Title: Metagenomic Next-Generation Sequencing Detects Pulmonary Pathogens in Hematopoietic Cellular Transplant Patients with Acute Respiratory Illnesses

Running Title: mNGS Detects Respiratory Pathogens in HCT Patients

Authors: Langelier, C[1]*; Zinter, MS[2]*; Kalantar, K[3]; Yanik, GA[4]; Christenson, S[5]; Odonovan, B[3]; White, C[3]; Wilson, M[6]; Sapru, A[2]; Dvorak, CC[7]; Miller, S[8]; Chiu, CY[1,8]; DeRisi, JL[3]

* 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
[8] Department of Laboratory Medicine, University of California, San Francisco - School of Medicine, San Francisco, CA, USA

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 and the Pediatric Blood and Marrow Transplant Foundation (Zinter MS), NHLBI U10HL069330 (Yanik GA), NHLBI K2HL123778 (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).

Conflicts of Interest: None

Keywords: hematopoietic stem cell transplantation; stem cell transplantation; bone marrow transplantation; pneumonia; infection; genomic sequencing;

Descriptor: 10.4 Diagnosis of Infections

Word Count: 3466

At A Glance Commentary: Unbiased metagenomic next-generation sequencing (mNGS) may enable precision diagnosis of respiratory infections in HCT patients by simultaneously detecting microbial pathogens, transcriptional biomarkers of the host's response and endogenous lung microbiota. This approach may facilitate targeted antimicrobial therapy and provide a new approach for distinguishing infectious from non-infectious post-HCT respiratory illness.

ABSTRACT:

BACKGROUND: Lower respiratory tract infections (LRTI) are a leading cause of mortality in hematopoietic cell transplant (HCT) recipients. Current microbiologic diagnostics often fail to identify etiologic pathogens, leading to diagnostic uncertainty and precluding the implementation of targeted therapies. To directly address the need for improved LRTI diagnostics, we undertook this study to evaluate the potential utility of metagenomic next generation sequencing (mNGS) approaches for detecting LRTI in the HCT population.

METHODS: We enrolled 22 post-HCT adults ages 19-69 years with acute respiratory illnesses who underwent bronchoalveolar lavage (BAL) at the University of Michigan between January 2012 and May 2013. Unbiased mNGS was performed on BAL fluid to detect microbes and assess host response. Results were compared to those obtained by standard clinical microbiology testing.

RESULTS: Unbiased mNGS detected all microbes identified by standard testing (human metapneumovirus, respiratory syncytial virus, *Stenotrophomonas maltophilia,* human herpesvirus 6 and cytomegalovirus). 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)* and findings were confirmed by independent PCR testing. mNGS identified microbes of unlikely or uncertain pathogenicity in 10 patients with clinical evidence of non-infectious respiratory conditions. Patients with respiratory pathogens were found to have significantly increased expression of immunity related gene biomarkers relative to those without ($p=0.022$) as well as lower alpha diversity of their respiratory microbial communities ($p=0.017$).

CONCLUSIONS: Compared to conventional diagnostics, host/pathogen mNGS enhanced detection of microbial pathogens in BAL fluid from HCT patients. Furthermore, this approach simultaneously evaluated the association between identified microbes and the expression of innate immunity gene biomarkers. Host/pathogen mNGS holds promise for precision diagnosis of post-HCT respiratory infection.

INTRODUCTION:

Lower respiratory tract infections (LRTI) are a leading cause of both hospitalization and mortality in hematopoietic cellular transplantation (HCT) recipients (1-3). This problem is underscored by autopsy studies showing that 30% of HCT patients had previously undetected pulmonary pathogens that likely contributed to death (4, 5). Despite this, the etiologic pathogens remain unidentified in most cases due to the limitations of current microbiologic tests in terms of sensitivity, speed and limited number of available targets (6).

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 (7-10). 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 (11, 12).

The limitations of current microbiologic tests drive excess use of empiric broad-spectrum antimicrobials, which potentiates the emergence of drug resistance and increases risk of *Clostridium difficile* infection (13). In some situations, empiric regimens may lack activity against the underlying microbe, resulting in delayed effective treatment, disease progression, and consequent adverse outcomes (6, 14). Furthermore, in transplant patients, concern for GVHD may compel clinicians to initiate empiric immunosuppressive agents that could inadvertently exacerbate disease in the presence of unrecognized infection (15).

Previously, microarray approaches proved useful for broadening the scope of pathogens detectable in a single LRTI assay, but were limited to detection of *a priori* selected targets (16). Metagenomic next-generation sequencing (mNGS) now offers significantly enhanced diagnostic capabilities by providing a culture-independent, comprehensive measurement of the microbial composition (17-22). This technology permits the simultaneous detection of bacterial, viral and fungal pathogens without introducing bias associated with fixed-target PCR or serologic assays (16, 18, 20, 22). By capturing both microbial and human RNA, mNGS also permits simultaneous transcriptional profiling of the host immunologic response associated with the presence and type of infection, and thus provides complementary insight regarding pathogenesis (6, 23-25). 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 randomly chosen 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 clinical definitions of community or hospital-acquired pneumonia that included cough plus fever, sputum production, breathlessness, wheeze, chest discomfort or pain and had evidence

of lower respiratory tract involvement on chest imaging. In some cases patients had other equally probable non-infectious explanations for respiratory symptoms (26-30). 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. The UCSF study team was blinded to the characteristics and outcomes associated with each patient were blinded to the study team until after mNGS data analysis was completed.

Clinical Microbiologic Testing

During the period of study enrollment, standard-of-care clinical microbiological diagnostic testing for respiratory infections included bacterial, mycobacterial and fungal cultures, CMV shell viral 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 nasophayngeal (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 used to selectively deplete human mitochondrial cDNAs from the tagmented library, thus enriching the mNGS libraries for non-human sequences (31). The final RNA sequencing (RNAseq) and DNA sequencing (DNAseq) libraries underwent 135 nucleotide paired-end sequencing on an Illumina HiSeq 4000 Please refer to **Supplemental Text** for further description.

