Pulmonary Metagenomic Sequencing Suggests Missed Infections in Immunocompromised Children

Matt S. Zinter,[1] Christopher C. Dvorak,[2] Madeline Y. Mayday,[1] Kensho Iwanaga,[3] Ngoc P. Ly,[3] Meghan E. McGarry,[3] Gwynne D. Church,[3] Lauren E. Faricy,[4] Courtney M. Rowan,[5] Janet R. Hume,[6] Marie E. Steiner,[6,7] Emily D. Crawford,[8,9] Charles Langelier,[10] Katrina Kalantar,[9] Eric D. Chow,[9] Steve Miller,[11] Kristen Shimano,[2] Alexis Melton,[2] Gregory A. Yanik,[12] Anil Sapru,[1,13] and Joseph L. DeRisi[8,9]

Divisions of [1]Critical Care, [2]Allergy, Immunology, and Blood & Marrow Transplantation, and [3]Pulmonology, Department of Pediatrics, Benioff Children's Hospital, University of California, San Francisco School of Medicine; [4]Division of Pulmonology, Department of Pediatrics, University of Vermont School of Medicine, Burlington; [5]Division of Critical Care, Department of Pediatrics, Riley Hospital for Children, Indiana University School of Medicine, Indianapolis; Divisions of [6]Critical Care and [7]Hematology/Oncology, Department of Pediatrics, Masonic Children's Hospital, University of Minnesota School of Medicine, Minneapolis; [8]Chan Zuckerberg Biohub, and [9]Department of Biochemistry & Biophysics, [10]Division of Infectious Diseases, Department of Internal Medicine, and [11]Department of Laboratory Medicine, University of California–San Francisco School of Medicine; [12]Division of Oncology, Department of Pediatrics, Motts Children's Hospital, University of Michigan School of Medicine, Ann Arbor; and [13]Division of Critical Care, Department of Pediatrics, Mattel Children's Hospital, University of California–Los Angeles, Geffen School of Medicine

Background. Despite improved diagnostics, pulmonary pathogens in immunocompromised children frequently evade detection, leading to significant mortality. Therefore, we aimed 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, parallel RNA/DNA extraction, and metagenomic sequencing. Sequencing reads were aligned to the National Center for Biotechnology Information 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.

Results. We identified a rich cross-domain pulmonary microbiome that contained 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, 0.61; interquartile range [IQR], 0.33–0.72 vs median, 0.96; IQR, 0.94–0.96; $P < .001$). Potential pathogens were detected in half of samples previously negative by clinical diagnostics, demonstrating increased sensitivity for missed pulmonary pathogens ($P < .001$).

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.

Keywords. intensive care units, pediatric; immunocompromised host; metagenomics; respiratory tract infections; microbiota.

Last year in the United States, approximately 15 000 children were diagnosed with cancer, 2000 underwent solid organ transplantation, and 2500 underwent hematopoietic cell transplantation (HCT) for an increasingly broad set of life-threatening diseases [1–3]. Despite improved safety of antineoplastic and transplantation-based therapies, the risk of infectious complications such as pneumonia remains high [4, 5]. Due to the inhibitory effect of antimicrobial pretreatment on culture growth, impaired serologic immunity, and the limited preselected targets of multiplex assays, current microbiologic diagnostics frequently fail to identify pathogenic organisms [6]. The significant mortality associated with undiagnosed pulmonary infections is evident in postmortem case series of pediatric HCT patients, in whom previously undetected pulmonary pathogens have been identified in 30%–50% [7].

Unlike assays that target the 16S and/or 28S/ITS ribosomal RNA amplicons, unbiased metagenomic next-generation sequencing (mNGS) can detect bacteria, viruses, and fungi and has shown promising results for diagnosing neurologic and ocular infections [8–10]. However, the identification of filamentous molds such as *Aspergillus* spp. remains difficult due to thick extracellular matrices and the relatively small inoculum required to induce disease [11–14]. 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 [15].

Therefore, we conducted a pilot study aimed to develop a highly sensitive mNGS assay capable of detecting myriad pulmonary bacterial, fungal, and viral pathogens and to test that assay on a retrospective cohort of immunocompromised children who underwent 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 identify potential pulmonary pathogens in this high-risk population.

METHODS

Development of Optimized mNGS Assay
We created mock-positive bronchoalveolar lavages (BALs) by spiking *Aspergillus niger* broth into aliquots of BAL containing known *Haemophilus influenzae/Human Adenovirus B* coinfection (Supplementary Material 1). An *Aspergillus* species was chosen as an optimization benchmark given its thick polysaccharide cell wall and extreme clinical importance of invasive mold in this patient population [16].

Sample Preparation
We tested nucleic acid extraction conditions by combining 200 μL of mock-positive BAL with either 600 μL DirectZol, 600 μL lysis buffer, or 200 μL DNA/RNA Shield (Zymo), followed by mechanical homogenization with either 0.1 mm or 0.5 mm glass bashing beads (Omni) for 2, 5, or 8 cycles of 25 seconds, bashing at 30 Hz with 60 seconds rest on ice between each cycle (TissuerLyser II, Qiagen). Samples homogenized in DNA/RNA shield also underwent enzymatic mycolysis with either 0.15 mg or 0.38 mg proteinase K at 23°C for 30 or 60 minutes (Zymo) or with 0.4 mg, 1.2 mg, 4 mg, or 8 mg 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 parallel DNA/RNA extraction (Zymo ZR-Duet DNA/RNA MiniPrep Kit). *Aspergillus niger* nucleic acid yield was measured using an orthogonal digital droplet polymerase chain reaction (PCR) assay with pan-*Aspergillus* primers (Supplementary Material 2) [17]. RNA and DNA sequencing libraries were prepared in parallel (New England Biolabs NEBNext Ultra-II Library Prep) and underwent 125 nucleotide paired-end sequencing on an Illumina HiSeq 4000 instrument (Supplementary Material 3).

