Title: Metagenomic Sequencing Detects Respiratory Pathogens in Hematopoietic Cellular Transplant Patients

Running Title: mNGS Detects Respiratory Pathogens in HCT Patients

Authors: Langelier, C1*; Zinter, MS2*; Kalantar, K3; Yanik, GA4; Christenson, S5; Odonovan, B3; White, C3; Wilson, M6; Sapru, A2; Dvorak, CC7; Miller, S8; Chiu, CY1,8; DeRisi, JL3,9

* Langelier C and Zinter MS share first authorship.

Affiliations:

1 Division of Infectious Diseases, Department of Medicine, University of California, San Francisco, CA, USA
2 Division of Critical Care Medicine, Department of Pediatrics, University of California, San Francisco - School of Medicine, San Francisco, CA, USA
3 Department of Biochemistry and Biophysics, University of California, San Francisco, San Francisco, CA, USA
4 Division of Hematology Oncology, Department of Pediatrics, University of Michigan, Ann Arbor – School of Medicine, Ann Arbor, MI, USA
5 Division of Pulmonary, Critical Care, Sleep and Allergy, Department of Medicine, University of California, San Francisco, USA
6 Weill Institute for Neurosciences, Department of Neurology, University of California, San Francisco, San Francisco, CA, USA
7 Division of Allergy, Immunology, and Blood and Marrow Transplantation, Department of Pediatrics, University of California, San Francisco - School of Medicine, San Francisco, CA, USA
Department of Laboratory Medicine, University of California, San Francisco - School of Medicine, San Francisco, CA, USA

Chan Zuckerberg Biohub, San Francisco CA

**Corresponding author:**

Charles Langelier, MD, PhD
Division of Infectious Diseases
Department of Medicine
1700 4th St, Room 403
San Francisco CA, 94158
chaz.langelier@ucsf.edu

**Funding:**

NHLBI K12HL119997 (Langelier C), NICHD K12HD000850, the Pediatric Blood and Marrow Transplant Foundation, the National Marrow Donor Program Amy Strelzer Manasevit Grant (Zinter MS), NHLBI U10HL069330 (Yanik GA), NHLBI K23HL123778 (Christensen S), NINDS K08NS096117 (Wilson, M), NHLBI R01HL114484 (Sapru A), NIAID U54AI082973 (Dvorak CC), NHLBI R01HL105704 and NIAID R21AI120977 (Chiu, CY), NIAID P01AI091575 and the Chan Zuckerberg Initiative (DeRisi JL).

**Author Contributions:**

*Study concept and design:* Langelier, Zinter, Sapru, Dvorak, Yanik, DeRisi.

*Acquisition of data:* Langelier, Zinter.

*Analysis and interpretation of data:* Langelier, Zinter, Kalantar, White, O'Donovan, Christenson, DeRisi, Miller, Chiu.
Drafting of the manuscript: Langelier, Zinter.

Critical revision of the manuscript for important intellectual content: Langelier, Zinter, Miller, Chiu, DeRisi.

Statistical analysis: Langelier, Zinter, Christenson, Wilson, O'Donovan, Kalantar, DeRisi.

Administrative, technical, or material support: Langelier, Zinter, Sapru, Yanik, Dvorak, DeRisi.

Study supervision: DeRisi, Sapru, Yanik.

Approval of final manuscript: Langelier, Zinter, Kalantar, Yanik, White, O’Brien, Wilson, Miller, Chiu, Sapru, Yanik, Dvorak, DeRisi

Commentary: Lower respiratory tract infections are a leading cause of death in hematopoietic cellular transplant recipients, but the microbiologic etiology of these infections is frequently unknown. The limits of current diagnostics preclude targeted antimicrobial treatments, resulting in excess morbidity and mortality. Metagenomic next-generation sequencing may enable precision diagnosis of respiratory infections in hematopoietic cellular transplant patients by simultaneously detecting microbial pathogens, transcriptional biomarkers of the host response, and the pulmonary microbiome. This approach may facilitate targeted antimicrobial prescription and provide a new method for distinguishing infectious and non-infectious post-transplant respiratory illness.

Conflicts of Interest: None

Descriptor: 10.04 Diagnosis of Infections

Word Count: 3476
ABSTRACT

RATIONALE: Current microbiologic diagnostics often fail to identify the etiology of lower respiratory tract infections (LRTI) in hematopoietic cellular transplant recipients (HCT), which precludes the implementation of targeted therapies.

OBJECTIVES: To address the need for improved LRTI diagnostics, we evaluated the utility of metagenomic next generation sequencing (mNGS) of bronchoalveolar lavage (BAL) to detect microbial pathogens in HCT patients with acute respiratory illnesses.

