Noninvasive molecular fingerprinting of host–microbiome interactions in neonates

Sharon M. Donovan, Mei Wang, Marcia H. Monaco, Camilia R. Martin, Laurie A. Davidson, Ivan Ivanov, Robert S. Chapkin

Department of Food Science and Human Nutrition, University of Illinois, Urbana, IL 61801, USA

Department of Neonatology, Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, MA 02215, USA

Department of Nutrition & Food Science and Program in Integrative Nutrition and Complex Diseases, Texas A&M University, College Station, TX 77843-2253, USA

Department of Veterinary Physiology and Pharmacology, Texas A&M University, College Station, TX 77843-2253, USA

Center for Translational Environmental Health Research, Texas A&M University, College Station, TX 77843-2253, USA

Abstract
The early postnatal period is a critical window for intestinal and immune maturation. Intestinal development and microbiome diversity and composition differ between breast- (BF) and formula-fed (FF) infants. Mechanistic examination into host–microbe relationships in healthy infants has been hindered by ethical constraints surrounding tissue biopsies. Thus, a statistically rigorous analytical framework to simultaneously examine both host and microbial responses to dietary/environmental factors using exfoliated intestinal epithelial cells was developed. Differential expression of >1200 genes, including genes regulating intestinal proliferation, differentiation and barrier function, was observed between BF and FF term infants. Canonical correlation analysis uncovered a relationship between microbiome virulence genes and host immunity and defense genes. Lastly, exfoliated cells from preterm and term infants were compared. Pathways associated with immune cell function and inflammation were up-regulated in preterm, whereas cell growth-related genes were up-regulated in the term infants. Thus, coordinate measurement of the transcriptomes of exfoliated epithelial cells and microbiome allows inquiry into mutualistic host–microbe interactions in the infant, which can be used to prospectively study gut development or, retrospectively, to identify potential triggers of disease in banked samples.

Keywords: Infant, Intestine, Nutrition, Microbiome, Gene expression, Exfoliated cell

1. Introduction
Intestinal epithelial cells and the commensal microbiota are in close and intimate contact. Studies emanating from germ free and gnotobiotic animals have provided conclusive evidence of the critical role of the intestinal microbiota in regulating gut development and gene expression [1,2]. Mucosal and systemic immunity [3], the enteric nervous system [4], gut brain axis [5] and host metabolism [6,7]. Recent studies have dispelled the concept that amniotic fluid and meconium are sterile under normal conditions [8]. Meconium, which is formed primarily by ingestion of amniotic fluid by the fetus in utero, also contains exfoliated intestinal cells and mucus. The meconium microbiome is influenced by maternal factors, including clinical conditions [9] and probiotic use [10], and may impact child health outcomes [8,9,11]. Thus, host–microbe interactions and education of the neonatal immune system begin in the womb [8].

Immediately after delivery, the human infant acquires a much more complex microbiota, whose composition is influenced by an interplay between genetic and environmental factors [12], of which nutrition is a key component [13]. At the same time, the gastrointestinal tract undergoes rapid structural and functional adaptation, which differs between breast-fed (BF) and formula-fed (FF) infants [14,15]. Although human milk contains growth factors and bioactive proteins and lipids that may directly promote the growth of the gastrointestinal tract [16,17], we speculated that dissimilarities in the composition of the microbiota between breast- and
Our long-term goal is to determine the role of host–microbe interactions within the neonatal intestine on infant development and to define how these cross-functional communications are affected by diet. Among the components of human milk that shape the composition of the microbiota are the human milk oligosaccharides (HMO). The HMO are comprised of a mixture of up to 200 complex oligosaccharides that constitute the third most predominant component of human milk [19]. The HMO content and composition is influenced by the mothers’ genetics (FUT-2 secretor status and Lewis blood group) [20], preterm delivery [21] and, to a lesser degree, the stage of lactation, where sialic acid containing HMO decline, while fucosylated HMO increase or stay constant over the course of lactation [19]. The potential physiological roles of HMO for the developing infant is far reaching in that their multifunctional actions range from regulation of intestinal cell proliferation, functional differentiation and apoptosis [22,23], gene expression [24], immune function [25–27], pathogen protection [28,29], and prebiotic activities, including serving as substrates for fermentation [30,31] and promoting growth of specific bifidobacteria [32], bacteroides [33] and Lactobacillus [34] species (reviewed in [35]).

