ORIGINAL ARTICLE

Rule-Out Outbreak: 24-Hour Metagenomic Next-Generation Sequencing for Characterizing Respiratory Virus Source for Infection Prevention

Alexander L. Greninger,[1,2] Alpana Waghmare,[2,3] Amanda Adler,[3] Xuan Qin,[1] Janet L. Crowley,[1] Janet A. Englund,[3] Jane M. Kuypers,[1] Keith R. Jerome,[1,2] and Danielle M. Zerr[2,3]

[1]Department of Laboratory Medicine, University of Washington, Seattle; [2]Vaccine and Infectious Disease Division, Fred Hutchinson Cancer Research Center, Seattle, Washington; and [3]Seattle Children's Research Institute, Washington

Background. Metagenomic next-generation sequencing (mNGS) has been used to uncover unusual causes of infectious diseases but has not been used routinely for the investigation of putative nosocomial outbreaks. Here, we describe the use of mNGS during investigation of a cluster of human rhinovirus (HRV)-positive infections on a high-risk pulmonary ward.

Methods. We performed mNGS on 6 midnasal turbinate swabs from 4 case-patients and 10 swabs from 9 control outpatients that tested positive for enterovirus/rhinovirus by the FilmArray system.

Results. HRV reads were recovered in 15 (94%) of the 16 samples sequenced. Phylogenetic analysis of HRV whole genomes from the 4 case-patients and 5 outpatient controls along with partial genomes from additional outpatient controls revealed that isolates from the case-patients were not directly related and that the 2 closest case HRV genomes had an estimated time to most recent common ancestor of 172 years. Our turnaround time from receipt of the sample to phylogenetic analysis was 24 hours.

Conclusions. We found the use of mNGS downstream of a rapid polymerase chain reaction respiratory panel during an investigation of 4 hospital-acquired rhinovirus infections to rapidly dispel concern of a single-source transmission event.

Keywords. HPIV3; human parainfluenza 3 virus; mNGS; next-generation sequencing; respiratory virus; rule out outbreak.

Respiratory illness is a major burden throughout the world, and rhinoviruses are the most common pathogen detected in cases of adult community-acquired pneumonia and pediatric respiratory disease [1, 2]. Most diagnostic respiratory polymerase chain reaction (PCR) panels fail to discern beyond the genus level for enteroviruses and rhinoviruses, which leaves the possibility that genotype-specific clinical phenomena are missed. Metagenomic next-generation sequencing (mNGS) of clinical samples is a promising approach for rapidly recovering pathogen sequences both for detection and epidemiological determination of transmission [3, 4].

Here, we describe mNGS of enterovirus/rhinovirus samples downstream of a rapid PCR respiratory panel to dispel concern of a possible single-source transmission event. Convenience sampling of existing specimens from outpatients who tested positive for rhinovirus by a rapid PCR panel provided a ready-made set of controls. This approach requires approximately 24 hours from sampling to receipt of results to ascertain whether PCR panel–positive isolates are linked.

OUTBREAK DESCRIPTION

Seattle Children's Hospital is a 371-bed quaternary care pediatric facility located in Seattle, Washington, that serves patients aged 0 to 21 years with a wide variety of acute and chronic health issues. Respiratory virus testing on midnasal turbinate swabs is performed routinely for symptomatic patients using the FilmArray assay (BioFire Diagnostics, Salt Lake City, UT).

Routine infection prevention surveillance identified 4 cases of hospital-acquired rhinovirus infection on a general medical unit over a 3-week period (Figure 1A). Hospital-acquired enterovirus/rhinovirus was defined for patients who tested positive for enterovirus/rhinovirus and developed clinical symptoms >3 days after initial hospital admission. Respiratory virus testing was performed at the discretion of the medical teams caring for the patients, and in all cases, testing was prompted by increased work of breathing and/or increased respiratory secretions.

