Improving Diagnostic Precision: Urine Proteomics Identifies Promising Biomarkers for Necrotizing Enterocolitis

Stephen Mackay¹, Lauren C. Frazer¹, Grace K. Bailey¹, Claire M. Miller¹, Qingqing Gong², Olivia N. DeWitt², Misty Good¹*¹

¹Division of Neonatal-Perinatal Medicine, Department of Pediatrics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 27517

²Division of Newborn Medicine, Department of Pediatrics, Washington University School of Medicine, St. Louis, MO 63110

Correspondence:
Misty Good, MD, MS
Division Chief, Neonatal-Perinatal Medicine
Department of Pediatrics
The University of North Carolina at Chapel Hill
101 Manning Drive Campus Box 7596
Chapel Hill, NC 27599
Office: 984.974.5063, Fax: 984.974.7857
Email: mistygood@unc.edu

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.
Abstract

Background: Necrotizing enterocolitis (NEC) is a severe intestinal disease that primarily impacts preterm infants. Current diagnostic tools are inadequate, so urine proteomics was performed for patients with and without NEC to identify putative biomarkers.

Research design and methods: The abundance of urinary proteins detected using an aptamer-based microarray was compared for infants with NEC (n=20) and controls, age-matched (n=8) or self-matched (n=12). Spearman r correlation and hierarchical cluster analysis were performed. The area under the curve (AUC) was calculated for receiver operator characteristic curves (ROC).

Results: Ninety-nine proteins differed in NEC vs. controls based on median fold change (Log$_2$ ± 1.1) and significance (P < 0.05). Patterns of abundance were consistent for both types of matching, and samples clustered based on NEC severity. Two panels were built to differentiate between infants with and without NEC. Panel 1 included proteins associated with inflammation/NEC and produced by the intestinal epithelium (REG1B, REG3A, FABP2, DEFA5, AUC 0.90). Panel 2 consisted of proteins with the largest fold change between NEC vs. controls and the highest individual AUC values (REG1B, SSBP1, CRYZL1, ITM2B, IL36B, IL36RN, AUC 0.98).

Conclusions: Urine proteins significantly differ between infants with and without NEC, which supports their potential as future biomarkers.

Keywords: Aptamer-based microarray, Biomarker, Necrotizing enterocolitis, Neonatal Intestine, Preterm, Urine Proteomics.
Background

Necrotizing enterocolitis (NEC) is a rapidly progressive intestinal disease that occurs primarily in preterm and low birth weight neonates and has an incidence of ~7% for neonates born at < 32 weeks gestation [1]. The first symptoms of NEC, such as feeding intolerance and abdominal distension, can be difficult to distinguish from other etiologies; however, within hours of disease onset, an infant with NEC can progress to needing emergency surgery or death. The rapidity of this clinical decompensation is why an early and accurate diagnosis of NEC is essential. Currently, NEC is diagnosed using imaging modalities such as abdominal X-rays and/or ultrasound combined with physical exams and laboratory tests, but these tools have inadequate sensitivity and specificity, particularly early in the disease course [2]. Unfortunately, no biomarkers for NEC have been implemented in routine clinical practice, which could improve upon some of the current diagnostic limitations.

Identification of biomarkers for NEC has been hindered by a multitude of factors, with the primary barrier being an incomplete understanding of disease pathogenesis. NEC is thought to result from a poorly controlled immune response in the neonatal intestine in the setting of a dysbiotic microbiome. This inflammation leads to intestinal epithelial injury, loss of mucosal barrier integrity, and, ultimately, tissue necrosis [3-6]. The precise factors that induce and exacerbate disease remain an area of active investigation, and that knowledge barrier continues to impede biomarker development.

Several potential biomarkers have been identified for NEC, including proteins/peptides [7,8], metabolites [9], and non-coding RNAs [10,11], but reproducibility and validation of the biomarkers in large cohorts have not been conclusive. In addition, some urinary biomarkers for
NEC have been identified, such as fatty acid binding protein-2 (FABP2) [12]; however, these markers are not in use clinically, as normalization of urine samples has been problematic [12].

Optimally, a biomarker for NEC would identify infants early in the disease process and reliably differentiate between neonates with NEC and other intestinal or systemic inflammatory pathologies. In addition, an ideal biomarker would be validated for use with biological specimens that are readily available and non-invasively acquired, given the difficulties associated with acquiring specimens from medically fragile preterm neonates. To overcome the difficulty in obtaining a large number of infants with NEC at one center, we established the multi-center NEC Biorepository to identify biomarkers of disease [13,14].

With improvements in technology, the identification of new biomarkers for NEC is now a distinct possibility. Traditional methods for biomarker identification include quantitative and qualitative analysis of proteins and metabolites using untargeted gas or liquid chromatography coupled with tandem mass spectrometry (LC-MS/MS). Recently, aptamer-based proteomic screens using large, highly selective libraries have been utilized to screen samples for biomarkers in various tissue and biological specimens, including serum, stool, and urine [15-17]. Compared to LC-MS/MS, aptamer-based scans are better able to detect small proteins [18], which presents the opportunity to detect new biomarkers for NEC.

In this study, we utilized aptamer-based proteomics technology (SomaScan®) to measure the relative abundance of over 7000 proteins in the urine of infants with NEC and controls. We identified ninety-nine proteins that differed between the groups and performed a detailed analysis of protein abundance, patterns across samples, disease severity, and diagnostic utility. In addition, we identified two protein panels and a pair of proteins that effectively differentiate between infants with NEC and controls. This study provides the
foundation for future work investigating urine biomarkers for NEC in a larger population of preterm infants and provides an advance toward identifying urgently needed biomarkers for this devastating disease.

Methods

Study design

Samples were prospectively collected from infants admitted to the St. Louis Children’s Hospital Neonatal Intensive Care Unit (NICU) in St. Louis, Missouri, USA. Enrollment of infants 22 to 42 weeks gestation was conducted according to protocols approved by the Washington University in St. Louis School of Medicine Institutional Review Board (IRB protocol numbers 201706182 and 201802101). Infants with major congenital anomalies were excluded. For the current study, we included infants initially enrolled as controls and then subsequently developed NEC (self-matched, n = 12). We also analyzed a cohort of age-matched infants (n = 8 pairs), which included infants enrolled at the time of NEC diagnosis (n = 8) and their age-matched controls (n = 8). Infants with urinary tract infections were excluded. The cohort was comprised of infants born between 24 and 36 weeks gestation (n = 28) (Table 1) and modified Bell’s staging was used to categorize NEC severity [19-21].

Sample collection

Urine samples were collected by placing sterile gauze in a clean diaper. Urine-saturated gauze was then collected at a routine care time, placed in a 10 mL syringe, and stored at 4°C until processed by laboratory staff 1-2 times per day. Urine was subsequently extracted from the
gauze by squeezing the syringe, aliquoted into cryotubes, and stored at -80°C until analysis.

Samples with visible stool contamination were discarded.

Proteomics assays

Sample preparation for SomaScan® and ELISA

Frozen urine samples (-80°C) were thawed in a water bath for 15 min at 37°C and mixed by pipetting before combining 216 µL urine with 9 µL of 1 M Tris-HCl to a final concentration of 40 mM. The samples, in two 100 µL volumes, were desalted by buffer exchange using a Zeba™ 7 kDa MWCO Spin column desalting plate (ThermoFisher Scientific, Waltham, MA) according to the manufacturer's guidelines. Desalting plates reached room temp before centrifuging at 1000 x g for 2 min to remove the storage buffer. A 250 µL volume of assay buffer was added to the resin beds and centrifuged at 1000 x g for 2 min. This was repeated three times. Protein concentration was determined using a Pierce™ Micro BCA™ protein assay kit (ThermoFisher Scientific, Waltham, MA) according to the manufacturer’s instructions.