We hypothesize that nutrition is a central regulator of host–microbe interactions in early life. As noted above, the composition of the microbiota of BF and FF infants differs in terms of overall diversity as well as composition [12,13,18]. Epidemiological studies have demonstrated that human milk protects against common infectious diseases in infancy (otitis media, respiratory syncytial virus, urinary tract infection), necrotizing enterocolitis (NEC) in preterm infants as well as immune-mediated disorders in later childhood, including allergy, asthma, atopic dermatitis, inflammatory bowel disease, Celiac Disease, Type 1 and Type 2 diabetes mellitus, and leukemia (ALL and AML) [36]. Recently, Walker [37] proposed that a diverse balanced microbiota is necessary for the development of an appropriate innate and adaptive immune response. This is further supported by studies associating dysbiosis in early life with immune-mediated childhood disorders [38–40] and obesity [41,42]. Dysbioses can arise from common pediatric practices, including preterm delivery, formula feeding, cesarean section, and use of antibiotics [42,43] (Fig. 1). Interestingly, cesarean section [43] and antibiotic use [44] are independently associated with an increased susceptibility to immune-mediated disease, potentially through dysregulation of host immune homeostasis [44,45]. It is important to note that all of these practices are amenable to changes in clinical protocols, and, as such, should be a priority for pediatric practice.

Given the evidence that early life nutritional exposures program long-term health outcomes, potentially through host–microbe interactions, our research group set out to systematically integrate genomic data from both the infant (host mucosa) and gut microbiota in order to define host gene–diet interactions within the context of the structure and operations of gut microbial communities. Until recently, no investigators had comprehensively profiled intestinal gene expression during early postnatal development due to limited availability of intestinal tissue from healthy infants. Thus, the potential for exfoliated epithelial cells to provide a non-invasive readout of intestinal gene expression was investigated [46–48].

2. Use of exfoliated cells to assess host gene expression

Each day, ~1/3rd to 1/6th of normal adult epithelial cells are shed [49], which corresponds to ~10 billion (10^10) cells per day. Exfoliation of intestinal epithelial cells from the villus tips in the small intestine and crypt surface in the colon is a form of programmed cell death induced by anchorage-dependent cells detaching from the surrounding extracellular matrix. Detachment also induces autophagy, which is a survival mechanism to loss of nutrients [51]. The exfoliated cells enter into a quiescent state and appear to maintain viability for differing lengths of time depending on the sources of cells. For example, quiescent exfoliated epithelial cells without signs of apoptosis were recovered in gastric fluid aspirates obtained from preterm infants [52]. Furthermore, exfoliated quiescent epithelial cells can be cultured, evidenced by the ability to use exfoliated cells to form lumen in 3-dimensional epithelial cell culture [47,48], suggesting that detachment-induced autophagy contributes to the viability of these cells.

This vast reservoir of host cells generated by exfoliation sparked interest from both basic and clinical translational investigators due to their potential utility to non-invasively assess cellular markers of gastrointestinal disease, predominantly colon cancer [53,54]. Subsequently, exfoliated epithelial cells had been used as sentinels of in vivo exposure to nutritional regimens [55,56] or as markers of disease states, including cancer in adults [57,58] and children with...
inflammatory bowel disease [59]. More recently, the feasibility of identifying protein markers and amplifying genes by polymerase chain reaction was demonstrated in exfoliated gastric cells from preterm infants [52,60]. Kaeffer and colleagues studied isolated exfoliated cells obtained from gastric aspirates [52,60] and stool samples [52] obtained from preterm infants. The gastric exfoliated cells were confirmed to be of epithelial origin by cytokeratin 18 expression and mRNA for beta-actin, clock genes and SLC26-A7-negative cells were positive for H+/K+-ATPase. Importantly, when glucocorticoid receptor (NRC31), as glucocorticoids are known that human milk protects preterm infants from the development of NEC, one of the most common causes of morbidity and mortality in very low birth weight infants [62,63]. We postulate that induction of this gene may provide a mechanism whereby the intestines of premature infants fed human milk are less likely to tolerate episodes of hypoxia and are thereby less likely to develop NEC. In addition to being the best single classifier, EPAS-1 was an even stronger predictor when combined with other genes involved in cell proliferation and apoptosis [46]. Interestingly, the most highly ranked gene for identifying BF versus FF infants was Endothelial PAS-domain-containing protein 1 (EPAS-1; also known as Hypoxia-inducible factor-2α [HIF-2α]). It is well known that human milk protects preterm infants from the development of NEC, one of the most common causes of morbidity and mortality in very low birth weight infants [62,63]. We postulate that induction of this gene may provide a mechanism whereby the intestines of premature infants fed human milk are less likely to tolerate episodes of hypoxia and are thereby less likely to develop NEC. In addition to being the best single classifier, EPAS-1 was an even stronger predictor when combined with other genes in ten separate 2- or 3-gene combinations (an example is shown in Fig. 2), thus providing potential biomarkers for nutritional modulation of gut development [46].