METHODS: We enrolled 22 post-HCT adults ages 19-69 years with acute respiratory illnesses who underwent BAL at the University of Michigan between January 2012 and May 2013. mNGS was performed on BAL fluid to detect microbes and simultaneously assess the host transcriptional response. Results were compared against conventional microbiologic assays.

MEASUREMENTS & MAIN RESULTS: mNGS demonstrated 100% sensitivity for detecting respiratory microbes (human metapneumovirus, respiratory syncytial virus, Stenotrophomonas maltophilia, human herpesvirus 6 and cytomegalovirus) when compared to standard testing. Previously unrecognized LRTI pathogens were identified in six patients for whom standard testing was negative (human coronavirus 229E, human rhinovirus A, Corynebacterium propinquum and Streptococcus mitis); findings were confirmed by independent PCR and 16S rRNA sequencing. Relative to patients without infection, patients with infection had increased expression of immunity related genes (p=0.022) and significantly lower diversity of their respiratory microbiome (p=0.017).
CONCLUSIONS: Compared to conventional diagnostics, mNGS enhanced detection of pathogens in BAL fluid from HCT patients. Furthermore, our results suggest that combining unbiased microbial pathogen detection with assessment of host gene biomarkers of immune response may hold promise for enhancing the diagnosis of post-HCT respiratory infections.

Abstract Word Count: 249

Keywords: next generation sequencing, pneumonia, lower respiratory tract infection, bone marrow transplant, transcriptome
INTRODUCTION:

Lower respiratory tract infections (LRTI) are a leading reason for hospitalization and mortality in hematopoietic cellular transplant (HCT) recipients (1, 2). This problem is underscored by autopsy studies showing that previously undetected pulmonary pathogens may contribute to death in 30% of HCT recipients (3, 4). Despite this, the etiologic pathogens remain unidentified in most cases of LRTI, due largely to the limitations of current microbiologic tests in terms of sensitivity, speed and breadth of assay targets (5).

Diagnosis of LRTI in HCT recipients is particularly challenging due to high rates of non-infectious inflammatory conditions such as graft versus host disease (GVHD) that can drive pulmonary inflammation, induce fever and mimic infection (6). Furthermore, the diagnostic yield of traditional tests is reduced in HCT patients due to antimicrobial prophylaxis, reduced antibody titers, and infections from uncommon opportunistic microorganisms (7).

The limitations of current microbiologic tests drive excessive use of empiric broad-spectrum antimicrobials, which potentiates the emergence of drug resistance and increases risk of Clostridium difficile infection (8). In some situations, empiric regimens may lack activity against an underlying microbe, leading to treatment failure, disease progression, and consequent adverse outcomes (5, 9). Furthermore, in transplant patients, concern for GVHD may compel clinicians to initiate empiric immunosuppressive agents that could inadvertently exacerbate disease in the presence of an unrecognized infection (10).

Previously, microarray approaches proved useful for broadening the scope of pathogens detectable in a single LRTI assay; some microarrays have even incorporated conserved sequences from all known viruses, allowing detection of novel viral strains (11). Metagenomic next-generation sequencing (mNGS) inherently offers enhanced diagnostic capabilities by providing a culture-independent, comprehensive measurement of the microbial composition of clinical samples (12, 13). This technology permits the simultaneous detection of bacterial, viral and fungal pathogens without introducing bias associated with fixed-target PCR or serologic
assays (11, 12, 14). By capturing both microbial and human RNA, mNGS also permits simultaneous transcriptional profiling of the host immunologic response, which can provide complementary insight regarding the presence and type of infection (15, 16). Due to the clear need for enhanced respiratory pathogen diagnostics in HCT recipients, we undertook this study examining the utility of host/pathogen mNGS for detection of LRTI pathogens in HCT patients hospitalized with acute respiratory illnesses.

MATERIALS AND METHODS:

**Subjects:**

This retrospective study evaluated sequentially enrolled adult HCT recipients who underwent bronchoscopy with bronchoalveolar lavage (BAL) at the University of Michigan Medical Center between January 25th, 2012 and May 20th, 2013. All patients met established clinical definitions of community-acquired or hospital-acquired pneumonia and had two or more of the following: cough, fever, chills, dyspnea, pleuritic chest pain, crackles, or bronchial breathing; in addition, an opacity, infiltrate or nodules on chest radiograph was required (17-19). In some cases, patients had other equally probable non-infectious explanations for respiratory symptoms. All subjects had routine microbiological testing of BAL fluid as part of standard of care diagnostic workup for respiratory illness. Patients consented to permit banking of surplus BAL specimens in a biorepository at -80°C in accordance with University of Michigan IRB protocol HUM00043287. De-identified samples were then transported to UCSF for mNGS.