SomaScan® proteomic assay

The differential abundance of proteins in urine samples was measured using an aptamer-based SomaScan® 7,000 protein microarray assay kit v4.1 (SomaLogic®, Boulder, CO, USA) as previously described [22]. Briefly, the microarray screened proteins using “Slow Off-Rate Modified Aptamers” (SOMAmers®) that were selected against protein epitopes using the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) [22]. The scan utilized a
resolution of 5 µm and detected Cy3 fluorescence expressed as relative fluorescence units (RFU). Off-scanner raw signal values were calibrated, standardized, and scaled at multiple dilution categories, namely S1 (61% of proteins), S2 (~38.5%), and S3 (0.4%), where S1 targets low abundant proteins, and S3 targets proteins close to saturation. RFU readouts were analyzed using the Agilent Feature Extraction v10.7.3.1 (Agilent Technologies, Santa Clara, CA). Differential abundance was calculated using the SomaScan® statistical analysis tool v4.1 and fitted as a linear model of the signal data and an empirical analysis by a Bayesian statistical test for group comparisons.

ELISA

Proteins selected for ELISA validation included REG1B (EH389RB, ThermoFisher Scientific, Waltham, MA), DEFA5 (HUFI01588, AssayGenie, Dublin, Ireland), FABP2 (EHFABP2, ThermoFisher Scientific, Waltham, MA), and REG3A (EH390RB, ThermoFisher Scientific, Waltham, MA). Urine samples were prepared, desalted, and quantified as described above. For each sample, triplicate measurements of 0.2, 2, 10, or 20 µg total protein were plated based on anticipated protein abundance to fit within the range of the standard curve for each ELISA.

Ingenuity Pathway Analysis

Network analysis and pathway enrichment were predicted using Ingenuity Pathway Analysis (IPA) (QIAGEN digital insights, Redwood City, CA, USA, accessed on June 20, 2023).
IPA was used to generate a core analysis using all 99 differentially abundant proteins (Table 2) based on differential median fold change expression (Log$_2$ ± 1.1) and Mann-Whitney U-test significance ($P < 0.05$) cut-offs.

**Statistical analysis**

Proteins above a Log$_2$ median fold change RFU of 1.1 and a statistical significance of $P < 0.05$ for the comparison between samples from infants with NEC and controls were selected for detailed analysis. Samples were analyzed for the influence of multiple comparisons using a false discovery rate (FDR) of < 0.05. Volcano plots, hierarchical cluster analysis, and Spearman’s $r$ correlation matrices were performed using Log$_2$ RFU median fold change values. Log$_{10}$ RFU values were used for comparative statistical analysis for violin plots and receiver operator characteristics (ROC). Statistical significance was determined by a Mann-Whitney U-test, non-parametric Wilcoxon test, or Kruskal Wallis as indicated. Combined ROC curves were generated using multiple logistic regression. Comparative statistical analysis was performed using GraphPad 9.3.1 (San Diego, USA). Hierarchical clustering was generated in R-studio version 4.2.0 using the “heatmap.2” package by Ward clustering.

**Results**

**Cohort characteristics**

The cohort included infants ($n = 28$) born between 24 and 36 weeks gestation and divided into age- and self-matched cohorts (Table 1). For the age-matched cohort, infants were
enrolled at the time of NEC diagnosis \((n = 8)\) and matched with infants who did not develop NEC (controls, \(n = 8\)). For the self-matched cohort, infants were initially enrolled as controls but then developed NEC \((n = 12)\). Samples obtained before NEC diagnosis were used as the control for this group, with six days being the shortest interval between a control sample and a diagnosis of NEC.

**Qualitative and quantitative proteomic analysis of urine proteins**

A total of 7596 proteins were screened in the urine samples using an aptamer-based proteomics microarray (SomaScan®). For the age-matched cohort, 150 urine proteins were significantly different between the infants with NEC and controls, based on Log$_2$ median fold change RFU of ± 1.1 and Mann-Whitney U-test of \(P < 0.05\) (Figure 1A). For the self-matched cohort, 360 urine proteins were differentially abundant in the control compared to the NEC samples (Figure 1B). There were 99 proteins shared between the age- and self-matched groups, and these will be the focus of the remainder of the analysis (Figure 1C). Of these 99 proteins, 35 were increased, and 64 were decreased between infants with NEC and their respective controls. A summary of the identified proteins in the combined cohort is provided in Table S1. Due to the small sample size of the cohort, only 10 proteins of the 7596 in the microarray survived multiple testing correction by false discovery rate (FDR < 0.05). Of the proteins that survived this correction, interleukin 36B (IL36B, FDR = 0.024) and interleukin 36 receptor antagonist (IL36RN, FDR = 0.024) were in the subset of proteins that were significantly different between the infants with NEC and controls.
We next identified two protein panels of interest for further analysis. Protein panel one consisted of four proteins chosen for their primary localization within the intestine and previously observed association with NEC and/or inflammation (Table 2). These include regenerating family member 1 beta (REG1B), regenerating islet-derived protein 3 alpha (REG3A), intestinal fatty acid binding protein 2 (FABP2), and defensin 5 alpha (DEFA5). Protein panel 2 consisted of proteins chosen based on the magnitude of the fold change between patients with NEC and controls, as well as the best area under the curve (AUC) values (Table 2). Panel 2 included REG1B (also in panel 1), single-stranded DNA binding protein 1 (SSBP1), crystallin zeta-like 1 (CRYZL1), integral membrane protein 2B (ITM2B), IL36B, and IL36RN.

We analyzed the association between the relative protein levels for the most differentially abundant proteins using an inter-protein Spearman’s $r$ correlation matrix (Figure 2A). With a focus on our protein panels, we found that for panel 1, DEFA5 and REG3A had the strongest correlation ($r = 0.83$) (Figure 2B). In addition, REG3A ($r = 0.57$) and DEFA5 ($r = 0.46$) had strong positive correlations to REG1B (Figure 2B). For protein panel 2, there was a strong negative correlation with IL36RN with REG1B ($r = -0.63$) (Figure 2C). In addition, SSBP1 correlated positively with ITM2B ($r = 0.62$) and negatively with CRYZL1 ($r = -0.52$) and IL36B ($r = -0.45$) (Figure 2C).

We next used hierarchical cluster analysis to visualize patterns in protein abundance across patient pairs and disease severity (Figure 3). We found that the pattern of relative fold change of the proteins was consistent across samples. In addition, age-matched and self-matched samples did not cluster together, but samples did cluster based on disease severity, with noted clustering of samples from patients with Bell’s stage III NEC (Figure 3).
Analysis of protein abundance and comparison based on disease severity.

Violin plots were used to visualize the protein abundance in urine samples for the proteins in panel 1, with significant differences noted between controls and patients with NEC (Figure 4A). We further analyzed the differences in urine protein abundance based on disease severity as defined by modified Bell’s stage (Figure 4B). REG1B was significantly increased in infants with NEC Bell’s stage II compared to controls ($P < 0.01$) as well as in Bell’s stage III vs. controls ($P < 0.01$). REG3A and DEFA5 were significantly increased in the urine of infants with NEC Bell’s stage III vs. controls (REG3A: $P < 0.0001$, DEFA5: $P < 0.0005$) and in Bell’s stage II compared to Bell’s stage III (REG3A: $P < 0.01$, DEFA5: $P < 0.01$). FABP2 was only significantly increased Bell’s stage III vs. controls ($P < 0.05$).

ELISAs were used to validate the relative protein abundance detected with the protein microarray. We selected the proteins in panel 1, REG1B, REG3A, FABP2, and DEFA5 (Supplementary Figure S1) for validation. REG1B (Figure S1A) and DEFA5 (Figure S1C) were both significantly increased in NEC relative to controls; however, REG3A (Figure S1B) and FABP2 (Figure S1D) were not significantly different in NEC compared to controls (Supplementary Table S2). This likely occurred because levels of the urine proteins were below the limit of detection for the ELISAs for some patients, which resulted in an inability to detect differences between patients with NEC and controls. These assays were limited by the volume of urine obtainable from preterm neonates and the low concentration of the proteins of interest. Thus, the microarray data was used for the remainder of the analysis due to increased sensitivity at lower protein concentrations.
Analysis of the ability of these proteins to serve as potential biomarkers was performed by calculating the AUC (Table 2 and Figure 4C). Values above 0.7 are considered acceptable, and above 0.8 are considered excellent for a diagnostic test. REG1B (0.82, 95% CI 0.69-0.95) and REG3A (0.79, 95% CI 0.69-0.95) were the most effective at differentiating between patients with NEC and controls (Figure 4C).

