**Section A**

RNA sequencing is a high-throughput next-generation sequencing technology that provides a snapshot of the transcriptome by analysing the quantity and sequence of RNA molecules. It enables researchers to collect quantitative and qualitative data. RNA-seq involves several key steps: consisting of the extraction of RNA, conversion into cDNA, and subsequent preparation of the cDNA library with adapters. It holds major advantages over other technologies, like its ability to detect novel transcripts, unannotated RNA and differences in gene expression that explore cell-to-cell heterogeneity, elucidating the functional links between gene expression and diseases. A study aimed to assess host transcriptome responses to SARS-CoV-2 infection against other respiratory viruses, wherein the findings resulted in further understanding of SARS-CoV-2 pathogenesis to allow informed therapeutic research to be developed (1).

In order to ensure accurate and reproducible results, an effective RNA-seq experiment must include the necessary controls, replication, and minimised confounding variables. **Uninfected (untreated) controls** establish the baseline, allowing observed transcriptional differences to be ascribed to infection or therapy rather than intrinsic variability. **Mock-treated controls**, which replicate all handling and vehicle exposure without the active agent, disentangle procedural and solvent effects from the experimental factor of interest. Additionally, replicates form the backbone of reliable RNA-seq studies with a major importance to specifically biological replicates. Additionally, replicates form the backbone of reliable RNA-seq studies with a major importance to specifically biological replicates. These are independent samples derived from different individuals or organisms for each experimental group, capturing probable natural variability. The generally accepted minimum number of biological replicates for RNA-seq is three, though four to six replicates are often recommended to increase statistical power, especially when detecting more subtle gene expression differences. Technical replicates, in which multiple sequencing runs or library preparations from the same biological sample, assess the reproducibility of the technical workflow. Furthermore, sample quality plays an important role in pre-processing of RNA, ideally where tools like Bioanalyzer are used to assess the integrity of RNA via a RNA Integrity Number (RIN) (1). Where high RIN values (≥7–8) indicate minimal degradation and are generally selected for downstream library preparation. A combination of these replicates allows robust conclusions to be drawn from the RNA-seq data (3).

Raw reads are subjected to quality control checks post sequencing to test for adaptor contamination, GC content, and base quality. Programs are used to trim low-quality reads and adaptor sequences, and aligners such as HISAT2 are used to align cleaned reads to a reference genome or transcriptome. These aligners map reads to known gene characteristics while considering splice junctions. A count matrix is generated by quantifying the aligned reads into gene-level counts, where each row denotes a gene. Genes with extremely low read counts add noise and lower statistical power, so across all samples, they are filtered out. This carefully curated count data serves as the input for DESeq2's subsequent statistical analysis, a widely used R package for RNA-seq differential expression. It performs normalisation, models counts, and runs statistical tests to find genes that change between groups.

**Section B**

1. Creating the dataset

# Create a reduced dataset containing only RSV and SARS-CoV-2 infections

dat3 <- as.matrix(dat[,c(14:17, 22:27)])

rownames(dat3) <- dat$Gene

# Define sample conditions for metadata

sample\_table <- data.frame(sample\_id=colnames(dat3),

condition=c('Mock\_RSV','Mock\_RSV','RSV','RSV',

'Mock\_Covid','Mock\_Covid','Mock\_Covid',

'Covid','Covid','Covid'))

# Construct a DESeq2 dataset from the count matrix and metadata

dds3 <- DESeqDataSetFromMatrix(dat3, colData=sample\_table, design=~condition)

dds3 <- DESeq(dds3)

Only RSV and SARS-CoV-2 infections with corresponding controls were included in the dataset. A DESeq2 dataset object (dds3) was produced and processed to allow for downstream differential expression analysis, and a metadata table was made to define sample conditions.

