1-1-2019

Using Host Response For Respiratory Virus Discovery And Surveillance

Amelia Erin Hanron
ameliahanron@gmail.com

Using Host Response for Respiratory Virus Discovery and Surveillance

Amelia Hanron, MLS (ASCP)

Master of Public Health 2019
Yale School of Public Health
First Reader: Dr. Melinda Pettigrew
Second Reader: Dr. Ellen Foxman
Location: Foxman Lab, Yale New Haven Hospital

Abstract

INTRODUCTION: In 2015 lower respiratory infections (LRI) were the fifth leading cause of death and the leading cause of death for children under five years old. Despite increasing advances in viral detection technology, etiology is never established for a large portion of viral respiratory infections. The most recent of such advancements- next-generation sequencing (NGS)- has greatly improved the ability to discover and/or identify rare or novel viruses. However, NGS platforms are still not feasible in a clinical laboratory due to cost, complexity, personnel, etc. The ability to screen for unknown viruses using technology that is already present in most clinical laboratories would offer an efficient, cost-effective way to determine which samples may benefit from further testing with NGS.

OBJECTIVES: Establish a methodology for general screening of clinical respiratory samples for unknown or unidentified viruses.

METHODS: Clinical nasopharyngeal swabs collected in January 2017 were examined based on the results of the hospital's virus PCR panel; samples testing negative for all viruses on the panel were screened for the inflammatory host biomarker for viral infection CXCL10 using an ELISA. CXCL10-positive samples were tested for viruses not included on the panel to ensure the presence of a "true unknown" virus. Potential NGS approaches were concurrently investigated using rhinovirus-positive samples from the same population.

RESULTS: Out of 251 patients with negative viral results from the clinical laboratory, 60 were found to express a high level of CXCL10 in their sample, indicating a likely viral infection. Twenty-eight of these were found to contain coronaviruses, and the remaining 32 were declared "unknown." These unknown samples will undergo further testing through a variety of techniques to determine the identity of any virus present in the sample.

CONCLUSION: Protein host biomarker CXCL10 is produced in human respiratory epithelial cells in the presence of viral infection. Screening of respiratory samples for CXCL10 provides a pan-viral test that does not require knowledge of a specific pathogen; this study demonstrated a feasible workflow that could be used to screen large numbers of clinical respiratory samples for the presence of unknown viruses using technology that is already widely used in clinical laboratories. This screening assay could serve as a low-cost way to identify samples containing viral pathogens that otherwise would not be detected; such samples could then be sequenced to identify the infectious agent.

Acknowledgements

The sequencing portion of this project was made possible by a collaboration between the Foxman Lab and Scott Handley's laboratory at Washington University, which has developed a bioinformatic pipeline for viral discovery.

Thanks to Valia Mihaylova for invaluable guidance and assistance in laboratory procedures

Table of Contents

List of Tables

List of Figures

Background

Burden of respiratory disease

Respiratory infections are a leading cause of morbidity, mortality, and disability, and present a substantial economic burden in the form of hospitalizations and health care costs.[1,2] Lower respiratory tract infections (LRIs) make up the majority of serious clinical cases, and are responsible for up to 4 million annual deaths globally.[2] LRIs are the cause of more global deaths than HIV, TB, and malaria combined, and disproportionally affect children under the age of 5.[1,2] Upper respiratory infections (URIs), while not generally as clinically severe, still account for millions of outpatient visits and over 20 million lost days of work or school.[3] In addition, patients with URIs may develop complications, particularly in the presence of respiratory comorbidities like asthma or COPD.[1]

While LRIs like pneumonia may be caused by bacteria, only about 10-15% of URIs are attributed to bacterial infections, with the majority attributed to viral infections.[4] Viruses pose a unique challenge in detection and diagnostics, and very often the etiologies of LRIs are never conclusively established (Graf et.al estimated that only 20-60% of cases of community-acquired pneumonia (CAP) have a detectable pathogen).[5] There are several reasons for this: viral genomes (particularly in RNA viruses) are highly prone to mutations, single nucleotide polymorphisms (SNPs), and other forms of genetic diversity, which can lead to difficulty in identification.[6,7,8] In addition, unlike bacteria or fungi there is no common genetic feature that all viruses share, making the creation of a pan-viral assay much more challenging.[9,10] Thus, the inherent issue with clinical virology assays is that one must know which specific pathogen is being looked for when testing. With so many respiratory infections lacking a detectable pathogen even after exhaustive

testing, it follows that there must be yet-unidentified viruses circulating in humans and causing clinical disease.

Virus detection and discovery methods

Culture

Researchers and clinicians have developed different methods for virus discovery and identification over the years, the most traditional being viral culture. In viral culture, a sample from a patient or other source is inoculated into a variety of established cell lines, and any observable cytopathic effects are evaluated.[6,9,11] There are many benefits to cell cultures, and they are often used in conjunction with newer, more complex assays.[12] Cultured cells and supernatants are a source of isolated, concentrated virus that can then be used for electron microscopy, serology assays, and sequencing methods.[6,9,13] Culturing is the best way to isolate and amplify a virus, and can be done without knowing what kind of virus may be present.[9,13]

The limits of cell cultures lie in the specific cell lines that are inoculated. Different pathogens thrive in different environmental conditions and often have specific requirements in order to survive and replicate.[12] Many viruses may not grow on traditionally-used cell lines, or may not show any visible cytopathic effects. [6,12,14] Therefore, a negative culture does not necessarily indicate that the original sample contains no virus. This can sometimes be solved by culturing cells specific to the original environment, like human airway epithelial cells (HAE) for the growth of respiratory viruses.[14] Another potential issue in viral cell culture is the possibility of mutations after several passes; selective pressure can result for specific strains or types of virus that are most equipped for the cell lines being used (although this issue is arguably more relevant for downstream research such as pathogenicity determination and assay development).[6,7,12]

Polymerase Chain Reaction

The polymerase chain reaction (PCR) is one of the most common viral detection methods used in clinical laboratories. Briefly, samples are combined with small genetic segments (primers) complementary to the gene or area of interest that a specific pathogen is known to contain. This allows for the exponential replication of the target sequence, the concentration of which can be measured and quantified to determine the relative amount of that sequence – and therefore that pathogen – in a sample. PCR (or reverse-transcription PCR for RNA viruses) assays in clinical settings are generally sensitive enough to detect very low levels of pathogen.

The downside of PCR, when it comes to virus discovery, is that the sequence of the pathogen must be known in order to create pathogen-specific primers and probes. Primers may be made that recognize a target common to a virus family or species, but even then the investigator must have an idea of what kind of virus may be present in the sample.[15] Many clinical laboratories have multiplexed PCR panels that allow for the testing of many viruses in one assay. This lowers the workload of testing, but still only tests for specific viruses, and may miss viruses not on the panel or even particularly divergent strains of panel viruses.[16,17]

Although perhaps not sufficient on its own, PCR technology is still an integral part of newer sequencing methods for viral discovery: it is used to identify certain genomic fragments, amplify target sequences, and other similar steps as explained in the next section.