Bioinformatics Pipeline
Resultant.fastq files were processed using a previously described pipeline that consisted of several open-source components [8, 9]. Briefly, reads underwent iterative removal of host (Hg38/PanTro), low-quality, low-complexity (Lempel–Ziv–Welch [LZW] compression ratio >0.45), and redundant sequences using STAR, Bowtie2,

PriceSeqFilter, and CD-HIT-DUP [18–21]. The remaining sequences were aligned to the National Center for Biotechnology Information (NCBI) nonredundant nucleotide database using GSNAPL for assignment of taxonomic IDs [22]. Microbes were described as potentially pathogenic or typically nonpathogenic based on a priori literature review (Supplementary Material 4).

Controls
To assess microbial contaminants in our reagents and laboratory environment, we sequenced 8 control samples containing spiked-in HeLa RNA (Supplementary Material 5).

Analysis
The prevalence of each microbe in each respiratory sample was described using 2 criteria: abundance relative to other microbes in the same sample, wherein we normalized sequencing reads per million total sequencing reads (rpm), and (2) abundance 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). Given the anticipated wide array of respiratory bacteria and the paucity of knowledge regarding the significance of low-level viruses and fungi, we aimed to maximize specificity for bacterial pathogens and to maximize sensitivity for viral and fungal pathogens [15, 23, 24]. Therefore, we defined microbial outliers as those with Z-score ≥2 and ≥10 rpm (bacteria) or ≥1 rpm (viruses/fungi). The Simpson diversity index was used to associate the loss of bacterial diversity with the presence of outlier microbes (Supplementary Material 6) [25].

Validation
Respiratory samples with outlier pathogens that were not identified on clinical testing were sent to reference laboratories for orthogonal confirmatory testing (Supplementary Material 7).

Patients
To test the optimized mNGS assay, we prospectively screened and approached immunocompromised patients age ≤25 years undergoing clinically indicated lower respiratory sampling 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. Patients were enrolled with consent. After respiratory samples were collected for clinical purposes, excess volume was separated, placed on dry ice within 10 minutes of the original procedure, and banked at −70°C until processing (Supplementary Material 8). This study was approved by each site's institutional review board.

RESULTS

Development of Optimized mNGS Assay
Iterative optimizations demonstrated that mechanical homogenization of BAL using 0.5-mm glass bashing beads for 5 cycles

in DNA/RNA shield without mycolytic enzymes yielded the highest quantity of *A. niger* nucleic acid (Supplementary Material 9, Supplementary Figure 1a–1d). This extraction protocol performed similarly when applied to clinical isolates of *A. fumigatus* (Supplementary Figure 1e). When sequencing RNA to a depth of 25 million reads per 200 μL BAL, this protocol improved the lower limit of detection (LLOD) of *A. niger* by approximately 100-fold (59.60 colony-forming units [CFU] preoptimization; 95% confidence interval [CI], 37.70–95.36 vs 0.42 CFU postoptimization; 95% CI, 0.12–1.40; paired *T* test *P* < .001; Supplementary Figure 2). However, sequencing simultaneously extracted DNA to the same depth yielded a 10-fold inferior *A. niger* LLOD (6.13 CFU; 95% CI, 4.16–9.04; paired *T* test *P* < .001; Figure 1). The optimization did not change the detection of *H. influenzae* or *Human Adenovirus B* (*T* test *P* = .343 and *P* = .420, respectively).

Application of mNGS Assay

Using the optimized protocol described above, we conducted mNGS on 41 clinical samples obtained from 34 patients (Table 1). Sequencing results are summarized below, with sequencing quality reported in Supplementary Tables 1 and 2, raw sequencing files available in NCBI dbGaP (https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs001684.v1.p1), and detailed patient results in Appendix 1. Negative controls demonstrated minimal environmental contamination (<1 rpm of pathogenic bacteria and no pathogenic fungi or viruses Supplementary Table 3).

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, 81.7% of bacterial genera were identified at <10 rpm and within 2 standard deviations of the cohort mean (Z-score < 2; Figure 2A). These included potentially pathogenic bacterial genera such as *Escherichia*, *Klebsiella*, *Pseudomonas*, *Staphylococcus*, *Stenotrophomonas*, and *Streptococcus*, which were identified in nearly all samples at levels well above those detected in the negative controls (38, 39, 40, 38, 41, and 40 of 41 samples, respectively). Conversely, only 0.4% of detected bacterial genera met outlier criteria. These outliers included 10 potentially pathogenic bacteria identified in 13/41 patient samples (*Corynebacterium, Enterobacter, Enterococcus, Escherichia, Haemophilus, Klebsiella, Mycoplasma, Pseudomonas, Staphylococcus,* and *Streptococcus*). Samples with outlier bacterial pathogens had significantly depressed alpha-diversity of the bacterial microbiome (median, 0.61; interquartile range [IQR], 0.33–0.72 vs 0.96; IQR, 0.94–0.96; *P* < .001; Figure 3).

