using an RNAqueous kit (Life Technologies, Carlsbad, CA) following the manufacturer's protocol, followed by DNase treatment with DNAFree (Life Technologies). RNA was quantified by Nanodrop spectrophotometer (Thermo Fisher Scientific, Waltham, MA), and quality determined using the Nano6000 chip on a Bioanalyzer 2100 (Agilent Technologies, Santa Clara, CA). Sequencing libraries were generated using 250 ng of RNA and the TruSeq RNA Sample Preparation kit (Illumina) as per the manufacturer's protocol. Sequencing libraries were prepared using 250 ng of RNA and the TruSeq RNA Sample Preparation Kit (Illumina) as previously described (16, 20). Only tissue samples with a RNA integrity number (RIN) > 8 were sequenced.

For AC and DC contents (host exfoliome), polyA$^+$ RNA was isolated as previously described (16, 20). Briefly, RNA was extracted using a commercially available kit (Active Motif, Carslbad, CA), quantified by Nanodrop, and quality assessed via Bioanalyzer 2100. Each sample (30 ng) was processed with the NuGen Ovation 3′-DGE kit (San Carlos, CA) to convert RNA into cDNA, followed by the NuGen Encore Rapid Library kit. Briefly, 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 NEBNext Library Quant kit for Illumina (NEB, Ipswich, MA) and assayed on an Agilent DNA High Sensitivity Chip to confirm sizing and exclusion of

adapter dimers. In the absence of a poly A^+ region, bacterial mRNA was not reverse transcribed (16, 20).

Sequencing Pipelines and Data Analysis

Samples from AC and DC mucosa and corresponding AC and DC stool contents were sequenced using the Illumina HiSeq 2500 platform (Illumina, San Diego, CA) at 50 bases single end. The IL mucosa was sequenced on the NextSeq 500 platform (Illumina, San Diego, CA) at 75 bases single end. Sequencing data were examined for quality control with FastQC. RNA sequence reads were mapped using STAR (Spliced Transcript Alignment to a Reference) aligner 2.4.0j (21) in combination with a customized Ensembl reference Sscrofa 11.1.91. Mucosal samples were mapped using default parameters. To determine optimal mapping parameters for the intestinal exfoliome, stool samples were compared after mapping using various Lread settings and trimming methods. Read lengths (Lread) settings of 0.30, 0.35, 0.40, 0.45, 0.50, 0.55, and 0.60 were assessed relative to a default setting of 0.66. Trimming was performed using cutadapt version 1.17 (22) with Python 2.7 to remove Illumina adaptor sequence (agatcggaagagcacacgtctgaactccagtca) along with bases exhibiting a quality score below Q26 only. To assess how trimming removed both Illumina adaptors, the quality score was compared with untrimmed samples. All trimmed outputs were required to have a minimum of 18 bases. Overall, the trimming had little effect on the mapping results. Thus, an Lread value of 0.45 was chosen since it resulted in an increased number of unique annotated reads over other values (Supplemental Fig. S6; see https://doi.org/10.6084/m9.figshare.13473798). Reads were quantified using HTSeq-count 0.11.1 (23). The datasets utilized for the current study are available via the NCBI bioproject (Accession No. PRJNA626016) http://www.ncbi.nlm.nih.gov/bioproject/. Statistical analysis was performed using R (v. 3.6.3) statistical software. The Integrative Genome Viewer (IGV) version 2.8.4 (24, 25) was used to generate images of the read mappings to protein coding genes of the *Sus scrofa* 11.1 genome downloaded from Ensembl.org. For the purpose of comparing gene expression in epithelial tissue and the exfoliome in AC and DC contents, biomarkers for each of the cell types and tissues were selected as previously described (16, 20, 26).

Pathway Analysis

To investigate potential relationships among the expressed genes, an analysis of upstream regulators was performed using the Ingenuity Pathway Analysis program (Qiagen, Redwood City, CA). This in silico analysis is based on prior knowledge of expected effects between transcriptional regulators and their target genes stored in the Ingenuity Knowledge Base as we have previously described (16, 20).

RESULTS

RNA Quality from Exfoliome versus Tissue

RNA quality was assessed via Bioanalyzer for both mucosa and intestinal stool contents. The mean RNA integrity number (RIN) from mucosa was 9.3 (range, 7.9–10) and 7.4 (range, 5.7–8.7) from AC and DC stool contents. RNA samples from IL contents had a mean RIN <5 and were therefore excluded from further analysis. Representative bioanalyzer traces from RNA isolated from piglet mucosa (18S and 28S rRNA) and AC and DC contents (16S and 23S rRNA) are shown in Supplemental Fig. S1 (see https://doi.org/10.6084/m9.figshare.13473801). The AC and DC stool content profiles are consistent with our previous findings (16), indicating an abundance of bacterial RNA. Therefore, we included an oligo dT probe in the first step of library construction for the purpose of selectively targeting host (piglet) transcripts for cDNA production and subsequent library construction. In total, 15,523 eukaryotic genes (including protein coding genes, ribosomal RNA, small nucleolar RNA, microRNA and other long non-coding RNAs) were analyzed from IL (n = 7), AC (n = 8), and DC (n = 7–8) samples. A detailed description of the initial and ending read-counts per sample type at major steps along the analytical RNA-Seq data pipeline is shown in Supplemental Fig. S2 (see https://doi.org/10.6084/m9.figshare.13473819).

Comparison of Gene Expression in Intestinal Mucosa and the Exfoliome

Sequencing of the samples revealed that reads for IL mucosa, AC mucosa, DC mucosa, and exfoliated cells from the AC and DC mapped to an average of 12,925, 12,902, 13,018, 10,647, and 10,179 genes per sample, respectively (Table 1).

As previously shown (15), some exfoliated cell transcripts consisted of hundreds of reads that accumulated in a narrow region in or close to the 3′-UTR (e.g., *AQP8, FABP2, HMGCR*). However, full exonic coverage was also observed in select genes (e.g., *RBP2* and *SLC20A1*). Representative genomic datasets are provided in Supplemental Fig. S3 (see https://doi.org/10.6084/m9.figshare.13473816), Supplemental Fig. S4 (see https://doi.

Table 1. *Total number of sequences and assigned genes in mucosa and luminal contents of 22-day-old formula-fed piglets*

	Mucosal Tissue			Luminal Contents (Exfoliome)	
	IL	AC	DC	AC	DC
Sequencing reads*	33,928,174 ± 16,888,897	26,941,595 ± 9,668,556	32,749,474 ± 13,566,353	63,446,871 ± 8,391,442	54,539,162 ± 5,269,020
Reads mapped per sample†	32,270,088 ± 16,003,942	25,673,809 ± 9,409,321	31,175,002 ± 12,926,952	26,295,112 ± 17,378,726	17,338,860 ± 11,885,168
Genes mapped per sample‡	12,925 ± 376	12,902 ± 245	13,018 ± 244	10,647 ± 908	10,179 ± 1,122

Means ± SD (standard deviation). SD was calculated for 7–8 samples for each location and data source. AC, ascending colon; DC, descending colon; IL, ileum. *Number of reads from FASTQ file. †Number of reads mapped per sample represent the number of mapped (at least once and multiple locations not more than 10) to the reference genome. ‡Number of genes mapped per sample represent the number of genes that have at least one raw count.

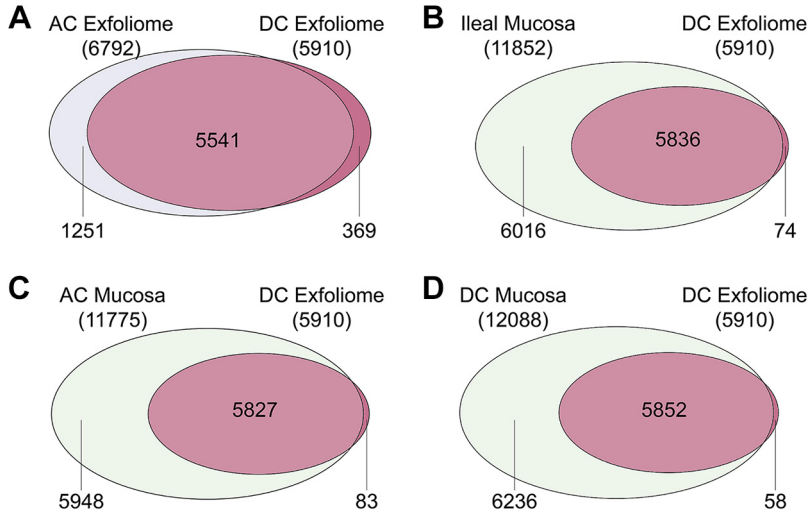

Figure 1. Venn diagrams representing detectable genes in exfoliated cells and each region of the colonic mucosa. Genes having >3 raw counts in >50% of the samples in each group were classified as expressed/detectable. 5,541 genes were detected in exfoliated cells from both ascending colon (AC) and descending colon (DC) (*A*); 5,836 genes were detected in IL mucosa and DC exfoliome (*B*); 5,827 genes were detected in both AC mucosa and DC exfoliome (*C*); and 5,852 genes were detected in both DC mucosa and DC exfoliome (*D*). A total of 10,000 genes were detected in the exfoliome with a minimum cut off of a single read. Of these, only those genes with at least three counts in more than 50% of the samples were utilized for downstream analyses.

org/10.6084/m9.figshare.13473795), and Supplemental Fig. S5 (see https://doi.org/10.6084/m9.figshare.13473804).

Comparisons of global gene signatures of exfoliated cells from the two colonic sites (AC and DC) and the mucosa profiles from three regions (ileum, AC and DC) in the intestine versus exfoliated cells from DC are shown in Fig. 1. Only genes having more than three raw counts in more than half of the samples in each group were classified as expressed/detectable. A total of 5,541 genes were detected in exfoliated cells in both AC and DC, accounting for 81.6% and 93.8% of genes in AC exfoliome and DC exfoliome, respectively (Fig. 1*A*). The intersection of detected genes in the DC exfoliome and three mucosa regions were calculated (Fig. 1, *B*–*D*).

The intersection of detected genes between the three mucosa regions and DC exfoliome indicated an ~99% overlap (Fig. 2). On average, 49% of the genes in each of the tissue mucosa samples were represented in the exfoliome samples. Specifically, 49.2% of genes in the ileal mucosa were represented in DC exfoliome samples, 49.5% of the genes in AC mucosa were represented in DC exfoliome samples, and 48.4% of genes in the DC mucosa were represented in DC exfoliome samples. According to our computational pipeline depicted on Supplemental Fig. S2, we performed a filtering step before examining gene expression of the host exfoliome and mucosal regions. When examining the *Sus scrofa* 11.1 genome, we detected 25,880 unique Ensembl gene IDs. Among these, there were 8,256 gene IDs without a gene name and 2,104 gene IDs where two or more had the same gene name. We attributed these issues to the pig genome not being annotated as well as the human or mouse genomes. As a result of these considerations, 15,523 genes that passed the filtering step were used for the computational pipeline subsequent steps. Data filtering was followed by a normalization step using the upper quartile method. Because the raw data counts in the mucosal samples were markedly higher than those from the exfoliome, we normalized the mucosal data separately from the exfoliome data. The distribution of raw counts and normalized counts per sample are plotted in Supplemental Fig. S7 (https://doi.org/10.6084/m9.figshare.14379014). In the following step, we determined that 5,789 out of those 15,523 genes were simultaneously detected in the DC exfoliome and all mucosa regions. A principal component analysis (PCA) plot using normalized counts of 5,798 genes is shown in Fig. 3. PCA clearly separated mucosa samples from exfoliome samples as well as ileum from colon (AC and DC) samples. In addition, to evaluate correlations between the datasets (IL mucosa, AC mucosa, DC mucosa, AC exfoliome, and DC exfoliome), normalized data were averaged across 7–8 samples per dataset and scatter plots including the estimation of the correlation trends reported in Fig. 4.

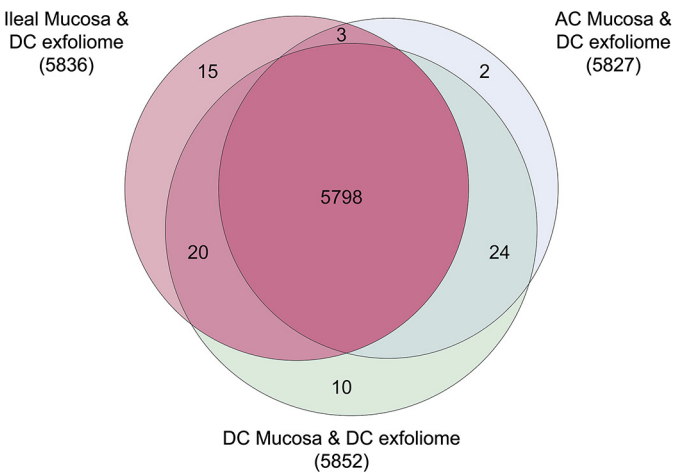

Figure 2. Venn diagrams representing the intersection of detectable genes in each region of colonic mucosa and exfoliated cells from the descending colon (DC). Genes having >3 raw counts in >50% of the samples in each group were classified as expressed/detectable. A total of 5,798 genes were detected in all three intersections, accounting for more than 99% of the genes in each region.

Identification of Anatomical and Cell-Type Source of Transcripts

Next, the anatomical origin of the transcripts expressed in the exfoliome was investigated (Fig. 5). Genes expressed predominantly in specific anatomic regions of the gastrointestinal tract (stomach, pancreas, small intestine, colon) were detectable in exfoliated cells (Fig. 5*A*). In addition, the types

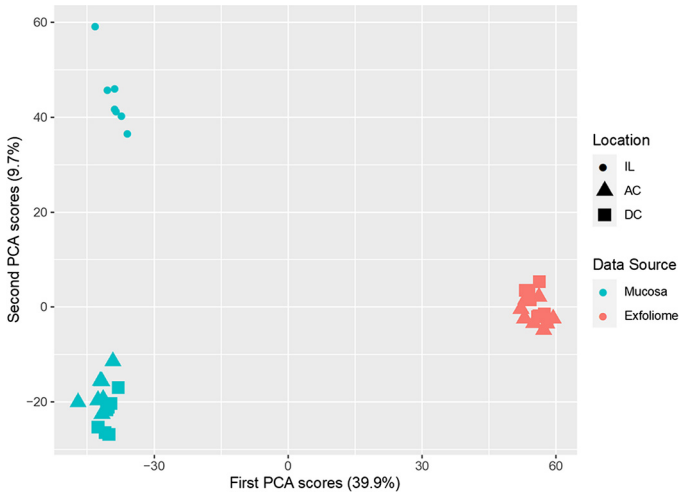

Figure 3. Principal component analysis plot showing visual clustering of ileal mucosa, colon mucosa (ascending colon, AC and descending colon, DC), and exfoliome data (contents from AC and DC). All mucosa and exfoliome data were normalized separately using the upperquartile method for 5,798 genes, with a data dimension of 38 by 5,798. A similar gene expression profile was observed between AC and DC, but was distinct with respect to the ileal location and between the mucosa and exfoliome.

of intestinal cells represented in the exfoliome were identified using marker genes known to be exclusively or highly expressed in certain cell types (stem cells, Goblet cells, etc.). Markers for all intestinal epithelial cell types were expressed in either AC or DC exfoliome, or both (Fig. 5*B*), indicating that the exfoliome represents a diverse array of cell types arising from both the small and large intestine.