Next Generation Sequencing

Sequencing all or part of an organism's genome and comparing it to reference sequences has become a promising technology in virus discovery and characterization. Metagenomics using next generation sequencing (NGS) technologies allows for the detection and identification of all genetic information present in a sample; the process has been used to characterize the viral population of environmental samples such as ocean water or waste.[6,11,18] NGS can be used to amplify and sequence an unknown viral genome, which can then be compared to reference sequences using bioinformatic software.[10,19]

The benefit of NGS is that it is sequence-independent; that is, it does not require *a priori* knowledge of the identity or sequence of any viruses in the sample.[5,15] Samples go through an enrichment (of viral particles) and/or depletion (of host DNA) process and are fragmented and amplified using PCR or similar technology. Those fragments are then sequenced in parallel, creating millions of "reads" of short DNA sequences that are segments of the entire sequence of the original sample. These reads are then analyzed and aligned to known reference sequences.

There are inherent challenges present when using NGS for viral discovery. First, in clinical samples the amount of viral nucleic acid is usually much lower than host DNA and RNA.[11,18,19] This means that, without certain preparation steps, any viral reads will be vastly outnumbered and therefore more difficult to detect and identify.[7,10] For this reason, samples are usually put through either a viral enrichment process or a background nucleic acid depletion procedure.[19,20] Enrichment methods include ultracentrifugation and filtration to concentrate viral particles and

remove non-viral nucleic acid.[11] Depletion is generally done with a DNAse step during the sample preparation process.[9,11,12]

Sequencing is only as useful as its reference library. Reads are identified by comparing them to known sequences in vast reference libraries.[20,21] This poses a problem when dealing with a truly novel virus or particularly divergent strains of known viruses; a matching sequence may not be found or identified as such in the available databases, making conclusions about the identity of sequences virus difficult.[9,20] Genomic material for which no match is found (and has been confirmed as non-human, non-fungal, and non-bacterial) is sometimes referred to as viral "dark matter," and it makes up a significant amount of reads in many viral metagenomics studies.[9,20,21] Most viruses that have been discovered to date using NGS had enough homology to known viruses to allow for the preliminary information on the family or species, which can then inform further identification processes.[4,12,17]

While still largely confined to research laboratories, there are many clinical reasons to develop methods to identify a virus using NGS. During an ongoing epidemic, identifying the causative agent early allows for the quick development of faster PCR tests to identify cases and inform the beginning of vaccine or antiviral work.[6,22] Deep sequencing also has the potential to identify distinct subtypes or SNPs that confer different resistance profiles or enhanced virulence.[5,10] Sequencing has enabled or aided in the discovery of many clinically significant viruses, including human metapneumovirus (hMPV), human bocavirus (HBoV), distinct subtypes of coronavirus, and others.[4,23,24]

Screening before NGS

Next generation sequencing is a complex process that comes with often prohibitive costs and the need for advanced computing power and training. For this reason, it is largely used for research – either in metagenomics or clinical case studies – rather than as a diagnostic tool. At this point in the technology, it would be impractical to implement NGS as a general diagnostic test to be used in the same context as PCR panels or immunoassays.

CXCL10

CXCL10 is a pro-inflammatory chemokine produced by many different kinds of cells in the body in response to infection or cell damage.[4,23,24] It serves as a chemoattractant, recruiting CD4+ T-cells, monocytes, and NK cells and inducing further cytokine secretion.[25,26] Its role in inflammation has been studied in many different contexts: it has been examined as a marker of increased risk of allogenic graft rejection, hepatitis-induced liver fibrosis, pre-term labor, and HIV progression.[26–28] Previous work from the Foxman Lab showed that CXCL10 may be useful as a host-derived biomarker for viral diagnostics, as it is highly expressed in respiratory epithelial cells in the presence of an active viral infection.[16] That study used nasopharyngeal swabs and showed that CXCL10 mRNA and protein levels were a good predictor of the presence of one or more viruses.[16] Importantly, that study also showed that bacterial infections did not result in the same activation of CXCL10, confirming that this marker may be effective in distinguishing between viral and bacterial respiratory infections.[16]

CXCL10 as a screening method

The goal of this project was to begin to bridge the gap between the clinical laboratory and the power of next generation sequencing, by developing a method to screen clinical samples to identify those containing truly unknown viruses and evaluate potential workflows to identify those viruses using NGS. We sought to develop a methodology to utilize CXCL10 as a way to screen patient samples for unknown or unidentifiable respiratory viruses. To this end we examined clinical respiratory samples (nasopharyngeal swabs) that had tested negative for all viruses on the clinical virus PCR panel to attempt to determine if any contained a virus that is not included in that panel, and so went undetected in the clinical laboratory. These relevant samples can then continue on to NGS.

Table 1: Respiratory Viruses in Yale New Haven Hospital semi-quantitative PCR Panel
Adenovirus
Human Metapneumovirus (hMPV)
Rhinovirus (RV)
Influenza A and B
Respiratory Syncytial Virus (RSV)
Parainfluenza 1-3, **4*(**PIV 1-4)
Coronavirus*
**added to panel after beginning of study*

Table 1: Viruses included in YNHH respiratory panel

Methods

Samples

1,109 samples of frozen viral transport media from nasopharyngeal swabs were supplied by the clinical laboratory at Yale New Haven Hospital (YNHH). Samples were collected in January 2017, and transport media was kept at -80°C. Additionally, individual test information was provided by the clinical laboratory to break down all viral respiratory tests that had been run on

the provided nasopharyngeal swabs. Samples had been tested using a clinical PCR panel consisting of Influenza A and B, adenovirus, respiratory syncytial virus (RSV), rhinovirus (RV), Parainfluenza viruses 1-3 (PIV), and human metapneumovirus (hMPV).

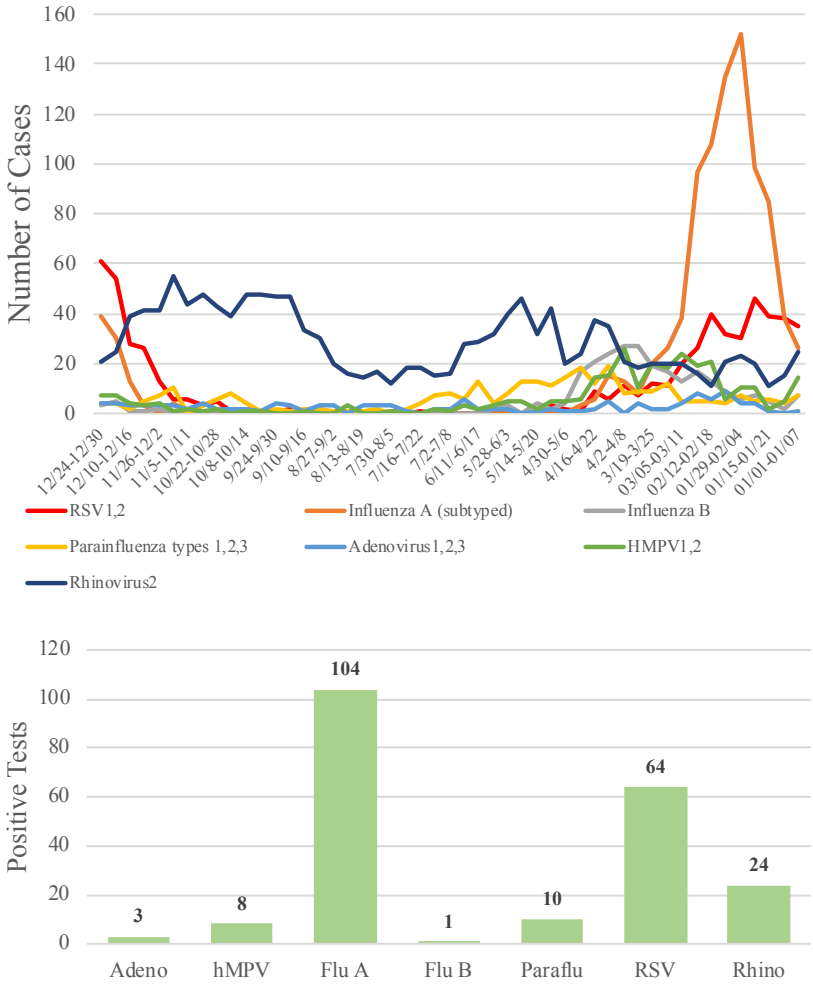

*Figure 1: URI cases at YNHH in 2017, by week **(a)**; Positive tests per virus, week 4 **(b)***

