

RESEARCH ARTICLE

Published September 12, 2018

FECAL HOST TRANSCRIPTOMICS FOR NON-INVASIVE HUMAN MUCOSAL IMMUNE PROFILING: PROOF OF CONCEPT IN *CLOSTRIDIUM DIFFICILE* INFECTION

AUTHORS

Robert Schlberg^{1,2}, Amanda Barrett³, Kornelia Edes¹, Michael Graves³, Litty Paul¹, Jenna Rychert², Bert K. Lopansri^{3,4*}, Daniel T. Leung^{1,3*}

ASSOCIATED INSTITUTIONS

¹Department of Pathology, University of Utah School of Medicine, Salt Lake City, Utah

²ARUP Laboratories, Salt Lake City, Utah

³Division of Infectious Diseases, University of Utah School of Medicine, Salt Lake City, Utah

⁴Division of Infectious Diseases and Clinical Epidemiology, Intermountain Medical Center, Murray, Utah

*B.K.L. and D.T.L. contributed equally to this work

CORRESPONDING AUTHOR

Daniel T. Leung
Division of Infectious Diseases
University of Utah School of Medicine
26 N. Medical Dr., Wintrobe 517
Salt Lake City, UT 84132
(801) 581-8804
daniel.leung@utah.edu

DOI

10.20411/pai.v3i2.250

SUGGESTED CITATION

Schlberg R, Barrett A, Edes K, Graves M, Paul L, Rychert J, Lopansri BK, Leung DT. Fecal Host Transcriptomics for Non-invasive Human Mucosal Immune Profiling: Proof of Concept in *Clostridium difficile* Infection. *Pathogens and Immunity*. 2018;3(2):164-80. doi: 10.20411/pai.v3i2.250

ABSTRACT

Background: Host factors play an important role in pathogenesis and disease outcome in *Clostridium difficile* infection (CDI), and characterization of these responses could uncover potential host biomarkers to complement existing microbe-based diagnostics.

Methods: We extracted RNA from fecal samples of patients with CDI and profiled human mRNA using amplicon-based next-generation sequencing (NGS). We compared the fecal host mRNA transcript expression profiles of patients with CDI to controls with non-CDI diarrhea.

Results: We found that the ratio of human actin gamma 1 (ACTG1) to 16S ribosomal RNA (rRNA) was highly correlated with NGS quality as measured by percentage of reads on target. Patients with CDI could be differentiated from those with non-CDI diarrhea based on their fecal mRNA expression profiles using principal component analysis. Among the most differentially expressed genes were ones related to immune response (IL23A, IL34) and actin-cytoskeleton function (TNNT1, MYL4, SMTN, MYBPC3, all adjusted P -values $< 1 \times 10^{-3}$).

Conclusions: In this proof-of-concept study, we used host fecal transcriptomics for non-invasive profiling of the mucosal immune response in CDI. We identified differentially expressed genes with biological plausibility based on animal and cell culture models. This demonstrates the potential of fecal transcriptomics to uncover host-based biomarkers for enteric infections.

Keywords: *Clostridium difficile*; host transcriptomics; fecal transcriptomics; actin-cytoskeleton; mucosal immune response

INTRODUCTION

Clostridium (or *Clostridioides*) *difficile* is a spore-forming, toxin-producing bacterium that causes diarrhea and colitis, most often associated with antibiotic use. It is the most common hospital-associated infection in the United States, with an estimated 453,000 cases and more than 29,000 deaths annually [1]. While *C. difficile* infection (CDI) is most often acquired in healthcare settings, community-associated infections are increasing [2, 3]. The most commonly used diagnostics for CDI are based on detection of the toxin or toxin genes. Due to their high sensitivity and rapid turnaround times, nucleic acid amplification tests (NAATs) have been widely adopted for the diagnosis of CDI. However, NAAT-based detection of *C. difficile* genomic DNA is unable to differentiate true CDI from colonization in the presence of alternative causes of diarrhea. Adoption of NAATs has led to an increase in CDI diagnoses worldwide, and concerns of over-diagnosis of CDI resulting in unnecessary use of antibiotics have sparked renewed interest in improved diagnostic strategies [4, 5].

Host biomarkers have the potential to complement microbe-based tests in the diagnostic algorithm. Unfortunately, currently available fecal biomarkers, including lactoferrin and calprotectin, have on the whole been unable to differentiate CDI from other causes of diarrhea [6, 7]. Furthermore, host factors play an important role in pathogenesis and disease outcome in CDI [8], and characterization of the host response could uncover novel biomarkers. Animal models have demonstrated that mice deficient for NOD1 [9], MyD88 [10], TLR4 [11], and innate lymphoid cells [12] are more susceptible to CDI, and that mice lacking IL-23 are more protected from CDI [13]. However, these findings have not been thoroughly tested in human studies. Limited hypothesis-driven studies in humans have identified several host factors associated with disease, such as

lower levels of antibody against *C. difficile* toxin A [14, 15], IL-8 polymorphism [16], and elevated fecal cytokines [17, 18].

Attempts at investigating intestinal immune responses in human CDI have been limited by the ethical and logistical challenges of invasive procedures. However, recent technological advances have enabled the noninvasive profiling of the mucosal immune responses by host transcriptomic profiling of fecal specimens using RNA-seq, microarrays, and PCR [19-21]. Thus, the objective of this study was to develop and optimize an unbiased method for examining the fecal host transcriptome in patients with *C. difficile* infection.

METHODS

Patient Selection

We collected de-identified fecal samples from adult patients whose feces were sent to the Intermountain Healthcare microbiology laboratory for *C. difficile* testing. Patients were defined to have CDI by positive results for GDH and toxin enzyme immunoassay (*C. diff* Quik Chek Complete, Alere). Controls were randomly selected from diarrheal samples sent for *C. difficile* testing but had negative GDH and toxin enzyme immunoassay results. Samples were de-identified and only data regarding age, gender, CDI status (positive or negative), and CDI severity were linked to the stools by a study number. Severe disease was determined by chart review using the modified University of Illinois criteria excluding the presence of pseudomembranous colitis [22, 23]. The study protocol was submitted to the Institutional Review Boards of the University of Utah and Intermountain Healthcare and deemed to be exempt from review.

Sample Collection and Preservation

The primary stool specimens were kept refrigerated after testing. After a positive toxin result, an aliquot of up to 4 mL of residual feces, not used for microbiologic testing, was placed into 4 mL of RNAProtect Cell Reagent (Qiagen) and stored at -80°C until RNA extraction.

