The use of next generation sequencing in the diagnosis and typing of respiratory infections

Fiona Thorburna,*, Susan Bennettb, Sejal Modhaa, David Murdochc, Rory Gunsonb, Pablo R. Murciaa

a MRC-University of Glasgow Centre for Virus Research, Garscube Campus, 464 Bearsden Road, Glasgow, G61 1QH, UK
b West Of Scotland Specialist Virology Centre, Glasgow Royal Infirmary, Alexandra Parade, Glasgow, G31 2ER, UK
c Department of Pathology, University of Otago, Christchurch, New Zealand

ARTICLE INFO
Article history:
Received 22 April 2015
Received in revised form 2 June 2015
Accepted 15 June 2015

Keywords:
Next generation sequencing
Virus detection:
Viral respiratory infection
Diagnostics

ABSTRACT

Background: Molecular assays are the gold standard methods used to diagnose viral respiratory pathogens. Pitfalls associated with this technique include limits to the number of targeted pathogens, the requirement for continuous monitoring to ensure sensitivity/specificity is maintained and the need to evolve to include emerging pathogens. Introducing target independent next generation sequencing (NGS) could resolve these issues and revolutionise respiratory viral diagnostics.

Objectives: To compare the sensitivity and specificity of target independent NGS against the current standard diagnostic test.

Study design: Diagnostic RT-PCR of clinical samples was carried out in parallel with target independent NGS. NGS sequences were analyzed to determine the proportion with viral origin and consensus sequences were used to establish viral genotypes and serotypes where applicable.

Results: 89 nasopharyngeal swabs were tested. A viral pathogen was detected in 43% of samples by NGS and 54% by RT-PCR. All NGS viral detections were confirmed by RT-PCR.

Conclusions: Target independent NGS can detect viral pathogens in clinical samples. Where viruses were detected by RT-PCR alone the Ct value was higher than those detected by both assays, suggesting an NGS detection cut-off – Ct = 32. The sensitivity and specificity of NGS compared with RT-PCR was 78% and 80% respectively. This is lower than current diagnostic assays but NGS provided full genome sequences in some cases, allowing determination of viral subtype and serotype. Sequencing technology is improving rapidly and it is likely that within a short period of time sequencing depth will increase in-turn improving test sensitivity.

© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

1. Background

Virus specific molecular assays such as real-time PCR (RT-PCR) are now considered the gold standard in the diagnosis of viral respiratory tract infections. They are rapid, relatively inexpensive and offer increased sensitivity and specificity over prior techniques such as virus culture and direct immunofluorescence. Assays can be developed quickly to detect novel/emerging pathogens and can be combined to identify multiple microbiological pathogens in a single test. Yet there is a limit to the number of targets, usually up to four, which can be included in an in-house test before compromising test sensitivity. As a result, diagnostic laboratories must develop a panel of multiplex tests in order to detect the whole range of pathogens. Also, as for all PCR based assays, detection is based on targeting conserved regions of the pathogen genome and mutations can lead to reduced sensitivity or false negative results. Furthermore, only the targeted pathogens included in the assay will be identified, therefore atypical or emerging pathogens will generally evade detection by PCR. Although commercial PCR based tests [1] are available that overcome some of the pitfalls associated with in-house tests, they remain PCR based technologies and as a result suffer from the same sequence based pitfalls outlined above.

Abbreviations: RT-PCR, real-time polymerase chain reaction; NGS, next generation sequencing; NPS, nasopharyngeal swab; VTM, viral transport medium; HRV, human rhinovirus; IFA, influenza A; IFB, influenza B; RSV, respiratory syncytial virus; ADV, adenovirus; hMPV, human metapneumovirus; PIV-1-4, parainfluenza virus 1-4; HCoV, human coronavirus; WoSSVC, West of Scotland Specialist Virology Center; HEV, human enterovirus; Ct, cycle threshold; BLAST, basic local alignment search tool; TRT, turn-around time.

