

# SCIENTIFIC REPORTS



OPEN

## A method to identify respiratory virus infections in clinical samples using next-generation sequencing

Talia Kustin<sup>1</sup>, Guy Ling<sup>1</sup>, Sivan Sharabi<sup>2,3</sup>, Daniela Ram<sup>2</sup>, Nehemya Friedman<sup>2,3</sup>, Neta Zuckerman<sup>2</sup>, Efrat Dahan Bucris<sup>2</sup>, Aharon Glatman-Freedman<sup>3,4</sup>, Adi Stern<sup>1</sup> & Michal Mandelboim<sup>2,3</sup>

Respiratory virus infections are very common. Such infections impose an enormous economic burden and occasionally lead to death. Furthermore, every few decades, respiratory virus pandemics emerge, putting the entire world population at risk. Thus, there is an urgent need to quickly and precisely identify the infecting agent in a clinical setting. However, in many patients with influenza-like symptoms (ILS) the identity of the underlying pathogen remains unknown. In addition, it takes time and effort to individually identify the virus responsible for the ILS. Here, we present a new next-generation sequencing (NGS)-based method that enables rapid and robust identification of pathogens in a pool of clinical samples without the need for specific primers. The method is aimed at rapidly uncovering a potentially common pathogen affecting many samples with an unidentified source of disease.

Respiratory virus infections are very common around the world, especially during the winter season. Such infections impose an enormous economic burden and can lead to death, especially among the elderly. Respiratory virus infections are very easily transmitted from one person to another and cause a global pandemic every few decades<sup>1,2</sup>. Thus, to prevent their rapid spread and minimize mortality rates, it is essential to diagnose infection as soon as possible.

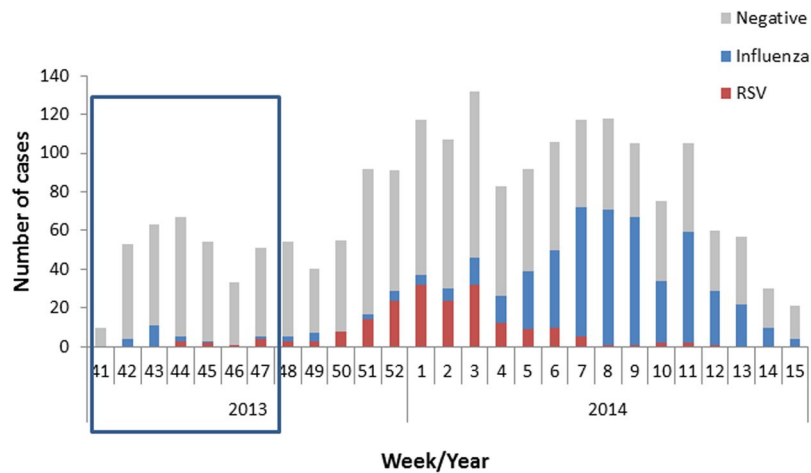
Several laboratory methods, such as enzyme-linked immunosorbent assay, nucleic acid hybridization, and polymerase chain reaction (PCR), are utilized to detect respiratory virus infections<sup>3</sup>. However, these assays are highly dependent on previous knowledge of the virus, and thus are ineffective in identifying new pathogens and are less effective in identifying novel variants of a known pathogen<sup>4</sup>. Mainly, these methods rely on knowledge about the viral sequence at hand, and hence relevant nucleic acid probes (primers) are used. Accordingly, rare viruses that are not commonly examined in the laboratory, will not be detected using these conventional methods.

Next-generation sequencing (NGS) technologies enable simultaneous sequencing of large numbers of samples, and can be used for the unbiased detection and characterization of multiple agents in a single sample. Use of NGS eliminates the need for prior knowledge of virus genomic sequences and provides advantages over traditional methods such as PCR amplification or microarray hybridization, that are dependent on target-specific primers<sup>3</sup>. To date, NGS has been instrumental in the discovery of novel viruses and characterization of viral communities. NGS was used to discover a new arenavirus<sup>5</sup>, a new ebola virus<sup>6</sup> and zika virus<sup>7</sup>. NGS is also applied for the characterization of viruses in the environment<sup>8,9</sup>, in animals<sup>10</sup> and in humans<sup>11,12</sup>. Here, we describe a new NGS-based method designed to identify viruses in a pool of clinical samples.