Infection prevention and the unit leadership performed an investigation of the hospital-acquired cases in this high-risk
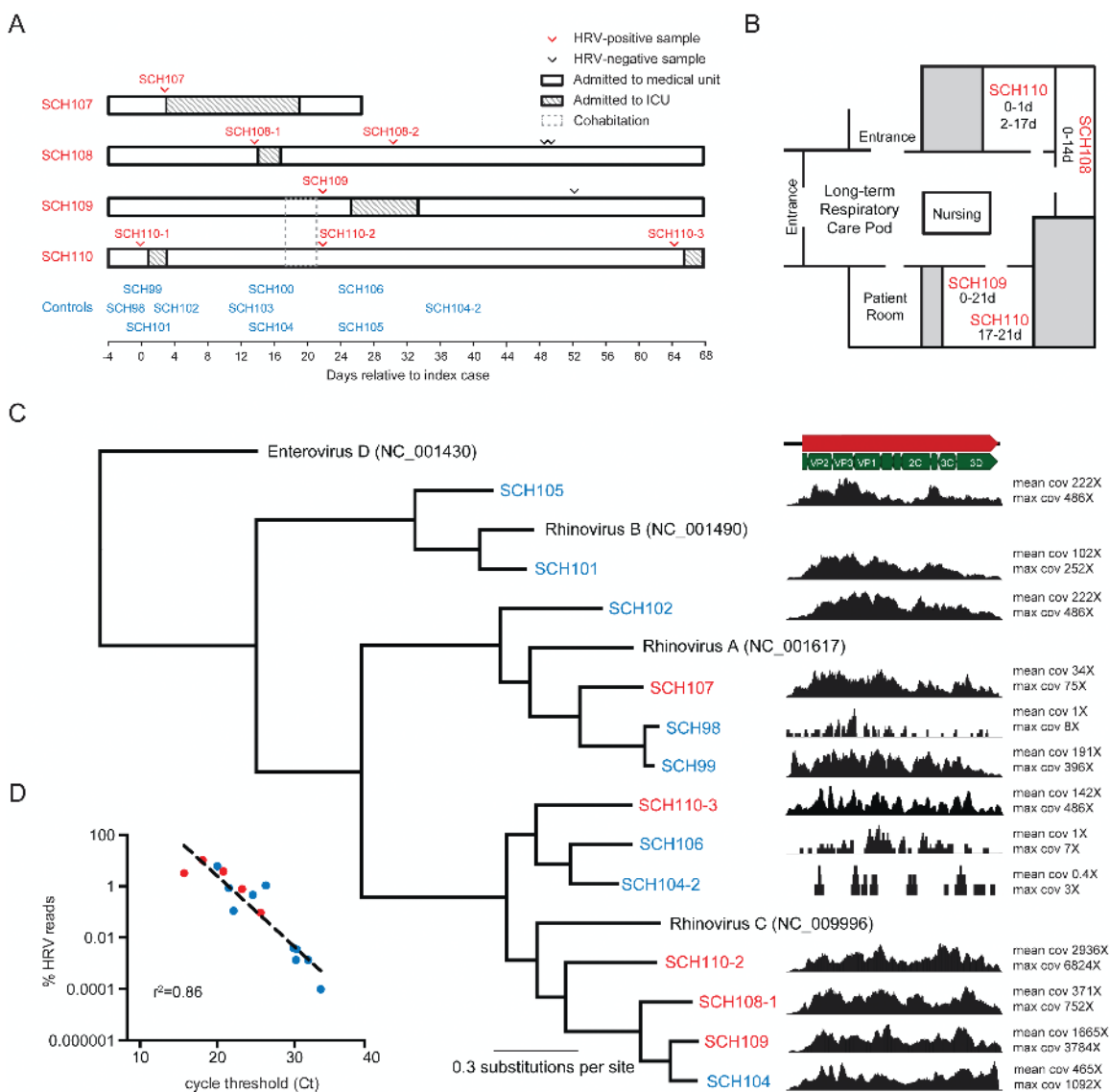
Figure 1. Phylogenetic analysis reveals that the rhinovirus strains are not part of a recent transmission chain. (A) Four long-term inpatients admitted to the intensive care unit (ICU) in associated with human rhinovirus (HRV) positivity within a 3-week period were investigated along with 9 community controls who tested positive for HRV during the same time period. Case-patients are highlighted in red, and community controls are highlighted in blue. Sample SCH110-1 tested positive for HRV by a FilmArray panel but was not available for sequencing. Symptom-onset dates were the same as the testing dates. (B) Location of case-patients SCH108, SCH109, and SCH110, who were colocated on the same long-term respiratory care pod within the medical unit over 2 weeks, which led to concern for a single-source transmission event. Patients SCH109 and SCH110 shared the same room for 4 days before each of them tested positive for HRV. (C) Bayesian phylogenetic tree and normalized coverage maps of partial and complete rhinovirus genomes recovered in this study along with mean and maximum coverage (cov) for each rhinovirus. Enterovirus D was used as an outgroup, and each HRV type species is listed with its NCBI accession number. Posterior probabilities for each branch were 100% and are not depicted. Each hospital-acquired infection strain was more than 1500 changes away from each other, which is consistent with an estimated 172 years of genetic distance between the 2 closest related viruses. (D) Inverse relationship between percent HRV reads recovered in a sequencing library versus quantitative reverse transcription polymerase chain reaction cycle threshold.

