Long-read RNA sequencing identifies polyadenylation elongation and differential transcript usage of host transcripts during SARS-CoV-2 *in vitro* infection

Jessie J.-Y. Chang[1], Josie Gleeson[2], Daniel Rawlinson[1], Miranda E. Pitt[1], Ricardo De Paoli-Iseppi[2], Chenxi Zhou[3], Francesca L. Mordant[1], Sarah L. Londrigan[1], Michael B. Clark[2], Kanta Subbarao[1,4], Timothy P. Stinear[1] and Lachlan J.M. Coin[1,3,5,6,7*]

[1]Department of Microbiology and Immunology, University of Melbourne at The Peter Doherty Institute for Infection and Immunity, Melbourne, VIC 3000, Australia

[2]Centre for Stem Cell Systems, Department of Anatomy and Physiology, University of Melbourne, Melbourne, VIC 3010, Australia

[3]Department of Clinical Pathology, University of Melbourne, Melbourne, VIC 3000, Australia

[4]WHO Collaborating Centre for Reference and Research on Influenza, Peter Doherty Institute for Infection and Immunity, Melbourne, VIC 3000, Australia

[5]Department of Infectious Disease, Imperial College London, London SW7 2AZ, UK

[6]Institute for Molecular Bioscience, University of Queensland, Brisbane, QLD 4072, Australia

[7]Lead contact

***Correspondence: lachlan.coin@unimelb.edu.au**

Abstract

Better methods to interrogate host-pathogen interactions during Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) infections are imperative to help understand and prevent this disease. Here we implemented RNA-sequencing (RNA-seq) combined with the Oxford Nanopore Technologies (ONT) long-reads to measure differential host gene expression, transcript polyadenylation and isoform usage within various epithelial cell lines permissive and non-permissive for SARS-CoV-2 infection. SARS-CoV-2-infected and mock-infected Vero (African green monkey kidney epithelial cells), Calu-3 (human lung adenocarcinoma epithelial cells), Caco-2 (human colorectal adenocarcinoma epithelial cells) and A549 (human lung carcinoma epithelial cells) were analysed over time (0, 2, 24, 48 hours). Differential polyadenylation was found to occur in both infected Calu-3 and Vero cells during a late time point (48 hpi), with Gene Ontology (GO) terms such as viral transcription and translation shown to be significantly enriched in Calu-3 data. Poly(A) tails showed increased lengths in the majority of the differentially polyadenylated transcripts in Calu-3 and Vero cell lines (up to ~101 nt in mean poly(A) length, padj = 0.029). Of these genes, ribosomal protein genes such as *RPS4X* and *RPS6* also showed downregulation in expression levels, suggesting the importance of ribosomal protein genes during infection. Furthermore, differential transcript usage was identified in Caco-2, Calu-3 and Vero cells, including transcripts of genes such as *GSDMB* and *KPNA2*, which have previously been implicated in SARS-CoV-2 infections. Overall, these results highlight the potential role of differential polyadenylation and transcript usage in host immune response or viral manipulation of host mechanisms during infection, and therefore, showcase the value of long-read sequencing in identifying less-explored host responses to disease.

41 **Introduction**

42 The Severe Acute Respiratory Corona Virus 2 (SARS-CoV-2) was first discovered in Wuhan, China at the
43 end of 2019 and is the causative agent of the global Coronavirus Disease 2019 (COVID-19) pandemic. The
44 World Health Organisation (WHO) reported over 5.1 million deaths and over 258 million confirmed cases
45 globally as of late November 2021, and the global health, social and economic burden due to this disease
46 continues to grow. Extensive research on this virus has been carried out since the first discovery of the
47 pathogen. Nevertheless, continued exploration of the host response during an infection with SARS-CoV-2
48 is imperative for developing novel therapeutics, diagnostics, and prophylactics.

49

50 The host response to SARS-CoV-2 infection has been comprehensively studied within the past two years.
51 This includes transcriptomic studies of the host using RNA sequencing (RNA-seq) from *in vitro* infections
52 of cell lines/primary cells, *in vivo* infection models in ferrets as well as clinical samples from infected
53 patients (Blanco-Melo et al., 2020; Mick et al., 2020; Wu et al., 2020). Of these, *in vitro* SARS-CoV-2
54 infection studies using continuous cell lines have been commonly used, due to the simplicity of the model.
55 Vero (African green monkey kidney epithelial) cells are known for their high susceptibility to SARS-CoV-
56 2, due to their defective interferon I responses (Emeny & Morgan, 1979). However, due to the lack of
57 biological relevance using these cells, human epithelial cells have mostly been used for assessing host
58 responses instead of Vero cells, such as Calu-3 (human lung adenocarcinoma epithelial), Caco-2 (human
59 colorectal adenocarcinoma epithelial) and A549 (human lung carcinoma epithelial) cells. SARS-CoV-2-
60 infected Calu-3 cells exhibited upregulation of genes involved in innate immune response to viral infections
61 such as *IFIT2*, *OAS2*, or *IFNB1*, similar to the responses elicited by the SARS-CoV-1 virus (Blanco-Melo
62 et al., 2020; Wyler et al., 2021). Also, in both Calu-3 and Caco-2 cells, genes involved in response to
63 Endoplasmic Reticulum (ER) stress and mitogen-activated protein (MAP) kinases were upregulated during
64 infection (Wyler et al., 2021). However, responses between Calu-3 and Caco-2 were found to be cell-
65 specific. Caco-2 cells lacked in innate immune responses when infected with SARS-CoV-1/2 (Chen et al.,
66 2021; Shuai et al., 2020; Wyler et al., 2021), and have shown fewer changes at the gene (Wyler et al., 2021)
67 and protein level (Saccon et al., 2021) compared to Calu-3 cells. Furthermore, A549 cells have shown lack
68 of susceptibility to SARS-CoV-2, despite being a human airway epithelial cell line like Calu-3 cells
69 (Blanco-Melo et al., 2020; Harcourt et al., 2020). This has been attributed to the lack of the main entry
70 receptor of SARS-CoV-2 - Angiotensin-Converting Enzyme 2 (*ACE2*) – on the surface of these cells.
71 However, air-liquid interface culturing or *ACE2*-expressing A549 (A549-hACE2) cells enhanced the
72 susceptibility to SARS-CoV-2 (Sasaki et al., 2021; Xie et al., 2020). Overall, host responses appeared to

vary between different epithelial cell lines and were dependent on the multiplicity of infection (MOI) of the virus in A549-h*ACE2* cells (Blanco-Melo et al., 2020).

Most RNA-seq data reported in the literature have been generated using short-read sequencing methods such as Illumina sequencing (Bibert et al., 2021; Islam et al., 2021). In these studies, differential expression and Gene Ontology (GO)/ Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analyses have been the main outcomes. Short-read RNA-seq is an effective technique for measuring differential mRNA abundance. However, utilising a long-read sequencing platform provides the ability to discern other functionally significant mRNA features such as length of the poly(A) tails, alternative splicing, and differential isoform usage (de Jong et al., 2017; De Paoli-Iseppi, Gleeson, & Clark, 2021; Gleeson et al., 2021; Workman et al., 2019). These additional mRNA features have been linked with different disease states (Curinha, Oliveira Braz, Pereira-Castro, Cruz, & Moreira, 2014; Dick et al., 2020; Tazi, Bakkour, & Stamm, 2009). However, these events have not been studied in depth for infectious diseases, especially with SARS-CoV-2 infections. An ability to measure full-length transcripts, polyadenylation status and isoform usage would permit significantly enriched insights into host responses to viral infection than standard RNA-seq methods allow.

Here we report the use of RNA-seq methods from the Oxford Nanopore Technologies (ONT) platform (direct RNA, direct cDNA and PCR cDNA) to carry out an in-depth investigation into the host response to SARS-CoV-2 *in vitro*. The responses were visualised throughout a time-course (0, 2, 24 and 48 hours post infection (hpi)) using four epithelial cell lines (Vero, Calu-3, Caco-2 and A549). Previously we performed a comprehensive analysis of the viral response for some of these datasets (Chang et al., 2021). In this current study, we investigated differential polyadenylation and transcript usage between infected and mock control cells. Additionally, we were interested in whether long-read differential expression analysis conveyed similar differential expression results to short-read RNA-seq studies shown in literature. Overall, our study demonstrated the value of long-read sequencing in identifying less-explored host responses to disease.

103 **Methods**

104 **Data availability**

105 ONT sequencing data (direct RNA and direct cDNA) for this study from cell lines (Vero, Caco-2 and Calu-
106 3) was derived from our previous work (Chang et al., 2021), and is currently publicly available at NCBI
107 repository BioProject PRJNA675370. Additional datasets were generated for this study including PCR
108 cDNA datasets for cell lines (Vero, Caco-2, Calu-3 and A549) and the direct RNA and direct cDNA datasets
109 for A549. These datasets are also available at NCBI repository BioProject PRJNA675370. The results of
110 individual analyses are available at Figtree DOI: 10.6084/m9.figshare.17139995 (differential expression),
111 10.6084/m9.figshare.16841794 (differential polyadenylation) and 10.6084/m9.figshare.17140007
112 (differential transcript usage).

113

114 **Experimental methods**

115 **Cell culture and RNA extraction/preparation**

116 Cell culture and RNA extraction/preparation methods have been described previously (Chang et al., 2021)
117 for Calu-3 (human lung adenocarcinoma epithelial - ATCC HTB-55), Caco-2 (human colorectal
118 adenocarcinoma epithelial - ATCC HTB-37) and Vero (African green monkey kidney epithelial - ATCC
119 CCL-81) cells. For this current study, we additionally cultured A549 (human lung carcinoma epithelial –
120 ATCC CCL-185) cells to supplement our main data, using similar methods. Briefly, A549, Vero, Calu-3
121 and Caco-2 cell lines were cultured in T75 flasks and maintained at 37 °C and 5% (v/v) CO_2. A549 cells
122 were cultured with Ham's F-12K (Kaighn's) Medium (Gibco) supplemented with 10% FBS, 4 mM L-
123 glutamine (Media Preparation Unit, The Peter Doherty Institute for Infection and Immunity (Doherty
124 Institute)), 100 IU penicillin, 10 µg streptomycin/mL, 1X non-essential amino acids (Gibco-BRL) and 50
125 µM B-mercaptoethanol (Life Technologies). All cell lines were seeded in 4 x 6-well tissue-culture plates
126 and maintained at 70-80% confluency for infection. Three wells of the 6-well plates were infected with
127 SARS-CoV-2 (Australia/VIC01/2020) at a MOI of 0.1 and the remaining wells were used as mock controls
128 for four time points (0, 2, 24 and 48 hpi). Total cellular RNA was extracted with the RNeasy Mini Kit
129 (Qiagen), treated with the Turbo DNAse-free Kit (Invitrogen) and purified with RNAClean XP magnetic
130 beads (Beckman Coulter). The final resulting RNA was eluted in nuclease-free water. Quality control was
131 carried out using NanoDrop 2000C (Thermo Fisher Scientific), Bioanalyzer 2100 (Agilent Technologies)
132 and Qubit 4 Fluorometer (Invitrogen).

133

134 **Library preparation and sequencing**

135 Library preparation and sequencing methods have been described previously (Chang et al., 2021). Briefly,
136 RNA from mock control and infected cells harvested at 0, 2, 24 and 48 hpi from Caco-2, Calu-3 and Vero
137 cells was sequenced with the ONT Direct cDNA Sequencing Kit (SQK-DCS109) in conjunction with the
138 Native Barcoding Kit (EXP-NBD104). RNA harvested at 2, 24 and 48 hpi was sequenced with the Direct
139 RNA Sequencing Kit (SQK-RNA002) by pooling the RNA from replicate wells. For this current study,
140 RNA from A549 cells was sequenced as per our previous work with minor modifications in the number of
141 time points sequenced, to supplement our main data. The ONT Direct RNA Sequencing Kit (SQK-
142 RNA002) was used to prepare 6 µg of pooled total RNA (2 µg RNA from each replicate well) from control
143 and infected cells at the 24 hpi. The Direct cDNA Sequencing Kit (SQK-DCS109) was used in conjunction
144 with the Native Barcoding Kit (EXP-NBD104) to prepare 3 µg of total RNA from all control and infected
145 replicates separately at both 0 and 24 hpi time points. All direct RNA and direct cDNA libraries were loaded
146 onto a R9.4.1 MinION flow cell and sequenced for 72 hrs using an ONT MinION or GridION. Additionally,
147 PCR cDNA long-read sequencing was carried out with RNA from all four cell lines (Vero, Caco-2, Calu-3
148 and A549 cells) using the following methods: cDNA libraries were constructed with the PCR-cDNA
149 Sequencing (SQK-PCS109) and PCR Barcoding (SQK-PBK004) kits using the supplied protocol. RNA
150 samples from 0 and 24 (hpi) were randomised and multiplexed for sequencing in groups of six using
151 sequential barcodes. 100 ng of sample RNA was used for cDNA synthesis. Transcripts were amplified by
152 PCR and barcodes added using the specified cycling conditions with a 7 min extension time and 13x cycles.
153 Amplified samples were individually cleaned using 0.5x AMPure XP beads (Beckman Coulter) and
154 quantified using a Qubit 4 Fluorometer (Invitrogen). The length distribution was determined via the
155 TapeStation 4200 (Agilent Technologies) before pooling. Equimolar amounts of each barcoded sample
156 were pooled to a total of 100 – 200 fmol (assuming median transcript size = 1.1 kb). 100 fmol of final
157 libraries were loaded onto a R9.4.1 MinION flow cell and sequenced for 72 hrs on an ONT GridION. Run
158 metrics were monitored live and if active pores dropped below 200, any remaining library was loaded
159 following a nuclease flush. Synthetic 'sequin' RNA standards, provided in two mixes (A and B) (Hardwick
160 et al., 2016), were added to each sample in direct RNA and PCR cDNA libraries. Mix A and B sequins,
161 diluted 1:250 (approximately 6-10% of estimated total mRNA), were added to infected and control samples,
162 respectively.

