Simultaneous Viral Whole-Genome Sequencing and Differential Expression Profiling in Respiratory Syncytial Virus Infection of Infants

Gu-Lung Lin[1,a], Tanya Golubchik[2,a], Simon Drysdale[1,3], Daniel O'Connor[1], Kimberley Jefferies[1], Anthony Brown[4], Mariateresa de Cesare[5], David Bonsall[2,5], M Azim Ansari[5], Jeroen Aerssens[6], Louis Bont[7], Peter Openshaw[8], Federico Martinón-Torres[9,10], Rory Bowden[5,b], Andrew J Pollard[1,b]

[1]Oxford Vaccine Group, Department of Paediatrics, University of Oxford, Oxford OX3 7LE, UK, and the National Institute for Health Research Oxford Biomedical Research Centre, Oxford OX4 2PG, UK.

[2]Big Data Institute, Nuffield Department of Medicine, University of Oxford, Oxford OX3 7LF, UK.

[3]Department of Paediatrics, St George's University Hospitals NHS Foundation Trust, London SW17 0QT, UK.

[4]Peter Medawar Building for Pathogen Research, University of Oxford, Oxford OX1 3SY, UK.

[5]Wellcome Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, UK.

[6]Translational Biomarkers, Infectious Diseases Therapeutic Area, Janssen Pharmaceutica NV, 2340 Beerse, Belgium.

[7]Department of Pediatrics, Wilhelmina Children's Hospital, University Medical Center Utrecht, 3584 EA Utrecht, Netherlands, and the ReSViNET Foundation, Zeist, Netherlands.

[8]National Heart and Lung Institute, Imperial College London, London SW7 2DD, UK.

[9]Translational Pediatrics and Infectious Diseases, Hospital Clínico Universitario de Santiago de Compostela, 15706 Santiago de Compostela, Spain.

[10]Genetics, Vaccines, Infectious Diseases, and Pediatrics Research Group (GENVIP), Instituto de Investigación Sanitaria de Santiago de Compostela, 15706 Santiago de Compostela, Spain.

[a]G.-L. L. and T. G. contributed equally to this work.

[b]R. B. and A. J. P. contributed equally to this work.

Correspondence: G.-L. Lin, MD, Oxford Vaccine Group, Centre for Clinical Vaccinology and Tropical Medicine, Churchill Hospital, Oxford OX3 7LE, UK (gu-lung.lin@paediatrics.ox.ac.uk)

Summary

This study used targeted metagenomics to sequence clinical RSV samples, reconstructed the phylogeny of the infecting viruses, and detected differential gene expression between two RSV subgroups. This methodology can relate viral genetics to clinical phenotype and facilitate population-level viral surveillance.

Abstract

Targeted metagenomics using strand-specific libraries with target enrichment is a sensitive, generalized approach to pathogen sequencing and transcriptome profiling. Using this method, we recovered 13 (76%) complete human respiratory syncytial virus (RSV) genomes from 17 clinical respiratory samples, reconstructed the phylogeny of the infecting viruses, and detected differential gene expression between two RSV subgroups, specifically, a lower expression of the P gene and a higher expression of the M2 gene in RSV-A than in RSV-B. This methodology can help to relate viral genetics to clinical phenotype and facilitate ongoing population-level RSV surveillance and vaccine development.

Keywords: respiratory syncytial virus; transcriptome profiling; RNA sequencing; whole-genome sequencing; differential gene expression.

BACKGROUND

Methods to provide high-throughput, whole-genome sequences for specific viruses without a bespoke assay would have substantial value in epidemiological and evolutionary studies. For example, considering the global impact of human respiratory syncytial virus (RSV) infection, the lack of a simple, cost-effective assay means that relatively few complete sequences have been determined, and little is known about potential links between viral genetics and pathogenicity. In this study we demonstrate that a generalized approach to pathogen sequencing, which we term targeted metagenomics, can provide a quantitative sequencing assay for a chosen virus that, because it relies only on library construction and enrichment (two processes that do not depend on the particular target pathogen), can be applied to new viruses without the requirement for optimization. In addition, the methodology enables virus transcriptome profiling to detect genotype-specific patterns of transcription of potential phenotypic significance.