RNA Extraction

Fecal samples were thawed, vortexed, and filtered through gauze with addition of 5-10 mL of PBS. The filtrate was pelleted by centrifugation with the gauze still in the tube for 5 minutes at 500g; then the gauze was discarded, the sample vortexed to resuspend the pellet, and the suspension was filtered through a 40 µm cell strainer (Fisherbrand). The resulting filtrate was centrifuged again at 500g, and the supernatant discarded. The remaining pellet was mixed with lysis buffer (PureLink RNA Mini Kit, LifeTechnologies), and cells were lysed by vortexing for 2 minutes in tubes containing 0.5 mm glass beads (PowerBead Tubes, MoBio). Equal volumes of 70% ethanol were added, the mix centrifuged, and the supernatant transferred to a PureLink spin column. The remainder of the RNA isolation steps was completed according to the PureLink RNA Mini Kit manufacturer's instructions, including use of on-column PureLink DNase treatment.

Quantitative RT-PCR

RNA was quantified by Qubit and Nanodrop 1000 (both Thermo Fisher Scientific). An aliquot of 8 µL of extracted RNA was reverse transcribed with SuperScript IV and VILO master mix (Invitrogen) primed with random hexamers. To assess integrity and yield of human mRNA, transcripts for a human housekeeping gene (actin gamma 1, ACTG1, Hs03044422_g1) and bacterial

16S rRNA (16S, Ba04230899_s1, both Thermo Fisher Scientific) were quantified by quantitative reverse transcriptase PCR (RT-qPCR), TaqMan Fast Advanced Master Mix (Thermo Fisher Scientific), and a QuantStudio 3 real-time thermocycler (Thermo Fisher Scientific). Amplification conditions were 50°C for 2 minutes, 95°C for 20 seconds, followed by 40 cycles of 95°C for 1 second and 60°C for 20 seconds. Commercial human RNA (Human Control RNA, Thermo Fisher Scientific) and total RNA extracted from a pure culture of *Escherichia coli* (DH5 α) were used as standards. A ratio of ACTG1 to 16S was calculated and compared to quality transcriptomic data.

Library Preparation for Transcriptome Sequencing

Libraries were prepared with the Ion AmpliSeq Transcriptome Human Gene Expression kit (Thermo Fisher Scientific, version 1) using 10 or 100 ng of total RNA, depending on human mRNA abundance. Commercial human RNA was used as control. Targets were amplified with the Ion AmpliSeq Transcriptome Human Gene Expression Core Panel targeting >20,000 genes, followed by partial digestion of primers, ligation of barcoded adapters, and library amplification for 18 cycles. Amplified libraries were eluted in 30 μ L of low TE buffer after purification and quantified using the Ion Library TaqMan Quantitation Kit (Thermo Fisher Scientific). Libraries from 12 samples with ACTG1:16S ratios of above 10^{-4} were randomly selected for sequencing on the Ion Torrent Proton system using a P1 chip according to the manufacturer's instructions (Thermo Fisher Scientific).

Analysis of Transcriptome Sequencing Data

The resulting sequencing data was analyzed using the Torrent Suite (Thermo Fisher Scientific, version 5.0.4 and the human reference genome build hg19) with default analysis parameters providing sequencing read counts for each of the targeted genes. The resulting read count matrices were analyzed using DESeq2 [24] within R (R Foundation for Statistical Computing, Vienna, Austria; version 3.3.2). Genes with *P*-values adjusted for multiple testing (Benjamini and Hochberg, generated within DESeq2) <0.05 were included in subsequent analyses. For quality control and to assess adequacy of calculated percentage of reads on target, these resulting sequencing reads were also analyzed with Taxonomer [25], and reads binned as human were quantified and compared to percentage of reads on target provided by the Torrent Suite software. Functional annotations were derived using DAVID [26] analyses using Benjamini-adjusted *P*-values.

RESULTS

Human Feces mRNA Yield and Effect on Transcriptome Profiling

The AmpliSeq Transcriptome Human Gene Expression kit is intended for analysis of 10 ng of human RNA. To assess limitations of analyzing mixed bacterial and human RNA from fecal samples, we quantified the proportion of human ACTG1 mRNA and bacterial 16S RNA in 34 fecal samples by qPCR. Estimated ACTG1 mRNA and 16S rRNA (ACTG1:16S) ratios were calculated to assess human mRNA concentrations. ACTG1:16S ratios spanned 5 orders of magnitude (Figure 1A). RNA from fecal specimens with ACTG1:16S ratios above the median (6.1×10^{-4}) were randomly selected for transcriptome analysis (Figure 1A, highlighted in red and blue). A mean of 4.7×10^6 reads (SD = 1.9×10^6) were generated per sample with a mean 86.0% of base calls having quality scores of \geq Q20 (SD 1.5%). On average, 58.5% of reads contained expected sequences or *reads on target* (ie, sequencing reads aligning to the expected regions of the >20,000 targeted genes, SD

13.1%) and 34.5% of targeted genes were detected (SD 9.2%, Supplementary Figure 1). To assess accuracy of data analysis with the Torrent Suite when applying the method to mixed RNA with a high abundance of non-human RNA, we also analyzed the sequencing reads with a metagenomics data analysis tool able to enumerate human mRNA reads in highly-mixed data (Taxonomer) [25]. The proportion of human mRNA reads as determined by Taxonomer was highly correlated with the percentage of reads on target as determined by the Torrent Suite (Supplementary Figure 2). Subsequent analyses are based on results of the Torrent Suite. The ACTG1:16S ratio correlated with the percentage of valid reads ($R^2=0.73$, Figure 1B) supporting its use as a predictor for sample suitability for transcriptome analysis.

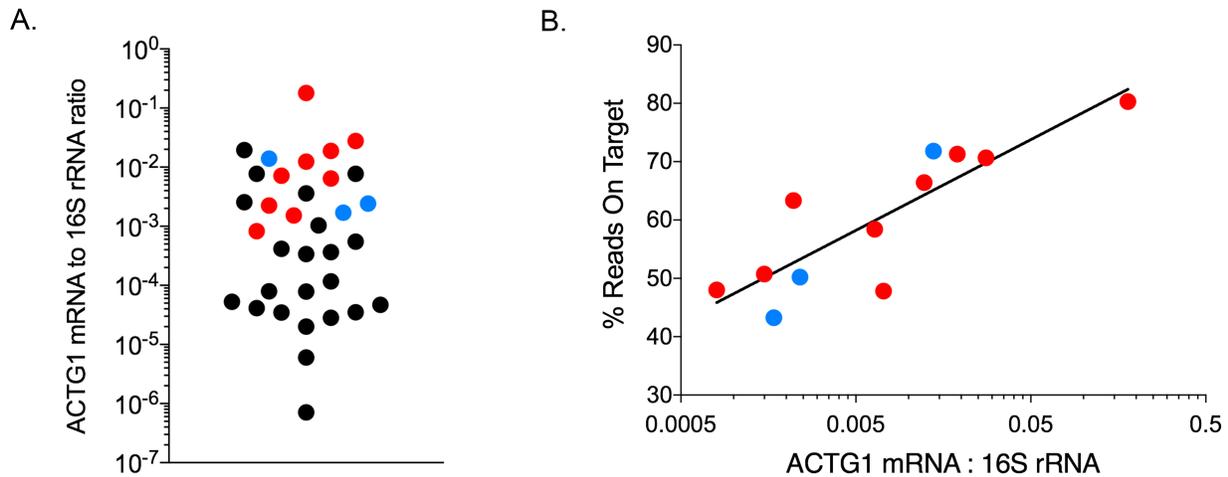


Figure 1. Quality control of extracted RNA and sequencing Data. A) ACTG1 mRNA to 16S rRNA ratios for 34 fecal samples; samples highlighted in red (CDI) and blue (diarrhea of other causes) were selected for sequencing. B) ACTG1:16S ratios correlated with expected sequences (*reads on target*, Pearson R^2 0.396; Goodness of Fit (semilog) $R^2=0.73$).