163

164

165

166 **Data analysis**

167 Publicly available data from our previous work (Chang et al., 2021) in combination with data generated
168 from this study were analysed using Spartan (Meade, Lafayette, Sauter, & Tosello, 2017) and Nectar from
169 the Australia Research Data Commons.

170

171 **Basecalling, alignment and generating counts files**

172 All FAST5 files were basecalled using standalone *Guppy* v3.5.2
173 (https://community.nanoporetech.com/sso/login?next_url=%2Fdownloads), except PCR cDNA data from
174 Vero cells which were live-basecalled using *Guppy* v3.2.8. All resulting FASTQ data were mapped using
175 *Minimap2* v2.17 (Heng Li, 2018). Direct RNA-seq data was mapped to the combined genome (consisting
176 of human/African green monkey genome from Ensembl (release 100), SARS-CoV-2 Australia virus
177 (Australia/VIC01/2020, NCBI:MT007544.1) and the RNA sequin decoy chromosome genome (Hardwick
178 et al., 2016) with the default direct RNA parameters '-ax splice -uf -k14 --secondary=no' and for all cDNA
179 datasets '-ax splice –secondary=no'. All data were mapped to the respective combined transcriptome using
180 the following parameters – '-ax map-ont'. The resulting BAM files were sorted and indexed using *Samtools*
181 v1.9 (H. Li et al., 2009). Counts files were generated using *Featurecounts* v2.0.0 (Liao, Smyth, & Shi,
182 2014) for genome-mapped cDNA data, and with *Salmon* v0.13.1 (Patro, Duggal, Love, Irizarry, &
183 Kingsford, 2017) for transcriptome-mapped cDNA data.

184

185 **Differential expression analysis**

186 *DESeq2* was used to identify differentially expressed genes/transcripts from direct cDNA data. A minimum
187 expression threshold of five reads per gene/transcript across all the replicates was used. Comparisons
188 between control and infected cells were made per time point (0, 2, 24, 48 hpi) with standard methods. Also,
189 the changes between time points (0-2, 0-24, 2-24, 24-48 hpi) were compared, where the interaction term
190 between control and infected across time points were found using a method by Steven Ge (https://rstudio-
191 pubs-static.s3.amazonaws.com/329027_593046fb6d7a427da6b2c538caf601e1.html#example-4-two-
192 conditionss-three-genotpes-with-interaction-terms). This is a more sensitive method as it calculates the
193 changes between time points in infected cells while accounting for the changes in the expression level in
194 the control cells. All genes/transcripts with padj < 0.05 were regarded as significantly differentially
195 expressed. The heatmap of differentially expressed genes in Caco-2 and Calu-3 at 24 and 48 hpi were
196 generated using *multiGO*. The filters utilised were padj < 0.05, GO p-value < 0.0001, scaled by row, with

197 ten maximum GO terms. Columns with more than 80% of NA's and rows with more than 10% of NA's
198 were excluded.

199

200 **Poly(A) tail length analysis**

201 Two tools were used for poly(A) tail length analysis; *nanopolish* (Simpson et al., 2017) and *tailfindr*
202 (Krause et al., 2019). For the *nanopolish* analysis, all Caco-2, Calu-3 and Vero direct RNA BAM files
203 mapped to the combined reference genome (host, sequin, virus) were indexed with the *nanopolish* v0.13.2
204 'index' function with the command 'nanopolish index -d $FAST5 -s $SEQUENCING_SUMMARY
205 $FASTQ'. The poly(A) tail lengths of each read were estimated using the 'polya' function with default
206 parameters 'nanopolish polya --reads $FASTQ --bam $SORTED_BAM --genome
207 $COMBINED_REFERENCE_GENOME > output.tsv'. The host reference sequence names were extracted
208 from the reference file with the following command 'cat $REFERENCE_GENOME | grep '>' | cut -d ' ' -f
209 1 | cut -f 2 -d '>''. Using this name file, the host reads were extracted from the final *nanopolish* TSV file by
210 this command 'awk 'NR==FNR{A[$1]; next} $2 in A' $NAMES.TSV $TSV'.

211

212 After the poly(A) lengths were determined, duplicates were removed, and the outputs were merged with a
213 file generated by an in-house pipeline – *npTranscript* - which allowed read names to be associated with
214 Ensembl ID's. The data were grouped per Ensembl ID and whether they were mitochondrial or non-
215 mitochondrial genes, and the median poly(A) lengths were calculated. Differential polyadenylation
216 between the overall median lengths of control vs infected cells per cell line were determined with p-values
217 using Wilcoxon's test of ranks for Ensembl ID's with more than one entry.

218

219 As the *nanopolish* results revealed evidence of differential polyadenylation between the overall median of
220 control and infected poly(A) lengths, *tailfindr* analysis was utilised to gather more evidence at a gene level.
221 For *tailfindr* analysis, direct cDNA datasets from Vero, Calu-3 and Caco-2 were utilised. Basecalled FAST5
222 files were subsetted by read ID's derived from demultiplexed FASTQ files. Each replicate was passed
223 through *tailfindr* v0.1.0 separately. The median poly(A) and poly(T) lengths were calculated per gene and
224 grouped by whether they were mitochondrial or non-mitochondrial.

225

226 The Pearson product-moment correlations between the median poly(A) and poly(T) lengths per gene from
227 *tailfindr* analyses were compared for 2, 24 and 48 hpi datasets from Caco-2, Calu-3 and Vero datasets.
228 Additionally, the Spearman's correlations between the *tailfindr* poly(T)/(A) and *nanopolish* poly(A)
229 median lengths per gene (in control and infected cells) were compared for Calu-3 48 hpi datasets via the
230 'cor.test' function in the *stats* package in R.

231

232 *tailfindr* poly(T) results were used for the main polyadenylation linear mixed-model analysis as replicate
233 information was able to be preserved and showed higher correlation to *nanopolish* poly(A) lengths
234 compared with *tailfindr* poly(A) lengths. The raw poly(T) lengths were log-transformed due to the right-
235 skew distribution and data with at least 6 entries were selected. Then, the package *lmerTest* v3.1-3
236 (Kuznetsova, Brockhoff, & Christensen, 2017) was used to derive a linear mixed-effects regression (lmer)
237 and therefore calculate the effect of SARS-CoV-2 infection compared with control mock-infected cells.
238 The p-values were generated per gene and Benjamini-Hochberg adjusted using the 'p.adjust' function in R,
239 which were filtered by padj < 0.05. Raincloud plots were generated for median poly(T) lengths of each
240 gene with increased poly(A) length in the Calu-3 48 hpi dataset in both conditions (control and infected)
241 using using *ggplot2* v3.3.4 (Wickham, 2016) to replicate the raincloud plots generated by the *raincloudplots*
242 package in R (Allen et al., 2021).

243

244 To test whether the same significant genes in the mixed-model analysis appeared in *nanopolish* Calu-3 48
245 hpi poly(A) data, the raw tail lengths were log-transformed and the median lengths per gene were tested
246 between control and infected cells using Wilcoxon's test of ranks, where p-values were adjusted using
247 Benjamini-Hochberg adjustment as above.

248

249

250 **Differential transcript usage analysis**

251 Counts from *Salmon* using transcriptome-mapped BAM files were used to determine the differential
252 transcript usage of transcripts between control and infected conditions for each cell line and time point.
253 The counts were input into *DRIMSeq* v1.16.1 (Nowicka & Robinson, 2016) and filtered by conditions
254 (min_samps_gene_expr = 6, min_samps_feature_expr = 3, min_gene_expr = 10, min_feature_expr = 10).
255 The output was used for stage-wise analysis using *StageR* v1.10.0 (Van Den Berge, Soneson, Robinson,
256 & Clement, 2017), where the final list of significant genes and transcripts was filtered by padj < 0.05.

257 **GO and KEGG pathway analysis**

258 Significant biological GO biological terms and KEGG pathways were identified with genes that were found
259 to be significantly differentially expressed and polyadenylated in the analyses above. For differential
260 expression analysis, genes found to be differentially expressed in direct cDNA datasets for each condition
261 and time point were used for analysis. For differential polyadenylation analysis, genes that were found to
262 be increased and decreased in poly(A) length in Calu-3 48 hpi direct cDNA dataset were used for analysis.
263 All pathway analyses were carried out using a novel shiny-app *multiGO*
264 (http://coinlab.mdhs.unimelb.edu.au/multigo). *multiGO* uses a hypergeometric test against a background of
265 all genes included in the *GO annotation database* v100. For differential expression, thresholds of padj <
266 0.05 and enrichment p-value < 1E-6 in at least one dataset were used for generating the GO plot, and
267 thresholds of padj < 0.05 and enrichment p-value < 0.0001 were used for generating the KEGG plot. Non-
268 significant bubbles were also shown
269 (http://coinlab.mdhs.unimelb.edu.au/multigo/?subdir=multigo/multiGO&file=DESeq2.zip). All terms with
270 padj < 0.05, enrichment p-value < 0.05 and at least two genes were deemed as significant for the analysis.
271 For differential polyadenylation and differential polyadenylation vs expression analyses, thresholds of padj
272 < 0.05 and enrichment p-value < 0.0001 were used
273 (http://coinlab.mdhs.unimelb.edu.au/multigo/?subdir=multigo/multiGO&file=DP_48hpi_no_filter_merge
274 d.zip).

275

276 **Differential expression vs differential polyadenylation**

277 Using a hypergeometric test, the probability of obtaining greater than or equal to two genes overlapping
278 between the differential expression and polyadenylation analyses were tested. Counts of downregulated
279 genes (padj < 0.05) from Calu-3 48 hpi datasets from *DESeq2* analysis were set as m=253. Counts of genes
280 with elongated poly(A) tails were set as k=13. Genes which were both downregulated and increased in
281 poly(A) length were set as n=2. The total background count was set as N=15,426. The code used for the
282 probability calculation was 'phyper(x,k,15426-k,m,lower.tail=F) + dhyper(x,k,15426-k,m)' and was
283 carried out in R. Both mitochondrial and non-mitochondrial genes were included in this calculation.

284

285

286

287

288 **Results**

289 **Viral burden changes between different cell lines and over time**

290 We utilised the percentage of mapped reads to host or virus to assess the level of infection in each cell line
291 across different datasets and over time (**Data S1**) to add to our earlier study (Chang et al., 2021). Between
292 0 and 2 hpi, the percentage of viral reads were minimal (< 0.1% of all reads) across each of the cell lines.
293 At 24 hpi, differences between the cell lines started to appear, with Vero cells leading in infection with
294 ~45% of reads mapping to virus, followed by Caco-2 (~2.3%), Calu-3 (~2%), and A549 (< 0.01%), as
295 measured with direct cDNA datasets (**Data S1**). The relative proportions of viral transcripts between these
296 four cell lines were aligned with the results in literature at the 24 hpi (Saccon et al., 2021). The final time
297 point (48 hpi) showed the greatest per-cell-line infection in Caco-2 (~12.5%) and Calu-3 (~3.7%) cells but
298 lowered in percentage in Vero cells (~25%) compared with 24 hpi. These results agreed with the idea that
299 the infection peaked at 24 hpi in Vero cells as shown by our previous study (Chang et al., 2021).
300 Interestingly, the percentage of reads mapping to virus were markedly higher in direct RNA datasets
301 compared with the direct cDNA and PCR cDNA datasets at the 24 hpi (**Table 1**). The reason for this may
302 be due to the direct RNA method involving the sequencing of the mRNA molecule, instead of the reverse-
303 transcribed cDNA, as in the direct cDNA and PCR cDNA methods. This would remove any biases caused
304 by the reverse-transcription. These results suggested that measuring viral infection using more than one
305 ONT RNA-seq approach may be more beneficial to accurately gauge the level of viral RNA in the sample.

306

307 **Table 1. Proportions of average viral reads in 24 hpi datasets in Vero, Calu-3, Caco-2 and A549 cell lines.** Related to **Data**
308 **S1**.