### Materials and Methods

**Patients and samples.** As part of community influenza surveillance, conducted in collaboration with the Israel Center for Disease Control (ICDC), clinical combined nose throat swab samples were collected from 300 patients presenting with influenza-like illness (ILI), during the winter season spanning between October 2013 and December 2013. The samples were collected using  $\Sigma$ -Virocult<sup>®</sup> M40-A Compliant kit (MWE, UK),

<sup>1</sup>School of Molecular Cell Biology and Biotechnology, George S. Wise Faculty of Life Sciences Tel Aviv University, Tel Aviv, Israel. <sup>2</sup>Central Virology Laboratory, Ministry of Health, Chaim Sheba Medical Center, Tel-Hashomer, Ramat-Gan, Israel. <sup>3</sup>Department of Epidemiology and Preventive Medicine, School of Public Health, Sackler Faculty of Medicine, Tel-Aviv University, Tel-Aviv, Israel. <sup>4</sup>The Israel Center for Disease Control, Israel Ministry of Health, Tel-Hashomer, Israel. Adi Stern and Michal Mandelboim contributed equally. Correspondence and requests for materials should be addressed to A.S. (email: [sternadi@post.tau.ac.il](mailto:sternadi@post.tau.ac.il)) or M.M. (email: [michalman@sheba.health.gov.il](mailto:michalman@sheba.health.gov.il))



**Figure 1.** RSV and influenza infection throughout the winter season. The number of clinical samples infected with influenza virus or RSV in the 2013–2014 winter season in Israel. The week number and year are shown on the X axis. The 300 samples addressed in this paper are highlighted in the box.

suspended in 2 ml M40 medium, and stored up to 48 hours in 4 °C. Following preparation, samples were stored in –70 °C. 500 microliters were used in the first preparation and the rest was kept in –70 °C. Then, for step 3 (Fig. 2), we used the original 54 samples and controls that were stored in –70 °C.

**Ethical considerations.** Community influenza surveillance, including combined nose throat samples, is performed under the Public Health Ordinance enacted in Israel. The IRB of the Sheba Medical Center approved the research: Helsinki Number 4379-17-SMC. Informed consent is not required.

**Nucleic acid extraction and qRT-PCR.** Total nucleic acid content was extracted from 500 microliters of the Virocult medium using NucliSENS easyMAG (BioMerieux, France). The presence of respiratory viruses was determined by TaqMan Chemistry using the ABI 7500 instrument. As part of community influenza surveillance, RNA viruses detection was performed using a panel of real-time reverse transcription-polymerase chain reaction (rRT-PCR) assays, as previously described: influenza A and B<sup>13</sup>, influenza A(H1N1)pdm09<sup>14,15</sup> and respiratory syncytial virus (RSV)<sup>16</sup>. Then, samples were tested for presence of: human metapneumovirus<sup>17</sup>, enterovirus<sup>18</sup>, rhinovirus<sup>19</sup>, parainfluenza 3<sup>20</sup> and coronaviruses<sup>21</sup>. RT-PCR was performed to detect DNA virus adenovirus<sup>22</sup> and bocavirus<sup>23</sup>. For the RNA rRT-PCR assays, the Ambion Ag-Path master mix (Life Technologies, Carlsbad, CA) was used, whereas for the DNA assays, ABgene Absolute Blue (Thermo, UK) was used. Both were used according to the manufacturer's instructions.

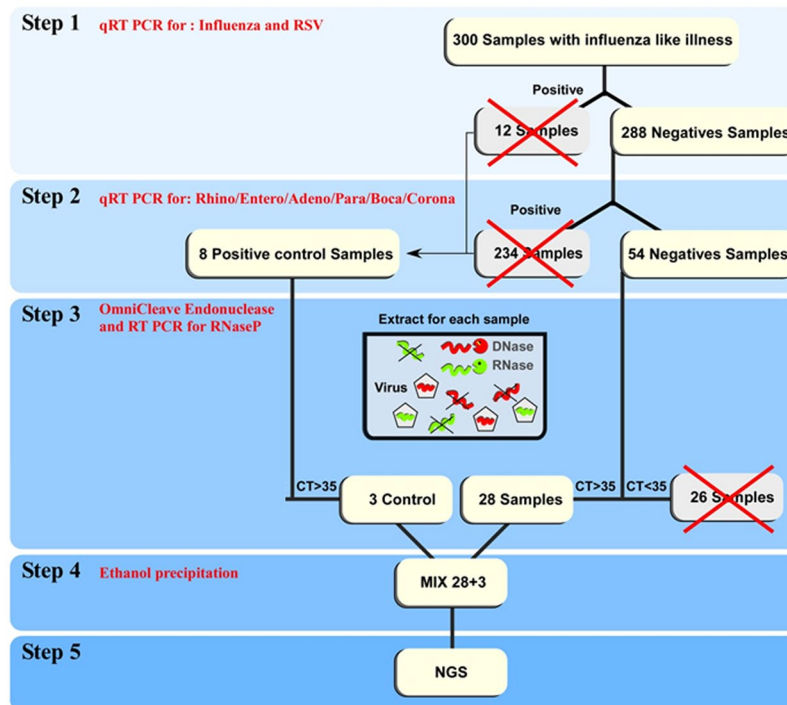
Torque teno virus (TTV) was detected by qRT-PCR<sup>24</sup>, and parainfluenza 1 virus was detected by RT-PCR, as previously described<sup>25</sup>. Cucumber green mottle mosaic virus (CGMMV) was detected using the following primers: forward 5'ATGGCAAACATTAATGAACAAATCAA3' and reverse 5'TCTATCTGGAAAGTTGGAAGAGGTC3'

The bacteria *Streptococcus pneumoniae* was detected by PCR as previously described<sup>26</sup> and using the following primers: forward 5'AAATGCGCGGTGAAGCAAAGG3' and reverse 5'GACCAGTAGCAGCTTGGAACG3'. *Staphylococcus aureus* was detected by PCR as previously described<sup>27</sup> and using the following primers: forward 5'CAAAAGACTGGGTAAACGTCGC3' and reverse 5'CGGTCCGTTTGCATTGCAAATGG3'. *Pseudomonas putida* was detected by PCR as previously described<sup>28</sup> and using the following primers: forward 5'GACAGTCGCTG CCTGTTTTTG3' and reverse 5'AGTCGATCTGTAAGCCGGTTTTG3'. Influenza C virus<sup>29</sup> and Parainfluenza 4 virus<sup>30</sup> were detected by RT-PCR on the samples after the second run of RNA library.