Pathogen Detection Bioinformatics

Detection of microbial pathogens leveraged a custom bioinformatics pipeline that incorporates Bayesian modeling to discriminate pathogens from background microbial sequences (32). Briefly, this analytical approach involved first filtering for quality and complexity, next extracting sequences aligning to the human genome (NCBI GRC h38) using STAR (33) and then subsequently removing reads aligning to non-fungal eukaryotes, phage, nonhuman tropic viruses using Bowtie2 (34) as well as contaminants in water controls. The identities of the remaining reads were determined by querying the NCBI nucleotide (nt) and non-redundant protein (nr) databases using GSNAP-L. Please refer to **Supplemental Text** for further description.

For each subject, microbial taxa at the genus level were assigned a microbial significance score based on the product of (1) the reads per million RNA transcripts sequenced (rpM) and (2) a composite genus Z-score reflecting the relative abundance of transcripts in the nt and nr databases compared to all other patients in the cohort: score = rpM x (Z_{nt} + Z_{nr}). To reduce spurious alignments, we required (1) bacteria and fungi detected by RNAseq to also have detectable genomes by DNAseq, (2) microbes to be detected by both nt and nr databases, and (3) microbes to have a Z_{nt} and Z_{nr} greater than zero. Viral pathogens not previously identified by standard clinical testing were confirmed by independent specific PCR testing. 16s rRNA gene sequencing (University of Washington) was used to confirm identification of top ranking bacterial pathogens at the genus level. Finally, alpha diversity of the respiratory microbiome in each subject was assessed using the Simpson Diversity Index (SDI) (35).

Definitions

Subjects were grouped into one of three categories based on microbiologic results. **Confirmed pathogen** if both clinical testing and mNGS identified a microbe with established pathogenicity in the lungs and the pathogen scored 10-fold greater than any other microbe identified by mNGS (26). **New potential pathogen** if previously occult microbes were detected by mNGS with literature evidence of pathogenicity in the lungs and with a score 10-fold greater than any other microbe of the same type, in terms of virus, bacteria or fungus. **Unlikely or uncertain pathogen** if microbes detected by any method were: a) not known to exist as pulmonary pathogens, b) potential pathogens but scoring 10-fold less than the top-ranking microbe as part of a polymicrobial sample, c) determined to be clinically insignificant by the treatment team, d) known to be a common laboratory contaminant, or e) a DNA virus of uncertain pathogenicity present in low abundance.

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 (36) 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 (37) and compared in terms of total normalized expression between subjects using the nonparametric Wilcoxon Rank Sum test (**Supplemental Text and Table S4**). 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 (38, 39).

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 (**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, **Table S1**). 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 3 subjects and chronic GVHD was present in 13 subjects. At the time of BAL, immune modulating agents were used for GVHD prophylaxis or treatment in 15 allogeneic recipients.

Fever was present in 57% 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 (**Table 1**). Seven subjects died between 51 and 3024 days later and their primary causes of death were GVHD (n=3), relapsed malignancy (n=2), and idiopathic pulmonary failure (n=1).

Antimicrobial use

All study subjects received antimicrobial prophylaxis prior to symptom onset and bronchoscopy. Empiric antibiotics were used in 16 patients including vancomycin (n=4), cefepime (n=4), vancomycin + cefepime (n=5), cefepime + tobramycin (n=1), and vancomycin + cefepime + tobramycin (n=1) (**Table 1**). Targeted antimicrobials were used in three subjects (TMP-SMZ for *Stenotrophomonas maltophilia*, foscarnet for HHV-6 and ribavirin for RSV). Four patients received no antimicrobials aside from pre-existing prophylactic agents.

Clinical Microbiologic Findings

Standard-of-care clinical BAL diagnostics performed at the study hospital (see methods) identified microbes in seven 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 through 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 (**Table S2**). mNGS identified all 7 microbes found by standard clinical diagnostics. In total, RNAseq identified 10 RNA viruses, all of which have established pathogenicity in the lungs and five DNA viruses of questionable pulmonary pathogenicity. DNAseq identified the genomes of these five DNA viruses as well as five others including CMV and HHV-6, which have been associated with pneumonitis. 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 (40-42). mNGS captured entire viral genomes for five patients at an average read depth of 3500 fold (**Figure S1)**.

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 (43, 44). We thus evaluated alpha diversity of actively replicating microbes between subjects by SDI and found that subjects with confirmed pathogens had significantly lower alpha 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, **Figure 2** and **Table S3**). 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, **Figure 2** and **Table S3**).

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 the heterogeneity of the infectious processes (bacterial and viral) and possible host responses (anti-bacterial, anti-viral) involved, 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 (p=0.022, **Figure 3, Table S4**). When patients with confirmed or potential new pathogens were compared to those without a potential or confirmed infection, the relationship trended towards significance (94.6 (IQR 76.5-10.5, n=11) vs. 33.1 (IQR 20.7-75.1, n=7), p=0.0634), but reached significance (p = 0.014) if an individual subject with relatively lower abundance HRV-A sequences was grouped amongst those with microbes of uncertain or unlikely pathogenicity (**Figure 3, Table S4**).

DISCUSSION:

In this proof of concept study, we demonstrate that unbiased mNGS can simultaneously detect pathogens and the host response in HCT patients with acute respiratory diseases. mNGS of respiratory fluid RNA and DNA permitted identification of new potential viral and bacterial pathogens in 6 of the 15 (40%) subjects with otherwise negative conventional testing. In addition, 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 describe a novel diagnostic approach that not only identifies respiratory microbes but also simultaneously assays the host's immunologic response and airway microbiome composition to provide complementary data regarding the clinical significance of the detected agent.

When benchmarked against the conventional standard of care testing performed during the time of study enrollment, mNGS yielded 100% sensitivity for microbial detection. In addition, mNGS confirmed clinically suspected LRTI in one patient with a positive HMPV nasopharyngeal PCR but negative BAL testing. CMV was identified by both viral culture and mNGS, however this finding was considered unrelated to the patient's respiratory illness by the treating physicians. Interestingly, 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 (45, 46).