**Clinical Microbiologic Testing**

During the period of study enrollment, standard of care clinical microbiologic diagnostic testing for respiratory pathogens included bacterial, mycobacterial and fungal cultures, CMV culture, *Aspergillus* galactomannan assay, multiplex PCR for influenza A/B, respiratory syncytial
virus (RSV) and human metapneumovirus (HMPV), human herpesvirus-6 (HHV-6) PCR, and silver stain for *Pneumocystis jiroveci*. Additional studies included nasopharyngeal (NP) swab PCR for influenza A/B, RSV, and HMPV, blood cultures, serum PCR for herpes simplex virus (HSV) types 1 and 2, varicella-zoster virus (VZV), Epstein-Barr virus (EBV), cytomegalovirus (CMV) and HHV-6, *Clostridium difficile* toxin stool PCR, and select additional studies on BAL per the clinical judgment of the treating physicians.

**Metagenomic Next-Generation Sequencing**

Total nucleic acid was extracted from 250µl of patient BAL fluid using bead-based lysis and the Zymo Viral DNA/RNA Kit (Zymo Research). RNA and DNA were isolated separately using DNAse or RNase, respectively, and the former was reverse transcribed to generate complementary DNA (cDNA). DNA and cDNA then underwent adapter addition and barcoding using the Nextera system (Illumina). Depletion of abundant sequences by hybridization (DASH) was employed to selectively deplete human mitochondrial cDNA, thus enriching for microbial transcripts (20). The final RNAseq and DNA sequencing (DNAseq) libraries underwent 135 nucleotide paired-end Illumina sequencing. Methodology is described further in the online data supplement.

**Pathogen Detection Bioinformatics**

Detection of microbial pathogens leveraged a custom bioinformatics pipeline to discriminate pathogens from background microbial sequences in clinical samples. This methodology is detailed in the Online Data Supplement (21). Briefly, this analytical approach involved first filtering for quality and complexity, next extracting sequences aligning to the human and Pan troglodytes genomes (NCBI GRC h38 and UCSC PanTro4) using STAR (22) and then subsequently removing reads aligning to non-fungal eukaryotes and phage using
Bowtie2 (23). The identities of the remaining reads were determined by querying the NCBI nucleotide (nt) and non-redundant protein (nr) databases using GSNAP-L.

To mitigate spurious taxonomic assignments due to laboratory background contaminants, we calculated a standard Z-score for each microbial genus relative to a control group of 17 “no-template” water-only controls (detailed in Table E1) derived from sequencing runs performed in our laboratory over the past 2 years (24). Z-scores were calculated based on unique reads aligned to either NCBI nt or nr, per million sequence input reads (rpM_{nt}, rpM_{nr}).

The respiratory tract is not a sterile environment, and as such, we devised a ranking score to aid in the stratification of the inherently complex alignment data. This metric involved multiplying the number of reads aligned per million reads sequenced by the sum of the nt and nr Z-scores for each microbial genus identified in each patient: score = rpM_{nt} \times (Z_{nt} + Z_{nr}). To further improve the specificity of microbial assignments, we required 1) microbes to be detected in both nucleotide (nt) and protein (nr) database alignments; 2) microbes to have a $Z_{nt}$ and $Z_{nr} > 0$; and 3) bacteria and fungi detected by RNAseq to also have detectable genomic sequence by DNAseq. In addition, we permitted DNA viruses to be detected on only DNAseq and RNA viruses to be detected on only RNAseq. For each patient, any microbes that passed these requirements were then processed using the following decision tree, summarized in Figure 1.

**Definitions**

Microbes were considered **confirmed pathogens** if a) both clinical testing and mNGS identified the microbe, b) there existed literature evidence of microbial pathogenicity in the lungs, and c) the microbe score was at least two-fold greater than that of any other microbe of the same type (virus, bacteria or fungus) identified in the patient (17). Microbes were considered **new potential pathogens** if a) mNGS alone identified the microbe, b) there existed literature evidence of microbial pathogenicity in the lungs, and c) the microbe score was two-fold greater than any other microbe of the same type (virus, bacteria or fungus) in the patient. New
potential pathogens were confirmed by independent specific PCR testing for viruses and 16s rRNA gene Sanger sequencing for bacteria. Finally, microbes were considered **unlikely or uncertain pathogens** if they a) lacked robust literature evidence of microbial pathogenicity in the lungs, b) had a score that was less than two-fold that of the top-ranking microbe, suggesting a polymicrobial sample, c) determined to be clinically insignificant by the treatment team, or d) were a DNA virus of uncertain pathogenicity present in low abundance (< 5rpM).

**Microbial diversity**

Alpha diversity of the respiratory microbiome in each subject was assessed using the Simpson Diversity Index (SDI) and compared between patients with a) confirmed pathogens, b) unlikely or uncertain pathogens and c) new potential pathogens using the nonparametric Wilcoxon Rank Sum test (25).

