

TITLE: Pulmonary Metagenomic Sequencing Suggests Missed Infections in Immunocompromised Children

AUTHORS: Zinter, MS;¹ Dvorak, CC;² Mayday, MY;¹ Iwanaga, K;³ Ly, NP;³ McGarry, ME;³ Church, GD;³ Faricy, LE;³ Rowan, CM;⁴ Hume, JR;⁵ Steiner, ME;^{5,6} Crawford, ED;^{7,8} Langelier, C;⁹ Kalantar, K,⁸ Chow, ED;⁸ Miller, S;¹⁰ Shimano, K;² Melton, A;² Yanik, GA;¹¹ Sapru, A,^{1,12} DeRisi, JL.^{7,8}

AFFILIATIONS:

¹ UCSF School of Medicine, Benioff Children's Hospital, Department of Pediatrics, Division of Critical Care, San Francisco, CA, USA

² UCSF School of Medicine, Benioff Children's Hospital, Department of Pediatrics, Division of Allergy, Immunology, and Blood & Marrow Transplantation, San Francisco, CA, USA

³ UCSF School of Medicine, Benioff Children's Hospital, Department of Pediatrics, Division of Pulmonology, San Francisco, CA, USA

⁴ Indiana University School of Medicine, Riley Hospital for Children, Department of Pediatrics, Division of Critical Care, Indianapolis, IN, USA

⁵ University of Minnesota School of Medicine, Masonic Children's Hospital, Department of Pediatrics, Division of Critical Care, Minneapolis, MN, USA

⁶ University of Minnesota School of Medicine, Masonic Children's Hospital, Department of Pediatrics, Division of Hematology/Oncology, Minneapolis, MN, USA

⁷ Chan Zuckerberg Biohub, San Francisco, CA, USA

⁸ UCSF School of Medicine, Department of Biochemistry & Biophysics, San Francisco, CA, USA

⁹ UCSF School of Medicine, Department of Internal Medicine, Division of Infectious Diseases, San Francisco, CA, USA

¹⁰ UCSF School of Medicine, Department of Laboratory Medicine, San Francisco, CA, USA

¹¹ University of Michigan School of Medicine, Mott's Children's Hospital, Department of Pediatrics, Division of Oncology, Ann Arbor, MI, USA

¹² UCLA Geffen School of Medicine, Mattel Children's Hospital, Department of Pediatrics, Division of Critical Care, Los Angeles, CA, USA

CONTACT: Matt Zinter, MD, University of California - San Francisco, Department of Pediatrics, Division of Critical Care Medicine, 550 16th St, San Francisco, CA 94143, USA, matt.zinter@ucsf.edu

AUTHOR CONTRIBUTIONS:

Study concept and design: MSZ, CCD, AS, JDL

Acquisition of data: MSZ, CCD, MYM, KI, NPL, MEM, GDC, LEF, CMW, JRH, MES, SM, KS, AM, GAY, AS, JDL

Analysis and interpretation of data: MSZ, CCD, MYM, EDC, CL, KK, EDC, SM, JDL

Drafting of the manuscript: MSZ, CCD, MYM, JDL

Critical revision of the manuscript for important intellectual content: MSZ, CCD, MYM, KI, NPL, MEM, GDC, LEF, CMW, JRH, MES, EDC, CL, KK, EDC, SM, KS, AM, GAY, AS, JDL

Statistical analysis: MSZ, MYM, KK, JDL

Administrative, technical, or material support: MSZ, CCD, SM, AS, JDL

Study supervision: MSZ, CCD, AS, JDL

Approval of final manuscript: MSZ, CCD, MYM, KI, NPL, MEM, GDC, LEF, CMW, JRH, MES, EDC, CL, KK, EDC, SM, KS, AM, GAY, AS, JDL

FUNDING: This work was supported by the NIH NICHD K12HD000850, the Pediatric Blood and Marrow Transplant Foundation, and the National Marrow Donor Program Amy Strelzer Manasevit Grant (Zinter); the NIH NHLBI K23HL085526 and R01HL114484 (Sapru); the Chan Zuckerberg Biohub (DeRisi).

RUNNING TITLE: Pulmonary mNGS in Immunocompromised Children

DESCRIPTOR: 10.11 Pediatrics: Respiratory Infections

WORD COUNT: 3,500

SUPPLEMENTAL DATA: This article has an online data supplement. Non-human reads from each patient are deposited in the NCBI SRA (BioProject ID PRJNA437169).

ABSTRACT

RATIONALE: Despite improved diagnostics, pulmonary pathogens in immunocompromised children frequently evade detection, leading to significant morbidity and mortality.

OBJECTIVES: To develop a highly sensitive metagenomic next generation sequencing (mNGS) assay capable of evaluating the pulmonary microbiome and identifying diverse pathogens in the lungs of immunocompromised children.

METHODS: We collected 41 lower respiratory specimens from 34 immunocompromised children undergoing evaluation for pulmonary disease at 3 children's hospitals from 2014-2016. Samples underwent mechanical homogenization, paired RNA/DNA extraction, and metagenomic sequencing. Sequencing reads were aligned to the NCBI nucleotide reference database to determine taxonomic identities. Statistical outliers were determined based on abundance within each sample and relative to other samples in the cohort.

MEASUREMENTS & MAIN RESULTS: We identified a rich cross-domain pulmonary microbiome containing bacteria, fungi, RNA viruses, and DNA viruses in each patient. Potentially pathogenic bacteria were ubiquitous among samples but could be distinguished as possible causes of disease by parsing for outlier organisms. Samples with bacterial outliers had significantly depressed alpha-diversity (median 0.58, IQR 0.33-0.62 vs. median 0.94, IQR 0.93-0.95, $p < 0.001$). Potential pathogens were detected in half of samples previously negative by clinical diagnostics, demonstrating increased sensitivity for missed pulmonary pathogens ($p < 0.001$).

CONCLUSIONS: An optimized mNGS assay for pulmonary microbes demonstrates significant inoculation of the lower airways of immunocompromised children with diverse bacteria, fungi, and viruses. Potential pathogens can be identified based on absolute and relative abundance. Ongoing investigation is needed to determine the pathogenic significance of outlier microbes in the lungs of immunocompromised children with pulmonary disease.

ABSTRACT WORD COUNT: 248

MeSH KEYWORDS: Intensive Care Units, Pediatric; Immunocompromised Host; Metagenomics; Respiratory Tract Infections; Microbiota

BACKGROUND

Last year in the United States, approximately 15,000 children were diagnosed with cancer, 2,000 children underwent solid organ transplantation (SOT), and 2,500 children underwent hematopoietic cell transplantation (HCT) for an increasingly broad set of life-threatening diseases (1-3). While significant progress has been made in improving the safety of anti-neoplastic and transplantation-based therapies, the risk of infectious complications remains high (4). In particular, lower respiratory tract infections in children who have undergone HCT are associated with mortality rates exceeding 45% (5, 6). Current microbiologic diagnostics are inadequate and fail to identify a pathogenic organism in over one third of such cases (7). Autopsy case series have identified previously undetected pulmonary pathogens in 30-50% of pediatric HCT patients, underscoring current diagnostic limitations and the significant mortality associated with undiagnosed pulmonary infections in such patients (8).