Amongst the cases with negative clinical testing, mNGS identified six patients harboring transcriptionally active microbes described as respiratory pathogens in the literature (47-52). HCOV-229E was identified in two patients and HRV-A was identified in three patients, one of whom was also culture-positive for *Stenotrophomonas maltophila.* The presence of these viruses, confirmed by independent PCR testing, was presumably missed because HRV and HCOV were not represented on the multiplex PCR panel used at the study hospital. The fact that many clinical laboratories still do not routinely test for these viruses highlights the utility of a single unbiased mNGS assay for detecting respiratory microbes. 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 permits detection of complete viral genomes (**Figure S1**), an attribute that can enable genotyping, detection of resistance mutations, and epidemiologic tracking of disease outbreaks (53).

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 the most abundant BAL microbe in one patient with idiopathic acute respiratory symptoms (52, 54). *Corynebacterium propriniquum*, one of the few virulent *Corynebacterium sp.* associated with LRTI, comprised the majority of microbial transcripts in another subject's BAL fluid (44, 47-49). 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 uncertain or unlikely 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.

Metagenomic NGS identified a number of 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 (55-57). 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). The 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 in one of these subjects, and while this virus has been associated with respiratory infections in immunocompromised patients, evidence supporting its role as a pathogen is lacking (58).

With respect to bacteria of uncertain or unlikely pathogenicity, several taxa representing common constituents of the oropharyngeal or respiratory microbiome,

including *Rothia*, *Prevotella* and *Actinomyces* and others, were identified (**Table S2)**. The majority were found at relatively low or equal abundance by both RNAseq and DNAseq, suggesting that they did not represent dominant, actively replicating pathogens but instead normal flora (**Table S2)**. *Lactobacillus* was identified in one patient with a Gram-positive rod bacteremia that was not further speciated. It is possible that mNGS may have value as a complementary method for ruling out infection or de-escalating antimicrobials in patients with negative culture or PCR-based assays and microbes of unlikely pathogenicity detected by mNGS (59).

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

Because asymptomatic carriage of respiratory pathogens is well described (60-62), 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=0.016, **Figure 2, Table S3**), presumably reflecting dominance of actively replicating pathogens (43, 44).

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 (6, 63). We found that that simultaneous assessment of host gene expression by mNGS can potentially serve as a biomarker for distinguishing infection from non-infection (e.g. colonization). Expression of a multi-gene immune response composite 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 maltophila*, 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 infection and tissue injury (60, 62). Additional prospective studies are needed to validate this metric and develop robust gene classifiers that differentiate LRTI from non-infectious lung disease in HCT recipients with acute respiratory symptoms.

By enhancing pathogen detection or by confirming negative results of conventional assays, mNGS may prove useful for reducing excessive broadspectrum antibiotic which is often driven by suspicion for occult bacterial pathogens missed by standard testing. For instance, subject 1, whose testing identified only HMPV, received empiric vancomycin and cefepime due to suspected, but 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 18 study subjects who received broad-spectrum antibiotics in the absence of detectable bacterial pathogens.

Many HCT recipients with unexplainable acute respiratory illnesses receive a diagnosis of Idiopathic Pneumonia Syndrome (IPS), a post-HCT condition that is characterized by acute lung injury and clinical features of pneumonia in the absence of a detectable pathogen (9). In our cohort, 17 (77%) subjects fulfilled IPS criteria and of these, mNGS identified possible pathogens in six (35%) cases, consistent with findings from a recent report (64).

This proof of concept study was limited by a relatively small sample size, only one case of culture-positive bacterial LRTI, no cases of fungal infection, and focused exclusively on subjects with inflammatory airway disease. 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 provide the human transcriptome coverage needed for statistically robust differential gene expression analyses, although we could rigorously evaluate a composite metric of immunity genes. Future studies that include a healthy control group without inflammatory lung disease 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 genome 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. Subsequent evaluation in a larger prospective cohort will be needed to validate these findings, define a genomic classifier for LRTI and evaluate the clinical impact of mNGS respiratory infection diagnosis in HCT recipients and other high-risk populations.

TABLE LEGENDS

Table 1: Clinical Characteristics Enrolled Patients

Clinical and microbiological data including duration and characteristics of patient respiratory symptoms. Chest CT was obtained for all subjects except 1, 3, 4, 8, 9 and 14 who received chest X-rays. Abbreviations: **Radiography**: B/L, bilateral; LUL, left upper lobe; RUL, right upper lobe. **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 score and then classified as either 1) confirmed pathogens, 2) new potential pathogens or 3) uncertain or unlikely pathogens.

Figure 2. BAL Microbial Diversity is Negatively 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), new potential pathogen (blue circles) or unlikely or uncertain pathogen (black squares). Patients with confirmed pathogens had significantly lower alpha 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 S3**.

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 of is plotted on the y-axis. Subjects are grouped according to confirmed pathogen (red triangles), new potential pathogen (blue circles) or unlikely or uncertain pathogen (black squares). Patients with confirmed pathogens had significantly higher multi-gene metric relative to patients with only microbes of unlikely pathogenicity (33.1, IQR20.7-75.1, n=7 vs. 94.9, IQR 93.8-105.6, n=6, p=0.0223). Raw data are listed in **Table S4**.

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.

(2) 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.

(3) Hutspardol S, Essa M, Richardson S, Schechter T, Ali M, Krueger J, Fujii H, Egeler RM, Gassas A. Significant transplant-related mortality from respiratory virus infections within the first 100 days in children post hematopoietic stem cell transplantation. *Biol Blood Marrow Transplant* 2015.

(4) 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.

(5) Zaas AK, Garner BH, Tsalik EL, Burke T, Woods CW, Ginsburg GS. The current epidemiology and clinical decisions surrounding acute respiratory infections. *Trends Mol Med* 2014;20:579-588.

(6) Ferrara JL, Deeg HJ. Graft-versus-host disease. *N Engl J Med* 1991;324:667-674.

(7) Panoskaltsis-Mortari A, Griese M, Madtes DK, Belperio JA, Haddad IY, Folz RJ, Cooke KR, American Thoracic Society Committee on Idiopathic Pneumonia Syndrome. An official American Thoracic Society research statement: noninfectious lung injury after hematopoietic stem cell transplantation: idiopathic pneumonia syndrome. *Am J Respir Crit Care Med* 2011;183:1262-1279.

(8) Cooke KR, Yanik G. Acute lung injury after allogeneic stem cell transplantation: is the lung a target of acute graft-versus-host disease? *Bone Marrow Transplant* 2004;34:753-765.