pulmonary ward to identify ill staff or family members and other risk factors that might have increased the risk of enterovirus/rhinovirus transmission to patients. Patients SCH108, SCH109, and SCH110 were located in 1 pod of the medical unit at the time of their initial positive test result, and patient SCH107 was located in another pod on the same unit (Figure 1B). Patient SCH110 tested positive for human rhinovirus (HRV) on day 0 of the cluster and 17 to 21 days later shared a room with patient SCH109. Patient

SCH110's original respiratory symptoms had resolved before sharing the room with patient SCH109. On day 22 of the cluster, patient SCH109's first test was positive and patient SCH110's second test was positive. One parent and several healthcare workers were noted to have mild respiratory symptoms and wore masks when having contact with patients on the unit.

To further investigate the cluster, mNGS was performed on case and control samples. Control samples were

identified from inpatient and outpatient children with community- acquired enterovirus/rhinovirus infection who tested positive in the same month as the cluster. All available respiratory samples obtained from 4 case-patients with hospital-acquired HRV and 9 control patients were sent to the University of Washington Virology Laboratory for mNGS. These samples included the initial positive samples from 3 case-patients (the initial positive sample from patient SCH110 was not available) and 9 controls and 3 subsequent positive samples obtained from 2 case-patients (SCH108-2, SCH110-2, and SCH110-3) and 1 subsequent positive sample from a control (SCH104-2) (Figure 1A). Information on patient demographics, room locations, symptoms, and outcomes was also collected.

mNGS AND QUANTITATIVE REVERSE-TRANSCRIPTION PCR

mNGS libraries were constructed as described previously [3]. Viral transport medium (500 µL) from middle turbinate respiratory swabs was spun in a 0.45-µm filter, and RNA was extracted using a Zymo viral RNA kit. RNA was DNase treated, and double-stranded complementary DNA was constructed using random hexamers, SuperScript III reverse transcriptase (ThermoFisher), and Sequenase 2.0 DNA polymerase (Affymetrix) and subjected to Nextera XT tagmentation for Illumina sequencing library generation. Sequencing libraries were run on a single 1 × 185-bp MiSeq system to achieve approximately 2 million reads per sample, and no two libraries shared either of their dual-indexed barcodes. The depth of sequencing was chosen on the basis of previous experience for rapid metagenomic sequencing and the need to multiplex samples into 1 run of the Illumina MiSeq system [3]. Sequencing reads were aligned to a concatenated reference genome containing enteroviruses A through D and rhinoviruses A through C and iteratively assembled on the basis of mapped reads into complete genomes in Geneious 9.1. Sequences were deposited in GenBank under accession numbers KY189313 through KY189321. Phylogenetic analyses were performed in MrBayes using default parameters and BEAST using 11 000 000 generations with a relaxed clock with log-normal distribution, 3-codon partitioning with a Bayesian SkyGrid tree prior. We included rhinovirus genomes in the BEAST phylogenetic analysis that comprised the 8 top BLASTn hits with full genomes and collection dates available for each of the viruses sequenced in this study. To correlate rhinovirus reads versus rhinovirus copy numbers (estimated by the PCR cycle threshold value), double-stranded complementary DNA was tested by a real-time reverse transcription PCR (RT-PCR) assay using primers and a probe targeting the rhinovirus 5′ untranslated region [5].