RSV, an unsegmented, single-stranded, negative-sense RNA virus with a 15.2-kb genome, consisting of 10 genes, is the leading pathogen causing lower respiratory tract infection among young children, globally responsible for an estimated 33.1 million episodes of lower respiratory tract infection, 3.2 million hospitalizations, and up to 149,400 deaths in children <5 years of age annually [1]. RSV is classified into two antigenically and phylogenetically distinct subgroups: RSV-A and RSV-B. A detailed understanding of RSV clinical infection at the molecular level could aid evaluation of viral responses to immune pressure from new vaccines or monoclonal antibodies that are under development. In this study we adopt a targeted sequencing approach, ve-SEQ [2], designed for the detection of pathogen sequences, and adapt existing computational pipelines for use in RSV genomics and transcriptomics.

METHODS

Detailed methods are provided in the Supplementary Methods.

Clinical studies and samples

All available nasopharyngeal swabs collected from a cohort of RSV-infected infants <1 year of age enrolled in the clinical studies of the REspiratory Syncytial virus Consortium in EUrope project (ClinicalTrials.gov identifier: NCT03627572 and NCT03756766) from Santiago de Compostela, Spain, and Oxford, UK, during the 2017–18 RSV season, were included in this study. The protocol was approved by the local ethics committee at each center. The parents or guardians of all participants provided written, informed consent.

Targeted metagenomic sequencing

NucliSENS easyMAG (BioMérieux, Marcy-l'Étoile, France) was used to extract 25 μL of total nucleic acids from 500 μL of each swab sample, following the manufacturer's instructions. Sequencing libraries were constructed from 8 μL of each nucleic acid sample, following the ve-SEQ protocol applied to HIV [3] with modifications. Briefly, dual indexed Illumina libraries were constructed using the SMARTer Stranded Total RNA-Seq Kit v2—Pico Input Mammalian (Takara Bio USA, California, US) without RNA fragmentation. 2 μL of library from each sample was pooled for sequence enrichment by a probe set targeting over 100 bacteria and viruses, including RSV (Supplementary Table 1) [4]. PCR of 16 cycles was carried out for post-capture amplification and the final product was purified by Agencourt AMPure XP (Beckman Coulter, California, US).

Sequencing was performed on the Illumina MiSeq platform (Illumina, California, US), generating 265-bp paired-end reads.

Viral load measurement

Viral load was measured by quantitative RT-PCR from the same nucleic acid samples used for sequencing. PCR assays were prepared in triplicate using the genesig RSV Standard Kit (Primerdesign, Southampton, UK), targeting the N gene, and performed on the Applied Biosystems StepOnePlus Real-Time PCR System (Thermo Fisher Scientific, Massachusetts, US) with amplification conditions of 55˚C for 10 min, 95˚C for 2 min, followed by 50 cycles of 95˚C for 10 sec and 60˚C for 60 sec.

Viral genome and transcriptome reconstruction

shiver [5] was used to reconstruct RSV genomes and transcriptomes from the de novo assemblies of sequencing reads generated by either IVA (v 1.0.8) [6] or SPAdes (v 3.13.0) [7] with the deduplication option enabled. A minimum of two unique reads were required to support consensus base-calls.

Each pre-deduplicated binary alignment/map (BAM) file was divided into 2 files: one containing read pairs mapped in the antisense orientation (genomic reads, SAM flags: 83 or 163), and the other in the sense orientation (positive-strand reads, SAM flags: 99 or 147). The 2 files were then deduplicated and processed using shiver to generate consensus sequences and per-base coverage and composition.

Phylogenetic analysis

Maximum-likelihood phylogenies were reconstructed from complete RSV sequences generated in this study with another 60 RSV whole-genome sequences downloaded from GenBank (Supplementary Table 4) using RAxML (v 8.2.12) [8] with the general time reversible model and gamma-distributed rate heterogeneity among sites.