(9) Blair JE, Logan JL. Coccidioidomycosis in solid organ transplantation. *Clin Infect Dis* 2001;33:1536-1544.

(10) Gartner BC, Kortmann K, Schafer M, Mueller-Lantzsch N, Sester U, Kaul H, Pees H. No correlation in Epstein-Barr virus reactivation between serological parameters and viral load. *J Clin Microbiol* 2000;38:2458.

(11) Brown K, Valenta K, Fisman D, Simor A, Daneman N. Hospital ward antibiotic prescribing and the risks of Clostridium difficile infection. *JAMA Intern Med* 2015;175:626-633.

(12) Fridkin S, Baggs J, Fagan R, Magill S, Pollack LA, Malpiedi P, Slayton R, Khader K, Rubin MA, Jones M, Samore MH, Dumyati G, Dodds-Ashley E, Meek J, Yousey-Hindes K, Jernigan J, Shehab N, Herrera R, McDonald CL, Schneider A, Srinivasan A, Centers for Disease Control and Prevention (CDC). Vital signs: improving antibiotic use among hospitalized patients. *MMWR Morb Mortal Wkly Rep* 2014;63:194-200.

(13) 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.

(14) Kistler A, Avila PC, Rouskin S, Wang D, Ward T, Yagi S, Schnurr D, Ganem D, DeRisi JL, Boushey HA. Pan-viral screening of respiratory tract infections in adults with and without asthma reveals unexpected human coronavirus and human rhinovirus diversity. *J Infect Dis* 2007;196:817-825.

(15) Yozwiak NL, Skewes-Cox P, Stenglein MD, Balmaseda A, Harris E, DeRisi JL. Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* 2012;6:e1485.

(16) Chiu CY. Viral pathogen discovery. *Curr Opin Microbiol* 2013;16:468-478.

(17) Suarez NM, Bunsow E, Falsey AR, Walsh EE, Mejias A, Ramilo O. Superiority of Transcriptional Profiling Over Procalcitonin for Distinguishing Bacterial From Viral Lower Respiratory Tract Infections in Hospitalized Adults. *J Infect Dis* 2015;212:213-222.

(18) Tsalik EL, Henao R, Nichols M, Burke T, Ko ER, McClain MT, Hudson LL, Mazur A, Freeman DH, Veldman T, Langley RJ, Quackenbush EB, Glickman SW, Cairns CB, Jaehne AK, Rivers EP, Otero RM, Zaas AK, Kingsmore SF, Lucas J, Fowler VG,Jr, Carin L, Ginsburg GS, Woods CW. Host gene expression classifiers diagnose acute respiratory illness etiology. *Sci Transl Med* 2016;8:322ra11.

(19) Jain S, Self WH, Wunderink RG, Fakhran S, Balk R, Bramley AM, Reed C, Grijalva CG, Anderson EJ, Courtney DM, Chappell JD, Qi C, Hart EM, Carroll F, Trabue C, Donnelly HK, Williams DJ, Zhu Y, Arnold SR, Ampofo K, Waterer GW, Levine M, Lindstrom S, Winchell JM, Katz JM, Erdman D, Schneider E, Hicks LA, McCullers JA, Pavia AT, Edwards KM, Finelli L, CDC EPIC Study Team. Community-Acquired Pneumonia Requiring Hospitalization among U.S. Adults. *N Engl J Med* 2015;373:415-427.

(20) Musher DM, Thorner AR. Community-acquired pneumonia. *N Engl J Med* 2014;371:1619-1628.

(21) Kalil AC, Metersky ML, Klompas M, Muscedere J, Sweeney DA, Palmer LB, Napolitano LM, O'Grady NP, Bartlett JG, Carratala J, El Solh AA, Ewig S, Fey PD, File

TM,Jr, Restrepo MI, Roberts JA, Waterer GW, Cruse P, Knight SL, Brozek JL. Management of Adults With Hospital-acquired and Ventilator-associated Pneumonia: 2016 Clinical Practice Guidelines by the Infectious Diseases Society of America and the American Thoracic Society. *Clin Infect Dis* 2016;63:e61-e111.

(22) 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.

(23) 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.

(24) 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.

(25) Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357-359.

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

(27) Paulson JN, Stine OC, Bravo HC, Pop M. Differential abundance analysis for microbial marker-gene surveys. *Nat Methods* 2013;10:1200-1202.

(28) Liberzon A, Birger C, Thorvaldsdottir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst* 2015;1:417-425.

(29) McNab F, Mayer-Barber K, Sher A, Wack A, O'Garra A. Type I interferons in infectious disease. *Nat Rev Immunol* 2015;15:87-103.

(30) Jones SA. Directing transition from innate to acquired immunity: defining a role for IL-6. *J Immunol* 2005;175:3463-3468.

(31) Diez-Aguilar M, Ruiz-Garbajosa P, Fernandez-Olmos A, Guisado P, Del Campo R, Quereda C, Canton R, Meseguer MA. Non-diphtheriae Corynebacterium species: an emerging respiratory pathogen. *Eur J Clin Microbiol Infect Dis* 2013;32:769-772.

(32) Abreu NA, Nagalingam NA, Song Y, Roediger FC, Pletcher SD, Goldberg AN, Lynch SV. Sinus microbiome diversity depletion and Corynebacterium tuberculostearicum enrichment mediates rhinosinusitis. *Sci Transl Med* 2012;4:151ra124.

(33) 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.

(34) Machado CM, Dulley FL, Boas LS, Castelli JB, Macedo MC, Silva RL, Pallota R, Saboya RS, Pannuti CS. CMV pneumonia in allogeneic BMT recipients undergoing early treatment of pre-emptive ganciclovir therapy. *Bone Marrow Transplant* 2000;26:413-417.

(35) Aspin MM, Gallez-Hawkins GM, Giugni TD, Tegtmeier B, Lang DJ, Schmidt GM, Forman SJ, Zaia JA. Comparison of plasma PCR and bronchoalveolar lavage fluid culture for detection of cytomegalovirus infection in adult bone marrow transplant recipients. *J Clin Microbiol* 1994;32:2266-2269.