### 3. Stool-derived eukaryotic mRNA can be used to non-invasively assess the impact of nutrition on intestinal gene expression in term infants

In a proof-of-principle study, stool samples containing exfoliated host cells and luminal bacteria were collected from 3-month-old exclusively BF or FF infants [46]. A total of 1214 genes were significantly differentially-expressed between BF and FF infants, however, we focused our analyses on the 146 genes that were included in a list of 529 genes that we had a priori hypothesized could be differentially expressed based on prior knowledge. Analysis of gene networks reflected broad differences with respect to Signal Transduction (WNT, NOTCH, TGF-β), Cytoskeletal Remodeling; Cell Adhesion and Immune Response [46]. Linear Discriminant Analysis (LDA) was used to identify genes that best “classified” or discriminated BF from FF infants and the top up- and down-regulated genes and their fold-changes are shown in Table 2. Although these human milk-regulated genes were identified in term infants, they could have particular importance to infants with impaired gut development and function, including preterm infants. These genes included glucocorticoid receptor (NRC31), as glucocorticoids are key contributors to gut maturation, ZO1 (TJP1) a critical tight junction protein, and a number of genes encoding proteins involved in cell–cell interactions, including integrins, cadherins and syntaxin, and proteins involved in cell proliferation and apoptosis [46]. Interestingly, the most highly ranked gene for identifying BF versus FF infants was Endothelial PAS-domain-containing protein 1 (EPAS-1; also known as Hypoxia-inducible factor-2α [HIF-2α]). It is well known that human milk protects preterm infants from the development of NEC, one of the most common causes of morbidity and mortality in very low birth weight infants [62,63]. We postulate that induction of this gene may provide a mechanism whereby the intestines of premature infants fed human milk are less likely to tolerate episodes of hypoxia and are thereby less likely to develop NEC. In addition to being the best single classifier, EPAS-1 was an even stronger predictor when combined with other genes in ten separate 2- or 3-gene combinations (an example is shown in Fig. 2), thus providing potential biomarkers for nutritional modulation of gut development [46].

### 4. Stool-derived eukaryotic mRNA and intestinal microbiota DNA can be used to non-invasively evaluate host–microbe interactions in the intestine of term infants

In an extension of this work, we created a novel methodology designed to assess the multivariate relationship between the

![Table 1](image-url)
### Table 2

Early nutrition differentially regulates epithelial gene expression. Listed are the most highly up- and down-regulated genes in breastfed (BF) versus formula-fed (FF) infants.