**Sample preparation.** 200 microliters of the original samples that were stored in –70 °C in the Virocult medium were freeze-thawed three times and then centrifuged at high speed for 20 minutes. OmniCleave Endonuclease (250U; Epicenter Distributor Madison, WI USA, OC7850K) was added to the supernatants in the presence of 2.5 mM MgCl<sub>2</sub> for 1 h at 37 °C. The genome was extracted from the samples using NucliSENS easyMAG. RNaseP content, quantified by qRT-PCR, served as an indicator of the amount of human genome remaining in the sample<sup>31</sup>. All samples in which the RNaseP was above 35 Ct qRT-PCR were pooled together.

**Library preparation and sequencing.** The Agilent RNA 6000 Nano Kit (Life Technologies, Waldbronn, Germany) was used to evaluate total RNA quality. RNA concentration was measured using the Qubit RNA BR Assay kit (Life Technologies). The library was processed for NGS sample preparation using SMARTer Stranded RNA-Seq Kits (Clontech, a Takara Bio Company, CA, USA), Strand-Specific Library Construction for Transcriptome Analysis on Illumina Platforms.

The Agilent High Sensitivity DNA Kit with the Agilent 2100 Bioanalyzer (Life Technologies, Waldbronn, Germany) was used to size, quantify and quality control DNA sequencing libraries. NGS was performed using a MiSeq v2 kit (500 cycles) (Illumina, San Diego, CA). After automated cluster generation in MiSeq, the sequencing was processed and genomic sequence reads were obtained.



**Figure 2.** Schematic presentation of the workflow of the described method.

For the detection of RNA viruses, sample extracts were pretreated with RNase free DNase (promega), the Agilent RNA 6000 Nano Kit (Life Technologies, Waldbronn, Germany) was used to evaluate total RNA quality. RNA concentration was measured using the Qubit RNA HS Assay kit (Life Technologies). RNA libraries were constructed for NGS using SMARTer Stranded RNA-Seq Kit according to manufacturer instructions (Clontech, a Takara Bio Company, CA, USA).

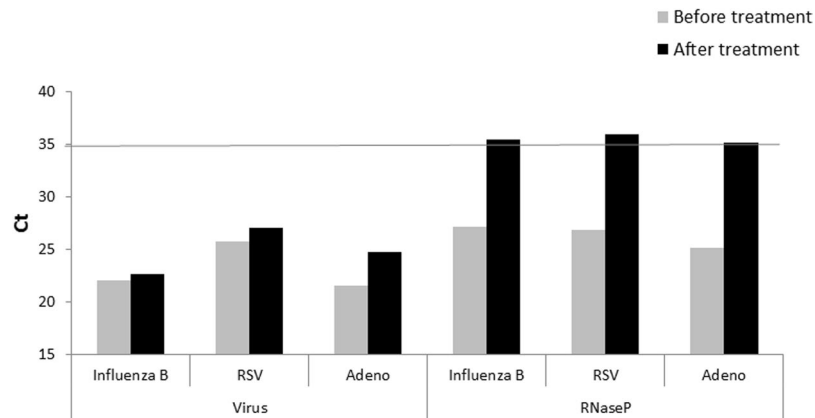
For DNA viruses, sample extracts were pretreated with RNase (JGI), DNA concentration was measured using the Qubit DNA HS Assay kit. DNA libraries were constructed using Nextera XT kit (Illumina, San Diego, CA) according to manufacturer instructions. RNA and DNA libraries size distribution were analyzed using Tape Station 2.0 (Agilent). Libraries were normalized according to the median fragment size measured by Tape Station and library concentration measured by Qubit. Sequencing was processed with Miseq v3 kit (600 cycles) (Illumina, San Diego, CA). After automated cluster generation, the genomic sequence reads (FASTQ) files were obtained for further analysis.

**Computational analysis of NGS data.** Bowtie 2.0<sup>32</sup> was used to filter out reads that mapped either to: (a) the human genome (hg19) or to the human transcriptome, both downloaded from Ensembl ([www.ensembl.org](http://www.ensembl.org)), or (b) to bacterial genomes listed in the human microbiome project (downloaded from <https://hmpdacc.org/>). Remaining reads were queried using Bowtie 2.0 against a virus-only database (downloaded from NCBI - <ftp://ftp.ncbi.nih.gov/refseq/release/viral/>) and the hits were validated with blastn and blastx. We omitted reads that mapped to viruses with bowtie2 but did not map to the same virus with blast. In order to examine reads that did not map to any database described above, we used the Velvet de-novo assembly program<sup>33</sup> with default settings to create larger “contigs” of data. Finally, contigs were queried against the nt (non-redundant nucleotide) and nr (non-redundant protein) databases at NCBI, using blastn and blastx, respectively, with default parameters.