Cell line	Direct RNA (%)	Direct cDNA (%)	PCR cDNA (%)
Vero	74	45	55
Calu-3	4	2	3
Caco-2	4	2	3
A549	0.02	0.01	0.01

309

310

311

312

313 **Cell-type specific changes in host gene expression *in vitro* following virus infection using long-read**
314 **sequencing**

315 The host responses to SARS-CoV-2 have been extensively studied at the gene and protein expression level
316 (Saccon et al., 2021; Shah, Firmal, Alam, Ganguly, & Chattopadhyay, 2020). As long-read sequencing
317 enables full-length transcripts to be sequenced unlike short-read sequencing, we were interested in whether
318 our long-read differential expression results would reveal similar results to existing studies (Blanco-Melo
319 et al., 2020; Chen et al., 2021; Saccon et al., 2021; Sun et al., 2021). The direct cDNA datasets were used
320 for differential expression analysis as it included data from all four time points (0, 2, 24, 48 hpi) in Calu-3,
321 Caco-2 and Vero cells, and two time points (0 and 24 hpi) in A549 cells.

322

323 It is well-known that A549 cells are invulnerable to SARS-CoV-2, due to the lack of *ACE2* receptors
324 (Blanco-Melo et al., 2020). However, *ACE2* is expressed in varying degrees in different human tissues (M.-
325 Y. Li, Li, Zhang, & Wang, 2020) and is expressed relatively poorly in the respiratory tract (M.-Y. Li et al.,
326 2020). Previous studies have shown the expression of *ACE2* at the gene (Puray-Chavez et al., 2021) and
327 protein (Saccon et al., 2021) level in Vero, Calu-3 and Caco-2 cells. Furthermore, the importance of the
328 protease *TMPRSS2* during SARS-CoV-2 has been noted (Hoffmann et al., 2020). Our long-read data were
329 in line with some of these results, where no transcripts mapped to *ACE2* and *TMPRSS2* in A549 cells.
330 However, we observed the absence/low expression of *ACE2* (< 5 reads per replicate) and *TMPRSS2* (< 25
331 reads per replicate) genes across all our susceptible cell lines. The presence of these transcripts correlated
332 with the viral burden observed in each cell line from our previous study (**Data S1**) (Chang et al., 2021).

333

334 In all cell lines, the earlier infection time points (0 and 2 hpi) showed little significant differential expression
335 as expected, given the short period of infection in which host responses could be elicited. We observed an
336 increase in significantly upregulated genes in Calu-3 and Vero cell lines at 24 hpi (**Figure 1**). At the final
337 time point (48 hpi), we noted an increase in downregulated genes as well as the presence of upregulated
338 genes in Calu-3, Vero and Caco-2 cell lines. In line with previous studies (Saccon et al., 2021; Wyler et al.,
339 2021), while Calu-3 and Vero cells exhibited clear changes in transcriptional activity throughout the final
340 two time points, Caco-2 cells revealed little differential expression activity (**Figure 1 & Table 2**). These
341 results were recapitulated in a second measurement of gene expression changes (**Figure S1**). In this
342 combined analysis, the differences between the host gene expression of control and infected cells across
343 two time points were measured (interaction term – see **Methods**) as opposed to differences at each
344 individual time point.

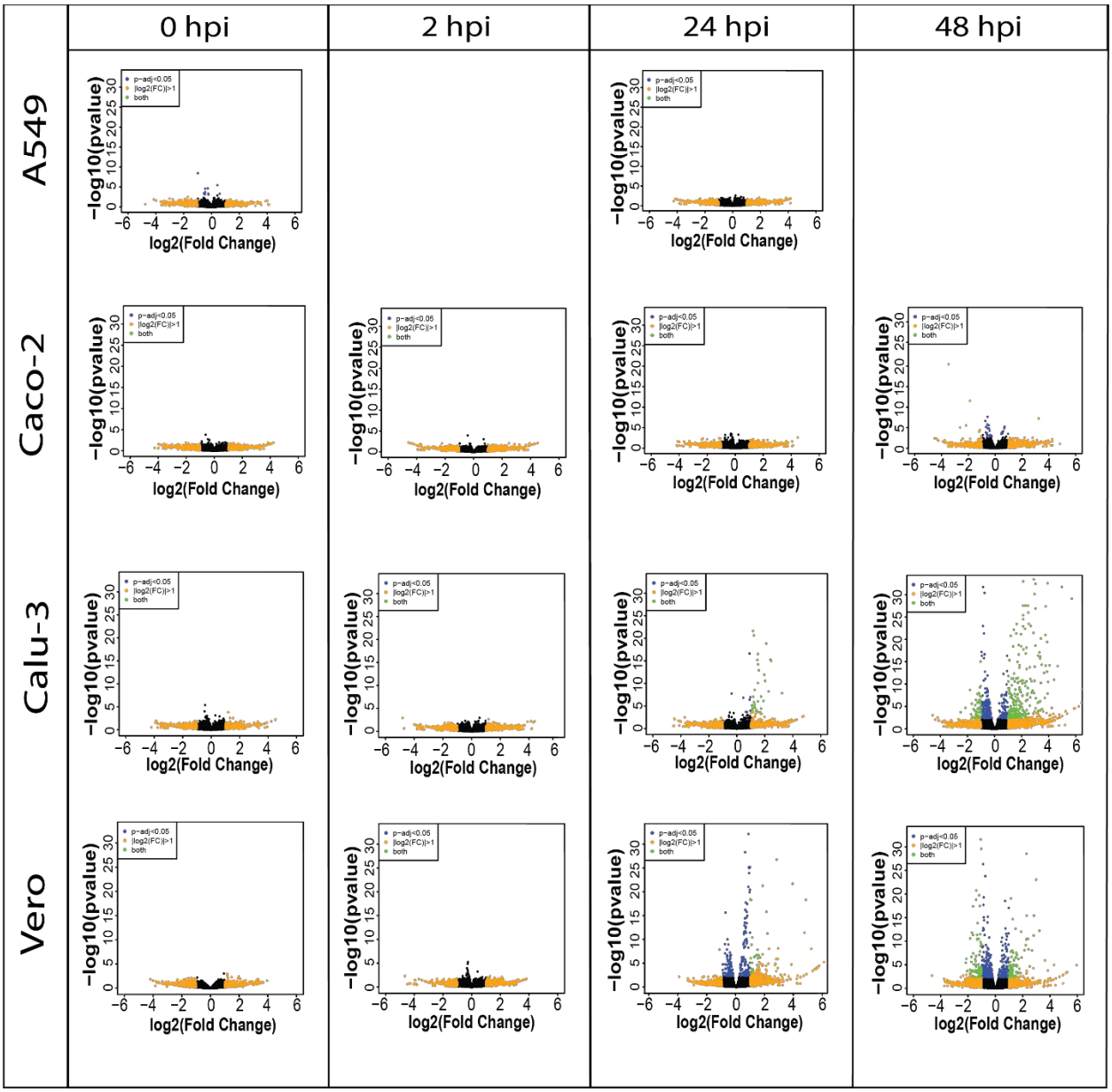


Figure 1. Volcano plots show the difference in expression level between control and infected cells per cell line (A549, Caco-2, Calu-3 and Vero) in direct cDNA datasets using *DESeq2*. X-axis represents log2FC and Y-axis displays –log10 p-adjusted value (padj). • padj < 0.05 (blue), • |log2FC |> 1 (orange), • both (green). Increased number of differentially expressed genes are shown in latter time points in Caco-2, Calu-3 and Vero cell lines, and a lack of differential expression was observed in A549 cells. Related to **Figure S1 & Table 2**.

355 **Table 2. The number of significantly differentially expressed genes (padj < 0.05) between control and infected cells in A549,**
356 **Caco-2, Calu-3 and Vero cells over 2, 24 and 48 hpi in direct cDNA datasets.** Related to **Figure 1 & S1**.

Cell line	Time point	Up	Down
A549	24	0	0
Caco-2	2	0	0
Caco-2	24	0	0
Caco-2	48	8	17
Calu-3	2	0	0
Calu-3	24	41	1
Calu-3	48	371	253
Vero	2	0	0
Vero	24	134	111
Vero	48	210	290

357

358

359

360 **Calu-3 and Caco-2 cells show distinct gene expression level patterns**

361 As the initial differential expression results showed differences between Calu-3 and Caco-2 cell lines, we
362 then investigated the similarity of gene expression patterns between the two cell lines at the 24 and 48 hpi
363 via a heatmap (**Figure 2**). Following the results in literature (Chen et al., 2021; Wyler et al., 2021), our
364 results also showed higher relative expression of interferon-related genes such as *IFI6* and *IFITM3* in
365 infected Calu-3 cells compared with Caco-2 cells. This was observed especially at the 48 hpi time point. In
366 contrast, Caco-2 cells revealed higher expression of ribosomal protein genes as well as mitochondrial genes
367 compared with Calu-3 cells. Therefore, our long-read results supported the idea that Caco-2 and Calu-3
368 cells have distinct gene expression level patterns, and that Caco-2 cells have diminished innate immune
369 responses in contrast to Calu-3 cells.

370

371

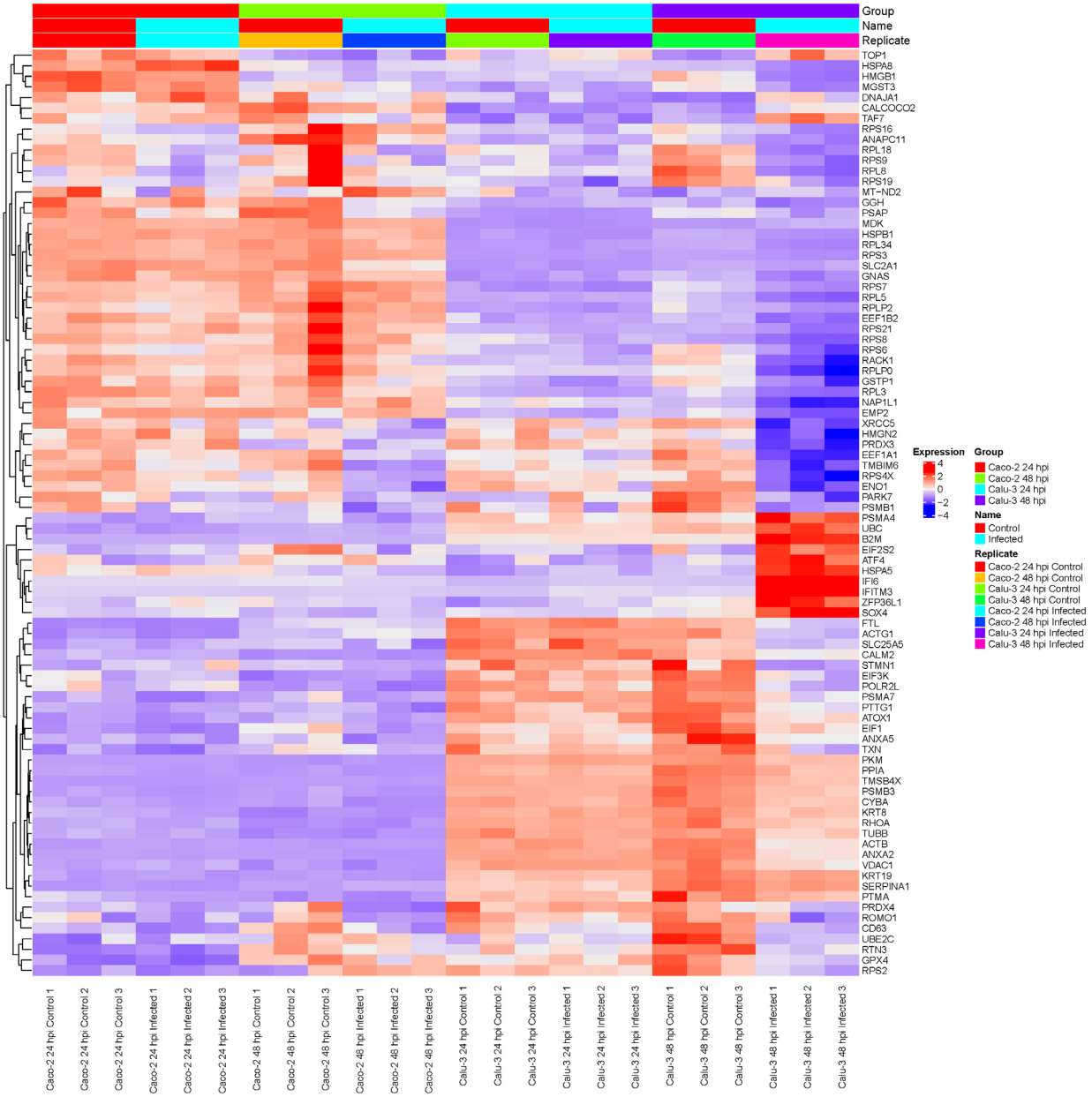


Figure 2. Heatmap of relative gene expression in Caco-2 and Calu-3 cells at 24 and 48 hpi using direct cDNA datasets reveal distinct gene expression profiles in each cell line, analysed by *DESeq2* and visualized by *multiGO*. The data was filtered by padj < 0.05, enrichment p-value < 0.0001. The expression levels were scaled per row and organised based on relevant GO terms. In Calu-3 cells, higher relative expression of interferon-response genes was observed compared with Caco-2 cells. In contrast, Caco-2 cells showed higher relative expression of ribosomal protein and mitochondrial genes. Related to **Figure 1 & S1 & Table 1**.