<table>
<thead>
<tr>
<th>Gene symbol and name</th>
<th>Function of encoded gene</th>
<th>Fold change (BF/FF)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Up-regulated in BF compared to FF</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NR3C1; glucocorticoid receptor</td>
<td>Transcription factor that binds to glucocorticoid response elements in the promoters of glucocorticoid responsive genes</td>
<td>5.5</td>
</tr>
<tr>
<td>BAD; BCL-2-associated agonist of apoptosis</td>
<td>Protein positively regulates cell apoptosis by forming heterodimers with BCL-X and BCL-2, and reverting their death repressor activity</td>
<td>4.0</td>
</tr>
<tr>
<td>PCDH7; protocadherin</td>
<td>Gene product is an integral membrane protein that is thought to function in cell–cell recognition and adhesion</td>
<td>3.9</td>
</tr>
<tr>
<td>EPAS1; endothelial PAS domain protein 1</td>
<td>Transcription factor involved in the induction of genes regulated by oxygen, which is induced as oxygen levels fall</td>
<td>3.3</td>
</tr>
<tr>
<td>NRS5A2; nuclear receptor subfamily 5, group A, member 2 or liver receptor homolog-1 (LRH-1)</td>
<td>LRH-1 is important for maintaining pluripotency of stem cells during embryonic development</td>
<td>2.8</td>
</tr>
<tr>
<td>MYB; v-myb avian myeloblastosis viral oncogene homolog</td>
<td>Plays an important role in the control of proliferation and differentiation</td>
<td>2.8</td>
</tr>
<tr>
<td>Stx2; Syntaxin-2</td>
<td>Gene product regulates epithelial–mesenchymal interactions and epithelial cell morphogenesis and activation</td>
<td>2.5</td>
</tr>
<tr>
<td>ITG82; integrin beta-2 (CD18)</td>
<td>ICAM Receptor; gene product is known to participate in cell adhesion as well as cell-surface mediated signaling</td>
<td>2.5</td>
</tr>
<tr>
<td>TJF1; ZO-1</td>
<td>The N-terminal may be involved in transducing a signal required for tight junction assembly, while the C-terminal may have specific properties of tight junctions</td>
<td>2.2</td>
</tr>
<tr>
<td><strong>Down-regulated in BF compared to FF</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FOXE3</td>
<td>Forkhead box protein E3; cell proliferation</td>
<td>0.32</td>
</tr>
<tr>
<td>EGF</td>
<td>EGF Receptor; cell proliferation and migration</td>
<td>0.33</td>
</tr>
<tr>
<td>WNT7B</td>
<td>Wingless-type MMRV integration site family, 7B; cell fate and cell patterning</td>
<td>0.58</td>
</tr>
<tr>
<td>LYZ6</td>
<td>Lysozyme-like 6; bacteriostatic factor</td>
<td>0.70</td>
</tr>
<tr>
<td>HIF3A</td>
<td>Hypoxia-inducible factor 3 alpha; negative regulator of hypoxia inducible gene expression</td>
<td>0.71</td>
</tr>
<tr>
<td>DAPK1</td>
<td>Death-associated protein kinase 1; induction of apoptosis</td>
<td>0.73</td>
</tr>
<tr>
<td>APC</td>
<td>Adenomatous polyposis coli; tumor suppressor</td>
<td>0.77</td>
</tr>
<tr>
<td>TGFBI</td>
<td>TGFBI; transforming growth factor beta 1; cell growth</td>
<td>0.80</td>
</tr>
<tr>
<td>SPARC</td>
<td>Secreted protein, acidic, cysteine-rich (osteonectin); cell proliferation</td>
<td>0.82</td>
</tr>
</tbody>
</table>

From: Chapkin et al. [40].

---

microbiome metagenomic functional profile and the host transcriptome [47] as shown in Fig. 3. By examining the multivariate structure underlying the bacterial metagenome or metatranscriptome and gut exfoliated cell transcriptome, our approach leverages richer and fuller information content compared to analyses focusing on single data sets (e.g., only host transcriptome data or only bacterial metatranscriptome data) and only single variables (e.g., gene by gene differential expression testing). We propose that this “integrative” strategy will help identify intestinal genes that are responsive to diet and influenced by factors known to cause dysbiosis in term and preterm infants (e.g. route of delivery, antibiotic use, formula feeding, parenteral nutrition).