## Results

**Sample selection and work outline.** In the beginning of the winter season, the identity of the viruses causing ILS, other than RSV and influenza, was unknown (Fig. 1). Later in the winter two epidemic waves were noticed, RSV and then influenza. However, a higher percentage of patients still suffered from a respiratory infection of unknown etiology, with clinical symptoms compatible with viral infection (Fig. 1). We therefore set out to develop a general and robust approach to identify the unknown viruses responsible for ILS. Three possibilities were considered: (1) The cause of the ILS is an unknown virus or a known virus not yet associated with respiratory illnesses; (2) the cause of the ILS is a virus not commonly tested in the laboratory; (3) the cause of the ILS is a known virus that causes respiratory infection that cannot be identified in the routine RT PCR tests due to variability.

The workflow of the approach is shown in Fig. 2. Initially, qRT-PCR for influenza and RSV was performed on 300 samples obtained from the beginning of the 2013–2014 winter season (Fig. 1); 12 samples were positive. The remaining 288 samples (Fig. 2), were then assayed by qRT-PCR for: rhinovirus, enterovirus, adenovirus, parainfluenza 3, bocavirus and coronaviruses; 234 samples were found to be positive (Fig. 2). In the next step, the



**Figure 3.** qRT-PCR for viruses and RNaseP. Ct results before and after Omnicleave treatment for the 3 controls tested viruses and for RNaseP.

remaining 54 samples and 8 positive control samples (2 influenza A, 2 influenza B, 2 RSV and 2 adenovirus) were treated with OmniCleave, which destroys free DNA and RNA, leaving the viral genome unaffected, as it is protected by a capsid. qRT-PCR of all samples for the cellular gene RNaseP was performed. Following treatment, the control virus amount remained unchanged, while RNaseP levels were reduced (resulting in higher cycle threshold (Ct) Fig. 3). The treatment efficiency threshold was set at Ct RNaseP > 35. We selected Ct > 35 since this indicates low levels of human genome that will not interfere with the viral identification. A level of Ct > 35 is equivalent to less than one copy of human DNA. The Ct of all control viruses, but two, was above 35. In one influenza B sample the Ct was 34.01 and in one adenovirus sample the Ct was 34.81. For the mixed sequencing we took one sample of each control virus.

Following this, 28 clinical samples were left in addition to 3 control samples (Influenza B, RSV and adenovirus) (Fig. 2). The samples were pooled and NGS was performed (Fig. 2).

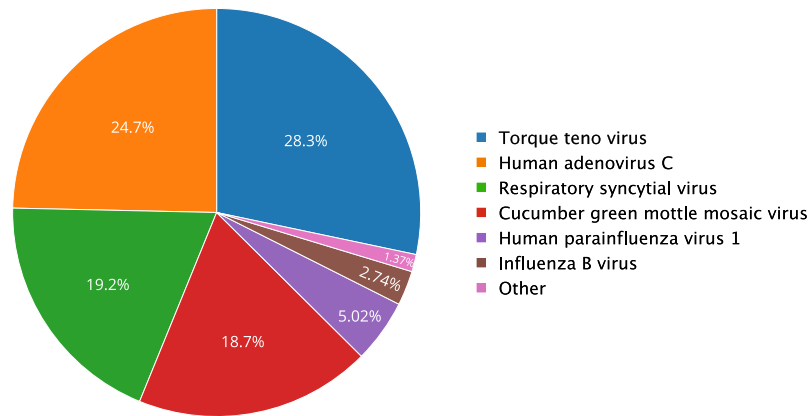
**Next-generation sequencing of the pooled samples.** The resulting NGS data consisted of ~9 M paired-end reads of 150 bp DNA fragments. Reads similar to human or known bacterial genomes were filtered out, and the remaining reads were tested for similarity to viruses from known databases, as previously done<sup>11,34</sup>. Among the 9M reads, 27% found to be human genomic material, and ~61% mapped to bacterial genomes were filtered out, leaving ~1.1 million reads to be further analyzed. When applying Bowtie 2.0 to match the remaining reads to known viruses, approximately 34,000 mapped to known viruses, with 29,000 reads mapping to phi X 174 bacteriophage, which is a control sample of Illumina sequencing. The remaining ~1 million unmapped reads were assembled into larger segments (contigs). Assuming that reads from a pathogen will be abundant enough in the sample to create somewhat long contigs, 8387 contigs with more than 350 nucleotides were detected and used to query the non-redundant nucleotide and protein databases in NCBI. The vast majority of contigs mapped to a bacterial or a phage source. The remaining viral-mapped contigs were mapped to influenza B and Torque teno virus (TTV) and one contig mapped to human picobirnavirus. There was a small number of contigs that mapped to plant, fungi, animal and protists.