381 **GO and KEGG analyses reveal similarities between Calu-3, Caco-2 and Vero cells**

382 To investigate the cell-specific gene expression changes at a deeper level, we wondered whether any
383 enrichment of pathways was shared between multiple cell types. By utilising the genes which were
384 significantly differentially expressed in the direct cDNA data, GO biological and KEGG pathway analyses
385 were carried out using a new visualisation tool *multiGO* (http://coinlab.mdhs.unimelb.edu.au/multigo)
386 (**Figures 3 & S2**).

387

388 Amongst many enriched GO biological terms, *reactive oxygen species (ROS) metabolic process* was
389 enriched in all three SARS-CoV-2 susceptible cell lines. Also, we found that only *neutrophil degranulation*
390 was commonly enriched exclusively in the two human cell lines and absent in Vero cells. In contrast, a
391 greater number of terms were shared between Calu-3 and Vero cell lines. These terms included *response to*
392 *virus*, *positive regulation of interferon-alpha production* and *translation* (**Figure 3 & Data S2**).

393

394 As expected, the Calu-3 cell line showed an increase in innate immune responses, with the strongest GO
395 enrichment for various terms associated with host immune responses to pathogens. This included terms
396 such as *defense response to virus* and *type I interferon signalling pathway* (**Figure 3 & Data S2**). As shown
397 above with the gene expression results, these responses were either absent or lacking in Caco-2 cells
398 compared with Calu-3 cells. Additionally, some unique GO terms were enriched in Vero cells. This
399 included *positive regulation of establishment of protein localization to telomere* (**Figure 3 & Data S2**).

400

401 Similarly, Calu-3 cells presented with the strongest significant enrichment of KEGG pathways (**Figure S2**
402 **& Data S2**). In our data, the enriched pathways were related to viral infections such as *influenza A* (*MX1,*
403 *OAS1, OAS2, OAS3, RSAD2, STAT1*) and *measles* (*MX1, OAS1, OAS2, OAS3, STAT1*). These pathways
404 were upregulated at the 24 and 48 hpi time points as well as between 2 vs 24 hpi and 24 vs 48 hpi datasets
405 in infected cells compared with control cells (**Figure S2**). *DDX58* was also observed as upregulated in these
406 pathways in the same datasets except for 2 vs 24 hpi. This has also been shown in influenza A studies
407 (Watson et al., 2020) and the gene has been shown to encode a cytosolic sensor for other coronaviruses ((J.
408 Li, Liu, & Zhang, 2010). The *coronavirus disease* pathway was enriched in both Calu-3 and Vero cells
409 (**Table 3**). Also, as shown previously in various infected epithelial cell lines (Martinelli, Akhmedov, &
410 Kwee, 2021), pathways related to neurological diseases such as Alzheimer's, Parkinson's and Huntington's
411 diseases were found to be enriched in both Vero and Calu-3 cells. The majority of genes in these pathways

412 were downregulated (**Figure 3 & Data S2**). Overall, our long-read RNA-seq data were aligned with results
413 from previous studies which utilised short-read RNA-seq.

414

415

416 **Table 3. Genes involved in the *coronavirus pathway* shown via KEGG pathway analysis.** Related to **Figure S2**.

Cell line	Time point	Common genes
Calu-3	24	*ISG15, MX1, OAS1, OAS2, OAS3, STAT1*
	48	
	2 vs 24	
	24 vs 48	
Vero	24	*FAU, RPL27, RPL28, RPLP2, RPS14, RPS16, UBA52*
	2 vs 24	
Vero	48	*NFKBIA, RPS8, STAT1, HBEGF*
	24 vs 48	

417

418

419

420

421

Figure 3. GO biological terms of differentially expressed genes in Calu-3, Caco-2 and Vero direct cDNA datasets analysed by *DESeq2* and visualised with *multiGO*. Results include datasets (in order); Caco-2 24 vs 48, Caco-2 48, Calu-3 2 vs 24, Calu-3 24 vs 48, Calu-3 24, Calu-3 48, Vero 2 vs 24, Vero 24 vs 48, Vero 24 and Vero 48 hpi. Commonly enriched GO terms across the three cell lines involved the *ROS metabolic process*. The bubble size and colour indicate the -log10 enrichment p-value and fraction of upregulated genes, respectively. Thresholds of padj < 0.05 and enrichment p-value < 1E-6 in at least one dataset were used for generating the plot, where insignificant bubbles are also shown on the plot. All terms with padj < 0.05, enrichment p-value < 0.05 involving at least two genes were deemed as significant for the analysis. Related to **Figure S2 & Data S2**.

435 **Lengths of host mRNA poly(A) tails change during SARS-CoV-2 infection**

436 Polyadenylation has been previously shown in literature to be critical for many different cellular functions.
437 The process promotes stabilisation of the RNA transcript (Beckel-Mitchener, 2002), trafficking into the
438 cytoplasm (Fuke & Ohno, 2007), and translation into proteins (Park, Yi, Kim, Chang, & V, 2016).
439 Furthermore, 3' UTRs can include binding sites for RNA-binding proteins (RBPs) (Gebauer, Preiss, &
440 Hentze, 2012) and microRNAs (miRNAs) (R. C. Lee, Feinbaum, & Ambros, 1993), which contribute to
441 gene expression. However, only a small number of studies exploring changes in host poly(A) lengths during
442 infections have been carried out to this date (Y. J. Lee & Glaunsinger, 2009). Therefore, we were interested
443 in whether infection of cells with SARS-CoV-2 would elicit changes in polyadenylation of transcripts
444 compared with control cells. The median poly(A) tail lengths of mitochondrial and non-mitochondrial
445 transcripts were compared between control and infected cells at 2, 24 and 48 hpi with two different methods:
446 *nanopolish* and *tailfindr*.

447

448 Firstly, *nanopolish* was used to analyse non-replicate direct RNA datasets (**Table 4**). Although the medians
449 for each condition were similar in some datasets, we observed a significant ($p < 0.05$, Wilcoxon's test of
450 ranks two-tailed approach) poly(A) tail length increase in host non-mitochondrial RNA of infected cells
451 compared with control cells in the 24 and 48 hpi in all susceptible cell lines. No significant change was
452 observed in mitochondrial RNA.

453

454

455

456

457

458

459

460

461

462

463

464 **Table 4. Significance of differential polyadenylation in Caco-2, Calu-3 and Vero cell lines measured by Wilcoxon's test of**
465 **ranks following *nanopolish* analysis of direct RNA datasets.** Significant differential polyadenylation (compared with matched
466 uninfected sample) was observed in all non-mitochondrial datasets ($p < 0.05$), except for the Calu-3 2 hpi dataset (in bold). Absence
467 of significant differential polyadenylation in mitochondrial transcripts was also observed. Related to **Figure S3-4 & Tables S1-2**.

Cell line	Time point	Type (Mitochondrial/Non-Mitochondrial)	Wilcoxon's test of ranks p-value
Caco-2	2	Mitochondrial	W = 623, p-value = 0.5875
Caco-2	24	Mitochondrial	W = 647, p-value = 0.8504
Caco-2	48	Mitochondrial	W = 657, p-value = 0.4629
Caco-2	**2**	**Non-mitochondrial**	**W = 89144178, p-value = 0.005541**
Caco-2	**24**	**Non-mitochondrial**	**W = 99425841, p-value = 0.00266**
Caco-2	**48**	**Non-mitochondrial**	**W = 91874180, p-value < 2.2e-16**
Calu-3	2	Mitochondrial	W = 648, p-value = 1
Calu-3	24	Mitochondrial	W = 603, p-value = 0.9163
Calu-3	48	Mitochondrial	W = 625, p-value = 0.4283
Calu-3	2	Non-mitochondrial	W = 86343421, p-value = 0.06655
Calu-3	**24**	**Non-mitochondrial**	**W = 92324958, p-value < 2.2e-16**
Calu-3	**48**	**Non-mitochondrial**	**W = 56602100, p-value < 2.2e-16**
Vero	2	Mitochondrial	W = 647, p-value = 0.5388
Vero	24	Mitochondrial	W = 686, p-value = 0.3935
Vero	48	Mitochondrial	W = 565, p-value = 0.7233
Vero	**2**	**Non-mitochondrial**	**W = 36539892, p-value = 1.149e-05**
Vero	**24**	**Non-mitochondrial**	**W = 32813975, p-value = 3.194e-15**
Vero	**48**	**Non-mitochondrial**	**W = 34564286, p-value = 1.505e-09**

468

469

470 As *nanopolish* results only used data from non-replicate direct RNA datasets, a second approach was
471 utilised. This involved direct cDNA datasets with triplicates for each condition using *tailfindr* to confirm
472 the results of *nanopolish* at the gene level. As the direct cDNA dataset is double-stranded, either strand of
473 the cDNA can be sequenced. Therefore, information on both poly(A) and poly(T) lengths were obtained,
474 which were weakly correlated (**Figure S3 & Table S1**). When the number of differentially polyadenylated
475 transcripts between control and infected cells were compared with *nanopolish* and *tailfindr* poly(A) and
476 poly(T) methods with the Calu-3 48 hpi non-mitochondrial data (Wilcoxon's test, padj < 0.05), the *tailfindr*
477 poly(A) dataset showed no significant differential polyadenylation (**Table S2**). The lack of significance in
478 the *tailfindr* poly(A) data can be explained by the fact that less data was available from the poly(A) dataset
479 compared to the poly(T) dataset. The full-length reads were comprised of 0.5% of poly(A) and 99.5% of
480 poly(T) strands, which made up ~58% of the total number of detected reads with valid Ensembl ID's. This
481 may be attributed to the process of ONT direct cDNA sequencing, where the motor protein is situated on

482 the 5' end of each strand. This means that the poly(T) sequence is sequenced first, whereas the poly(A)
483 sequence is sequenced last for each respective strand. Therefore, this method of sequencing would lead to
484 higher quantity and accuracy of poly(T) sequences compared with poly(A) sequences (**Table S2**). To test
485 this proposition, we compared the median lengths of poly(A/T) tails per gene between *nanopolish* and
486 *tailfindr* Calu-3 48 hpi datasets via the Spearman's correlation test. Weak significant positive correlations
487 between *nanopolish* poly(A) and *tailfindr* poly(T) datasets were observed ($r = 0.12\text{-}0.27$, p-value < 0.05).
488 In contrast, *nanopolish* poly(A) and *tailfindr* poly(A) data were not significantly correlated (**Figure S4**),
489 leading us to choose poly(T) length as a proxy for direct RNA-inferred poly(A) length.

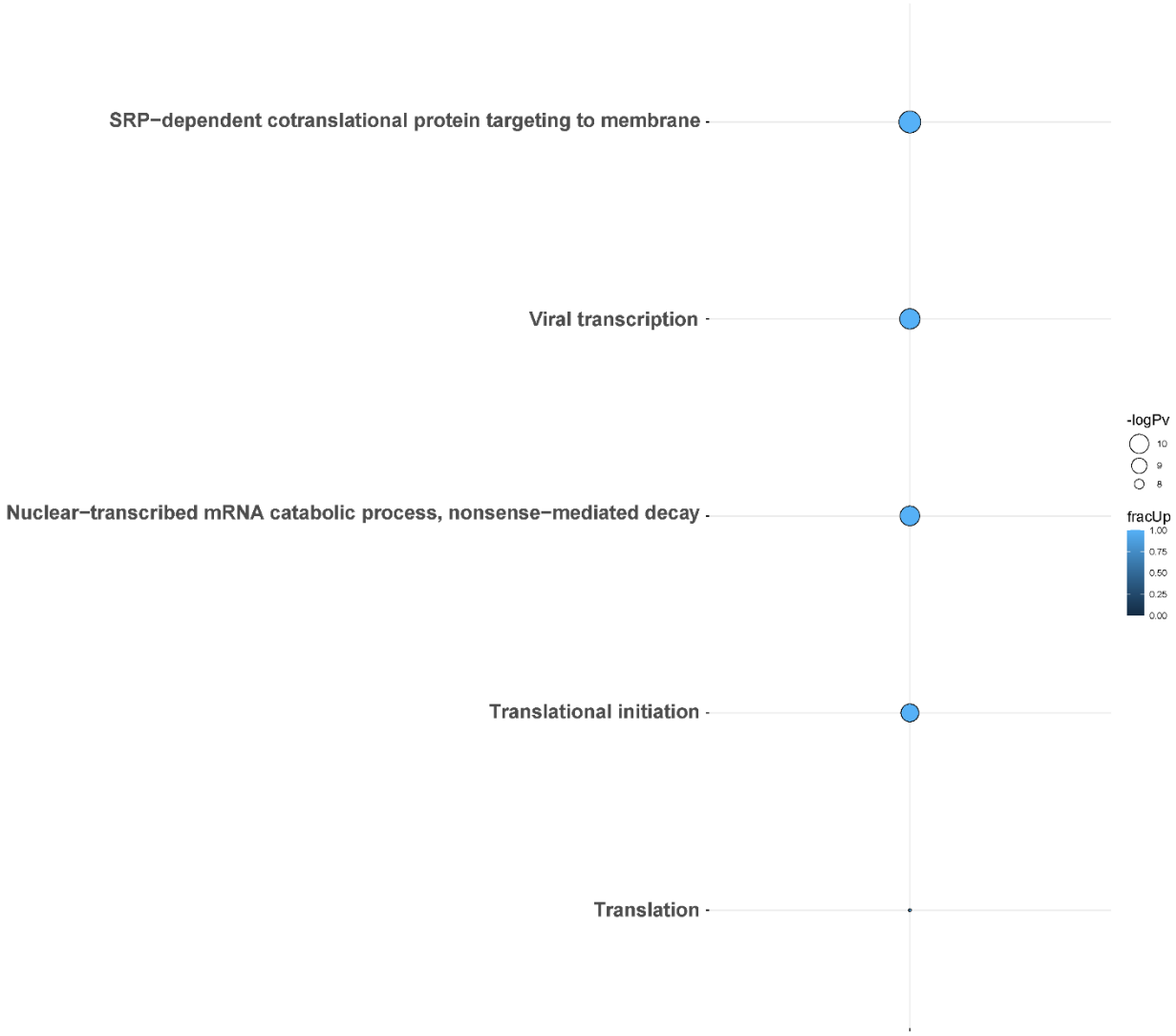
490

491

492 **Transcripts of ribosomal protein genes are elongated during SARS-CoV-2 infection**

