Investigation of Respiratory Syncytial Virus Outbreak on an Adult Stem Cell Transplant Unit using Metagenomics Next-generation Sequencing

Yijun Zhu[a], Teresa R. Zembower[b], Kristen Metzger[c], Zhengdeng Lei[d], Stefan J. Green[d], Chao Qi[e]

Department of Pathology, University of Chicago, Chicago, IL, USA [a]; Division of Infectious Diseases, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA [b]; Infection Prevention Analytics Consultant, Northwestern Memorial Healthcare, Chicago, IL, USA [c], Research Resources Center, University of Illinois at Chicago, Chicago, IL, USA [d]; Department of Pathology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA [e]

-Running Title: RSV outbreak

Work performed at Northwestern Memorial Hospital
251 E. Huron St.
Chicago, IL 60611

Address correspondence to:

Yijun Zhu
Department of Pathology
The University of Chicago
5720 S. Drexel Avenue
Chicago, IL 60637
Yijun.Zhu@uchospitals.edu

Abstract

A viral whole genome sequencing strategy, based on PCR amplification followed by shotgun sequencing, was used to compare RSV-B genomes from 16 patients and HCWs suspected to be involved in the nosocomial RSV-B outbreak in a hematology-oncology and stem cell transplant unit with the RSV-B genomes from patients unrelated to the outbreak. Phylogenetic analysis identified a cluster of 11 patients and HCWs with an identical RSV-B strain and distinguished them from the patients unrelated to the outbreak.

Introduction

Respiratory Syncytial Virus (RSV) is well known to cause significant morbidity and mortality in pediatric populations, especially in premature or very young infants, those with chronic heart or lung disease and the immunosuppressed (1-3). In adults, the overall burden of RSV infection is similar to the prevalence of influenza A illness in elderly and high-risk adults (4). In transplant recipients, RSV is the second leading cause of respiratory virus infections (5). Progression of RSV to involve the lower respiratory tract is commonly associated with immunocompromised status. In allogeneic stem cell transplant recipients, lower respiratory tract RSV infection has been reported to have a mortality rate of up to 70% (6, 7).

Transmission of RSV occurs via direct viral inoculation of the eye and/or nose and by indirect inoculation after contact with contaminated fomites (8). Outbreaks of RSV have occurred in a variety of health care settings, including infant wards, adult hematology and transplant units, and outpatient cancer centers (9, 10). Infected visitors or health care workers, and patients with prolonged viral shedding can serve as a source for an outbreak. A recent study showed that prolonged viral shedding over 30 days in patients with hematological disorders was mostly commonly associated with RSV compared with other respiratory viral pathogens (11).

Timely identification of the outbreak source is critical to allow implementation of infection control measures. Whole genome sequencing (WGS) with next generation sequencing technology has been increasingly applied to assess the epidemiological link between bacteria implicated in outbreaks (12-14). In the studies with bacterial pathogens, the targeted pathogen was first recovered by culture, and then whole genome sequencing was performed with the nucleic acid extracted from the pure isolates. Using WGS for investigation of viral molecular

epidemiology is more difficult due to less frequent use of viral culture for virus detection. A few studies reported analysis of genetic diversity of global and local strains of Influenza and RSV with WGS technology (15, 16). Only one recent study reported an investigation of hospital transmission of human parainfluenza virus 3 on a general medicine unit (17). In this study, we characterized a RSV outbreak in an adult stem cell transplant unit with WGS using the samples collected for respiratory virus PCR. The sequence data were used to identify genetic variation among patient- and healthcare worker-associated strains, and in turn to define transmission pathways.

Material and Methods

Setting - Northwestern Memorial Hospital (NMH) is an 894-bed tertiary care academic medical center in Chicago, Illinois. The hematology/oncology and stem cell transplant unit consists of 36 single-occupancy rooms. The majority of patients in the unit are pre- or post-hematopoietic stem cell transplantation (HSCT). These include patients undergoing conditioning for HSCT, receiving HSCT, and undergoing monitoring during the pre-engraftment period. Patients with HSCT-associated conditions, such as graft versus host disease, as well as patients with hematologic malignancies undergoing induction or consolidation chemotherapy before HSCT, are also frequently admitted to this unit.