* Corresponding author.
E-mail address: Fiona_thorburn@hotmail.com (F. Thorburn).

http://dx.doi.org/10.1016/j.jcv.2015.06.082
1386-6532/© 2015 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).
Introducing NGS into a diagnostic setting may revolutionize the investigation of respiratory infections. Combining sequence independent amplification with NGS will potentially detect viral and non-viral pathogens within a clinical specimen without actively targeting them, while simultaneously analyzing the genetic sequence. NGS is established in virus discovery, whole genome studies and metagenome studies [2–4] thus the simultaneous detection of multiple different pathogens with this technique is possible. However the efficacy and feasibility of employing such techniques in a diagnostic setting requires further study.

2. Objectives

Here we present a pilot study that compares current diagnostic techniques, namely RT-PCR with NGS in the detection of RNA viruses in respiratory samples from individuals symptomatic of a respiratory illness.

3. Study design

3.1. Samples

Eighty nine nasopharyngeal swabs (NPS) were collected from adults with upper respiratory tract infections between May 2010 and October 2011. Samples were collected as part of the VIDARIS trial, a random subset of which were used in this study. It should be noted that over half of the participants in this trial were vaccinated against influenza. Ethical approval was provided by the Upper South B Regional Ethics Committee. All participants provided written informed consent [5]. Swabs were stored in viral transport media (VTM) at −80°C until testing. The VTM was thawed at 37°C and centrifuged at 1500 × g for 10 min to remove debris. Total nucleic acids were extracted from 200 μl of the supernatant (MagJet Viral DNA and RNA kit, Thermo Scientific) and eluted in 100 μl of water.

3.2. Next generation sequencing method

A 20 μl aliquot of the extract was treated with DNAse 4U (Turbo DNAsese 2U/μl, Life technologies) for 30 min at 37°C. RNA was purified from the reaction using RNAclean XP beads (Agencourt), eluted in 15 μl of water and reverse transcription carried out using Maxima Minus H (Thermofisher) at 50°C for 60 min with 0.2pM primer FR26RV-N (5′ GCC GGA GCT CTG CAG ATA TC 3′). Second strand cDNA was synthesised (NEBNext mRNA 2nd Strand Synthesis, New England Biolabs) and the reaction purified with Ampure XP beads (Agencourt). Sequence-independent single primer amplification (SISPA) was carried out with the Advantage 2 PCR kit (Clontech) and 0.2pM primer FR20RV (5′ GCC GGA GCT CTG CAG ATA TC 3′). The PCR product was purified with Ampure XP beads (Agencourt) and quantified (Qubit HS DNA, Life Technologies). 1 ng of cDNA was used to prepare barcoded sequencing libraries with the Nextera XT DNA Sample Prep kit (Illumina) and indices from the Nextera XT Index Kit as per the manufacturer’s instructions. Up to 24 sample libraries were pooled for sequencing run and 151 bp paired-end reads were generated on the Illumina MiSeq.

3.3. Bioinformatic analysis

Sequencing adapters and low quality sequencing reads were removed (Trim Galore!, Babraham Bioinformatics) and low-complexity reads filtered out (PrinSeq [6]). High quality paired-end sequences were retained for downstream analyses. These sequences were mapped to a database containing a human genome and cDNA references, to remove host sequences. Unmapped sequences were entered into the Metamos pipeline [7] which employs multiple de novo assemblers with k-mer optimization to assemble contigs. The contigs from the most effective assembly were then taxonomically classified using the Basic Local Alignment Search Tool (BLAST) against the GenBank nucleotide and non-redundant databases (cut off E value 0.001). Identical sequences between samples were removed using BowTie and unique sequences were retained for further analysis. Sequenced reads were then mapped back to the top taxonomic hit for each sample and visualized using Tablet [8], to quantify viral reads within each sample and generate a consensus sequence. Where appropriate, greater than 90% of reference genome coverage, the consensus sequences were aligned with known reference sequences and phylogenetic analysis carried out using MEGA6 [9]. Taxonomic hits were compared with the results of the diagnostic qRT-PCR (Table 1).