We performed a similar analysis for the proteins in panel 2. The RFU data for the urine samples from patients with NEC and controls is shown in Figure 5A. We also analyzed the ability of these proteins to differentiate between patients based on NEC severity (Figure 5B). We found that CRYZL1, ITM2B, IL36B, and IL36RN were significantly different between infants with Bell’s stage II NEC relative to controls ($P < 0.05$). SSBP1 ($P < 0.005$) and IL36B ($P < 0.05$) were significantly different between infants with NEC Bell’s Stage II and III.

ROC curves were generated for all proteins in panel 2. All of the proteins in panel 2 were effective at differentiating between infants with NEC and controls (Figure 5C). IL36B (0.86, 95% CI 0.74-0.98) and IL36RN (0.91, 95% CI 0.82-0.99) had the best AUC values and have the potential to serve as excellent diagnostic tests.

Diagnostic utility of protein panels

We generated combined ROC curves for groups of proteins to determine if this would improve their diagnostic utility as urine biomarkers (Figure 6). We found that protein panel 1 (Figure 6A, AUC 0.9, 95% CI 0.8-1), protein panel 2 (Figure 6B, AUC 0.98, 95% CI 0.95-1), and the combination of IL-36B and IL36RN (Figure 6C, AUC 0.94, 95% CI 0.86-1) had an outstanding ability to differentiate between infants with NEC and controls.
Upstream regulators and statistical enrichment

Ingenuity pathway analysis (IPA) was used to predict activated or inhibited upstream regulators and signaling pathways. IPA predicts differentially activated regulators measured by downstream enrichment scores (overlap $P$-value) and activation states (activation z-score) [23].

Upstream regulator analysis (URA) predicts direct relationships between upstream regulators and downstream molecules using the protein dataset. Causal analysis (CNA) tools further predict upstream master regulators that may have multiple indirect relationships, nodes, and links. Upstream regulators of interest (linked to inflammation) by URA and CNA were predicted and summarized in Table 3.

URA predicted several direct regulators based on the downstream status of the differentially abundant proteins identified in the urine samples. This includes regulators of lipid metabolism, sterol regulatory element binding protein 1 (SREBF1, $z = 1.95$, $P = 0.00015$), and retinoic acid receptor alpha (RXRA, $z = 1.94$, $P = 0.00794$) (Table 3). In addition, the polyfunctional cytokine IL-22 ($z = 1.96$, $P = 0.00091$) was predicted as activated based on the differential abundance of REG1B (+), REG3A (+), LBP (+), and TGM3 (-) (Supplementary Figure S2, Table 3).

CNA includes indirect relationships and linked nodes to identify upstream-master regulators (Table 3). Predicted master regulators included: TNF receptor-associated factor 1B (TRAF2) ($z = 2.353$, network bias corrected $P = 0.0003$), which plays a role in NFκB and JNK signaling pathways [24], mRNA decay activator protein ZFP36 ($z = 2.263$, network bias corrected $P = 0.0153$), which is involved in cytokine degradation and regulates the TNF-alpha feedback loop through mRNA destabilization [25,26], Endoribonuclease ZC3H12A ($z = 2.191$,
network bias corrected $P = 0.043$), and IL-22 ($z = 2.0$, network bias corrected $P = 0.0171$).

STAT3, which is activated by IL-22 and other cytokines, was upstream of several differentially abundant proteins in this study (Supplementary Figure S2, Table 3).

Discussion

In this study, we used an aptamer-based protein microarray to successfully identify a subset of urine proteins that differentiated between infants with and without NEC. The validity of our data was supported by the fact that patterns of relative protein abundance in the urine were consistent across both age- and self-matched cohorts. In addition, the protein levels in the urine of these patient cohorts clustered based on the severity of NEC. Our identified biomarker panels had excellent diagnostic potential based on AUC values, supporting further studies investigating these proteins of interest as potential biomarkers.

Our study focused on urine as the biological source for biomarker testing, given that it is an easily obtainable specimen even for tiny preterm neonates. Unfortunately, the detection of biomarkers in urine adds additional analytical challenges due to sample-to-sample biological variability. This includes significant variability in hydration states, glomerular filtration rates (GFR), and differential abundance in carrier proteins and metabolites [27]. The accurate quantitation of differentially abundant proteins measured in urine is likely improved using a reference marker. Typical urine reference markers require a constitutively expressed gene or metabolite unaffected by the disease. Studies have tried to identify a reliable normalization reference using urinary proteins or peptides to adjust for urinary dilution [7]. These have included serum albumin, collagen, and urine creatinine; however, protein and metabolite
variability based on the developmental age of neonates complicates normalization [28]. For example, urine albumin and creatinine concentrations in neonates have been shown to be inconsistent in preterm infants [29]. The consistency of our protein abundance patterns with two different patient matching strategies and across a range of corrected gestational ages indicates that these technical issues were likely minimized by using equivalent protein loading, as was performed in our studies.

Based on published data, the likely intestinal origin of many of the proteins we detected provides another degree of complexity, as it requires proteins or peptides to enter the bloodstream from the intestine and be filtered by the glomeruli in the kidney to be identified in the urine. In addition, biomarkers need to distinguish NEC from non-specific inflammation in the kidneys and urinary tract to have clinical significance. In this study, known biomarkers for urinary tract infections [30] and acute kidney injury [31], specifically NGAL (Lipocalin-2) and CST3 (Cystatin C), were not significantly abundant. In addition, the consistency of protein abundance patterns across matching strategies and the clustering of patients based on disease severity supports a NEC-specific protein signature independent of renal complications.

In this study, we developed two protein panels and identified a pair of proteins with excellent AUC values for diagnostic studies. Using a protein panel increases the power of a diagnostic test to accurately discriminate between healthy and diseased states. Although a single disease biomarker would be optimal, this is likely unfeasible given the complexity of the pathogenesis of NEC. Below, we discuss the specific proteins we incorporated into our panels.

REG1B and REG3A
Regenerating islet-derived proteins (REGs) have multifunctional roles in cellular processes, including promoting differentiation, increasing cellular proliferation, preventing apoptosis, and enhancing host defense [32]. REG proteins (18-19 kDa) have been discovered in multiple tissues and diseases, and they are typically expressed at sites of inflammation [33]. REG1B and REG3A recognize bacterial peptidoglycan and reduce bacterial membrane integrity [34]. They are increased in inflammatory bowel disease (IBD), specifically ulcerative colitis (UC) [35]. REG1B expression is also increased in the intestine during NEC [36]. In addition, increased REG1B expression drives intestinal crypt regeneration [37]. REG3A has anti-apoptotic effects and increases enterocyte viability by reducing caspase cleavage [38]. REG1B [39] and REG3A [40] have been identified as potential stool biomarkers for enteropathy and environmental enteric dysfunction (EED). Finally, REG1B and REG3A are induced by the IL-22 signaling pathway [41], which is protective in the intestine of mice during experimental NEC [42].

FABP2

Fatty acid binding protein 2 (FABP2) is a 15 kDa enterocyte-specific carrier protein involved in intracellular fatty acid translocation across the gut barrier [43]. FABPs are subdivided into intestinal (FABP2), hepatic (FABP1), and adipose (FABP4) variants based on localization. FABP2, FABP1, and FABP4 were all differentially abundant in our proteomic assay (Supplementary Table 1). FABP2 has previously been identified as differentially abundant during intestinal barrier dysfunction [44] and, specifically, NEC in several biomarker studies of stool and urine [45]. A clinical metadata study showed a high FABP2 level was an indicator of NEC, but many studies were limited in sample size and had difficulty consistently normalizing urine samples by creatinine [12].
DEFA5

Human α-defensin (DEFA5) (10 kDa) is a powerful antimicrobial host defense protein [46]. It is primarily expressed by intestinal Paneth cells and stored intracellularly as a zymogen. After secretion, DEFA5 is activated by trypsin and regulated by Paneth cell-derived serine protease inhibitors, which play a role in selectively maintaining gut microbiome diversity. One study has linked increased expression with urinary tract infections [47]. Increased DEFA5 has previously been detected in the stool of infants with NEC via proteomics [48] and in intestinal samples resected from infants with NEC via RNA sequencing [49]. In addition, altered stool DEFA5 methylation patterns have been observed during NEC [50].