493 Specific genes involved in differential polyadenylation were investigated using a linear mixed-model
494 method using *tailfindr* outputs. The most interesting dataset was Calu-3 48 hpi, where twelve genes were
495 found to be significantly increased in poly(A) length (up to ~101 nt in mean poly(A) length) in the infected
496 cells compared with control cells (*UQCRC1, RPL30, RPS12, RPL13, KRT17, CXCL8, RPS6, ZBTB44,*
497 *MIEN1, RPS4X, RPL10 and* a lncRNA-*ENSG00000273149*) (**Figure S5**). Using *multiGO*, GO biological
498 terms of genes with increased poly(A) length were found, which included *viral transcription* (*RPS12,*
499 *RPL30, RPS6, RPL13, RPS4X, RPL10*) (enrichment p-value < 0.0001) (**Figure 4 & Data S3**). KEGG
500 pathways of these genes included the *coronavirus disease* and *ribosome* pathways (**Figure S6 & Data S3**).
501 This suggests that poly(A) tail elongation may be directly linked to SARS-CoV-2 infections, as opposed to
502 a randomly occurring event. A small number of mitochondrial genes were also found to be differentially
503 polyadenylated (including *ENSG00000198888/MT-MD1*) in both Calu-3 and Vero cell lines. Additionally,
504 among the twelve genes which were found to be increased in poly(A) length in *tailfindr* mixed-model
505 analysis, eight genes were also found to be significantly increased in poly(A) length in *nanopolish* analysis
506 after log-transformation and p-value adjustment (*ENSG00000273149, RPS12, RPL30, RPS6, RPL13,*
507 *MIEN1, RPS4X, RPL10,* padj < 0.05). This confirmed the robustness of these results, which increased the
508 confidence of true poly(A) elongation in these eight genes. The other four genes were unable to be detected
509 in *nanopolish* datasets even when padj thresholds were relaxed.

510



512 **Figure 4. GO biological terms from genes with differential poly(A) tail length in *tailfindr* poly(T) mixed-model analyses**
513 **from the Calu-3 48 hpi direct cDNA dataset.** Genes involved in *viral transcription, translation, translational initiation, SRP-*
514 *mediated cotranslational protein targeting to membrane and nuclear-transcribed mRNA catabolic process* and *nonsense-mediated*
515 *decay* pathways were increased in poly(A) tail length after infection. The bubble size and colour indicate the -log10 enrichment p-
516 values and the fraction of genes with increased polyadenylation, respectively. Thresholds of padj < 0.05, enrichment p-value <
517 0.0001 were used. Only bubbles which meet the thresholds are shown. Related to **Figures S5-6 & Data S3**.

518

519

520 **Ribosomal protein genes *RPS4X* and *RPS6* show increased poly(A) tail lengths and downregulated**
521 **expression upon infection**

522

523 We next investigated whether there was any relationship between differential polyadenylation and
524 differential expression results in response to infection. Interestingly, when comparing the GO terms which
525 were shared among the differential polyadenylation and differential expression results of Calu-3 48 hpi
526 direct-cDNA datasets, the results showed an apparent correlation between the two analyses. The enriched
527 GO terms were composed of genes with mainly increased poly(A) tail lengths and decreased expression
528 levels after infection (**Figure 5**). Upon closer inspection, we found that many of the genes involved in these
529 GO terms were associated with the ribosome. Of note, two genes (*RPS4X* and *RPS6*) which contributed to
530 all GO terms, both showed an increase in poly(A) length and downregulated gene expression. This overlap
531 was significant (hypergeometric test, p=0.018). When the KEGG pathways were compared in a similar
532 manner, we observed that the *coronavirus disease* pathway was shared between differential polyadenylation
533 and expression datasets (**Figure S7**). Unlike the GO terms, most genes had an increase in poly(A) tail length
534 but were upregulated in differential expression levels, although many ribosomal genes were downregulated
535 in the same dataset. For example, *CXCL8* had an increased poly(A) length after infection and was
536 upregulated in the expression level results, unlike the ribosomal protein genes described above. This
537 suggested that the correlation between increased poly(A) tails and decreased expression levels were shown
538 specifically in the ribosome-related protein genes and indicated the importance of ribosomal protein genes
539 during SARS-CoV-2 infections.

540

541

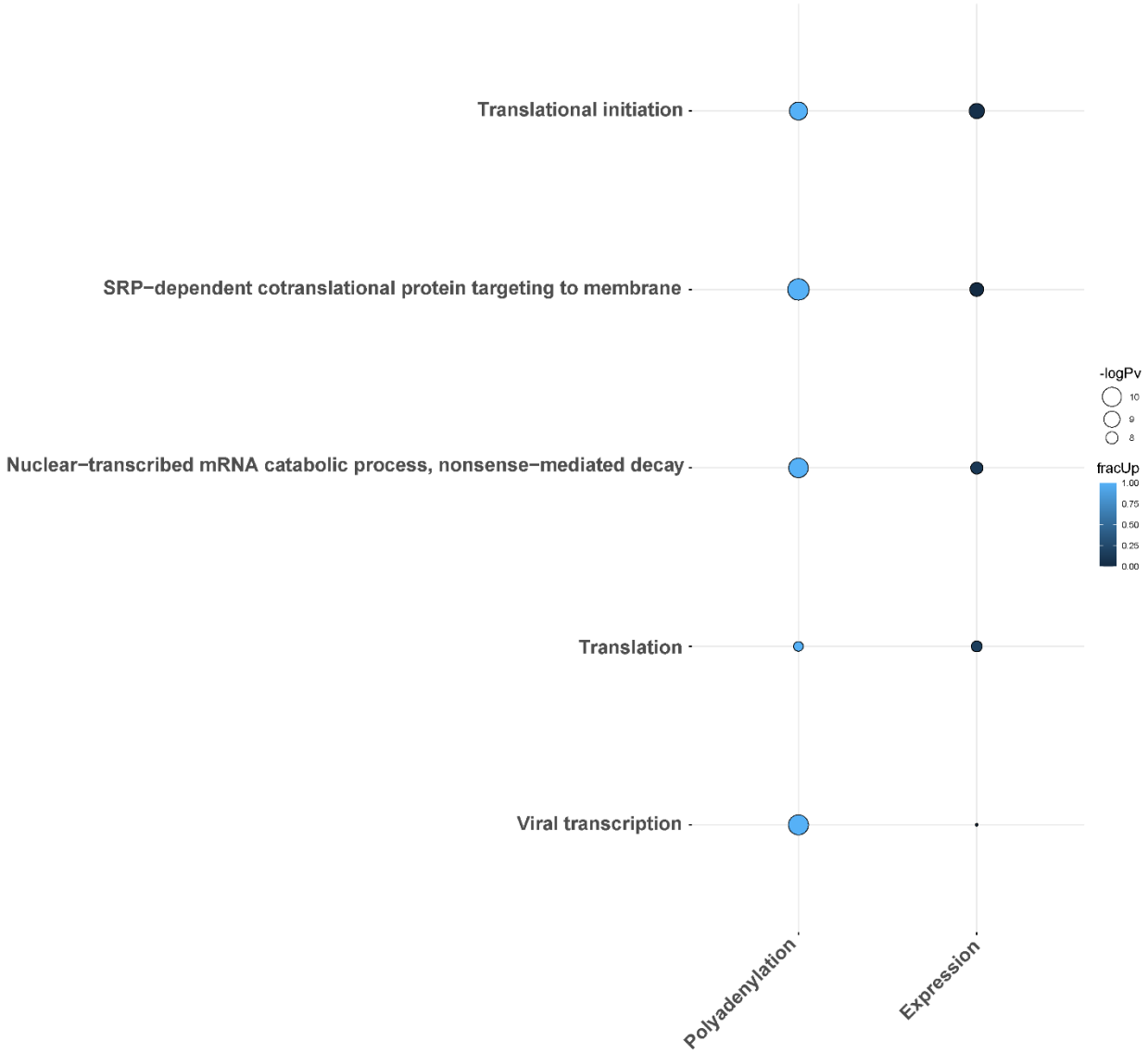


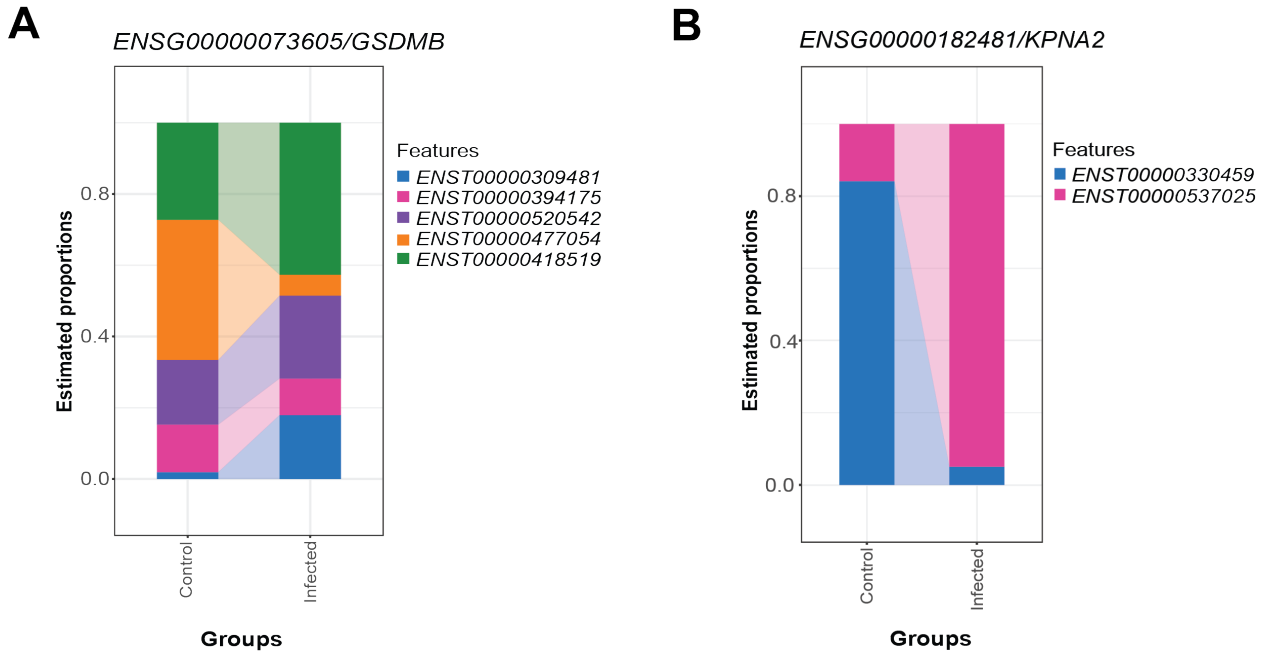
Figure 5. Direction of differentially polyadenylated and expressed genes belonging to common GO biological terms in the two analysis methods using the Calu-3 48 hpi direct cDNA dataset. The plot shows a potential correlation in increased poly(A) tail length and downregulation in gene expression. The bubble size and colour indicate the -log10 enrichment p-values and fraction of upregulated genes/genes with increased polyadenylation, respectively. Thresholds of padj < 0.05, enrichment p-value < 0.0001 were used. Only bubbles which meet the thresholds are shown. Related to **Figures S7 & Data S4**.

552 **Differential transcript usage occurs between control and infected cells during SARS-CoV-2 infection**

553 Differential transcript usage is the differential presence of transcripts between different conditions
554 measured via identifying the proportion of each transcript against the total pool of transcripts and is another
555 valuable feature of ONT RNA-seq. Using *DRIMSeq* and *StageR*, significant differential transcript usage
556 was observed in all three SARS-CoV-2 susceptible cell lines (Calu-3, Caco-2 and Vero) between infected
557 and mock-control cells. These events were observed in three time points (2, 24, 48 hpi) in Caco-2, two time
558 points in Calu-3 (2, 48 hpi) and one time point in Vero cells (24 hpi). This included a processed transcript
559 - *SLC37A4-205* - in the Caco-2 2 hpi dataset and a retained intron transcript - *GSDMB-208* – in the Calu-3
560 48 hpi dataset (**Figure 6 & Table 5**). These results suggested that non-protein-coding transcripts could also
561 show differential usage as with protein-coding transcripts. Furthermore, these results revealed that
562 differential transcript usage events were not specific to a given time point and may have also occurred in a
563 cell-specific manner.