Recent advances in metagenomic next generation sequencing (mNGS) have shown promising results in diagnosing neurologic, ocular, and other infections (9-11). However, the identification of fastidious pathogens such as *Aspergillus* and other filamentous molds remain difficult due to thick extracellular matrices and the relatively small inoculum required to induce disease (12-15). Unfortunately, off-the-shelf assays for respiratory biospecimens have proven inadequate to survey the variety of organisms present in thick and mucoid respiratory secretions. As such, unlike the better-characterized microbiomes of the human gastrointestinal tract and nasopharynx, data describing the composition of the pulmonary microbiome are sparse and insufficient to reliably discriminate between health and disease (16).

Therefore, we conducted a pilot study aimed to (1) develop and optimize a highly sensitive mNGS assay capable of detecting pulmonary bacterial, fungal, and viral pathogens clinically relevant to immunocompromised children, and (2) test this mNGS assay on a cohort of immunocompromised children undergoing lower respiratory tract sampling as evaluation for suspected pulmonary infection. We hypothesized that an optimized mNGS assay could improve characterization of the pulmonary microbiome and aid in the identification of potential pulmonary pathogens in this high-risk population.

METHODS

Development of Optimized mNGS Assay: Our laboratory has previously characterized high-throughput mNGS protocols for use with human cerebrospinal and intravitreal fluid (9-11). Given the challenges of diagnosing pulmonary infections, we created a set of mock-positive BALs by spiking a pre-determined number of colony forming units (CFUs) of an *Aspergillus niger* clinical isolate into aliquots of BAL with a previously identified *Haemophilus influenzae*/*Human Adenovirus B* co-infection (**Supplemental Text 1**). *Aspergillus* was chosen as an optimization benchmark given its thick polysaccharide cell wall and extreme clinical importance in this patient population.

Sample Preparation: We tested nucleic acid extraction conditions by combining 200 μ L of mock-positive BAL with either 600 μ L DirectZol, 600 μ L Lysis Buffer, or 200 μ L DNA/RNA Shield (Zymo), followed by mechanical homogenization with either 0.1mm or 0.5mm glass bashing beads (Omni) for 2, 5, or 8 cycles of 25 seconds bashing at 30Hz with 60 seconds rest on ice between each cycle (TissuerLyser II, Qiagen). Samples homogenized in DNA/RNA shield also underwent enzymatic mycolysis with either 0.15mg or 0.38mg Proteinase K at 23°C for 30 minutes or 60 minutes (Zymo), or with 0.4mg, 1.2mg, 4mg, or 8mg Yatalase (Takara Bio Inc) at 23°C or 37°C for 60 or 90 minutes. Subsequently, all samples underwent 10 minutes of centrifugation at 4°C and the supernatant was used for paired DNA/RNA extraction (Zymo ZR-Duet DNA/RNA MiniPrep Kit). *Aspergillus* nucleic acid yield was assessed using an orthogonal digital droplet PCR assay with *Aspergillus*-specific primers (ddPCR, **Supplemental Text 2**) (17). Subsequently, sequencing libraries were prepared using the New England Biolabs NEBNext RNA/DNA Ultra-II Library Prep Kit and underwent 125 nucleotide paired-end sequencing on an Illumina HiSeq 4000 instrument (**Supplemental Text 3**).

Bioinformatics Pipeline: Resultant .fastq files were processed using a previously described pipeline consisting of several open-source components (9, 10). Briefly, reads underwent iterative removal of host (Hg38/PanTro), low quality, low complexity (LZW compression ratio >0.45), and redundant sequences using STAR, Bowtie2, PriceSeqFilter, and CD-HIT-DUP (18-21). The remaining sequences were then

aligned to the NCBI non-redundant nucleotide database using GSNAPL for assignment of bacterial, fungal, viral and other taxonomic IDs (22). Microbes were described as potentially pathogenic or typically non-pathogenic based on *a priori* literature review (**Supplemental Text 4**).

Analysis: To avoid spurious alignments and sequencing artifacts, we only included read pairs where both ends aligned to the same taxonomic ID. The prevalence of each microbe was described (1) *relative to other microbes in the same sample*, wherein we normalized sequencing reads per million total sequencing reads (rpm); and (2) *relative to the same microbe in other samples in the cohort*, wherein we normalized sequencing reads as the number of standard deviations above or below the mean \log_{10} -transformed rpm for the total cohort (Z-score) (23). Outliers were defined as microbes with RNA alignments that were both abundant within a sample ($\geq 10\%$ of all bacterial reads in a sample, or >1 rpm for fungi and viruses) and ≥ 2 standard deviations above the cohort mean for that particular microbe (Z-score ≥ 2). A subset of outliers were confirmed via orthogonal assays. The Simpson's Diversity Index was used to associate the loss of bacterial diversity with the presence of outlier microbes (detailed methods in **Supplemental Text 5**) (24). Sequencing reads are deposited in the NCBI dbGaP (BioProject ID PRJNA437169).

Patients: To test the optimized mNGS assay, we prospectively collected 41 lower respiratory samples from 34 immunocompromised patients age ≤ 25 years who underwent clinically-indicated lower respiratory diagnostics between September 2014 and April 2016 at the University of California San Francisco Benioff Children's Hospital, Indiana University Riley Hospital for Children, and the University of Minnesota Masonic Children's Hospital (**Supplemental Text 6**). Each sample was labeled as having (1) a pathogen causing respiratory disease, (2) a pathogen not causing respiratory disease, or (3) no pathogen according to interpretation of clinical microbiologic testing as documented by the treating physician. This study was approved by each site's Institutional Review Board with consent provided by all patients or their surrogates.

RESULTS

Development of Optimized mNGS Assay:

Iterative optimizations demonstrated that mechanical homogenization of BAL using 0.5mm inert bashing beads for 5 cycles in a DNA/RNA shield without mycolytic enzymes was superior to all other approaches for isolation of microbial nucleic acid (**Supplemental Text 7, Supplemental Figures 1a-d**). At a depth of 25 million reads per 200uL, this reduced the lower limit of detection (LLOD) of RNA sequencing from 59.60 *Aspergillus* CFU (95% CI 37.70-95.36) to 0.42 *Aspergillus* CFU (95% CI 0.12-1.40, paired T-test $p < 0.001$, **Supplemental Figure 2**), demonstrating an approximately 100-fold increased sensitivity for *Aspergillus*. In comparison, the optimized DNA sequencing assay was able to detect as few as 6.13 *Aspergillus* CFU (95% CI 4.16-9.04), suggesting RNAseq to be approximately 10-fold more sensitive than DNAseq for detecting *Aspergillus* nucleic acid in respiratory specimens (paired T-test $p < 0.001$, **Figure 1**). The optimization did not change the detection of *H.influenzae* or *Human Adenovirus B* (T-test $p = 0.343$ and $p = 0.420$, respectively). These results suggest that only the optimized RNAseq protocol had a sufficiently dynamic range to detect *Aspergillus* nucleic acid in quantities analogous to rare growth in culture.

Application of mNGS Assay:

We enrolled 34 patients and collected 41 lower respiratory samples for this pilot project (**Table 1**). Using the optimized extraction protocol described above, we conducted mNGS resulting in 21.2 ± 4.2 million reads per sample. After quality filtering and removal of human reads, the remaining reads aligning to bacteria, fungi, and viruses were analyzed further as discussed below. Detailed reports for each patient exist in **Appendix 1**.