Functional Signatures

To investigate the potential functional signatures of mucosal and exfoliated cell transcriptomes, we compared mapped genes in the IL and DC mucosal and DC exfoliome to genes previously identified in human and rodent intestinal cells by Wang et al. (27). Heat maps showing the average counts of each gene class across the three samples (IL and DC mucosa and DC exfoliome) are shown in Fig. 6. Genes associated with lipid, bile acid, vitamin, water, amino acid, inorganic solute, organic solute, sugar, metal ion, and nucleotide absorption are shown in Fig. 6*A*. Among the genes involved in nutrient absorption and transport, only 13% (16 of 121) of the genes queried in IL or DC mucosa were not detected in the exfoliome. In some cases, genes were strongly expressed in all three datasets, including HMG-CoA reductase (*HMGCR*) and aquaporin-8 (*AQP8*). Importantly, several genes that were expressed in the IL mucosa, but not in the DC mucosa, were detected in the exfoliome, including fatty acid binding protein2 (*FABP2*), ATP binding cassette subfamily G member 5 (*ABCG5*), and vanin 1 (*VNN1*).

The average expression level of genes linked to specific hormones produced by enteroendocrine cells is shown in Fig. 6*B*. Expression of the majority (9 of 12) of the queried genes was detected in the exfoliome. Regional differences in mucosal expression were observed for several genes (*CCK* and *GIP*), which were expressed in the IL mucosa, but not DC mucosa or DC exfoliome. Several genes (*INSL5*, *PYY*, and *TAC1*) were most highly expressed in the DC mucosa and were also present in the DC exfoliome. Interestingly, the expression of amphiregulin (*AREG*) was higher in the DC exfoliome relative to mucosal sites. Lastly, only one of the 23 immune-related genes queried was not detected in the DC exfoliome (Fig. 6*C*). A heatmap of signature genes for goblet cells is shown in Supplemental Fig. S8 (see https://doi.org/10.6084/m9.figshare.13473807). Two genes commonly associated with goblet cells, trefoil factor 3 (*TFF3*), and regenerating family member 4 (REG4) were detected in IL and DC mucosa and DC exfoliome in all piglets.

Analysis of Predicted Upstream Regulators

Ingenuity IPA Upstream Regulator Analysis was used to identify the potential upstream regulators responsible for gene expression in the experimental dataset (28). A total of 121 upstream regulators were identified that were enriched in the dataset (*P* value of overlap) and for which a mechanistic network could be formed (Supplemental Table S1; see https://doi.org/10.6084/m9.figshare.14379002). For this process, the number of genes that are downstream from regulators in the network were identified. Since we did not have a treatment comparison, we could not predict whether the upstream regulators were activated or inhibited. In IPA, a mechanistic network, is a plausible set of connected upstream regulators that can work together to elicit the gene expression observed in a dataset. The regulators represented a variety of molecule types, including transcription factors, cytokines, growth factors, enzymes, G-protein coupled receptors, kinases, and micro RNA. The top 15 upstream regulators sorted by numbers of genes in the mechanistic network are listed in Table 2 and their downstream target molecules are shown in Supplemental Table S1. In addition, a number of microRNA (miRNA) were identified (Supplemental Table S1).

Potential regulators included cytokines [tumor necrosis factor (TNF), interleukin (IL)-1B, IL-2, IL-4, and IL-6], hormones and growth factors [insulin-like growth factor-I (IGF-I), epidermal growth factor (EGF), insulin, leptin (LEP), transforming growth factor-B1 (TGFB1), and nuclear receptors for glucocorticoids (NR3C1) and estrogen (ESR1)]. Transcription factors previously associated with the regulation of intestinal gene expression were identified, including signal transducers and activators of transcription (STAT) -1, -3, -5A, and -6, CCAAT enhancer-binding protein (CEBP)-α and -β, and WNT-3a and 4 (Supplemental Table S1).

The IPA Path Designer Graphical Representation was used to generate figures of relationships between the two upstream regulators of genes, TNF and LEP, and their potential down-stream targets (Fig. 7). These were selected based on the statistical strength of the proposed relationships (*P* value of overlap) and richness of the mechanistic networks. Potential cross-talk between TNF and LEP is also apparent as both TNF and LEP are linked to INS (Fig. 7*A*) and LEP is linked to TNF (Fig. 7*B*). In addition, they are linked to other potential upstream regulators identified in Table 2, including IL-6, EGF, IL-1B, NR3C1, and ESR1. In these figures, solid and dashed lines represent stimulatory and inhibitory relationships, respectively. The majority of relationships for TNF and LEP are proposed to be inhibitory, with the exception of LEP signaling to the leptin receptor (LEPR) (Fig. 7*B*).

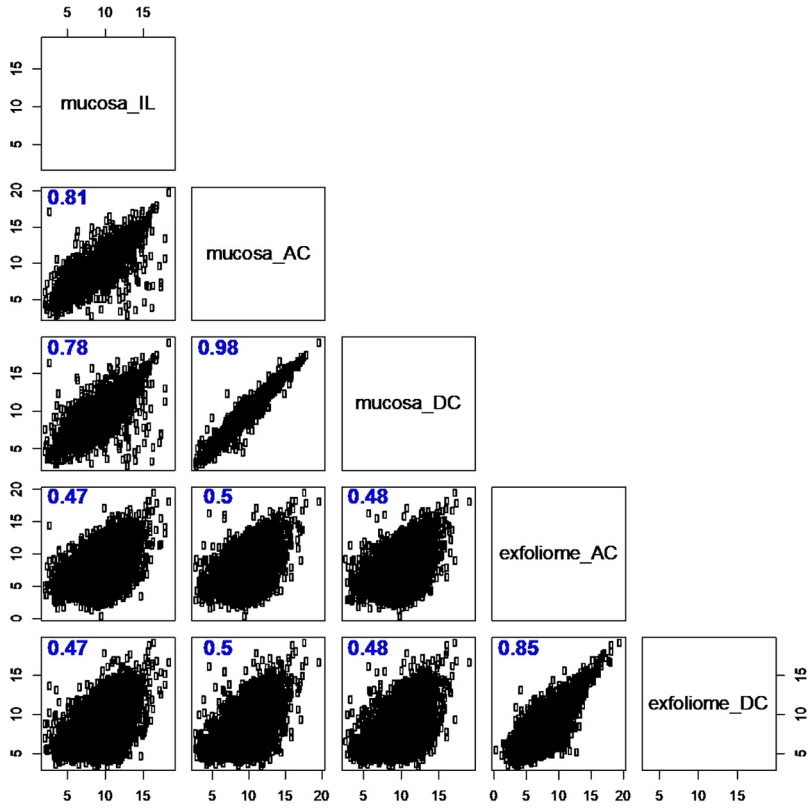

Figure 4. Pairwise scatter plots of log(2) transformed averaged normalized counts across 7–8 samples for each dataset. Three mucosa sources were compared: Ileum (IL), ascending colon (AC) and descending colon (DC), and the exfoliome from AC and DC contents. A strong and significant correlation between mucosa sites and to a lesser extent between mucosa sites and the exfoliome was observed. The log(2) transformed data was calculated as log2(averaged normalized counts + 1).

DISCUSSION

The gastrointestinal tract is a complex organ comprised on numerous cell types that coordinate signals from the environment to the host. For example, epithelial cells form a barrier between the luminal contents and the host and also digest and absorb nutrients. Enteroendocrine cells sense luminal contents and secrete hormones that coordinate digestive processes and communicate with the brain to signal hunger and satiety. Lastly, multiple immune cells types are responsible for responding to signals from commensal and pathogenic microbes and for inducing tolerance to allergens (27). Intestinal growth and functional maturation and mucosal immune development are influenced by diet and signals from the microbiota (6, 10, 12). However, assessing these processes over time, as well as, in responses to dietary, prebiotic or probiotic interventions has been hampered by the absence of sensitive, noninvasive approaches to monitor gut function. This is particularly true for healthy infants, in which sampling of intestinal tissue would be unethical (12). To overcome this challenge, we have demonstrated the feasibility of measuring transcriptomic profiles of intact sloughed epithelial cells in the stool of term (13) and preterm human infants (15) as a noninvasive approach to identify putative molecular biomarkers of dietary and developmental effects on maturation of intestinal function.

A limitation of our prior work was that the precise origin of exfoliated cells was not defined. Based on the presence of mRNA for genes associated with discrete epithelial cell types in the exfoliome, such as lactase and sucrase-isomaltase (absorptive enterocytes), chromogranin A (enteroendocrine cells), and lysozyme (paneth cells), we proposed that the exfoliome contained transcriptomic signatures of both the small and large intestinal cells (13). Further evidence from our laboratory obtained from adult mice (16) and horses (29) indicated that transcriptional signatures from the small intestinal mucosa were present in the fecal exfoliome. Based on these observations, the goal herein was to compare the transcriptome of exfoliated cells present in the AC and DC contents with that of the underlying mucosa obtained from the IL, AC, and DC of piglets. We hypothesized that genes in the transcriptome of exfoliated cells in the DC would include gene signatures of mucosa in more proximal regions of the gut.

Consistent with our previous observations (15, 16), transcripts from the exfoliome accumulated in or close to the 3′-UTR, providing sufficient length reads to identify gene signatures. The average numbers of genes detected in the AC and DC contents (~10,300 genes) was not markedly less than that observed in the IL, AC, and DC mucosa (12,950 genes). Over 5,500 genes were present in both the AC and DC exfoliome, representing 81.6% and 93.8% of total genes in these two regions, respectively. In addition, 49.2%, 49.5%, and 48% of the IL, AC, and DC mucosal genes were detected in the DC exfoliome, supporting our hypothesis that the fecal host transcriptome in the DC carries molecular signatures of cells exfoliated in more proximal regions of the gut.

To further test our hypothesis, we sought to identify the anatomical origin of transcripts present in the exfoliome using marker genes known to be exclusively or highly

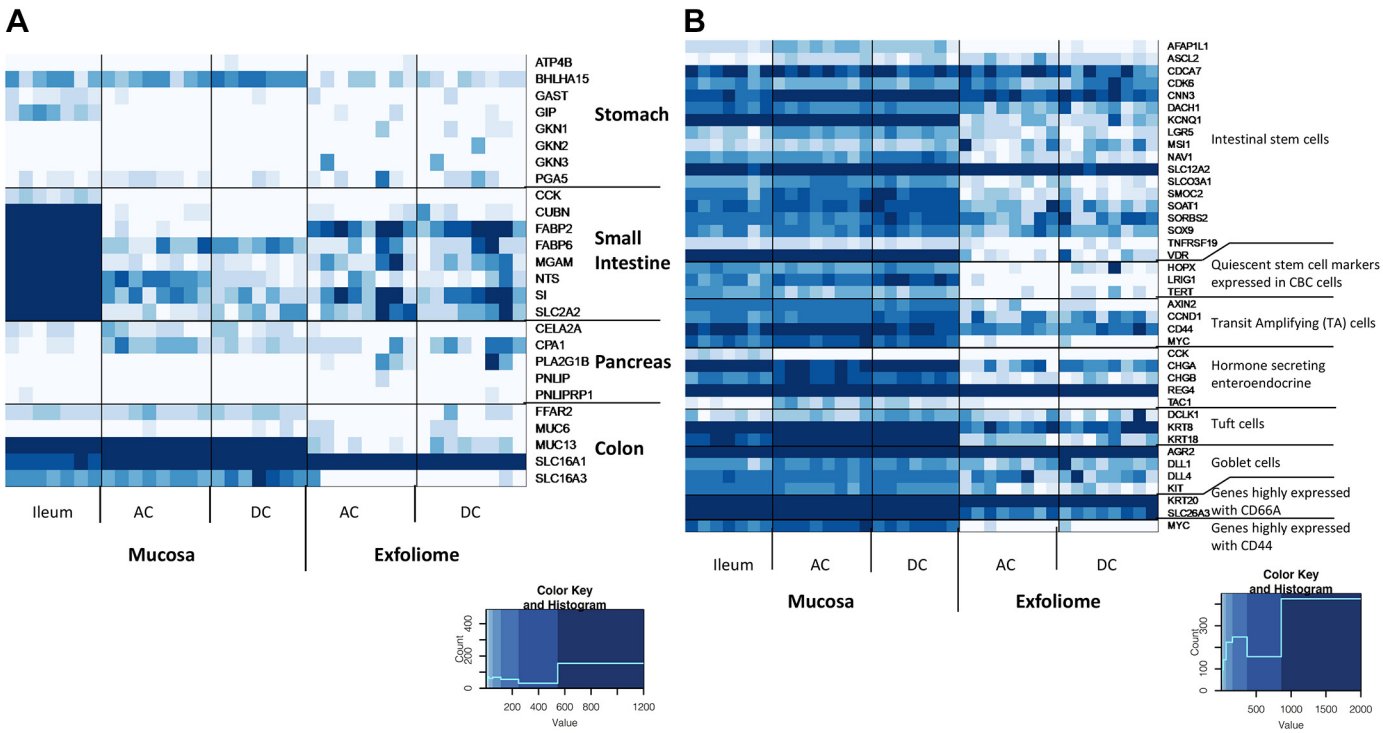

Figure 5. Heatmaps of genes counts from each sample and each data source. The exfoliome signature arises from cells sloughed from both the small and large intestine. Counts were normalized for all tissue and exfoliome data using the upperquartile method. *A*: genes reported to be primarily expressed at specific anatomic locations (stomach, pancreas, small intestine, colon). *B*: cell-specific genes.

expressed by specific cell types. The AC and DC exfoliome signatures mapped predominantly to markers of small intestine and large intestinal cells, with fewer markers from the stomach or pancreas, which is consistent with our previous findings (15, 16). As expected, all genes associated with the small intestine were strongly expressed in the IL mucosa and to a lesser degree in the AC and DC mucosa. A number of the small intestinal genes were detected at higher levels of