Differential expression analysis

Samples with complete RSV genome assembly were used for differential expression analysis, following the RNAseq123 workflow [9]. Briefly, HTSeq (v 0.11.2) [10] was used to calculate the number of positive-strand reads mapped to each RSV gene from pre-deduplicated BAM files. Linear models were constructed to compare gene expression levels between RSV subgroups, severity, and host sex groups from the normalized and transformed count data. Severity was defined using the ReSVinet scale [11]; a score of 0–7 was defined as mild, a score of 8–13, moderate, and a score of 14–20, severe. Finally, empirical Bayes moderated t-statistics were applied to compute the significance of differential expression with a minimum $\log_2$ fold change of 1.2.

Statistical analysis

All statistical analyses were performed using R. P-values or adjusted P-values <0.05 were considered significant.

RESULTS

We constructed RNA-seq libraries from 17 RSV-positive samples, including 6 from Oxford, UK and 11 from Santiago de Compostela, Spain (Supplementary Table 2). The enriched libraries yielded a median of 30% (0.02–75%) RSV reads, originating from both the positive (transcriptome and replicative intermediate) and negative (genomic) strands at a ratio of 5–11:1 (Supplementary Table 3 and Data). Complete RSV genomes were recovered from 13 (76%) of the samples, including all samples with viral load >2,000 copies/mL or that generated >10,000 unique (deduplicated) RSV reads (Figure 1A), with a sensitivity, achieved without further optimization, that makes the methodology suitable for studying a wide variety of infections.

Using phylogenetic analysis, we identified four RSV-A (genotype ON1) and nine RSV-B (genotype BA9) samples among the 13 completely sequenced genomes (Figure 1B,C). In all but one case, the nearest neighbor sequence for each new genome was another sample from our study, although not necessarily from the same country, reflecting the ongoing evolution and limited number of co-circulating RSV lineages.

Read coverage plots revealed characteristic, strand-specific patterns of read depth along the genome and antigenome. In contrast to the genomic strand (Figure 2A, left), positive-strand read coverage was marked by read depths in intergenic regions ~1–10% of those in coding regions (Figure 2A, right), recapitulating the known mechanism of RSV transcription, with a single initiation point at the 3' end of the genome and polyadenylation-termination at each gene-end signal before reinitiating at the next gene-start signal [12]. Positive-strand reads spanning gene boundaries presumably represent antigenomes and read-through transcripts. A general pattern of decreasing coverage of successively more-5' genes reflects the dissociation of a fraction of polymerase complexes before reinitiation [12], leading to progressively weaker transcription. Under this model, due to an overlap between the M2 and L genes, expression of the 5'-most gene, L, requires polymerase to terminate at the gene-end signal of the preceding M2 gene, and then to move backwards to the L gene-start signal before reinitiation [12]. This inference is supported by a jump in read depth at the start of the L gene (Figure 2B). However, in light of this basic model, a more detailed examination of relative expression levels also provides indirect support for more complex, post-transcriptional control of transcript abundance: rather than uniformly decreasing transcript levels, transcripts from the NS2, P, and G genes were more abundant than those from the genes preceding them (Figure 2C), in contrast with patterns of transcription in cell-free systems [12].

Noting marked variation among isolates in the proportions of positive-strand reads assigned to each gene (Supplementary Figure 1), we undertook a formal analysis of differential gene expression. RSV subgroup, but not host sex or disease severity, was significantly associated with differences in specific transcript levels (Supplementary Figure 2). The mean level of M2 transcripts in RSV-A (samples GB-01, -03, -05, and -06) was 240% of that in RSV-B (GB-02, -04, and all Spanish isolates), while the expression of the P gene in RSV-A was 57% of that in RSV-B (Figure 2D). While inter-sample variation was marked, perhaps reflecting clinical heterogeneity between samples, other genes were not significantly differentially expressed.

The M2 gene is known to contain two overlapping open reading frames (ORFs): M2-1, whose product interacts with the P gene product (phosphoprotein) to induce antitermination (promoting transcription of full-length subgenomic mRNAs), and M2-2, whose protein conversely acts to favor production of antigenomes over full-length transcripts [12]. Under the accepted model of RSV molecular biology, both ORFs belong to a single transcriptional unit. Given their contrasting roles, we scrutinized our transcription data more carefully to see whether the change in M2 transcription affected both ORFs. In our data (Figure 2B), (i) read depth corresponding to transcription drops by an aggregate 0.5–1 log-fold downstream of M2-1 but within M2-2; (ii) levels of transcription in the remainder of M2-2 are similar to intergenic levels in the rest of the genome; (iii) the signal for enhanced transcription of the M2 gene in RSV-A is confined to the M2-1 region (Figure 2D).