**Detection of known viruses.** We next focused on the 5,000 reads that mapped to known viruses. Of these, ~2000 mapped to bacterial phages and ~3000 to other viruses. From the latter reads, three viruses (adenovirus, RSV and influenza B) that were included as positive samples in our pooled sample, were detected (Fig. 4). Notably, the average coverage of the positive control viruses was quite low: sixteen reads per base for adenovirus and one read per base for RSV and influenza B. Additional viruses identified included parainfluenza virus 1, TTV, a virus which is frequently present in humans, and a cucumber green mottle mosaic virus (CGMMV), present in vegetables like cucumber and watermelon and is not known to cause illness in humans (Fig. 4 and Table 1).

To validate that the newly identified viruses were indeed present in the tested samples and to determine precisely which sample was infected with each of these viruses, qRT-PCR was then performed on the individual samples. As can be seen in Fig. 5, TTV was detected in 23 of 31 samples (20 unidentified samples and the 3 control samples). TTV titer was high in two unidentified samples and low in the rest of the samples detected.

The control viruses that were included in the pooled assay sample were identified only in the control samples (Fig. 5). One patient (number 10, Fig. 5) was infected with parainfluenza 1 virus and three patients with CGMMV, that is not known to cause illness in humans (Fig. 5). Human picobirnavirus was detected in one of the contigs assembled. This virus causes gastroenteritis symptoms and is not associated with ILS<sup>35</sup>.

**Detection of possible bacterial agents.** Although our study is focused on patients with viral symptoms rather than bacterial symptoms it is reasonable to assume that some of the patients were infected with bacterial agents. Our sequencing resulted in ~5.4 million reads that were mapped to bacteria. 50% of the bacterial reads were mapped to the *Pseudomonas* genus, which is mostly non-pathogenic and if pathogenic it is usually due to opportunistic infection<sup>36</sup>. The other genus detected with high prevalence was *Streptococcus* with 22% of the reads.



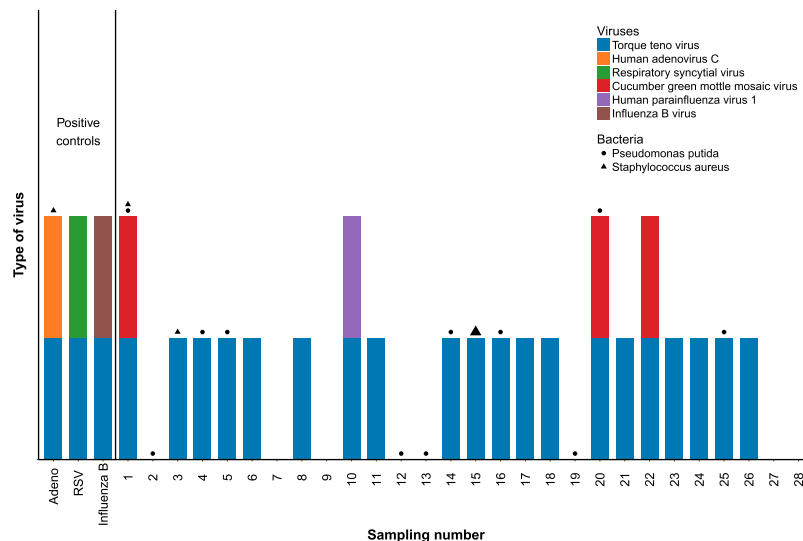
**Figure 4.** Viruses identified by NGS in 2013–2014 pooled clinical ILS samples. The chart includes viruses with more than 30 reads mapped to them, and read counts per virus are normalized by genome length of the virus.

Virus	Number of mapped reads
Human adenovirus C	1904
Respiratory syncytial virus	627
Cucumber green mottle mosaic virus	258
Human parainfluenza virus 1	161
Torque teno virus	122
Torque teno mini virus	78
Influenza B virus	78
Torque teno midi virus	12
Moloney murine leukemia virus	6
Apis mellifera filamentous virus	5
Human herpesvirus 7	5
Emiliana huxleyi virus 86	4
Megavirus chiliensis	3
Prunus necrotic ringspot virus	2
Tomato brown rugose fruit virus	2
Tobacco mosaic virus	2
Human herpesvirus 4	2
Bovine respiratory syncytial virus	2
Avian leukosis virus - RSA	1
Chrysoideixis chalcites nucleopolyhedrovirus	1
Fowlpox virus	1
Human papillomavirus type 5	1
Human papillomavirus type 9	1
Variola virus	1

**Table 1.** Number of reads for each virus. Number of mapped reads for each virus discovered in the next generation sequencing.

Some species from this genus such as *Streptococcus pneumoniae* (216 K reads) can cause disease with influenza like symptoms. Another possible causative agent of ILS that was detected is *Staphylococcus aureus* that had 170 K reads mapped to it. It is important to acknowledge that the bacterial results obtained here do not represent the full repertoire of bacteria in the samples due to the OmniCleave treatment that destroyed most of the bacterial genomes.