2. Volcano plots

# Results for RSV vs mock control

rsv\_results <- results(dds3, contrast=c('condition','RSV','Mock\_RSV'))

# Results for SARS-CoV-2 vs mock control

covid\_results <- results(dds3, contrast=c('condition','Covid','Mock\_Covid'))

# Define significance thresholds

cols <- ifelse(rsv\_results$padj <= 0.05 & abs(rsv\_results$log2FoldChange) >= 1, "red", "black")

# Plot volcano plot for RSV

plot(rsv\_results$log2FoldChange, -log10(rsv\_results$padj),

col=cols, pch=16,

xlab="Log2 Fold Change", ylab="-Log10 adjusted p-value",

main="RSV vs Mock Volcano Plot")

abline(h=-log10(0.05), col="blue", lty=2)

abline(v=c(-1,1), col="blue", lty=2)

Differential expression results were obtained by comparing infected samples against mock controls. A volcano plot was generated for RSV, where genes with significant adjusted p-values and large fold-changes are highlighted. Threshold lines indicate the significance cut-off (horizontal) and fold-change cut-offs (vertical), allowing easy identification of strongly regulated genes. A similar plot was produced for SARS-CoV-2.

3. Interferon Heatmap

# Transform count data to stabilise variance

rld2 <- rlog(dds3, blind=TRUE)

# Extract variance-stabilised expression for interferon-related genes

ifn\_genes <- c("IFNA1","IFNA2","IFNB1","IFNL1","IFNL2","IFNL3") # example list

ifn\_mat <- assay(rld2)[ifn\_genes, ]

# Generate clustered heatmap

pheatmap(ifn\_mat,

cluster\_rows=TRUE, cluster\_cols=TRUE,

show\_rownames=TRUE, show\_colnames=TRUE,

color=colorRampPalette(rev(brewer.pal(9,"RdBu")))(255),

main="Interferon Gene Expression: RSV vs SARS-CoV-2")

To compare host interferon responses, variance-stabilised counts were extracted for selected interferon genes. A heatmap was generated to visualise expression differences between RSV and SARS-CoV-2 samples. Hierarchical clustering highlights the grouping of conditions, making it clear how interferon signalling differs between the two viral infections.

**Section C**

A graph with red green and blue dots

AI-generated content may be incorrect.

Figure 1: PCA plot of A549 transcriptomes post-infection with RSV and SARS-CoV-2 and IAV.

A screen shot of a graph

AI-generated content may be incorrect.A diagram of a heatmap

AI-generated content may be incorrect.

Figure 2: Sample-to-sample distance heatmap of infections Figure 3: Heatmap of interferon-stimulated gene expression in RSV- and SARS-CoV-2–infected A549 cells.

A graph of a graph with red dots

AI-generated content may be incorrect.A graph of red and black dots

AI-generated content may be incorrect.

Figure 4. Volcano plots of differential gene expression in A549 cells infected with RSV or SARS-CoV-2.

A graph of a virus

AI-generated content may be incorrect.

Figure 5: A summarised volcano plot representing gene expression between infected cells with RSV in comparison to infection with SARS-CoV-2.

We analysed a subset of the A549 RNA‑seq data corresponding to respiratory virus infections reported (1). Our goal was to compare global transcriptional responses across Respiratory Syncytial Virus (RSV) and SARS‑CoV‑2 infections, and to evaluate the magnitude and character of their interferon‑stimulated gene (ISG) programs. The first principal component (PC1; 58% of variance) separates RSV samples sharply from their mock controls (Figure 1), indicating a coherent shift in the A549 transcriptome upon RSV infection (Figure 4). SARS-CoV-2 replicates also separate from mock, but chiefly along PC2 (~10–15% of variance) and with a smaller displacement, consistent with a more muted response. The large gap between RSV and mock captures widespread up‑regulation of immune pathways documented below. The ordering along PC1 therefore provides a first, unbiased ranking of response magnitude: RSV > SARS-CoV-2. However, the second principal component (PC2; 19.8% variance) portrays a smaller separation in SARS-CoV-2, suggesting lesser transcriptional alterations due to the proportional variance observed.

A block‑diagonal structure mirroring the PCA was observed on the sample-to-sample distance matrix (Figure 2). Darker squares (low distance) occur along the diagonal within each virus group, confirming replicate consistency. RSV forms a distinct block furthest from mock, while SARS-CoV-2 shows moderate distances from its mock controls and substantially smaller distances than RSV–mock. Thus, RSV strongly alters the transcriptome while SARS-CoV-2 induces a milder shift, mirroring the interferon heatmap. Together, the PCA and distance heatmap establish that infection status—rather than batch—is the dominant driver of transcriptional variance.