Fungi

Relative to bacterial alignments, fungal alignments were significantly less prevalent in the cohort. Further, 92.1% of fungal genera were quantified below 1 rpm (Figure 2B). Evidence supporting these alignments is limited by the rarity of the reads themselves. Only 3.4% of fungal genera met outlier criteria; these included potentially pathogenic fungi identified in 7/41 patient samples (*Alternaria, Aspergillus, Candida,*
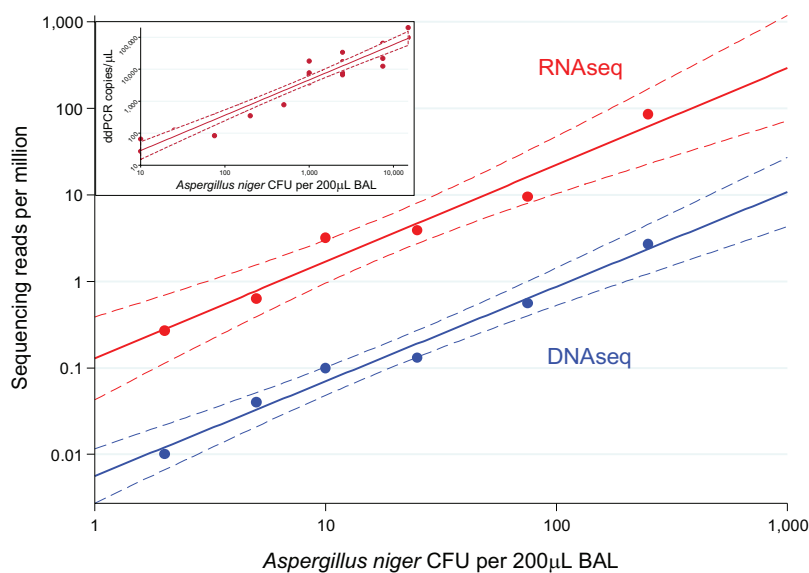

Figure 1. *Aspergillus niger* lower limit of detection by optimized next-generation sequencing. At the lower limit of detection (LLOD) of RNA sequencing (RNAseq), the optimized assay was able to detect as few as 0.42 *Aspergillus niger* colony-forming units (CFU; 95% confidence interval [CI], 0.12–1.40), whereas at the LLOD of DNAseq, the optimized assay was able to detect as few as 6.13 *A. niger* CFU (95% CI, 4.16–9.04; paired *T* test *P* < .001). Red data represent RNAseq and blue data represent DNAseq. Dotted lines represent the 95% CIs for each linear regression. *(Insert)* Parallel detection of *A. niger* RNA using digital droplet polymerase chain reaction assay. As nucleic acid bioavailability may vary across *Aspergillus* species, these results may not be directly extrapolated to other *Aspergillus* species and other medically relevant molds. Abbreviations: BAL, bronchoalveolar lavage; CFU, colony-forming unit; ddPCR, droplet digital polymerase chain reaction; seq, sequencing.

Table 1.   Characteristics of Enrolled Patients

Demographics (n = 34 patients)	Descriptor
Age (median years, IQR)[a]	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[b]	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 pediatric intensive care unit admission	21 (51%)
Required >24 hours mechanical ventilation	17 (41%)
Hospital death[c]	10 (29%)

Abbreviations: BAL, bronchoalveolar lavage; ETT, endotracheal tube; HCT, hematopoietic cell transplantation; IQR, interquartile range.

[a]Age at first specimen collection.

[b]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).

[c]Hospital death n = 10/34 (29%).

Cladosporium, *Cryptococcus*, *Fusarium*, and *Pneumocystis*). While *Aspergillus* RNA levels were detectable in 14/41 patient samples, only 1 was positive for invasive pulmonary aspergillosis (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. *Aspergillus* RNA levels did not correlate with growth in culture (*T* test *P* = .148) but demonstrated weak association with BAL galactomannan (*P* < .001; Supplementary Figure 3).

Viruses

In contrast with bacteria and fungi, a significantly larger portion of RNA alignments to viral genera were present at high abundance, and 30.5% met outlier criteria (Figure 2C and 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* met outlier criteria, an additional 3 *Rhinoviruses* were abundant but had Z-scores between +0.92 and +1.78. Additionally, 2 patients had viral coinfections (*Parainfluenza-3* and *Influenza-C*; *Adenovirus-C* and *Rhinovirus-A*). 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 samples, respectively).

Comparison to Clinical Testing

Clinical testing identified causative pathogens in 41.4% of samples (n = 17; Figure 4). Of these, 11 were concordantly identified as outliers by mNGS, and 3 of these 11 contained outlier quantities of RNA aligning to a second previously undetected potential copathogen (*Bocavirus*, *Corynebacterium*, and *Influenza-C*). An additional 3 were identified by mNGS but were not classified as outliers (2 cases of *Aspergillus* diagnosed by galactomannan and 1 case of *Rhinovirus-A*), and another 3 had a different outlier pathogen identified by mNGS (*Coronavirus 229E* twice and *Coronavirus OC43* once). Clinical testing did not identify any pathogens in 58.5% of samples (n = 24). Here, mNGS was able to identify statistically outlying potential pathogens in 11/24 cases, including a variety of bacteria (ie, *P. aeruginosa*, *E. cloacae*, *M. pneumoniae*), fungi (ie, *C. glabrata*), and viruses (ie, *Rhinovirus-A)*.