Out of the 1,109 samples, 732 had been tested with a complete clinical PCR panel (see **Table 1** for a list of included viruses). Of those, 588 were negative (80.3%). Samples were organized by week; week 4, from January 22 to January 31, contained the most samples (n=266), so these samples were chosen for further analysis (see **Figure 1a** and **1b** for a weekly breakdown of URI cases at YNHH). Some samples in the provided record did not have a corresponding sample or were mislabeled, bringing the final number of analyzed samples to 251. Data was collected on these samples including patient demographics, information on the encounter in which the sample was taken, and comorbidities. Samples were deidentified and chart review procedure was approved by Yale Human Investigations Committee. Symptoms

were pulled from ICD-9 codes and chart notes from the encounters for which the specific samples were collected.

Additionally, information was collected from 33 samples from the original 1,109 that were positive for rhinovirus (RV) according to the clinical virus panel. These samples were used in initial optimization and quality control testing.

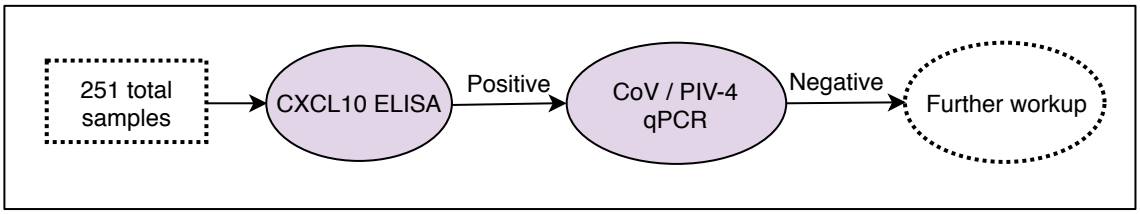

Figure 2: Workflow for PCR-negative samples

The workflow for these samples is shown in **Figure 2.** First, samples were tested for the presence of CXCL10 by ELISA. Those that tested positive for CXCL10 were screened for coronavirus and parainfluenzavirus-4 with qPCR (these are common respiratory viruses that were not included in the hospital's panel), and those that were negative were considered true unknowns and were carried on for further investigation

Method Development and Sample Preparation

Experiments using rhinovirus (RV)-positive samples were done to determine the optimal lysis buffer and storage conditions that would result in the highest RNA yield. Thawed samples were prepared using the following methods: thawed samples with added Qiagen AVL buffer and BME (Sigma) that were tested without being refrozen (referred to as "fresh"), and samples that went through one freeze-thaw after preparation and before testing. Freeze-thaw samples were designated as frozen with no additives, frozen with AVL and BME, and frozen with AVL and

RNALater RNA stabilization reagent (Qiagen). qPCR was then performed using primers for HPRT, β-actin, CXCL10, and rhinovirus/enterovirus (Invitrogen). Relative amounts were calculated based on β-actin, and are shown for both high and low levels of rhinovirus in **Figure 3** (based on CTs from clinical laboratory results)**.** Based on these results, it was decided that going forward, samples would be prepared with AVL and BME, as it would not be possible to use fresh samples

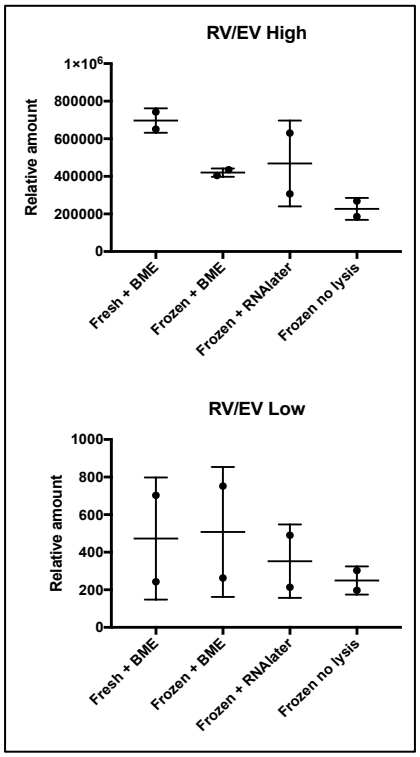

Figure 3: Relative amounts of rhino/enterovirus present in rhinovirus-positive samples under different storage conditions

In the final workflow used for this project, the following aliquots were made from each sample: the original sample (viral transport media) was thawed on ice, and two tubes of 140 µl were made for RNA isolation, to which 560 µl of AVL Lysis buffer (Qiagen) was added. 400 µl of original sample was spun down to remove cell debris; 70 µl of supernatant was added to a flat-bottom 96-well plate for immunoassay. The remaining supernatant was transferred into a new tube to be used as extra material if needed. All aliquots, if not immediately used, were frozen at -80° C.

RNA extraction

Viral RNA extraction was carried out on samples using the Qiagen QIAmp Viral RNA Mini kit with no added carrier RNA. Samples had been frozen with AVL lysis buffer as described above, which is the first step of the Qiagen protocol, so those aliquots could be simply be thawed before

proceeding with the protocol. Extraction was done per kit instructions, with the addition of a 15-minute DNAse (Qiagen) incubation step before elution. Samples were eluted in two elutions of 30 µl, the first of which was used to make cDNA. Remaining eluates were stored at -80°C.

cDNA was prepared from extracted RNA using the Bio-Rad iScript cDNA synthesis kit, with 4 µl of Reaction Mix and 1 µl reverse transcriptase added to 10 µl water and 5 µl sample. This was run on a Bio-Rad T100 Thermal Cycler under manufacturer-recommended conditions. cDNA was stored at -80° when not in use.

CXCL10 ELISA

Samples were screened for CXCL10 using the R&D Systems Human CXCL10/IP-10 DuoSet ELISA per manufacturer's instruction. For each 96-well plate, the assay was performed over the course of three days. Detection was performed on the final day, and the plates were read using a VersaMax Microplate Reader (Molecular Devices) at 450 nm.

Standard curves were generated using the manufacturer-provided concentrations of the standards, and used to calculated CXCL10 concentrations (in pg/ml). The lower limit of the linear range was 32.25 pg/ml. Samples were run in 1:5 dilutions, making the lower limit of detection 61.25 pg/ml for original samples.

qPCR

Because the clinical PCR panel did not include coronavirus or parainfluenza-4 virus at the time that these samples were analyzed, SYBR green qPCR (Bio-Rad) was performed on CXCL10-

positive samples with CoV and PIV-4 primers.[29] Samples that were positive for CXCL10 and negative for CoV/PIV-4 were considered "true unknowns" to be further investigated. Some CXCL10 samples were also tested for ß-actin to ensure sample integrity.