4. Results

4.1. Next generation sequencing

The average number of sequences generated per sample was ~660,640 (range 30,872–1,278,122) after quality trimming and filtering. Viral contigs were found in 53/89 samples but following removal of duplicate reads this was reduced to 46/89. In a subset of samples (n = 8), there were fewer than 10 unique viral reads detected by the NGS assay alone. Due to the low number of reads we deemed these to be negative by NGS. The viral sequences detected in the remaining 38 samples belonged to the Picornaviridae, Coronaviridae, Paramyxoviridae and Orthomyxoviridae (Table 1). No mixed infections were detected by NGS.

Picornaviruses were most frequently detected (n = 21) and classified as HRV in 20/21 and enterovirus (HEV) in 1/21 cases. These could be subdivided into 3 rhinovirus species, A (11/21), B (4/21), C (5/21) and HEV D. Picornaviruses sequences generated by NGS showed high similarity at the nucleotide level to reference genomes available in the NCBI database, allowing us to assign a serotype in all but one cases (Table 1). The extent of reference genome coverage and phylogenetic similarity to reference sequences are shown in Figs. 1 and 2, respectively. Numerous HRV serotypes and an HEV-D68 were detected. Human coronaviruses (HCoV) were detected in nine samples and were found to belong to the following types: HCoV 229E (4/9), HCoV NL63 (3/9) and HCoV OC43 (2/9). Paramyxoviruses were detected in seven samples. These included hMPV-B (2/7), PIV-3 (2/7), RSV-A (2/7) and RSV-B (1/7). An Orthomyxovirus was detected in one sample and typed as Influenza A H3N2.

All viruses identified by NGS were confirmed by qRT-PCR. However, in eleven cases, virus was identified by RT-PCR only. This included the following viruses, ADV (1/11), PIV-2 (1/11), hMPV (1/11), RSV (2/11), HCoV (3/11) and HRV (3/11). One sample was found by RT-PCR to contain a mixture of both ADV and HRV. The NGS method failed to detect the ADV in this sample. Where NGS confirmed the findings of the RT-PCR assay the Ct values were sig-
Table 1
A summary of sequenced reads, real-time PCR and NGS data, including RT-PCR results for all positive samples and the accession number of top taxonomic hits from BLAST along with reference assembly details.