Orthogonal Validation

Statistical approaches to separate commensals from pathogens are inherently imperfect; therefore, we undertook orthogonal validation as an independent means to verify these results. Organisms detected by mNGS but not clinical testing were validated with commercially available Clinical Laboratory Improvement Amendments (CLIA)-approved assays performed
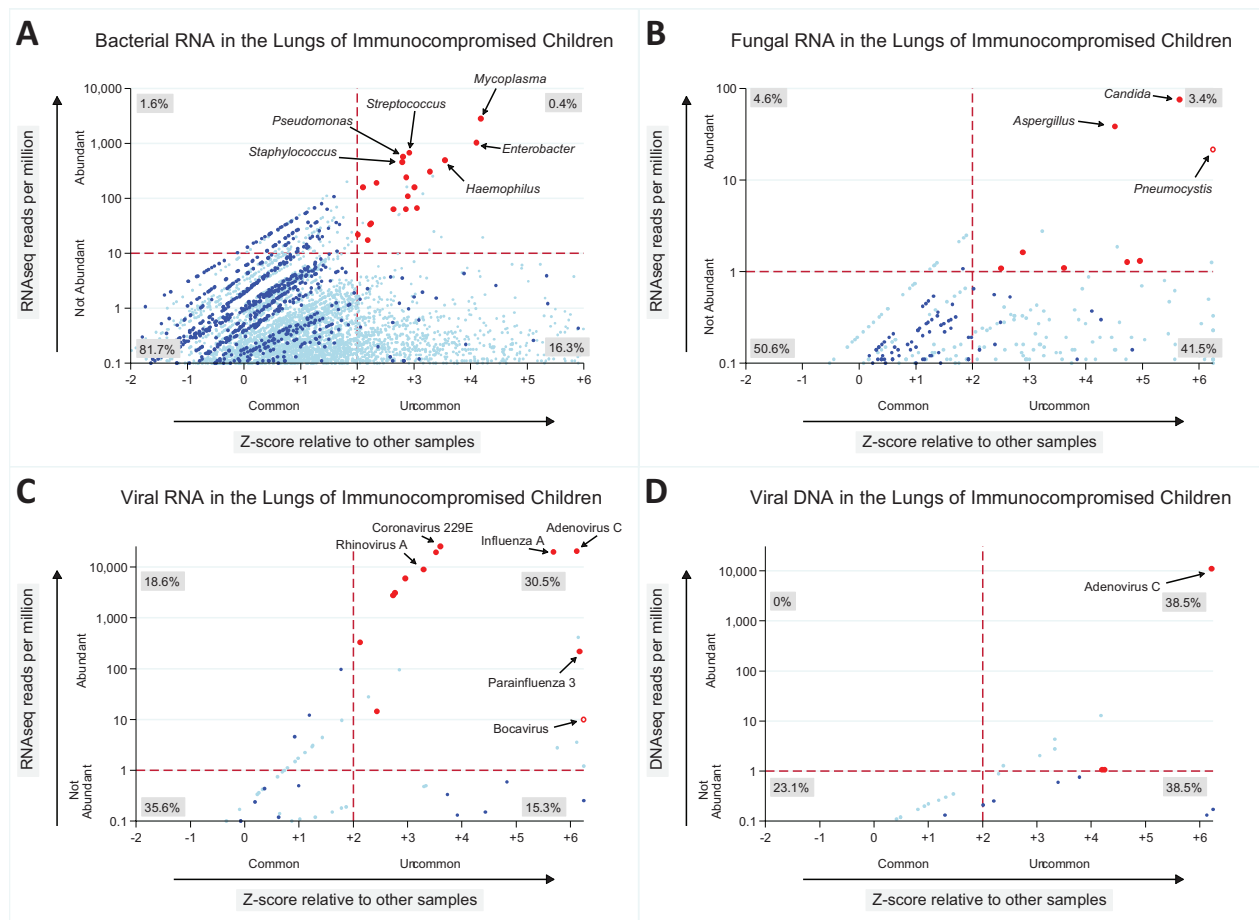
Figure 2. Microbial alignments detected in the lungs of immunocompromised children. Red dots represent potentially pathogenic microbes that are both abundant (≥10 rpm for bacteria or ≥1 rpm for fungi or viruses) and identified at levels greater than most other samples in the cohort (Z-score ≥2). Hollow 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 nonpathogenic microbes. Subplots show *(A)* all bacteria, *(B)* fungi, *(C)* RNA viruses, and *(D)* DNA viruses 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 ($\log_{10}[0.01\text{rpm}] = -2$). Abbreviation: seq, sequencing.

on aliquots of the original unprocessed sample (Supplementary Material 7). All validation tests were concordant with mNGS with 1 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 reverse-transcription primer sets followed by Sanger sequencing. These data demonstrate that mNGS has significantly greater sensitivity for detecting potential pulmonary pathogens than current clinical diagnostics (McNemar's *P* < .001).

DISCUSSION

In this study, we developed and optimized an 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 population. Further, by comparing the quantity of

microbial nucleic acid to that of other microbes within a sample and to other samples within the cohort, we were able to identify outlying potential pathogens in approximately half of clinically negative samples.