Initial Sequence Investigation

The next step in a process for identifying unknown viruses may be NGS. To begin to explore the best options for this step, eight samples were selected. These included four clinical samples that were known to be RV-positive (one of which was coinfected with RSV), two RV-infected cell culture samples (positive controls), and two RV-negative cell culture samples (negative controls). RNA isolated from these samples using the above protocol was sent to the Yale Center for Genomic Analysis (YCGA). YCGA performed a low-input library generation protocol (using the Kapa Biosystems ribo-depletion kit) and sequenced samples using Illumina technology. The resulting data was sent to both an academic collaborator and a commercial pathogen identification company for analysis.

RNA isolation and sequencing was chosen over DNA or total nucleic acid based on the high percentage of RNA respiratory viruses, while recognizing that sequencing RNA may also recognize DNA viruses that are actively being transcribed. Ribo-depletion was performed to eliminate host nucleic acids and increase sensitivity to any viral particles.

Results

Results of the procedures are summarized in **Figure 4.** Out of 251 patient samples, 60 (23.9%) were found to be CXCL10-positive by immunoassay. Of those 60, 28 (11.2% of original 251

samples) were found to be positive for CoV, using primers that included three different strains (229E, NL63, OC43). The remaining 32 (12.7%) CXCL10-positive samples tested negative for CoV and PIV-4 via qPCR, and were considered true unknowns.

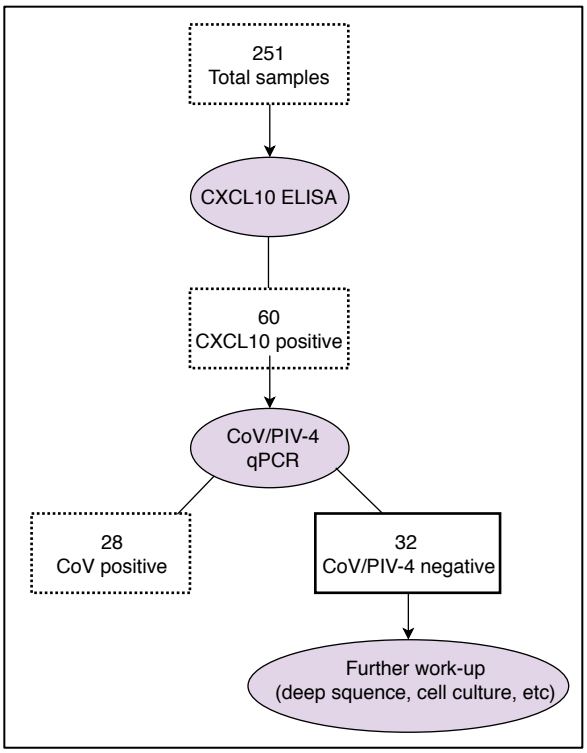

Figure 4: Summary of results

The 251 samples that made up our study population are described in **Table 2.** The average age of the patients associated with the samples was 59.9. 43.0% were over the age of 65, with 79.2% being over the age of 45. Females accounted for 61.0% of the samples, and the majority were from Caucasian patients (57%). Most samples came from inpatients (80.5%), many of whom were admitted from the emergency department. 53.4% of patients presented with respiratory symptoms and 19.5% were febrile. 10.4% reported cardiac related symptoms, such as chest pain or pressure, tachycardia, etc. It's important to note that these symptoms are also commonly seen in respiratory infections and should not be considered solely indicative of a cardiovascular problem. The most common comorbidities (33.5%) were respiratory diseases

Table 2. *Description of samples (n=251)*

	N (%)*
Age	
<5	9 (3.6)
6-15	2 (0.8)
16-25	13 (5.2)
26-45	28 (11.2)
46-55	35 (13.9)
56-65	56 (22.3)
>65	108 (43.0)
Gender	
Male	98 (39.0)
Female	153 (61.0)
Race/Ethnicity	
White	143 (57.0)
Black	67 (26.7)
Hispanic	32 (12.8)
Other/Unknown	9 (3.6)
Patient Status	
Inpatient	202 (80.5)
Outpatient	23 (9.2)
ED	21 (8.4)
Unknown	5 (2.0)
Presenting symptoms	
Respiratory	134 (53.4)
Fever	49 (19.5)
Cardiac	26 (10.4)
Altered mental state	22 (8.8)
Fatigue	8 (3.6)
Other	95 (37.8)
Comorbidities	
Respiratory	84 (33.5)
Cardiovascular	67 (26.7)
Diabetes	60 (23.9)
Cancer	51 (20.3)
Liver/kidney disease	48 (19.1)
Other	81 32.3)
Ordering Department	
General medicine	148 (59.0)
ICU/Surgery	35 (13.9)
Oncology	22 (8.8)
ED	20 (8.0)
Outpatient	16 (6.4)
Other/Unknown	10 (4.0)

**Percentages may not add up to 100% due to multiple symptoms/comorbidities per single patient*

Table 3. *Description of PCR-negative samples: CXCL10 negative vs CXCL10 positive ("unknowns")*

	PCR neg/CXCL10 neg (n=191) N (%)†	**PCR neg/CXCL10 pos (n=32)** N (%)†	**p-value**
Age			0.038
<5	4 (2.1)	3 (9.4)	
6-15	2 (1.1)	0 (0.0)	
16-25	7 (3.7)	5 (15.6)	
26-45	21 (11.0)	3 (9.4)	
46-55	29 (15.2)	4 (12.5)	
56-65	42 (22.0)	5 (15.6)	
>65	86 (45.0)	12 (37.5)	
Gender			0.67
Male	76 (39.8)	14 (43.8)	
Female	115 (60.2)	18 (56.3)	
Race/Ethnicity			0.16
White	105 (55.0)	24 (75.0)	
Black	52 (27.2)	6 (18.8)	
Hispanic	25 (13.1)	2 (6.3)	
Other/Unknown	9 (4.7)	0 (0.0)	
Patient Status			0.68
Inpatient	157 (82.2)	25 (78.1)	
Outpatient	15 (7.9)	4 (12.5)	
ED	15 (7.9)	3 (9.4)	
Unknown	4 (2.1)	0 (0.0)	
Presenting symptoms*			
Respiratory symptoms	95 (49.7)	20 (62.5)	0.18
Fever	37 (19.4)	6 (18.8)	0.93
Cardiac	20 (10.5)	1 (3.1)	0.32
Altered mental state	16 (8.4)	4 (12.5)	0.50
Fatigue	3 (1.6)	5 (15.6)	<0.001
Other	65 (34.0)	14 (43.8)	0.29
Comorbidities*			
Respiratory	68 (33.6)	10 (31.3)	0.63
Cardiovascular	57 (29.8)	5 (15.6)	0.097
Diabetes	49 (25.7)	5 (15.6)	0.24
Cancer	37 (19.4)	8 (25.0)	0.46
Liver/kidney disease	42 (22.0)	4 (12.5)	0.22
Other	62 (32.5)	9 (28.1)	0.63
Ordering Department			0.92
General medicine	113 (59.2)	17 (53.1)	
ICU/Surgery	27 (14.1)	6 (18.8)	
Oncology	18 (9.4)	2 (6.3)	
ED	15 (7.9)	3 (9.4)	
Other/Unknown	7 (3.7)	2 (6.3)	
Outpatient	11 (5.8)	2 (6.3)	

Comparisons were performed using a global χ^2 test except where indicated with (*), where individual χ^2 was used.
(†) *Some percentages may not add up to 100% due to multiple symptoms/comorbidities per single patient*

(including asthma and COPD), followed by cardiovascular disease (26.7%), and diabetes (23.9%).