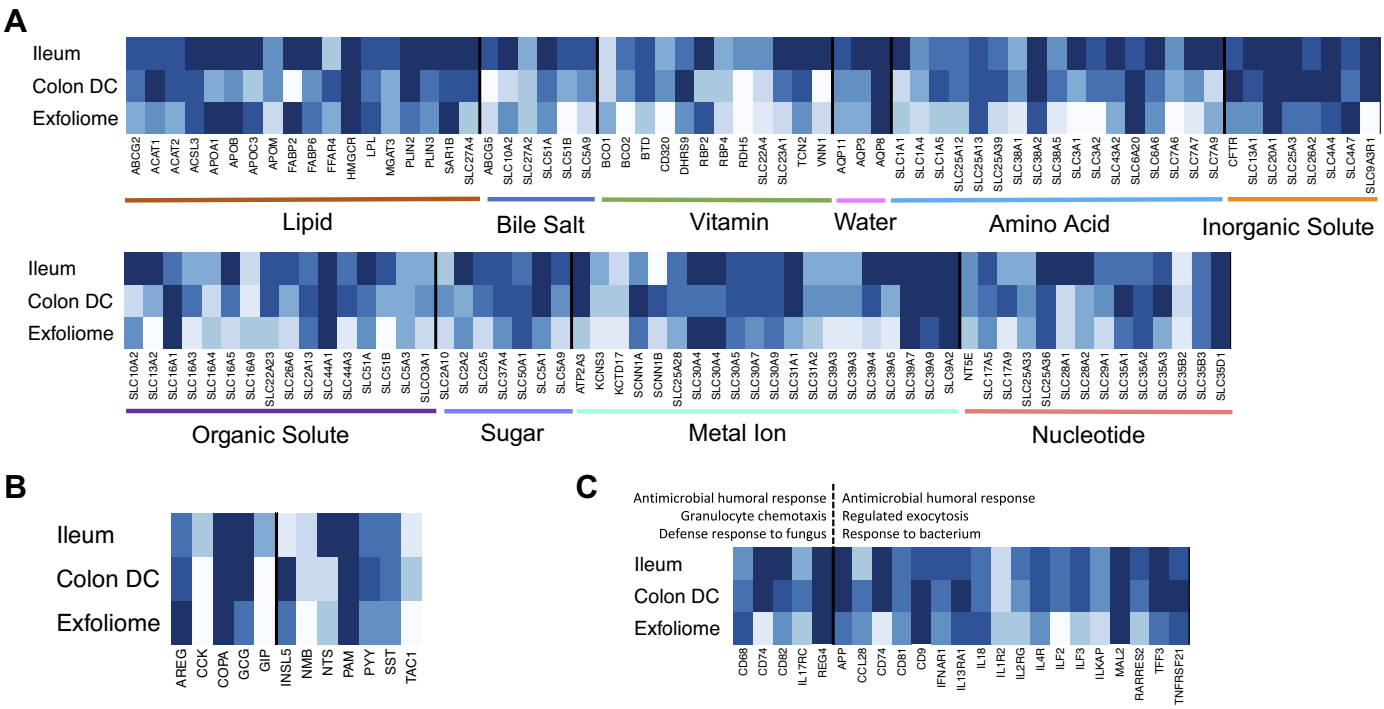

Figure 6. Heatmaps showing the average counts of each gene class across samples from three data sources: two different mucosa locations (Ileum and descending colon, DC) and DC exfoliome. Nutrient absorption and transport genes (*A*); hormone genes (*B*), and immune-related genes (*C*) are presented.

Table 2. *Top 15 upstream regulators of genes detected in mucosa and exfoliated intestinal epithelial cells predicted from mechanistic network analysis conducted in ingenuity pathway analysis*

Upstream Regulator	Molecule Type	*P* Value of Overlap*	Mechanistic Network *A*(*B*)†
TNF	Cytokine	4.39×10^{-09}	126 (21)
IGF1	Growth factor	9.67×10^{-07}	118 (18)
IL-6	Cytokine	3.73×10^{-07}	119 (21)
EGF	Growth factor	5.44×10^{-05}	108 (16)
Insulin	Group	3.24×10^{-6}	106 (19)
IL-1B	Cytokine	5.49×10^{-08}	103 (15)
LEP	Growth factor	2.81×10^{-13}	102 (21)
HNF1A	Transcription regulator	3.58×10^{-07}	100 (18)
P38 MAPK	Group	6.26×10^{-03}	99 (18)
NR3C1	Ligand-dependent nuclear receptor	2.21×10^{-06}	98 (9)
IL-4	Cytokine	4.48×10^{-04}	95 (14)
IL-2	Cytokine	2.82×10^{-04}	94 (18)
TGFB1	Growth factor	3.85×10^{-04}	93 (14)
ESR1	Ligand-dependent nuclear receptor	1.09×10^{-08}	91 (10)

ESR1, estrogen; IGF, insulin-like growth factor-I; IL-1B, interleukin 1B; LEP, leptin; NR3C1, nuclear receptors for glucocorticoids; TGFB1, transforming growth factor-B1; TNF, tumor necrosis factor. *Overlap *P* value indicates whether there is a statistically significant overlap between the dataset genes and the genes regulated by a transcription factor using Fisher's exact test. †Mechanistic network value is a list plausible sets of connected upstream regulators that can work together to elicit the gene expression changes observed in a dataset. The value shown as *A* (*B*) in which *A* is the genes that are downstream from regulators (*B*) in the network. For more information, see: https://qiagen.secure.force.com/KnowledgeBase/KnowledgeIPAPage?id=kA41i000000L5lzCAC.

expression in the AC and DC exfoliome. In particular, FABP2 was strongly expressed in the AC and DC exfoliome, with low or undetectable levels in the AC and DC mucosa. This observation suggests that the source of the signal was generated from cells exfoliated in the IL rather than from the underlying mucosa in the colon. FABPs are the most abundant cytoplasmic proteins in the small intestine and play pivotal roles in intestinal lipid transport. FABP2 (I-FABP) expression is highest in the distal half of small intestine. FABP2 shows a preference for transporting saturated long-chain fatty acids, but targeted gene deletion of the *FABP2* gene in rodents results in sexually dimorphic metabolic disturbances and changes in gene expression, supporting broader roles for this protein (30). During tissue injury, FABP-2 is released and can be measured in the serum and urine and has been proposed as a luminally expressed biomarker of intestinal injury (31, 32). Previous studies have shown that I-FABP may be a valid and promising serologic biomarker for early diagnosis of necrotizing enterocolitis (NEC) in preterm infants. Prospective clinical trials have shown that both plasma and urinary I-FABP levels specifically identify NEC in preterm infants before appearance of diagnostic radiological signs suggestive for NEC (33). Moreover, serial I-FABP measurements accurately predict development of complicated disease (33) and concentrations of I-FABP in plasma and urine at the onset NEC have been correlated with the length of intestinal resection during laparotomy (34). Overall, a recent meta-analysis reported that I-FABP is a valid serologic biomarker for early diagnosis in NEC for premature neonates with a moderate accuracy (35). Although it remains to be demonstrated, we propose that measurement of I-FABP mRNA expression in the fecal exfoliome may be a clinically relevant early biomarker of risk of NEC.

We were also interested in probing for biomarkers of intestinal function in the exfoliome using gene lists generated by Wang et al. (27). These investigators utilized single-cell RNA-seq analyses of 14,537 epithelial cells surgically obtained human ileum, colon, and rectum. We found that 87% of the 121 genes reported to be involved in nutrient absorption and transport were detected in the exfoliome. In addition to FABP-2, other genes were present in the DC exfoliome and IL mucosa, but not DC exfoliome. Furthermore, most of the genes for hormones expressed by enteroendocrine cells (75%) and nearly all (95%) of the immune genes probed were present in the exfoliome. Taken together, the high level of coverage of genes identified in human intestinal samples, the detection of regional differences in mucosal gene expression, and specificity of gene expression in the mucosa versus the exfoliome provides strong evidence that the fecal transcriptome provides insight into transcriptional activity in both the small and large intestine.

Finally, we sought to identify potential upstream regulators responsible for gene expression in the experimental dataset. Over 100 putative upstream regulators were enriched in the dataset and for which a mechanistic network could be formed. As expected, key transcription factors that regulate intestinal gene expression were identified, including STAT-1, -3, -5 A, -5B and -6, CEBP-α, CEBP-β, and WNT-3a and 4 (36). STATs are a family of latent cytoplasmic transcription factors that convey signals from numerous cytokines and growth factors to the nucleus (37). Both STAT1 and STAT3 are important regulators of genes that are involved in cell survival (BCL-x, survivin, caspases) and cell proliferation (c-Myc, p21, cyclin D1), their deregulation significantly impacts the homeostasis of intestinal tissues (38). The C/EBP transcription factor family is involved in intestinal differentiation. Both CEBP-α and -β are localized to intestinal epithelial cells. CEBP-α is mainly found in the crypts, whereas CEBP-β in present in both villi and crypts (39). In rodents, there is a surge of intestinal expression of CEBP-α in the late fetal phase ~24 h before morphological maturation and the onset of expression of numerous epithelial genes (40). However, morphological development, cellular proliferation, or the onset of expression of a cluster of epithelial mRNAs expression were unaffected in CEBP-α null mice (41), suggesting that CEBP-α has no essential role. Since other C/EBP isoforms are present in the developing intestine, it is possible that

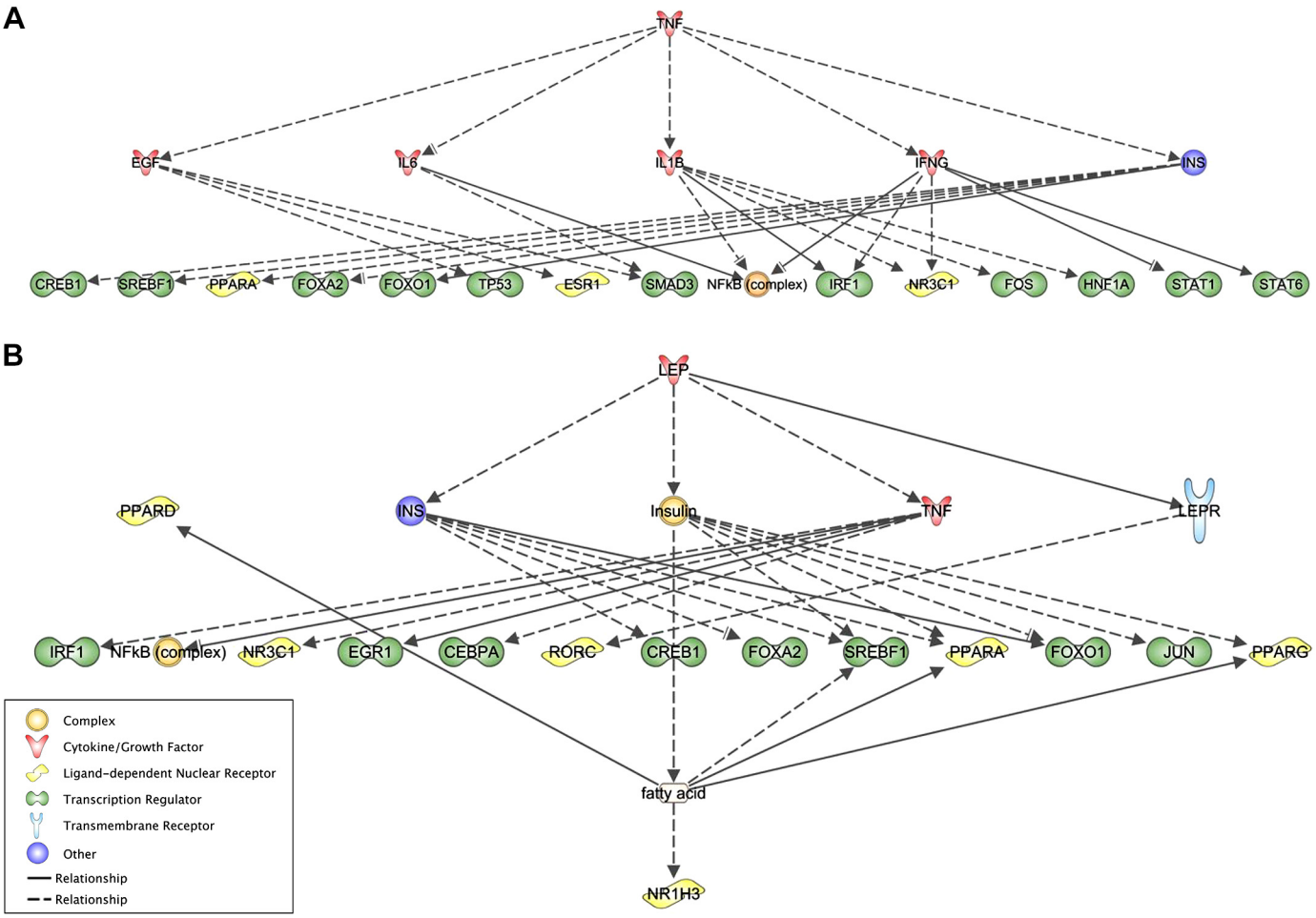

Figure 7. Mechanistic networks of relationships between potential upstream regulators of genes identified in mucosa and exfoliated epithelial cells for tumor necrosis factor (TNF; *A*) and leptin (LEP; *B*). Molecules are represented as nodes, and the biological relationship between two nodes is represented as an edge (line). All edges are supported by at least one reference from the literature or from canonical information stored in the Ingenuity Knowledge Base.

there is a generic requirement for a member of the C/EBP family.

The Wnt/β-catenin pathway plays a crucial role in development and renewal of the intestinal epithelium (36). Mariadason et al. (42) used DNA microarrays to query epithelial cell gene expression along the crypt-villus axis in mice. The set of 1,113 differentially expressed genes was significantly enriched for genes involved in cell cycle progression, RNA processing, and translation (downregulated as cells migrate up the crypt) and genes involved in cytoskeleton assembly and lipid uptake (upregulated during maturation). They confirmed that Wnt signaling was maximal in the proliferative compartment (42). In a subsequent study, the same group reported a significant correlation observed between overall proteomic changes and corresponding gene expression changes along the crypt-villus axis, indicating intestinal cell maturation is primarily regulated at the transcriptional level (43).