DISCUSSION

In this study, we demonstrate the simultaneous recovery of RSV genomes and transcriptomes directly from clinical samples, without a prerequisite for culture or purification of viral particles. Using strand-specific libraries with target enrichment, we were able to examine separately the RSV genome and transcriptome sequences, and to relate viral gene expression to the phylogenetic subgroup of the infecting virus. We identified intriguing differences between RSV-A and RSV-B transcriptional profiles, which could provide novel pathogenetic and therapeutic insights.

The nonconcordant expressions of the ORFs M2-1 and M2-2 lead us to hypothesize that the two ORFs may represent separate transcriptional units, and that the M2-2 protein may be translated from a low-abundance RNA of unknown length (including possibly read-through of M2-1 mRNA or genomic intermediate) and to conclude that the M2 differential expression signal relates specifically to M2-1. Since the M2-1 protein facilitates the synthesis of polycistronic read-through mRNAs [12], a testable hypothesis consistent with our transcriptomics result is that the RSV-A samples should have less marked declines in positive-strand read depth within intergenic regions than RSV-B due to more read-through mRNAs.

Differential expression of the P gene between subgroups is also intriguing because the direct interaction between the M2-1 and P proteins [13] means that the contrasting changes in gene expression could combine in complex ways. For example, a skew towards transcription in RSV-A could foreseeably affect infection at both the cellular and whole-host levels, and may lead to generally higher prevalence of RSV-A in annual epidemics and proposed greater transmissibility of RSV-A than RSV-B [14]. However, during the study period of 2017–2018, RSV-B was predominant over RSV-A, accounting for 72% of the RSV samples collected in our multi-center study. Although no study has found differences in transcriptional machinery that would potentially cause the differential expressions between subgroups, we speculate that sequence differences or polymerase dissociation in intergenic regions may play a role.

Limitations of this study include the modest sample size, but our study design was robust to technical and analytical bias (see Supplementary Methods). Another limitation is that we did not perform quantitative PCR to support the findings of differential expression. Nevertheless, RNA-seq has been shown to be highly accurate for quantifying transcriptomes, providing estimates of gene expression levels closely correlated with measurements using quantitative PCR [15].

More generally, our methodology facilitates population-level RSV surveillance in the context of current efforts in global RSV surveillance (https://www.who.int/influenza/rsv/who_rsv_surveillance_2nd_phase/en/) and vaccine development. With our approach, it is straightforward to relate viral load, sequence variation, and transcriptional profiles from sequencing data in order to investigate the epidemiological and phenotypic consequences (e.g. host biomarkers) of virus evolution. Targeted metagenomics provides a powerful methodology for characterizing sets of pathogens that may co-occur in a single sample type.

FIGURE LEGEND

Figure 1. A, Sequencing yield and virus load. Yield of unique, RSV-mapped reads is associated with viral load and proportion of genome assembly. Shaded area around the regression line represents the 95% confidence interval. B,C, Maximum-likelihood phylogenies of RSV subgroups A and B. Phylogenies were reconstructed using RAxML from consensus sequences of the 13 complete genomes, augmented with 30 reference genomes representing sampled diversity of each of RSV-A and RSV-B. Sequences from this study are colored in red (UK) or blue (Spain) and genotypes are colored according to the legend. Reference taxa are labelled with country (AU, Australia; BE, Belgium; BR, Brazil; CN, China; DE, Germany; ES, Spain; GB, United Kingdom; IT, Italy; JO, Jordan; KE, Kenya; NL, Netherlands; NZ, New Zealand; PE, Peru; PH, Philippines; TH, Thailand; US, United States of America) and year of sampling. Scale bar shows nucleotide substitutions per site. The phylogenies were rooted to the oldest sampled RSV sequences, dating from 1977.