RESULTS

The 4 case-patients ranged in age from 3 months to 7 years (median, 0.96 years). All of them had an underlying medical condition that required prolonged hospitalization, and all of them were admitted to the hospital for at least 1 month before their first positive enterovirus/rhinovirus test result (median, 126 days). Three case-patients had a history of chronic lung disease, and 1 had a history of congenital nasal cavity malformation. All 4 case-patients were transferred to the intensive care unit as a result of respiratory distress after initial detection of enterovirus/rhinovirus (Figure 1A). Three case-patients continued to test positive for rhinovirus/enterovirus after their initial positive test (median duration of positivity, 39 days [range, 16–64 days]), although repeat testing was not performed in a standardized manner. Acute respiratory symptoms ultimately resolved in all the patients.

Control patients ranged in age from 1 month to 13 years (median, 5.85 years). Two patients were admitted to the general medical floor; both of them had a chronic medical condition (cystic fibrosis and sick cell anemia). Neither the case-patients nor the controls had a history of travel.

The initial rhinovirus/enterovirus-positive sample from 1 case-patient (SCH110) was not available for sequencing; sequencing was performed on the positive samples obtained at days 22 and 64. Samples with 10 or fewer HRV reads were not included in the phylogenetic analysis (Figure 1C), including 2 controls, SCH100 and SCH103, and 1 repeat case sample SCH108-2. A median of 1 808 370 (range, 434 462–3 139 792) adapter/quality-trimmed reads were recovered per sample, and a median of 4252 (range, 0–214 464) reads aligned to HRV (Table 1). The percent rhinovirus reads recovered was inversely proportional to the rhinovirus quantitative RT-PCR cycle threshold (Ct) for each sample (Figure 1D). A total of 10 complete HRV genomes were assembled from samples from all 4 case-patients, 1 repeat enterovirus/rhinovirus-positive sample from a case-patient, and 5 control samples. In addition, 3 partial sequences from outpatient controls, including 1 repeat enterovirus/rhinovirus positive sample (SCH104-2), were included in the phylogenetic analysis (Figure 1A).

In the samples from the 4 case-patients, 3 HRV-C and 1 HRV-A sequence were found, whereas in the initial samples from the controls, 2 HRV-C, 2 HRV-B, and 5 HRV-A sequences were found. Repeat sampling from 1 case-patient and 1 control (SCH110-3 and SCH104-2, respectively) revealed that they both were HRV-C members, as did the preceding positive samples recovered from each patient, but aligned only 66% to 68% by nucleotide, indicating new infection with HRV-C. The clinical findings associated with SCH110-3 were consistent with the acquisition of a new HRV strain because it was obtained at the onset of new respiratory symptoms and readmission to

Table 1. Hospital-Acquired HRV Cases and Community Control Samples Sequenced in This Study