Using this approach, phyla-level differences in the microbiota of BF and FF infants were observed [47]. Microbiota functional characteristics were mapped to functional SEED categories. Because of the hierarchical structure of the SEED classification system, aggregating reads into coarser classifications provided for a more informed analysis. virulence was the one SEED category that differed between the bacterial metagenome of BF and FF infants [47]. To examine the intrinsic relationship between host and microbiome, host transcriptome and metagenomic data were combined and integrated using the multivariate technique of canonical correlation analysis (CCA) [47]. We examined whether a relationship existed between microbiota virulence genes and sets of host immunity and defense genes (n = 660), intestinal biology genes (n = 660) or a random genes (n = 459). A robust multivariate structure relating microbiota virulence genes and host immunity and defense genes was observed. Seven of the top eleven immunity and defense host genes that were related to the microbiota were down-regulated in BF versus FF infants, including ALOX5, a lipoxygenase involved in arachidonic acid and leukotriene synthesis, the cytokine IL1α, and binding proteins for natural killer cells (KLRF1), T-lymphocytes (AOC3) and LPS (BPI1) [47]. These findings indicate that the overall impact of breastfeeding was to reduce inflammatory genes in the gut potentially promoting tolerance to the luminal microbes.

### 5. Stool-derived eukaryotic mRNA can be used to non-invasively assess the impact of preterm delivery on intestinal gene expression

We next applied our exfoliated cell transcriptomic approach to the preterm infant [48]. Preterm birth, which affected 15 million...
children in 2010, is a major determinant of neonatal morbidity and is the second leading cause of death in children under 5 years [64,65]. Infants born at <32 weeks gestational age (GA) are faced with an unique set of challenges due to their developmental immaturity. In the U.S., the costs associated with preterm birth were more than $26.2 billion in 2005 [66]. Emerging evidence has clearly demonstrated direct and interactive links between diet, the intestinal microbiome and immune development [67,68]. Many common diseases afflicting preterm infants are associated with dysregulated immune function [69–73].

To determine whether sufficient exfoliated epithelial cells could be isolated from preterm infants and whether they would inform developmental differences, host transcript abundance in healthy full term (>38 weeks GA) and extremely preterm (24–30 weeks GA) infants were measured using RNA-Seq [48]. Approximately, 5500 genes were detected (FPKM > 1) on average in both preterm and term samples. Several key observations were made. First, gene expression in preterm infants was more heterogeneous amongst themselves and compared to term infants. Second, exfoliated cells express genes associated with specific intestinal cell types including absorptive enterocytes (lactase and sucrose–isomaltase), Goblet cells (mucin-2), enteroendocrine cells (chromogranin A), and Paneth cells (lysozyme). Lastly, the transcriptional landscape is dramatically altered in the preterm versus term infant intestine [48].

Gene pathways that were over-expressed in preterm versus term or term versus preterm intestine were evaluated. Although none of the infants were clinically ill at the time the stool samples were collected, preterm over-expressed genes related to immune function. Several cytokines, including IL-1α and IL-33 were up-regulated in preterm versus term. In addition, several genes that regulate the expression of cytokines and other immune genes were expressed at 3- (NFKB1a) to 6-fold (CASP1) higher levels in preterm versus term infant exfoliated cells [48]. This is consistent with...
which is a hallmark of NEC [76]. The excessive inflammatory response of the immature intestine appeared to be in part due to a developmental under-expression of IkB coupled with over-expression of the inflammatory genes (SIGIRR, IRAK-M, A-20 and TOLLIP) in fetal intestine relative to other children [75]. Thus, it appears that immaturity of the intestinal innate immune response may contribute to excessive inflammation in the intestine in response to colonizing bacteria, which is a hallmark of NEC [76].

In contrast, in term infants, up-regulated immune genes were involved in balancing the immune system, e.g., promoting T-cell development (LCP2; 3.6-fold greater than preterm), while inhibiting macrophage activation (LENG9; 16-fold greater than preterm). The majority of genes were involved in cell turnover, by regulating proliferation and apoptosis. One of the most highly differentially-expressed genes was an anti-apoptotic factor (MTRNR2L6; 5-fold higher in term than preterm). Another interesting gene was SP3 (~2-fold higher in term than preterm), which is a transcription factor that can be regulated through short-chain fatty acid induced acetylation [77,78], potentially supporting the role of products of microbial metabolism in regulating normal gut growth in term infants.

The underlying reason for differences in intestinal gene expression between preterm and term infants is multifactorial and may be a result of the developmental immaturity of the preterm gut, coupled with specific environmental exposures that are unique to the post-natal course of this population. Although well-controlled studies are needed to evaluate these exposures, this study highlights the potential of using the described noninvasive technology. We anticipate that this approach will allow investigators to elucidate how diet and bedside clinical management of this immature population influences intestinal development and immune ontogeny over time and has the potential for generating comprehensive, diagnostic gene sets for the noninvasive identification/prediction of different intestinal phenotypes in infants.