To determine which samples were infected with each of these bacterial strains we performed PCR in the individual samples. *Pseudomonas putida*, which was the most prevalent *Pseudomonas* species found, was present in 12 samples in low quantities; *Staphylococcus aureus* was detected with a strong signal in one sample (patient 15, Fig. 5) and with a weak signal in two other samples and in the adenovirus positive control. *Streptococcus pneumoniae* was not detected in our samples (Fig. 5). These results imply that at least in one patient (sample 15) the ILS symptoms may have been a result of bacterial infection by *Staphylococcus aureus*. It is quite possible that there are more bacterial strains that cause ILS symptoms in our samples that were not detected.



**Figure 5.** Verification of virus and bacteria presence in each of the samples. The presence of each of the viruses identified (Fig. 4) was verified in each sample by qRT-PCR/PCR. Viruses are shown in the colored bars and bacteria are shown with the shaped symbols, the size of the symbol corresponds with the strength of detection.

**Improvement of the method.** We were concerned that our initial  $>35$  cutoff for sample inclusion might have been too stringent so we performed a second round of NGS sequencing, this time including all 54 negative samples. We split the pooled samples in two and created separate samples for RNA viruses and DNA viruses (using treatments of DNase and RNase respectively). We did this to maximize library quality and the probability for viral detection. The sequencing results for the RNA viruses sample revealed three additional viruses that were not detected in the first sequencing. Those viruses are parainfluenza 4 virus and influenza C virus that cause ILS symptoms and rotavirus A which causes diarrhoeal disease. Parainfluenza 4 and influenza C were identified by RT-PCR in two clinical samples (parainfluenza 4 in patient number 49 and influenza C in patient 30). The DNA viruses sample didn't reveal presence of previously undetected viruses.

## Discussion

Infections with respiratory viruses are common and affect practically the entire population. Such infections impose a heavy economic burden, due to loss of working days, and they can be life-threatening. Thus, it is critical to monitor such infections from an early stage. However, occasionally the virus responsible for the infection cannot be identified. For example, we show here, that at the start of our screen, no common respiratory virus could be identified in 54 out of 300 (18%) samples collected from patients with ILS. Furthermore, new methods are needed to rapidly detect viruses present in a pool of samples to enable rapid treatment of the population in case of pandemics. Here, we presented a new NGS-based method to identify the viruses present in these samples.

In this method, all unknown samples were pooled with positive controls, enabling virus identification in a single experiment, rendering the approach rapid, robust, efficient and economic. It takes less than one day to perform the qRT-PCR and one day to generate the library for sequencing. The bioinformatics analysis runs for a few hours. Next, the presence of the identified virus in each of the samples can be rapidly determined using qRT-PCR.

Out of 54 samples that were sequenced, our method discovered infection with human parainfluenza 1 virus, human parainfluenza 4 virus and influenza C virus, each in one of the patients. The original clinical diagnosis of our samples was infection with a viral agent, based on their symptoms. Nevertheless, our method also detected a sample that was most likely infected with a bacterial agent rather than a viral one. Interestingly, TTV was identified in many samples. TTV is the first human virus identified to have a circular negative-stranded DNA genome<sup>37</sup>. Its genome is approximately 3.8k nucleotides (nt) long<sup>38</sup>. TTV frequently circulates in many geographic regions<sup>24,38,39</sup>. Although this virus has a very high prevalence in the general population, neither its interaction with its hosts nor its direct involvement in the etiology of specific diseases are fully understood. Nevertheless, a significant correlation between TTV loads and airflow limitation within the peripheral airways, as well as between severity of bronchiectasis and decrease of lung function, has been observed<sup>40</sup>. Thus, it is possible that TTV was the cause of the ILS. Further studies will be necessary to confirm that TTV is indeed responsible for the ILS and to understand why TTV is present in some individuals (even in the healthy) but not in others.

Detection of CGMMV demonstrates that the method can identify viruses present in human diet providing further indication of the validity and accuracy of the presented method.

The virus responsible for the infection could not be certainly identified in 51 samples. The following hypotheses may be raised: (a) the source of disease was bacterial rather than viral, as suggested herein for one sample that may have been infected with *Staphylococcus aureus*. If this is indeed the case then for the other unidentified samples, the bacterial DNA may have been destroyed when using the OmniCleave endonuclease (see Methods), (b) the source of the disease was TTV, (c) the source of the disease was viral but due to low virus titer it was

undetectable. Accordingly, we suggest in the future improving the sample preparation process and striving for higher sequencing depth when aiming to discover a new virus or a novel virus variant.

In summary, we present a rapid and efficient method for detection of novel, uncommon and non-tested pathogens associated with disease that might be useful in a clinical setting. With the advancement of NGS technology, we expect this approach to become increasingly useful.