SSBP1

Single-stranded DNA binding protein 1 (SSBP1) is a ubiquitously expressed protein involved in mitochondrial biogenesis. Mitochondrial damage, through the presence of mitochondrial DNA and proteins in urine, has been used as a marker for tissue damage during inflammation [51]. In addition, SSBP1, through heat shock factor 1 (HSF1) signaling, has been linked to increased expression of heat shock protein 70 (Hsp70) [52]. Hsp70 protects cells from proteolytic stress and regulates pro-inflammatory Toll-like receptor (TLR)-induced cytokines [53]. In enterocytes, activation of TLR4 leads to upregulation of Hsp70, and Hsp70 signaling has been identified as a potential counter-regulatory mechanism that opposes TLR4 signaling during NEC [54].

CRYZL1

Crystallin zeta-like 1 (CRYZL1) is ubiquitously expressed and shares sequence homology with zeta crystallin or quinone oxidoreductase 1 (NQO1). While the role of CRYZL1 is
unclear, NQO1 protects against cellular damage by acting as a redox switch interacting with multiple proteins, mRNA, and NADH [55]. Upregulation of Nqo1 expression has been associated with protection from NEC-like injury in rats [56].

*ITM2B*

Integral membrane protein 2B (ITM2B) is found in all tissue types. Its exact function needs to be delineated, although it is thought to play a role in apoptosis, mitochondrial homeostasis, and processing of amyloid precursor proteins [57].

*IL36B and IL36RN*

Interleukin 36 (IL-36) proteins are a family of inflammatory cytokines and consist of agonists IL36 (α, β, γ) and a receptor antagonist, IL36RN (IL36Ra). IL-36 dysregulation is associated with inflammatory diseases, including psoriasis, arthritis, and IBD [58]. In addition, IL-36 regulates both immune and non-immune cells in a feedback loop where IL-22, IL-17A, and TNFα induce all three IL-36 family cytokines, and IFNγ is selective for inducing IL36B [59]. Interestingly, in mice with NEC, increased IL36B was detected only in the ileum, and an increased IL36B: IL36RN ratio was associated with NEC [60]. In our study, the relative decrease in IL36RN is consistent with increased inflammation, whereas the reduction in IL36B is not expected and needs further exploration in mechanistic studies.

*Upstream regulator network analysis*

IPA identified several potential upstream and mechanistic regulators. Notably, IL-22 was predicted to be an upstream regulator based on the downstream differential abundance of REG1B, REG3A, LPS-binding protein (LBP), and transglutaminase 3 (TGM3). IL-22, through
STAT3-dependent signaling [61], is involved in intestinal barrier protection and induces proliferation in intestinal stem cells and goblet cell proliferation in the crypts [62]. Of note, STAT3 is activated by cytokines other than IL-22, such as IL-6 and IL-10 family cytokines [63]. The redundancy in these pathways makes it difficult to pinpoint the upstream mediators without mechanistic data; however, the increased abundance of these intestinal proteins in the urine of infants with NEC indicates a potential increase in signaling via STAT3 activating cytokines and/or is reflective of a weakened gut barrier in patients with NEC. Additionally, hepcidin (HAMP), which had a 3-fold increased abundance in the samples from infants with NEC, is an antimicrobial iron sequestering protein and is also positively regulated by IL-22 [64]. Moreover, IL-22 is a complex and polyfunctional cytokine. We have previously shown that administration of recombinant IL-22 is protective against intestinal injury during experimental NEC in mice [42]; however, further studies are needed to determine the balance between protective and inflammatory roles for this cytokine in NEC.

Conclusion

Due to the rapid and unpredictable onset of NEC, a biomarker needs to be readily available, non-invasive, rapidly tested, and able to identify NEC prior to the advanced Bell’s stages. In this study, we identify proteins of interest that have the potential to serve as biomarkers that are so urgently needed to improve the care of infants with NEC. While our results need to be validated in a large cohort of infants with diverse physiologic and inflammatory conditions, this study will serve as an important foundation for future studies on a larger, diverse cohort.
List of abbreviations

Necrotizing enterocolitis (NEC), receiver operator characteristics (ROC), area under the curve (AUC), relative fluorescence units (RFU), false discovery rate (FDR), Ingenuity pathway analysis (IPA), Upstream regulator analysis (URA), and Causal analysis (CNA).

Declarations

Ethics approval and consent to participate

After consent was obtained, samples were prospectively collected from infants admitted to the St. Louis Children's Hospital Neonatal Intensive Care Unit (NICU) in St. Louis, Missouri, USA. Enrollment was conducted according to protocols approved by the Washington University in St. Louis School of Medicine Institutional Review Board (IRB protocol numbers 201706182 and 201802101).

Funding

LCF is supported by a Thrasher Research Fund Early Career Award, UNC School of Medicine Physician Scientist Training Program Faculty Award, and UNC Children's Development Early Career Investigator Grant through the generous support of donors to UNC. MG is supported by National Institutes of Health (NIH) grants R01DK124614, R01DK118568, R01HD105301, R44HD110306, the Chan Zuckerberg Initiative Grant number 2022-316749, and the University of North Carolina at Chapel Hill Department of Pediatrics.
Availability of data and materials

All data generated or analyzed during this study are included in this published article and its supplementary information files.

Disclosure statement

The authors report that there are no competing interests to declare.

Authors contributions

Manuscript preparation by SM, LF, CM, and MG. Critical review of the manuscript by all authors. Data analysis by SM and LF. Research coordination and sample processing by QG, OD, GB, and MG. Study design by SM, LF, and MG.

Acknowledgements

We want to thank the neonates and families who participated in our study. We would also like to thank the staff at St. Louis Children’s Hospital Neonatal Intensive Care Unit in St. Louis, Missouri, USA, for all their hard work in helping to collect samples, which made this study possible.
Figure Legends

Graphical abstract. Overview of study findings. Created with Biorender.com

Figure 1. Volcano plots depicting the relative abundance of urine proteins in infants with and without NEC. Urine protein levels were compared between infants with and without NEC in the (A) Age-matched cohort (n = 8 pairs), (B) Self-matched cohort (n = 12 pairs), or (C) both groups combined (n = 20 pairs). Differentially abundant proteins are presented on a volcano plot with a median fold change of RFU values (Log₂ ± 1.1) and P-value (−Log₁₀ Mann Whitney U-test, P = 0.05) depicted as red lines. Proteins of interest are labeled red (Protein panel 1) or blue (Protein panel 2). REG1B was included in both panels.

Figure 2. Correlation between relative protein levels in the urine of infants with and without NEC. (A) Spearman's r correlation was used to determine the correlation of relative protein abundance for the entire cohort (n = 20 pairs). The directionality of the median fold change for NEC vs. non-NEC is depicted by red (increased) and blue (decreased) arrows. Protein panel 1 (●). Protein panel 2 (○). REG1B was included in both panels. Correlation matrices were generated for protein panels 1 (B) and 2 (C).

Figure 3. Relative protein levels in matched urine samples for pairs of infants with and without NEC cluster based on Bell's Stage, a marker of NEC severity. Hierarchical cluster analysis demonstrating median fold change in protein abundance for sample pairs. The
Directionality of the median fold change for NEC vs. control is depicted by red (increased) and blue (decreased) arrows. Protein panel 1 (●). Protein panel 2 (○). The black box delineates patients with Bell’s Stage III NEC.

Figure 4. Urine proteins in protein panel 1 differ between infants with NEC and controls, as well as based on Bell’s stage. The proteins in panel 1 include REG1B, REG3A, DEFA5, and FABP2. (A) Comparison of protein abundance for infants with NEC and controls. Line indicates median. Dots indicate individual infants. * P < 0.05, ** P < 0.01 by non-parametric Wilcoxon test. (B) Samples were stratified based on disease severity using modified Bell’s stage. Line indicates median. Dots indicate individual infants. ns, P > 0.05, * P < 0.05, ** P < 0.01, *** P < 0.005, **** P < 0.0001 by Kruskal Wallis. (C) ROC curves with AUC values and 95% CI.