Of all of the potential upstream regulators identified, the cytokine TNF-α showed the highest *P* value of overlap and formed the strongest network of connected upstream regulators that can work together to elicit the gene expression changes observed in our dataset, with 126 potential downstream targets. This was somewhat surprising since these samples were obtained from healthy piglets and TNF-α is typically associated with inflammation. However, recent data suggest that TNF participates in pleiotropic activities with implications in various cellular processes, including proliferation and differentiation (44). Moreover, an anti-inflammatory role of TNF, primarily via the induction of local glucocorticoids synthesis in the intestinal epithelium has been reported (45). Thus, the outcome of TNF receptor (TNFR) signaling largely depends on various factors, e.g., the tissue receptor composition TNFR1 versus TNFR2 and the precise cellular context or tissue type, which will determine cellular fate (46). As shown in Supplemental Table S1, some of the downstream targets of TNF-α were related to inflammation and apoptosis (e.g., caspases, cytokines), others were markers of cellular proliferation (e.g., BMP2, SMAD1) and differentiated function, including glucose (SLC2A2), sodium (SCNN1A, SCNN1B), amino acid (SLC1A4), and short-chain fatty acid (SLC16A5) transporters. Thus, additional research in regard to the potential

pleiotropic effects of TNF-α in the developing intestine is warranted.

In conclusion, the intestinal epithelium represents a complex combination of specialized cellular components, structural organization, as well as fine-tuned maintenance and renewal mechanisms that ensure its barrier and absorptive function (47). The well-being of the gut is intrinsically linked to the overall nutritional status and health of the host, and perturbations to this homeostasis can have severe impact on local and systemic health (48). However, the lack of noninvasive approaches to repeatedly access intestinal tissue along the GI tract has hampered our ability to study normal gut development or assess clinical responses to dietary or medical interventions. Exfoliation of ~10 billion intestinal epithelial cells from the villus tips in the small intestine and crypt surface in the colon is an active biochemical process linked to intestinal epithelial homeostasis (29). This vast reservoir of host cells generated by exfoliation has sparked interest from both basic and clinical translational investigators due to their potential utility to noninvasively assess cellular markers of gastrointestinal disease (12). Our data demonstrate for the first time that the complement of exfoliated cells present in stool reflects mRNA signatures of cells originating from both the proximal and distal bowel. In addition, markers of numerous epithelial cell types may provide insight into intestinal function. Application of this technique would enable researchers and clinicians to assess gut gene expression within an individual before initiating an intervention, at the end of the intervention, and following a washout. Our research group has also shown that genomic data from both the host mucosa and gut microbiota can be integrated in order to define host gene-diet interactions within the context of the structure and operations of gut microbial communities (14). In the future, we anticipate that the fecal exfoliome could also be used to personalize nutritional and medical interventions based on an individual's gene signature and microbiome.

SUPPLEMENTAL MATERIALS

The supplemental material can be found at the following sites:
Table S1: https://doi.org/10.6084/m9.figshare.14379002.
Fig. S1: https://doi.org/10.6084/m9.figshare.13473801.
Fig. S2: https://doi.org/10.6084/m9.figshare.13473819.
Fig. S3: https://doi.org/10.6084/m9.figshare.13473816.
Fig. S4: https://doi.org/10.6084/m9.figshare.13473795.
Fig. S4: https://doi.org/10.6084/m9.figshare.13473795.
Fig. S5: https://doi.org/10.6084/m9.figshare.13473804.
Fig. S6: https://doi.org/10.6084/m9.figshare.13473798.
Fig. S7: https://doi.org/10.6084/m9.figshare.14379014.
Fig. S8: https://doi.org/10.6084/m9.figshare.13473807.

DATA AVAILABILITY

The datasets utilized for the current study are available via the NCBI bioproject (Accession No. PRJNA626016) http://www.ncbi.nlm.nih.gov/bioproject/.

ACKNOWLEDGMENTS

The authors thank Dr. Marcia Monaco and Dr. Mei Wang for assistance in conducting the animal study and Rachel Wright for generating the graphical abstract.

GRANTS

This work received support from National Institutes of Health Grants RO1DK107561 (to S. M. Donovan), R35CA197707 (to R. S. Chapkin), and T32CA090301 (to G. Yoon, R. S. Chapkin), the Allen Endowed Chair in Nutrition & Chronic Disease Prevention (to R. S. Chapkin), and the Hagler Institute for Advanced Study at Texas A&M University (to S. M. Donovan).

DISCLOSURES

No conflicts of interest, financial or otherwise, are declared by the authors.

AUTHOR CONTRIBUTIONS

I.I., S.M.D., and R.S.C. conceived and designed research; G.Y., L.A.D., J.S.G., and D.A.M. performed experiments; G.Y., L.A.D., J.S.G., D.A.M., and I.I. analyzed data; G.Y., L.A.D., I.I., S.M.D., and R.S.C. interpreted results of experiments; G.Y. and D.A.M. prepared figures; G.Y. drafted manuscript; G.Y., L.A.D., J.S.G., D.A.M., I.I., S.M.D., and R.S.C. edited and revised manuscript; G.Y., L.A.D., J.S.G., D.A.M., I.I., S.M.D., and R.S.C. approved final version of manuscript.

REFERENCES

1. **Selma-Royo M**, **Tarrazó M**, **García-Mantrana I**, **Gómez-Gallego C**, **Salminen S**, **Collado MC.** Shaping microbiota during the first 1000 days of life. *Adv Exp Med Biol* 1125: 3–24, 2019. doi:10.1007/5584_2018_312.
2. **Charbonneau MR**, **Blanton LV**, **DiGiulio DB**, **Relman DA**, **Lebrilla CB**, **Mills DA**, **Gordon JI.** A microbial perspective of human developmental biology. *Nature* 535: 48–55, 2016. doi:10.1038/nature18845.
3. **Schwarzer M.** Gut microbiota: puppeteer of the host juvenile growth. *Curr Opin Clin Nutr Metab Care* 21: 179–183, 2018. doi:10.1097/MCO.0000000000000463.
4. **Wells JM**, **Brummer RJ**, **Derrien M**, **MacDonald TT**, **Troost F**, **Cani PD**, **Theodorou V**, **Dekker J**, **Méheust A**, **de Vos WM**, **Mercenier A**, **Nauta A**, **Garcia-Rodenas CL.** Homeostasis of the gut barrier and potential biomarkers. *Am J Physiol Gastrointest Liver Physiol* 312: G171–G193, 2017. doi:10.1152/ajpgi.00048.2015.
5. **Laforest-Lapointe I**, **Arrieta MC.** Patterns of early-life gut microbial colonization during human immune development: an ecological perspective. *Front Immunol.* 8: 788, 2017. doi:10.3389/fimmu.2017.00788.
6. **Wang M**, **Monaco MH**, **Donovan SM.** Impact of early gut microbiota on immune and metabolic development and function. *Semin Fetal Neonatal Med* 21: 380–387, 2016. doi:10.1016/j.siny.2016.04.004.
7. **Codagnone MG**, **Spichak S**, **O'Mahony SM**, **O'Leary OF**, **Clarke G**, **Stanton C**, **Dinan TG**, **Cryan JF.** Programming bugs: microbiota and the developmental origins of brain health and disease. *Biol Psychiatry* 85: 150–163, 2019. doi:10.1016/j.biopsych.2018.06.014.
8. **Stewart CJ**, **Ajami NJ**, **O'Brien JL**, **Hutchinson DS**, **Smith DP**, **Wong MC**, **Ross MC**, **Lloyd RE**, **Doddapaneni H**, **Metcalf GA**, **Muzny D**, **Gibbs RA**, **Vatanen T**, **Huttenhower C**, **Xavier RJ**, **Rewers M**, **Hagopian W**, **Toppari J**, **Ziegler AG**, **She JX**, **Akolkar B**, **Lernmark A**, **Hyoty H**, **Vehik K**, **Krischer JP**, **Petrosino JF.** Temporal development of the gut microbiome in early childhood from the TEDDY study. *Nature* 562: 583–588, 2018. doi:10.1038/s41586-018-0617-x.
9. **Donovan SM.** Evolution of the gut microbiome in infancy within an ecological context. *Curr Opin Clin Nutr Metab Care* 23: 223–227, 2020. doi:10.1097/MCO.0000000000000650.
10. **Le Doare K**, **Holder B**, **Bassett A**, **Pannaraj PS.** Mother's milk: a purposeful contribution to the development of the infant microbiota and immunity. *Front Immunol* 9: 361, 2018. doi:10.3389/fimmu.2018.00361.
11. **Bode L**, **Raman AS**, **Murch SH**, **Rollins NC**, **Gordon JI.** Understanding the mother-breastmilk-infant "triad". *Science* 367: 1070–1072, 2020. doi:10.1126/science.aaw6147.
12. **Donovan SM**, **Wang M**, **Monaco MH**, **Martin CR**, **Davidson LA**, **Ivanov I**, **Chapkin RS.** Noninvasive molecular fingerprinting of host-

microbiome interactions in neonates. *FEBS Lett* 588: 4112–4119, 2014. doi:10.1016/j.febslet.2014.07.008.