**Host gene expression analyses**

RNA transcripts aligning to the human genome were captured by our computational pipeline as described above. Cumulative sum scaling normalization of protein-coding gene transcripts was carried out (26) and genes expressed in fewer than 30% of samples, or as outliers in only 10% of samples, were removed. Using a supervised approach, pathways related to immune functionality were selected *a priori* from the Molecular Signatures Database (27) and compared in terms of total normalized expression between subjects using the nonparametric Wilcoxon Rank Sum test as described in the online data supplement. These pathways contained gene biomarkers associated with antiviral response, interferons α, β, and γ, IL-6/JAK/STAT5 signaling and adaptive immunity, and have all been strongly associated with the immune response to infection (28, 29).
RESULTS:

Cohort characteristics and clinical outcomes

We enrolled 22 HCT recipients hospitalized for acute respiratory symptoms aged 19-69 years at a median 356 days post-transplant, as summarized in Table 1. Their most common transplant indications were leukemia (n=12) and lymphoma (n=8). The majority of patients received allogeneic HCT (n=20) and/or myeloablative conditioning (n=20), as summarized in Table E2. The median absolute neutrophil count (ANC) at the time of BAL was 3.8/µL (IQR 2.1-5.4) and engraftment was achieved at the time of BAL in all patients. Acute GVHD was present in three subjects and chronic GVHD was present in 13 subjects. At the time of BAL, 15 subjects were receiving immune modulating agents.

Fever was present in 13 of 22 (59%) of patients, all patients met systemic inflammatory response syndrome criteria, and no patients had septic shock. Twelve patients required supplemental oxygen, three required mechanical ventilation and all survived their illnesses except for one subject who died 10 days after BAL due to lymphoma relapse. Seven of the 22 (32%) subjects died between 51 and 3024 days post-BAL and their primary causes of death were GVHD (n=3), relapsed malignancy (n=2), idiopathic respiratory failure (n=1), and unknown (n=1).

Antimicrobial use

All study subjects received antimicrobial prophylaxis prior to symptom onset and bronchoscopy. Empiric broad spectrum antibiotics were used in 16 of the 22 (73%) patients including vancomycin (n=4), cefepime (n=4), vancomycin + cefepime (n=5), cefepime + tobramycin (n=1), and vancomycin + cefepime + tobramycin (n=1) (Table 1). Pathogen-targeted antimicrobials were used in three subjects (14%). Four patients (15%) received no antimicrobials aside from pre-existing prophylactic agents.
Clinical Microbiologic Findings

Standard of care clinical BAL diagnostics performed at the study hospital (described above in Methods) identified microbes in seven of 22 (32%) patients, of which six were thought to represent etiologic pathogens by the treating physicians. These included HMPV (n=2), RSV (n=2), HHV-6 (n=1) and Stenotrophomonas maltophilia (n=1). CMV was identified by shell vial culture in subject 37 but thought to represent incidental carriage because of symptom resolution in the absence of intervention prior to the return of testing.

Next Generation Sequencing Findings

An average of 49 million paired-end sequencing reads were generated from each BAL sample, of which <1% were microbial. Sequencing statistics and pipeline output for each patient are described in Table E3. mNGS identified all seven microbes found by standard clinical diagnostics, demonstrating 100% sensitivity for pathogen detection. In total, RNAseq identified 10 RNA viruses, all of which have established pathogenicity in the lungs, and the RNA intermediates of five DNA viruses that have uncertain pulmonary pathogenicity. DNAseq identified the genomes of these same five DNA viruses as well as five others including CMV and HHV-6, which have been associated with pneumonitis. mNGS captured entire viral genomes for five patients at an average read depth of 3500-fold, as shown in Figure E1. Bacteria known to exist contextually as either pathogens or commensals, including Stenotrophomonas maltophilia, Streptococcus mitis, and Corynebacterium propinquum were identified, as were genera not typically considered pathogenic (30-32). Specificity of mNGS could not be assessed because this cohort lacked patients with proven non-infectious respiratory illnesses.

Relationship Between Respiratory Microbial Diversity and Detection of Pathogens

Loss of diversity within respiratory tract microbial communities has been proposed as an ecological marker of infection (33, 34). We thus evaluated alpha diversity of actively replicating
microbes identified by RNAseq using the SDI and found that subjects with confirmed pathogens had significantly lower diversity relative to patients with only microbes of unlikely or uncertain pathogenicity (0.34, IQR 0.15-0.64, n=6 vs. 0.92, IQR 0.86-0.93, n=10, p=0.017, Figure 2 and Table E4). Reduced diversity was also observed if subjects with confirmed or potential new pathogens were compared together against those without (0.41, IQR 0.20-0.55, n=12 vs. 0.92, IQR 0.86-0.93, n=10, p<0.001).