Transcriptome Profiling in *C. difficile*-Positive and *C. difficile*-Negative Fecal Samples

Of the 12 samples sequenced, 9 (75%) were from patients with CDI and 3 (25%) were from patients with diarrhea of other causes (Table 1). Of the 9 samples from CDI patients, the median age was 62 years (range 17-91), 5 were female, 5 had recurrent episodes, and none met Zar criteria for severe CDI [23]. A total of 20,046 genes (96.3%) had ≥ 1 read in at least 1 sample, with 6388 genes (31.9%) being differentially expressed between patients with CDI and diarrhea of other causes based on an adjusted P -value of <0.05 (DESeq2). To identify genes that were consistently and maximally differentially expressed, we limited subsequent analyses to genes with $<80\%$ coefficient of variation within each of the 2 groups and >10 -fold difference between group means ($n=922$). Relative expression levels of the top 50 of these differentially expressed genes are shown in Figure 2A. Among these genes were several with a known role in the pathogenesis of CDI, including IL23A (encoding for the alpha subunit of interleukin 23), IL34 (interleukin 34), TNNT1 (slow skeletal troponin type 1), MYBPC3 (myosin-binding protein C isoform 3), and NMRAL1 (NmrA like redox sensor 1). Of the 9 patients with CDI, 1 showed an mRNA expression profile that was more similar to those of the controls (CDI08), 3 had intermediary mRNA expression profiles (CDI23, CDI26, CDI29), while the remaining 5 (CDI10, CDI24, CDI27, CDI35, CDI36) were

characterized by high expression levels of most of the differentially expressed genes (Figure 2A). In a principal component analysis of these expression profiles >84% of the variance was explained by principal component (PC) 1, which also separated the 3 groups outlined above (Figure 2B). There were no clear differences between groups in age, gender, or recurrence status.

Table 1. Demographic and clinical data for patients with CDI and controls with non-CDI diarrhea

Sample ID	Group	Recurrent Episode	Age	Gender
08	CDI	Yes	73	M
10	CDI	No	70	M
23	CDI	Yes	62	M
24	CDI	No	17	M
26	CDI	Yes	46	F
27	CDI	Yes	80	F
29	CDI	No	80	F
35	CDI	Yes	31	F
36	CDI	No	19	F
N9	Control	-	30	F
N10	Control	-	52	F
N11	Control	-	64	M

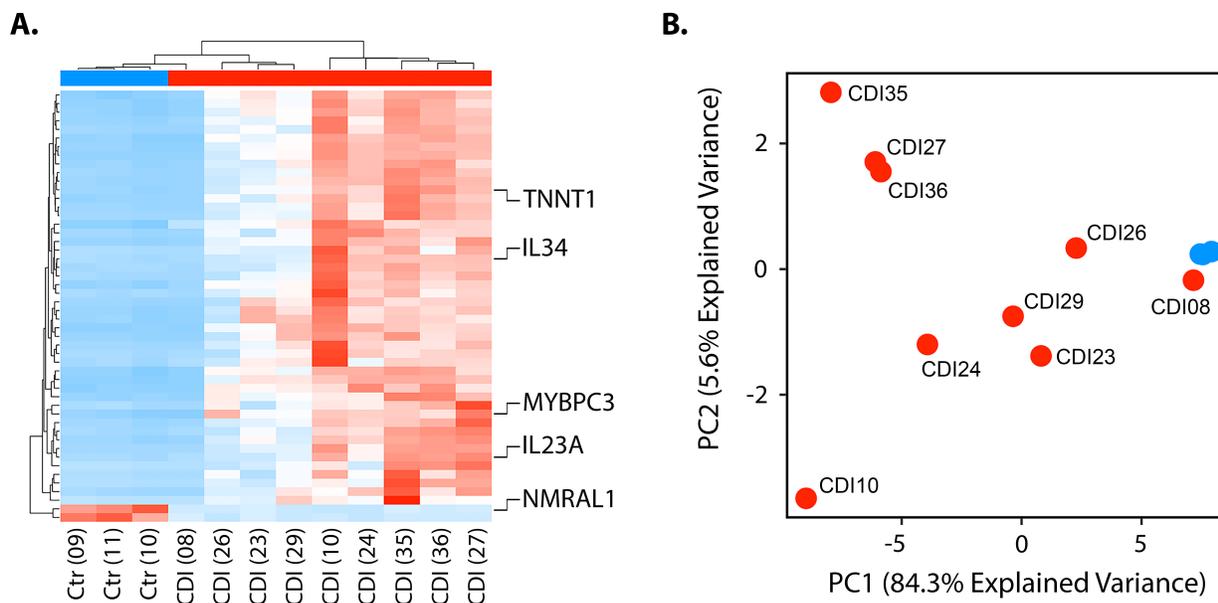


Figure 2. mRNA expression in patients with CDI vs controls. A) Relative expression levels of the top 50 of these differentially expressed genes between patients with CDI and diarrhea of other causes. B) Principal component analysis demonstrated 84.3% of the variance explained by principal component (PC) 1.

DAVID analysis of the top 50 differentially expressed genes (Supplementary Table 1) showed 5 functional annotations to be significantly enriched (Figure 3). All 5 functional annotations ('cytoskeletal protein binding', 'sarcomere', 'myofibril', 'contractile fiber part', and 'actin cytoskeleton') play a known role in the pathogenesis of CDI. The 7 genes linked to these functional annotations are shown in Supplementary Table 2 and include TNKS1BP1, JUP, TNNT1, SPTBN1, C22ORF28, and BCL2L11.

