Journal of Clinical Virology 69 (2015) 96–100

Contents lists available at ScienceDirect

Journal of Clinical Virology

journal homepage: www.elsevier.com/locate/jcv

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

Fiona Thorburn[a,*], Susan Bennett[b], Sejal Modha[a], David Murdoch[c], Rory Gunson[b], Pablo R. Murcia[a]

[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.

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

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 DNAse 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 TCN NNN NN 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 per 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 optimisation 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 BedTools 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).

3.4. Diagnostic in-house RT-PCR methods

40 μl of the nucleic acid extract was then screened for human rhinovirus (HRV), influenza A/B (IFA/IFB), respiratory syncytial virus (RSV), adenovirus (ADV), human metapneumovirus (hMPV), parainfluenzavirus 1–4 (PIV 1–4), coronaviruses (HCoV) HKU1, NL63, OC43 and 229E and Mycoplasma pneumonia using the routine diagnostic qRT-PCR at the West of Scotland Specialist Virology Centre (WoSSVC) as previously described [10].

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. Picornavirus 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-