(36) Ison MG, Hayden FG, Kaiser L, Corey L, Boeckh M. Rhinovirus infections in hematopoietic stem cell transplant recipients with pneumonia. *Clin Infect Dis* 2003;36:1139-1143.

(37) Hakki M, Rattray RM, Press RD. The clinical impact of coronavirus infection in patients with hematologic malignancies and hematopoietic stem cell transplant recipients. *J Clin Virol* 2015;68:1-5.

(38) Bernard K. The genus corynebacterium and other medically relevant coryneform-like bacteria. *J Clin Microbiol* 2012;50:3152-3158.

(39) Motomura K, Masaki H, Terada M, Onizuka T, Shimogama S, Furumoto A, Asoh N, Watanabe K, Oishi K, Nagatake T. Three adult cases with Corynebacterium propinquum respiratory infections in a community hospital. *Kansenshogaku Zasshi* 2004;78:277-282.

(40) Furumoto A, Masaki H, Onidzuka T, Degawa S, Yamaryo T, Shimogama S, Watanabe K, Oishi K, Nagatake T. A case of community-acquired pneumonia caused by Corynebacterium propinquum. *Kansenshogaku Zasshi* 2003;77:456-460.

(41) Marrie TJ. Bacteremic community-acquired pneumonia due to viridans group streptococci. *Clin Invest Med* 1993;16:38-44.

(42) Sherry NL, Porter JL, Seemann T, Watkins A, Stinear TP, Howden BP. Outbreak investigation using high-throughput genome sequencing within a diagnostic microbiology laboratory. *J Clin Microbiol* 2013;51:1396-1401.

(43) Tunkel AR, Sepkowitz KA. Infections caused by viridans streptococci in patients with neutropenia. *Clin Infect Dis* 2002;34:1524-1529.

(44) De Vlaminck I, Khush KK, Strehl C, Kohli B, Luikart H, Neff NF, Okamoto J, Snyder TM, Cornfield DN, Nicolls MR, Weill D, Bernstein D, Valantine HA, Quake SR. Temporal response of the human virome to immunosuppression and antiviral therapy. *Cell* 2013;155:1178-1187.

(45) Focosi D, Maggi F, Albani M, Macera L, Ricci V, Gragnani S, Di Beo S, Ghimenti M, Antonelli G, Bendinelli M, Pistello M, Ceccherini-Nelli L, Petrini M. Torquetenovirus

viremia kinetics after autologous stem cell transplantation are predictable and may serve as a surrogate marker of functional immune reconstitution. *J Clin Virol* 2010;47:189-192.

(46) Babakir-Mina M, Ciccozzi M, Perno CF, Ciotti M. The human polyomaviruses KI and WU: virological background and clinical implications. *APMIS* 2013;121:746-754.

(47) Langelier C, Christenson SA. An Expression of Clinical Significance: Exploring the Human Genome to Understand the Variable Response to Rhinovirus. *Am J Respir Crit Care Med* 2016;193:710-712.

(48) Jansen RR, Wieringa J, Koekkoek SM, Visser CE, Pajkrt D, Molenkamp R, de Jong MD, Schinkel J. Frequent detection of respiratory viruses without symptoms: toward defining clinically relevant cutoff values. *J Clin Microbiol* 2011;49:2631-2636.

(49) Heinonen S, Jartti T, Garcia C, Oliva S, Smitherman C, Anguiano E, de Steenhuijsen Piters WA, Vuorinen T, Ruuskanen O, Dimo B, Suarez NM, Pascual V, Ramilo O, Mejias A. Rhinovirus Detection in Symptomatic and Asymptomatic Children: Value of Host Transcriptome Analysis. *Am J Respir Crit Care Med* 2016;193:772-782.

(50) Seo S, Renaud C, Kuypers JM, Chiu CY, Huang ML, Samayoa E, Xie H, Yu G, Fisher CE, Gooley TA, Miller S, Hackman RC, Myerson D, Sedlak RH, Kim YJ, Fukuda T, Fredricks DN, Madtes DK, Jerome KR, Boeckh M. Idiopathic pneumonia syndrome after hematopoietic cell transplantation: evidence of occult infectious etiologies. *Blood* 2015;125:3789-3797.