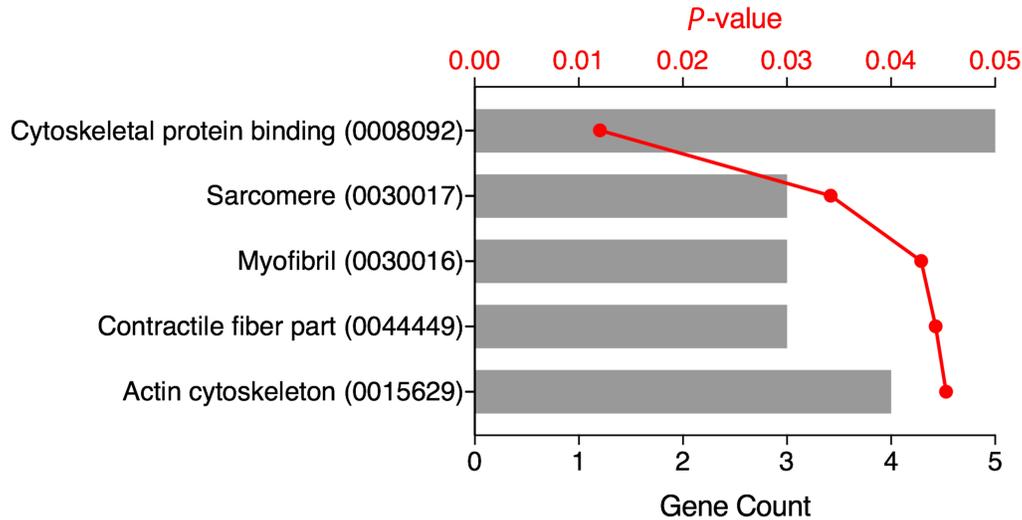


Figure 3. Functional Annotation. Gene Ontology (GO) terms and accession numbers (in parenthesis) for the top 50 differentially expressed genes (DAVID analysis, GO terms with Benjamini-adjusted P -values < 0.05) are shown with the number of genes (grey bar) and adjusted P -value (red).

DISCUSSION

Clinical manifestations of *Clostridium difficile* infection range from asymptomatic carriage to fulminant, pseudomembranous colitis and are mainly mediated by the effects of toxins A and/or B on intestinal epithelial cells. The expanding reservoir of *C. difficile* beyond the healthcare setting into community-based sources [3], and increasing use of NAAT tests to detect *C. difficile* and other gastrointestinal pathogens [27], has created challenges in accurately distinguishing CDI from other causes of diarrhea due to frequently overlapping clinical syndromes. When used in patients with known alternative causes of diarrhea (eg, laxative use) [28], a positive toxin result can lead to overestimation of disease and inappropriate treatment. Given the wide range of clinical manifestations of CDI, characterizing the host transcriptional response to infection has the potential to improve our understanding of the disease process and to identify expression profiles that correlate with outcome.

We demonstrate the potential use of host mRNA-targeted amplicon-based sequencing of fecal samples to identify gene transcripts enriched in patients with CDI compared to controls with *C. difficile*-negative diarrhea. We found upregulation of genes associated with cytokine responses previously shown to be important in *C. difficile* pathogenesis, including IL23A, which is associated with increased colonic inflammation and mortality in animal models of CDI [13, 29, 30], and IL34, which is upregulated in tissues of patients with inflammatory bowel disease and in mouse

models of experimental colitis [31-33]. Intriguingly, we also found upregulation of genes related to actin-cytoskeleton function, including TNNT1, JUP, SPTBN1, and MYBPC3. The toxins of *C. difficile* act through disturbance of the intestinal cell cytoskeleton resulting in apoptosis [34, 35], and hypervirulent strains of *C. difficile* inhibit actin polymerization through production of an actin-ADP-ribosylating toxin [36]. Identification of such biologically plausible targets thus provides proof-of-concept for the use of host fecal transcriptomics to probe human mucosal immune responses.

Host transcript profiling has the potential to complement existing microbe-targeted molecular tests to improve specificity for detecting clinically relevant *C. difficile* infections. Existing tests have centered on microbial toxin detection in stools by functional assays (ie, cell cytotoxicity assay), detection of toxins A and/or B antigen, or NAATs for toxin genes (*tcdA* and/or *tcdB*). NAAT testing has improved our ability to rapidly detect genes for Toxins A and/or B, which represents a significant advance over detection of free toxin by antigen detection methods [37]. An accumulating body of evidence, however, questions the benefit of NAAT testing alone [38]. Specificity of NAAT decreases when clinical manifestations are taken into account [39]. Additionally, clinical outcomes in patients solely with PCR-positive results are similar to patients with negative results for *C. difficile* by both toxin assay and NAAT [40]. Updated *C. difficile* guidelines discourage testing with NAATs alone and recommend combining NAAT with toxin detection in multistep algorithms in conjunction with efforts to avoid inappropriate testing that might identify asymptomatic carriers of toxigenic *C. difficile* [41]. These recommendations were considered to be weak with low quality of evidence so new approaches are needed.

Host mRNA can be detected in feces by nucleic acid amplification tests and next-generation sequencing [20, 42-45]. However, host RNA contributes only a minute fraction to the total RNA present in fecal specimens with most of the total RNA being of bacterial origin (largely ribosomal RNA) [46]. Quantifying expression levels of host mRNA in this overwhelming background of bacterial RNA is technically challenging. In addition, the integrity of host cells and nucleic acid may be compromised. Next-generation sequencing technologies based on sequence-specific amplification of short regions of hundreds to thousands of mRNA transcripts (amplicon sequencing) provides a solution to this dilemma. Similar methods have been used for transcriptional profiling of RNA that is in low abundance or highly fragmented RNA, such as RNA extracted from formalin-fixed tissue [47, 48].

Our results provide proof-of-concept for the use of amplicon sequencing to noninvasively probe the mucosal host immune response to *C. difficile* infection. If confirmed in larger studies, this strategy will also enable noninvasive profiling and diagnosis of other gastrointestinal diseases such as other intestinal infections, inflammatory bowel disease, and colon cancer. In respiratory tract infections, a similar approach has led to the discovery of host mRNA expression signatures that can be used for diagnostic purposes and to differentiate bacterial from viral infections [49-52]. After defining a minimal set of maximally informative transcripts, quantitative reverse transcription PCR panels can be designed for faster, cheaper, and more scalable testing and development of tests that can be implemented in clinical laboratories [53]. While fecal samples provide unique technical challenges for nucleic acid extraction and detection, applications in noninvasive gastrointestinal cancer detection and multiplex panels for gastrointestinal pathogens have demonstrated that these can be overcome. Eliminating the need for colonoscopy and biopsy to assess the

mucosal immune response provides a substantial benefit from our approach and an incentive to further optimize testing strategies.

Our study has several limitations. First, despite the large number of samples collected, only a small number of samples were analyzed by next-generation sequencing because of variable human mRNA yields (as measured by ACTG1:16S ratios), resource constraints, and the exploratory nature of this study. Despite the limited sample size, the biological plausibility of results (genes with a known role in pathogenesis were differentially expressed) demonstrates the power of this noninvasive approach. Second, our convenience sampling of stool specimens with >4 mL and with higher levels of ACTG1 mRNA may have introduced a selection bias towards a certain phenotype of diarrhea caused by *C. difficile*. Future studies will include a carefully selected patient population with clinically confirmed *C. difficile* colitis, patients with diarrhea with a positive *C. difficile* test with an alternative cause of diarrhea identified, and randomly selected specimens to address this limitation. Third, further improvements in enriching human cells and/or RNA in stool specimens will improve sensitivity for expression analyses of less abundant genes and allow for a larger proportion of specimens to be analyzed. Lastly, we did not account for antibiotic use among patients, which may influence the host transcriptome. Despite these limitations, our study provides proof-of-concept that amplicon-based next generation sequencing of fecal samples can be used to probe host gene expression for immune profiling and biomarker discovery.