Outbreak and sample collection - The outbreak was detected on March 5, 2015, when an infection preventionist (IP) recognized a cluster of 3 patients who tested positive for RSV-B in the HSCT unit. Upon further investigation, 11 additional patients with a diagnosis of RSV-B infection were identified during or shortly after admission to the units from February 23 to April 22, 2015. Screening of all asymptomatic health care workers (HCWs) and unit staff for RSV by

nasopharyngeal respiratory virus PCR identified five HCWs carrying RSV-B; thus 19 individuals were implicated in this RSV outbreak.

Nucleic acid was extracted from 200 µl of viral transport medium containing nasopharyngeal samples from the 14 RSV-B positive patients and five HCWs submitted for routine clinical care for respiratory virus PCR using Qiangen Symphony automated extraction system with the QIAsymphony DSP Virus/pathogen kit (Qiagen, Inc. Hilden, Germany).

Library construction and DNA sequencing - Twenty-five overlapping pairs of primers (Table 1) were designed to amplify 600-700 bp amplicons based on the complete genome sequence of the Human respiratory syncytial virus wildtype strain B1 (Genbank accession number, AF013254.1). Reverse transcription (RT) - PCR was performed with SuperScript III RT-PCR system containing Platinum *Taq* DNA polymerase with random primer (Invitrogen). Each fragment was amplified with RSV specific primers with the condition 94°C for 3 min, 30 cycles of 94°C for 30s, 55°C for 30s, and 68°C for 30s. The 25 PCR products from each sample were quantified and pooled together for next-generation sequencing. Pooled amplicons were prepared for sequencing using the Nextera XT DNA library preparation kit (Illumina), according to the manufacturer's instruction. Barcoded libraries were pooled together, and sequenced using an Illumina MiSeq sequencer, employing V2 chemistry with paired-end 2x250 base reads. Sequencing was performed at the Research Resources Center at University of Illinois at Chicago (UIC). Demultiplexing of the sequence data was performed on instrument. Data have been deposited at NCBI's Sequence Read Archive (SRA) under the BioProject accession number PRJNA371804.

Data analysis - Sequence data were processed with the software package SPANDx (Sarovich et al. 2014), using the RSV genome (AF013254) as a reference (18). SPANDx is an open-source, high-throughput, comparative genomic analysis tool for haploid organisms. It incorporates Burrows-Wheeler Aligner (BWA) for read alignment mapping, SAMtools for read filtering and parsing, BEDTools for genetic locus presence/absence (P/A) determination, Picard (http://picard.sourceforge.net) for data filtering, the Genome Analysis Tool Kit (GATK) for realignment around insertion-deletion (indel) regions, base quality score recalibration, variant determination, data filtering and improved insertion-deletion calling, VCFtools for single-nucleotide polymorphism (SNP) and indel matrix construction, and SnpEff for variant annotation (19-24). Coverage of the reference genome by the reads from the samples ranged from 97% to 99%. SPANDx analysis was performed in UIC's Center for Research Informatics (CRI).

The final output from the SPANDx pipeline was used for phylogenetic analyses, based on genomic positions which varied in at least one of the 24 genomes compared. A total of 244 positions were identified and used for phylogeny, employing neighbor-joining (NJ), maximum likelihood and Bayesian analyses. A neighbor-joining phylogenetic tree was constructed with aligned sequences; the scale of analysis is the number of base differences per sequence (out of 244 positions). In addition, an initial tree for the heuristic search was obtained automatically by applying NJ and BioNJ algorithms to a matrix of pairwise distances estimated using the Maximum Composite Likelihood (MCL) approach, and then selecting the topology with superior log likelihood value. The robustness of both NJ and MCL inferred tree topologies was evaluated by 1000 bootstrap re-samplings of the data. The phylogenetic trees were compared, and nodes

that were not supported by bootstrap values of 70 % or higher for at least one of the methods were treated as polytomies. In addition, Bayesian analyses were performed on the aligned sequence data by running five simultaneous chains (four heated, one cold) for one million generations, sampling every 1000 generations. The selected model was the general time reversible (GTR) using empirical base frequencies and estimating the shape of the gamma distribution and proportion of invariant sites from the data. A resulting 50 % majority-rule consensus tree (after discarding the burn-in of 25 % of the generations) was determined to calculate the posterior probabilities for each node. In all cases, the split-differential at 1 million generations was below 0.01. NJ and MCL phylogenetic analyses were performed using the software package MEGA6 (25), and Bayesian analyses were performed using the software package MrBayes v3.1.2 (26).