Figure 5. Urine proteins in protein panel 2 differ between infants with NEC and controls, as well as based on Bell’s stage. (A) Comparison of protein abundance for infants with NEC and controls. Line indicates median. Dots indicate individual infants. * P < 0.05, ** P < 0.01, *** P < 0.005, **** P < 0.0001 by non-parametric Wilcoxon test. (B) Samples were stratified based on disease severity using modified Bell’s stage. Line indicates median. Dots indicate individual infants. * P < 0.05, ** P < 0.01, *** P < 0.005 by Kruskal Wallis. (C) ROC curves with AUC values and 95% CI.
Figure 6. Combinations of select urine proteins very effectively discriminate between infants with and without NEC. Combined ROC curves were generated for (A) Protein panel 1 (REG1B, REG3A, DEFA5, and FABP2), (B) Protein panel 2 (REG1B, SSBP1, CRYZL1, ITM2B, IL36B, and IL36RN) and (C) IL-36 and IL36RN. AUC values with the 95% CI are shown.

Supplementary Figure S1. ELISA validation of urine protein levels for controls and infants with NEC. (A) REG1B, (B) REG3A, (C) DEFA5, and (D) FABP2 were measured in the urine of all infants in the study via ELISA. Samples with protein levels below the limit of detection were censored. Bar indicates median. n = number of patients with detectable urine protein. ns, P > 0.05, * P < 0.05, ** P < 0.01 via Mann Whitney U-test.

Supplementary Figure S2. IL-22 and STAT3 are upstream regulators of proteins that are differentially abundant in the urine of infants with and without NEC. (A) Upstream regulator analysis (URA) predicted IL-22 as a direct regulator based on the downstream state of REG1B (+), REG3A (+), LBP (+), and TGM3 (-). (B) Causal network analysis (CNA) mapped indirect relationships, which included IL-22 and STAT3. Proteins marked in red and green were increased and decreased, respectively. Orange and blue proteins and interactions indicate predicted activation and inhibition, respectively. Yellow and grey lines indicate interactions that are inconsistent or not predicted with the state of the downstream molecule. Dashed and solid lines indicate indirect and direct relationships, respectively.
References


[37] Peterson KM, Guo X, Elkhalloung AG, et al. The expression of REG 1A and REG 1B is increased during acute amebic colitis. Parasitol Int. 2011 Sep;60(3):296-300.


the Early Stage of Disease. Journal of Immunology Research. 2020
2020/03/03;2020:3074313.


Table 1. Study cohort characteristics

<table>
<thead>
<tr>
<th>Sample ID</th>
<th>NEC/Control</th>
<th>Gestational age range (weeks)</th>
<th>Biological sex</th>
<th>Birth weight (g)</th>
<th>Route of delivery</th>
<th>NEC</th>
<th>Highest Bell’s stage</th>
<th>Radiologic Findings</th>
<th>Surgical NEC</th>
<th>Final disposition</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Age-Matched</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AM 1 N</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>480</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB</td>
<td>PN</td>
<td></td>
<td>Death, LD</td>
</tr>
<tr>
<td>AM 1 C</td>
<td></td>
<td>33 - 37</td>
<td>Female</td>
<td>1940</td>
<td>V</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 2 N</td>
<td></td>
<td>33 - 37</td>
<td>Female</td>
<td>2360</td>
<td>CS</td>
<td>Yes</td>
<td>IIA</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 2 C</td>
<td></td>
<td>33 - 37</td>
<td>Female</td>
<td>1740</td>
<td>CS</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 3 N</td>
<td></td>
<td>33 - 37</td>
<td>Female</td>
<td>2330</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB, PN, PVG</td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 3 C</td>
<td></td>
<td>33 - 37</td>
<td>Male</td>
<td>2360</td>
<td>CS</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 4 N</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>640</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 4 C</td>
<td></td>
<td>23 - 27</td>
<td>Female</td>
<td>710</td>
<td>CS</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 5 N</td>
<td></td>
<td>28 - 32</td>
<td>Female</td>
<td>1190</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB</td>
<td>PN, PVG</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 5 C</td>
<td></td>
<td>33 - 37</td>
<td>Female</td>
<td>2030</td>
<td>V</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 6 N</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>700</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB</td>
<td>G, PR, PD</td>
<td></td>
<td>Death, LD</td>
</tr>
<tr>
<td>AM 6 C</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>1150</td>
<td>V</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 7 N</td>
<td></td>
<td>28 - 32</td>
<td>Female</td>
<td>1870</td>
<td>CS</td>
<td>Yes</td>
<td>IIA</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 7 C</td>
<td></td>
<td>28 - 32</td>
<td>Female</td>
<td>2320</td>
<td>V</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 8 N</td>
<td></td>
<td>23 - 27</td>
<td>Female</td>
<td>790</td>
<td>V</td>
<td>Yes</td>
<td>IIIB</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>AM 8 C</td>
<td></td>
<td>28 - 32</td>
<td>Male</td>
<td>1580</td>
<td>V</td>
<td>No</td>
<td></td>
<td></td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td><strong>Self-Matched</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SM 9 C+N</td>
<td></td>
<td>28 - 32</td>
<td>Male</td>
<td>1220</td>
<td>CS</td>
<td>Yes</td>
<td>IIIA</td>
<td>PN, PVG</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>SM10 C+N</td>
<td></td>
<td>23 - 27</td>
<td>Female</td>
<td>760</td>
<td>V</td>
<td>Yes</td>
<td>IIIB, PN, PVG, P, PD</td>
<td></td>
<td></td>
<td>Death, NEC</td>
</tr>
<tr>
<td>SM11 C+N</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>760</td>
<td>CS</td>
<td>Yes</td>
<td>IIIA</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>SM12 C+N</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>790</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB, P</td>
<td>PR, PD</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>SM13 C+N</td>
<td></td>
<td>28 - 32</td>
<td>Female</td>
<td>1775</td>
<td>V</td>
<td>Yes</td>
<td>II</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>SM14 C+N</td>
<td></td>
<td>28 - 32</td>
<td>Female</td>
<td>1780</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>SM15 C+N</td>
<td></td>
<td>23 - 27</td>
<td>Female</td>
<td>690</td>
<td>CS</td>
<td>Yes</td>
<td>II</td>
<td></td>
<td></td>
<td>Study closed</td>
</tr>
<tr>
<td>SM16 C+N</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>430</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB</td>
<td>PN</td>
<td></td>
<td>Death, LD</td>
</tr>
<tr>
<td>SM17 C+N</td>
<td></td>
<td>28 - 32</td>
<td>Male</td>
<td>1770</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>SM18 C+N</td>
<td></td>
<td>23 - 27</td>
<td>Female</td>
<td>710</td>
<td>CS</td>
<td>Yes</td>
<td>IIA</td>
<td>PN</td>
<td></td>
<td>Discharged</td>
</tr>
<tr>
<td>SM19 C+N</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>530</td>
<td>CS</td>
<td>Yes</td>
<td>IIIA</td>
<td>G</td>
<td></td>
<td>Death, LD</td>
</tr>
<tr>
<td>SM20 C+N</td>
<td></td>
<td>23 - 27</td>
<td>Male</td>
<td>660</td>
<td>CS</td>
<td>Yes</td>
<td>IIIB</td>
<td>PVG, PR</td>
<td></td>
<td>Study closed</td>
</tr>
</tbody>
</table>

Delivery route: CS = cesarean section, V = vaginal; Surgical NEC: PR = partial resection, PD = peritoneal drain; Disposition: LD = lung disease, Radiological findings: PN = pneumatosis, PVG = portal venous gas, P = pneumoperitoneum, G = gasless abdomen, Study closed = PI relocation.
Table 2. Classification of differentially abundant urine proteins identified by the SomaScan® (7K) assay.