Figure 2. Coverage plots and differential gene expression. A, Read coverage of complete RSV genomes (left) and transcriptomes (right). Plots are normalized read depth per million mapped reads in each category. Legend below the plots represents the position and length of each RSV gene. B, Positive-strand read coverage across the M2 gene. There is little evidence for read depth dropping at the end of the M2 gene; instead coverage drops within the open reading frame (ORF) M2-2, which is inconsistent with a single transcript covering both ORFs M2-1 and M2-2. Shaded areas represent gene or open reading frame overlaps. C, Transcription by gene and sample. Read counts are normalized by total read count per sample and gene length. D, Differential expression of M2 and P genes and ORFs M2-1 and M2-2. RSV-A has a higher expression of the M2 gene (fold change = 2.38, 95% CI = 1.44 to 3.95, adjusted P = 0.032) and a lower expression of the P gene (fold change = 0.57, 95% CI = 0.47 to 0.70, adjusted P = 0.009)

than RSV-B. Within the M2 gene, RSV-A has a higher expression of M2-1 than RSV-B (fold change = 2.29, 95% CI = 1.49 to 3.53, adjusted P = 0.017), while the expression of M2-2 is comparable between RSV-A and RSV-B (adjusted P = 0.703). Center line of each box denotes the median; box limits, first and third quartiles; whiskers, the highest and lowest values within 1.5 times the interquartile range from the box limits; and outlying points, outliers. CPM denotes counts per million.

Footnote page

Data availability

All raw sequencing reads are accessible at the European Nucleotide Archive (ENA) under the study accession PRJEB34042 (https://www.ebi.ac.uk/ena/data/view/PRJEB34042). The complete RSV sequences generated in this study have been deposited in GenBank with the accession numbers LR699315, LR699726, LR699734, and LR699736–44.

Potential conflicts of interest

A.J.P. is a National Institute for Health Research Senior Investigator with funding from the British Research Council. J.A. is an employee of Janssen Pharmaceutica NV. The views expressed in this article are those of the authors and may not be understood or quoted as being made on behalf of or reflecting the position of the organizations with which the authors are employed/affiliated.

Financial support

This work was supported by the National Institute for Health Research (NIHR) Oxford Biomedical Research Centre, the NIHR Thames Valley and South Midlands Clinical Research Network, the British Research Council, and the REspiratory Syncytial virus Consortium in EUrope (RESCEU) project. RESCEU has received funding from the Innovative Medicines Initiative 2 Joint Undertaking under grant agreement No 116019 (http://www.imi.europa.eu). This Joint Undertaking receives support from the European Union’s Horizon 2020 research and innovation program and European Federation of Pharmaceutical Industries and Associations.

Author contributions

G.-L.L., T.G., D.B., R.B., and A.J.P. conceived and designed the work. G.-L.L, S.D., K.J., J.A., L.B., P.O., F.M.-T., and A.J.P. conducted and supervised the clinical studies. M.A.A. designed the probe set that were used for capture. G.-L.L., A.B., and M.d.C. performed the experiments. G.-L.L., T.G., D.O'C., and R.B. analyzed and interpreted the data. G.-L.L. drafted the manuscript and T.G., R.B., and A.J.P. substantively revised it. All authors have approved the submitted version and agreed to submit the manuscript.

Presented in part: 7th Applied Bioinformatics and Public Health Microbiology Conference in Hinxton, UK, between 5th and 7th of June, 2019.

References

1. Shi T, McAllister DA, O'Brien KL, et al. Global, regional, and national disease burden estimates of acute lower respiratory infections due to respiratory syncytial virus in young children in 2015: a systematic review and modelling study. Lancet **2017**; 390:946–58.

2. Bonsall D, Ansari MA, Ip C, et al. ve-SEQ: Robust, unbiased enrichment for streamlined detection and whole-genome sequencing of HCV and other highly diverse pathogens. F1000Res **2015**; 4:1062.

3. Bonsall D, Golubchik T, de Cesare M, et al. A comprehensive genomics solution for HIV surveillance and clinical monitoring in a global health setting. bioRxiv 397083 [Preprint]. August 28, 2018 [cited 2020 Mar 10]. Available from: https://doi.org/10.1101/397083.