13. **Chapkin RS**, **Zhao C**, **Ivanov I**, **Davidson LA**, **Goldsby JS**, **Lupton JR**, **Mathai RA**, **Monaco MH**, **Rai D**, **Russell WM**, **Donovan SM**, **Dougherty ER.** Non-invasive stool-based detection of infant gastrointestinal development using gene expression profiles from exfoliated epithelial cells. *Am J Physiol Gastrointest Liver Physiol* 298: G582–G589, 2010. doi:10.1152/ajpgi.00004.2010.
14. **Schwartz S**, **Friedberg I**, **Ivanov IV**, **Davidson LA**, **Goldsby JS**, **Dahl DB**, **Herman D**, **Wang M**, **Donovan SM**, **Chapkin RS.** A metagenomic study of diet-dependent interaction between gut microbiota and host in infants reveals differences in developmental and immune responses. *Genome Biol* 13: R32, 2012. doi:10.1186/gb-2012-13-4-r32.
15. **Knight JM**, **Davidson LA**, **Herman D**, **Martin CR**, **Goldsby JS**, **Ivanov IV**, **Donovan SM**, **Chapkin RS.** Non-invasive analysis of intestinal development in preterm and term infants using RNA-sequencing. *Sci Rep* 4: 5453, 2014. doi:10.1038/srep05453.
16. **Whitfield-Cargile CM**, **Cohen ND**, **He K**, **Ivanov I**, **Goldsby JS**, **Chamoun-Emanuelli A**, **Weeks BR**, **Davidson LA**, **Chapkin RS.** The non-invasive exfoliated transcriptome (exfoliome) reflects the tissue-level transcriptome in a mouse model of NSAID enteropathy. *Sci Rep* 7: 14687, 2017. doi:10.1038/s41598-017-13999-5.
17. **Burrin D**, **Sangild PT**, **Stoll B**, **Thymann T**, **Buddington R**, **Marini J**, **Olutoye O**, **Shulman RJ.** Translational advances in pediatric nutrition and gastroenterology: new insights from pig models. *Annu Rev Anim Biosci* 8: 321–324, 2020. doi:10.1146/annurev-animal-020518-115142.
18. **Mudd AT**, **Dilger RA.** Early-life nutrition and neurodevelopment: use of the piglet as a translational model. *Adv Nutr* 8: 92–104, 2017. doi:10.3945/an.116.013243.
19. **Wang M**, **Radlowski EC**, **Monaco MH**, **Fahey GC Jr**, **Gaskins HR**, **Donovan SM.** Mode of delivery and early nutrition modulate microbial colonization and fermentation products in the neonatal piglet. *J Nutr* 143: 795–803, 2013. doi:10.3945/jn.112.173096.
20. **Lampe JW**, **Kim E**, **Levy L**, **Davidson LA**, **Goldsby JS**, **Miles FL**, **Navarro SL**, **Randolph TW**, **Zhao N**, **Ivanov I**, **Kaz AM**, **Damman C**, **Hockenbery DM**, **Hullar MAJ**, **Chapkin RS.** Colonic mucosal and exfoliome transcriptomic profiling and fecal microbiome response to a flaxseed lignan extract intervention in humans. *Am J Clin Nutr* 110: 377–390, 2019. doi:10.1093/ajcn/nqy325.
21. **Dobin A**, **Davis CA**, **Schlesinger F**, **Drenkow J**, **Zaleski C**, **Jha S**, **Batut P**, **Chaisson M**, **Gingeras TR.** STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 29: 15–21, 2013. doi:10.1093/bioinformatics/bts635.
22. **Martin M.** Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J* 17: 10–12, 2011. doi:10.14806/ej.17.1.200
23. **Anders S**, **Pyl PT**, **Huber W.** HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics* 31: 166–169, 2015. doi:10.1093/bioinformatics/btu638.
24. **Robinson JT**, **Thorvaldsdóttir H**, **Winckler W**, **Guttman M**, **Lander ES**, **Getz G**, **Mesirov JP.** Integrative Genomics Viewer. *Nat Biotechnol* 29: 24–26, 2011. doi:10.1038/nbt.1754.
25. **Thorvaldsdóttir H**, **Robinson JT**, **Mesirov JP.** Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. *Brief Bioinform* 14: 178–192, 2013. doi:10.1093/bib/bbs017.
26. **Franzén O**, **Gan LM**, **Björkegren JLM.** PanglaoDB: a web server for exploration of mouse and human single-cell RNA sequencing data. *Database (Oxford)* 2019: baz046, 2019. doi:10.1093/database/baz046.
27. **Wang Y**, **Song W**, **Wang J**, **Wang T**, **Xiong X**, **Qi Z**, **Fu W**, **Yang X**, **Chen YG.** Single-cell transcriptome analysis reveals differential nutrient absorption functions in human intestine. *J Exp Med* 217: e20191130, 2020. doi:10.1084/jem.20191130.
28. **Krämer A**, **Green J**, **Pollard J Jr**, **Tugendreich S.** Causal analysis approaches in Ingenuity Pathway Analysis. *Bioinformatics* 30: 523–530, 2014. doi:10.1093/bioinformatics/btt703.
29. **Coleman MC**, **Whitfield-Cargile C**, **Cohen ND**, **Goldsby JL**, **Davidson L**, **Chamoun-Emanuelli AM**, **Ivanov I**, **Eades S**, **Ing N**, **Chapkin RS.** Non-invasive evaluation of the equine gastrointestinal mucosal transcriptome. *PLoS One* 15: e0229797, 2020. doi:10.1371/journal.pone.0229797.
30. **Chen Y**, **Agellon LB.** Distinct alteration of gene expression programs in the small intestine of male and female mice in response to ablation of intestinal *Fabp* genes. *Genes (Basel)* 11: 943, 2020. doi:10.3390/genes11080943.
31. **Cummins G**, **Yung DE**, **Cox BF**, **Koulaouzidis A**, **Desmulliez MPY**, **Cochran S.** Luminally expressed gastrointestinal biomarkers. *Expert Rev Gastroenterol Hepatol* 11: 1119–1134, 2017. doi:10.1080/17474124.2017.1373017.
32. **Pelsers MM**, **Hermens WT**, **Glatz JF.** Fatty acid-binding proteins as plasma markers of tissue injury. *Clin Chim Acta* 352: 15–35, 2005. doi:10.1016/j.cccn.2004.09.001.
33. **Schurink M**, **Kooi EM**, **Hulzebos CV**, **Kox RG**, **Groen H**, **Heineman E**, **Bos AF**, **Hulscher JB.** Intestinal fatty acid-binding protein as a diagnostic marker for complicated and uncomplicated necrotizing enterocolitis: a prospective cohort study. *PLoS One* 10: e0121336, 2015. doi:10.1371/journal.pone.0121336.
34. **Heida FH**, **Hulscher JB**, **Schurink M**, **Timmer A**, **Kooi EM**, **Bos AF**, **Bruggink JL**, **Kasper DC**, **Pones M**, **Benkoe T.** Intestinal fatty acid-binding protein levels in necrotizing enterocolitis correlate with extent of necrotic bowel: results from a multicenter study. *J Pediatr Surg* 50: 1115–1118, 2015. doi:10.1016/j.jpedsurg.2014.11.037.
35. **Cheng S**, **Yu J**, **Zhou M**, **Tu Y**, **Lu Q.** Serologic intestinal-fatty acid binding protein in necrotizing enterocolitis diagnosis: a meta-analysis. *Biomed Res Int* 2015: 156704, 2015. doi:10.1155/2015/156704.
36. **Das D**, **Fletcher RB**, **Ngai J.** Cellular mechanisms of epithelial stem cell self-renewal and differentiation during homeostasis and repair. *Wiley Interdiscip Rev Dev Biol* 9: e361, 2020. doi:10.1002/wdev.361.
37. **Bromberg JF**, **Horvath CM**, **Wen Z**, **Schreiber RD**, **Darnell JE Jr.** Transcriptionally active Stat1 is required for the antiproliferative effects of both interferon alpha and interferon gamma**.** *Proc Natl Acad Sci USA* 93: 7673–7678, 1996. doi:10.1073/pnas.93.15.7673.
38. **Klampfer L.** The role of signal transducers and activators of transcription in colon cancer. *Front Biosci* 13: 2888–2899, 2008. doi:10.2741/2893.
39. **Wang H**, **Qu X**, **De Plaen IG**, **Hsueh W.** Platelet-activating factor and endotoxin activate CCAAT/enhancer binding protein in rat small intestine. *Br J Pharmacol* 133: 713–721, 2001. doi:10.1038/sj.bjp.0704102.
40. **Montgomery RK**, **Rings EH**, **Thompson JF**, **Schuijt CC**, **Aras KM**, **Wielenga VJ**, **Kothe MJ**, **Büller HA**, **Grand RJ.** Increased C/EBP in fetal rat small intestine precedes initiation of differentiation marker mRNA synthesis. *Am J Physiol Gastrointes Liver Physiol* 272: G534–G544, 1997. doi:10.1152/ajpgi.1997.272.3.G534.
41. **Oesterreicher TJ**, **Leeper LL**, **Finegold MJ**, **Darlington GJ**, **Henning SJ.** Intestinal maturation in mice lacking CCAAT/enhancer-binding protein alpha (C/EPBα). *Biochem J* 330: 1165–1171, 1998. doi:10.1042/bj3301165.
42. **Mariadason JM**, **Nicholas C**, **L'Italien KE**, **Zhuang M**, **Smartt HJ**, **Heerdt BG**, **Yang W**, **Corner GA**, **Wilson AJ**, **Klampfer L**, **Arango D**, **Augenlicht LH.** Gene expression profiling of intestinal epithelial cell maturation along the crypt-villus axis. *Gastroenterology* 128: 1081–1088, 2005. doi:10.1053/j.gastro.2005.01.054.
43. **Chang J**, **Chance MR**, **Nicholas C**, **Ahmed N**, **Guilmeau S**, **Flandez M**, **Wang D**, **Byun DS**, **Nasser S**, **Albanese JM**, **Corner GA**, **Heerdt BG**, **Wilson AJ**, **Augenlicht LH**, **Mariadason JM.** Proteomic changes during intestinal cell maturation *in vivo*. *J Proteomics* 71: 530–546, 2008. doi:10.1016/j.jprot.2008.08.003.
44. **Delgado ME**, **Brunner T.** The many faces of tumor necrosis factor signaling in the intestinal epithelium. *Genes Immun* 20: 609–626, 2019. doi:10.1038/s41435-019-0057-0.
45. **Noti M**, **Corazza N**, **Tuffin G**, **Schoonjans K**, **Brunner T.** Lipopolysaccharide induces intestinal glucocorticoid synthesis in a TNFα-dependent manner. *FASEB J* 24: 1340–1346, 2010. doi:10.1096/fj.09-140913.
46. **Wang K**, **Han G**, **Dou Y**, **Wang Y**, **Liu G**, **Wang R**, **Xiao H**, **Li X**, **Hou C**, **Shen B**, **Guo R**, **Li Y**, **Shi Y**, **Chen G.** Opposite role of tumor necrosis factor receptors in dextran sulfate sodium-induced colitis in mice. *PLoS One* 7: e52924, 2012. doi:10.1371/journal.pone.0052924.
47. **Vancamelbeke M**, **Vermeire S.** The intestinal barrier: a fundamental role in health and disease. *Expert Rev Gastroenterol Hepatol* 11: 821–834, 2017. doi:10.1080/17474124.2017.1343143.
48. **Kaeffer B.** Survival of exfoliated epithelial cells: a delicate balance between anoikis and apoptosis. *J Biomed Biotechnol Article ID* 2011: 534139, 2011. doi:10.1155/2011/534139.

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, S.M. Donovan, R.S. Chapkin

Table S1. Upstream regulators of genes detected in mucosa and exfoliated intestinal epithelial cells predicted from mechanistic network analysis conducted in IPA.

Upstream Regulator	Molecule Type	P-value of Overlap[a]	Mechanistic Network[b]	Target Molecules in Dataset
TNF	cytokine	4.39E-10	126 (21)	ABCG2, ANPEP, APOA1, AQP3, AXIN1, AXIN2, BMP2, CASP10, CASP3, CASP6, CASP7, CASP8, CCK, CCL28, CD82, CFTR, FABP1, FZD5, GSK3B, HMGCR, HSPA5, IER3, IL18, IL1R2, IL4R, LPL, PIGR, PLIN2, RARRES2, S100A10, SCNN1A, SCNN1B, SLC16A5, SLC1A4, SLC20A1, SLC22A4, SLC2A2, SLC35B2, SLC43A2, SMAD1, SMAD3, TAC1, TNFRSF21
IGF1	growth factor	9.67E-07	118 (18)	ANPEP, BMP2, CASP3, CD74, HMGCR, HSPA5, IER3, IL18, IL2RG, IL4R, LPL, NUPR1, PYY, SLC20A1, SST
IL6	cytokine	3.73E-07	119 (21)	ABCG2, ANPEP, APOA1, APOB, AREG, AXIN1, AXIN2, BMP2, CASP3, CASP7, CCK, CD74, CD82, GCG, HSPA5, IL17RC, IL2RG, IL4R, LPL, SCNN1A, SLC2A2, SLC7A7, SST, TAC1
EGF	growth factor	5.44E-05	108 (16)	ABCG2, ANPEP, APOA1, AQP3, AREG, CASP3, GSK3B, IER3, IL1R2, S100A10, SCNN1B, SLC4A7, SLC9A2, SMAD1, SMAD3
Insulin (INS)	group	3.24E-6	106 (19)	ACAT1, ALDOB, APOA1, APOM, BMPR2, CCK, CYCS, GCG, GSK3B, HMGCR, HSPA5, LPL, PLIN3, RBP4, RDH5, SLC10A2, SLC29A1, SLC2A2, SLC2A5, SLC37A4, SLC38A2, SLC3A1
IL1B	cytokine	5.49E-08	103 (15)	ABCG2, APOB, BMP2, CASP10, CASP3, CCL28, CD74, CD82, CFTR, FABP1, GSK3B, HSPA5, IER3, IL18, IL1R2, PIGR, S100A10, SCNN1A, SCNN1B, SLC10A2, SLC20A1, SLC22A4, SLC2A2, SLC37A4, SLC3A1, SMAD3, TAC1
LEP	growth factor	2.81E-13	102 (21)	ABCG5, ACAT1, ACSL3, ALDOB, APOA1, APOA4, APOM, AQP3, CASP10, CASP3, CASP6, CASP7, CCK, GCG, GSK3B, HMGCR, HSPA5, IL1R2, LPL, NTS, PLIN2, SLC13A2, SLC16A1, SLC27A4, SLC2A2, SST
HNF1A	transcription regulator	3.58E-07	100 (18)	ACAT2, ALDOB, ANPEP, APOB, APOC3, APOM, AQP3, CCL28, FABP1, HMGCR, HSPA5, IFNAR1, PIGR, SLC10A2, SLC2A2, SLC37A4, SLC5A1, SLCO3A1
P38 MAPK	group	6.26E-03	99 (18)	BMP2, CCK, CD82, GCG, HSPA5, IER3, SCNN1A, SMAD2, SMAD3
NR3C1	ligand-dependent nuclear receptor	2.21E-06	98 (9)	ACAT1, BMP2, CASP10, CASP6, CASP7, HMGCS2, IER3, IL18, PIGR, PLIN2, SCNN1A, SLC10A2, SLC25A33, SLC2A5, SLC38A1, SLC7A6, SMAD1, TAC1, TNFRSF21
IL4	cytokine	4.48E-04	95 (14)	CASP3, CASP6, CD74, HMGCR, IER3, IFNAR1, IL13RA1, IL18, IL1R2, IL4R, KCTD12, LGALS2, LPL, LRP1, PIGR, PLIN2, S100A10, SDF2L1, SLC23A1, SLC29A1, SLC4A7, TFF3

IL2	cytokine	2.82E-04	94 (18)	AREG, BMP2, CASP3, CD74, CD9, CYCS, HSPA5, IER3, IL18, IL1R2, IL2RG, IL4R, NT5E, SLC3A2, TNFRSF21
TGFB1	growth factor	7.90E-07	93 (14)	ABCG2, ACSL3, ALDOB, ANPEP, APOB, AQP11, AQP8, AREG, BMP2, CASP3, CASP8, GSN, HSPA5, IER3, IL18, IL4R, LPL, MGAT3, NT5E, NUPR1, S100A10, SAR1B, SLC16A3, SLC16A9, SLC20A1, SLC35A1, SLC3A2, SLC51B, SMAD2, SMAD3, TCN2
INS	other	1.20E-05	92 (17)	APOA1, AQP3, GCG, HMGCR, HSPA5, IER3, LPL, PLIN2, SLC2A2
ESR1	ligand-dependent nuclear receptor	1.09E-08	91 (10)	ABCG2, ABCG5, APOA1, AQP3, AREG, ATP2A3, AXIN2, BMP2, BMPR2, CASP10, CASP6, CD74, COPA, FABP1, GSN, HMGCR, HMGCS2, IER3, LPL, PLIN2, RBP4, RDH5, SCNN1B, SLC16A1, SLC16A3, SLC1A4, SLC25A36, SLC26A2, SLC35A1, SLC38A2, SLC39A4, SLC3A2, SLC44A1, SLC6A6, SLC9A2, SLC9A3R1, SMAD2, SMAD3, SMAD5, TAC1, TFF3
Ins1	other	4.24E-06	91 (20)	APOA1, APOC3, CCK, GCG, HMGCR, LDHB, LPL, LRP1, PLIN2, SLC28A1, SLC28A2, SLC2A2, SLC37A4
LDL	complex	7.56E-07	89 (12)	BMP2, CASP10, CASP3, CASP6, CASP7, CASP8, HMGCR, HSPA5, IL17RC, LRP1, PLIN2, SMAD3, VNN1
NR0B2	ligand-dependent nuclear receptor	7.33E-03	88 (16)	APOA1, APOM, HMGCR, SLC10A2
STAT1	transcription regulator	9.76E-04	87 (10)	CASP3, CASP6, CASP8, IL18, KCTD12, S100A10, SLC2A2, SLC51B, SMAD2, SMAD3
FGF2	growth factor	4.40E-03	87 (15)	ANPEP, AQP3, AREG, AXIN2, BMP2, CCK, GSK3B, S100A10, SLC20A1, TAC1
TP53	transcription regulator	6.15E-03	86 (10)	ABCG2, ACAT1, ACSL3, APOA1, AQP3, AREG, CASP3, CASP6, CASP8, CD81, CD82, DHRS9, DPEP1, GSK3B, GSN, HMGCR, HSPA5, IER3, IL4R, LDHB, NUPR1, SCNN1A, SLC16A1, SLC25A13, SLC5A3, SLC6A6, TCN2, TMSB10/TMSB4X
PPARG	ligand-dependent nuclear receptor	8.56E-06	84 (17)	ABCG2, APOA1, AQP3, CYCS, FABP1, FABP2, HMGCR, HMGCS2, LPL, PLIN2, RARRES2, RBP4, SLC27A4, SLC2A2, SLC44A1, VNN1
IFNG	cytokine	1.42E-08	82 (11)	AQP11, AREG, CASP10, CASP3, CASP6, CASP7, CASP8, CCL28, CD74, CFTR, HMGCR, HSPA5, IER3, IL18, IL4R, KCTD12, LPL, NUPR1, PAM, PIGR, S100A10, SCNN1A, SCNN1B, SLC16A3, SLC16A9, SLC28A1, SLC28A2, SLC29A1, SLC2A2, SLC37A4, SLC3A1, SLC4A4, SLC6A6, SMAD1, SMAD3, TAC1
NR1H3	ligand-dependent nuclear receptor	8.82E-04	83 (18)	ABCG5, ACSL3, APOA1, APOC3, APOM, DHRS9, HMGCR, LPL