Bacteria: The vast majority of taxa derived from bacterial alignments were present at low abundance and in quantities similar across all samples in the cohort. Specifically, 95.9% of bacterial genera composed <10% of the total bacterial alignments in each sample and were present at levels within 2 standard deviations of the cohort mean (Z -score <2; **Figure 2a**). These included potentially pathogenic bacteria such as *Escherichia*, *Klebsiella*, *Pseudomonas*, *Staphylococcus*, *Stenotrophomonas*, and *Streptococcus*, which were identified in nearly all patient samples (38, 39, 40, 38, 41, and 40 of 41 samples, respectively).

Conversely, only 0.2% of detected bacterial genera composed $\geq 10\%$ of their respective pulmonary microbiome and were present at levels greater than 2 standard deviations above the cohort mean (Z -score ≥ 2). These outliers included 9 potentially pathogenic bacteria identified in 13/41 patient samples (*Corynebacterium*, *Enterobacter*, *Escherichia*, *Fusobacterium*, *Haemophilus*, *Mycoplasma*, *Pseudomonas*, *Rothia*, and *Staphylococcus*). These results suggest that parsing pulmonary bacterial reads according to both absolute and relative abundance is an effective way to identify potentially pathogenic bacteria from microbially-rich respiratory samples. Samples with outlier bacterial pathogens had depressed diversity of the bacterial microbiome (median 0.58, IQR 0.33-0.62 vs. median 0.94, IQR 0.93-0.95, $p < 0.001$, **Supplemental Figure 3**), and Simpson's Diversity Index cutoffs of ≥ 0.8 or ≥ 0.9 showed 87.1% (95% CI 74.7-93.9) and 96.2% (95% CI 79.0-99.4) negative predictive value for the presence of an outlier bacterial pathogen, suggesting that the identification of bacterial dysbiosis may be a useful screen for recognizing possible bacterial infections.

Fungi: Relative to bacterial alignments, fungal alignments were significantly less prevalent in the cohort. 92.1% of fungal genera were detected at less than a single read per million reads mapped (**Figure 2b**). Approximately half of all identified fungal genera were of low abundance and were present at levels similar to other samples in the cohort (50.6%); an additional 41.5% were of low abundance but were rare in the cohort. Evidence supporting these alignments is limited by the rarity of the reads themselves. Only 3.4% of fungal genera were both abundant and had a Z -score ≥ 2 . These outliers included 7 potentially pathogenic

fungi identified a total of 8 times in 7/41 patient samples (*Alternaria*, *Aspergillus*, *Candida*, *Cladosporium*, *Cryptococcus*, *Fusarium*, and *Pneumocystis*).

Next, to contextualize the burden of *Aspergillus* in each sample, we estimated the number of *Aspergillus* CFU according to the linear regression associating *Aspergillus* CFU with RNA transcripts derived from our spike-in experiments. *Aspergillus* RNA levels correlated with ≥ 1 CFU in 14/41 samples, suggesting that the lower airways of immunocompromised children are frequently exposed to quantities of *Aspergillus* RNA consistent with low numbers of CFUs. Of these 14 samples, one patient sample was positive for IPA by both culture and galactomannan assay (**Figure 2b, labeled**). Although the remaining 13 were all culture-negative, 12/13 had received fungicidal or fungistatic anti-*Aspergillus* pharmacotherapy within 48 hours of sample collection, suggesting that empiric antifungal pharmacotherapy significantly confounds the association between *Aspergillus* RNA and growth in culture. While *Aspergillus* RNA sequencing reads did not correlate with growth in culture (T-test $p=0.148$), they did correlate with BAL galactomannan (regression correlation coefficient 0.74, 95% CI 0.43-1.06, $p<0.001$ **Supplemental Figure 4**). Samples with *Aspergillus* RNA equivalent to ≥ 1 CFU had a positive likelihood ratio of 6.0 for having galactomannan ≥ 0.5 (95% CI 2.5-14.7, $p=0.001$). These data suggest that while *Aspergillus* growth in culture lacks sensitivity for detecting IPA, mNGS may augment the diagnostic utility of the galactomannan assay.

Viruses: In this cohort, detected respiratory viruses were dominated by RNA viruses, with *Adenovirus* being the only significant outlying DNA virus. In contrast with bacteria and fungi, a significantly larger portion of RNA alignments to viral genera were present at high abundance, with the majority (30.5%) present in quantities significantly greater than the cohort mean ($Z \geq 2$; **Figure 2c, 2d**). Communicable respiratory viruses with abundant alignments were identified in approximately one third of all patient samples (13/41) and included *Adenoviruses A and C*, *Bocavirus*, *Coronaviruses 229E and OC43*, *Influenzaviruses A and C*, *Parainfluenzavirus 3*, and *Rhinoviruses A and C*. Although 4 *Rhinoviruses* were classified as outliers (Z -score ≥ 2), an additional 3 *Rhinoviruses* were abundant but had Z -scores between

+0.92 and +1.78. Additionally, two patients had viral co-infections (*Parainfluenza-3* and *Influenza-C*; *Adenovirus-C* and *Rhinovirus-A*).

Approximately half of viral genera were detected below 1 read per million sequencing reads (50.9%). These included additional communicable respiratory viruses at low abundance in 15% of the cohort (6/41). Although there were no cases of clinically suspected herpesvirus pneumonitis, *Epstein-Barr virus*, *Cytomegalovirus*, *HHV-6*, and *HHV-7* were identified in low abundance in 9/41 samples. Viral genera of uncertain or unlikely pathogenicity were also identified in 21/41 samples and included *Papillomaviruses*, *WU* and *KI Polyomaviruses*, and *Torquetenoviruses* (2, 5, and 18 of 41 samples, respectively).

Comparison to Clinical Testing: Clinical testing identified causative pathogens in 36.6% of samples (n=15), and each of these were concordantly identified by mNGS, although 4 were not classified as outliers (**Figure 3**). mNGS also identified 4 previously undetected potential co-pathogens in this group (*Bocavirus*, *Influenza-C*, and *Coronavirus* twice). Clinical testing identified non-causative pathogens in 12.2% of samples (n=5); 4 were concordantly identified by mNGS and none were ranked as outliers. However, abundant transcripts aligning to *Human Coronavirus 229E* provided an alternative explanation for pulmonary disease in two of these patients. Clinical testing did not identify any pathogens in 51.2% of samples (n=21). Here, mNGS was able to identify statistically outlying potential pathogens in 10/21 cases. These potential pathogens included a variety of bacteria (ie: *P.aeruginosa*, *E.cloacae*, *M.pneumoniae*), fungi (ie: *C.glabrata*), and viruses (ie: *Rhinovirus-A*, *Influenza-A*). Among patients whose clinical testing did *not* identify a causative pathogen, we observed a non-significant trend towards greater mortality in patients with an outlier pathogen detected on mNGS (5/11 vs. 1/12, Fisher p=0.068).

Orthogonal Validation: Organisms detected by mNGS but missed during clinical testing were validated with commercially available CLIA-approved assays, including multiplex respiratory viral PCR,

Mycoplasma-specific PCR, and 16S bacterial and 28S/ITS fungal amplicon sequencing (**Supplemental Text 5**). All validation tests were concordant with mNGS with one exception: 28S/ITS amplicon DNA sequencing failed to identify *C.glabrata* in Sample 37. For this sample, we confirmed the presence of this organism with 3 separate species-specific RT-PCR primer sets followed by Sanger sequencing in-house, confirming the initial mNGS report of *C.glabrata*. As there are numerous complementary bioinformatics approaches for analyzing metagenomic datasets, we tested the reproducibility of our approach by re-analyzing our data using a parallel method developed in-house that weights organisms by both abundance and concordance of paired RNA and DNA sequencing reads without Z-scores (**Supplemental Table 1**). The results of this analytical approach and our primary analytical approach showed strong agreement. In sum, these data demonstrate that mNGS has significantly greater sensitivity for detecting potential pulmonary pathogens than current clinical diagnostics (McNemar's $p < 0.001$).