4. Goh C, Golubchik T, Ansari MA, et al. Targeted metagenomic sequencing enhances the identification of pathogens associated with acute infection. bioRxiv 716902 [Preprint]. July 28, 2019 [cited 2020 Mar 10]. Available from: https://doi.org/10.1101/716902.

5. Wymant C, Blanquart F, Golubchik T, et al. Easy and accurate reconstruction of whole HIV genomes from short-read sequence data with shiver. Virus Evol **2018**; 4:vey007.

6. Hunt M, Gall A, Ong SH, et al. IVA: accurate de novo assembly of RNA virus genomes. Bioinformatics **2015**; 31:2374–6.

7. Bankevich A, Nurk S, Antipov D, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol **2012**; 19:455–77.

8. Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics **2014**; 30:1312–3.

9. Law CW, Alhamdoosh M, Su S, et al. RNA-seq analysis is easy as 1-2-3 with limma, Glimma and edgeR. F1000Res **2016**; 5.

10. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics **2015**; 31:166–9.

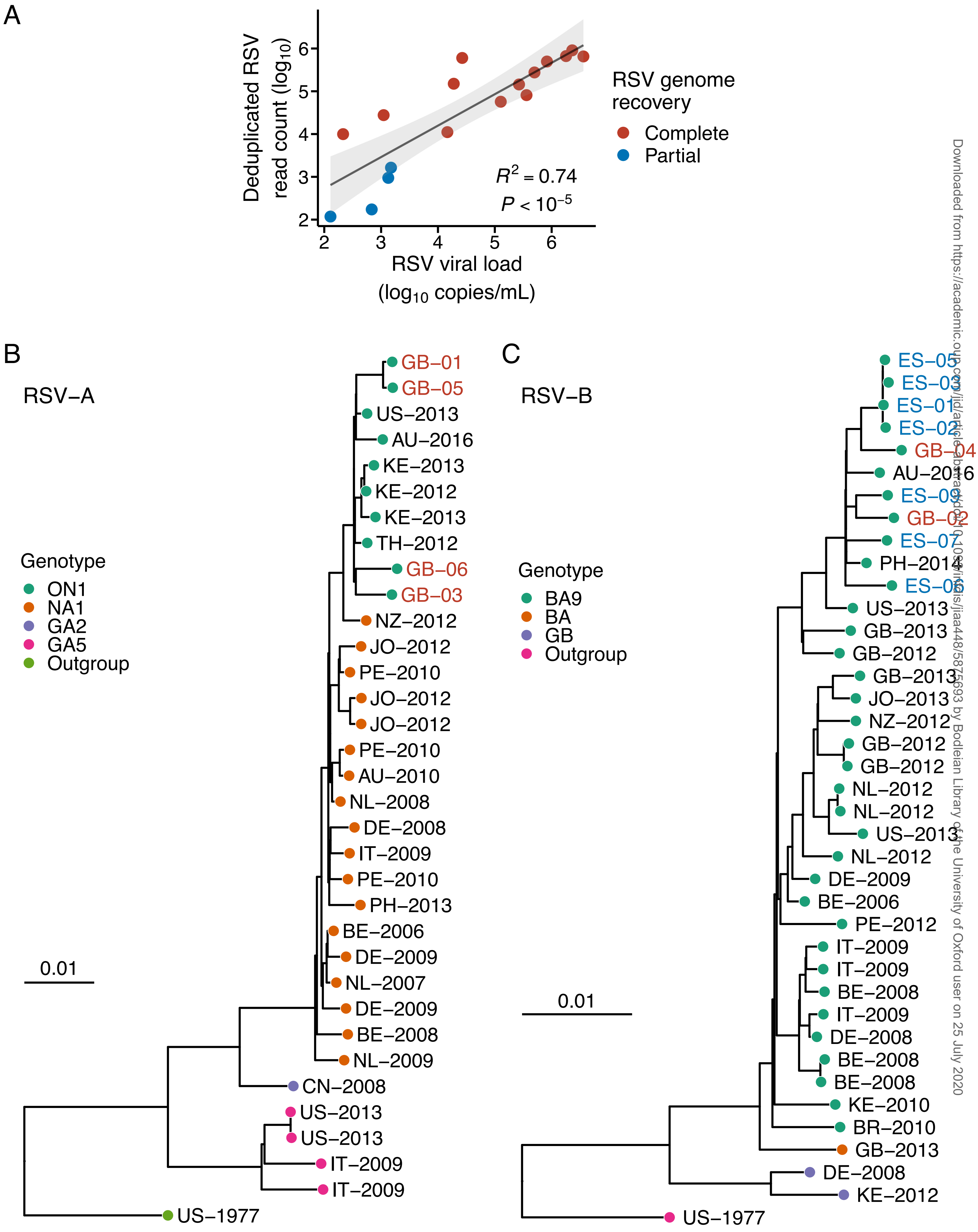
11. Justicia-Grande AJ, Pardo-Seco J, Cebey-López M, et al. Development and validation of a new clinical scale for infants with acute respiratory infection: the ReSVinet Scale. PLoS One **2016**; 11:e0157665.