ESR2	ligand-dependent nuclear receptor	7.47E-04	82 (7)	ANPEP, AQP3, BMP2, BMPR2, LPL, MGAT3, PLIN2, RBP4, SLC9A3R1, SMAD2, SMAD3, TAC1, TSPAN13
IL10	cytokine	6.95E-04	82 (11)	BMP2, CASP3, CASP8, CD74, CLCA1, IL18, IL1R2, IL2RG, IL4R, NT5E, PAM
SMAD3	transcription regulator	5.40E-06	80 (13)	APOA1, APOA4, APOB, APOC3, AREG, AXIN2, BMP2, HSPA5, SLC51B, SMAD1, SMAD3, SMAD5
LDLR	transporter	1.31E-04	80 (16)	APOB, AXIN2, BMP2, DHRS9, HMGCR, HSPA5, IER3, LPL, LRP1
RXRA	ligand-dependent nuclear receptor	1.13E-08	79 (17)	ABCG2, ACSL3, APOA1, APOC3, APOM, FABP1, FABP2, FABP6, HMGCS2, IER3, LPL, RBP2, SLC10A2, SLC2A2, SLC51A, SLC51B
CD36	transmembrane receptor	2.13E-06	79 (15)	APOA1, APOA4, APOB, AREG, FABP1, HSPA5, LPL, LRP1
NFkB (complex)	complex	4.58E-06	78 (12)	BMP2, CASP7, CASP8, CCK, CD74, CD82, CFTR, DVL3, FABP6, IER3, IL18, SLC16A3, SLC22A4, SLC2A2, SLC2A5, SLC37A4, SLC3A1, SMAD3, TAC1
STAT6	transcription regulator	2.62E-03	78 (8)	AREG, CASP6, FABP1, HMGCR, HMGCS2, IL4R, KCTD12, S100A10, SLC4A7, TFF3
ADIPOQ	other	9.66E-04	77 (12)	BMP2, CASP6, GSK3B, HMGCS2, IL4R, LPL, SLC27A2
RARA	ligand-dependent nuclear receptor	5.09E-03	77 (12)	APOA1, APOC3, APOM, CD9, RBP2, SLC17A9, SLC25A36, SLC2A2, SMAD3
PPARD	ligand-dependent nuclear receptor	8.52E-09	76 (14)	APOA1, APOA4, APOB, AQP3, FABP1, FCGBP, HMGCS2, LDHB, LPL, PLIN2, SLC10A2, SLC27A2, SLC2A10, SLC2A2, TFF3
CG	complex	2.26E-05	76 (5)	AREG, BMP2, HMGCR, IER3, IL1R2, IL4R, KHK, S100A10, SLC20A1, SLC4A4, SMAD2, TAC1, TSPAN13
APOA1	transporter	1.70E-03	75 (15)	APOA1, APOC3, HMGCR, HSPA5
CETP	enzyme	7.98E-10	75 (16)	APOA1, APOB, HMGCR, HSPA5, LPL, PLIN2, PLIN3
LEPR	transmembrane receptor	6.77E-05	74 (17)	ABCG5, APOA1, APOA4, APOB, CASP8, HMGCR, IL1R2, LPL, SLC2A2
PPARGC1A	transcription regulator	2.01E-05	72 (12)	ACAT1, APOA4, APOC3, CYCS, GSK3B, HMGCS2, IL18, IL1R2, IL4R, LDHB, LPL, PLIN2, RBP4
IRF1	transcription regulator	6.33E-03	71 (7)	CASP3, CASP7, CASP8, CFTR, IL18, PIGR
PPARA	ligand-dependent nuclear receptor	2.19E-11	69 (17)	ACAT1, ACAT2, ALDOB, APOA1, APOA4, APOC3, APOM, AQP3, AQP8, CYCS, FABP1, FABP2, HMGCR, HMGCS2, KHK, LPL, PLIN2, RBP2, RBP4, SLC10A2, SLC27A2, SLC27A4, VNN1

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, S.M. Donovan, R.S. Chapkin

HTT	transcription regulator	7.88E-04	68 (11)	ACAT2, APOA1, CASP3, CCK, CD74, CD9, CYCS, GSN, HMGCR, HSPA5, KCTD17, LDHB, NTS, RBP4, SST, TAC1
ABCA1	transporter	5.70E-05	67 (13)	APOA1, APOB, APOM, HMGCR, HSPA5
FOXO1	transcription regulator	6.09E-05	67 (12)	ALDOB, APOC3, APOM, CASP6, CASP8, CYCS, FABP2, HMGCR, HSPA5, IER3, IL18, LPL, SLC25A12, SLC2A2
SREBF1	transcription regulator	1.59E-05	65 (15)	APOC3, CYCS, FABP6, HMGCR, HSPA5, IL1R2, LPL, NUPR1, SLC20A1, SLC22A4, SLC2A2
FOXA2	transcription regulator	9.87E-07	65 (8)	ALDOB, APOA1, APOB, APOC3, APOM, CLCA1, GCG, HMGCS2, LPL, PYY, SLC2A2, SST
EGR1	transcription regulator	1.46E-03	64 (7)	APOA1, AREG, CASP3, CASP6, CASP7, CASP8, HMGCR
IL1	group	6.93E-05	63 (7)	ABCG2, APOC3, AREG, BMP2, FZD5, HMGCR, IL1R2, IL4R, LRP1, PIGR, RBP4, TAC1
CSF1	cytokine	4.47E-04	63 (8)	CD74, CD9, HMGCR, HSPA5, IL18, LPL, SLC16A3, SLC1A5, SLC29A1
AKT1	kinase	1.35E-03	61 (8)	ACAT2, APOC3, BMP2, CASP3, CD74, HMGCR, LDHB, SLC3A2, SLC4A7
IRS2	enzyme	5.99E-03	61 (9)	HMGCR, LPL, SLC2A2
Vegf	group	5.51E-03	59 (7)	ANPEP, APOM, AXIN2, BMP2, CASP3, CASP8, HMGCS2, IL17RC, IL18, SLC20A1, SMAD5
SMAD4	transcription regulator	3.15E-04	58 (10)	APOA1, APOA4, APOB, APOC3, AREG, BMP2, IER3, SMAD2, SMAD3
NCOA2	transcription regulator	6.13E-07	57 (8)	APOA4, APOM, BMP2, HMGCR, KHK, LPL, PLIN2, SLC2A2, SLC3A1
HDAC5	transcription regulator	5.46E-03	57 (7)	CASP8, HMGCR, HMGCS2, LPL
NPC1	transporter	7.04E-03	56 (8)	APOB, CD74, HMGCR, LPL
NR2F1	ligand-dependent nuclear receptor	2.26E-05	56 (8)	APOA1, APOA4, APOC3, LPL, RBP2
CEBPA	transcription regulator	2.17E-04	55 (10)	ANPEP, APOA4, APOB, APOC3, CYCS, GSK3B, HMGCR, HSPA5, IER3, LPL, PLIN2, RARRES2, TAC1
Growth hormone	group	1.10E-03	54 (7)	BMP2, FZD5, IER3, LPL, SCNN1A, SLC7A9, SST, TFF3
Pka	complex	8.76E-03	53 (8)	AREG, CCK, HMGCR, NTS, SST
IGF2	growth factor	4.01E-03	52 (7)	AREG, BMP2, CASP6, IER3, SLC20A1, SLC38A2

YAP1	transcription regulator	2.90E-05	52 (6)	ACAT1, ACAT2, AREG, BMP2, CASP3, CASP8, HMGCR, HMGCS2, LPL, NUPR1, SMAD2
SCARB1	transporter	9.04E-04	52 (11)	ABCG5, APOA1, APOB, HMGCR
CYP19A1	enzyme	4.77E-03	51 (8)	GSK3B, LPL, PLIN2, SLC27A2
APOE	transporter	7.00E-04	51 (10)	ACAT1, APOA1, APOB, CASP3, HMGCR, HSPA5, LPL, LRP1, SLC27A4, SMAD1
PDX1	transcription regulator	2.79E-04	50 (7)	ALDOB, ATP2A3, GIP, HSPA5, IER3, SLC2A2, SLC6A6, SST
VDR	transcription regulator	8.97E-06	50 (9)	FABP6, IER3, IL18, LPL, NUPR1, PLIN2, PYY, SLC10A2, SLC51A, SLC51B, SLC7A7
KRAS	enzyme	6.63E-03	49 (5)	ANPEP, AREG, AXIN2, CASP3, CD74, CFTR, GSN, HSPA5, IER3, NT5E, SLC20A1, SLC29A1, SLC3A2
HNF1B	transcription regulator	4.37E-03	49 (4)	ALDOB, AXIN2, SLC29A1, SLC2A2, SLC5A1
HNF4A	transcription regulator	1.37E-04	49 (4)	ABCG5, ACAT1, ALDOB, ANPEP, APOA1, APOA4, APOB, APOC3, APOM, AQP3, AQP8, AXIN2, FABP1, FABP2, GSK3B, GSN, HSPA5, IFNAR1, MAL2, RBP2, SLC17A5, SLC25A13, SLC27A2, SLC2A2, SLC30A7, SLC31A1, SLC35A1, SLC35A2, SLC35A3, SLC35D1, SLC37A4, SLC38A1, SLC39A7, SLC39A9, SLC44A1, SLC5A3
RXRG	ligand-dependent nuclear receptor	9.74E-06	49 (9)	APOA1, APOC3, FABP1, LPL, RBP2
SP1	transcription regulator	1.74E-04	49 (8)	APOA1, APOC3, ATP2A3, CASP3, FABP6, HMGCR, HSPA5, IL4R, LPL, PYY, SLC22A4, SLC31A1, SLC4A7, SLC5A1, SMAD3
GIP	other	4.17E-04	48 (7)	GCG, GIP, LPL
NR1H4	ligand-dependent nuclear receptor	3.78E-09	47 (13)	ABCG5, APOA1, APOC3, APOM, FABP1, FABP2, FABP6, HSPA5, RARRES2, SLC10A2, SLC13A1, SLC51A, SLC51B
EPAS1	transcription regulator	2.17E-04	47 (7)	ABCG2, AREG, FABP2, HSPA5, LPL, NT5E, PLIN2, SLC16A4, SLC29A1
BDNF	growth factor	8.90E-03	46 (8)	BMP2, CASP3, HMGCR, HSPA5, NMB, S100A10, SST, TAC1
USP7	peptidase	1.87E-03	44 (7)	FABP1, PLIN2, SLC2A2
CHGA	other	1.57E-03	44 (7)	GCG, HSPA5
CEBPB	transcription regulator	6.49E-03	44 (7)	APOB, APOC3, CFTR, GSK3B, IER3, NUPR1, SLC38A2, SMAD2, SST, TAC1
CYB5R4	enzyme	6.85E-04	43 (7)	ACSL3, HSPA5, LPL

NR2F2	ligand-dependent nuclear receptor	1.99E-05	43 (7)	APOA1, APOA4, APOC3, LPL, RBP2
ATG5	other	1.87E-03	41 (4)	CASP3, CASP7, HSPA5
SCP2	transporter	5.86E-04	41 (6)	FABP1, HMGCR, PLIN2
MMP2	peptidase	3.87E-03	41 (4)	BMP2, IL2RG, IL4R
ACSS2	enzyme	3.76E-05	38 (7)	FABP1, FABP2, HMGCR, SLC27A4
CD3	complex	1.63E-03	37 (4)	AQP3, CASP3, CASP6, CASP8, CD74, CYCS, HSPA5, IL2RG, IL4R, ILF2, LRP1, PLIN2, SLC5A3, SPINK4
SMAD7	transcription regulator	2.30E-03	37 (6)	ALDOB, APOC3, AREG, BMP2, BMPR2, SMAD3
SREBF2	transcription regulator	7.01E-05	37 (5)	FABP6, HMGCR, LRP1, PLIN2, RARRES2, TFF3
Esrra	transcription regulator	4.55E-05	37 (5)	ACAT1, ALDOB, APOA4, CYCS, FABP1, FABP2, SCNN1A
FGF19	growth factor	1.06E-04	37 (6)	AQP8, AREG, HMGCR, SLC10A2, SLC2A2, SLC2A5
LCAT	enzyme	5.38E-05	36 (6)	APOA1, AXIN2, HMGCR, HSPA5
IL13	cytokine	3.42E-05	36 (4)	CASP3, CASP6, CASP7, CASP8, CLCA1, GSN, IL13RA1, IL18, IL1R2, PLIN2, SLC26A6, SLC7A7, TFF3, VNN1
CTNNB1	transcription regulator	6.81E-04	34 (3)	ACAT1, ACAT2, AXIN2, BMP2, CASP7, DPEP1, HMGCR, HMGCS2, ILF3, LPL, PLIN2, PLIN3, RBP4, SCNN1B, SLC17A9, SLC1A5, SLC26A2, SLC38A5, SLC9A3R1
BMP7	growth factor	4.48E-03	33 (6)	BMP2, BMPR2, CYCS, SMAD1, SMAD2, SMAD5
AHR	ligand-dependent nuclear receptor	5.18E-03	32 (2)	ABCG2, AREG, CASP8, HMGCR, HSPA5, SLC16A5, SLC1A1, SLC1A4, TFF3
NCOA6	transcription regulator	1.83E-06	32 (6)	ABCG5, HSPA5, LPL, SST
SOST	other	3.01E-03	31 (4)	AXIN2, LPL, PLIN2
CPE	peptidase	8.71E-05	30 (6)	CCK, HMGCR, PAM, TAC1
INSIG1	other	3.03E-04	29 (4)	HMGCR, HMGCS2, LPL, PLIN2, SLC2A2, VNN1
NR1H2	ligand-dependent nuclear receptor	2.19E-03	28 (5)	ABCG5, HMGCR, LPL, SLC2A2, SMAD2
AVP	other	8.68E-03	28 (5)	HSPA5, SCNN1B, SST

LPL	enzyme	1.12E-04	28 (4)	APOA1, APOA4, APOB, CASP3, LPL
MTTP	transporter	1.13E-05	28 (5)	ABCG5, APOA4, APOB, HMGCR, HSPA5
FABP1	transporter	1.19E-03	27 (5)	ABCG5, FABP6, SLC10A2
Ncoa6	transcription regulator	4.80E-03	27 (5)	AREG, LPL
MAPT	other	7.03E-03	25 (3)	ABCG2, ACAT1, CASP3, CYCS, GSK3B, HSPA5, LDHB, LRP1, SLC16A3
JUN	transcription regulator	1.47E-03	25 (4)	ACAT2, APOC3, APOM, CYCS, GIP, GSK3B, NTS, S100A10, SLC10A2, SLC38A2, SLC6A6, SST
FOS	transcription regulator	1.87E-03	25 (4)	ABCG5, AQP3, CASP3, CYCS, FZD5, GSK3B, HSPA5, LPL, NTS, S100A10, SLC10A2, SMAD5, TAC1
ABL1	kinase	5.22E-03	24 (3)	BMP2, IER3, SST
IFT88	other	2.76E-03	24 (3)	ACAT1, ACAT2, HMGCR
STAT5A	transcription regulator	1.11E-03	24 (3)	AXIN2, CASP3, CASP6, CASP8, MAL2, SLC2A10, SLC7A7, TFF3, TSPAN13
BGN	other	1.68E-03	24 (3)	AXIN2, BMP2, SMAD3
WNT4	cytokine	1.42E-02	24 (3)	AXIN2, BMP2, PLIN2
NOD2	other	4.94E-04	23 (3)	AREG, AXIN2, BMP2, IER3, IL18
CLU	other	2.52E-03	23 (3)	HSPA5, SMAD2, SMAD3
ATF2	transcription regulator	1.46E-03	21 (3)	APOC3, GIP, NTS, NUPR1, SST
CFH	other	1.13E-03	20 (3)	APOA1, APOB
POU5F1	transcription regulator	2.24E-04	17 (2)	AXIN1, B3GNT7, CASP10, CASP3, CASP6, CASP7, FABP1, FABP2, SLC27A4, SMAD1, TNFRSF21
LMO2	transcription regulator	1.03E-05	15 (2)	AXIN2, BMPR2, CASP3, FZD5, GSN, ITM2C, KCTD17, NT5E, SLC16A1, SLC16A5, SLC5A9, SMAD3
ERN1	kinase	1.58E-03	11 (2)	APOB, HSPA5, PLIN2, SDF2L1, SLC20A1, SLC25A28
DAXX	transcription regulator	4.17E-04	10 (2)	CASP10, CASP3, CASP8

*The overlap p-value measures whether there is a statistically significant overlap between the dataset genes and the genes regulated by a transcription factor using Fisher's Exact Test.