Table 3 shows characteristics of the 223 PCR-negative samples (the original 251 minus the 28 that tested CoV positive), comparing those that were CXCL10-positive (i.e. our unknowns) vs those that were CXCL10-negative (those that are presumably virus-negative)**.** There was a statistically significant difference in age between the two groups ($p=0.038$); in general, more CXCL10-positive patients were under the age of 25, although the highest percentage of patients in both populations were over the age of 65. The only other significant difference between the populations was that CXCL10-positive patients were more likely to report fatigue as one of their symptoms upon presentation ($p<0.001$).

Briefly, 37.5% of CXCL10-positive samples (n=32) were over the age of 65, with 9.4% under the age of five. Like the CXCL10-negative population, the majority of patients were female (56.3%), white (75.0%), and were inpatients at the time of sample collection (78.1%). 62.5% of these patients presented with respiratory symptoms of some kind, and 31.3% had respiratory comorbidities. The most common comorbidity after respiratory among these 32 patients was cancer (25.0%). Breakdowns of what was included in each category in symptoms and comorbidities can be found in **Supplementary Tables 1** and **2.**

At the time that these samples were collected, YNHH's clinical laboratory did not test for parainfluenza-4 virus or any coronaviruses. Because of the high incidence demonstrated by this study, the YNHH lab has now added PIV-4 and CoV to its standard PCR panel.

CXCL10 Concentration

For the purpose of this study, any discernible CXCL10 (that is, above the lower limit of detection of 161.25 pg/ml) was considered positive. CXCL10 concentrations for unknowns as well as those that tested coronavirus-positive, sorted highest to lowest, are shown in **Figure 5.** The two populations had similar distributions of CXCL10 concentrations, with the exception of the two highest CoV samples which had higher concentrations than the highest unknowns. These comparable CXCL10 levels between known virus-positive samples and "unknowns" further support the hypothesis that the PCR-negative "unknown" samples contain a virus. Interpretation of CXCL10 levels will depend on its use; likely, a qualitative positive/negative result would serve the purpose of most assays likely to utilize this method. In that case, a cut-off value would need to be established to define a "positive" result.

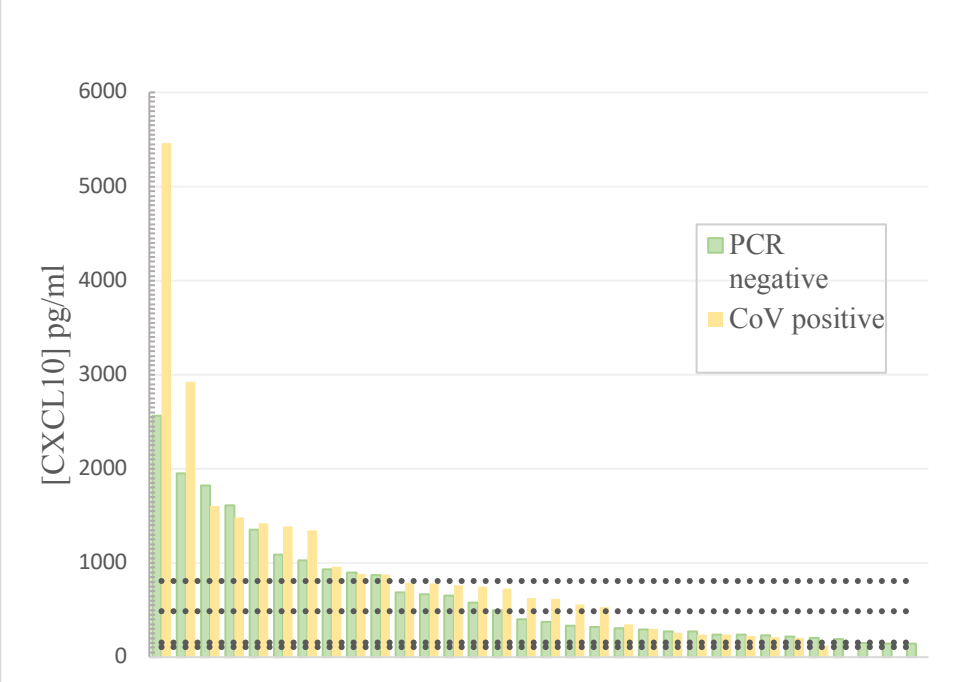

Figure 5: [CXCL10] for "unknown" samples and coronavirus positive samples. Dotted lines represent potential cut offs that result in sensitivities of 50%, 60%, 76%, and 81%. Sensitivities estimated based on data from Landry and Foxman (2018)

Figure 5 shows possible cut-off concentrations resulting in different sensitivities using data from Landry and Foxman.[16] Cut-offs of 107 pg/ml, 155 pg/ml, 488 pg/ml, and 809 pg/ml result in sensitivities of 50%, 60%, 76%, and 81%, respectively. Applying these cut-offs to the 32

CXCL10-positive samples from this study would result in the following percentages of our original "unknowns" to be considered positive: 100% (32/32), 91% (29/32), 47% (15/32), and 31% (10/32) with 50%, 60%, 76%, and 81% sensitivity, respectively. It is important to note that Landry and Foxman used a different technology to quantify CXCL10 than did this study (Bio-Plex immunoassay vs ELISA). Further studies with samples containing known concentrations of specific viruses would be needed to better determine a cut-off based on specificity and sensitivity.

Sequencing Results

The commercial company detected some but not all of the viruses known to be present in the virus-positive samples, and those that were detected generally had very few reads. The report

Sample	Organism		No. of reads
Positive Control 1	**Rhinovirus A**		20
	Pseudomonas veronii	Environmental bacteria	6,372
Positive Control 2	**Rhinovirus A**		20
	Pseudomonas pseudoalcaligenes	Environmental bacteria	5,408
RV-positive 1	*Acanthamoeba royreba*	Non-pathogenic amoeba	4,095
	Treponema pedis	Bovine pathogen	2,747
RV-positive 2	*Zygosaccharomyces bailii*	Fungi involved in food spoilage	5,114
	Macrocystidia cucumis	Inedible mushroom	1,151
RV-positive 3	*Macrocystidia cucumis*	Inedible mushroom	3,711
RV-positive 4	**RSV**		6,518
	Rhinovirus B		31
	Corynebacterium propinquum	Normal flora	92,148
	Brachybacterium muris	Non-pathogenic bacteria isolated from mouse liver	5,298
(Negative Control 1)	*Leucobacter sp.*	Environmental microbacteria	5,452
(Negative Control 2)	*Zygosaccharomyces bailii*	Fungi involved in food spoilage	5,687

Table 4: Number of reads of expected clinical virus (if detected) and most abundant organism for six rhinovirus-positive and two negative control samples from commercial pathogen-identification report.

also listed a myriad of other organisms in both the rhinovirus-positive samples and the negative control cells. These included a murine malaria pathogen, an oceanic diatom, plant pathogens, and one fungus that is actually an edible mushroom (See **Table 4)**. It is important to note that this company's methods are proprietary, and they largely work with DNA rather than RNA. Our bioinformatics collaborators in the Handley group at Washington University in St Louis did detect rhinovirus in all six positive samples, although the sensitivity was still low (number of rhinovirus reads ranged from 1-528). The Handley group did not report environmental contamination at the same level as the commercial company (data not shown).