<table>
<thead>
<tr>
<th>Protein name</th>
<th>UniProt ID</th>
<th>Median fold change</th>
<th>Consensus tissue enrichment</th>
<th>Consensus localization enrichment</th>
<th>Role in inflammation</th>
</tr>
</thead>
<tbody>
<tr>
<td>REG1B</td>
<td>P48304</td>
<td>9.2</td>
<td>Intestine, pancreas</td>
<td>Secreted locally</td>
<td>▲▲▲▲ Putative antimicrobial properties. Regulated by STAT3. Upregulated in NEC. 0.82 (0.69-0.95) [36]</td>
</tr>
<tr>
<td>SSBP1</td>
<td>Q04837</td>
<td>5.7</td>
<td>Ubiquitous</td>
<td>Mitochondrial</td>
<td>▲▲▲▲ Critical regulator in mitochondrial DNA maintenance. Protects cells against proteotoxic stress 0.83 (0.70-0.96) [52]</td>
</tr>
<tr>
<td>CRYZL1</td>
<td>O95825</td>
<td>3.7</td>
<td>Ubiquitous</td>
<td>Intracellular</td>
<td>▲▲▲▲ Lipopolysaccharide (LPS)-induced expression mitigates pro-inflammatory responses in human monocytes. 0.76 (0.59-0.92) [65]</td>
</tr>
<tr>
<td>F9</td>
<td>P00740</td>
<td>3.2</td>
<td>Liver</td>
<td>Secreted to blood</td>
<td>▲▲▲▲ An inflammatory response factor involved in the intrinsic blood coagulation pathway by activating factor X. 0.67 (0.50-0.84) [66]</td>
</tr>
<tr>
<td>FABP4</td>
<td>P15090</td>
<td>3.0</td>
<td>Adipocyte</td>
<td>Intracellular</td>
<td>▲▲▲▲ Promotes inflammation. Downregulation suppresses inflammation, apoptosis, and oxidative stress. 0.69 (0.52-0.85) [67]</td>
</tr>
<tr>
<td>HAMP</td>
<td>P01172</td>
<td>3.0</td>
<td>Liver</td>
<td>Secreted to blood</td>
<td>▲▲▲▲ Anti-inflammatory activity and anti-inflammatory IL-6 mediated-on sequestration response. 0.76 (0.61-0.91) [46]</td>
</tr>
<tr>
<td>DEF5A</td>
<td>Q01523</td>
<td>3.0</td>
<td>Intestine</td>
<td>Locally secreted</td>
<td>▲▲▲▲ Paneth cell anti-microbial peptide. Induces secretion of IL-8 by intestinal epithelial cells. Increased in NEC. 0.66 (0.49-0.83) [71]</td>
</tr>
<tr>
<td>FABP2</td>
<td>P12104</td>
<td>3.0</td>
<td>Intestine</td>
<td>Intracellular</td>
<td>▲▲▲▲ Intestinal fatty acid binding protein. Putative biomarker which correlates with NEC in urine and serum. 0.68 (0.51-0.84) [69]</td>
</tr>
<tr>
<td>FABP1</td>
<td>P07148</td>
<td>3.0</td>
<td>Liver, intestine</td>
<td>Intracellular</td>
<td>▲▲▲▲ Liver fatty acid binding protein. Putative urine marker for NEC in urine. 0.63 (0.49-0.80) [70]</td>
</tr>
<tr>
<td>CHOL1</td>
<td>P36222</td>
<td>3.0</td>
<td>Macrophages</td>
<td>Secreted to blood</td>
<td>▲▲▲▲ Pro-inflammatory signaling in colitis by AKT, TNFα, and IL-6. A putative marker for intestinal inflammation. 0.71 (0.55-0.87) [71]</td>
</tr>
<tr>
<td>REG3A</td>
<td>Q06141</td>
<td>2.6</td>
<td>Intestine</td>
<td>Secreted locally</td>
<td>▲▲▲▲ Increased transcription observed in NEC. Antimicrobial peptide. Regulated by IL-17. 0.79 (0.66-0.93) [49]</td>
</tr>
<tr>
<td>C2</td>
<td>P06681</td>
<td>2.6</td>
<td>Liver</td>
<td>Secreted to blood</td>
<td>▲▲▲▲ MASP-2, an enzyme that regulates C2 and C4 activity, is associated with the development of NEC. 0.65 (0.48-0.82) [72]</td>
</tr>
</tbody>
</table>

**Increased**

**Decreased**

<table>
<thead>
<tr>
<th>Protein name</th>
<th>UniProt ID</th>
<th>Median fold change</th>
<th>Consensus tissue enrichment</th>
<th>Consensus localization enrichment</th>
<th>Role in inflammation</th>
</tr>
</thead>
<tbody>
<tr>
<td>ITM2B</td>
<td>Q9Y287</td>
<td>-9.8</td>
<td>Ubiquitous</td>
<td>Membrane</td>
<td>▲▲▲▲ Regulates anti-microbial amyloid-beta peptide aggregation. 0.79 (0.65-0.94) [57]</td>
</tr>
<tr>
<td>A2ML1</td>
<td>A8K2U0</td>
<td>-7.0</td>
<td>Squamous epithelium</td>
<td>Secreted locally</td>
<td>▲▲▲▲ Inactivates digestive proteases by covalent binding. Putative urine marker to distinguish NEC severity. 0.74 (0.59-0.89) [73]</td>
</tr>
<tr>
<td>IL36B</td>
<td>Q9N2H7</td>
<td>-6.1</td>
<td>Keratinocyte,</td>
<td>Secreted locally</td>
<td>▲▲▲▲ A pro-inflammatory cytokine that triggers MAPK and NF-κB signaling pathways. 0.86 (0.75-0.98) [58]</td>
</tr>
<tr>
<td>TGM3</td>
<td>Q08188</td>
<td>-4.3</td>
<td>Squamous epithelium</td>
<td>Secreted to blood</td>
<td>▲▲▲▲ Involved in innate immunity and barrier integrity. 0.71 (0.54-0.87) [74]</td>
</tr>
<tr>
<td>REXO2</td>
<td>Q9Y3B8</td>
<td>-3.7</td>
<td>Mitochondrial</td>
<td>Intracellular</td>
<td>▲▲▲▲ A putative anti-viral exoribonuclease that preferentially degrades two nucleotide RNA and DNA. 0.78 (0.63-0.92) [75]</td>
</tr>
<tr>
<td>ADPR5</td>
<td>Q9NX46</td>
<td>-3.7</td>
<td>Ubiquitous</td>
<td>Intracellular</td>
<td>▲▲▲▲ Post-translational modification of ribose to proteins, ribosylation linked to immune response modulation. 0.70 (0.54-0.87) [76]</td>
</tr>
<tr>
<td>KLK10</td>
<td>O43240</td>
<td>-3.5</td>
<td>Ubiquitous</td>
<td>Secreted locally</td>
<td>▲▲▲▲ Serine protease which can inhibit endothelial inflammation in blood vessels. 0.75 (0.60-0.90) [77]</td>
</tr>
<tr>
<td>SERPINB13</td>
<td>Q9U1V8</td>
<td>-3.2</td>
<td>Squamous epithelium</td>
<td>Intracellular</td>
<td>▲▲▲▲ Anti-inflammatory serine protease inhibitor. 0.65 (0.48-0.82) [78]</td>
</tr>
<tr>
<td>ITPK1</td>
<td>Q13572</td>
<td>-3.0</td>
<td>Ubiquitous</td>
<td>Mitochondrial</td>
<td>▲▲▲▲ Protects against TNFα-induced apoptosis and MLKL-mediated necroptosis. 0.81 (0.65-0.96) [79]</td>
</tr>
<tr>
<td>SERPINB3</td>
<td>P29508</td>
<td>-3.0</td>
<td>Squamous epithelium</td>
<td>Intracellular</td>
<td>▲▲▲▲ Cysteine protease inhibitor. Modulates immune response and regulates necrosis. 0.72 (0.55-0.88) [80]</td>
</tr>
<tr>
<td>UCM</td>
<td>Q8WVF2</td>
<td>-3.0</td>
<td>Chondrocytes</td>
<td>Extracellular matrix</td>
<td>▲▲▲▲ A secretory γ-carboxylglutamate that regulates fibriilin 2. 0.65 (0.48-0.82) [81]</td>
</tr>
<tr>
<td>PGRMC1</td>
<td>O00264</td>
<td>-3.0</td>
<td>Ubiquitous</td>
<td>Membrane, Mitochondrial</td>
<td>▲▲▲▲ Reduces lipotoxicity through fatty acid oxidation and mitochondrial respiration. 0.72 (0.55-0.88) [82]</td>
</tr>
<tr>
<td>EPS8L2</td>
<td>Q9H6S3</td>
<td>-2.8</td>
<td>Ubiquitous</td>
<td>Intracellular</td>
<td>▲▲▲▲ An analog of the EGFR receptor substrate has unknown involvement in the actin cytoskeleton remodeling. 0.83 (0.70-0.96) [83]</td>
</tr>
<tr>
<td>PPPR1R2</td>
<td>P41236</td>
<td>-2.8</td>
<td>Ubiquitous</td>
<td>Intracellular</td>
<td>▲▲▲▲ Enhances Tn17 differentiation and regulates TLR/ NF-κB, and MAPK signaling. 0.79 (0.65-0.93) [84]</td>
</tr>
<tr>
<td>PODXL2</td>
<td>Q9NN53</td>
<td>-2.8</td>
<td>CNS and pancreas</td>
<td>Intracellular</td>
<td>▲▲▲▲ Mediates leukocyte migration by acting as a ligand for vascular selectins E-α, P-α, and L-selectins. 0.73 (0.57-0.90) [85]</td>
</tr>
<tr>
<td>PADI1</td>
<td>Q9ULC8</td>
<td>-2.8</td>
<td>Keratinocytes</td>
<td>Intracellular</td>
<td>▲▲▲▲ Involved in the differentiation of immune cells. Regulated by NF-κB. 0.64 (0.47-0.81) [86]</td>
</tr>
<tr>
<td>IL36RN</td>
<td>Q9UBH0</td>
<td>-2.6</td>
<td>Keratinocytes</td>
<td>Intracellular</td>
<td>▲▲▲▲ Antagonist of the IL-36 signaling in epithelial barriers. High IL-36:IL36RN ratios observed in NEC. 0.91 (0.82-0.99) [87]</td>
</tr>
<tr>
<td>ENO2</td>
<td>P09104</td>
<td>-2.6</td>
<td>CNS and pancreas</td>
<td>Intracellular</td>
<td>▲▲▲▲ Gamma enolase is the predominant CNS dimer. Elevated in serum of preterm infants. 0.78 (0.63-0.92) [87]</td>
</tr>
<tr>
<td>COL5A1</td>
<td>P12109</td>
<td>-2.6</td>
<td>Connective tissue</td>
<td>Extracellular matrix</td>
<td>▲▲▲▲ Myofibroblasts deposit collagen VI to enhance T-cell adhesion to regulate intestinal inflammation. 0.73 (0.57-0.89) [88]</td>
</tr>
<tr>
<td>TUF1</td>
<td>Q9NXX1</td>
<td>-2.6</td>
<td>Secreted locally</td>
<td>Intracellular</td>
<td>▲▲▲▲ Involved in mesenchymal stem cell function and adaptation to hypoxia. 0.73 (0.57-0.89) [89]</td>
</tr>
<tr>
<td>MYG1</td>
<td>Q9HB07</td>
<td>-2.6</td>
<td>Ubiquitous</td>
<td>Intracellular</td>
<td>▲▲▲▲ Exonuclease which cleaves in situ at specific transcripts. 0.71 (0.55-0.88) [90]</td>
</tr>
<tr>
<td>PYGL</td>
<td>P06737</td>
<td>-2.6</td>
<td>Ubiquitous, liver</td>
<td>Intracellular</td>
<td>▲▲▲▲ Strongly linked to inflammation through glycogen metabolism 0.70 (0.54-0.87) [91]</td>
</tr>
<tr>
<td>CAB39</td>
<td>Q9Y376</td>
<td>-2.6</td>
<td>Ubiquitous</td>
<td>Intracellular</td>
<td>▲▲▲▲ Component of a complex that binds and activates STK11/LKB1 (tumor suppressor). 0.70 (0.53-0.86) [92]</td>
</tr>
<tr>
<td>CTNMB1P1</td>
<td>Q9NSA3</td>
<td>-2.6</td>
<td>Ubiquitous</td>
<td>Intracellular</td>
<td>▲▲▲▲ Enterocyte proliferation is impaired by dampened WNT/β-catenin signaling in NEC. 0.67 (0.50-0.83) [93]</td>
</tr>
</tbody>
</table>
### Table 3. Ingenuity pathway upstream regulator analysis

