National Institute for Public Health and the Environment
Ministry of Health, Welfare and Sport

A target hybridization-based next-generation sequencing assay enhances surveillance of seasonal respiratory pathogens: a retrospective assessment (ERVINGS)

Thierry K. S. Janssens[1], Shing H. Zhan[2], Pieter Overduin[1], Sepideh M. Alamouti[2], Brian S. Kwok[2], Dennis Schmitz[1], Mohammad A. Qadir[2], and Adam Meijer[1,*]

1 Centre for Infectious Diseases Research , Diagnostics and *laboratory* Surveillance, National Institute for Public Health and the Environment , Bilthoven, The Netherlands

2 Fusion Genomics Corporation, Burnaby, BC, Canada

* adam.meijer@rivm.nl

Introduction

Contemporary monitoring of respiratory pathogens heavily relies on conventional molecular and phenotypic testing.

Conventional "gold standard" tests are directed at specific targets and require multiplexing but also need regular updates and validations.

Whole metagenomics sequencing (WMGS) enables comprehensive and less biased detection and simultaneous molecular characterization of pathogens without prior sequence knowledge .

Bottlenecks for the implementation of WMGS in routine surveillance are affordability due to high requisite sequence depth and complexity of the analysis. This is especially valid in **low-density samples**, such as respiratory swabs.

Target enrichment via probe capture facilitates WMGS by reducing the required sequence depth and streamlining the analysis and ***allows*** *for* the concurrent detection and molecular characterization of multiple pathogens of interest.

Aims

Evaluation and validation of the OneTest™ URP RNA kit, a commercial targeted enrichment-based metagenomics assay to detect and characterize respiratory pathogens in routine surveillance and clinical analysis: ***molecular first***.

Methods

Total RNA from 155 respiratory samples from 2006 *through* 2008 and 2012 *through* 2017 underwent OneTest™ URP RNA library prep and probe capture. QuantumProbes™ are machine-learning developed probes that capture of a taxonomically broad and genetically diverse spectrum of pathogens to achieve ~9000x enrichment of the taxonomically and phenotypically informative genes.

Samples were deliberately selected to cover a range of types and loads of respiratory pathogens from the Dutch surveillance program for influenza-like illness and acute respiratory infection. The samples were previously screened by "gold standard" assays including PCR, virus isolation and Sanger sequencing for typing.

The libraries were sequenced using Illumina NextSeq 2x150 bp (~1-2Mreads/sample), reads were preprocessed and mapped to reference sequences, iteratively assembled and aligned to the closest reference. Pathogen loads were estimated as Fragments Per Kilobase per Million mapped reads (FPKM) and coverage breadth and correlated to RT-qPCR Cycle threshold (Ct) values. Maximum-likelihood phylogenetic analysis of the consensus sequences was conducted to characterize some of the detected pathogens.

Results

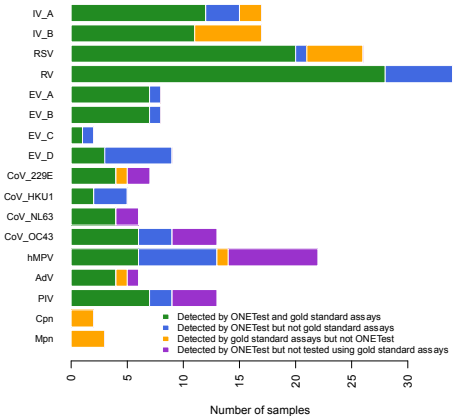

Fig. 1: Comparison of the pathogen calls between the gold standard assays and the OneTest™ assay.