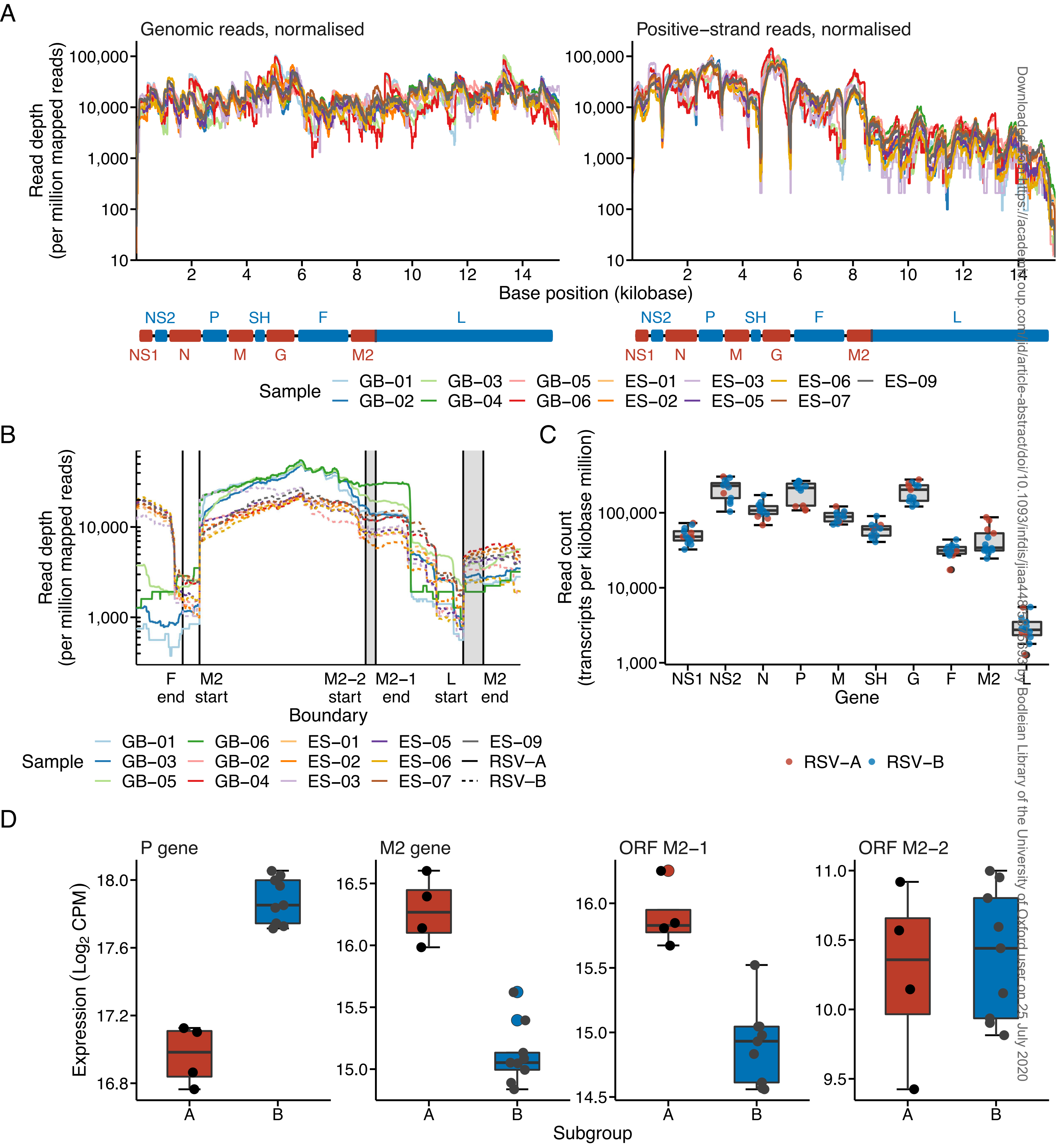
12. Collins PL, Fearns R, Graham BS. Respiratory syncytial virus: virology, reverse genetics, and pathogenesis of disease. Curr Top Microbiol Immunol **2013**; 372:3–38.