#### Causal Network Analysis

<table>
<thead>
<tr>
<th>Master Regulator</th>
<th>Protein Name</th>
<th>UniProt</th>
<th>Molecule Type</th>
<th>Predicted Activation</th>
<th>Activation z-score</th>
<th>P-value of overlap</th>
<th>Network bias-corrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>TNFRSF1B</td>
<td>TNF receptor-associated factor 2</td>
<td>Q12933</td>
<td>Transmembrane receptor</td>
<td>Activated</td>
<td>2.353</td>
<td>1.0E-09</td>
<td>0.0003</td>
</tr>
<tr>
<td>ZFP36</td>
<td>miRNA decay activator protein ZFP36</td>
<td>P26651</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>2.263</td>
<td>7.4E-07</td>
<td>0.0153</td>
</tr>
<tr>
<td>ZC3H12A</td>
<td>Endoribonuclease ZC3H12A</td>
<td>Q50168</td>
<td>Enzyme</td>
<td>Activated</td>
<td>2.191</td>
<td>1.7E-05</td>
<td>0.0430</td>
</tr>
<tr>
<td>IL22</td>
<td>Interleukin-22</td>
<td>Q9G2X6</td>
<td>Cytokine</td>
<td>Activated</td>
<td>2.000</td>
<td>6.9E-04</td>
<td>0.0171</td>
</tr>
<tr>
<td>USP11</td>
<td>Ubiquitin carboxyl-terminal hydrolase 11</td>
<td>P51784</td>
<td>Peptidase</td>
<td>Inhibited</td>
<td>-2.121</td>
<td>9.6E-07</td>
<td>0.0140</td>
</tr>
<tr>
<td>EML4-ALK</td>
<td>Tyrosine-protein kinase receptor</td>
<td>J7MA22</td>
<td>Fusion gene/product</td>
<td>Inhibited</td>
<td>-2.137</td>
<td>9.6E-08</td>
<td>0.0018</td>
</tr>
<tr>
<td>ALDH1A3</td>
<td>Aldehyde dehydrogenase family 1 member A3</td>
<td>P47895</td>
<td>Enzyme</td>
<td>Inhibited</td>
<td>-2.236</td>
<td>8.0E-06</td>
<td>0.0005</td>
</tr>
<tr>
<td>GRM1</td>
<td>Metabotropic glutamate receptor 1</td>
<td>Q13225</td>
<td>G-protein coupled receptor</td>
<td>Inhibited</td>
<td>-2.263</td>
<td>1.8E-05</td>
<td>0.0263</td>
</tr>
<tr>
<td>PLCE1</td>
<td>1-phosphatidylinositol 4,5-bisphosphatasephodiesterase E1</td>
<td>Q9P212</td>
<td>Enzyme</td>
<td>Inhibited</td>
<td>-2.401</td>
<td>6.4E-06</td>
<td>0.0274</td>
</tr>
<tr>
<td>RNF11</td>
<td>RING finger protein 11</td>
<td>Q9Y3C5</td>
<td>Enzyme</td>
<td>Inhibited</td>
<td>-2.414</td>
<td>4.3E-06</td>
<td>0.0435</td>
</tr>
<tr>
<td>SEPTIN9</td>
<td>Septin-9</td>
<td>Q2UH68</td>
<td>Enzyme</td>
<td>Inhibited</td>
<td>-2.502</td>
<td>7.3E-06</td>
<td>0.0178</td>
</tr>
</tbody>
</table>