Results and Discussion

Genome phylogeny analysis – Sixteen of 19 nasopharyngeal samples generated sufficient RSV PCR product for sequencing. Eleven samples were from patients (R1, R2, R5, R8, R10, R11, R12, R13, R14, R15, R16) and five samples (R3, R4, R6, R7, R9) were collected from healthcare works (HCWs) who were in contact with the RSV positive patients during the outbreak. A viral WGS strategy, based on PCR amplification followed by shotgun sequencing, was used to determine the phylogenetic relatedness of the 16 RSV strains from the suspected outbreak cluster and 8 strains (RC1 to RC8) amplified from outpatient samples.

Phylogenetic analysis identified a cluster of 11 RSV strains with identical genome (Fig. 1). These strains were recovered from eight patients (R1, R2, R5, R8, R10, R12, R14, R15) and three

HCWs (R4, R6, R9). The rest of patients, including R11, R13, R16, and two HCWs (R3 and R7) involved in the outbreak were infected with unrelated strains. As expected, the eight RSV strains (RC1 to RC8) recovered from the outpatients were diverse and were different from both the outbreak strain and the other strains circulating on the stem cell unit.

Quantitatively differentiate non-outbreak strains - A cross-tabulation of SNPs comparing RSV from the 24 samples is shown in figure 2. No SNP differences were observed among the strains shared by patients and HCWs involved in the outbreak, suggesting a common source outbreak and person-to person transmission of one RSV strain among the patients and HCWs. R3, R7, R11, R13, and R16 were different from the potential transmission cluster with the dissimilarity ranging from 10 to 11 SNPs. Strains carried by the non-outbreak patients were distinct from the strain shared by the transmission group with sequence dissimilarity ranging from 9 to 43 SNPs.

Epidemiologic investigation – Description of the outbreak and patient characteristics has been described previously (27). Spot map of the 36-bed stem cell transplant unit with patients infected with RSV is showed in figure 3. All eight patients who shared the outbreak strain, including R1, R2, R5, R8, R10, R12, R14, and R15, were from the unit's west wing. The three patients with RSV strains distinct from the outbreak strain, including R11, R13, and R16, were from the unit's east wing. While three HCWs involved in the outbreak carried the outbreak strain R5, R6, R9, the other two HCWs had unrelated non-outbreak strains. The data supports that this outbreak included eight patients from west wing of the unit and three HCWs. The other three patients from the east wing of the unite and two HCW implicated in the outbreak were not part of the outbreak (figure 3).

Traditionally RSV genotyping was focused on analysis of complete or partial sequence of the attachment glycoprotein (G) (28). G-protein mediates virus binding to cells. During natural propagation of RSV in infected patients, sequence changes accumulate quickly especially in the C-terminal, second hypervariable region of the protein (29). Complete or partial G gene sequences are commonly utilized to distinguish the two RSV groups (A and B), and the various genotypes within each group. G gene sequence-based RSV genotyping has been widely used for global epidemiological study of RSV. Evidence indicates that genetic diversity based the entire RSV genome is necessary for investigation of transmission of RSV strains over short periods of time. Agti, et al. compared RSV genomes with identical G genes and suggested that genotyping based on the whole genome distinguished the RSV strains with identical G genes and increased the sensitivity for tracking RSV transmission over a short period of time (15).

In our study, comparison of RSV-B genomes from 16 patients and HCWs suspected to be involved in the nosocomial RSV-B outbreak in a hematology-oncology and stem cell transplant unit with the RSV-B genomes from patients unrelated to the outbreak clearly identified a cluster of 11 patients and HCWs with an identical RSV-B strain and distinguished them from the patients unrelated to the outbreak. Investigation of patient geographic location provided additional evidence to support the genetic relatedness of RSV-B genome revealed by the WGS typing, demonstrating the existence of a common cause for the outbreak.

The information of patient contact history of each HCWs during the outbreak was not available. The five HCWs implicated in the outbreak included a social work, Stem Cell Coordinator, a food service worker, Palliative Care physician, and a nurse. They likely worked on both west and east sides of the unite. In this outbreak three HCWs were infected with the

outbreak strain. Although index personnel for the outbreak was difficult to determine, the HCWs were important source for pathogen spread. Most hospitals have policy to keep HCWs with acute respiratory symptoms away from direct patient contact. This practice is not sufficient to prevent transmission of pathogen from asymptomatic carriers. The practical process to reduce the risk of transmission from a colonized personnel in a high risk patient population requires further investigation.