564

565



567 **Figure 6. Differential transcript usage in the Calu-3 48 hpi dataset. A)** Differential estimated proportions of transcripts of
568 *ENSG00000073605/GSDMB* between control and infected cells. **B)** Differential estimated proportions of transcripts of
569 *ENSG00000182481/KPNA2* between control and infected cells. Related to **Table 5**.

570

571 **Table 5. List of isoforms with significant differential usage in Caco-2, Calu-3 and Vero.** Related to **Figure 6**.

Cell line	Time point	Isoforms with differential transcript usage
Caco-2	2	*IPO5-201, IPO5-230, SLC37A4-205, SLC37A4-217*
Caco-2	24	*NACA-221, SERF2-201*
Caco-2	48	*PKM-204, RPL4-201, RPL4-215*
Calu-3	2	*ATP13A3-201, STIL-201, STIL-202, TRA2A-202, TRA2A-211*
Calu-3	48	*SPTBN1-207, GSDMB-208, KPNA2-201, KPNA2-202, RAE1-202, RAE1-203, ATIC-205, ATIC-207, ADK-202, ADK-206, ATRAID-201, ATRAID-207, ANKRD12-201, ANKRD12-203, WARS1-202, WARS1-204*
Vero	24	*ARPC1B-202, ARPC1B-201*

572

573

574

575

576

577

578

579

580

581

582

583

584

585 **Discussion**

586 Long-read sequencing enabled the detection of differential polyadenylation, transcript usage and gene
587 expression level changes within *in vitro* SARS-CoV-2 infection models. Firstly, median poly(A) tail lengths
588 between control and infected cells in direct RNA data were estimated using *nanopolish*. This showed that
589 the non-mitochondrial median poly(A) tail lengths were significantly increased in all three cell lines (Caco-
590 2, Calu-3 and Vero) at the 24 and 48 hpi (**Table 4**). These results suggested that infection with SARS-CoV-
591 2 may cause an increase in the poly(A) lengths of non-mitochondrial transcripts. We explored this further
592 using *tailfindr*. The results from the mixed-effects model analysis showed poly(A) tail elongation after
593 infection in Calu-3 and Vero cells at the 48 hpi. In Calu-3 cells, six genes were involved in viral
594 transcription (*RPS12, RPL30, RPS6, RPL13, RPS4X, RPL10*) (**Figure 6**). This indicated that
595 polyadenylation may play a role in aiding the virus to generate viral mRNA for further protein production
596 or to replicate during infection with SARS-CoV-2. This group of genes is involved in the formation of the
597 ribosome, which is required for protein synthesis. The result is relevant as the SARS-CoV-2 non-structural
598 protein Nsp1 binds to the 40S subunit of the ribosome and inhibits translation initiation of cellular mRNA
599 (Lapointe et al., 2021; Schubert et al., 2020). Ribosomal proteins have also been known to be associated
600 with viral transcription/replication as the host ribosomal machinery needs to be utilised to produce viral
601 proteins for these processes (Stern-Ginossar, Thompson, Mathews, & Mohr, 2019). Therefore, further
602 investigation is warranted to explore the link between increased polyadenylation in host cells after infection
603 with SARS-CoV-2. It would also be of interest to study whether host defense ability decreases with elevated
604 polyadenylation of transcripts related to viral transcription and the ribosome. A lncRNA and a small number
605 of mitochondrial transcripts were also observed with an elongated poly(A) length in infected cells compared
606 with control cells. This is of interest as it suggests that not only the protein-coding genes may be able to
607 play a role in host responses to SARS-CoV-2. Furthermore, as elongated poly(A) tails were observed at a
608 late time point (48 hpi) in Vero cells, it suggested that differential polyadenylation may be more likely to
609 occur at later stages of infection compared with early stages.

610

611 We also explored the observation that many of the genes involved in the commonly enriched GO terms and
612 KEGG pathways were increased in poly(A) tail length and decreased in gene expression (**Figures 5 & S7**).
613 Interestingly, the majority of genes which were involved in both of these observations were ribosomal
614 proteins, such as *RPS4X* and *RPS6* which encode for proteins in the 40S ribosomal subunit. In contrast, a
615 non-ribosomal gene *CXCL8* had increases in both poly(A) length and expression level after infection, which
616 suggested that this correlation belonged exclusively to the ribosomal protein genes. This was an interesting
617 observation as decreased expression of ribosomal proteins in response to SARS-CoV-2 infections have

618 been observed previously (Lieberman et al., 2020), due to the effect of global suppression of ribosomal
619 activity initiated by the virus. However, the increase in poly(A) lengths in the same transcripts was
620 unexpected, as elongation of poly(A) tails are indicative of increase in stability, in contrast to the decrease
621 in expression levels. Why these observations were uniquely presented in these ribosomal protein genes is
622 currently unclear. However, we speculate this may be due to the competition between viral-driven
623 expression downregulation and host-driven post-transcriptional regulation for increased stability of mRNA.
624 In some cases, aberrant polyadenylation has been linked to aid the destruction of eukaryotic mRNA (Y. J.
625 Lee & Glaunsinger, 2009), which may provide an alternative explanation for this phenomenon.

626

627 Supporting the importance of the ribosome during SARS-CoV-2 infection, the *translation* GO term was
628 enriched in infected Calu-3 and Vero cells (**Figure 3**). Among the genes involved in translation, the *EIF1*
629 gene encodes for the eukaryotic translation initiation factor 1, which partakes in translation initiation in
630 eukaryotes by forming a part of the 43S preinitiation complex along with the 40S ribosomal subunit. As
631 mentioned earlier, according to Lapointe et al., (2021), this factor may enhance the binding of SARS-CoV-
632 2 Nsp1 protein to the host 40S subunit, perhaps via changing the conformation of the mRNA entry channel.
633 This facilitates host translation inhibition by competing with host mRNA, which has been shown to also
634 decrease translation of viral mRNA. However, another study showed that the binding of Nsp1 to the 40S
635 subunit induced preferential translation of viral mRNA over host mRNA (Schubert et al., 2020). These
636 findings may explain the downregulation the *EIF1* as a host response to viral infection.

637

638 Differential transcript usage has not yet been extensively studied with SARS-CoV-2 infections but has been
639 useful for studying other illnesses like Parkinson's disease (Dick et al., 2020). Differential transcript usage
640 was observed in all three SARS-CoV-2 susceptible cell lines studied – Caco-2, Calu-3 and Vero, where
641 protein-coding, processed and retained-intron transcripts were involved. Calu-3 48 hpi data showed the
642 greatest number of genes that had undergone differential transcript usage, where a retained-intron transcript
643 *GSDMB-208* showed differential usage in infected cells compared with control cells (**Figure 6**). SARS-
644 CoV-2 induces pyroptosis in human monocytes (Ferreira et al., 2021), and the Gasdermin family of proteins
645 has been implicated in cell death where the granzyme-mediated cleavage of GSDMB can activate
646 pyroptosis (Z. Zhou et al., 2020). Furthermore, *KPNA2* transcripts also showed differential usage in Calu-
647 3 cells at 48 hpi. *KPNA2* is an importin that is bound by ORF6 of the virus to block nuclear IRF3 and ISGF3
648 to antagonise IFN-1 production and signalling (Xia et al., 2020). This suggested that transcripts with
649 differential usage may be involved in important pathways contributing to host responses towards viral

650 infection or the evasion of these responses by the virus. Hence, the specific activity of each transcript as
651 opposed to the activity at the gene level should be further investigated.

652

653 Overall, our long-read sequencing datasets agreed with differential expression studies in the literature. In
654 agreement with previous studies, our differential expression analysis using direct cDNA datasets showed
655 varied host gene expression activity upon infection in different cell types (**Figure 1**). Although Vero cells
656 are imperfect *in vitro* models for SARS-CoV-2 infection, we note that the *ROS metabolic pathway* was
657 enriched in our direct cDNA data across the SARS-CoV-2 susceptible cell lines (**Figures 3 & S2**). The
658 mitochondria can be linked to the ROS metabolic pathway as it produces ROS which can induce increased
659 oxidative stress in cells, potentially leading to cell death (Orrenius, 2007). These results increased support
660 for the idea that mitochondrial processes are important during these infections (Codo et al., 2020; Singh et
661 al., 2021).

662

663 We also observed the downregulation of pathways involved in neurological pathologies such as *Parkinson's*
664 *disease* (Martinelli et al., 2021) (**Figure S2**). The fact that some enriched pathways were involved in non-
665 respiratory, neurological illnesses suggests potential modes of action for SARS-CoV-2 co-morbidities.
666 There are now increasing numbers of studies reporting on the relevance between COVID-19 and other
667 diseases. This includes clinical data where patients with COVID-19 can develop neurological problems
668 which are not only non-specific (e.g. headaches), but also varied, including such maladies as viral
669 meningitis, encephalitis (Moriguchi et al., 2020), olfactory and gustatory dysfunction (Luers et al., 2020),
670 and dementia-related symptoms similar to Alzheimer's disease (Y. Zhou et al., 2021). However, the
671 relevance of this pathway in non-neuronal cells is potentially limited.

672

673

674

675

676

677

678

679 **Limitations & future directions**

680 In our study, we utilised *in vitro* models of SARS-CoV-2 infections using continuous cell lines with a low
681 MOI of 0.1, which may hinder the biological relevance of these results. However, our results confirmed
682 that the use of ONT RNA-seq methods enabled the detection of full-length isoforms, differential
683 polyadenylation and transcript usage. This provides evidence to pursue further investigations with more
684 sophisticated models such as air-liquid-interface cultured organoids from healthy human nasal swabs or *in*
685 *vivo* models such as ferrets. The ideal MOI should also be found via optimisation studies.

686

687 Moreover, our results showed that the *nanopolish* and *tailfindr* methods had significant weak positive
688 correlations ($r < 0.3$, p-value < 0.05) in the median poly(A) lengths from *nanopolish* and median poly(T)
689 lengths from *tailfindr* (**Figure S4**). However, the median poly(A) lengths of *tailfindr* showed non-
690 significant correlations with *nanopolish* poly(A) data. The increased significance of poly(T) transcripts may
691 have occurred because more data was available from the poly(T) dataset. As our results supported similar
692 findings from Krause et al. (2019), we speculate that these discrepancies between poly(A) and poly(T)
693 datasets using ONT direct cDNA sequencing may arise in future studies. We acknowledge that our data is
694 preliminary and the correlation between *nanopolish* and *tailfindr* data should be tested via direct RNA
695 datasets with replicates to validate these findings.

696

697 Functional work should also be carried out to further validate the results of this study. For differential
698 expression analysis, knock-down experiments within the same cell lines using CRISPR technology may be
699 utilised to evaluate the effects of differentially expressed genes identified in this study. Furthermore,
700 functional work for polyadenylation may be approached by utilising cell lines with plasmids containing
701 gene sequences of interest followed by a poly(A) sequence of varying lengths. Assays such as measuring
702 viral titre and further RNA-seq may be used to test the effects of these alterations after infection with SARS-
703 CoV-2.

704

705 **Conclusions**

706 Overall, by utilising three ONT RNA-seq methodologies, we generated an in-depth characterisation of
707 differential expression, polyadenylation and differential transcript usage of cell lines infected *in vitro* by
708 SARS-CoV-2. Unravelling the pathways associated with duration of infection and responses of differing
709 cell types using long-read methods will provide novel insights into the pathogenesis of SARS-CoV-2.

Funding

This research was supported by a University of Melbourne "Driving research momentum" award (to LC) and NHMRC EU project grant (GNT1195743 to LC). KS was supported by an NHMRC Investigator grant. The Melbourne WHO Collaborating Centre for Reference and Research on Influenza was supported by the Australian Government Department of Health. MC was supported by an Australian National Health and Medical Research Council Investigator Fellowship (APP1196841). JC was supported by the Miller Foundation and the Australian Government Research Training Programme (RTP) scholarship.

List of Figures

Figure 1. Volcano plots show the difference in expression level between control and infected cells per cell line (A549, Caco-2, Calu-3 and Vero) in direct cDNA datasets using *DESeq2*. X-axis represents log2FC and Y-axis displays –log10 p-adjusted value (padj). • padj < 0.05 (blue), • |log2FC |> 1 (orange), • both (green). Increased number of differentially expressed genes are shown in latter time points in Caco-2, Calu-3 and Vero cell lines, and a lack of differential expression was observed in A549 cells. Related to **Figure S1 & Table 2**.