Analysis of Host Gene Expression

We hypothesized that gene expression from respiratory fluids obtained at the site of infection would distinguish patients with LRTI from those with non-infectious respiratory diseases. To test this idea, we evaluated a priori selected gene biomarkers of innate and adaptive immune responses using the total normalized expression of all genes. Given suspected between-subject immunologic response heterogeneity due to differences in infection type and degree of immunologic compromise, we created a composite metric based on the sum normalized gene expression of all biomarkers. We found significantly increased expression in patients with confirmed LRTI pathogens versus those without (94.9, IQR 93.8-105.6, n=6 vs. 33.1, IQR 20.7-75.1, n=7, p=0.022), as shown in Figure 3 and Table E5. When patients with confirmed or potential new pathogens were compared together against all others, the relationship trended towards significance (94.6, IQR 76.5-105.6, n=11 vs. 33.1, IQR 20.7-75.1, n=7, p=0.0634), and reached significance (p = 0.014) if a subject with a relatively lower abundance of HRV-A transcripts was instead grouped with patients found to have microbes of unlikely or uncertain pathogenicity. Raw gene counts are provided in the additional online supplemental data table.
DISCUSSION:

In this proof of concept study, we demonstrate that mNGS can simultaneously detect pathogens and the host response in HCT patients with acute respiratory illnesses. When benchmarked against the conventional standard of care testing performed during the time of study enrollment, mNGS demonstrated 100% sensitivity for microbial detection and permitted identification of potential new viral and bacterial pathogens in six of 15 (40%) subjects with otherwise negative testing. We observed that the presence of respiratory pathogens was characterized by increased expression of host immune response genes and lower airway microbial diversity. Our results suggest that simultaneous mNGS detection of respiratory pathogens, the host’s immunologic response and the airway microbiome may provide complementary data that could inform the clinical significance of a detected pathogen.

Amongst the cases with negative clinical testing, mNGS identified six patients harboring transcriptionally active microbes with recognized respiratory pathogenicity (35-40). With respect to viruses, HCOV-229E was identified in two patients and HRV-A was identified in three patients, one of whom was also culture-positive for *Stenotrophomonas maltophilia*. Notably, HRV and HCOV were not represented on the multiplex PCR panel used at the study hospital; similar limited panels are still in use at many hospitals today. Unlike rapid antigen or multiplex PCR assays, mNGS is not limited to a fixed number of pre-specified targets on a panel, obviating the need to order multiple independent diagnostic tests. Furthermore, mNGS can detect entire viral genomes, as illustrated in Figure E1, an attribute that can enable genotyping, detection of resistance mutations, and epidemiologic tracking of disease outbreaks (41).

With respect to previously unidentified bacterial pathogens, mNGS detected *Streptococcus mitis*, an oropharyngeal microbe known to cause bacteremia and acute respiratory distress in HCT recipients, as a potential pulmonary pathogen in one patient (40, 42). *Corynebacterium proriniquum*, one of the few virulent *Corynebacterium* species associated with LRTI, comprised the majority of microbial transcripts in another subject’s BAL
fluid (43). Both patients received empiric vancomycin, which has activity against *Streptococcus* and *Corynebacteria* spp., and recovered from their acute respiratory illnesses. Findings were independently confirmed at the genus level by 16s rRNA gene sequencing.

Ten patients with negative conventional testing were found to have microbes of unlikely or uncertain pathogenicity. Notably, each of these patients had potential alternative explanations for their respiratory symptoms. While one of these patients had bacteremia/sepsis, the remaining nine had acute and/or chronic GVHD, underlining the importance of non-infectious alloreactive inflammation in post-HCT pulmonary complications.

mNGS identified several viruses with DNA genomes, however only five of these also had well-defined evidence of active replication marked by detectable RNA transcripts (HSV, human papilloma virus and torque teno viruses (TTV, n=3). TTV was the most commonly detected DNA virus, a finding consistent with prior reports demonstrating an increased prevalence of this presumptively innocuous constituent of the human virome in immunosuppressed patients (44). Herpesviridae genomic DNA in the absence of viral transcripts was identified in five patients and included HHV-6 (n=1), CMV (n=2), HSV (n=1) and EBV (n=1). One subject, who died 10 days post-BAL due to relapsed lymphoma, had EBV DNA detected in the setting of EBV viremia. WU polyoma virus was detected by DNA sequencing another case, and while this virus has been associated with respiratory infections in immunocompromised patients, evidence supporting its role as a pathogen is lacking (45).

CMV was identified by both viral culture and mNGS in one subject, however this finding was considered unrelated to the patient’s respiratory illness by the treating physicians. Notably, while CMV was detected by DNAseq, no RNA transcripts were identified, suggesting that it may have represented incidental carriage as opposed to a transcriptionally active pathogen (46). This patient was also found to harbor *Pseudomonas fluorescens*, a described colonizer of the human respiratory tract (47), as well as several other taxa representing common constituents of
the oropharyngeal or respiratory microbiome, including *Rothia*, *Prevotella*, and *Actinomyces* spp., as detailed in Table E3.