## References

- Ross, A. G., Crowe, S. M. & Tyndall, M. W. Planning for the Next Global Pandemic. *Int J Infect Dis* **38**, 89–94, <https://doi.org/10.1016/j.ijid.2015.07.016> (2015).
- Welfare, W. & Wright, E. Planning for the unexpected: Ebola virus, Zika virus, what's next? *Br J Hosp Med (Lond)* **77**, 704–707, <https://doi.org/10.12968/hmed.2016.77.12.704> (2016).
- Mackay, I. M., Arden, K. E. & Nitsche, A. Real-time PCR in virology. *Nucleic Acids Res* **30**, 1292–1305 (2002).
- Bronze, M. S., Huycke, M. M., Machado, L. J., Voskuhl, G. W. & Greenfield, R. A. Viral agents as biological weapons and agents of bioterrorism. *Am J Med Sci* **323**, 316–325, doi:S0002-9629(15)34447-5 (2002).
- Palacios, G. *et al.* A new arenavirus in a cluster of fatal transplant-associated diseases. *N Engl J Med* **358**, 991–998, <https://doi.org/10.1056/NEJMoa073785> (2008).
- Towner, J. S. *et al.* Newly discovered ebola virus associated with hemorrhagic fever outbreak in Uganda. *PLoS Pathog* **4**, e1000212, <https://doi.org/10.1371/journal.ppat.1000212> (2008).
- Quick, J. *et al.* Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat Protoc* **12**, 1261–1276, <https://doi.org/10.1038/nprot.2017.066> (2017).
- Bibby, K., Viau, E. & Peccia, J. Viral metagenome analysis to guide human pathogen monitoring in environmental samples. *Letts Appl Microbiol* **52**, 386–392, <https://doi.org/10.1111/j.1472-765X.2011.03014.x> (2011).
- Wong, K., Fong, T. T., Bibby, K. & Molina, M. Application of enteric viruses for fecal pollution source tracking in environmental waters. *Environ Int* **45**, 151–164, <https://doi.org/10.1016/j.envint.2012.02.009> (2012).
- Ge, X. *et al.* Metagenomic analysis of viruses from bat fecal samples reveals many novel viruses in insectivorous bats in China. *J Virol* **86**, 4620–4630, <https://doi.org/10.1128/JVI.06671-11> (2012).
- Nakamura, S. *et al.* Direct metagenomic detection of viral pathogens in nasal and fecal specimens using an unbiased high-throughput sequencing approach. *PLoS One* **4**, e4219, <https://doi.org/10.1371/journal.pone.0004219> (2009).
- Yozwiak, N. L. *et al.* Virus identification in unknown tropical febrile illness cases using deep sequencing. *PLoS Negl Trop Dis* **6**, e1485, <https://doi.org/10.1371/journal.pntd.0001485> (2012).
- Hindiyeh, M. *et al.* Evaluation of BioStar FLU OIA assay for rapid detection of influenza A and B viruses in respiratory specimens. *J Clin Virol* **17**, 119–126, doi:S1386653200000810 (2000).
- Hindiyeh, M. *et al.* Rapid detection of influenza A pandemic (H1N1) 2009 virus neuraminidase resistance mutation H275Y by real-time reverse transcriptase PCR. *J Clin Microbiol* **48**, 1884–1887, <https://doi.org/10.1128/JCM.02540-09> (2010).
- Itoh, Y. *et al.* *In vitro* and *in vivo* characterization of new swine-origin H1N1 influenza viruses. *Nature* **460**, 1021–1025, <https://doi.org/10.1038/nature08260> (2009).
- Hu, A., Colella, M., Tam, J. S., Rappaport, R. & Cheng, S. M. Simultaneous detection, subgrouping, and quantitation of respiratory syncytial virus A and B by real-time PCR. *J Clin Microbiol* **41**, 149–154 (2003).
- Regev, L. *et al.* Characterization of human metapneumovirus infections in Israel. *J Clin Microbiol* **44**, 1484–1489, <https://doi.org/10.1128/JCM.44.4.1484-1489.2006> (2006).
- Verstrepen, W. A., Kuhn, S., Kockx, M. M., Van De Vyvere, M. E. & Mertens, A. H. Rapid detection of enterovirus RNA in cerebrospinal fluid specimens with a novel single-tube real-time reverse transcription-PCR assay. *J Clin Microbiol* **39**, 4093–4096, <https://doi.org/10.1128/JCM.39.11.4093-4096.2001> (2001).
- Lu, X. *et al.* Real-time reverse transcription-PCR assay for comprehensive detection of human rhinoviruses. *J Clin Microbiol* **46**, 533–539, doi:JCM.01739-07.1128/JCM.01739-07 (2008).
- Hu, A. *et al.* Development of a real-time RT-PCR assay for detection and quantitation of parainfluenza virus 3. *J Virol Methods* **130**, 145–148, <https://doi.org/10.1016/j.jviromet.2005.06.014> (2005).
- Gerna, G. *et al.* Human respiratory coronavirus HKU1 versus other coronavirus infections in Italian hospitalised patients. *J Clin Virol* **38**, 244–250, <https://doi.org/10.1016/j.jcv.2006.12.008> (2007).
- Heim, A., Ebnet, C., Harste, G. & Pring-Akerblom, P. Rapid and quantitative detection of human adenovirus DNA by real-time PCR. *J Med Virol* **70**, 228–239, <https://doi.org/10.1002/jmv.10382> (2003).
- Hindiyeh, M. Y. *et al.* High rate of human bocavirus and adenovirus coinfection in hospitalized Israeli children. *J Clin Microbiol* **46**, 334–337, <https://doi.org/10.1128/JCM.01618-07> (2008).
- Tokita, H. *et al.* High TT virus load as an independent factor associated with the occurrence of hepatocellular carcinoma among patients with hepatitis C virus-related chronic liver disease. *J Med Virol* **67**, 501–509, <https://doi.org/10.1002/jmv.10129> (2002).
- Echevarria, J. E., Erdman, D. D., Swierkosz, E. M., Holloway, B. P. & Anderson, L. J. Simultaneous detection and identification of human parainfluenza viruses 1, 2, and 3 from clinical samples by multiplex PCR. *J Clin Microbiol* **36**, 1388–1391 (1998).
- Stralin, K., Backman, A., Holmberg, H., Fredlund, H. & Olcen, P. Design of a multiplex PCR for *Streptococcus pneumoniae*, *Haemophilus influenzae*, *Mycoplasma pneumoniae* and *Chlamydia pneumoniae* to be used on sputum samples. *APMIS* **113**, 99–111, <https://doi.org/10.1111/j.1600-0463.2005.apm1130203.x> (2005).
- Alarcon, B., Vicedo, B. & Aznar, R. PCR-based procedures for detection and quantification of *Staphylococcus aureus* and their application in food. *J Appl Microbiol* **100**, 352–364, <https://doi.org/10.1111/j.1365-2672.2005.02768.x> (2006).
- Yamamoto, S. & Harayama, S. PCR Amplification and Direct Sequencing of *gyrB* Genes with Universal Primers and Their Application to the Detection and Taxonomic Analysis of *Pseudomonas putida* Strains. *Appl Environ Microbiol* **61**, 3768 (1995).
- Pabbaraju, K. *et al.* Detection of influenza C virus by a real-time RT-PCR assay. *Influenza Other Respir Viruses* **7**, 954–960, <https://doi.org/10.1111/irv.12099> (2013).
- Aguilar, J. C. *et al.* Detection and identification of human parainfluenza viruses 1, 2, 3, and 4 in clinical samples of pediatric patients by multiplex reverse transcription-PCR. *J Clin Microbiol* **38**, 1191–1195 (2000).
- Singh, A. K. *et al.* Viral aetiology of acute lower respiratory tract illness in hospitalised paediatric patients of a tertiary hospital: one year prospective study. *Indian J Med Microbiol* **32**, 13–18, <https://doi.org/10.4103/0255-0857.124288> (2014).
- Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359, <https://doi.org/10.1038/nmeth.1923> (2012).
- Zerbino, D. R. & Birney, E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* **18**, 821–829, <https://doi.org/10.1101/gr.074492> (2008).
- Finkbeiner, S. R. *et al.* Metagenomic analysis of human diarrhea: viral detection and discovery. *PLoS Pathog* **4**, e1000011, <https://doi.org/10.1371/journal.ppat.1000011> (2008).
- Yoshinaga, T. A morphological study on the mechanism of diazoxide-hyperglycemia (clinical and experimental investigation). *Endokrinologie* **54**, 410–422 (1969).