DISCUSSION

In this study, we developed an optimized mNGS assay with adequate sensitivity to identify bacteria, fungi, and both RNA and DNA viruses within the lower respiratory tract of immunocompromised children. In doing so, we identified a rich molecular portrait of the pulmonary microbiome in this vulnerable patient population. Further, by comparing the quantity of microbial nucleic acid to other microbes within a sample and to other samples within the cohort, we were able to identify outlying potential pathogens in approximately half of previously-negative samples.

Due to inherent challenges in sampling the lower respiratory tract, the pulmonary microbiome was not one of the original sites sampled in the 2008 Human Microbiome Project and data regarding pulmonary microbial communities in health and disease have lagged decades behind similar analyses of human intestinal, cutaneous, and nasopharyngeal microbiomes (27). In this study, we detected that many pathogenic bacteria such as *Pseudomonas* and *Streptococcus* are ubiquitous and hence their abundance

needs to be contextualized by cohort-specific norms. By using cohort-specific Z-scores, we provide a template for discriminating normal from abnormal microbial burden in the lungs of immunocompromised children. For example, 100% of samples had detectable *Pseudomonas* RNA, but only Sample 29 had detectable *Pseudomonas* RNA more than 2 standard deviations above the cohort mean (**Appendix 1**).

While prior metagenomic sequencing assays for human biospecimens have demonstrated excellent detection of viral and bacterial nucleic acid, they have lacked sensitivity for detecting fastidious organisms such as filamentous mold (12, 13). This study confirms the need for aggressive mechanical homogenization in stabilizing media in order to detect molds such as *Aspergillus* while simultaneously preserving the detection of bacteria and viruses (25). Despite this, transcripts aligning to fungi remained significantly less ubiquitous than transcripts aligning to bacteria, perhaps due to relative organism abundance. Although the majority of commercial sequencing assays use DNA, our data demonstrate that RNA sequencing is >10 times as sensitive as DNA sequencing for the detection of such fastidious organisms, which we speculate may be due to high transcriptional activity of regions such as rRNA genes in active organisms (26).

Surprisingly, *Aspergillus* RNA equivalent to ≥ 1 CFU was detected in 34.1% of samples and no water controls, suggesting that the lungs of immunocompromised children are frequently exposed to low levels of *Aspergillus*. The 100-fold increase in *Aspergillus* nucleic acid after optimized mechanical homogenization suggests that the origin of the *Aspergillus* RNA is likely from intact fungal organisms in the lower respiratory tract of immunocompromised children, rather than from migration of extracellular nucleic acid down the respiratory tree. However, as only 10% of samples originated from patients with suspected IPA, we speculate that patient-specific factors such as immune reconstitution, alloreactive inflammation, and impaired mucociliary clearance are as important as inoculum exposure in determining which child might develop IPA (28). Although *Aspergillus* growth in culture appeared to be confounded by the frequent use of prophylactic and empiric antifungals, the quantity of *Aspergillus* sequencing reads was correlated with BAL galactomannan (29, 30). As nutrient availability and the kinetics of *Aspergillus* growth may affect both galactomannan release and bioavailability of *Aspergillus* nucleic acid, we advocate

that RNA/DNA-based assays might complement but should not replace antigen-based assays in the diagnosis of IPA (31, 32). While our mNGS assay was highly sensitive for *Aspergillus* detection, specificity for differentiating infection from colonization was limited; future studies pairing mNGS with host susceptibility and immune function may improve discrimination of colonization and invasive mycosis (33-36).

Consistent with emerging data in pediatrics, this work also detected herpesviruses such as *CMV* and *HHV-6* in low abundance across many children in the cohort (37, 38). Interestingly, a number of typically non-pathogenic viruses such as *Torquetenovirus* and *KI Polyomavirus* were present in over half the cohort. While data supporting the direct pathogenicity of these viruses are lacking, they may be a marker of immune function and a predictor of post-transplant infectious and alloimmune complications, including lung injury in general (39-41). Future studies aimed at characterizing longitudinal changes in the pulmonary virome of immunocompromised children are warranted.

By characterizing the distribution of pulmonary microbes within the cohort, we were able to identify statistically outlying potential pathogens in half of previously negative samples. Most cultivable bacteria and fungi that were not detected clinically were isolated from patients who had already received antimicrobial therapies, highlighting the importance of culture-independent techniques (16). In addition, several pathogenic viruses were not detected clinically due to omission from multiplex PCR assays (*Human coronavirus*, *Human bocavirus*, *Influenzavirus-C*) (42, 43). Unfortunately, a number of pathogens detected by mNGS currently lack targeted therapeutics, and as such, the development of pharmacologic and cell-based therapies remains a priority in this field (44-46). In addition, while this study suggests that many idiopathic pulmonary complications may indeed be triggered by infections, half of samples without pathogens identified by clinical testing were concordantly negative on mNGS, emphasizing the ongoing clinical significance of non-infectious pulmonary complications in this population. Future clinical validation of mNGS may demonstrate utility in safely excluding infections and allowing amplified

immunosuppression in patients with suspected non-infectious pulmonary syndromes such as hepatic veno-occlusive disease, thrombotic microangiopathy, and pulmonary graft-versus-host disease (47, 48).

Strengths and limitations:

Our study has several strengths. First, we optimized the identification of *Aspergillus* RNA and contextualized *Aspergillus* sequencing reads by comparing with both growth in culture and BAL galactomannan. Second, we proposed a logical analytical framework that ranks organism abundance both within a sample and relative to other samples; this framework allowed us to rapidly exclude more than 95% of identified bacteria. Third, we provide, to our knowledge, the first evaluation of the pulmonary microbiome in immunocompromised children.

This pilot study has several limitations. First, while using standard Z-scores was useful in de-emphasizing commonly abundant organisms (ie: *S.pneumoniae*), it may have overvalued uncommon pathogens with less abundant transcripts (ie: *CMV*) or undervalued frequently encountered pathogens (ie: *Rhinovirus*). Additional larger studies aimed at quantifying pulmonary microbial communities in immunocompromised children will naturally improve the value of Z-score analysis. Second, as with all mNGS assays, the identification of microbial nucleic acid does not directly confirm the presence of viable, live organisms, does not directly implicate that microbe as a contributor to pulmonary disease, and does not exclude less abundant organisms as potential contributors to pulmonary disease. In order to optimize patient outcomes, we advocate for ongoing multidisciplinary collaboration between intensivists, immunologists, oncologists, transplant physicians, infectious disease physicians, pulmonologists, bioinformaticians, and basic scientists.

CONCLUSIONS

In summary, we present a mNGS assay optimized to evaluate the pulmonary microbiome and detect potential pathogens in immunocompromised children. This assay is highly sensitive, reveals a rich bacterial, fungal, and viral pulmonary microbiome, and further identified potential pathogens in half of previously negative samples. As such, advanced organism detection offers the potential for early implementation of targeted therapy and the possibility for improved clinical outcomes in immunocompromised children. We invite the scientific and clinical community to participate in an ongoing multicenter collaborative clinical trial aimed at further refining this emerging technology (<https://clinicaltrials.gov/ct2/show/NCT02926612>).