Due to inherent challenges in sampling the lower respiratory tract, the pulmonary microbiome was not one of the original sites studied in the 2008 Human Microbiome Project, and its exploration has lagged decades behind similar analyses of human intestinal, cutaneous, and nasopharyngeal microbiomes [26]. In addition, studies of the pulmonary microbiome in children necessarily lack healthy matched controls due to the inherent risks of anesthesia and bronchoscopy [27]. In this study, we found that many potentially pathogenic bacteria such as *Pseudomonas* and *Streptococcus* are ubiquitous, and hence their abundance needs to be contextualized by cohort-specific norms. For example, 100% of samples had detectable *Pseudomonas* RNA, but only sample 29 had *Pseudomonas* RNA detected more than 2 standard deviations above the cohort mean. Normalizing
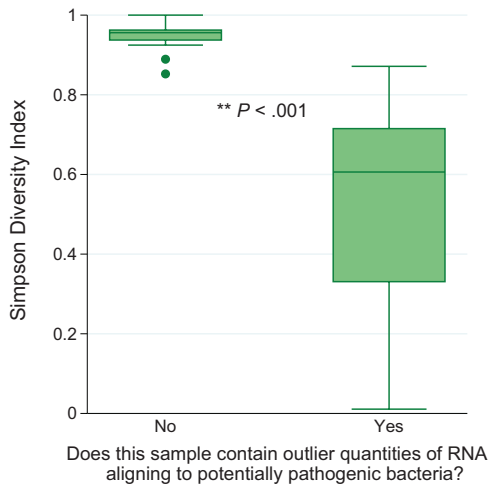
Figure 3. Respiratory samples with outlier pathogens have depressed bacterial alpha-diversity. Diversity of the bacterial microbiome was significantly decreased in samples with potentially pathogenic bacteria present at ≥10 rpm of the pulmonary bacterial microbiome and Z-score ≥2 (median, 0.61; interquartile range [IQR], 0.33–0.72; n = 13 vs median, 0.96; IQR, 0.94–0.96; n = 28; *P* < .001). Simpson diversity index cutoffs of ≥0.8 or ≥0.9 showed 90.3% (95% confidence interval, 77.6–96.2) and 100% 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.

population-dependent measurements is common in pediatrics and appears well suited to describe aspects of the pulmonary microbiome [28–30].

To date, the majority of metagenomic sequencing assays have targeted amplicons within the 16S or 28S rRNA subunits, allowing detection of pulmonary bacteria or fungi, but not both, and not viruses [31]. Recently, unbiased mNGS assays have allowed

detection of both bacterial and viral nucleic acid but have lacked ideal sensitivity for detecting filamentous mold [11, 12]. Although the European Aspergillus PCR Initiative has described ideal methodology for extracting fungal nucleic acid from blood, the optimal extraction conditions for respiratory specimens remain less well defined [32, 33]. 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 [34–36]. As the majority of commercial sequencing assays measure DNA, this study adds to the literature by demonstrating that RNA sequencing is >10 times more sensitive for the detection of such fastidious organisms, which we speculate may be due to high copy numbers of particular RNA templates present in active organisms [37].

Using optimized mechanical homogenization, *Aspergillus* RNA was detected in 34.1% of samples and no negative controls, suggesting that the lungs of immunocompromised children are frequently exposed to low levels of potentially viable *Aspergillus* organisms. These data are novel in the pediatric population and are congruous with surveillance data from neutropenic and nonneutropenic adults [38, 39]. However, as only 10% of samples originated from patients with suspected IPA, patient-specific factors such as antifungal pretreatment, immune reconstitution, alloreactive inflammation, and impaired mucociliary clearance remain crucial in determining which child might develop IPA [16]. While this study was not powered to assess performance characteristics of this assay for IPA, pan-fungal and *Aspergillus*-specific PCR have demonstrated 76%–79% sensitivity and 93%–95% specificity for probable/proven IPA [14, 40–47]. Combining nucleic acid tests with galactomannan can hasten diagnosis and improve
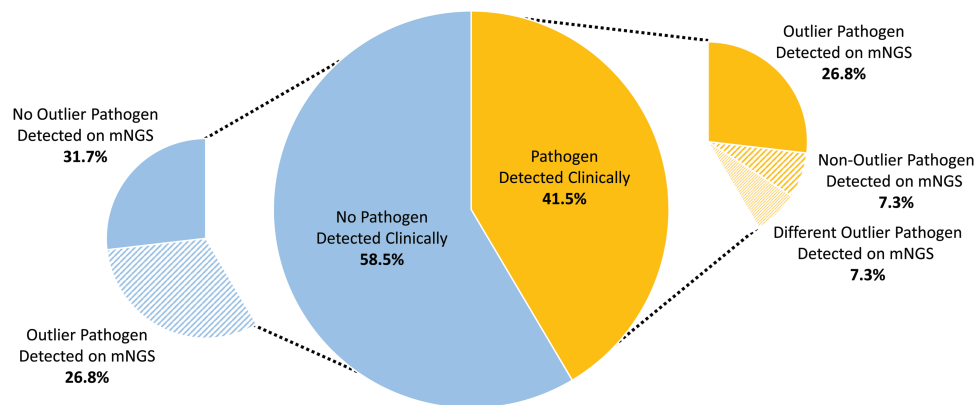

Figure 4. Comparison of clinical laboratory results vs metagenomic next-generation sequencing (mNGS) results. Clinical laboratory results were determined by review of medical charts. n = 17 patients had samples with a pathogen detected clinically, as determined by interpretation of clinical microbiologic testing by the treating physician. Of these, n = 11 had concordant pathogens of outlier quantities on mNGS (Adenovirus/Rhinovirus, Aspergillus fumigatus, Enterobacter cloacae, Escherichia coli, Haemophilus influenzae, Haemophilus influenzae/Parainfluenza virus, Mycoplasma pneumoniae [n = 2], Pneumocystis jirovecii/Rhinovirus-A, Rhinovirus-C, and Staphylococcus aureus); n = 3 had concordant pathogens identified on mNGS but not in outlier quantities (Aspergillus [n = 2] and Rhinovirus-A); and n = 3 had an alternative pathogen identified on mNGS (Human coronavirus *229E* [n = 2] and Human coronavirus *OC43*). n = 24 patients had samples without a pathogen detected clinically. Of these, n = 11 had a potential pathogen present in outlier quantities on mNGS (Candida glabrata, Cytomegalovirus, Cryptococcus [n = 2], Enterobacter cloacae, Human herpesvirus-6, Mycoplasma pneumoniae, Rhinovirus-A, Pseudomonas aeruginosa/Influenza-A, Staphylococcus epidermidis, and Streptococcus pneumoniae) and n = 13 did not. Abbreviation: mNGS, metagenomic next-generation sequencing.