In summary, in our small, pilot study, we demonstrate differential expression of several genes in patients with *C. difficile* infection compared to those with diarrhea negative for *C. difficile*. Larger numbers of representative samples from patients with known *C. difficile* infection and diarrhea of other causes with *C. difficile* carriage will need to be tested to confirm and validate our findings and assess the power of host transcriptomics to assist in the critical distinction between *C. difficile* infection and carriage.

FUNDING

Research reported in this publication was supported in part by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award numbers K08AI100923 and R01AI130378 [to D.T.L.]; the University of Utah Research Incentive Seed Grant Program [to D.T.L.]; and the University of Utah Center for Clinical and Translational Science Pilot Award Program [to R.S., B.K.L., and D.T.L.] funded by the National Center for Advancing Translational Sciences of the National Institutes of Health under Award Number UL1TR001067. The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

CONFLICTS OF INTEREST

R.S. is a coinventor on a patent application pending for Taxonomer, which has been licensed by IDbyDNA from the University of Utah. He owns equity in and consults for IDbyDNA. B.K.L. has received research support from OpGen, travel support for a speaking engagement from Luminex, and is a member of the scientific advisory board of Merck. All other authors report no potential conflicts of interest.

REFERENCES

1. Lessa FC, Mu Y, Bamberg WM, Beldavs ZG, Dumyati GK, Dunn JR, Farley MM, Holzbauer SM, Meek JI, Phipps EC, Wilson LE, Winston LG, Cohen JA, Limbago BM, Fridkin SK, Gerding DN, McDonald LC. Burden of *Clostridium difficile* infection in the United States. *N Engl J Med*. 2015;372(9):825-34. PubMed PMID: 25714160. doi: 10.1056/NEJMoa1408913
2. Chitnis AS, Holzbauer SM, Belflower RM, Winston LG, Bamberg WM, Lyons C, Farley MM, Dumyati GK, Wilson LE, Beldavs ZG, Dunn JR, Gould LH, MacCannell DR, Gerding DN, McDonald LC, Lessa FC. Epidemiology of community-associated *Clostridium difficile* infection, 2009 through 2011. *JAMA Intern Med*. 2013;173(14):1359-67. PubMed PMID: 23780507. doi: 10.1001/jamainternmed.2013.7056
3. Khanna S, Pardi DS, Aronson SL, Kammer PP, Orenstein R, St Sauver JL, Harmsen WS, Zinsmeister AR. The epidemiology of community-acquired *Clostridium difficile* infection: a population-based study. *Am J Gastroenterol*. 2012;107(1):89-95. PubMed PMID: 22108454. Pubmed Central PMCID: PMC3273904. doi: 10.1038/ajg.2011.398
4. Longtin Y, Trottier S, Brochu G, Paquet-Bolduc B, Garenc C, Loungnarath V, Beaulieu C, Goulet D, Longtin J. Impact of the type of diagnostic assay on *Clostridium difficile* infection and complication rates in a mandatory reporting program. *Clin Infect Dis*. 2013;56(1):67-73. PubMed PMID: 23011147. doi: 10.1093/cid/cis840
5. Williamson DA, Basu I, Freeman J, Swager T, Roberts SA. Improved detection of toxigenic *Clostridium difficile* using the Cepheid Xpert C *difficile* assay and impact on C *difficile* infection rates in a tertiary hospital: a double-edged sword. *Am J Infect Control*. 2013;41(3):270-2. PubMed PMID: 22939418. doi: 10.1016/j.ajic.2012.03.025
6. Usacheva EA, Jin JP, Peterson LR. Host response to *Clostridium difficile* infection: Diagnostics and detection. *J Glob Antimicrob Resist*. 2016;7:93-101. PubMed PMID: 27693863. doi: 10.1016/j.jgar.2016.08.002
7. Swale A, Miyajima F, Roberts P, Hall A, Little M, Beadsworth MB, Beeching NJ, Kolamunnage-Dona R, Parry CM, Pirmohamed M. Calprotectin and lactoferrin faecal levels in patients with *Clostridium difficile* infection (CDI): a prospective cohort study. *PLoS One*. 2014;9(8):e106118. PubMed PMID: 25170963. Pubmed Central PMCID: 4149523. doi: 10.1371/journal.pone.0106118
8. Loo VG, Bourgault AM, Poirier L, Lamothe F, Michaud S, Turgeon N, Toye B, Beaudoin A, Frost EH, Gilca R, Brassard P, Dendukuri N, Beliveau C, Oughton M, Brukner I, Dascal A. Host and pathogen factors for *Clostridium difficile* infection and colonization. *N Engl J Med*. 2011;365(18):1693-703. PubMed PMID: 22047560. doi: 10.1056/NEJMoa1012413
9. Hasegawa M, Yamazaki T, Kamada N, Tawaratsumida K, Kim YG, Nunez G, Inohara N. Nucleotide-binding oligomerization domain 1 mediates recognition of *Clostridium difficile* and induces neutrophil recruitment and protection against the pathogen. *J Immunol*. 2011;186(8):4872-80. PubMed PMID: 21411735. doi: 10.4049/jimmunol.1003761