REFERENCES:

- (1) Center for International Blood and Marrow Transplant, a contractor for the C.W. Bill Young Cell Transplantation Program operated through the U. S. Department of Health and Human Services, Health Resources and Services Administration, Healthcare Systems Bureau. U.S. Transplant Data by Age, Number of Transplants Reported in 2016 . *U S Transplant Data by Center Report* July 28, 2016;2017.
- (2) OPTN/SRTR 2015 Annual Data Report: Introduction. *Am J Transplant* 2017;17 Suppl 1:11-20.
- (3) Siegel RL, Miller KD, Jemal A. Cancer Statistics, 2017. *CA Cancer J Clin* 2017;67:7-30.
- (4) Kaya Z, Weiner DJ, Yilmaz D, Rowan J, Goyal RK. Lung function, pulmonary complications, and mortality after allogeneic blood and marrow transplantation in children. *Biol Blood Marrow Transplant* 2009;15:817-826.
- (5) Forslow U, Mattsson J, Ringden O, Klominek J, Remberger M. Decreasing mortality rate in early pneumonia following hematopoietic stem cell transplantation. *Scand J Infect Dis* 2006;38:970-976.
- (6) Zinter MS, Dvorak CC, Spicer A, Cowan MJ, Sapru A. New Insights Into Multicenter PICU Mortality Among Pediatric Hematopoietic Stem Cell Transplant Patients. *Crit Care Med* 2015;43:1986-1994.
- (7) Aguilar-Guisado M, Jimenez-Jambrina M, Espigado I, Rovira M, Martino R, Oriol A, Borrell N, Ruiz I, Martin-Davila P, de la Camara R, Salavert M, de la Torre J, Cisneros JM, Spanish Network for Research in Infectious Diseases. Pneumonia in allogeneic stem cell transplantation recipients: a multicenter prospective study. *Clin Transplant* 2011;25:E629-38.
- (8) Roychowdhury M, Pambuccian SE, Aslan DL, Jessurun J, Rose AG, Manivel JC, Gulbahce HE. Pulmonary complications after bone marrow transplantation: an autopsy study from a large transplantation center. *Arch Pathol Lab Med* 2005;129:366-371.

- (9) Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu G, Salamat SM, Somasekar S, Federman S, Miller S, Sokolic R, Garabedian E, Candotti F, Buckley RH, Reed KD, Meyer TL, Seroogy CM, Galloway R, Henderson SL, Gern JE, DeRisi JL, Chiu CY. Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing. *N Engl J Med* 2014.
- (10) Doan T, Wilson MR, Crawford ED, Chow ED, Khan LM, Knopp KA, O'Donovan BD, Xia D, Hacker JK, Stewart JM, Gonzales JA, Acharya NR, DeRisi JL. Illuminating uveitis: metagenomic deep sequencing identifies common and rare pathogens. *Genome Med* 2016;8:90-016-0344-6.
- (11) Langelier C, Zinter MS, Kalantar K, Yanik GA, Christenson S, O'Donovan B, White C, Wilson M, Sapru A, Dvorak CC, Miller S, Chiu CY, DeRisi JL. Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant Patients. *Am J Respir Crit Care Med* 2017.
- (12) Bittinger K, Charlson ES, Loy E, Shirley DJ, Haas AR, Laughlin A, Yi Y, Wu GD, Lewis JD, Frank I, Cantu E, Diamond JM, Christie JD, Collman RG, Bushman FD. Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. *Genome Biol* 2014;15:487.
- (13) Clarke EL, Lauder AP, Hofstaedter CE, Hwang Y, Fitzgerald AS, Imai I, Biernat W, Rekawiecki B, Majewska H, Dubaniewicz A, Litzky LA, Feldman MD, Bittinger K, Rossman MD, Patterson KC, Bushman FD, Collman RG. Microbial Lineages in Sarcoidosis: A Metagenomic Analysis Tailored for Low Microbial Content Samples. *Am J Respir Crit Care Med* 2017.
- (14) Heng SC, Morrissey O, Chen SC, Thursky K, Manser RL, Nation RL, Kong DC, Slavin M. Utility of bronchoalveolar lavage fluid galactomannan alone or in combination with PCR for the diagnosis of invasive aspergillosis in adult hematology patients: a systematic review and meta-analysis. *Crit Rev Microbiol* 2015;41:124-134.

- (15) Hoenigl M, Prattes J, Spiess B, Wagner J, Prueller F, Raggam RB, Posch V, Duettmann W, Hoenigl K, Wolfler A, Koidl C, Buzina W, Reinwald M, Thornton CR, Krause R, Buchheidt D. Performance of galactomannan, beta-d-glucan, Aspergillus lateral-flow device, conventional culture, and PCR tests with bronchoalveolar lavage fluid for diagnosis of invasive pulmonary aspergillosis. *J Clin Microbiol* 2014;52:2039-2045.
- (16) Dickson RP, Erb-Downward JR, Prescott HC, Martinez FJ, Curtis JL, Lama VN, Huffnagle GB. Analysis of culture-dependent versus culture-independent techniques for identification of bacteria in clinically obtained bronchoalveolar lavage fluid. *J Clin Microbiol* 2014;52:3605-3613.
- (17) Walsh TJ, Wissel MC, Grantham KJ, Petraitiene R, Petraitis V, Kasai M, Francesconi A, Cotton MP, Hughes JE, Greene L, Bacher JD, Manna P, Salomoni M, Kleiboeker SB, Reddy SK. Molecular detection and species-specific identification of medically important Aspergillus species by real-time PCR in experimental invasive pulmonary aspergillosis. *J Clin Microbiol* 2011;49:4150-4157.
- (18) Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15-21.
- (19) Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357-359.
- (20) Ruby JG, Bellare P, Derisi JL. PRICE: software for the targeted assembly of components of (Meta) genomic sequence data. *G3 (Bethesda)* 2013;3:865-880.
- (21) Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. *Bioinformatics* 2012;28:3150-3152.
- (22) Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. *Bioinformatics* 2010;26:873-881.

(23) Wilson MR, O'Donovan BD, Gelfand JM, Sample HA, Chow FC, Betjemann JP, Shah MP, Richie MB, Gorman MP, Hajj-Ali RA, Calabrese LH, Zorn KC, Greenlee JE, Blum JH, Green G, Khan LM, Banerji D, Langelier C, Bryson-Cahn C, Harrington W, Lingappa JR, Shanbhag NM, Green AJ, Brew BJ, Soldatos A, Strnad L, Doernberg SB, Jay CA, Douglas V, Josephson SA, DeRisi JL. Metagenomics for chronic meningitis: clarifying interpretation and diagnosis. *bioRxiv* 2017.

(24) Simpson E. **Measurement of Diversity**. *Nature* 1949;163:688.

(25) Leite GM, Magan N, Medina A. Comparison of different bead-beating RNA extraction strategies: an optimized method for filamentous fungi. *J Microbiol Methods* 2012;88:413-418.

(26) Gu W, Crawford ED, O'Donovan BD, Wilson MR, Chow ED, Retallack H, DeRisi JL. Depletion of Abundant Sequences by Hybridization (DASH): using Cas9 to remove unwanted high-abundance species in sequencing libraries and molecular counting applications. *Genome Biol* 2016;17:41-016-0904-5.