36. Fernandez, M. *et al.* Analysis of the pathogenic potential of nosocomial *Pseudomonas putida* strains. *Front Microbiol* **6**, 871, <https://doi.org/10.3389/fmicb.2015.00871> (2015).
37. Hino, S. TTV, a new human virus with single stranded circular DNA genome. *Rev Med Virol* **12**, 151–158, <https://doi.org/10.1002/rmv.351> (2002).
38. Hsiao, K. L., Wang, L. Y., Lin, C. L. & Liu, H. F. New Phylogenetic Groups of Torque Teno Virus Identified in Eastern Taiwan Indigenous. *PLoS One* **11**, e0149901, <https://doi.org/10.1371/journal.pone.0149901> (2016).
39. Peng, Y. H. *et al.* Analysis of the entire genomes of thirteen TT virus variants classifiable into the fourth and fifth genetic groups, isolated from viremic infants. *Arch Virol* **147**, 21–41 (2002).
40. Pifferi, M. *et al.* High torquetenovirus loads are correlated with bronchiectasis and peripheral airflow limitation in children. *Pediatr Infect Dis J* **25**, 804–808, <https://doi.org/10.1097/01.inf.0000232723.58355.f400006454-200609000-00010> (2006).

### Acknowledgements

This study was supported in part by a fellowship from the Edmond J. Safra Center for Bioinformatics at Tel-Aviv University [to T.K. and G.L.].

### Author Contributions

T.K. and G.L. performed the N.G.S. analysis, S.S. performed the molecular biology experiments, D.R., N.Z. and E.D.B. prepared the RNA and DNA libraries, N.F. helped in the molecular biology experiments, A.G.-F. responsible for the influenza survey, T.K., A.S. and M.M. wrote the paper and A.S. and M.M. supervised the project.

### Additional Information

**Competing Interests:** The authors declare no competing interests.

**Publisher's note:** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2019