Methods to obtain a higher RNA yield and as such prepare better libraries are discussed in the "Next Steps" section of this paper.

Discussion

Characterization of Unknowns

This project serves to introduce a series of procedures that improve the efficiency of identifying patients with unidentified respiratory viruses, and a potential pipeline that those samples may move through for such identification. CXCL10 is a proinflammatory cytokine that is released by cells during a viral infection. Its main function is as a chemoattractant for T-cells, but it may also play a role in other aspects of the immune response.[25] CXCL10 is one of a family of genes that are highly induced by interferon signaling, and is one of the most highly induced proteins in nasal epithelial cells in the presence of viral RNA.[16] Previous work from the Foxman laboratory showed that CXCL10 mRNA and protein levels are highly correlated with the presence of a viral infection, using nasopharyngeal swabs from symptomatic patients.[16] Because CXCL10 is

released as part of a specifically antiviral response, it is a promising option for a pan-viral marker for respiratory viral infection.

In this study, clinical nasopharyngeal swabs for which a clinical respiratory virus panel had been ordered were obtained for the month of January 2017, 144 of which tested positive for at least one virus on the clinical PCR panel and 588 of which no virus was detected. (The PCR panel included Influenza A and B, Parainfluenza 1-3, Adenovirus, Rhinovirus, RSV, and hMPV). A sub-population of 251 negative samples were then tested for CXCL10 to attempt to identify samples containing viruses that were not included on the hospital's panel. These may include rare or unusual viruses, divergent sub-species or strains of a common virus, a virus that is generally not pathogenic but may become so in certain settings (i.e. immunosuppression), or novel viruses. Sixty panel-negative samples (23.9%) tested positive for CXCL10 (>161 pg/ml based on standard curve) using a standard ELISA. Of those 60, 28 ultimately tested for other known viruses that were not on the original panel, but 32 remained undiagnosed (11.2% and 12.7% of the original population, respectively).

There was a small but significant difference ($p=0.038$) in age distributions when the 32 "unknowns" were compared with the rest of the PCR-negative/CXCL10-negative samples. While both populations had the highest percentage of people in the over-65 age group, a higher percentage of the remaining patients in the CXCL10-positive ("unknown") population were under the age of 25. Assuming that our assumption is correct – that the unknown samples do indeed contain a virus – it is reasonable to see more children represented, due to their potentially higher susceptibility to infection. Notably, unknown samples had more patients under the age of

five: 9.4%, compared to only 2.1% in CXCL10-negative samples. The only other statistically significant difference was that patients with CXCL10-positive/unknown samples were more likely to list fatigue as a symptom upon initial presentation ($p<0.001$). While it would be reasonable to attribute high levels of fatigue to virus-positive samples, if that were the case we would also likely expect to see significant differences in other symptoms (such as fever) between the two populations. This difference was not observed. However, all patients were judged to need a nasopharyngeal swab from the clinical staff – likely decisions on whether or not to screen for CXCL10 will be due to overall presentation rather than on specific symptoms. The populations were similar in all other variables, including race and gender, comorbidities, and immunosuppressed status (considered when patient met one or more of the following criteria: HIV-positive, diagnosed with cancer, currently pregnant, or recent transplant donor or recipient. Data not shown). This suggests that there is no factor shared by these 32 patients that may produce elevated CXCL10 levels in the absence of a viral infection.

Initial sequencing results

The next step in a potential virus identification pipeline is identification of the unknown pathogen. To this end, RNA from rhinovirus-positive clinical samples (isolated using the described method) as well as positive and negative cell culture controls was sequenced to determine optimal library preparation protocols and data analysis methods. RNA was chosen over DNA or total nucleic acid to take advantage of the fact that the majority of respiratory viral pathogens are RNA viruses; the use of a DNAse step and ribo-depletion method could then be used to increase sensitivity to RNA virus particles. Importantly, RNA sequencing may also be able to identify the presence of actively replicating DNA viruses.

As was expected, clinical samples had very low levels of RNA, generally under 20 ng/ml (data not shown). To ensure that an adequate number of reads could be detected, the first library preparation was carried out with a low-input amplification step. Low-input protocols involve additional amplification of original sample in order to increase the relative amount in the sample. However, subsequent data analyses (described below) reported a high amount of environmental contamination, possibly due to the nonspecific amplification performed during the low-input library preparation. For this reason, a low-input protocol was not used in subsequent library preparations. Subsequent sequencing of clinical samples without a low-input protocol greatly reduced background and false positive reads (data not shown).

It could be argued that the negative clinical PCR results for some or all of these samples could be due to low amounts of or highly degraded RNA, rather than the true absence of virus. To rule out this possibility, 23/32 "unknown" samples included primers for the house-keeping gene ß-actin. All were highly positive, indicating that sample RNA was of high enough integrity to be detected by qPCR. (data not shown).

Data analysis was performed concurrently by a commercial sequencing company and by the Hadley group at Washington University. The company did not detect rhinovirus in all of the RV-positive samples, and when it was detected it was in amounts much lower than what would be expected based on PCR results. Additionally, they reported the detection of a massive number of other microorganisms (bacteria, fungi, and protozoa), the large majority of which appeared to be either environmental contamination or misidentified reads. The Hadley group did not report the

same level of contamination, and was able to detect rhinovirus in all positive sample (as well as RSV in a coinfected sample).

There are several factors to be considered here. The Hadley group reported that water – even laboratory grade DI or RNAse-free water- often contains reads from viruses or bacteria. For this reason, in many sequencing labs it is standard to sequence the water used in the preparation processes, to establish a background that may then be subtracted from the final result. This was not done in this case, and likely is a source of at least some of the environmental contaminants detected. Secondly, there can be significant overlap between viral sequences and the human genome, which the Handley lab protocol takes into account. The protocol used by the commercial company is proprietary, and thus it is not known if these or other potential issues are considered as part of their procedure.

Clinical benefits

Identifying unknown viruses present in respiratory samples has several benefits. First, knowing what virus is present can affect treatment or follow-up decisions, and generally improve patient care. Secondly, identifying these viruses can give important information about what pathogens are circulating in the hospital (or wider) population, which may be useful for observing trends, adjusting testing or treatment protocols, or recognizing an epidemic or emerging virus in the early stage. This was illustrated in the course of this study, when the YNHH clinical laboratory added four new viruses (PIV-4 and three strains of CoV) to their standard panel based on how many of them were found over the course of this project.

This sequence of methods presents a way of efficiently screening large numbers of patients for unknown respiratory viruses using technology that is already well established in clinical laboratories. In one possible algorithm, any samples that are negative for the PCR panel would automatically be sent for a CXCL10 ELISA. Any samples positive on the ELISA would be considered unknown and could then be sent for sequencing. ELISA assays can be batched, so many samples can be run at once – potentially between 30-40 in a standard 96-well plate.