<table>
<thead>
<tr>
<th>Sample</th>
<th>Total Reads</th>
<th>% mapping to human reference</th>
<th>PCR</th>
<th>Ct</th>
<th>Taxonomy</th>
<th>Accession</th>
<th>% reads mapping</th>
<th>% genome covered</th>
<th>Nucleotide mismatch</th>
</tr>
</thead>
<tbody>
<tr>
<td>1C2</td>
<td>1107488</td>
<td>81.67</td>
<td>HRV</td>
<td>26.75</td>
<td>Enterovirus D-68</td>
<td>AB601885</td>
<td>1.46</td>
<td>14.9</td>
<td>1.4</td>
</tr>
<tr>
<td>2A3</td>
<td>1110727</td>
<td>88.23</td>
<td>HCoV 229E</td>
<td>18.14</td>
<td>HCoV 229E</td>
<td>JX503060</td>
<td>0.16</td>
<td>55.3</td>
<td>0.6</td>
</tr>
<tr>
<td>2C3</td>
<td>470840</td>
<td>91.93</td>
<td>HCoV 229E</td>
<td>25.83</td>
<td>HCoV 229E</td>
<td>JX503060</td>
<td>0.48</td>
<td>0.0</td>
<td>0.7</td>
</tr>
<tr>
<td>2D2</td>
<td>743683</td>
<td>90.66</td>
<td>HCoV NL63</td>
<td>27.17</td>
<td>HCoV NL63</td>
<td>JQ765569</td>
<td>0.97</td>
<td>33.9</td>
<td>0.5</td>
</tr>
<tr>
<td>1E7</td>
<td>426736</td>
<td>93.78</td>
<td>hMPV-B</td>
<td>25.83</td>
<td>hMPV-B</td>
<td>KF530112</td>
<td>0.38</td>
<td>33.9</td>
<td>0.5</td>
</tr>
<tr>
<td>1F8</td>
<td>644660</td>
<td>94.12</td>
<td>HCoV OC43</td>
<td>27.17</td>
<td>HCoV OC43</td>
<td>JN129835</td>
<td>2.70</td>
<td>97.6</td>
<td>0.7</td>
</tr>
<tr>
<td>1G6</td>
<td>81749</td>
<td>37.68</td>
<td>hMPV</td>
<td>25.83</td>
<td>hMPV-B</td>
<td>KJ673971</td>
<td>1.41</td>
<td>18.5</td>
<td>2.0</td>
</tr>
<tr>
<td>2A9</td>
<td>626886</td>
<td>84.12</td>
<td>hMPV</td>
<td>25.83</td>
<td>hMPV-B</td>
<td>KF530171</td>
<td>6.35</td>
<td>48.4</td>
<td>3.1</td>
</tr>
<tr>
<td>1C7</td>
<td>538412</td>
<td>69.88</td>
<td>HRV</td>
<td>24.93</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>4.76</td>
<td>87.4</td>
<td>1.2</td>
</tr>
<tr>
<td>1D4</td>
<td>1089553</td>
<td>23.19</td>
<td>HRV</td>
<td>19.81</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>74.67</td>
<td>99.6</td>
<td>2.2</td>
</tr>
<tr>
<td>1I5</td>
<td>728425</td>
<td>92.82</td>
<td>HRV</td>
<td>19.23</td>
<td>HRV-A2</td>
<td>X02316</td>
<td>5.32</td>
<td>97.0</td>
<td>3.7</td>
</tr>
<tr>
<td>2D5</td>
<td>992040</td>
<td>94.07</td>
<td>HRV</td>
<td>20.44</td>
<td>HRV-A2</td>
<td>JN837694</td>
<td>6.65</td>
<td>96.0</td>
<td>8.1</td>
</tr>
<tr>
<td>1B1</td>
<td>546282</td>
<td>20.17</td>
<td>HRV</td>
<td>19.78</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>32.03</td>
<td>100</td>
<td>1.9</td>
</tr>
<tr>
<td>1D1</td>
<td>1010455</td>
<td>59.82</td>
<td>HRV</td>
<td>20.68</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>32.03</td>
<td>100</td>
<td>1.9</td>
</tr>
<tr>
<td>1D4</td>
<td>1089553</td>
<td>23.19</td>
<td>HRV</td>
<td>19.81</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>74.67</td>
<td>99.6</td>
<td>2.2</td>
</tr>
<tr>
<td>1I8</td>
<td>1132106</td>
<td>99.02</td>
<td>HRV</td>
<td>22.34</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>0.01</td>
<td>69.6</td>
<td>5.6</td>
</tr>
<tr>
<td>1C3</td>
<td>1278307</td>
<td>47.23</td>
<td>HRV</td>
<td>28.15</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>0.31</td>
<td>20.9</td>
<td>8.7</td>
</tr>
<tr>
<td>1I2</td>
<td>710612</td>
<td>60.67</td>
<td>HRV</td>
<td>19.80</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>33.29</td>
<td>99.1</td>
<td>3.8</td>
</tr>
<tr>
<td>1B8</td>
<td>818055</td>
<td>96.78</td>
<td>HRV</td>
<td>29.11</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>0.01</td>
<td>69.6</td>
<td>5.6</td>
</tr>
<tr>
<td>1F7</td>
<td>236596</td>
<td>70.59</td>
<td>HRV</td>
<td>28.03</td>
<td>HRV-A1</td>
<td>JN837694</td>
<td>0.01</td>
<td>69.6</td>
<td>5.6</td>
</tr>
<tr>
<td>1E3</td>
<td>700028</td>
<td>93.78</td>
<td>Influenza A</td>
<td>22.34</td>
<td>Influenza A</td>
<td>KJ627329</td>
<td>33.29</td>
<td>99.1</td>
<td>3.8</td>
</tr>
<tr>
<td>1E1</td>
<td>758165</td>
<td>61.76</td>
<td>Influenza A</td>
<td>20.1</td>
<td>Influenza A</td>
<td>KJ627329</td>
<td>33.29</td>
<td>99.1</td>
<td>3.8</td>
</tr>
<tr>
<td>1B2</td>
<td>638257</td>
<td>93.07</td>
<td>Influenza A</td>
<td>20.1</td>
<td>Influenza A</td>
<td>KJ627329</td>
<td>33.29</td>
<td>99.1</td>
<td>3.8</td>
</tr>
<tr>
<td>1E5</td>
<td>714509</td>
<td>96.87</td>
<td>RSV</td>
<td>30.01</td>
<td>RSV</td>
<td>KF826849</td>
<td>0.03</td>
<td>10.8</td>
<td>2.0</td>
</tr>
<tr>
<td>1F7</td>
<td>236596</td>
<td>70.59</td>
<td>RSV</td>
<td>33.96</td>
<td>RSV</td>
<td>KF826849</td>
<td>0.03</td>
<td>10.8</td>
<td>2.0</td>
</tr>
</tbody>
</table>