clinical outcomes in some populations [48–51]; further, incorporation of immune function may also improve discrimination of colonization and invasive mycosis [52–56]. Importantly, for all nucleic acid–based tests, measurable *Aspergillus* yield varies among species and according to life stage, cell wall characteristics, and hydrophobicity [57]. Therefore, any future clinical mNGS assay will need to be benchmarked against numerous *Aspergillus* species as well as other medically relevant fungi in order to have the broadest utility.

Consistent with emerging data in pediatrics, we also detected herpesviruses such as *CMV* and *HHV-6* in low abundance among many children in the cohort [58, 59]. Interestingly, a number of typically nonpathogenic viruses such as *Torquetenovirus* and *KI Polyomavirus* were present in more than half of the cohort. While data supporting the direct pathogenicity of these viruses are lacking, they may indicate immune dysfunction and predict numerous post-transplant infectious and alloimmune complications [60–62]. 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 identified statistically outlying potential pathogens in half of clinically negative samples. Most newly detected cultivable organisms were isolated from patients who had been pretreated with antimicrobials, highlighting the importance of culture-independent techniques [15]. In addition, several newly detected viruses were not included in clinical multiplex PCR assays (*Human coronavirus*, *Human Bocavirus*, *Influenzavirus-C*) [63]. While this study suggests that many idiopathic pulmonary complications may be associated with infections, one quarter of samples were concordantly negative for pathogens on clinical and mNGS testing, emphasizing the ongoing significance of noninfectious pulmonary complications in this population. Future clinical validation of mNGS may demonstrate utility in safely excluding infections and allowing amplified immunomodulation in patients with predominantly inflammatory pulmonary syndromes [64].

Our study has several strengths. First, we optimized the extraction of *Aspergillus* spp. RNA while preserving detection of bacterial and viral nucleic acid. Second, we proposed a logical analytical framework that ranks organism abundance both within a sample and relative to other samples. Third, outlier pathogens identified by mNGS that were not detected clinically were subsequently validated by orthogonal assays. Fourth, we provide, to our knowledge, the first evaluation of the pulmonary microbiome in immunocompromised children.

Our study has several limitations. First, while the Z-score was useful in deemphasizing commonly abundant organisms (ie, *S. pneumoniae*), it may have overvalued uncommon organisms with less abundant transcripts (ie, *CMV*); additional larger studies will naturally strengthen the utility of Z-score analyses. Second, because the relationship between microbe quantity and

sequencing reads varies across organisms based on nucleic acid accessibility and the availability of annotated reference genomes, future clinical application of an mNGS assay will require validation on numerous clinically relevant species. Third, samples with abundant nucleic acid from human epithelial cells, leukocytes, viruses, and other sources may have reduced detection of sparse or fastidious microbes. Finally, 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. Future studies are needed to determine whether prospective use of mNGS in a clinically relevant time frame might affect patient management and improve outcomes. In order to optimize patient outcomes, we advocate for ongoing multidisciplinary collaboration among clinicians, laboratory scientists, and bioinformaticians.

CONCLUSIONS

In summary, we present an optimized mNGS assay that revealed a rich bacterial, fungal, and viral pulmonary microbiome in immunocompromised children and identified potential pathogens in half of clinically 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 [65].

Supplementary Data

Supplementary materials are available at *Clinical Infectious Diseases* online. Consisting of data provided by the authors to benefit the reader, the posted materials are not copyedited and are the sole responsibility of the authors, so questions or comments should be addressed to the corresponding author.