This method could also have potential for wider surveillance programs outside of a clinical setting. This is particularly relevant since respiratory viruses have been the cause of the most important outbreaks and epidemics of the last 20 years including SARS, MERS, and H1N1 influenza.[19]

While the PCR and ELISA process can be completed in a day or two, sequencing any unknown viruses likely still takes too much time to be of clinical help to the original patient. However, identifying and classifying the virus can be useful for future patients – sequencing allows for the creation of pathogen-specific primers which can then be used for a much faster identification, and provide a starting point for drug discovery or vaccine trials.[10,19,22]

Limitations

This project was limited by the samples that were available. These were collected from one week in January, and as such do not give information on the prevalence of unknown or unusual viruses during the rest of the year. Respiratory viruses are known to be seasonal, so likely the number of hospital-panel-negative/CXCL10 positive samples will vary. Similarly, these results may not be

generalizable to all populations, as the sample study population tended to be older, with more comorbidities than younger adults or children. However, this was taken to be representative of the specific population that would likely have their samples tested through this methodology. Extensive testing for respiratory viruses is generally only done in serious, inpatient cases, and the population that tends to be most at risk for serious respiratory infection is consistent with our study population. Similar screening studies at different seasonal points during the year and with specialized populations would be beneficial.

It may also be important to note that complete clinical data was not available or not collected for all of the patients in this population. In future studies, evaluating other laboratory results – particularly other microbiology results – could add valuable information to the categorization of this population.

The results of this study should not be taken to mean that this protocol will recognize all unknown viruses that are circulating. The vast majority of respiratory infections are self-limiting, and patients may only present to an outpatient clinic or not seek care at all. This method is useful in the population that is most at risk: inpatients who are particularly susceptible to severe respiratory infections, whether due to underlying disease or extremes of age. Additionally, this protocol may not distinguish samples in which there is a coinfection of an unknown virus with a clinically-recognized virus.

Finally, although CXCL10 is a known biomarker for viral infection, it may also be induced or triggered by other pathogens or disease processes.[28,30,31] CXCL10 has been suggested as a

biomarker for allogenic graft rejection and graft-versus-host disease (GVHD) and as a predictor of rapid progression of AIDS in HIV-infected patients. [27,28,31] Additionally, it has been shown to be upregulated in several non-viral disease processes, including cerebral malaria. *H. pylori* and *Mycoplasma* infections, chronic inflammatory placental lesions, various malignant tumors, and severe *Legionella* infections.[27,28,30] Most of these studies have measured plasma or serum concentrations, rather than swabs or other cellular samples as in this study. There is also evidence to show increased levels of CXCL10 during times of general immune system dysregulation, as in sepsis, multi-agent infections, or gastrointestinal dysbiosis.[25,28,30] More studies of CXCL10 concentrations in the presence of non-viral and/or non-respiratory infections are needed to further define the function of this protein, and care should be taken when using CXCL10 in diagnostic decisions, particularly in patients with severe or multifaceted illness.

Future Directions

While this methodology holds promise for identifying samples with unknown viruses, as demonstrated by the identification of coronavirus-positive samples in "negative" patient samples, the process of identifying viruses that may be present in those samples is ongoing. To this end, several parallel investigations are ongoing in the Foxman Lab. These include PCR for additional untested viruses such as coronavirus type HKU1 and polyomavirus, ruling out fungal infections, and methods to enrich and amplify the viral signal for more sensitive sequencing. This is being done using viral culture on human nasal epithelial (HNE) cells. Using cultured cells or supernatant will result in a sample that is much more concentrated and less contaminated by host DNA than a clinical sample. Different enrichment methods can serve this same purpose, and may be helpful if an unknown is proving difficult to culture.

In order to ensure that only truly unknown pathogens are being sent for sequencing, additional tests for etiology or coinfection should be done. This may include testing for additional viruses that are not routinely tested for, such as CMV or EBV. If sample preparation procedure includes a DNAse step (as ours did), it is advised that DNA is also isolated from the sample to test for the presence of DNA viruses; while adenovirus is included in most panels, human bocavirus is generally not tested for. In rare cases - for instance if a patient has a severe immunodeficiency- it may also be prudent to test for commonly pathogenic fungi. The overall goal is to be as confident as possible that the pathogen in a sample is truly unknown or divergent, and not one that is already established and simply tested for less frequently.

Conclusion

This study advances the development of a method for efficient screening of respiratory viruses with negative clinical tests for the presence of a possible unknown virus. In one proposed method clinical nasopharyngeal samples that test negative for all viruses included in a standard respiratory virus panel can be reflexed to an ELISA to detect the presence of CXCL10, which indicates an active viral infection. CXCL10-positive samples may then be further evaluated using NGS and other methods. It is currently not feasible nor would it be necessary to perform NGS on all respiratory samples; the CXCL10 assay could serve to screen for only samples most likely to contain truly unknown viruses that would benefit from sequencing or other further workup. Currently virus discovery/identification is carried out in research studies or based on a case study of one or two particular patients. A method to screen all clinical samples for the presence of an unknown virus may be the first step in making sequencing more feasible for

clinical settings. Additionally, this method could be used to streamline the virus discovery process or detection of emerging infections.

References

1. Marciniuk DD, Schraufnagel DE, European Respiratory Society. *The Global Impact of Respiratory Disease*.; 2017.

2. Mokdad AH, ed. Estimates of the global, regional, and national morbidity, mortality, and aetiologies of lower respiratory tract infections in 195 countries: a systematic analysis for the Global Burden of Disease Study 2015. *Lancet Infec Dis*. 2017;17:11333-11361.

3. Thomas M, Bomar P. Upper Respiratory Tract Infection. In: Treasure Island (FL): StatPearls Publishing; 2018.

4. Berry M, Gamieldien J, Fielding B. Identification of New Respiratory Viruses in the New Millennium. *Viruses*. 2015;7(3):996-1019. doi:10.3390/v7030996

5. Graf EH, Simmon KE, Tardif KD, et al. Unbiased Detection of Respiratory Viruses by Use of RNA Sequencing-Based Metagenomics: a Systematic Comparison to a Commercial PCR Panel. Caliendo AM, ed. *J Clin Microbiol*. 2016;54(4):1000-1007. doi:10.1128/JCM.03060-15

6. Datta S. Next-generation sequencing in clinical virology: Discovery of new viruses. *World J Virol*. 2015;4(3):265. doi:10.5501/wjv.v4.i3.265

7. Marston DA, McElhinney LM, Ellis RJ, et al. Next generation sequencing of viral RNA genomes. *BMC Genomics*. 2013;14(1):444. doi:10.1186/1471-2164-14-444

8. Rosenberg R. Detecting the emergence of novel, zoonotic viruses pathogenic to humans. *Cell Mol Life Sci*. 2015;72(6):1115-1125. doi:10.1007/s00018-014-1785-y

9. Krishnamurthy SR, Wang D. Origins and challenges of viral dark matter. *Virus Res*. 2017;239:136-142. doi:10.1016/j.virusres.2017.02.002

10. Hall RJ, Draper JL, Nielsen FGG, Dutilh BE. Beyond research: a primer for considerations on using viral metagenomics in the field and clinic. *Front Microbiol*. 2015;6. doi:10.3389/fmicb.2015.00224

11. Kumar A, Murthy S, Kapoor A. Evolution of selective-sequencing approaches for virus discovery and virome analysis. *Virus Res*. 2017;239:172-179. doi:10.1016/j.virusres.2017.06.005

12. Sridhar S, To KKW, Chan JFW, Lau SKP, Woo PCY, Yuen K-Y. A Systematic Approach to Novel Virus Discovery in Emerging Infectious Disease Outbreaks. *J Mol Diagn*. 2015;17(3):230-241. doi:10.1016/j.jmoldx.2014.12.002