10. Jarchum I, Liu M, Shi C, Equinda M, Pamer EG. Critical role for MyD88-mediated neutrophil recruitment during *Clostridium difficile* colitis. *Infect Immun*. 2012;80(9):2989-96. PubMed PMID: 22689818. Pubmed Central PMCID: PMC3418725. doi: 10.1128/IAI.00448-12
11. Ryan A, Lynch M, Smith SM, Amu S, Nel HJ, McCoy CE, Dowling JK, Draper E, O'Reilly V, McCarthy C, O'Brien J, Ni Eidhin D, O'Connell MJ, Keogh B, Morton CO, Rogers TR, Fallon PG, O'Neill LA, Kelleher D, Loscher CE. A role for TLR4 in *Clostridium difficile* infection and the recognition of surface layer proteins. *PLoS Pathog*. 2011;7(6):e1002076. PubMed PMID: 21738466. Pubmed Central PMCID: PMC3128122. doi: 10.1371/journal.ppat.1002076
12. Abt MC, Lewis BB, Caballero S, Xiong H, Carter RA, Susac B, Ling L, Leiner I, Pamer EG. Innate Immune Defenses Mediated by Two ILC Subsets Are Critical for Protection against Acute *Clostridium difficile* Infection. *Cell Host Microbe*. 2015;18(1):27-37. PubMed PMID: 26159718. Pubmed Central PMCID: PMC4537644. doi: 10.1016/j.chom.2015.06.011
13. Buonomo EL, Madan R, Pramoongjago P, Li L, Okusa MD, Petri WA, Jr. Role of interleukin 23 signaling in *Clostridium difficile* colitis. *J Infect Dis*. 2013;208(6):917-20. PubMed PMID: 23776194. Pubmed Central PMCID: PMC3749013. doi: 10.1093/infdis/jit277
14. Kyne L, Warny M, Qamar A, Kelly CP. Asymptomatic carriage of *Clostridium difficile* and serum levels of IgG antibody against toxin A. *N Engl J Med*. 2000;342(6):390-7. PubMed PMID: 10666429. doi: 10.1056/NEJM200002103420604
15. Kyne L, Warny M, Qamar A, Kelly CP. Association between antibody response to toxin A and protection against recurrent *Clostridium difficile* diarrhoea. *Lancet*. 2001;357(9251):189-93. PubMed PMID: 11213096. doi: 10.1016/S0140-6736(00)03592-3
16. Garey KW, Jiang ZD, Ghantaji S, Tam VH, Arora V, Dupont HL. A common polymorphism in the interleukin-8 gene promoter is associated with an increased risk for recurrent *Clostridium difficile* infection. *Clin Infect Dis*. 2010;51(12):1406-10. PubMed PMID: 21058913. doi: 10.1086/657398
17. El Feghaly RE, Stauber JL, Deych E, Gonzalez C, Tarr PI, Haslam DB. Markers of intestinal inflammation, not bacterial burden, correlate with clinical outcomes in *Clostridium difficile* infection. *Clin Infect Dis*. 2013;56(12):1713-21. PubMed PMID: 23487367. Pubmed Central PMCID: 3707425. doi: 10.1093/cid/cit147
18. El Feghaly RE, Stauber JL, Tarr PI, Haslam DB. Intestinal inflammatory biomarkers and outcome in pediatric *Clostridium difficile* infections. *J Pediatr*. 2013;163(6):1697-704 e2. PubMed PMID: 24011765. Pubmed Central PMCID: 4098967. doi: 10.1016/j.jpeds.2013.07.029
19. Cernada M, Bauerl C, Serna E, Collado MC, Martinez GP, Vento M. Sepsis in preterm infants causes alterations in mucosal gene expression and microbiota profiles compared to non-septic twins. *Sci Rep*. 2016;6:25497. PubMed PMID: 27180802. Pubmed Central PMCID: 4867619. doi: 10.1038/srep25497

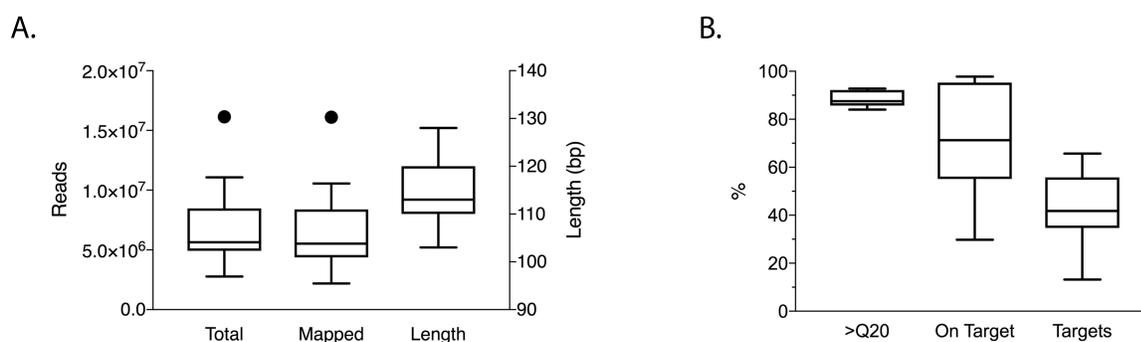
20. Knight JM, Davidson LA, Herman D, Martin CR, Goldsby JS, Ivanov IV, Donovan SM, Chapkin RS. Non-invasive analysis of intestinal development in preterm and term infants using RNA-Sequencing. *Sci Rep.* 2014;4:5453. PubMed PMID: 24965658. Pubmed Central PMCID: 4071321. doi: 10.1038/srep05453
21. Stauber J, Shaikh N, Ordiz MI, Tarr PI, Manary MJ. Droplet digital PCR quantifies host inflammatory transcripts in feces reliably and reproducibly. *Cell Immunol.* 2016;303:43-9. PubMed PMID: 27063479. Pubmed Central PMCID: 4863679. doi: 10.1016/j.cellimm.2016.03.007
22. Fujitani S, George WL, Murthy AR. Comparison of clinical severity score indices for *Clostridium difficile* infection. *Infect Control Hosp Epidemiol.* 2011;32(3):220-8. PubMed PMID: 21460506. doi: 10.1086/658336
23. Zar FA, Bakkanagari SR, Moorthi KM, Davis MB. A comparison of vancomycin and metronidazole for the treatment of *Clostridium difficile*-associated diarrhea, stratified by disease severity. *Clin Infect Dis.* 2007;45(3):302-7. PubMed PMID: 17599306. doi: 10.1086/519265
24. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 2014;15(12):550. PubMed PMID: 25516281. Pubmed Central PMCID: PMC4302049. doi: 10.1186/s13059-014-0550-8
25. Flygare S, Simmon K, Miller C, Qiao Y, Kennedy B, Di Sera T, Graf EH, Tardif KD, Kapusta A, Rynearson S, Stockmann C, Queen K, Tong S, Voelkerding KV, Blaschke A, Byington CL, Jain S, Pavia A, Ampofo K, Eilbeck K, Marth G, Yandell M, Schlaberg R. Taxonomer: an interactive metagenomics analysis portal for universal pathogen detection and host mRNA expression profiling. *Genome Biol.* 2016;17(1):111. PubMed PMID: 27224977. Pubmed Central PMCID: PMC4880956. doi: 10.1186/s13059-016-0969-1
26. Huang da W, Sherman BT, Lempicki RA. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc.* 2009;4(1):44-57. PubMed PMID: 19131956. doi: 10.1038/nprot.2008.211
27. Stockmann C, Rogatcheva M, Harrel B, Vaughn M, Crisp R, Poritz M, Thatcher S, Korgenski EK, Barney T, Daly J, Pavia AT. How well does physician selection of microbiologic tests identify *Clostridium difficile* and other pathogens in paediatric diarrhoea? Insights using multiplex PCR-based detection. *Clin Microbiol Infect.* 2015;21(2):179 e9-15. PubMed PMID: 25599941. Pubmed Central PMCID: PMC4330102. doi: 10.1016/j.cmi.2014.07.011
28. Truong CY, Gombar S, Wilson R, Sundararajan G, Tekic N, Holubar M, Shepard J, Madison A, Tompkins L, Shah N, Deresinski S, Schroeder LF, Banaei N. Real-Time Electronic Tracking of Diarrheal Episodes and Laxative Therapy Enables Verification of *Clostridium difficile* Clinical Testing Criteria and Reduction of *Clostridium difficile* Infection Rates. *J Clin Microbiol.* 2017;55(5):1276-84. PubMed PMID: 28250001. Pubmed Central PMCID: PMC5405247. doi: 10.1128/JCM.02319-16
29. Cowardin CA, Kuehne SA, Buonomo EL, Marie CS, Minton NP, Petri WA, Jr. Inflammatory activation contributes to interleukin-23 production in response to *Clostrid-*