Notes

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. Available at: https://bloodcell.transplant.hrsa.gov/research/transplant_data/transplant_activity_report/index.html. Accessed 1 July 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**; 15:817–26.
5. 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**; 43:1986–94.
6. Aguilar-Guisado M, Jiménez-Jambrina M, Espigado I, et al.; Spanish Network for Research in Infectious Diseases. Pneumonia in allogeneic stem cell transplantation recipients: a multicenter prospective study. Clin Transplant **2011**; 25:E629–38.
7. Roychowdhury M, Pambuccian SE, Aslan DL, et al. Pulmonary complications after bone marrow transplantation: an autopsy study from a large transplantation center. Arch Pathol Lab Med **2005**; 129:366–71.
8. Wilson MR, Naccache SN, Samayoa E, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. N Engl J Med **2014**; 370:2408–17.
9. Doan T, Wilson MR, Crawford ED, et al. Illuminating uveitis: metagenomic deep sequencing identifies common and rare pathogens. Genome Med **2016**; 8:90.
10. Langelier C, Zinter MS, Kalantar K, et al. Metagenomic sequencing detects respiratory pathogens in hematopoietic cellular transplant patients. Am J Respir Crit Care Med **2018**; 197:524–8.
11. Bittinger K, Charlson ES, Loy E, et al. Improved characterization of medically relevant fungi in the human respiratory tract using next-generation sequencing. Genome Biol **2014**; 15:487.
12. Clarke EL, Lauder AP, Hofstaedter CE, et al. Microbial lineages in sarcoidosis. A metagenomic analysis tailored for low-microbial content samples. Am J Respir Crit Care Med **2018**; 197:225–34.
13. Heng SC, Morrissey O, Chen SC, et al. 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**; 41:124–34.
14. Hoenigl M, Prattes J, Spiess B, et al. 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**; 52:2039–45.
15. Dickson RP, Erb-Downward JR, Prescott HC, et al. Analysis of culture-dependent versus culture-independent techniques for identification of bacteria in clinically obtained bronchoalveolar lavage fluid. J Clin Microbiol **2014**; 52:3605–13.
16. Wattier RL, Dvorak CC, Hoffman JA, et al. A prospective, international cohort study of invasive mold infections in children. J Pediatric Infect Dis Soc **2015**; 4:313–22.
17. Walsh TJ, Wissel MC, Grantham KJ, et al. Molecular detection and species-specific identification of medically important *Aspergillus* species by real-time PCR in experimental invasive pulmonary aspergillosis. J Clin Microbiol **2011**; 49:4150–7.
18. Dobin A, Davis CA, Schlesinger F, et al. 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**; 9:357–9.
20. Ruby JG, Bellare P, Derisi JL. PRICE: software for the targeted assembly of components of (meta) genomic sequence data. G3 (Bethesda) **2013**; 3:865–80.
21. Fu L, Niu B, Zhu Z, Wu S, Li W. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics **2012**; 28:3150–2.
22. Wu TD, Nacu S. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics **2010**; 26:873–81.
23. Cui L, Lucht L, Tipton L, et al. Topographic diversity of the respiratory tract mycobiome and alteration in HIV and lung disease. Am J Respir Crit Care Med **2015**; 191:932–42.
24. Abbas AA, Diamond JM, Chehoud C, et al. The perioperative lung transplant virome: torque teno viruses are elevated in donor lungs and show divergent dynamics in primary graft dysfunction. Am J Transplant **2017**; 17:1313–24.
25. Simpson E. Measurement of diversity. Nature **1949**; 163:688.
26. Aagaard K, Petrosino J, Keitel W, et al. The Human Microbiome Project strategy for comprehensive sampling of the human microbiome and why it matters. FASEB J **2013**; 27:1012–22.
27. de Blic J, Midulla F, Barbato A, et al. Bronchoalveolar lavage in children. ERS task force on bronchoalveolar lavage in children. European Respiratory Society. Eur Respir J **2000**; 15:217–31.
28. Chubb H, Simpson JM. The use of Z-scores in paediatric cardiology. Ann Pediatr Cardiol **2012**; 5:179–84.
29. Buxbaum NP, Robinson C, Sinaii N, et al. Impaired bone mineral density in pediatric patients with chronic graft-versus-host disease. Biol Blood Marrow Transplant **2018**; 24:1415–23.
30. Inokuchi M, Matsuo N, Takayama JI, Hasegawa T. BMI Z-score is the optimal measure of annual adiposity change in elementary school children. Ann Hum Biol **2011**; 38:747–51.
31. Moffatt MF, Cookson WO. The lung microbiome in health and disease. Clin Med (Lond) **2017**; 17:525–9.
32. White PL, Barnes RA. Isolation of nucleic acids for fungal diagnosis. Methods Mol Biol **2017**; 1508:223–47.
33. White PL, Bretagne S, Klingspor L, et al.; European Aspergillus PCR Initiative. *Aspergillus* PCR: one step closer to standardization. J Clin Microbiol **2010**; 48:1231–40.
34. Leite GM, Magan N, Medina Á. Comparison of different bead-beating RNA extraction strategies: an optimized method for filamentous fungi. J Microbiol Methods **2012**; 88:413–8.
35. Fredricks DN, Smith C, Meier A. Comparison of six DNA extraction methods for recovery of fungal DNA as assessed by quantitative PCR. J Clin Microbiol **2005**; 43:5122–8.
36. Griffiths LJ, Anyim M, Doffman SR, Wilks M, Millar MR, Agrawal SG. Comparison of DNA extraction methods for *Aspergillus fumigatus* using real-time PCR. J Med Microbiol **2006**; 55:1187–91.
37. Gu W, Crawford ED, O'Donovan BD, et al. 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.
38. Barton RC, Hobson RP, McLoughlin H, Morris A, Datta B. Assessment of the significance of respiratory culture of *Aspergillus* in the non-neutropenic patient. A critique of published diagnostic criteria. Eur J Clin Microbiol Infect Dis **2013**; 32:923–8.
39. Escribano P, Marcos-Zambrano LJ, Peláez T, et al. Sputum and bronchial secretion samples are equally useful as bronchoalveolar lavage samples for the diagnosis of invasive pulmonary aspergillosis in selected patients. Med Mycol **2015**; 53:235–40.
40. Tuon FF. A systematic literature review on the diagnosis of invasive aspergillosis using polymerase chain reaction (PCR) from bronchoalveolar lavage clinical samples. Rev Iberoam Micol **2007**; 24:89–94.
41. Mengoli C, Cruciani M, Barnes RA, Loeffler J, Donnelly JP. Use of PCR for diagnosis of invasive aspergillosis: systematic review and meta-analysis. Lancet Infect Dis **2009**; 9:89–96.
42. Avni T, Levy I, Sprecher H, Yahav D, Leibovici L, Paul M. Diagnostic accuracy of PCR alone compared to galactomannan in bronchoalveolar lavage fluid for diagnosis of invasive pulmonary aspergillosis: a systematic review. J Clin Microbiol **2012**; 50:3652–8.
43. Warris A, Lehrnbecher T. Progress in the diagnosis of invasive fungal disease in children. Curr Fungal Infect Rep **2017**; 11:35–44.
44. Hummel M, Spiess B, Roder J, et al. Detection of *Aspergillus* DNA by a nested PCR assay is able to improve the diagnosis of invasive aspergillosis in paediatric patients. J Med Microbiol **2009**; 58:1291–7.
45. Wehrle-Wieland E, Affolter K, Goldenberger D, et al. Diagnosis of invasive mold diseases in patients with hematological malignancies using *Aspergillus*, Mucorales, and panfungal PCR in BAL. Transpl Infect Dis **2018**; e12953.
46. Huppler AR, Fisher BT, Lehrnbecher T, Walsh TJ, Steinbach WJ. Role of molecular biomarkers in the diagnosis of invasive fungal diseases in children. J Pediatric Infect Dis Soc **2017**; 6:32–44.
47. Grancini A, Orlandi A, Lunghi G, et al. Evaluation of real time PCR *Aspergillus* spp. in bronchoalveolar lavage samples. New Microbiol **2018**; 41:67–70.
48. White PL, Wingard JR, Bretagne S, et al. *Aspergillus* polymerase chain reaction: systematic review of evidence for clinical use in comparison with antigen testing. Clin Infect Dis **2015**; 61:1293–303.