The purpose of the study is to determine whether WGS would be able to separate cases of RSV-B transmission from the patients with strains unrelated to the transmission in an outbreak over eight weeks. We showed that WGS was valuable for a local outbreak investigation. Identifying transmission and defining outbreak boundaries is critical information that allows implementation of appropriate infection control and prevention measures. Delay in detection of patients and HCWs involved in transmission due to lack of symptoms often results in propagation of the outbreak. In this outbreak, RSV-B transmission occurred in four patients in spite of the stringent infection control measures (?). WGS may offer a valuable solution to this problem, with its high throughput capacity and ability to screen and perform a quick genome typing analysis for a large number of patients and HCWs once a small cluster of patients appears.

References

1. **Griffiths C, Drews SJ, Marchant DJ.** 2017. Respiratory Syncytial Virus: Infection, Detection, and New Options for Prevention and Treatment. Clin Microbiol Rev **30:**277-319.

2. **Nair H, Nokes DJ, Gessner BD, Dherani M, Madhi SA, Singleton RJ, O'Brien KL, Roca A, Wright PF, Bruce N, Chandran A, Theodoratou E, Sutanto A, Sedyaningsih ER, Ngama M, Munywoki**

PK, Kartasasmita C, Simoes EA, Rudan I, Weber MW, Campbell H. 2010. Global burden of acute lower respiratory infections due to respiratory syncytial virus in young children: a systematic review and meta-analysis. Lancet **375:**1545-1555.

3. **Schobel SA, Stucker KM, Moore ML, Anderson LJ, Larkin EK, Shankar J, Bera J, Puri V, Shilts MH, Rosas-Salazar C, Halpin RA, Fedorova N, Shrivastava S, Stockwell TB, Peebles RS, Hartert TV, Das SR.** 2016. Respiratory Syncytial Virus whole-genome sequencing identifies convergent evolution of sequence duplication in the C-terminus of the G gene. Sci Rep **6:**26311.

4. **Falsey AR.** 2005. Respiratory syncytial virus infection in elderly and high-risk adults. Exp Lung Res **31 Suppl 1:**77.

5. **Couch RB, Englund JA, Whimbey E.** 1997. Respiratory viral infections in immunocompetent and immunocompromised persons. Am J Med **102:**2-9; discussion 25-26.

6. **Avetisyan G, Mattsson J, Sparrelid E, Ljungman P.** 2009. Respiratory syncytial virus infection in recipients of allogeneic stem-cell transplantation: a retrospective study of the incidence, clinical features, and outcome. Transplantation **88:**1222-1226.

7. **Shah JN, Chemaly RF.** 2011. Management of RSV infections in adult recipients of hematopoietic stem cell transplantation. Blood **117:**2755-2763.

8. **Dare RK, Talbot TR.** 2016. Health Care-Acquired Viral Respiratory Diseases. Infect Dis Clin North Am **30:**1053-1070.

9. **Chu HY, Englund JA, Podczervinski S, Kuypers J, Campbell AP, Boeckh M, Pergam SA, Casper C.** 2014. Nosocomial transmission of respiratory syncytial virus in an outpatient cancer center. Biol Blood Marrow Transplant **20:**844-851.

10. **Kassis C, Champlin RE, Hachem RY, Hosing C, Tarrand JJ, Perego CA, Neumann JL, Raad, II, Chemaly RF.** 2010. Detection and control of a nosocomial respiratory syncytial virus outbreak in

a stem cell transplantation unit: the role of palivizumab. Biol Blood Marrow Transplant **16:**1265-1271.

11. **Lehners N, Tabatabai J, Prifert C, Wedde M, Puthenparambil J, Weissbrich B, Biere B, Schweiger B, Egerer G, Schnitzler P.** 2016. Long-Term Shedding of Influenza Virus, Parainfluenza Virus, Respiratory Syncytial Virus and Nosocomial Epidemiology in Patients with Hematological Disorders. PLoS One **11:**e0148258.

12. **Dominguez SR, Anderson LJ, Kotter CV, Littlehorn CA, Arms LE, Dowell E, Todd JK, Frank DN.** 2016. Comparison of Whole-Genome Sequencing and Molecular-Epidemiological Techniques for Clostridium difficile Strain Typing. J Pediatric Infect Dis Soc **5:**329-332.