- ium difficile. *MBio*. 2015;6(1). PubMed PMID: 25626905. Pubmed Central PMCID: PMC4324312. doi: 10.1128/mBio.02386-14
30. McDermott AJ, Falkowski NR, McDonald RA, Pandit CR, Young VB, Huffnagle GB. Interleukin-23 (IL-23), independent of IL-17 and IL-22, drives neutrophil recruitment and innate inflammation during *Clostridium difficile* colitis in mice. *Immunology*. 2016;147(1):114-24. PubMed PMID: 26455347. Pubmed Central PMCID: PMC4693884. doi: 10.1111/imm.12545
 31. Zwicker S, Martinez GL, Bosma M, Gerling M, Clark R, Majster M, Soderman J, Almer S, Bostrom EA. Interleukin 34: a new modulator of human and experimental inflammatory bowel disease. *Clin Sci (Lond)*. 2015;129(3):281-90. PubMed PMID: 25896238. Pubmed Central PMCID: PMC4557398. doi: 10.1042/CS20150176
 32. Franze E, Monteleone I, Cupi ML, Mancina P, Caprioli F, Marafini I, Colantoni A, Ortenzi A, Laudisi F, Sica G, Sileri P, Pallone F, Monteleone G. Interleukin-34 sustains inflammatory pathways in the gut. *Clin Sci (Lond)*. 2015;129(3):271-80. PubMed PMID: 25800277. doi: 10.1042/CS20150132
 33. Franze E, Marafini I, De Simone V, Monteleone I, Caprioli F, Colantoni A, Ortenzi A, Crescenzi F, Izzo R, Sica G, Sileri P, Rossi P, Pallone F, Monteleone G. Interleukin-34 Induces Cc-chemokine Ligand 20 in Gut Epithelial Cells. *J Crohns Colitis*. 2016;10(1):87-94. PubMed PMID: 26449789. doi: 10.1093/ecco-jcc/jjv181
 34. Just I, Selzer J, Wilm M, von Eichel-Streiber C, Mann M, Aktories K. Glucosylation of Rho proteins by *Clostridium difficile* toxin B. *Nature*. 1995;375(6531):500-3. PubMed PMID: 7777059. doi: 10.1038/375500a0
 35. Riegler M, Sedivy R, Pothoulakis C, Hamilton G, Zacherl J, Bischof G, Cosentini E, Feil W, Schiessel R, LaMont JT, et al. *Clostridium difficile* toxin B is more potent than toxin A in damaging human colonic epithelium in vitro. *J Clin Invest*. 1995;95(5):2004-11. PubMed PMID: 7738167. Pubmed Central PMCID: PMC295778. doi: 10.1172/JCI117885
 36. Popoff MR, Rubin EJ, Gill DM, Boquet P. Actin-specific ADP-ribosyltransferase produced by a *Clostridium difficile* strain. *Infect Immun*. 1988;56(9):2299-306. PubMed PMID: 3137166. Pubmed Central PMCID: PMC259564.
 37. Swindells J, Brenwald N, Reading N, Oppenheim B. Evaluation of diagnostic tests for *Clostridium difficile* infection. *J Clin Microbiol*. 2010;48(2):606-8. PubMed PMID: 20032256. Pubmed Central PMCID: PMC2815642. doi: 10.1128/JCM.01579-09
 38. Planche TD, Davies KA, Coen PG, Finney JM, Monahan IM, Morris KA, O'Connor L, Oakley SJ, Pope CF, Wren MW, Shetty NP, Crook DW, Wilcox MH. Differences in outcome according to *Clostridium difficile* testing method: a prospective multicentre diagnostic validation study of *C difficile* infection. *Lancet Infect Dis*. 2013;13(11):936-45. PubMed PMID: 24007915. Pubmed Central PMCID: 3822406. doi: 10.1016/S1473-3099(13)70200-7
 39. Dubberke ER, Han Z, Bobo L, Hink T, Lawrence B, Copper S, Hoppe-Bauer J, Burnham CA, Dunne WM, Jr. Impact of clinical symptoms on interpretation of diagnostic assays for *Clostridium difficile* infections. *J Clin Microbiol*. 2011;49(8):2887-93.

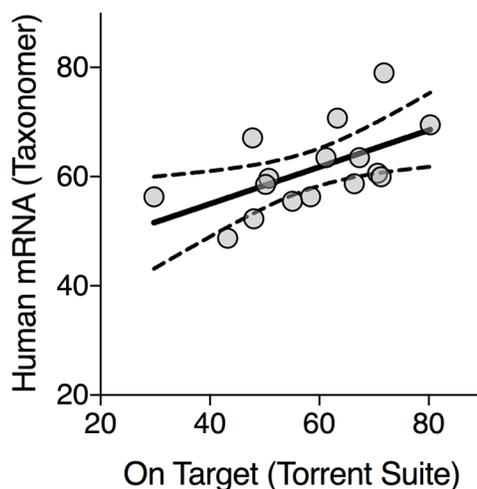
- PubMed PMID: 21697328. Pubmed Central PMCID: PMC3147743. doi: 10.1128/JCM.00891-11
40. Polage CR, Gyorke CE, Kennedy MA, Leslie JL, Chin DL, Wang S, Nguyen HH, Huang B, Tang YW, Lee LW, Kim K, Taylor S, Romano PS, Panacek EA, Goodell PB, Solnick JV, Cohen SH. Overdiagnosis of *Clostridium difficile* Infection in the Molecular Test Era. *JAMA Intern Med.* 2015;175(11):1792-801. PubMed PMID: 26348734. Pubmed Central PMCID: 4948649. doi: 10.1001/jamainternmed.2015.4114
 41. McDonald LC, Gerding DN, Johnson S, Bakken JS, Carroll KC, Coffin SE, Dubberke ER, Garey KW, Gould CV, Kelly C, Loo V, Shaklee Sammons J, Sandora TJ, Wilcox MH. Clinical Practice Guidelines for *Clostridium difficile* Infection in Adults and Children: 2017 Update by the Infectious Diseases Society of America (IDSA) and Society for Healthcare Epidemiology of America (SHEA). *Clin Infect Dis.* 2018;66(7):e1-e48. PubMed PMID: 29462280. doi: 10.1093/cid/cix1085
 42. Ahmed FE, James SI, Lysle DT, Dobbs LJ, Jr., Johnke RM, Flake G, Stockton P, Sinar DR, Naziri W, Evans MJ, Kovacs CJ, Allison RR. Improved methods for extracting RNA from exfoliated human colonocytes in stool and RT-PCR analysis. *Dig Dis Sci.* 2004;49(11-12):1889-98. PubMed PMID: 15628722.
 43. Alexander RJ, Raicht RF. Purification of total RNA from human stool samples. *Dig Dis Sci.* 1998;43(12):2652-8. PubMed PMID: 9881497.
 44. Bennett WE, Jr., Gonzalez-Rivera R, Shaikh N, Magrini V, Boykin M, Warner BB, Hamvas A, Tarr PI. A method for isolating and analyzing human mRNA from newborn stool. *J Immunol Methods.* 2009;349(1-2):56-60. PubMed PMID: 19660464. Pubmed Central PMCID: PMC2850193. doi: 10.1016/j.jim.2009.07.013
 45. Reck M, Tomasch J, Deng Z, Jarek M, Husemann P, Wagner-Dobler I, Consortium C. Stool metatranscriptomics: A technical guideline for mRNA stabilisation and isolation. *BMC Genomics.* 2015;16:494. PubMed PMID: 26140923. Pubmed Central PMCID: 4490624. doi: 10.1186/s12864-015-1694-y
 46. Giannoukos G, Ciulla DM, Huang K, Haas BJ, Izard J, Levin JZ, Livny J, Earl AM, Gevers D, Ward DV, Nusbaum C, Birren BW, Gnirke A. Efficient and robust RNA-seq process for cultured bacteria and complex community transcriptomes. *Genome Biol.* 2012;13(3):R23. PubMed PMID: 22455878. Pubmed Central PMCID: PMC3439974. doi: 10.1186/gb-2012-13-3-r23
 47. Li W, Turner A, Aggarwal P, Matter A, Storvick E, Arnett DK, Broeckel U. Comprehensive evaluation of AmpliSeq transcriptome, a novel targeted whole transcriptome RNA sequencing methodology for global gene expression analysis. *BMC Genomics.* 2015;16:1069. PubMed PMID: 26673413. Pubmed Central PMCID: PMC4681149. doi: 10.1186/s12864-015-2270-1
 48. FitzGerald LM, Jung CH, Wong EM, Joo JE, Gould JA, Vasic V, Bassett JK, O'Callaghan N, Nottle T, Pedersen J, Giles GG, Southey MC. Obtaining high quality transcriptome data from formalin-fixed, paraffin-embedded diagnostic prostate tumor specimens. *Lab Invest.* 2018. PubMed PMID: 29339835. doi: 10.1038/s41374-017-0001-8