4. Discussion

As a pilot study designed to assess the utility of NGS as a method for the detection of respiratory RNA viruses, we compared an in-house NGS method with an established in-house RT-PCR test using 89 respiratory samples. A viral pathogen was detected by the NGS method in 38 samples with the RT-PCR confirming these and detecting a further 11 viruses. Overall, based on these results, the NGS assay had a sensitivity of 77.55% (95% CI 63.37–88.21%) and specificity of 80.49% (95% CI 65.13–91.15%) or 100% (95% CI 91.31–100%) if considering samples with fewer than 10 unique reads are negative, compared to RT-PCR. It is possible that the sensitivity of NGS detection could vary between virus groups however the number of detections in this cohort was not powered to assess each individually.

The 11 NGS negative/RT-PCR positive samples contained virus at lower Cts compared to those positive by both assays, suggesting the current NGS method has a cut off in the region of Ct 32, approximately 1–2 logs less sensitive than the qRT-PCR method (data not shown). This finding is similar to that outlined by Prachayaprecha et al. At present, this level of sensitivity is not appropriate for diagnostic services to replace qRT-PCR.
Fig. 1. HRV genome coverage – a graphic visualization of the reference genome coverage following Bowtie2 alignment of sequencing reads to the top BLAST hit for each HRV positive sample.

Fig. 2. Neighbor-joining (NJ) phylogenetic tree of full-genome reference sequences and a subset of the NGS consensus sequences with full or near-full reference genome coverage. The branch annotations represent the bootstrap values (percentage of 1000 samples trees). Sequences generated in this study are shown as circles and squares.

Fig. 3. Threshold cycle (Ct) values respiratory of samples positive by RT-PCR and either confirmed or unconfirmed by the NGS assay. The bars indicate the mean and one standard deviation. An unpaired t-test demonstrates a significant difference between the two groups (p < 0.0001).

Fig. 4. The percentage of sequenced reads, after quality trimming, mapping to the taxonomic reference genome determined by BLAST. Linear regression of data, $R^2$ value = 0.27 ($p = 0.0009$). This is suggestive of a log relationship between the proportion of reads mapping and the Ct value.

sequencing could improve the sensitivity of NGS, either by reducing the multiplexing of samples (i.e., reducing the number of samples processed per sequencing run) or using an alternative platform with greater capacity such as the Illumina NextSeq or HiSeq. Reducing the contamination of host DNA through host DNA depletion holds the greatest promise [11] to increase the depth of viral and microbial DNA as we found up to 99% of reads obtained per sample were derived from the host.