13. Richard CA, Rincheval V, Lassoued S, et al. RSV hijacks cellular protein phosphatase 1 to regulate M2-1 phosphorylation and viral transcription. PLoS Pathog **2018**; 14:e1006920.

14. White LJ, Waris M, Cane PA, Nokes DJ, Medley GF. The transmission dynamics of groups A and B human respiratory syncytial virus (hRSV) in England & Wales and Finland: seasonality and cross-protection. Epidemiol Infect **2005**; 133:279–89.

15. Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet **2009**; 10:57–63.


A
Deduplicated RSV read count ($\log_{10}$)
6
5
4
3
2
2
3
4
5
6
RSV viral load ($\log_{10}$ copies/mL)
$R^2 = 0.74$
$P < 10^{-5}$
RSV genome recovery
● Complete
● Partial
B
RSV−A
Genotype
● ON1
● NA1
● GA2
● GA5
● Outgroup
0.01
GB−01
GB−05
US−2013
AU−2016
KE−2013
KE−2012
KE−2013
TH−2012
GB−06
GB−03
NZ−2012
JO−2012
PE−2010
JO−2012
JO−2012
PE−2010
AU−2010
NL−2008
DE−2008
IT−2009
PE−2010
PH−2013
BE−2006
DE−2009
NL−2007
DE−2009
BE−2008
NL−2009
CN−2008
US−2013
US−2013
IT−2009
IT−2009
US−1977
C
RSV−B
Genotype
● BA9
● BA
● GB
● Outgroup
0.01
ES−05
ES−03
ES−01
ES−02
GB−04
AU−2016
ES−09
GB−02
ES−07
PH−2014
ES−06
US−2013
GB−2013
GB−2012
GB−2013
JO−2013
NZ−2012
GB−2012
GB−2012
NL−2012
NL−2012
US−2013
NL−2012
DE−2009
BE−2006
PE−2012
IT−2009
IT−2009
BE−2008
IT−2009
DE−2008
BE−2008
BE−2008
KE−2010
BR−2010
GB−2013
DE−2008
KE−2012
US−1977


A
Genomic reads, normalised
Positive−strand reads, normalised
Read depth
(per million mapped reads)
100,000
10,000
1,000
100
10
2
4
6
8
10
12
14
Base position (kilobase)
NS2
P
SH
F
L
NS1
N
M
G
M2
Sample
GB−01
GB−02
GB−03
GB−04
GB−05
GB−06
ES−01
ES−02
ES−03
ES−05
ES−06
ES−07
ES−09
B
Read depth
(per million mapped reads)
10,000
1,000
F end
M2 start
M2−2 start
M2−1 end
L start
M2 end
Boundary
Sample
GB−01
GB−03
GB−05
GB−06
GB−02
GB−04
ES−01
ES−02
ES−03
ES−05
ES−06
ES−07
ES−09
RSV−A
RSV−B
C
Read count
(transcripts per kilobase million)
100,000
10,000
1,000
NS1
NS2
N
P
M
SH
G
F
M2
L
Gene
RSV−A
RSV−B
D
P gene
M2 gene
ORF M2−1
ORF M2−2
Expression (Log$_2$ CPM)
18.0
17.6
17.2
16.8
16.5
16.0
15.5
15.0
16.0
15.5
15.0
14.5
11.0
10.5
10.0
9.5
A
B
Subgroup