49. Bhattacharya S, Rosenberg AF, Peterson DR, Grzesik K, Baran AM, Ashton JM, Gill SR, Corbett AM, Holden-Wiltse J, Topham DJ, Walsh EE, Mariani TJ, Falsey AR. Transcriptomic Biomarkers to Discriminate Bacterial from Nonbacterial Infection in Adults Hospitalized with Respiratory Illness. *Sci Rep.* 2017;7(1):6548. PubMed PMID: 28747714. Pubmed Central PMCID: PMC5529430. doi: 10.1038/s41598-017-06738-3
50. McClain MT, Nicholson BP, Park LP, Liu TY, Hero AO, 3rd, Tsalik EL, Zaas AK, Veldman T, Hudson LL, Lambkin-Williams R, Gilbert A, Burke T, Nichols M, Ginsburg GS, Woods CW. A Genomic Signature of Influenza Infection Shows Potential for Presymptomatic Detection, Guiding Early Therapy, and Monitoring Clinical Responses. *Open Forum Infect Dis.* 2016;3(1):ofw007. PubMed PMID: 26933666. Pubmed Central PMCID: PMC4771939. doi: 10.1093/ofid/ofw007
51. Ramilo O, Mejias A, Mahajan P, Kuppermann N. RNA signature test to distinguish bacterial from viral infection. *J Pediatr.* 2017;182:401-4. PubMed PMID: 28237452. doi: 10.1016/j.jpeds.2016.12.066
52. Tsalik EL, Henao R, Nichols M, Burke T, Ko ER, McClain MT, Hudson LL, Mazur A, Freeman DH, Veldman T, Langley RJ, Quackenbush EB, Glickman SW, Cairns CB, Jaehne AK, Rivers EP, Otero RM, Zaas AK, Kingsmore SF, Lucas J, Fowler VG, Jr., Carin L, Ginsburg GS, Woods CW. Host gene expression classifiers diagnose acute respiratory illness etiology. *Sci Transl Med.* 2016;8(322):322ra11. PubMed PMID: 26791949. Pubmed Central PMCID: PMC4905578. doi: 10.1126/scitranslmed.aad6873
53. Zaas AK, Burke T, Chen M, McClain M, Nicholson B, Veldman T, Tsalik EL, Fowler V, Rivers EP, Otero R, Kingsmore SF, Voora D, Lucas J, Hero AO, Carin L, Woods CW, Ginsburg GS. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. *Sci Transl Med.* 2013;5(203):203ra126. PubMed PMID: 24048524. Pubmed Central PMCID: PMC4286889. doi: 10.1126/scitranslmed.3006280

SUPPLEMENTARY FIGURES



Supplementary Figure 1. Sequencing quality metrics for expression profiling of mRNA extracted from feces of patients with CDI and diarrhea of other causes.



Supplementary Figure 2. Correlation of sample composition and expression profiling by Taxonomer (y-axis, percentage of human mRNA reads) and Torrent Suite (x-axis, percentage reads on target).

SUPPLEMENTARY TABLES

Supplementary Table 1. Top 50 Genes Used for DAVID Analysis

UNKL	GPRC5C	WBP1L
FAM3A	NAALADL1	CDC26
AIP	HIPK2	FAM3C
PSMD11	MDK	SDSL
SNX4	POLD4	TARSL2
ARHGAP35	KCTD5	ALOX15
EXOSC8	NIPA1	TNNT1
SPTBN1	BCL2L11	TNKS1BP1
GTF2F1	EIF4A2	ARHGAP42
PLCB3	MPST	OTUD5
C22orf28	BMF	JUP
IFT122	SERPINB6	KLHDC8A
CDK8	EFNA3	DUSP23
MTA1	TARS	TUBB2A
RAP2A	SLC47A1	NFU1
SLC28A1	LZTS2	CHID1
NAA38	ANKRD11	

Supplementary Table 2. Enriched functional annotations of the top 50 differentially expressed genes (DAVID analysis) and included genes.

GO Term	GO Accession	Gene
Cytoskeletal protein binding	0008092	TNKS1BP1, JUP, TNNT1, SPTBN1, C22ORF28, BCL2L11
Sarcomere	0030017	JUP, TNNT1, SPTBN1
Myofibril	0030016	JUP, TNNT1, SPTBN1
Contractile fiber part	0044449	JUP, TNNT1, SPTBN1
Actin cytoskeleton	0015629	JUP, TNNT1, SPTBN1, BMF

FOOTNOTES

Submitted July 8, 2018 | Accepted August 31, 2018 | Published September 12, 2018

COPYRIGHT

© Pathogens and Immunity 2018

This work is licensed under a Creative Commons Attribution 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>