One of the 11 pathogens not detected by the NGS method was adenovirus. It is unclear whether this was due to the sensitivity cut off discussed above (Ct of 35.1) or due to the initial DNAse step outlined in the method. In future, if this is to be adapted for the detection of DNA viruses and other non viral causes of respiratory infection then other ways of enrichment will need to be sought. Perhaps target enrichment via hybridization, but again this may encounter the issue of missing novel or changed viruses.

Although the NGS method failed to detect all the positive samples, it did offer several advantages over the RT-PCR method. For example, the NGS provided more detailed typing information for the detected viruses including subtype data for RSV and hMPV as well as serotype data for the HRV and HEV. Such information is highly informative. As well as enabling real-time diagnostic data to be produced, NGS could simultaneously provide resistance testing (e.g., H275Y mutation of influenza A, conferring oseltamivir resistance) and typing data, as demonstrated here and previously [12,13]. This could be used to inform public health of circulating strains and could allow rapid detection of the emergence of novel subtypes (e.g., EV68 and ADV 14) or highlight potential outbreaks. In future, it may also detect viral polymorphisms associated with disease severity [14,15].

Similar to other studies we also found a correlation between NGS sequence reads and RT-PCR cycle thresholds [16,17]. This was particularly strong for rhinoviruses – mainly because these represented the majority of the pathogens detected. This suggests that sequence reads might be a suitable proxy for viral/target copies. This would enable laboratories/clinicians to interpret the clinical relevance of results [18]. Such data can also be used to infer prognosis or treatment response.

Viral co-infection detection by NGS could not be assessed in this cohort as only a single co-infection episode was detected by RT-PCR. The population under study were normally healthy adults who did not require health care interventions. There is little information on co-infections in such individuals but evidence suggests that viral co-infections occur more commonly in hospitalised individuals [19]. The detection of influenza was also lower than may have been expected which probably relates to the level vaccina-
tion amongst the cohort. Further studies including children and unvaccinated individuals would be of use.

The NGS approach did not detect positive results that were missed by the RT-PCR. Although this panel was small, our data supports other studies which have concluded that current respiratory panels are appropriate for the detection of the main causes of viral respiratory disease. A major barrier to the introduction of NGS as a routine diagnostic test is its cost and turnaround time. The current cost of sequencing is prohibitive as a diagnostic test when compared to RT-PCR but this has reduced in recent years, a trend which will probably continue in the future. Although not designed to detect bacteria the current method still detected bacterial genomes in nearly all samples, this finding shows that syndromic testing is a possibility. If viral, bacterial and fungal diagnostic tests can be combined into a single assay this would benefit the cost effectiveness of developing NGS as a diagnostic tool.

The turnaround-time of the above process was in the region of seven days, including preliminary data analysis whereas the turnaround time for RT-PCR methods is usually just a few hours. As a result, NGS is unlikely to become a routine diagnostic test in the near term. However there are many steps that could be accelerated with automation. The development of kit based library preparation methods has also resulted in a condensed process, reducing hands-on time. Technical advances that allow a greater depth of sequencing could obviate the need for enrichment processes. Furthermore third generation sequencers, such as the PacBio RS II, offer the potential for even more rapid sequencing. Taken together, these advances are likely to improve the TRT of NGS significantly.

Competing interests

None declared.

Funding

This study was funded by the UK Medical Research Council (Grant reference number G0801822).

Ethical approval

The study was approved by the Upper South B Regional Ethics Committee, New Zealand (URB/09/10/050) and all participants provided written informed consent.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.jcv.2015.06.082

References