Origin of Infection	Patient No.	Ct[a]	Total No. of Reads[a]	No. of HRV Reads[a]	GenBank Accession No.	Days Relative to Initial Hospital-Acquired Case	HRV Species
Community acquired	SCH98	29.9	1 947 654	77		−1	A
	SCH99	21.5	1 719 587	14 745	KY189313	0	A
	SCH100	31.8	745 319	10		14	A
	SCH101	24.6	1 344 804	6206	KY189314	1	B
	SCH102	26.3	1 197 551	12 796	KY189315	2	A
	SCH103	33.4	2 051 482	2		12	A
	SCH104	20	466 510	28 062	KY189316	15	C
	SCH104-2	30.2	1 562 207	21		35	C
	SCH105	22.1	1 693 577	1871	KY189317	25	B
	SCH106	30.3	1 897 153	66		26	C
Hospital acquired	SCH107	25.6	2 477 471	2297	KY189318	2	A
	SCH108	23.2	3 139 792	24 534	KY189319	14	C
	SCH108-2	37	2 150 557	0		30	NA
	SCH109	20.8	2 817 820	105 787	KY189320	22	C
	SCH110-2	18.1	2 050 558	214 464	KY189321	22	C
	SCH110-3	15.7	434 462	14 127	KY348786	64	C

Abbreviations: Ct, cycle threshold; HRV, human rhinovirus; NA, not available.

[a]Quantitative reverse transcription polymerase chain reaction cycle threshold values, total reads, and HRV reads were recovered for each sample sequenced in this study.

the ICU (day 64 of hospital stay). No members of enteroviruses A through D were recovered from our specimens.

The 3 complete HRV-C genomes from patients with hospital-acquired infection had 1575, 2277, and 2192 nucleotide differences between their respective genomes. The 2 closest HRV-C strains of concern (from case-patients SCH108-1 and SCH109) best aligned 79.0% by nucleotide to the 2006 HRV-C strain N4 from China (GenBank accession number GQ223227) and 96.6% by nucleotide to a 2009 HRV-C08 strain from the United States (GenBank accession number JQ245964), respectively. The estimated mean substitution rate was 1.93×10^{-3} substitutions per site per year (95% confidence interval, 6.59×10^{-4} to 4.04×10^{-3} substitutions per site per year) across the rhinovirus genomes analyzed here, which is consistent with previous estimates for HRV-C [6]. On the basis of this estimated substitution rate, the time to most recent common ancestor for the closest putative outbreak samples in this study, SCH108 and SCH109, was approximately 172 years (95% confidence interval, 67.4–306.6 years), consistent with not being part of a recent transmission chain.

DISCUSSION

We demonstrate here the ability of mNGS to rapidly dispel concern about a single-source transmission outbreak among a cluster of high-risk infants with underlying pulmonary disease who were diagnosed with rhinovirus/enterovirus infection. mNGS recovered HRV whole genomes from the relevant case samples, and HRV reads were recovered from all but 1 sample. In this particular cluster, HRV sequences from all the case-patients were very distantly related, and a nosocomial outbreak was

effectively ruled out. In other clusters, the recovery of whole genomes can allow for accurate inference of a transmission chain [3].

It is notable that 3 of the 4 inpatient rhinovirus infections were caused by HRV-C members, whereas only 1 of 9 outpatient rhinovirus infections was caused by an HRV-C species. HRV-C is found more commonly during the winter months and more often in lower respiratory tract illness in children than in adults [7]. Although HRV-C was more common in the case-patients, the potential association between HRV-C and worse clinical outcomes is not clear [7].

Rapid respiratory PCR panels offer turnaround times shorter than 2 hours and have revolutionized clinical virology. These panels provide sensitive detection through multiplex nested PCR with specificity based on melting temperature. Their adoption enables rapid actionable information for patient care, cohorting, and infection prevention purposes. However, providing this information leads to downstream questions, such as whether a rapid succession of positive results for the same virus in a particular clinical context is indicative of a community outbreak or a single-source hospital transmission. To answer these questions, rapid sequencing of whole genomes and phylogenetic analysis, along with complete databases in which to track these infections, are useful [8].