**Lung microbiome diversity and the host immune response as biomarkers of infection**

Because asymptomatic carriage of respiratory pathogens is well described (48, 49), establishing biomarkers of genuine infection is critical for determining whether a given microbiologic finding is clinically significant. Our findings suggest that respiratory tract microbial diversity may be such a biomarker. Specifically, we found that patients with LRTI pathogens had significantly lower alpha diversity versus those without (p=0.017, Figure 2, Table E5), presumably reflecting dominance of actively replicating pathogens (33, 34).

Our results also demonstrate that both mNGS and conventional methods identify clinically significant and insignificant microbes, and emphasize the need to assess the impact of a given microbiologic finding in the context of a patient’s immunologic response (5). Expression of a multi-gene immune response metric was significantly increased in patients with confirmed respiratory pathogens relative to those without, suggesting that despite significant and heterogeneous states of immune suppression, HCT recipients still exhibited immunologic biomarkers of active infection (p=0.022, Figure 3).

Amongst subjects with potential new pathogens, we observed that two of the three HRV-A positive subjects demonstrated the lowest expression of this immune response metric, while the remaining subject, who was co-infected with HRV-A and *Stenotrophomonas maltophilia*, had one of the highest values. This is consistent with prior reports demonstrating that HRV can induce a broad range of clinical disease severity, and that viral-bacterial co-infection can increase the severity of disease (48, 49). Amongst subjects with confirmed pathogens, the patient with clinically suspected HHV-6 pneumonitis was notably an outlier in terms of reduced immune metric expression and elevated airway diversity. The respiratory pathogenicity of HHV-6 is controversial, and our paired mNGS pathogen data demonstrating absence of viral
transcripts but detectable genomic DNA support a less virulent role for HHV-6 in this patient (50).

Broadspectrum antibiotic overuse is frequently driven by suspicion for occult pathogens missed by standard diagnostic testing. For instance, patient 1, whose conventional testing identified only HMPV, received empiric vancomycin and cefepime due to suspected occult bacterial infection and subsequently developed C. difficile colitis. mNGS confirmed the presence of HMPV and also demonstrated absence of bacterial pathogens in this patient. The theoretical availability of our mNGS findings during the actual period of patient hospitalization potentially could have informed more targeted antimicrobial use in the 18 study subjects who received broad spectrum antibiotics in the absence of detectable bacterial pathogens by mNGS.

While these results are highly encouraging, this proof-of-concept study suggests many routes for further improvement. First, this study was limited by a relatively small sample size, only one case of culture-positive bacterial LRTI, no fungal infections, and no subjects with clinically diagnosed non-infectious airway diseases. Future studies with larger cohorts will be needed to validate the sensitivity and specificity of mNGS for LRTI diagnosis in this population. Second, our limited sequencing depth did not yield the human transcriptome coverage that would be desired for robust differential gene expression analyses, although we were able to rigorously evaluate a composite metric of immunity genes. Future larger studies are needed to identify and validate gene classifiers that can distinguish LRTI from non-infectious airway diseases in HCT recipients.

CONCLUSIONS:

Here we leverage continued improvements in metagenomic sequencing to expand the capabilities of LRTI diagnostics in HCT recipients with acute respiratory illnesses. We demonstrate that compared to current microbial diagnostics, mNGS has a greater capacity for detecting microbes and an ability to couple pathogen detection with simultaneous profiling of the
host response and the airway microbiome. We suggest several ways in which future studies may improve on this proof-of-concept study, including validation in a larger prospective cohort of HCT recipients and other high-risk populations.
**TABLE LEGENDS**

Table 1: Clinical and Microbiologic Data. Abbreviations: Radiography: B/L, bilateral; LUL, left upper lobe; RUL, right upper lobe. Chest CT was obtained for all subjects except 1, 3, 4, 8, 9 and 14 who received chest X-rays. **Antimicrobials:** ACV, acyclovir; CIDV, cidofovir; CPM, cefepime; FOSC, foscarnet, FLUC, fluconazole; GCV, ganciclovir; MICA, micafungin; POSA, posaconazole; RIBV, ribavirin; TMP-SMZ, trimethoprim-sulfamethoxazole; TOBR, tobramycin; vACV, valacyclovir; VANC, vancomycin; VORI, voriconazole. **Respiratory Support:** HFNC, high flow nasal cannula; IPPV, invasive positive pressure ventilation; NC, nasal cannula. *Clinicians concluded that CMV in subject 37 was not the principal cause of respiratory disease.*