Figure 1

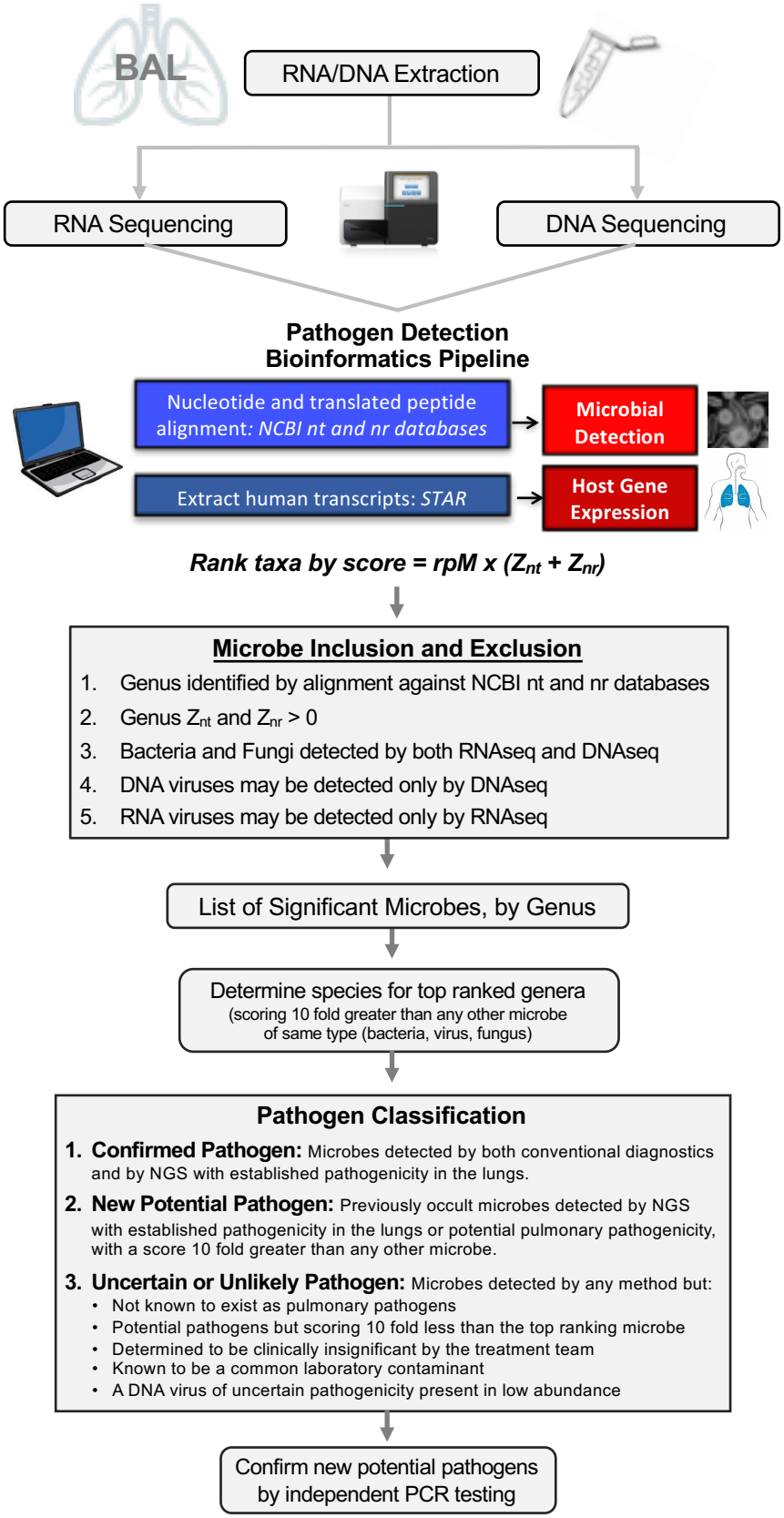
BAL
RNA/DNA Extraction
RNA Sequencing
DNA Sequencing
Pathogen Detection
Bioinformatics Pipeline
Nucleotide and translated peptide alignment: NCBI nt and nr databases
Microbial Detection
Extract human transcripts: STAR
Host Gene Expression
Rank taxa by score = rpM x (Z_{nt} + Z_{nr})
Microbe Inclusion and Exclusion
1. Genus identified by alignment against NCBI nt and nr databases
2. Genus Z_{nt} and Z_{nr} > 0
3. Bacteria and Fungi detected by both RNAseq and DNAseq
4. DNA viruses may be detected only by DNAseq
5. RNA viruses may be detected only by RNAseq
List of Significant Microbes, by Genus
Determine species for top ranked genera
(scoring 10 fold greater than any other microbe of same type (bacteria, virus, fungus)
Pathogen Classification
1. Confirmed Pathogen: Microbes detected by both conventional diagnostics and by NGS with established pathogenicity in the lungs.
2. New Potential Pathogen: Previously occult microbes detected by NGS with established pathogenicity in the lungs or potential pulmonary pathogenicity, with a score 10 fold greater than any other microbe.
3. Uncertain or Unlikely Pathogen: Microbes detected by any method but:
• Not known to exist as pulmonary pathogens
• Potential pathogens but scoring 10 fold less than the top ranking microbe
• Determined to be clinically insignificant by the treatment team
• Known to be a common laboratory contaminant
• A DNA virus of uncertain pathogenicity present in low abundance
Confirm new potential pathogens by independent PCR testing

Figure 2

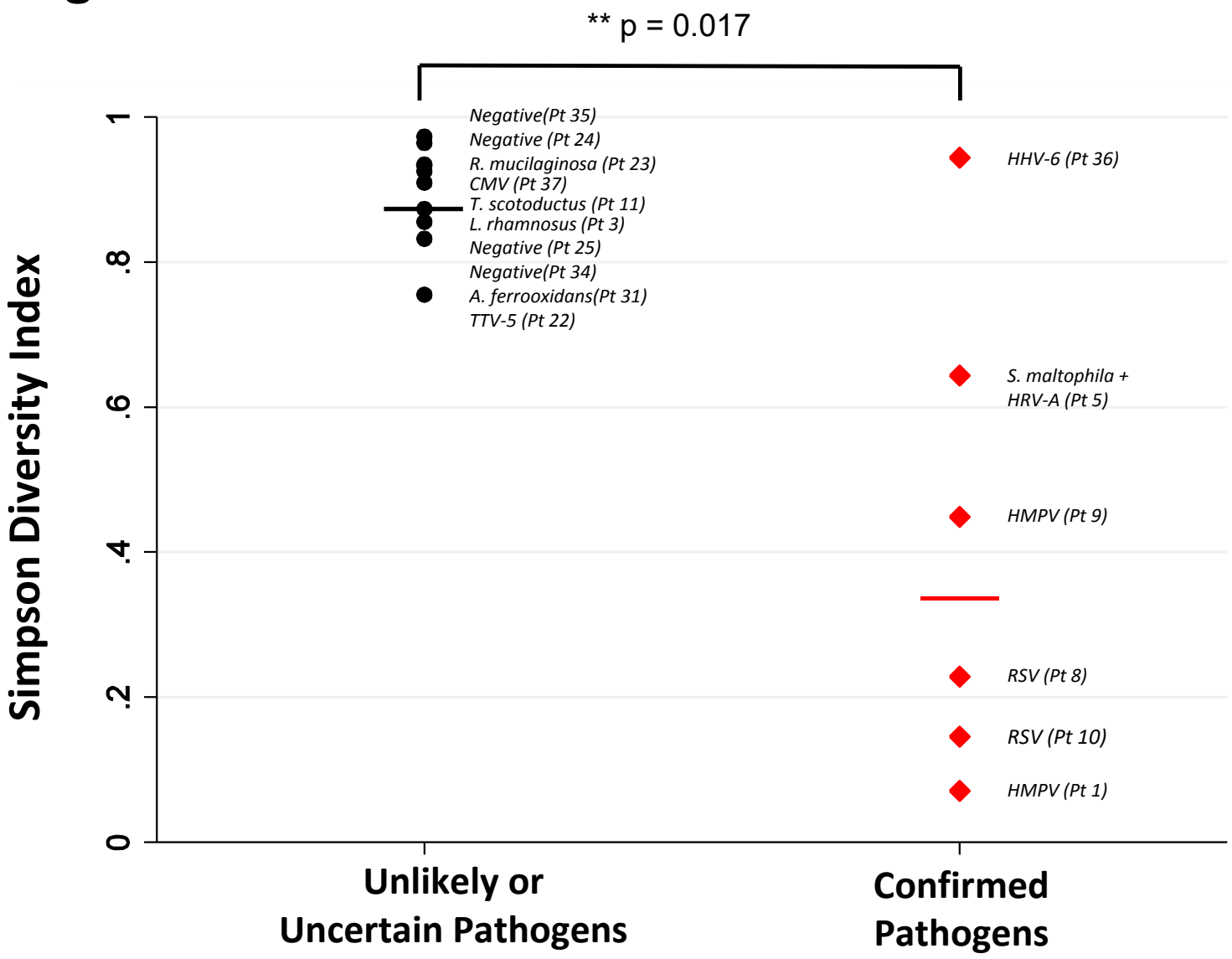
** p = 0.017
Simpson Diversity Index
0
.2
.4
.6
.8
1
Negative(Pt 35)
Negative (Pt 24)
R. mucilaginosa (Pt 23)
CMV (Pt 37)
T. scotoductus (Pt 11)
L. rhamnosus (Pt 3)
Negative (Pt 25)
Negative(Pt 34)
A. ferrooxidans(Pt 31)
TTV-5 (Pt 22)
HHV-6 (Pt 36)
S. maltophila +
HRV-A (Pt 5)
HMPV (Pt 9)
RSV (Pt 8)
RSV (Pt 10)
HMPV (Pt 1)
Unlikely or
Uncertain Pathogens
Confirmed
Pathogens