Figure 2. Heatmap of relative gene expression in Caco-2 and Calu-3 cells at 24 and 48 hpi using direct cDNA datasets reveal distinct gene expression profiles in each cell line, analysed by *DESeq2* and visualized by *multiGO*. The data was filtered by padj < 0.05, enrichment p-value < 0.0001. The expression levels were scaled per row and organised based on relevant GO terms. In Calu-3 cells, higher relative expression of interferon-response genes was observed compared with Caco-2 cells. In contrast, Caco-2 cells showed higher relative expression of ribosomal protein and mitochondrial genes. Related to **Figure 1 & S1 & Table 1**.

Figure 3. GO biological terms of differentially expressed genes in Calu-3, Caco-2 and Vero direct cDNA datasets analysed by *DESeq2* and visualised with *multiGO*. Results include datasets (in order); Caco-2 24 vs 48, Caco-2 48, Calu-3 2 vs 24, Calu-3 24 vs 48, Calu-3 24, Calu-3 48, Vero 2 vs 24, Vero 24 vs 48, Vero 24 and Vero 48 hpi. Commonly enriched GO terms across the three cell lines involved the *ROS metabolic process*. The bubble size and colour indicate the -log10 enrichment p-value and fraction of upregulated genes, respectively. Thresholds of padj < 0.05 and enrichment p-value < 1E-6 in at least one dataset were used for generating the plot, where insignificant bubbles are also shown on the plot. All terms with padj < 0.05, enrichment p-value < 0.05 involving at least two genes were deemed as significant for the analysis. Related to **Figure S2 & Data S2**.

Figure 4. GO biological terms from genes with differential poly(A) tail length in *tailfindr* poly(T) mixed-model analyses from the Calu-3 48 hpi direct cDNA dataset. Genes involved in *viral transcription, translation, translational initiation, SRP-mediated cotranslational protein targeting to membrane and nuclear-transcribed mRNA catabolic process* and *nonsense-mediated decay* pathways were increased in poly(A) tail length after infection. The bubble size and colour indicate the -log10 enrichment p-values and the fraction of genes with increased polyadenylation, respectively. Thresholds of padj < 0.05, enrichment p-value < 0.0001 were used. Only bubbles which meet the thresholds are shown. Related to **Figures S5-6 & Data S3**.

Figure 5. Direction of differentially polyadenylated and expressed genes belonging to common GO biological terms in the two analysis methods using the Calu-3 48 hpi direct cDNA dataset. The plot shows a potential correlation in increased poly(A) tail length and downregulation in gene expression. The bubble size and colour indicate the -log10 enrichment p-values and fraction of upregulated genes/genes with increased polyadenylation, respectively. Thresholds of padj < 0.05, enrichment p-value < 0.0001 were used. Only bubbles which meet the thresholds are shown. Related to **Figures S7 & Data S4**.

Figure 6. Differential transcript usage in the Calu-3 48 hpi dataset. A) Differential estimated proportions of transcripts of *ENSG00000073605/GSDMB* between control and infected cells. **B)** Differential estimated proportions of transcripts of *ENSG00000182481/KPNA2* between control and infected cells. Related to **Table 5**.

Supplementary items

Data S1. Relative percentages of mapped reads to host, virus or sequin genomes in all datasets

Data S2. *multiGO* outputs of GO biological and KEGG analyses from the direct cDNA differential expression results, including –log10 p-values from the enrichment analyses and genes involved in the GO terms/KEGG pathways

Data S3. *multiGO* outputs of GO biological and KEGG analyses from the Calu-3 48 hpi direct cDNA differential polyadenylation results, including –log10 p-values from the enrichment analyses and genes involved in the GO terms/KEGG pathways

Data S4. *multiGO* outputs of GO biological and KEGG analyses from the Calu-3 48 hpi direct cDNA differential polyadenylation vs expression results, including –log10 p-values from the enrichment analyses and genes involved in the GO terms/KEGG pathways

Table S1. Correlations between *tailfindr* median poly(A) and poly(T) lengths per gene using direct cDNA data. Weak positive correlations were observed between poly(A) and poly(T) datasets (r < 0.4), with all correlations being significant (p-value < 0.05, Pearson's correlation test). Related to **Figure S3 & Table 4**.

Table S2. Comparison of significantly differentially polyadenylated non-mitochondrial gene clusters in the Calu-3 48 hpi datasets between *nanopolish*, *tailfindr* poly(A) and poly(T) outputs (before log-transformation) using Wilcoxon's test of ranks. *nanopolish* and *tailfindr* poly(T) results were more comparable compared with *tailfindr* poly(A) results, as no significant polyadenylation was observed in *tailfindr* poly(A) data. Related to **Figure S3-4 & Table 4**.

Figure S1. Volcano plots showing changes in differential expression between control and infected cells across two time points in direct cDNA datasets using *DESeq2* and the interactive term. X-axis represents log2FC and Y-axis displays –log10 padj. • padj < 0.05, • |log2(FC)| > 1, • both. Plots reveal increase in differential expression at latter time points compared with earlier time points in Caco-2, Calu-3 and Vero cell lines, and a lack of significant differential expression was observed between A549 0 vs 24 hpi. Related to **Figure 1**.

Figure S2. KEGG pathways from differentially expressed genes in Calu-3, Caco-2 and Vero direct cDNA datasets analysed by *DESeq2*. Results include datasets (in order); Caco-2 24 vs 48, Caco-2 48, Calu-3 2 vs 24, Calu-3 24 vs 48, Calu-3 24, Calu-3 48, Vero 2 vs 24, Vero 24 vs 48, Vero 24 and Vero 48 hpi. Strongest KEGG pathway enrichment was observed in the Calu-3 cells including pathways such as *influenza A*, *measles* and *coronavirus disease*. The bubble size and colour indicate the -log10 enrichment p-values and fraction of upregulated genes, respectively. Thresholds of padj < 0.05 and enrichment p-value < 0.0001 in at least one dataset were used for generating the plot, and all terms with padj < 0.05 and enrichment p-value < 0.05 were deemed as significant for the analysis. Related to **Figure 3 & Data S2**.

Figure S3. Correlations between median poly(A/T) lengths from *tailfindr* analyses. Scatter plots comparing *tailfindr* poly(A) vs poly(T) datasets from Caco-2 **A)** 2 hpi, **B)** 24 hpi, **C)** 48 hpi, Calu-3 **D)** 2 hpi, **E)** 24 hpi, **F)** 48 hpi, Vero **G)** 2 hpi, **H)** 24 hpi**, I)** 48 hpi. Weak correlation between median poly(A) and poly(T) lengths were observed (R < 0.4, exact values along with p-values are indicated in **Table S1**), where each dot represents a gene. Related to **Tables 4 & S1-2**.

Figure S4. Correlations between median poly(A) lengths from *nanopolish* and poly(A/T) lengths from *tailfindr* analyses. Scatter plots comparing **A)** median *nanopolish* poly(A) and *tailfindr* poly(A) lengths in control cells **B)** median *nanopolish* poly(A) and *tailfindr* poly(T) lengths in control cells, **C)** median *nanopolish* poly(A) and *tailfindr* poly(A) lengths in infected cells and **D)** median *nanopolish* poly(A) and *tailfindr* poly(T) lengths in infected cells. Weak significant positive correlations were observed between the *nanopolish* poly(A) and *tailfindr* poly(T) datasets (r = 0.12-0.27, p-value < 0.05, Spearman's correlation test). No significant correlation was observed between *nanopolish* poly(A) and *tailfindr* poly(A) datasets (p-value > 0.05, Spearman's correlation test). Related to **Tables 4 & S2**.

Figure S5. Raincloud plots of raw, untransformed poly(T) tail lengths in twelve genes with increased poly(A) tail length in the *tailfindr* poly(T) mixed-model analysis with log-transformation. Genes include **A)** *UQCRC1*, **B)** *RPL30*, **C)** *RPS12*, **D)** *RPL13*, **E)** *KRT17*, **F)** *CXCL8*, **G)** *RPS6*, **H)** *ZBTB44*, **I)** *MIEN1*, **J)** *RPS4X*, **K)** *RPL10* and **L)** a lncRNA. Related to **Figure 4**.

Figure S6. KEGG pathways from genes with differential poly(A) tail length in *tailfindr* poly(T) mixed-model analyses from the Calu-3 48 hpi dataset. Genes involved in *coronavirus disease* and *ribosome* pathways were increased in poly(A) length after infection. The bubble size and colour indicate the -log10 enrichment p-values and the fraction of genes with

823 increased polyadenylation, respectively. Thresholds of padj < 0.05, enrichment p-value < 0.0001 were used. Only bubbles which
824 meet the thresholds are shown. Related to **Figure 4 & Data S3**.

825 **Figure S7. KEGG pathways of differentially polyadenylated and expressed genes from the Calu-3 48 hpi direct cDNA**
826 **dataset.** Only the *coronavirus disease* pathway was shared between the two analyses. The plot shows a potential correlation in
827 increased poly(A) tail length and downregulation in gene expression. The bubble size and colour indicate the -log10 enrichment p-
828 values and fraction of upregulated genes/genes with increased polyadenylation, respectively. Thresholds of padj < 0.05, enrichment
829 p-value < 0.0001 were used. Related to **Figure 5 & Data S4**.

830

831

832

833

834

835

836

837

838

839

840

841

842

843

844

845

846

847

848

849

850

References

Allen, M., Poggiali, D., Whitaker, K., Marshall, T. R., Van Langen, J., & Kievit, R. A. (2021). Raincloud plots: a multi-platform tool for robust data visualization. *Wellcome Open Research, 4*, 63. doi:10.12688/wellcomeopenres.15191.2

Beckel-Mitchener, A. C. (2002). Poly(A) Tail Length-dependent Stabilization of GAP-43 mRNA by the RNA-binding Protein HuD. *277*(31), 27996-28002. doi:10.1074/jbc.m201982200

Bibert, S., Guex, N., Lourenco, J., Brahier, T., Papadimitriou-Olivgeris, M., Damonti, L., . . . Reg, C. S. G. (2021). Transcriptomic Signature Differences Between SARS-CoV-2 and Influenza Virus Infected Patients. *Front Immunol, 12*, 666163. doi:10.3389/fimmu.2021.666163

Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W.-C., Uhl, S., Hoagland, D., Møller, R., . . . Tenoever, B. R. (2020). Imbalanced Host Response to SARS-CoV-2 Drives Development of COVID-19. *Cell, 181*(5), 1036-1045.e1039. doi:10.1016/j.cell.2020.04.026

Chang, J. J., Rawlinson, D., Pitt, M. E., Taiaroa, G., Gleeson, J., Zhou, C., . . . Coin, L. J. M. (2021). Transcriptional and epi-transcriptional dynamics of SARS-CoV-2 during cellular infection. *Cell Rep, 35*(6), 109108. doi:10.1016/j.celrep.2021.109108

Chen, F., Zhang, Y., Sucgang, R., Ramani, S., Corry, D., Kheradmand, F., & Creighton, C. J. (2021). Meta-analysis of host transcriptional responses to SARS-CoV-2 infection reveals their manifestation in human tumors. *Scientific Reports, 11*(1). doi:10.1038/s41598-021-82221-4

Codo, A. C., Davanzo, G. G., Monteiro, L. D. B., De Souza, G. F., Muraro, S. P., Virgilio-Da-Silva, J. V., . . . Moraes-Vieira, P. M. (2020). Elevated Glucose Levels Favor SARS-CoV-2 Infection and Monocyte Response through a HIF-1α/Glycolysis-Dependent Axis. *Cell Metabolism, 32*(3), 437-446.e435. doi:10.1016/j.cmet.2020.07.007

Curinha, A., Oliveira Braz, S., Pereira-Castro, I., Cruz, A., & Moreira, A. (2014). Implications of polyadenylation in health and disease. *Nucleus, 5*(6), 508-519. doi:10.4161/nucl.36360

de Jong, L. C., Cree, S., Lattimore, V., Wiggins, G. A. R., Spurdle, A. B., kConFab, I., . . . Walker, L. C. (2017). Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events. *Breast Cancer Res, 19*(1), 127. doi:10.1186/s13058-017-0919-1

De Paoli-Iseppi, R., Gleeson, J., & Clark, M. B. (2021). Isoform Age - Splice Isoform Profiling Using Long-Read Technologies. *Front Mol Biosci, 8*, 711733. doi:10.3389/fmolb.2021.711733

Dick, F., Nido, G. S., Alves, G. W., Tysnes, O.-B., Nilsen, G. H., Dölle, C., & Tzoulis, C. (2020). Differential transcript usage in the Parkinson's disease brain. *PLoS Genetics, 16*(11), e1009182. doi:10.1371/journal.pgen.1009182

Emeny, J. M., & Morgan, M. J. (1979). Regulation of the Interferon System: Evidence that Vero Cells have a Genetic Defect in Interferon Production. *Journal of General Virology, 43*(1), 247-252. doi:10.1099/0022-1317-43-1-247

Ferreira, A. C., Soares, V. C., De Azevedo-Quintanilha, I. G., Dias, S. D. S. G., Fintelman-Rodrigues, N., Sacramento, C. Q., . . . Souza, T. M. L. (2021). SARS-CoV-2 engages inflammasome and pyroptosis in human primary monocytes. *Cell Death Discovery, 7*(1). doi:10.1038/s41420-021-00428-w