#### Upstream Regulator analysis

<table>
<thead>
<tr>
<th>Upstream Regulator</th>
<th>Protein name</th>
<th>UniProt</th>
<th>Molecule Type</th>
<th>Predicted Activation</th>
<th>Activation z-score</th>
<th>P-value of overlap</th>
<th>Network bias-corrected P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Esrra</td>
<td>Steroid hormone receptor ERR1</td>
<td>P11474</td>
<td>Ligand-dependent nuclear</td>
<td>Activated</td>
<td>2.000</td>
<td>0.0003</td>
<td></td>
</tr>
<tr>
<td>IL22</td>
<td>Interleukin-22</td>
<td>Q9G2X6</td>
<td>Cytokine</td>
<td>Activated</td>
<td>1.957</td>
<td>0.0009</td>
<td></td>
</tr>
<tr>
<td>SREBF1</td>
<td>Sterol regulatory element-binding protein 1</td>
<td>P36956</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>1.950</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>RXRA</td>
<td>Retinoic acid receptor RXR-alpha</td>
<td>P19793</td>
<td>Ligand-dependent nuclear</td>
<td>Activated</td>
<td>1.941</td>
<td>0.0079</td>
<td></td>
</tr>
<tr>
<td>PGR</td>
<td>Progesterone receptor</td>
<td>P06401</td>
<td>Ligand-dependent nuclear</td>
<td>Activated</td>
<td>1.732</td>
<td>0.0499</td>
<td></td>
</tr>
<tr>
<td>STAT3</td>
<td>Signal transducer and activator of transcription 3</td>
<td>P40763</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>0.736</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>HIF1A</td>
<td>Hypoxia-inducible factor 1-alpha</td>
<td>Q16665</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>2.177</td>
<td>0.0013</td>
<td></td>
</tr>
<tr>
<td>TNF</td>
<td>Tumor necrosis factor</td>
<td>P01375</td>
<td>Cytokine</td>
<td>Activated</td>
<td>2.130</td>
<td>0.0026</td>
<td></td>
</tr>
<tr>
<td>KLF6</td>
<td>Kruppel-like factor 6</td>
<td>Q99612</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>2.000</td>
<td>0.0078</td>
<td></td>
</tr>
<tr>
<td>TGFBI</td>
<td>Transforming growth factor beta-1 proprotein</td>
<td>P01137</td>
<td>Growth factor</td>
<td>Activated</td>
<td>1.338</td>
<td>0.0018</td>
<td></td>
</tr>
<tr>
<td>IL1B</td>
<td>Interleukin-1 beta</td>
<td>P01584</td>
<td>Cytokine</td>
<td>Activated</td>
<td>1.277</td>
<td>0.0001</td>
<td></td>
</tr>
<tr>
<td>NFE2L2</td>
<td>Nuclear factor erythroid 2-related factor 2</td>
<td>Q16236</td>
<td>Transcription regulator</td>
<td>Activated</td>
<td>1.233</td>
<td>0.0198</td>
<td></td>
</tr>
<tr>
<td>IL10RA</td>
<td>Interleukin-10 receptor subunit alpha</td>
<td>Q13551</td>
<td>Transmembrane receptor</td>
<td>Activated</td>
<td>-1.000</td>
<td>0.0123</td>
<td></td>
</tr>
<tr>
<td>LPS</td>
<td>lipopolysaccharide</td>
<td>Q16661</td>
<td>Chemical drug</td>
<td>Activated</td>
<td>-0.942</td>
<td>0.0000</td>
<td></td>
</tr>
<tr>
<td>CTNNB1</td>
<td>Catenin beta-1</td>
<td>P35222</td>
<td>Cytokine</td>
<td>Activated</td>
<td>-0.686</td>
<td>0.0126</td>
<td></td>
</tr>
<tr>
<td>TGM2</td>
<td>Protein-glutamine gamma-glutamytransferase 2</td>
<td>P21980</td>
<td>Enzyme</td>
<td>Inhibited</td>
<td>-0.664</td>
<td>0.0005</td>
<td></td>
</tr>
<tr>
<td>AHR</td>
<td>Aryl hydrocarbon receptor</td>
<td>P35869</td>
<td>Ligand-dependent nuclear</td>
<td>Inhibited</td>
<td>-0.517</td>
<td>0.0001</td>
<td></td>
</tr>
</tbody>
</table>
Graphical abstract

Control vs. NEC in the context of mucus biogenesis and vasculature regulation.

Key proteins:
- REG1B
- FABP2
- REG3A
- DEFA5
- CRYZL1
- ITM2B
- IL-36B
- SSPB1
- IL-36RN

35 proteins upregulated in NEC
64 proteins downregulated in NEC

Renal System
Urinary Tract
Figure 1

A. Age-matched

B. Self-matched

C. Combined

Log$_2$ Median Fold Change vs. -Log$_{10}$ (P-value)

Genes with significant fold changes are highlighted:

- APOC1
- C1orf226
- C2
- CCL16
- CHI3L1
- COL6A1
- DTD1
- ELANE
- ENO2
- FABP1
- KLK11
- LBP
- LILRA3
- MYG1
- NECTIN1
- PRTN3
- REG3G
- REG4
- RETN
- UNC5C
- VSIG4

- A2ML1
- AFP
- ANGPTL1
- CD48
- CHIT1
- CNTN6
- FABP4
- HAMP
- HCLS1
- LRTM1
- OLFM2
- RET
- SERPINB13
- SERPINB3

Significance level: P < 0.05

All rights reserved. No reuse allowed without permission.
Figure 2

Protein Panel 1

<table>
<thead>
<tr>
<th></th>
<th>REG1B</th>
<th>REG3A</th>
<th>FABP2</th>
<th>DEFA5</th>
</tr>
</thead>
<tbody>
<tr>
<td>REG1B</td>
<td>1.00</td>
<td>0.57</td>
<td>0.10</td>
<td>0.46</td>
</tr>
<tr>
<td>REG3A</td>
<td>0.57</td>
<td>1.00</td>
<td>0.11</td>
<td>0.83</td>
</tr>
<tr>
<td>FABP2</td>
<td>0.10</td>
<td>0.11</td>
<td>1.00</td>
<td>0.47</td>
</tr>
<tr>
<td>DEFA5</td>
<td>0.46</td>
<td>0.83</td>
<td>0.47</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Protein Panel 2

<table>
<thead>
<tr>
<th></th>
<th>SSBP1</th>
<th>CRYZL1</th>
<th>ITM2B</th>
<th>IL36B</th>
<th>IL36RN</th>
<th>REG1B</th>
</tr>
</thead>
<tbody>
<tr>
<td>SSBP1</td>
<td>1.00</td>
<td>-0.52</td>
<td>0.62</td>
<td>-0.45</td>
<td>0.00</td>
<td>0.09</td>
</tr>
<tr>
<td>CRYZL1</td>
<td>-0.52</td>
<td>1.00</td>
<td>-0.30</td>
<td>-0.02</td>
<td>-0.10</td>
<td>-0.10</td>
</tr>
<tr>
<td>ITM2B</td>
<td>0.62</td>
<td>-0.30</td>
<td>1.00</td>
<td>-0.08</td>
<td>0.16</td>
<td>0.11</td>
</tr>
<tr>
<td>IL36B</td>
<td>-0.45</td>
<td>-0.02</td>
<td>0.86</td>
<td>1.00</td>
<td>-0.23</td>
<td>0.06</td>
</tr>
<tr>
<td>IL36RN</td>
<td>0.00</td>
<td>-0.10</td>
<td>0.16</td>
<td>-0.23</td>
<td>1.00</td>
<td>-0.63</td>
</tr>
<tr>
<td>REG1B</td>
<td>0.09</td>
<td>0.11</td>
<td>0.06</td>
<td>-0.63</td>
<td>1.00</td>
<td>0.00</td>
</tr>
</tbody>
</table>
Figure 4

A

REG1B

REG3A

DEFA5

FABP2

B

REG1B Log_{10} RFU

REG3A Log_{10} RFU

DEFA5 Log_{10} RFU

FABP2 Log_{10} RFU

Control   NEC

Control   NEC

Control   NEC

Control   NEC

C

Bell's stage

Bell's stage

Bell's stage

Bell's stage

Sensitivity%

Sensitivity%

Sensitivity%

Sensitivity%

AUC 0.82

AUC 0.79

AUC 0.66

AUC 0.68

95% CI 0.69 to 0.95

95% CI 0.65 to 0.93

95% CI 0.49 to 0.83

95% CI 0.51 to 0.84
Figure 6

A  Protein Panel 1
REG1B, REG3A, DEFA5, FABP2

B  Protein Panel 2
REG1B, SSBP1, CRYZL1, ITM2B, IL36B, IL36RN

C  IL-36 and IL-36RN

AUC  0.90
95% CI  0.80 to 1.0

AUC  0.98
95% CI  0.95 to 1.0

AUC  0.94
95% CI  0.86 to 1.0