Figure 3

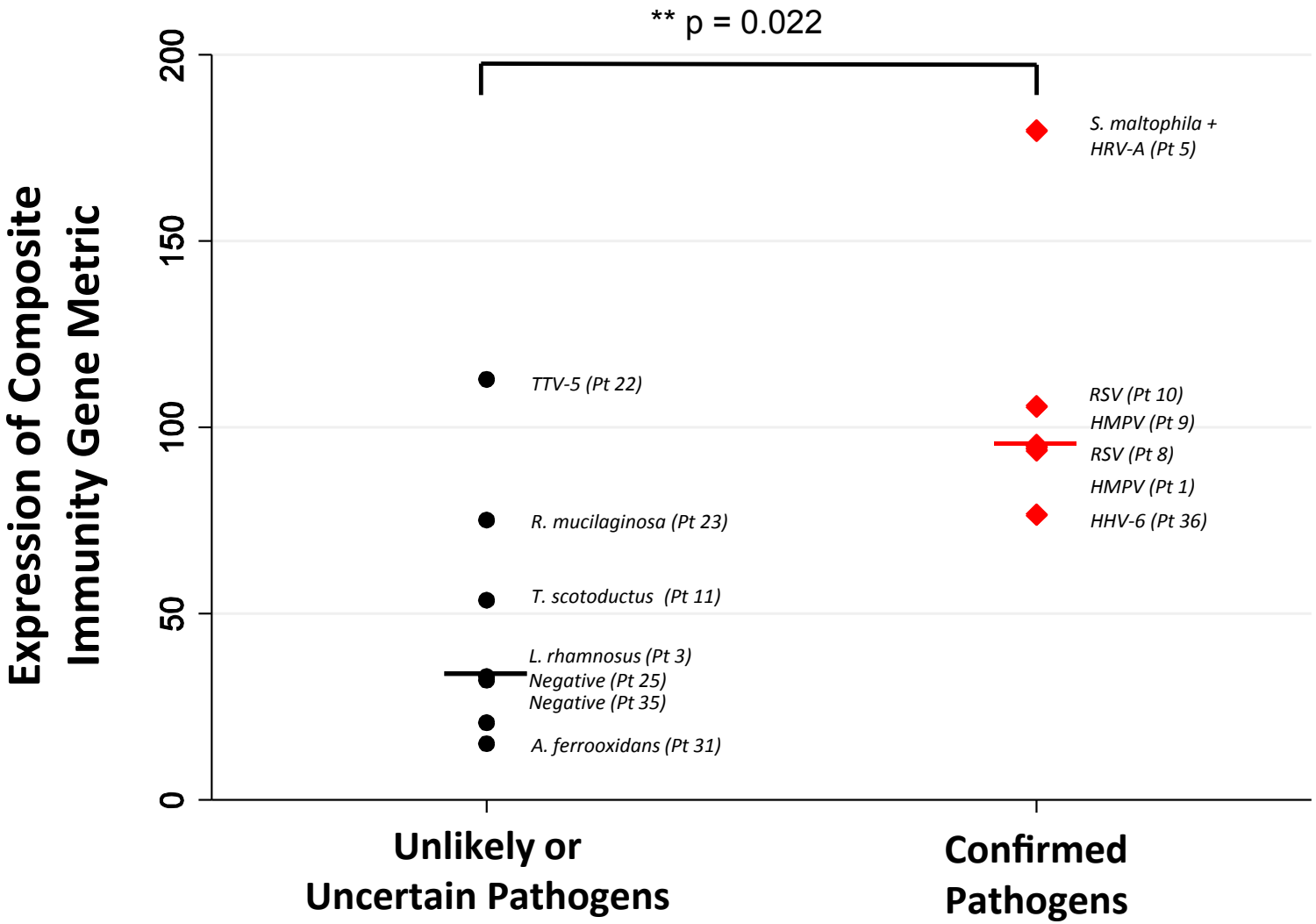
** p = 0.022
Expression of Composite Immunity Gene Metric
200
150
100
50
0
S. maltophila + HRV-A (Pt 5)
TTV-5 (Pt 22)
RSV (Pt 10)
HMPV (Pt 9)
RSV (Pt 8)
HMPV (Pt 1)
HHV-6 (Pt 36)
R. mucilaginosa (Pt 23)
T. scotoductus (Pt 11)
L. rhamnosus (Pt 3)
Negative (Pt 25)
Negative (Pt 35)
A. ferrooxidans (Pt 31)
Unlikely or Uncertain Pathogens
Confirmed Pathogens