**FIGURE LEGENDS**

Figure 1. Overview of Pathogen Detection Workflow. Total nucleic acid extracted from BAL fluid of HCT recipients underwent DNA and RNA sequencing. A custom bioinformatics pipeline simultaneously identified pathogens and assayed the human transcriptome. Pathogens meeting inclusion and exclusion criteria were ranked by microbial significance score [Score = \( r_p M_{nt} \times (Z_{nt} + Z_{nr}) \)]. To further improve the specificity of microbial assignments: we required 1) microbes to be detected in both nt and nr database alignments, 2) microbes to have a \( Z_{nt} \) and \( Z_{nr} \) greater than zero, and 3) bacteria and fungi detected by RNAseq to also have detectable genomic sequence by DNAseq. For each patient, any microbes that passed these requirements were then classified as either 1) confirmed pathogens, 2) new potential pathogens or 3) unlikely or uncertain pathogens.

Figure 2. BAL Microbial Diversity is Inversely Associated with Presence of a Transcriptionally Active Respiratory Pathogen. Each data point represents a single patient.
for whom the Simpson Diversity Index (SDI) is plotted on the y-axis. Subjects are grouped according to confirmed pathogen (red triangles) vs. unlikely or uncertain pathogen (black circles). Patients with confirmed pathogens had significantly lower diversity relative to patients with only microbes of unlikely pathogenicity (0.34, IQR 0.15-0.64, n=6, vs. 0.92, IQR 0.86-0.93, n=10, p=0.017). Raw data are listed in Table E4.

**Figure 3. Expression of a Host Immune Response Multi-Gene Metric Correlates with Detection of LRTI Pathogens.** Each data point represents a single patient for whom the composite immune response gene metric is plotted on the y-axis. Subjects are grouped according to confirmed pathogen (red triangles) vs. or unlikely or uncertain pathogen (black circles). Patients with confirmed pathogens had significantly higher multi-gene metric relative to patients with only microbes of unlikely pathogenicity (33.1, IQR 20.7-75.1, n=7 vs. 94.9, IQR 93.8-105.6, n=6, p=0.022). Raw data are listed in Table E5.
References

(1) Center for International Blood and Marrow Transplant Research (CIBMTR), National Marrow Donor Program (NMDP), European Blood and Marrow Transplant Group (EBMT), American Society of Blood and Marrow Transplantation (ASBMT), Canadian Blood and Marrow Transplant Group (CBMTG), Infectious Disease Society of America (IDSA), Society for Healthcare Epidemiology of America (SHEA), Association of Medical Microbiology and Infectious Diseases Canada (AMMI), Centers for Disease Control and Prevention (CDC). Guidelines for preventing infectious complications among hematopoietic cell transplant recipients: a global perspective. *Bone Marrow Transplant* 2009;44:453-558.


Table 1. Clinical and Microbiologic Data

<table>
<thead>
<tr>
<th>Pathogens identified by conventional and NGS diagnostics</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Patient ID</strong></td>
</tr>
<tr>
<td>---------------</td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
<tr>
<td>10</td>
</tr>
<tr>
<td>13</td>
</tr>
<tr>
<td>14</td>
</tr>
<tr>
<td>18</td>
</tr>
<tr>
<td>19</td>
</tr>
</tbody>
</table>

New potential pathogens identified by NGS

<table>
<thead>
<tr>
<th><strong>Patient ID</strong></th>
<th><strong>Symptoms and Duration</strong></th>
<th><strong>Chest Radiography</strong></th>
<th><strong>Standard BAL Microbiology</strong></th>
<th><strong>Top Scoring Microbes by nNGS</strong></th>
<th><strong>Non-BAL Microbiology</strong></th>
<th><strong>Antimicrobial Prophylaxis</strong></th>
<th><strong>Antimicrobial Treatment</strong></th>
<th><strong>GVHD</strong></th>
<th><strong>Respiratory Support</strong></th>
<th><strong>Cause of death</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>Cough, dyspnea</td>
<td>x 3 days B/L nodules</td>
<td>Negative</td>
<td>HCOV-229E</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>CFPM, VANC</td>
<td>aGVHD, cGVHD</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>7</td>
<td>Dyspnea</td>
<td>x 1 day B/L perihilar opacities</td>
<td>Negative</td>
<td>HCOV-229E</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>CFPM, VANC</td>
<td>aGVHD, cGVHD</td>
<td>None</td>
<td>IPPV x 4 days</td>
</tr>
<tr>
<td>10</td>
<td>Cough, dyspnea</td>
<td>x 2 days LLL opacities</td>
<td>Negative</td>
<td>HRV-A</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>aGVHD, cGVHD</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>18</td>
<td>Cough, dyspnea</td>
<td>x 2 days RML, RLL atelectasis</td>
<td>Negative</td>
<td>HRV-A</td>
<td>S. epidermidis (blood)</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>aGVHD, cGVHD</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>19</td>
<td>Dyspnea</td>
<td>x 4 wks RUL, LUL reticular opacities</td>
<td>Negative</td>
<td>Streptococcus millis</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>aGVHD, cGVHD</td>
<td>None</td>
<td>alive</td>
</tr>
</tbody>
</table>