Abbreviations:

IV: Influenza virus, RSV: respiratory syncytial virus, RV: rhinovirus, EV: enterovirus, CoV: coronavirus, hMPV: metapneumovirus, AdV: adenovirus, PIV: parainfluenzavirus, Cpn: *Chlamydia pneumoniae*, Mpn: *Mycoplasma pneumoniae*

Discussion and Conclusion

- High concordance of the OneTest™ RNA kit with "gold standard" assays.
- OneTest™ provides semi-quantitative information, like qPCR.
- Additional calls in OneTest™ help to close the diagnostic gap not covered by "gold standard" assays.
- OneTest™ enables simultaneously detection and sequence-based characterization, e.g. typing, phenotypic prediction and phylogenetics.
- Opportunity for application of OneTest™ in seasonal surveillance of respiratory syndromes.

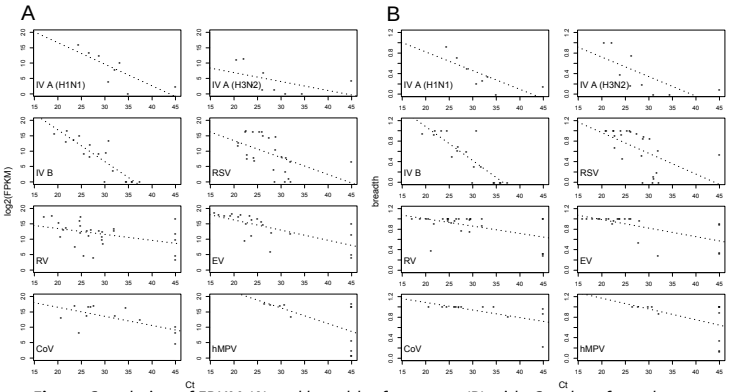

Fig. 2: Correlation of FPKM (A) and breadth of coverage (B) with Ct values for select pathogens. The target genes used were (IV: haemagglutinin, RSV and hMPV: fusion protein gene, RV and EV: VP1 capsid protein gene and CoV: nucleprotein gene) The data points at Ct=45 were considered negative.

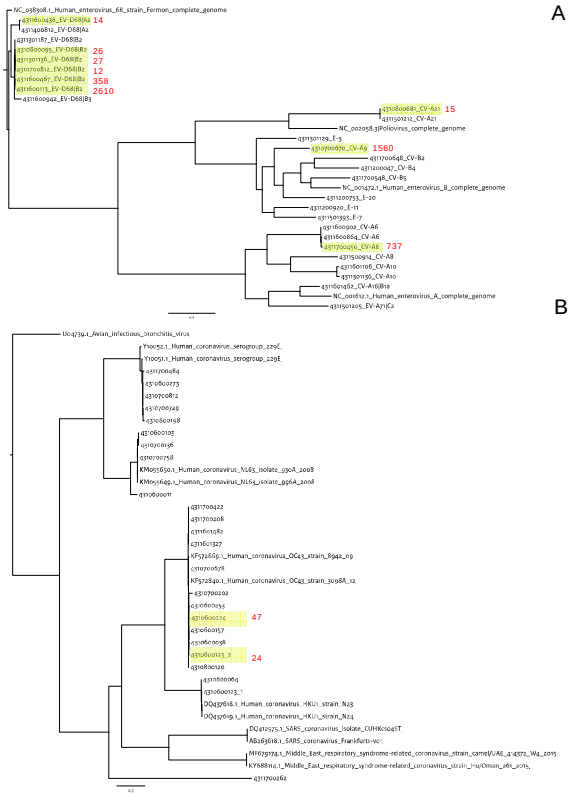

Fig. 3: ML tree (GTR+G model) on (A) Enterovirus VP1 capsid protein and (B) Coronavirus spike protein sequences, revealed a consistent typing pattern by comparison of the consensuses sequences with select reference sequences. Yellow marked sequences (with FPKM values in red) were not detected by the gold standard assays, despite reasonable coverage. This is likely caused by reduced primer and probe match in the gold standard assays. This was apparent for the Enterovirus D68 B2 clade.

Rijksinstituut voor Volksgezondheid en Milieu, Antonie van Leeuwenhoeklaan 9, 3721 MA Bilthoven, Nederland www.rivm.nl