In addition to the host gene networks that are affected by the developing microbiota, the metabolic products from both microbes and host can give rise to signaling and inflammatory pathways that can affect numerous organs such as the lung, liver and brain in addition to the intestinal tract [73]. Undoubtedly, the comprehensive analysis of microbiota, intestinal transcriptome and metabolites will provide an overall picture of the milieu that is associated with many of the major morbidities seen in preterm infants.

6. Eukaryotic mRNA-derived for meconium can be used to noninvasively assess the impact of maternal probiotic treatment on fetal intestinal immune gene expression in uterus

The long-held belief that the fetus existed in a sterile intrauterine environment and that introduction of bacteria into the amniotic fluid would lead to an adverse pregnancy outcome has been challenged by recent studies showing that the placenta, amniotic fluid and fetus contain a diversity bacteria and that these differ between term and preterm infants [79,80]. Although dysbiosis may trigger preterm delivery [11,81], other studies have described a relatively complex placenta microbiota comprised of nonpathogenic commensal microbiota from the Firmicutes, Tenericutes, Proteobacteria, Bacteroidetes, and Fusobacteria phyla [9,79,80] that, surprisingly, clusters more similarly to the maternal oral than vaginal or gut microbiota [79,82]. These results suggest that the placental and intrauterine bacteria do not ascend from the vagina, but may be delivered through the circulation [82]. Rautava [10] administered probiotics in order to investigate whether microbes in placenta or amniotic fluid affected fetal innate immune gene expression during late pregnancy and whether innate immune gene expression profiles in the placenta and the fetal gut may be modulated by dietary supplementation with specific probiotics. In a double-blind clinical trial, pregnant women were administered either placebo, Bifidobacterium lactis (B. lactis) or B. lactis and Lactobacillus rhamnosus GG (LGG) for 14 days before elective cesarean section at full term [10]. Bacterial DNA was detected in all placenta samples by PCR. Meconium samples were collected and host mRNA extracted from fetal exfoliated epithelial cells using the method of Chapkin and colleagues [46]. An association between the presence of microbial DNA in amniotic fluid and placenta and changes in toll like receptor (TLR)-related gene expression in the fetal intestine was observed. Additionally maternal probiotic supplementation significantly modulated the expression of TLR-related genes both in the placenta and in the fetal gut. Compared to mRNA expression in exfoliated cells of infants exposed to the placebo, TLR6 mRNA expression was down-regulated nearly 90% in exfoliated cells from infants whose mothers consumed B. lactis + LGG, whereas TLR7 was down-regulated 70% in infants of mothers administered B. lactis alone [10]. TLR are important mediators of innate immunity through their recognition of highly conserved microbial-associated molecular patterns. Specifically, TLR6 interacts with TLR2 to mediate cellular response to bacterial lipoproteins. TLR7 recognizes single-strand RNA in endosomes, which is a common feature of viral genomes [83]. These findings support the hypothesis that microbial programming begins in fetal life through host–microbe interactions in utero [8], which can be manipulated by maternal probiotic intervention [10].

7. Summary

We have recently validated a novel molecular methodology that utilizes stool samples containing intact sloughed epithelial cells to noninvasively quantify intestinal gene expression profiles in the developing human neonate, which is “an important first step towards a more comprehensive understanding of the biological mechanisms underlying the parallel development of the host and microbiome in early life” [84]. This approach enables repeated assessment of the same infant overtime to assess temporal changes in gene expression [40]. Furthermore, we have expanded upon this methodology by combining host gene expression with the bacterial metagenome from the same infant [47]. This approach is statistically rigorous and is sensitive to dietary intake [46,47] and the stage of gestation [48]. In conclusion, we propose that the investigation of host–microbiome interaction will fill an important gap in our understanding of the coordinated development of gut microbiota and the infant intestine. In the long-term, it is anticipated that nutritional strategies to improve the development of the microbiome and intestine will enhance the clinical care of high-risk infants.

References