(27) Aagaard K, Petrosino J, Keitel W, Watson M, Katancik J, Garcia N, Patel S, Cutting M, Madden T, Hamilton H, Harris E, Gevers D, Simone G, McInnes P, Versalovic J. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. *FASEB J* 2013;27:1012-1022.

(28) Wattier RL, Dvorak CC, Hoffman JA, Brozovich AA, Bin-Hussain I, Groll AH, Castagnola E, Knapp KM, Zaoutis TE, Gustafsson B, Sung L, Berman D, Halasa NB, Abzug MJ, Velegraki A, Sharma TS, Fisher BT, Steinbach WJ. A Prospective, International Cohort Study of Invasive Mold Infections in Children. *Journal of the Pediatric Infectious Diseases Society* 2014.

(29) Mennink-Kersten MA, Ruegebrink D, Wasei N, Melchers WJ, Verweij PE. In vitro release by *Aspergillus fumigatus* of galactofuranose antigens, 1,3-beta-D-glucan, and DNA, surrogate markers used for diagnosis of invasive aspergillosis. *J Clin Microbiol* 2006;44:1711-1718.

- (30) Khan ZU, Ahmad S, Theyyathel AM. Detection of *Aspergillus fumigatus*-specific DNA, (1-3)-beta-D-glucan and galactomannan in serum and bronchoalveolar lavage specimens of experimentally infected rats. *Mycoses* 2008;51:129-135.
- (31) Mennink-Kersten MA, Ruegebrink D, Verweij PE, Steinbach WJ. Calcineurin-dependent galactomannan release in *Aspergillus fumigatus*. *Eur J Clin Microbiol Infect Dis* 2011;30:551-553.
- (32) Morton CO, Loeffler J, De Luca A, Frost S, Kenny C, Duval S, Romani L, Rogers TR. Dynamics of extracellular release of *Aspergillus fumigatus* DNA and galactomannan during growth in blood and serum. *J Med Microbiol* 2010;59:408-413.
- (33) Carroll CS, Amankwa LN, Pinto LJ, Fuller JD, Moore MM. Detection of a Serum Siderophore by LC-MS/MS as a Potential Biomarker of Invasive Aspergillosis. *PLoS One* 2016;11:e0151260.
- (34) Brasier AR, Zhao Y, Spratt HM, Wiktorowicz JE, Ju H, Wheat LJ, Baden L, Stafford S, Wu Z, Issa N, Caliendo AM, Denning DW, Soman K, Clancy CJ, Nguyen MH, Sugrue MW, Alexander BD, Wingard JR. Improved Detection of Invasive Pulmonary Aspergillosis Arising during Leukemia Treatment Using a Panel of Host Response Proteins and Fungal Antigens. *PLoS One* 2015;10:e0143165.
- (35) Lucena CM, Rovira M, Gabarrus A, Filella X, Martinez C, Domingo R, Torres A, Agusti C. The clinical value of biomarkers in respiratory complications in hematopoietic SCT. *Bone Marrow Transplant* 2017;52:415-422.
- (36) Bidula S, Sexton DW, Yates M, Abdolrasouli A, Shah A, Wallis R, Reed A, Armstrong-James D, Schelenz S. H-ficolin binds *Aspergillus fumigatus* leading to activation of the lectin complement pathway and modulation of lung epithelial immune responses. *Immunology* 2015;146:281-291.

(37) Davila S, Halstead ES, Hall MW, Doctor A, Telford R, Holubkov R, Carcillo JA, Storch GA, Eunice Kennedy Shriver Collaborative Pediatric Critical Care Research Network Investigators. Viral DNAemia and Immune Suppression in Pediatric Sepsis. *Pediatr Crit Care Med* 2018;19:e14-e22.

(38) Verhoeven DH, Claas EC, Jol-van der Zijde CM, Thijssen JC, Lankester AC, Bredius RG, Putter H, Kroes AC, Egeler RM, Schilham MW, van Tol MJ. Reactivation of Human Herpes Virus-6 After Pediatric Stem Cell Transplantation: Risk Factors, Onset, Clinical Symptoms and Association With Severity of Acute Graft-Versus-Host Disease. *Pediatr Infect Dis J* 2015;34:1118-1127.

(39) Albert E, Solano C, Pascual T, Torres I, Macera L, Focosi D, Maggi F, Gimenez E, Amat P, Navarro D. Dynamics of Torque Teno virus plasma DNAemia in allogeneic stem cell transplant recipients. *J Clin Virol* 2017;94:22-28.

(40) Wohlfarth P, Leiner M, Schoergenhofer C, Hopfinger G, Goerzer I, Puchhammer-Stoeckl E, Rabitsch W. Torquetenovirus Dynamics and Immune Marker Properties in Patients Following Allogeneic Hematopoietic Stem Cell Transplantation: A Prospective Longitudinal Study. *Biol Blood Marrow Transplant* 2017.

(41) Wootton SC, Kim DS, Kondoh Y, Chen E, Lee JS, Song JW, Huh JW, Taniguchi H, Chiu C, Boushey H, Lancaster LH, Wolters PJ, DeRisi J, Ganem D, Collard HR. Viral infection in acute exacerbation of idiopathic pulmonary fibrosis. *Am J Respir Crit Care Med* 2011;183:1698-1702.

(42) Hayden RT, Gu Z, Rodriguez A, Tanioka L, Ying C, Morgenstern M, Bankowski MJ. Comparison of two broadly multiplexed PCR systems for viral detection in clinical respiratory tract specimens from immunocompromised children. *J Clin Virol* 2012;53:308-313.

(43) Salez N, Vabret A, Leruez-Ville M, Andreoletti L, Carrat F, Renois F, de Lamballerie X. Evaluation of Four Commercial Multiplex Molecular Tests for the Diagnosis of Acute Respiratory Infections. *PLoS One* 2015;10:e0130378.

(44) Papadopoulou A, Gerdemann U, Katari UL, Tzannou I, Liu H, Martinez C, Leung K, Carrum G, Gee AP, Vera JF, Krance RA, Brenner MK, Rooney CM, Heslop HE, Leen AM. Activity of broad-spectrum T cells as treatment for AdV, EBV, CMV, BKV, and HHV6 infections after HSCT. *Sci Transl Med* 2014;6:242ra83.

(45) Bollard CM, Heslop HE. T cells for viral infections after allogeneic hematopoietic stem cell transplant. *Blood* 2016;127:3331-3340.

(46) Papadopoulou A, Kaloyannidis P, Yannaki E, Cruz CR. Adoptive transfer of Aspergillus-specific T cells as a novel anti-fungal therapy for hematopoietic stem cell transplant recipients: Progress and challenges. *Crit Rev Oncol Hematol* 2016;98:62-72.

(47) Gower WA, Collaco JM, Mogayzel PJ, Jr. Pulmonary dysfunction in pediatric hematopoietic stem cell transplant patients: non-infectious and long-term complications. *Pediatr Blood Cancer* 2007;49:225-233.

(48) Yanik GA, Grupp SA, Pulsipher MA, Levine JE, Schultz KR, Wall DA, Langholz B, Dvorak CC, Alangaden K, Goyal RK, White ES, Collura JM, Skeens MA, Eid S, Pierce EM, Cooke KR. TNF receptor inhibitor therapy for the treatment of children with idiopathic pneumonia syndrome (IPS). A joint Pediatric Blood and Marrow Transplant Consortium (PBMTTC) and Children's Oncology Group (COG) study (ASCT0521). *Biol Blood Marrow Transplant* 2014.