†Mechanistic network value is a list plausible sets of connected upstream regulators that can work together to elicit the gene expression changes observed in a dataset. The value shown as A (B) in which A is the genes that are downstream from regulators (B) in the network. For more information, see: https://qiagen.secure.force.com/KnowledgeBase/KnowledgeIPAPage?id=kA41i000000L5lzCAC

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, S.M. Donovan, R.S. Chapkin

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, R.S. Chapkin, S.M. Donovan

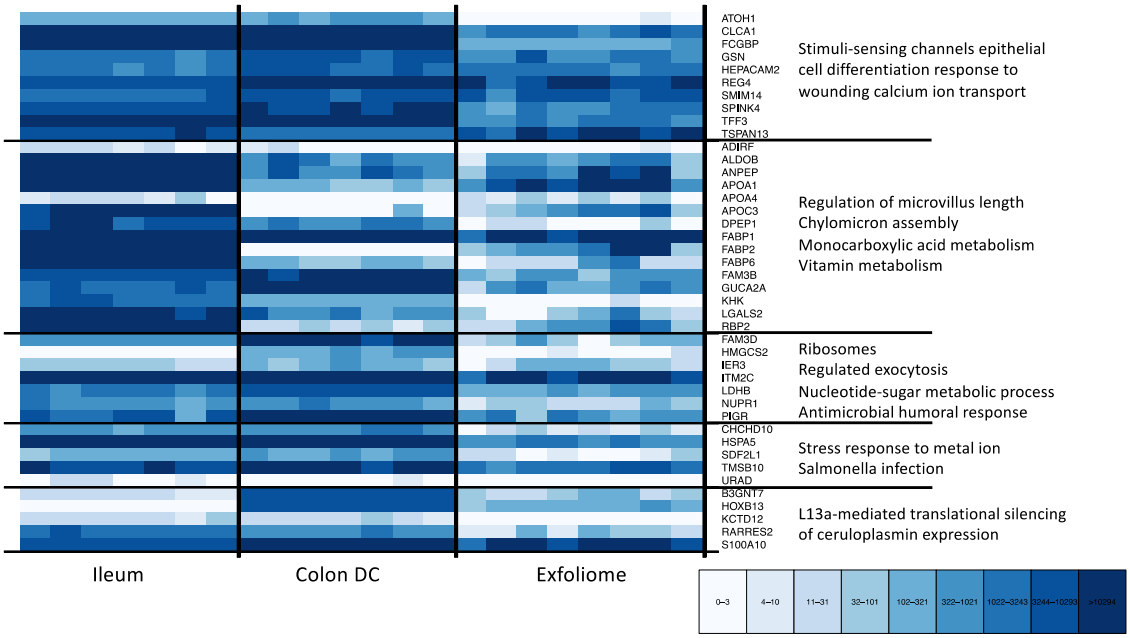

Fig. S8. Heatmap counts of signature genes in goblet cells and gene functional enrichment in three data sources: two different mucosa locations (IL and descending colon, DC) and DC exfoliome content. Color bar represents an exponential scale from 0 to 90% quantile of data value.

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, S.M. Donovan, R.S. Chapkin

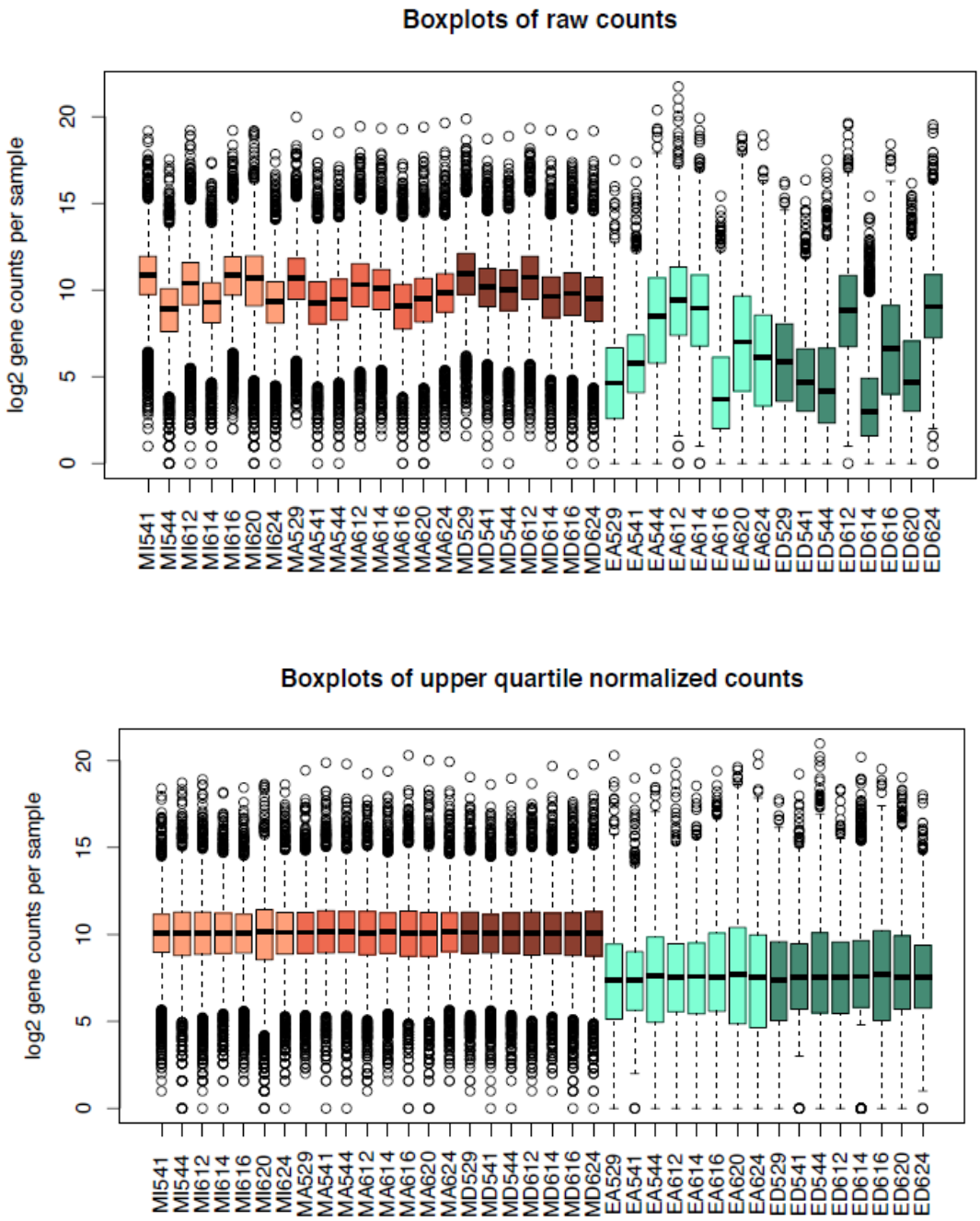

Fig. S7. Boxplots showing the distribution of log(2) transformed raw and upper quartile normalized counts. For each sample, (x-axis), log2(raw counts + 1) (***top panel***) and log2(normalized counts + 1) (***bottom panel***) are plotted in the y-axis for 15,523 genes. MI = mucosa data from Ileal location; MA = mucosa data from AC location; MD = mucosa data from DC location; EA = exfoliome data from AC location; ED = exfoliome data from DC location; Last three digits: each piglet ID.

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, R.S. Chapkin, S.M. Donovan

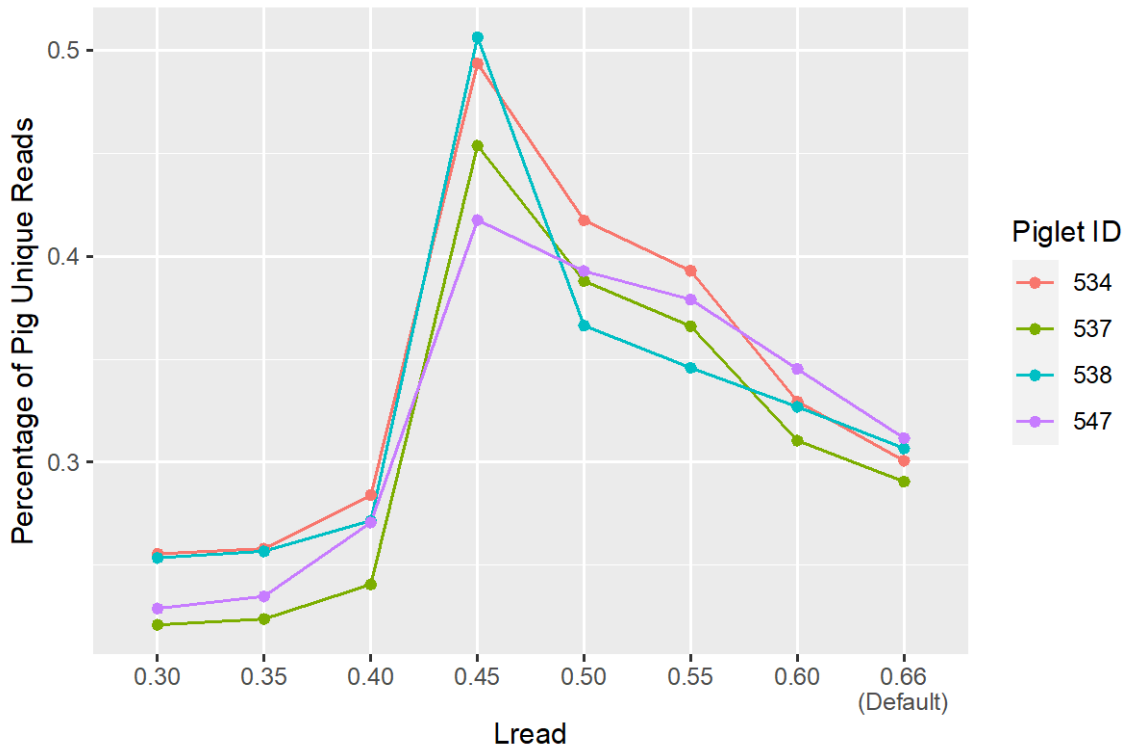

Fig. S6. Line graph showing the percentage of uniquely mapped reads among total number of mapped reads per exfoliome sample. Each color represents different piglet samples. Since the percentage of uniquely annotated reads is maximized at Lread = 0.45, 0.45 was selected for optimal Lread criteria of STAR RNA-seq aligner for piglet exfoliome data. Very similar patterns were observed regardless of the type of trimming.

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets

Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, R.S. Chapkin, S.M. Donovan

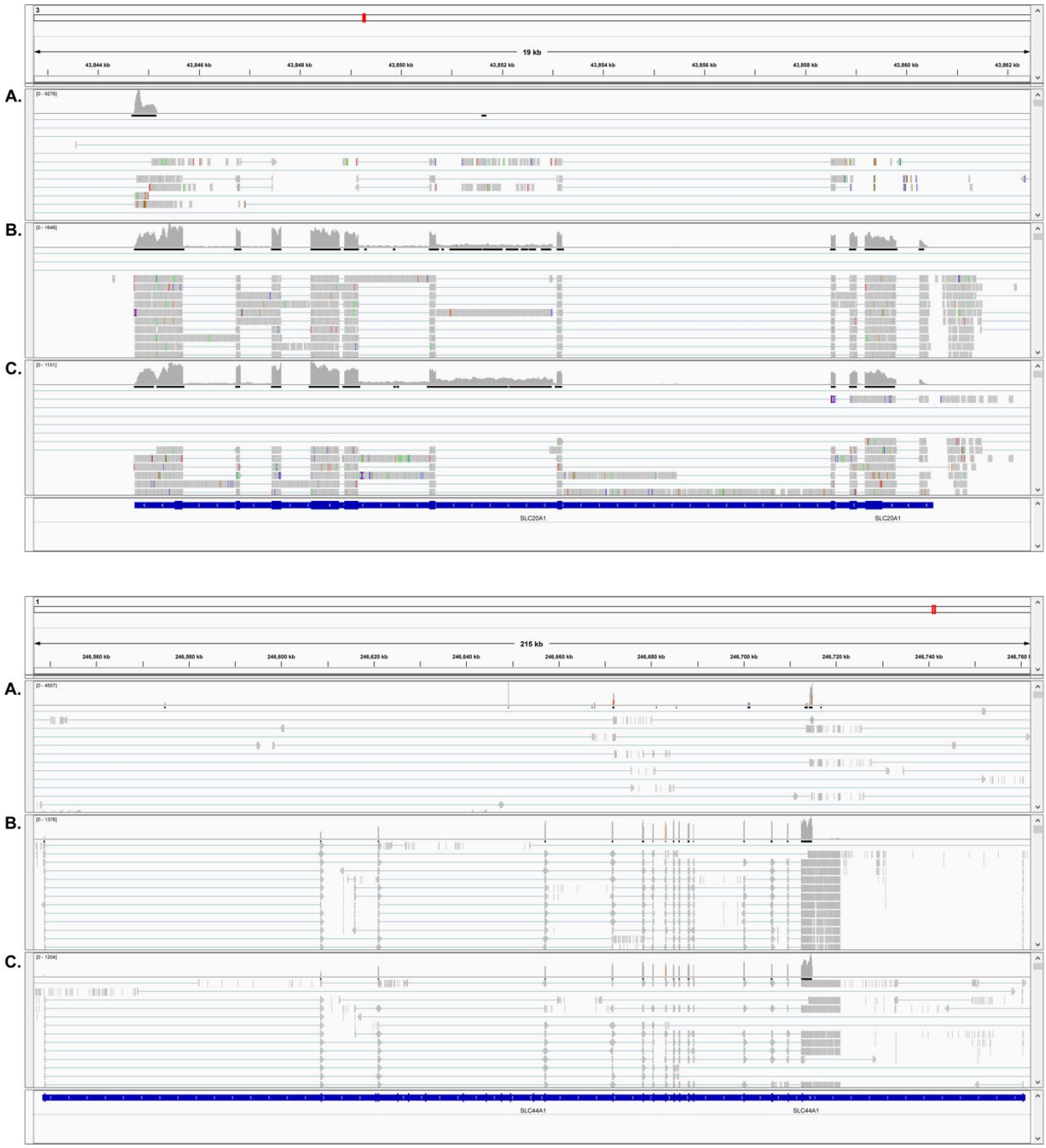

Fig. S5. Representative gene mapping images for sodium-dependent phosphate transporter 1 (*SLC20A1*) (top panel) and for choline transporter-like protein 1 (bottom panel). Figures were generated using Integrative Genome Viewer (IGV) version 2.8.4 and the Ensembl.org *Sus scrofa* 11.1 genome. Samples from the A. distal colonic exfoliome, B. distal colonic mucosa, and C. ileal mucosa were computationally merged.