13. **Koser CU, Holden MT, Ellington MJ, Cartwright EJ, Brown NM, Ogilvy-Stuart AL, Hsu LY, Chewapreecha C, Croucher NJ, Harris SR, Sanders M, Enright MC, Dougan G, Bentley SD, Parkhill J, Fraser LJ, Betley JR, Schulz-Trieglaff OB, Smith GP, Peacock SJ.** 2012. Rapid whole-genome sequencing for investigation of a neonatal MRSA outbreak. N Engl J Med **366:**2267-2275.

14. **Rusconi B, Sanjar F, Koenig SS, Mammel MK, Tarr PI, Eppinger M.** 2016. Whole Genome Sequencing for Genomics-Guided Investigations of Escherichia coli O157:H7 Outbreaks. Front Microbiol **7:**985.

15. **Agoti CN, Otieno JR, Munywoki PK, Mwihuri AG, Cane PA, Nokes DJ, Kellam P, Cotten M.** 2015. Local evolutionary patterns of human respiratory syncytial virus derived from whole-genome sequencing. J Virol **89:**3444-3454.

16. **Bose ME, He J, Shrivastava S, Nelson MI, Bera J, Halpin RA, Town CD, Lorenzi HA, Noyola DE, Falcone V, Gerna G, De Beenhouwer H, Videla C, Kok T, Venter M, Williams JV, Henrickson KJ.** 2015. Sequencing and analysis of globally obtained human respiratory syncytial virus A and B genomes. PLoS One **10:**e0120098.

17. **Greninger AL, Zerr DM, Qin X, Adler AL, Sampoleo R, Kuypers JM, Englund JA, Jerome KR.** 2017. Rapid Metagenomic Next-Generation Sequencing during an Investigation of Hospital-Acquired Human Parainfluenza Virus 3 Infections. J Clin Microbiol **55:**177-182.

18. **Sarovich, D.S. and Price, E.P.** 2014. SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. *BMC research notes*, **7***:* 618-629

19. **Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R, Genomes Project Analysis G.** 2011. The variant call format and VCFtools. Bioinformatics **27:**2156-2158.

20. **DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ.** 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet **43:**491-498.

21. **Li H, Durbin R.** 2010. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics **26:**589-595.

22. **Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S.** 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics **25:**2078-2079.

23. **McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, DePristo MA.** 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res **20:**1297-1303.

24. **Quinlan AR, Hall IM.** 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics **26:**841-842.

25. **Tamura K, Stecher G, Peterson D, Filipski A, Kumar S.** 2013. MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. Mol Biol Evol **30:**2725-2729.

26. **Ronquist F, Huelsenbeck JP.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics **19:**1572-1574.

27. **Kelly SG, Metzger K, Bolon MK, Silkaitis C, Mielnicki M, Cullen J, Rooney M, Blanke T, Tahboub A, Noskin GA, Zembower TR.** 2016. Respiratory syncytial virus outbreak on an adult stem cell transplant unit. Am J Infect Control **44:**1022-1026.

28. **Katzov-Eckert H, Botosso VF, Neto EA, Zanotto PM, consortium V.** 2012. Phylodynamics and dispersal of HRSV entails its permanence in the general population in between yearly outbreaks in children. PLoS One **7:**e41953.

29. **Sullender WM, Mufson MA, Anderson LJ, Wertz GW.** 1991. Genetic diversity of the attachment protein of subgroup B respiratory syncytial viruses. J Virol **65:**5425-5434.

Legends

Figure 1. **Phylogenetic tree reflecting the relationships of strains based on whole genome re-sequencing and analysis.** The tree topology was obtained from a boot-strapped neighbor-joining analysis. Nodes for which bootstrap values equaled or exceed 70 % are indicated by a numerical value. The bootstrap values derived from maximum likelihood analysis is also indicated (NJ/ML). Polytomies indicate branching points that were not consistently supported by bootstrap analyses. Nodes supported by Bayesian analysis, with posterior probability values greater than 95 %, are indicated with black circles. The node indicated by the arrow represents 11 identical genomes based on mapping to the reference genome (GenBank AF013254.1).

Figure 2. Sequence mismatches between pairs of isolate genomes. R1 to R6 represent RSV B genomes from the patients and HCWs implicated in the outbreak. RC1 to RC8 represent RSV B genomes from outpatients. Values represent counts of single nucleotide polymorphisms between pairs of RSV B genomes.