Fuke, H., & Ohno, M. (2007). Role of poly (A) tail as an identity element for mRNA nuclear export. *36*(3), 1037-1049. doi:10.1093/nar/gkm1120

Gebauer, F., Preiss, T., & Hentze, M. W. (2012). From Cis-Regulatory Elements to Complex RNPs and Back. *Cold Spring Harbor Perspectives in Biology, 4*(7), a012245-a012245. doi:10.1101/cshperspect.a012245

Gleeson, J., Leger, A., Prawer, Y. D. J., Lane, T. A., Harrison, P. J., Haerty, W., & Clark, M. B. (2021). Accurate expression quantification from nanopore direct RNA sequencing with NanoCount. *Nucleic Acids Research*. doi:10.1093/nar/gkab1129

Harcourt, J., Tamin, A., Lu, X., Kamili, S., Sakthivel, S. K., Murray, J., . . . Thornburg, N. J. (2020). Isolation and characterization of SARS-CoV-2 from the first US COVID-19 patient. 2020.2003.2002.972935. doi:10.1101/2020.03.02.972935 %J bioRxiv

Hardwick, S. A., Chen, W. Y., Wong, T., Deveson, I. W., Blackburn, J., Andersen, S. B., . . . Mercer, T. R. (2016). Spliced synthetic genes as internal controls in RNA sequencing experiments. *Nature Methods, 13*(9), 792-798. doi:10.1038/nmeth.3958

Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., . . . Pöhlmann, S. (2020). SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell, 181*(2), 271-280.e278. doi:10.1016/j.cell.2020.02.052

Islam, A. B. M. M. K., Khan, M. A.-A.-K., Ahmed, R., Hossain, M. S., Kabir, S. M. T., Islam, M. S., & Siddiki, A. M. A. M. Z. (2021). Transcriptome of nasopharyngeal samples from COVID-19 patients and a comparative analysis with other SARS-CoV-2 infection models reveal disparate host responses against SARS-CoV-2. *Journal of Translational Medicine, 19*(1). doi:10.1186/s12967-020-02695-0

Krause, M., Niazi, A. M., Labun, K., Torres Cleuren, Y. N., Muller, F. S., & Valen, E. (2019). tailfindr: alignment-free poly(A) length measurement for Oxford Nanopore RNA and DNA sequencing. *RNA, 25*(10), 1229-1241. doi:10.1261/rna.071332.119

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software, 82*(13), 1–26. doi:10.18637/jss.v082.i13

Lapointe, C. P., Grosely, R., Johnson, A. G., Wang, J., Fernández, I. S., & Puglisi, J. D. (2021). Dynamic competition between SARS-CoV-2 NSP1 and mRNA on the human ribosome inhibits translation initiation. *Proceedings of the National Academy of Sciences, 118*(6), e2017715118. doi:10.1073/pnas.2017715118

Lee, R. C., Feinbaum, R. L., & Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell, 75*(5), 843-854. doi:10.1016/0092-8674(93)90529-y

Lee, Y. J., & Glaunsinger, B. A. (2009). Aberrant Herpesvirus-Induced Polyadenylation Correlates With Cellular Messenger RNA Destruction. *PLoS Biology, 7*(5), e1000107. doi:10.1371/journal.pbio.1000107

Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics, 34*(18), 3094-3100. doi:10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., . . . Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics, 25*(16), 2078-2079. doi:10.1093/bioinformatics/btp352

Li, J., Liu, Y., & Zhang, X. (2010). Murine coronavirus induces type I interferon in oligodendrocytes through recognition by RIG-I and MDA5. *J Virol, 84*(13), 6472-6482. doi:10.1128/JVI.00016-10

Li, M.-Y., Li, L., Zhang, Y., & Wang, X.-S. (2020). Expression of the SARS-CoV-2 cell receptor gene ACE2 in a wide variety of human tissues. *Infectious Diseases of Poverty, 9*(1). doi:10.1186/s40249-020-00662-x

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics, 30*(7), 923-930. doi:10.1093/bioinformatics/btt656

Lieberman, N. A. P., Peddu, V., Xie, H., Shrestha, L., Huang, M.-L., Mears, M. C., . . . Greninger, A. L. (2020). In vivo antiviral host transcriptional response to SARS-CoV-2 by viral load, sex, and age. *PLoS Biology, 18*(9), e3000849. doi:10.1371/journal.pbio.3000849

Luers, J. C., Rokohl, A. C., Loreck, N., Wawer Matos, P. A., Augustin, M., Dewald, F., . . . Heindl, L. M. (2020). Olfactory and Gustatory Dysfunction in Coronavirus Disease 2019 (COVID-19). *Clinical Infectious Diseases, 71*(16), 2262-2264. doi:10.1093/cid/ciaa525

Martinelli, A., Akhmedov, M., & Kwee, I. (2021). Meta-analysis of MERS, SARS and COVID-19 in vitro Infection Datasets Reveals Common Patterns in Gene and Protein Expression. 2020.2009.2025.313510. doi:10.1101/2020.09.25.313510 %J bioRxiv

Meade, B., Lafayette, L., Sauter, G., & Tosello, D. (2017). Spartan HPC-Cloud Hybrid: Delivering Performance and Flexibility: University of Melbourne.

Mick, E., Kamm, J., Pisco, A. O., Ratnasiri, K., Babik, J. M., Castañeda, G., . . . Langelier, C. (2020). Upper airway gene expression reveals suppressed immune responses to SARS-CoV-2 compared with other respiratory viruses. *Nature Communications, 11*(1). doi:10.1038/s41467-020-19587-y

Moriguchi, T., Harii, N., Goto, J., Harada, D., Sugawara, H., Takamino, J., . . . Shimada, S. (2020). A first case of meningitis/encephalitis associated with SARS-Coronavirus-2. *International Journal of Infectious Diseases, 94*, 55-58. doi:10.1016/j.ijid.2020.03.062

Nowicka, M., & Robinson, M. D. (2016). DRIMSeq: a Dirichlet-multinomial framework for multivariate count outcomes in genomics. *F1000Research, 5*, 1356. doi:10.12688/f1000research.8900.2

Orrenius, S. (2007). Reactive Oxygen Species in Mitochondria-Mediated Cell Death. *Drug Metabolism Reviews, 39*(2-3), 443-455. doi:10.1080/03602530701468516

Park, J.-E., Yi, H., Kim, Y., Chang, H., & V. (2016). Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *62*(3), 462-471. doi:10.1016/j.molcel.2016.04.007

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods, 14*(4), 417-419. doi:10.1038/nmeth.4197

Puray-Chavez, M., Lapak, K. M., Schrank, T. P., Elliott, J. L., Bhatt, D. P., Agajanian, M. J., . . . Kutluay, S. B. (2021). Systematic analysis of SARS-CoV-2 infection of an ACE2-negative human airway cell. *Cell Reports, 36*(2), 109364. doi:10.1016/j.celrep.2021.109364

Saccon, E., Chen, X., Mikaeloff, F., Rodriguez, J. E., Szekely, L., Vinhas, B. S., . . . Gupta, S. (2021). Cell-type-resolved quantitative proteomics map of interferon response against SARS-CoV-2. *iScience, 24*(5), 102420. doi:10.1016/j.isci.2021.102420

Sasaki, M., Kishimoto, M., Itakura, Y., Tabata, K., Intaruck, K., Uemura, K., . . . Sawa, H. (2021). Air-liquid interphase culture confers SARS-CoV-2 susceptibility to A549 alveolar epithelial cells. *Biochem Biophys Res Commun, 577*, 146-151. doi:10.1016/j.bbrc.2021.09.015

Schubert, K., Karousis, E. D., Jomaa, A., Scaiola, A., Echeverria, B., Gurzeler, L.-A., . . . Ban, N. (2020). SARS-CoV-2 Nsp1 binds the ribosomal mRNA channel to inhibit translation. *Nature Structural & Molecular Biology, 27*(10), 959-966. doi:10.1038/s41594-020-0511-8

Shah, V. K., Firmal, P., Alam, A., Ganguly, D., & Chattopadhyay, S. (2020). Overview of Immune Response During SARS-CoV-2 Infection: Lessons From the Past. *Front Immunol, 11*, 1949. doi:10.3389/fimmu.2020.01949

Shuai, H., Chu, H., Hou, Y., Yang, D., Wang, Y., Hu, B., . . . Yuen, K.-Y. (2020). Differential immune activation profile of SARS-CoV-2 and SARS-CoV infection in human lung and intestinal cells: Implications for treatment with IFN-β and IFN inducer. *Journal of Infection, 81*(4), e1-e10. doi:10.1016/j.jinf.2020.07.016

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods, 14*(4), 407-410. doi:10.1038/nmeth.4184

Singh, K., Chen, Y.-C., Hassanzadeh, S., Han, K., Judy, J. T., Seifuddin, F., . . . Pirooznia, M. (2021). Network Analysis and Transcriptome Profiling Identify Autophagic and Mitochondrial Dysfunctions in SARS-CoV-2 Infection. *Frontiers in Genetics, 12*. doi:10.3389/fgene.2021.599261

Stern-Ginossar, N., Thompson, S. R., Mathews, M. B., & Mohr, I. (2019). Translational Control in Virus-Infected Cells. *Cold Spring Harbor Perspectives in Biology, 11*(3), a033001. doi:10.1101/cshperspect.a033001

Sun, G., Cui, Q., Garcia, G., Wang, C., Zhang, M., Arumugaswami, V., . . . Shi, Y. (2021). Comparative transcriptomic analysis of SARS-CoV-2 infected cell model systems reveals differential innate immune responses. *Scientific Reports, 11*(1). doi:10.1038/s41598-021-96462-w

Tazi, J., Bakkour, N., & Stamm, S. (2009). Alternative splicing and disease. *Biochim Biophys Acta, 1792*(1), 14-26. doi:10.1016/j.bbadis.2008.09.017

1000 Van Den Berge, K., Soneson, C., Robinson, M. D., & Clement, L. (2017). stageR: a general stage-wise
1001 method for controlling the gene-level false discovery rate in differential expression and
1002 differential transcript usage. *Genome Biology, 18*(1). doi:10.1186/s13059-017-1277-0
1003 Watson, A., Spalluto, C. M., Mccrae, C., Cellura, D., Burke, H., Cunoosamy, D., . . . Wilkinson, T.
1004 (2020). Dynamics of IFN-β Responses during Respiratory Viral Infection. Insights for
1005 Therapeutic Strategies. *American Journal of Respiratory and Critical Care Medicine, 201*(1), 83-
1006 94. doi:10.1164/rccm.201901-0214oc
1007 Wickham, H. (2016). *ggplot2: Elegant Graphics for Data Analysis.* In.
1008 Workman, R. E., Tang, A. D., Tang, P. S., Jain, M., Tyson, J. R., Razaghi, R., . . . Timp, W. (2019).
1009 Nanopore native RNA sequencing of a human poly(A) transcriptome. *Nature Methods, 16*(12),
1010 1297-1305. doi:10.1038/s41592-019-0617-2
1011 Wu, M., Chen, Y., Xia, H., Wang, C., Tan, C. Y., Cai, X., . . . Kasper, D. L. (2020). Transcriptional and
1012 proteomic insights into the host response in fatal COVID-19 cases. *Proceedings of the National*
1013 *Academy of Sciences, 117*(45), 28336-28343. doi:10.1073/pnas.2018030117
1014 Wyler, E., Mösbauer, K., Franke, V., Diag, A., Gottula, L. T., Arsiè, R., . . . Landthaler, M. (2021).
1015 Transcriptomic profiling of SARS-CoV-2 infected human cell lines identifies HSP90 as target for
1016 COVID-19 therapy. *iScience, 24*(3), 102151. doi:10.1016/j.isci.2021.102151
1017 Xia, H., Cao, Z., Xie, X., Zhang, X., Chen, J. Y.-C., Wang, H., . . . Shi, P.-Y. (2020). Evasion of Type I
1018 Interferon by SARS-CoV-2. *Cell Reports, 33*(1), 108234. doi:10.1016/j.celrep.2020.108234
1019 Xie, X., Muruato, A. E., Zhang, X., Lokugamage, K. G., Fontes-Garfias, C. R., Zou, J., . . . Shi, P.-Y.
1020 (2020). A nanoluciferase SARS-CoV-2 for rapid neutralization testing and screening of anti-
1021 infective drugs for COVID-19. *Nature Communications, 11*(1). doi:10.1038/s41467-020-19055-7
1022 Zhou, Y., Xu, J., Hou, Y., Leverenz, J. B., Kallianpur, A., Mehra, R., . . . Cheng, F. (2021). *Network*
1023 *medicine links SARS-CoV-2/COVID-19 infection to brain microvascular injury and*
1024 *neuroinflammation in dementia-like cognitive impairment*. Cold Spring Harbor Laboratory.
1025 Zhou, Z., He, H., Wang, K., Shi, X., Wang, Y., Su, Y., . . . Shao, F. (2020). Granzyme A from cytotoxic
1026 lymphocytes cleaves GSDMB to trigger pyroptosis in target cells. *Science, 368*(6494).
1027 doi:10.1126/science.aaz7548

1028