TABLE 1) Characteristics of Enrolled Patients

Demographics (n=34 patients)	
Age (median years, IQR)	11.2 (IQR 4.3-16.2)
Sex	
Female	16 (47%)
Male	18 (53%)
Race	
White	26 (76%)
Black	1 (3%)
Asian	1 (3%)
Hawaiian/Pacific Islander	1 (3%)
Other	1 (3%)
Unknown	4 (12%)
Ethnicity	
Hispanic/Latino	9 (26%)
Not Hispanic/Latino	24 (74%)
Primary Medical Condition	
Allogeneic HCT	20 (59%)
Autologous HCT	3 (9%)
Acute Leukemia (without HCT)	2 (6%)
Primary Immunodeficiency (without HCT)	4 (12%)
Severe Aplastic Anemia (without HCT)	2 (6%)
Solid Tumor (without HCT)	1 (3%)
Solid Organ Transplantation	2 (6%)
Clinical Course (n=41 episodes)	
Lower Respiratory Sample type	
BAL	33 (80%)
Mini-BAL	4 (10%)
ETT Aspirate	4 (10%)
Therapies (median number of therapies, IQR)	
Antibacterials	4 (1-5)
Antivirals	1 (0-2)
Antifungals	1 (0-2)
Immunomodulation	2 (0-2)
Patients with Identified Pathogen	
Any pathogen	13 (32%)
Bacteria only	5 (12%)
Fungi only	1 (2%)
Viruses only	4 (10%)
Multiple pathogens	3 (7%)
Outcomes	
Required PICU admission	21 (51%)
Required >24hr mechanical ventilation	17 (41%)
Hospital death	10 (29%)*

Legend: Age at 1st specimen collection. Indications for allogeneic HCT were acute leukemia (12/20), primary immunodeficiency (3/17), severe aplastic anemia (2/17), myeloproliferative/myelodysplastic disorder (2/17), and osteopetrosis (1/17). *Hospital death n=10/34 (29%).

FIGURE LEGENDS:

Figure 1: *Aspergillus* Lower Limit of Detection by Optimized Next Generation Sequencing

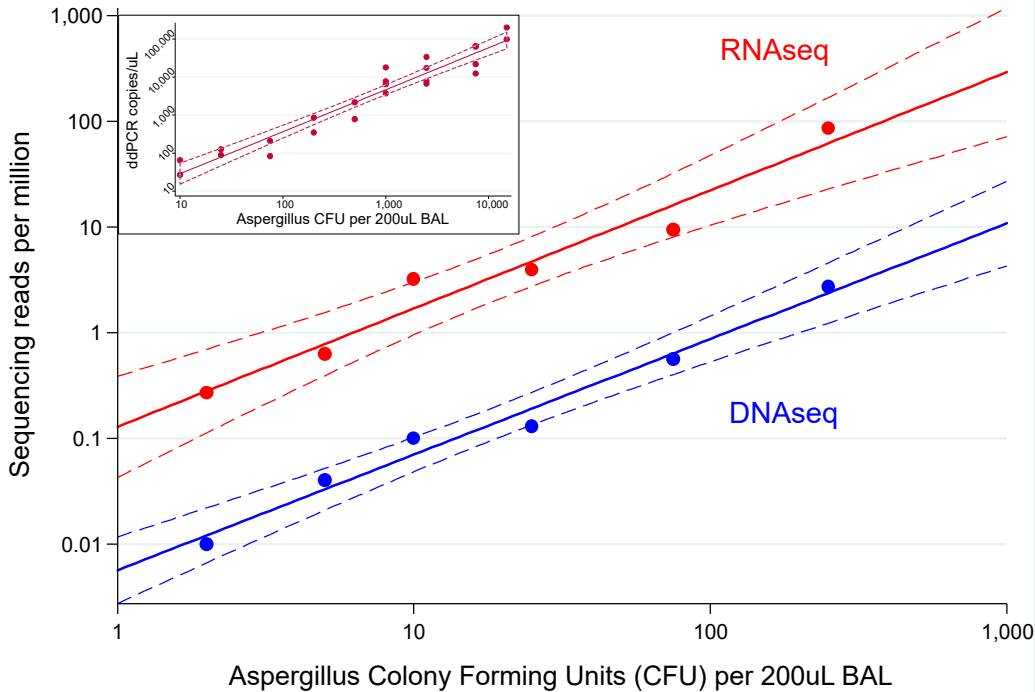
Legend: Main Figure: At the LLOD of RNAseq, the optimized assay was able to detect as few as 0.42 *Aspergillus* CFU (95% CI 0.12-1.40), whereas at the LLOD of DNAseq, the optimized assay was able to detect as few as 6.13 *Aspergillus* CFU (95% CI 4.16-9.04, paired T-test $p < 0.001$). Red data represent RNAseq and blue data represents DNAseq. Dotted lines represent the 95% confidence intervals for each linear regression. **Upper Left Insert:** Parallel detection of *Aspergillus* RNA using ddPCR.

Figure 2: Microbial Alignments Detected in the Lungs of Immunocompromised Children

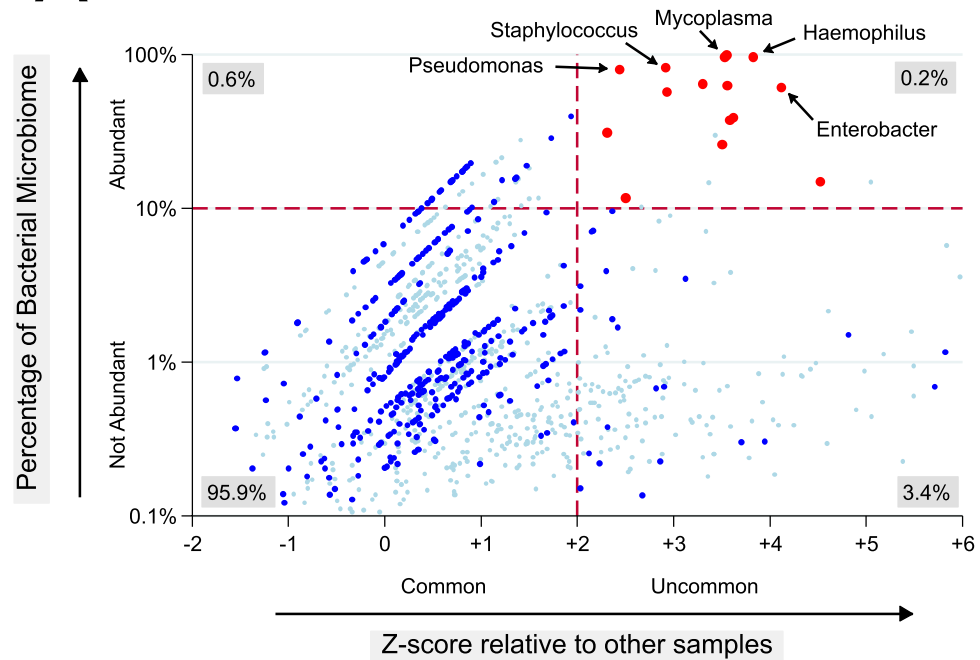
Legend: Red dots represent potentially pathogenic microbes that are both abundant ($\geq 10\%$ of all bacterial reads for bacteria, or ≥ 1 rpm for fungi or viruses) and identified at levels greater than most other samples in the cohort (Z -score ≥ 2). Hollow red dots indicating *Bocavirus* and *Pneumocystis* are used to indicate organisms observed only once in this cohort. Blue dots represent all other potentially pathogenic microbes; light blue dots represent typically non-pathogenic microbes. Subplots show all bacteria (**A**), fungi (**B**), RNA viruses (**C**), and DNA viruses (**D**) identified across all samples in the cohort. For the purpose of the Z -score calculation, the value of \log_{10} -transformed reads for undetected microbes was assumed to equal -2, just below the lower limit of detection for our sequencing depth ($\log_{10}[0.01\text{rpm}] = -2$).