Limitations of this study include its retrospective nature, the lack of availability of the initial HRV-positive respiratory sample for case-patient SCH110 for metagenomic sequencing, and the limited depth of sequencing performed here (median, 1.8 million reads). It should be noted that investigation for a single source for the suspected cluster described here could have been informed by VP1 or VP4-VP2 RT-PCR and Sanger sequencing, as has been described in previous studies of HRV nosocomial

transmission [9, 10]. On the basis of a molecular clock of approximately 2×10^{-3} substitutions per site per year estimated here, we would expect the limits of resolution of identical coding sequence genomes to be approximately 2 to 3 weeks, which would be enough to rule-in a single source of transmission for a cluster of hospital-acquired cases similar to the one described here [6]. Previous 5′ untranslated region sequencing of HRV associated with household transmission found 0 to 3 SNPs associated with iterative sampling of the same individual and viral transmission over a 3-week period [11]. Future studies are merited to define the empirical limits of resolution of whole-genome analysis among known respiratory virus transmission patterns such as those previously seen in household transmission studies [11].

The work described here comprises 6 hours of hands-on wet-laboratory work, 17 hours of sequencing time, and approximately 1 hour of analysis time, for a total of 24 hours from sample receipt to rule out an outbreak. The reagent cost was approximately $1500, not including the fixed costs of the MiSeq system. Nearly all steps described here could be automated. The technique described here can be used to detect any RNA virus in respiratory specimens; no specific primers or probes are used. mNGS data can help characterize the nature of hospital-associated transmission events and direct the control strategies that are used to prevent future cases and ongoing transmission. Negative findings can be informative and cost-saving by reducing concern about potential single-source transmission. Future research is needed to examine the role of new rapid sequencing

technologies to augment the use of rapid PCR in diagnostic virology.

Note

References

1. Jain S, Self WH, Wunderink RG, et al. Community-acquired pneumonia requiring hospitalization among U.S. Adults. N Engl J Med **2015**; 373:415–27.
2. Taylor S, Lopez P, Weckx L, et al. Respiratory viruses and influenza-like illness: epidemiology and outcomes in children aged 6 months to 10 years in a multi-country population sample. J Infect **2017**;74:29–41.
3. Greninger AL, Zerr DM, Qin X, et al. Rapid metagenomic next-generation sequencing during an investigation of hospital-acquired human parainfluenza virus 3 infections. J Clin Microbiol **2017**; 55:177–82.
4. Greninger AL, Naccache SN, Messacar K, et al. A novel outbreak enterovirus D68 strain associated with acute flaccid myelitis cases in the USA (2012–14): a retrospective cohort study. Lancet Infect Dis **2015**; 15:671–82.
5. Lu X, Holloway B, Dare RK, et al. Real-time reverse transcription-PCR assay for comprehensive detection of human rhinoviruses. J Clin Microbiol **2008**; 46:533–9.
6. Kuroda M, Niwa S, Sekizuka T, et al. Molecular evolution of the VP1, VP2, and VP3 genes in human rhinovirus species C. Sci Rep **2015**; 5:8185.
7. Royston L, Tapparel C. 2016. Rhinoviruses and respiratory enteroviruses: not as simple as ABC. Viruses **2016**; 8:pii:E16.
8. Greninger AL, Messacar K, Dunnebacke T, et al. Clinical metagenomic identification of *Balamuthia mandrillaris* encephalitis and assembly of the draft genome: the continuing case for reference genome sequencing. Genome Med **2015**; 7:113.
9. Reid AB, Anderson TL, Cooley L, et al. An outbreak of human rhinovirus species C infections in a neonatal intensive care unit. Pediatr Infect Dis J **2011**; 30:1096–5.
10. Reese SM, Thompson M, Price CS, Young HL. Evidence of nosocomial transmission of human rhinovirus in a neonatal intensive care unit. Am J Infect Control **2016**; 44:355–7.
11. Peltola V, Waris M, Osterback R, et al. Rhinovirus transmission within families with children: incidence of symptomatic and asymptomatic infections. J Infect Dis **2008**; 197:382–9.