49. Aguado JM, Vázquez L, Fernández-Ruiz M, et al.; PCRAGA Study Group; Spanish Stem Cell Transplantation Group; Study Group of Medical Mycology of the Spanish Society of Clinical Microbiology and Infectious Diseases; Spanish Network for Research in Infectious Diseases. Serum galactomannan versus a combination of galactomannan and polymerase chain reaction-based *Aspergillus* DNA detection for early therapy of invasive aspergillosis in high-risk hematological patients: a randomized controlled trial. Clin Infect Dis **2015**; 60:405–14.

50. Morrissey CO, Chen SC, Sorrell TC, et al.; Australasian Leukaemia Lymphoma Group and the Australia and New Zealand Mycology Interest Group. Galactomannan and PCR versus culture and histology for directing use of antifungal treatment for invasive aspergillosis in high-risk haematology patients: a randomised controlled trial. Lancet Infect Dis **2013**; 13:519–28.

51. Hardak E, Yigla M, Avivi I, Fruchter O, Sprecher H, Oren I. Impact of PCR-based diagnosis of invasive pulmonary aspergillosis on clinical outcome. Bone Marrow Transplant **2009**; 44:595–9.

52. 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.

53. Brasier AR, Zhao Y, Spratt HM, et al. 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.

54. Lucena CM, Rovira M, Gabarrús A, et al. The clinical value of biomarkers in respiratory complications in hematopoietic SCT. Bone Marrow Transplant **2017**; 52:415–22.

55. Bidula S, Sexton DW, Yates M, et al. H-ficolin binds *Aspergillus fumigatus* leading to activation of the lectin complement pathway and modulation of lung epithelial immune responses. Immunology **2015**; 146:281–91.

56. Heldt S, Prattes J, Eigl S, et al. Diagnosis of invasive aspergillosis in hematological malignancy patients: performance of cytokines, Asp LFD, and *Aspergillus*

57. Taghizadeh-Armaki M, Hedayati MT, Moqarabzadeh V, et al. Effect of involved *Aspergillus* species on galactomannan in bronchoalveolar lavage of patients with invasive aspergillosis. J Med Microbiol **2017**; 66:898–904.

58. Davila S, Halstead ES, Hall MW, et al.; Eunice Kennedy Shriver Collaborative Pediatric Critical Care Research Network Investigators. Viral DNAemia and immune suppression in pediatric sepsis. Pediatr Crit Care Med **2018**; 19:e14–22.

59. Verhoeven DH, Claas EC, Jol-van der Zijde CM, et al. 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**; 34:1118–27.

60. Albert E, Solano C, Pascual T, et al. Dynamics of torque teno virus plasma DNAemia in allogeneic stem cell transplant recipients. J Clin Virol **2017**; 94:22–8.

61. Wohlfarth P, Leiner M, Schoergenhofer C, et al. Torquetenovirus dynamics and immune marker properties in patients following allogeneic hematopoietic stem cell transplantation: a prospective longitudinal study. Biol Blood Marrow Transplant **2018**; 24:194–9.

62. Wootton SC, Kim DS, Kondoh Y, et al. Viral infection in acute exacerbation of idiopathic pulmonary fibrosis. Am J Respir Crit Care Med **2011**; 183:1698–702.

63. Salez N, Vabret A, Leruez-Ville M, et al. Evaluation of four commercial multiplex molecular tests for the diagnosis of acute respiratory infections. PLoS One **2015**; 10:e0130378.

64. Yanik GA, Mineishi S, Levine JE, et al. Soluble tumor necrosis factor receptor: Enbrel (etanercept) for subacute pulmonary dysfunction following allogeneic stem cell transplantation. Biol Blood Marrow Transplant **2012**; 18:1044–54.

65. Pathogen identification in pediatric hematopoietic stem cell transplant patients with suspected lower respiratory tract infection. **2016**. Available at: http://clinicaltrials.gov/ct2 (Identification No. NCT02926612).