Microbes of uncertain or unlikely pathogenicity

<table>
<thead>
<tr>
<th><strong>Patient ID</strong></th>
<th><strong>Symptoms and Duration</strong></th>
<th><strong>Chest Radiography</strong></th>
<th><strong>Standard BAL Microbiology</strong></th>
<th><strong>Top Scoring Microbes by nNGS</strong></th>
<th><strong>Non-BAL Microbiology</strong></th>
<th><strong>Antimicrobial Prophylaxis</strong></th>
<th><strong>Antimicrobial Treatment</strong></th>
<th><strong>GVHD</strong></th>
<th><strong>Respiratory Support</strong></th>
<th><strong>Cause of death</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>Fever</td>
<td>x 2 days B/L infiltrates</td>
<td>Negative</td>
<td>Flavobacterium sp.</td>
<td>Gram positive rod (blood)</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>GVHD, day 267</td>
</tr>
<tr>
<td>11</td>
<td>Cough, fever</td>
<td>x 2 days LLL opacities</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>GVHD, day 267</td>
</tr>
<tr>
<td>22</td>
<td>Dyspnea, rigors</td>
<td>x 1 day RUL, LUL peribronchial thickening, GG</td>
<td>Negative</td>
<td>Torque teno virus</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>23</td>
<td>Cough, dyspnea, fever</td>
<td>x 3 wks B/L GG, lymphadenopathy</td>
<td>Negative</td>
<td>Negative</td>
<td>EBV viremia</td>
<td>ACV, LEVO, MICA, TMP-SMZ, VORI</td>
<td>CFPM</td>
<td>aGVHD</td>
<td>IPPV x 10 days (until death)</td>
<td>No</td>
</tr>
<tr>
<td>24</td>
<td>Dyspnea, fever, week</td>
<td>x 4 days B/L peribronchial thickening, GG</td>
<td>Negative</td>
<td>Pyrenophora sp.</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>25</td>
<td>Cough, dyspnea, fever</td>
<td>x 2 wks Diffuse bronchiectasis, multifocal GG nodules</td>
<td>Negative</td>
<td>WU Polyoma Virus</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>31</td>
<td>Dyspnea</td>
<td>x 2 wks B/L GG bronchiectasis</td>
<td>Negative</td>
<td>Prevotella</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>34</td>
<td>Cough, dyspnea</td>
<td>x 2 wks Multifocal nodules, opacities, air trapping</td>
<td>Negative</td>
<td>Torque teno virus</td>
<td>Negative</td>
<td>ACV, LEVO, VORI</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>35</td>
<td>Cough, dyspnea</td>
<td>x 2 wks Multifocal GG, tree-in-bud opacities</td>
<td>Negative</td>
<td>Negative</td>
<td>Negative</td>
<td>ACV, LEVO</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>alive</td>
</tr>
<tr>
<td>37</td>
<td>Cough, dyspnea</td>
<td>x 8 days B/L GG</td>
<td>Negative</td>
<td>CMV</td>
<td>Negative</td>
<td>ACV, LEVO</td>
<td>VANC</td>
<td>No</td>
<td>None</td>
<td>Alive</td>
</tr>
</tbody>
</table>
Figure 1

BAL

RNA Sequencing + DNA Sequencing

\[
\text{Rank taxa by score} = rpM \times (Z_{nt} + Z_{nr})
\]

Inclusion Criteria
1. Genus identified by nt and nr database alignments
2. Genus Znt and Znr > 0
3. Bacteria and Fungi detected by RNAseq and DNAseq
4. DNA viruses may be detected only by DNAseq
5. RNA viruses may be detected only by RNAseq

Significant Microbe List

Confirmed Pathogens
New Potential Pathogens
Unlikely Pathogens

Confirm by PCR
Figure 2

** p=0.017

Simpson Diversity Index

Unlikely or Uncertain Pathogen

Confirmed Pathogen

Pt 35, 24
Pt 11, 23, 37, 3
Pt 25, 34, 31
Pt 22
Pt 36
Pt 5
Pt 9
Pt 8
Pt 10
Pt 1
Figure 3

**p = 0.022**

- Pt 5
- Pt 10
- Pt 9, 8, 1
- Pt 36
- Pt 22
- Pt 23
- Pt 11
- Pt 25, 3
- Pt 35, 31

Y-axis: Multigene Immunity Metric
X-axis: Unlikely or Uncertain Pathogen vs. Confirmed Pathogen