We next contrasted infected versus mock for RSV and for SARS-CoV-2 (Figure 4), then plotted both sets of genes on one volcano to emphasise direction and significance (Figure 5). Points coloured red and blue meet significance thresholds, and black points do not. RSV shows a strong right‑skew, while SARS-CoV-2 exhibits a more balanced distribution with fewer extreme positives and a noticeable set of down-regulated genes(2).In contrast, upon plotting the infected cells against each other, it revealed a small number of genes were significantly expressed (Figure 5), with a higher quantity seen in RSV-infected cells. The contrasting shapes recapitulate the PCA ranking and suggest that RSV drives a dominant antiviral program, whereas SARS-CoV-2 both underinduces ISGs and actively represses subsets of host genes.

To focus on antiviral signalling, we visualised normalised expression of representative ISGs (e.g., IFIT1/2, MX1, OAS1). Rows were scaled to emphasise relative differences. RSV samples display intense warm colours across most ISGs, denoting robust induction. In contrast, SARS-CoV-2 samples are predominantly cool‑coloured and cluster with mock (Figure 3), consistent with attenuated type I interferon signalling (2)(5). RSV shows clear induction for many ISGs, typically seen at a substantial magnitude. Dendrograms cluster RSV away from SARS-CoV-2 and mock —again aligning with the genome-wide patterns.

A549 cells result in a vigorous, coordinated antiviral program to RSV, and a comparatively weak response to SARS-CoV-2. The weak ISG signature in SARS-CoV-2‑CoV‑2 despite measurable differential expression elsewhere suggests viral strategies that limit interferon production or signalling. This conclusion justifies subsequent experiments, typical (1). workflow, such as direct cytokine quantification or time‑course profiling to detect delayed responses. Overall, our subset analysis reproduces the central theme of Figure 1: coronaviruses elicit a blunted interferon response relative to other respiratory viruses, with RSV provoking a much stronger ISG program.

**Section D**

A critical direction would be validating interferon pathway suppression by SARS-CoV-2 using proteomic or cytokine assays, as transcriptional data alone cannot confirm protein-level changes, building upon Blanco-Melo et al (5). While bulk RNA-seq, as used here, is powerful for quantifying global expression changes and comparing conditions, it averages signals across populations of cells, masking cellular heterogeneity. Necessitating single-cell RNA-seq (scRNA-seq) that reveals differential responses within subsets of infected versus bystander cells, identifies rare cell populations, and uncovers regulatory programs (4)—advantages particularly relevant for studying heterogeneous tissues such as the lung. However, scRNA-seq is more costly, computationally demanding, and often less sensitive for lowly expressed genes compared with bulk RNA-seq (6). Together, using both approaches strategically can provide a complementary and more complete picture of host–virus interactions.

**Word Count: 1470**

**References**

1 - Blanco-Melo D, Nilsson-Payant BE, Liu WC, et al. Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell*. 2020;181(5):1036-1045. doi: 10.1016/j.cell.2020.04.026.

2 - Montazersaheb, S., Hosseiniyan Khatibi, S.M., Hejazi, M.S., et al. COVID-19 infection: an overview on cytokine storm and related interventions. *Virol J*. 2022;19:92. doi:10.1186/s12985-022-01814-1.

3 - Schroeder A, Mueller O, Stocker S, et al. The RIN: an RNA integrity number for assigning integrity values to RNA measurements. *BMC Mol Biol*. 2006;7:3. doi:10.1186/1471-2199-7-3.

4 - Fios Genomics. Single-Cell vs Bulk RNA Sequencing [Internet]. 2023 Apr 28 [cited 2025 Aug 19]. Available from: <https://www.fiosgenomics.com/single-cell-vs-bulk-rna-sequencing/>

5 - Znaidia M, Demeret C, van der Werf S, Komarova AV. Characterization of SARS-CoV-2 evasion: interferon pathway and therapeutic options. Viruses. 2022 Jun 8 [cited 2025 Aug 19];14(6):1247. Available from: <https://www.mdpi.com/1999-4915/14/6/1247>

6 - Chen G, Ning B, Shi T. Single-Cell RNA-Seq Technologies and Related Computational Data Analysis. Front Genet. 2019 Apr 5 [cited 2025 Aug 19];10:317. Available from: <https://www.frontiersin.org/journals/genetics/articles/10.3389/fgene.2019.00317/full>