Table 1. Demographic, Clinical, Microbiology and Diagnostic Data

ID	Age	Sex	Symptoms, Duration		Chest Radiography	Conventional BAL Microbiology	Top Scoring Microbes on mNGS	Non-BAL Microbiology	Antimicrobial Prophylaxis	Antimicrobial Treatment	GVHD	Respiratory Support	Cause of death
Pathogens identified by conventional and NGS diagnostics													
1	61	F	Fever, hypoxia	x 3 days	RLL, LLL infiltrate	*HMPV*	*HMPV*	HMPV (NP), C. *difficile* (stool)	ACV, LEVO, MICA, TMP-SMZ	CFPM, VANC	aGVHD, cGVHD	NC	Relapse, day 253
5	47	M	Cough, dyspnea, fever	x 3 days	Multifocal nodules, bronchiectasis	*Stenotrophomonas malotophila*	*Stenotrophomonas maltophila + HRV-A*	Negative	ACV, LEVO, TMP-SMZ, VORI	TMP-SMZ, VANC	aGVHD, cGVHD	NC	GVHD, day 462
8	49	M	Dyspnea	x 12 days	B/L GGO	*RSV*	*RSV + Candida glabrata*	CMV (blood)	ACV, LEVO, PENT, POSA	CIDV,FOSC, GCV, VANC, (IVIG)	No	HFNC	Relapse, day 265
9	56	M	Cough	x 3 days	RUL, RML, LUL GGO, nodules	*Negative*	*HMPV*	HMPV (NP)	ACV, TMP-SMZ	None	cGVHD	None	n/a
10	51	F	Cough, fever	x 4 days	Bibasilar airspace disease	*RSV*	*RSV*	RSV (NP)	FLUC, LEVO, vACV	CFPM, RIBV, (IVIG)	aGVHD, cGVHD	NC	GVHD, day 3024
36	59	M	Dyspnea, fever	x 1 day	Diffuse septal thickening, GGO	*HHV-6*	*HHV-6*	HHV-6 (blood)	ACV, FLUC, LEVO, VORI	FOSC	No	None	n/a
New potential pathogens identified by NGS													
6	69	M	Cough, dyspnea	x 3 days	B/L nodules	*Negative*	*HCOV-229E*	Negative	TMP-SMZ, vACV, VORI	CFPM, TOBR, VANC	cGVHD	None	n/a
7	55	F	Dyspnea	x 1 day	B/L perihilar opacities	*Negative*	*HCOV-229E*	Negative	ACV, FLUC, LEVO, TMP-SMZ	CFPM, VANC, (IVIG)	aGVHD, cGVHD	IPPV x 4 days	GVHD, day 96
13	64	M	Cough, dyspnea	x 2 days	LLL opacities, air trapping	*Negative*	*Corynebacterium propiniquum*	Negative	ACV, FLUC, LEVO, TMP-SMZ	CFPM, VANC	No	IPPV x 3 days	Pulmonary dysfunction, day 51
14	58	M	Cough, fever	x 2 days	LLL opacities	*Negative*	*HRV-A*	Negative	TMP-SMZ	None	cGVHD	None	n/a
18	69	M	Cough, fever	x 2 days	RML, RLL atelectasis	*Negative*	*HRV-A*	*S. epidermidis* (blood)	FLUC, LEVO, vACV	CFPM, VANC	n/a	None	n/a
19	58	M	Dyspnea	x 4 wks	RUL, LUL reticular opacities	*Negative*	*Streptococcus mitis*	Negative	ACV, LEVO, VORI	VANC	aGVHD, cGVHD	None	n/a
Microbes of uncertain or unlikely pathogenicity													
3	47	M	Fever	x 2 days	B/L infiltrates	*Negative*	*Lactobacillus rhamnosus*	Gram positive rod (blood)	ACV, LEVO, TMP-SMZ, VORI	VANC	No	None	GVHD, day 267
11	53	M	Cough, fever	x 2 days	LLL opacities	*Negative*	*Thermus scotoductus*	*Negative*	ACV, FLUC, LEVO, TMP-SMZ	CFPM	aGVHD, cGVHD	None	n/a
22	56	M	Dyspnea, rigors	x 1 day	RUL, LUL peribronchial thickening, GG	*Negative*	*Torque teno virus 5*	Negative	ACV, LEVO, TMP-SMZ, VORI	CFPM	cGVHD	NC	n/a
23	36	F	Cough , dyspnea, fever	x 3 wks	B/L GG, lymphadenopath y	*Negative*	*Rhodotorula mucilaginosa*	EBV viremia	ACV, LEVO, MICA, TMP-SMZ, VORI	CFPM	aGVHD	IPPV x 10 days (until death)	No Autopsy, day 10
24	67	F	Dyspnea, fever	x 1 wk	B/L perivascular and airspace GG	*Negative*	*Negative*	Negative	ACV, TMP-SMZ, VORI	CFPM, TOBR	aGVHD, cGVHD	NC	n/a
25	19	M	Cough, dyspnea, fever	x 4 wks	Diffuse bronchiectasis, multifocal GG nodules	*Negative*	*Negative*	*Negative*	LEVO, VORI, TMP-SMZ	VANC	aGVHD, cGVHD	NC	n/a
31	59	M	Dyspnea	x 2 wks	B/L GG, bronchiectasis	*Negative*	*Acidithiobacillus ferrooxidans*	Negative	ACV, FLUC, LEVO, TMP-SMZ	CFPM	cGVHD	NC	n/a
34	56	F	Cough, dyspnea	x 2 wks	Multifocal nodules, opacities, air trapping	*Negative*	*Negative*	Negative	ACV, TMP-SMZ	None	cGVHD	None	n/a
35	62	F	Cough, dyspnea	x 2 days	Multifocal GG, tree-in-bud opacities	*Negative*	*Negative*	Negative	ACV, FLUC, LEVO	VANC	aGVHD, cGVHD	NC	n/a
37	58	M	Cough, dyspnea	x 8 days	B/L GG	*CMV*	*CMV*	Negative	TMP-SMZ	None	aGVHD, cGVHD	None	n/a