13. Lipkin WI, Firth C. Viral surveillance and discovery. *Curr Opin Virol*. 2013;3(2):199-204. doi:10.1016/j.coviro.2013.03.010

14. Jazaeri Farsani SM, Deijs M, Dijkman R, et al. Culturing of respiratory viruses in well-differentiated pseudostratified human airway epithelium as a tool to detect unknown viruses. *Influenza Other Respir Viruses*. 2015;9(1):51-57. doi:10.1111/irv.12297

15. Epstein JH, Anthony SJ. Viral discovery as a tool for pandemic preparedness: -EN- -FR- La découverte des virus, un outil en appui de la préparation aux pandémies -ES- El descubrimiento de virus como herramienta de preparación para pandemias. *Rev Sci Tech OIE*. 2017;36(2):499-512. doi:10.20506/rst.36.2.2669

16. Landry ML, Foxman EF. Antiviral Response in the Nasopharynx Identifies Patients With Respiratory Virus Infection. *J Infect Dis*. 2018;217(6):897-905. doi:10.1093/infdis/jix648

17. Grard G, Fair JN, Lee D, et al. A Novel Rhabdovirus Associated with Acute Hemorrhagic Fever in Central Africa. Wang D, ed. *PLoS Pathog*. 2012;8(9):e1002924. doi:10.1371/journal.ppat.1002924

18. Greninger AL. A decade of RNA virus metagenomics is (not) enough. *Virus Res*. 2018;244:218-229. doi:10.1016/j.virusres.2017.10.014

19. Chiu CY. Viral pathogen discovery. *Curr Opin Microbiol*. 2013;16(4):468-478. doi:10.1016/j.mib.2013.05.001

20. Law J, Jovel J, Patterson J, et al. Identification of Hepatotropic Viruses from Plasma Using Deep Sequencing: A Next Generation Diagnostic Tool. Shomron N, ed. *PLoS ONE*. 2013;8(4):e60595. doi:10.1371/journal.pone.0060595

21. Dutilh BE, Reyes A, Hall RJ, Whiteson KL. Editorial: Virus Discovery by Metagenomics: The (Im)possibilities. *Front Microbiol*. 2017;8. doi:10.3389/fmicb.2017.01710

22. Lipkin WI. The changing face of pathogen discovery and surveillance. *Nat Rev Microbiol*. 2013;11(2):133-141. doi:10.1038/nrmicro2949

23. Allander T, Tammi MT, Eriksson M, Bjerkner A, Tiveljung-Lindell A, Andersson B. From The Cover: Cloning of a human parvovirus by molecular screening of respiratory tract samples. *Proc Natl Acad Sci*. 2005;102(36):12891-12896. doi:10.1073/pnas.0504666102

24. van den Hoogen B, deJong JC, Groen J, et al. A newly discovered human pneumovirus isolated from young children with respiratory disease. *Nat Med*. 2001;7(6).

25. Zhao Q, Kim T, Pang J, et al. A novel function of CXCL10 in mediating monocyte production of proinflammatory cytokines. *J Leukoc Biol*. 2017;102(5):1271-1280. doi:10.1189/jlb.5A0717-302

26. Joshi D, Carey I, Foxton M, et al. CXCL10 levels identify individuals with rapid fibrosis at 12 months post-transplant for hepatitis C virus and predict treatment response. *Clin Transplant*. 2014;28(5):569-578. doi:10.1111/ctr.12354

27. Valverde-Villegas JM, de Medeiros RM, Ellwanger JH, et al. High CXCL10/IP-10 levels are a hallmark in the clinical evolution of the HIV infection. *Infect Genet Evol*. 2018;57:51-58. doi:10.1016/j.meegid.2017.11.002

28. Romero R, Chaemsaithong P, Chaiyasit N, et al. CXCL10 and IL-6: Markers of two different forms of intra-amniotic inflammation in preterm labor. *Am J Reprod Immunol*. 2017;78(1):e12685. doi:10.1111/aji.12685

29. Kuypers J, Martin ET, Heugel J, Wright N, Morrow R, Englund JA. Clinical Disease in Children Associated With Newly Described Coronavirus Subtypes. *PEDIATRICS*. 2007;119(1):e70-e76. doi:10.1542/peds.2006-1406

30. Liu M, Guo S, Hibbert JM, et al. CXCL10/IP-10 in infectious diseases pathogenesis and potential therapeutic implications. *Cytokine Growth Factor Rev*. July 2011. doi:10.1016/j.cytogfr.2011.06.001

31. Paczesny S, Abu Zaid M. CXCL10: most consistent cGVHD biomarker? *Blood*. 2016;127(24):2950-2951. doi:10.1182/blood-2016-04-709543

Supplemental Tables 1 & 2: *Breakdown of symptoms/comorbidities included in each category from Table 2 and Table 3. Symptoms and comorbidities extracted from patient chart notes.*

Symptoms	
Cardiac	Fibrillation
	Cardiac arrest
	Tachycardia
	Chest pain/pressure
Altered Mental State	Agitation
	Paranoia
	Altered mental state
	Psychosis
	Confusion
	Delirium
Respiratory	Low O2/desaturation
	Pleural effusion
	Respiratory failure/hypoxia
	Asthma
	Chronic Obstructive Pulmonary Disease (COPD)
	Respiratory Distress Syndrome (RDS)
	Pneumonia
	Nasopharyngitis
	Bronchiolitis
	Sore throat
	Sinusitis
	Shortness of breath/dyspnea
	Cough
Fatigue	Fatigue/lethargy
Fever	Fever
	Chills
Other	Hypo/hypertension
	Anemia
	Hypo/hyperglycemia
	Sickle Cell
	Rigors
	Hematuria
	Seizures
	Edema
	Acidosis
	Dehydration
	Abscess/ulcer
	Acute Kidney Injury (AKI)
	Rash
	Planned procedure
	Sepsis
	Lymphocytosis
	Fall
	Gastrointestinal symptoms
	Headache
	Body pain
	Malaise
	Weakness
	Dizziness/syncope

Comorbidities	
Liver/Kidney	Acute/Chronic Renal Failure (A/CRF)
	CKD
	End Stage Renal Disease (ESRD)
	Hepatitis
	Kidney failure
	Cirrhosis
Cardiovascular	Aortic aneurysm
	Coronary Artery Disease (CAD)
	Cardiac arrest
	Cardiomyopathy
	Chronic Heart Failure (CHF)
Respiratory	Cystic fibrosis
	Emphysema
	Interstitial Lung Disease (ILD)
	Pulmonary fibrosis
	Chronic respiratory failure
	Small airways disease
	Pleural effusion
	Pulmonary embolism
	Tuberculosis
	Asthma
	Chronic Obstructive Pulmonary Disease (COPD_
Other	Von Willebrand's Disease
	Grave's disease
	Guillain-Barre syndrome
	Graft vs Host Disease (GVHD)
	Kawasaki's Disease
	Lupus
	Polycystic Ovarian Syndrome (PCOS)
	Sarcoidosis
	Myelodysplastic Syndrome (MDS)
	Sickle cell
	Traumatic Brain Injury (TBI)
	Recent transplant
	Prematurity
	Cerebral palsy
	Parkinson's
	Schizophrenia
	Seizures
	CMV
	Smoker
	Drug/alcohol abuse
	Obesity
Diabetes	