Figure 3: Comparison of Clinical Laboratory Results vs. mNGS Results

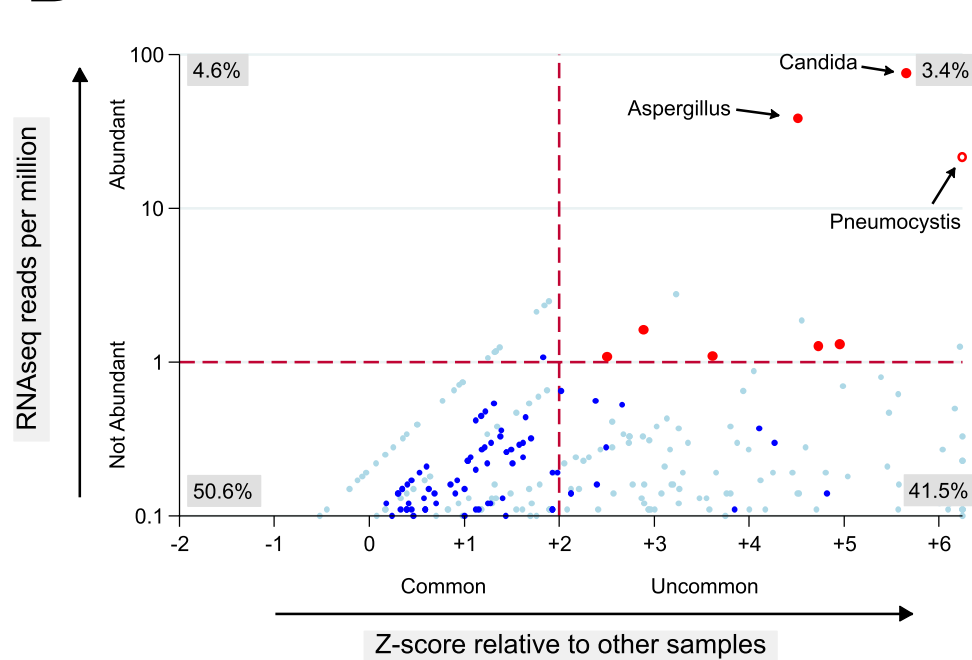
Legend: Clinical laboratory results were determined by review of medical charts. Pathogens reported by clinical laboratories but determined by the treating clinical team to *not be* causative of the patient's underlying respiratory disease were labeled as unlikely pathogens.



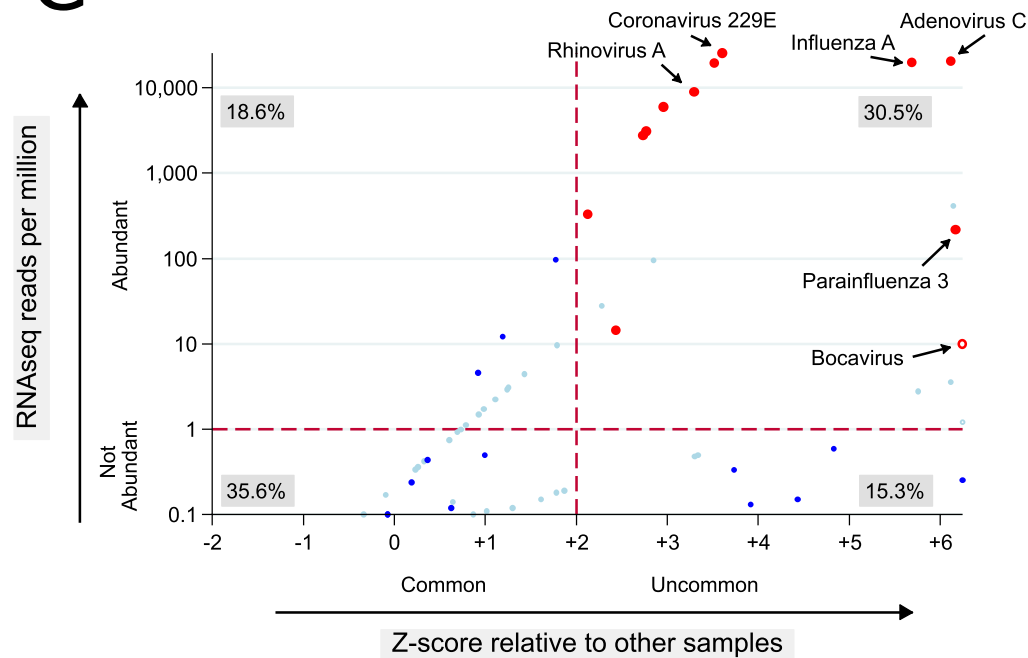
A Bacterial RNA in the Lungs of Immunocompromised Children



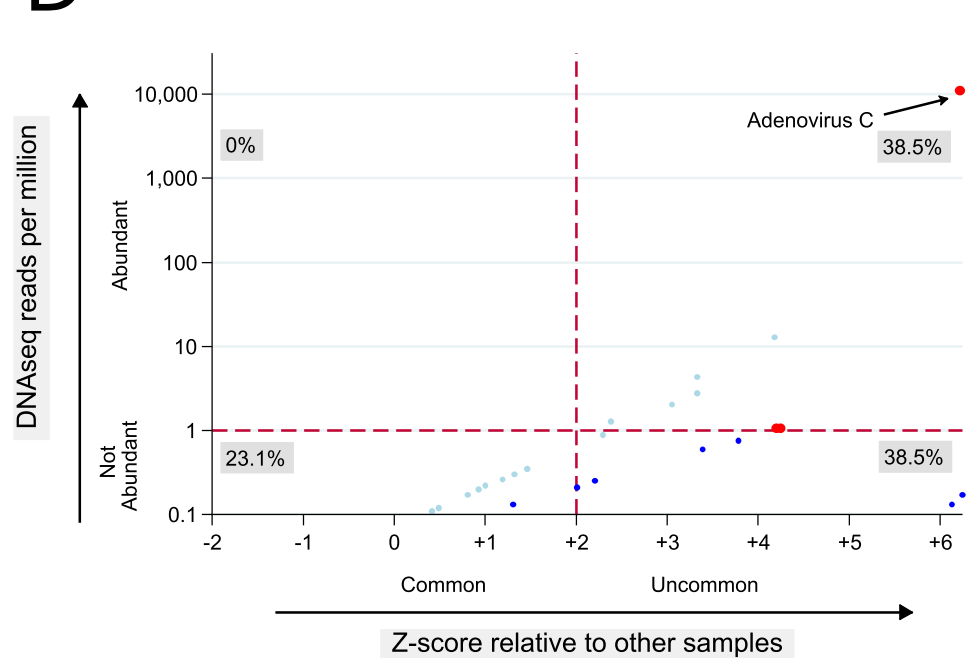
B Fungal RNA in the Lungs of Immunocompromised Children



C Viral RNA in the Lungs of Immunocompromised Children



D Viral DNA in the Lungs of Immunocompromised Children



Clinical Laboratory Results