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, R.S. Chapkin, S.M. Donovan

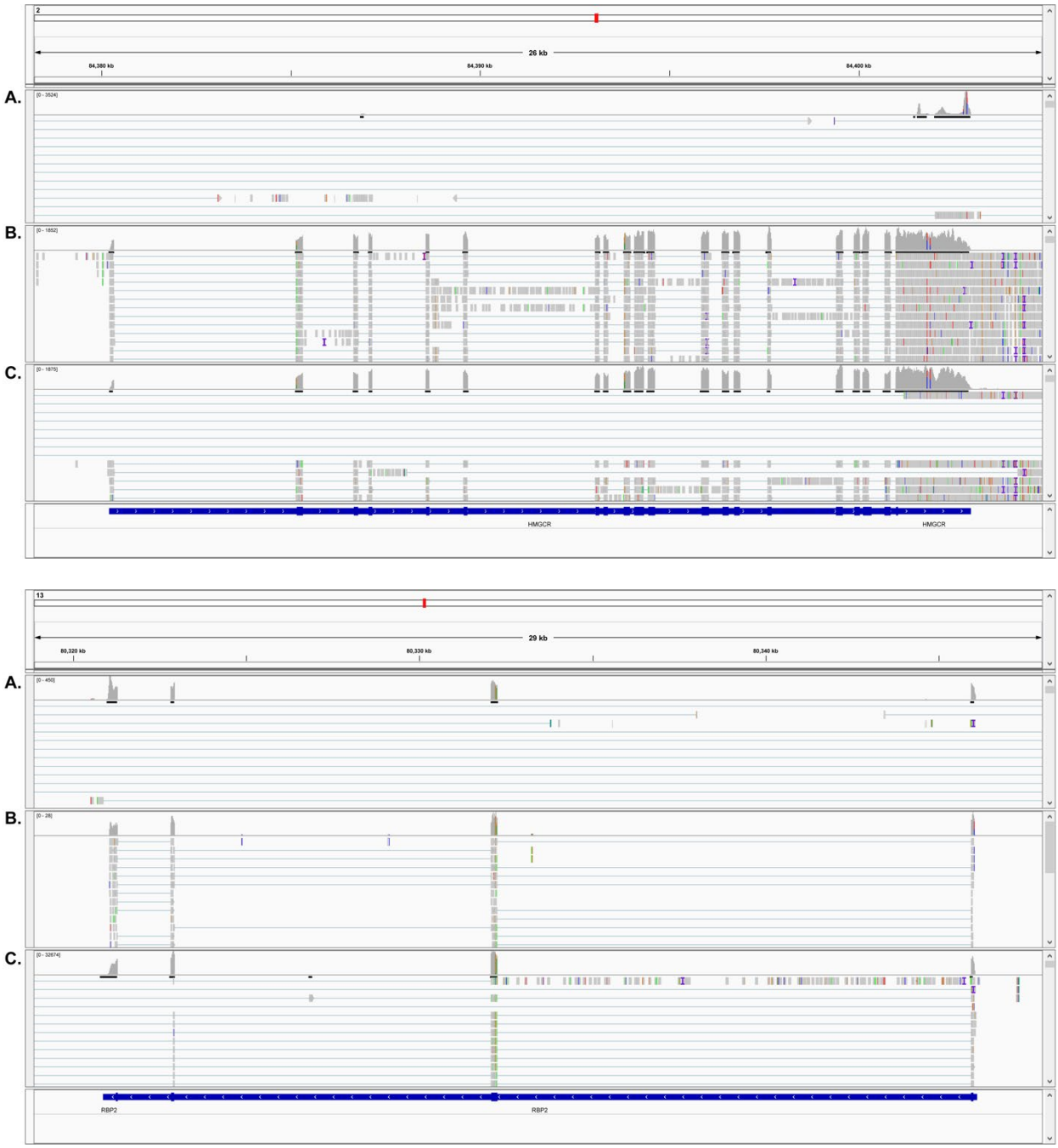

Fig. S4. Representative gene mapping images for HMG-CoA reductase (*HMGCR*) (top panel) and retinol binding p rotein 2 (*RBP2*) (bottom panel). Figures were generated using Integrative Genome Viewer (IGV) version 2.8.4 and the Ensembl.org *Sus scrofa* 11.1 genome. Samples from the A. distal colonic exfoliome, B. distal colonic mucosa, and C. ileal mucosa were computationally merged.

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, R.S. Chapkin, S.M. Donovan

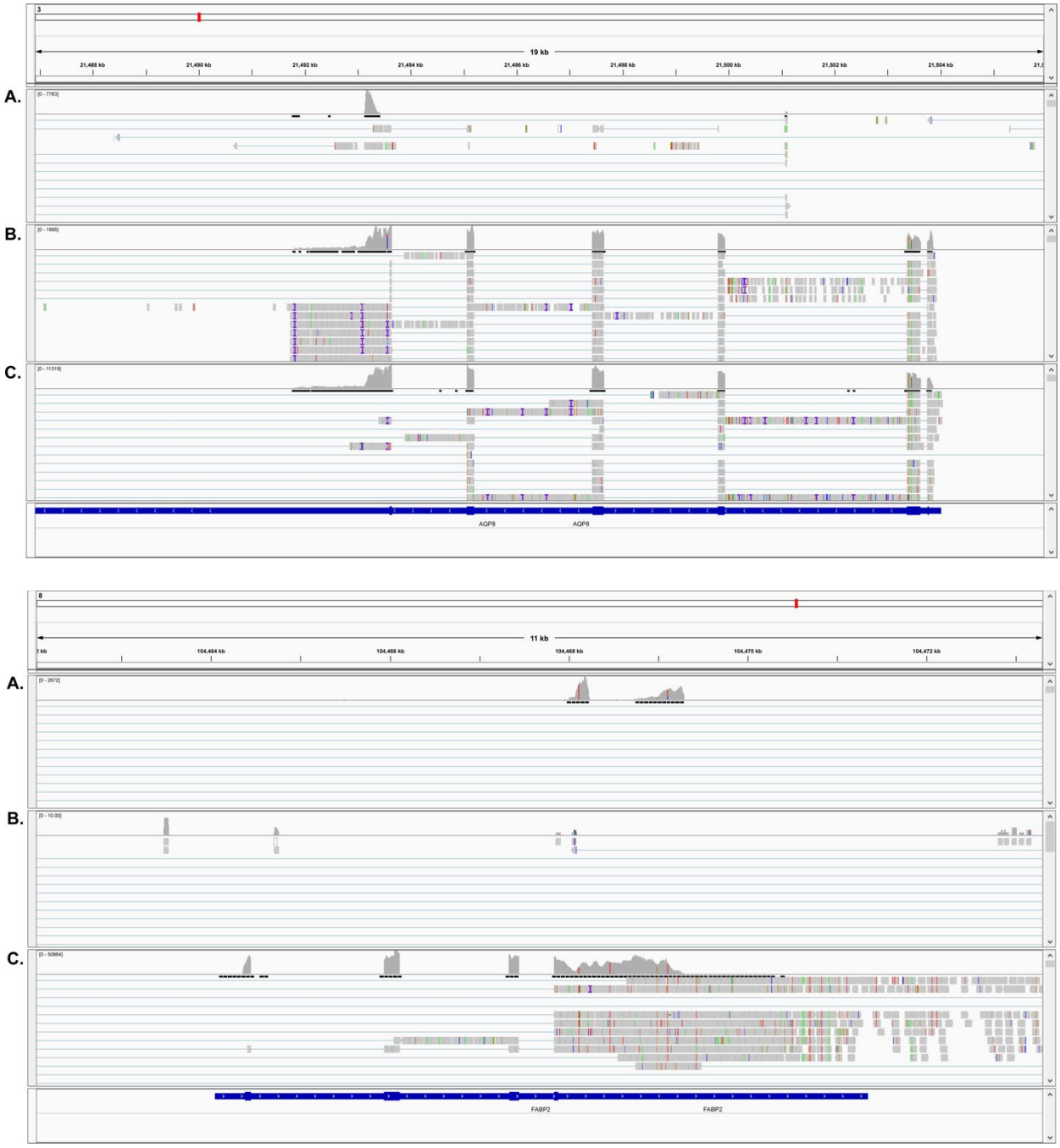

Fig. S3. Representative gene mapping images for Aquaporin (*AQP8*) (top panel) and fatty acid binding protein 2 (*FABP2*) (bottom panel). Figures were generated using Integrative Genome Viewer (IGV) version 2.8.4 and the Ensembl.org *Sus scrofa* 11.1 genome. Samples from the A. distal colonic exfoliome, B. distal colonic mucosa, and C. ileal mucosa were computationally merged.

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets

Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, R.S. Chapkin, S.M. Donovan

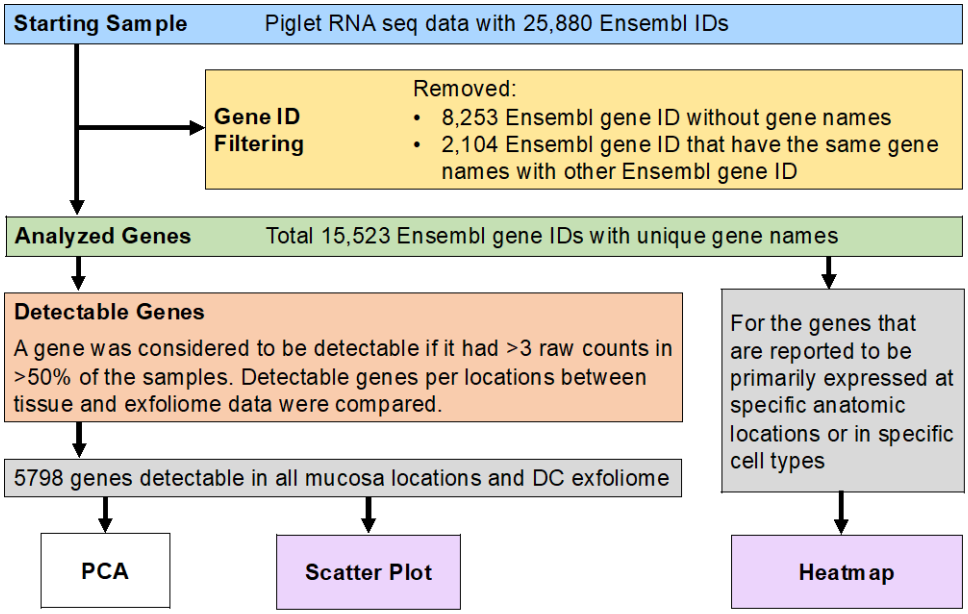

Fig. S2. Flow diagram of data processing and analysis pipeline. Abbreviations: DC (descending colon), PCA (Principal Component Analysis).

Title: Exfoliated Epithelial Cell Transcriptome Reflects Both Small and Large Intestinal Cell Signatures in Piglets
Authors: G. Yoon, L.A. Davidson, J.S. Goldsby, D.A. Mullens, I. Ivanov, R.S. Chapkin, S.M. Donovan

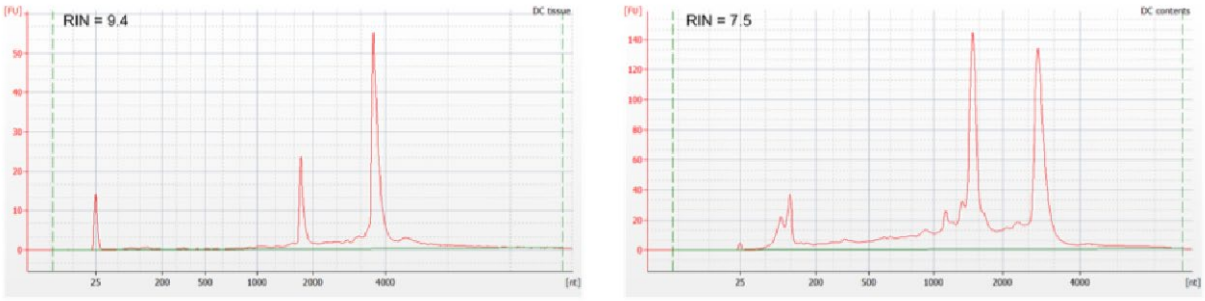

Fig. S1. Bioanalyzer profiles. Representative profiles of RNA quality from the Bioanalyzer are shown for DC tissue and the DC exfoliome. Note that the tissue contents contain 18S and 28S rRNA peaks as expected, while the exfoliome contents contain 16S and 23S rRNA peaks, indicating a large bacterial RNA presence.