الجامعة الافتراضية السورية
SYRIAN VIRTUAL UNIVERSITY

Syrian Virtual University
Masters in Bioinformatics

Gene Expression Responses to Viral Infection Using RNA-Sequencing: Differential Expression Analysis in SARS- CoV-2-Infected Organoids

By: Abdulhameed Al Khatib

Supervisor: Prof. Majd Al Jamali

Class of: S24

Table of Contents

List of Figures

Abstract

COVID-19, caused by the SARS-CoV-2 virus, is a complex, multi-organ disease with effects extending beyond the respiratory system to other vital organs such as the heart. This thesis investigates the transcriptomic responses in different tissues to uncover the molecular mechanisms driving COVID-19's systemic impact.

RNA sequencing (RNA-seq) was used to systematically profile transcriptional changes of hPSC-derived cells/organoids caused by SARS-CoV-2 infection. This study investigates the multi-organ impact of COVID-19. Using RNA sequencing, gene expression profiles were generated from cardiac and airway tissue samples obtained from infected and mock-treated controls. RNA-Seq analysis involved key steps such as Quality Control, Alignment to a reference genome, Gene quantification and Differential expression analysis to identify differentially expressed genes (DEGs).

Broad differential expression datasets were generated to explore transcriptional responses, providing a comprehensive overview of gene expression changes. Visualizations such as volcano plots and heatmaps highlighted key patterns, guiding further refinement. To ensure high-confidence results, stricter thresholds were applied, narrowing the datasets to focus on biologically significant genes.

The results revealed distinct patterns of upregulated and downregulated genes in both tissues, with a subset of overlapping DEGs suggesting shared pathways contributing to systemic inflammation. Tissue-specific analysis highlighted the enrichment of pathways related to immune response, cytokine signaling, and cell death in airway samples, while cardiac tissue exhibited significant enrichment in pathways associated with oxidative stress, fibrosis, and mitochondrial dysfunction. Venn diagram analysis further emphasized the tissue-specific and overlapping gene expression changes, providing insight into the multi-faceted nature of COVID-19 pathology.

This study highlights the utility of RNA-Seq in elucidating molecular mechanisms underlying COVID-19 as a multi-organ disease, offering valuable insights into potential therapeutic targets for mitigating its systemic effects.

Keywords: *RNA-Seq, COVID-19, SARS-CoV-2, Immune response, Differential Gene Expression (DEG), Read mapping, Alignment, Gene Ontology (GO) Enrichment, Biological Pathways, Gene Expression Patterns.*

Gene Expression Responses to Viral Infection Using RNA-Sequencing: Differential Expression Analysis in SARS- CoV-2-Infected Organoids.

1 Introduction

1.1 Background on SARS-CoV-2 and Its Impact on Human Health

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the virus responsible for the COVID-19 pandemic, belongs to the virus family *Coronaviridae*, the SARS-CoV-2 virus spread rapidly across countries and the COVID-19 outbreak was declared a pandemic by the World Health Organization (WHO) on March 11th 2020 (World Health Organization, 2020). As of 30 January 2022, over 370 million confirmed COVID-19 cases and more than 5.6 million deaths have been reported to World Health Organization (WHO) (World Health Organization, 2022).

Currently, SARS-CoV-2 has not disappeared and continues to prevail worldwide, with the ongoing risk of mutations and the potential for severe COVID-19.

SARS-CoV-2 is a single positive-stranded RNA enveloped virus that replicates in epithelial cells. During SARS-CoV-2 infection SARS-CoV-2 binds the host ACE2 receptor through its spike protein, and enters the cells by fusion of the viral membrane with the epithelial cell membrane or by endocytosis. After binding, the spike protein can be cleaved by TMPRSS2, a host membrane serine protease, facilitating viral entry. Then, the virus replicates inside epithelial cells and produces newly synthetized viral particles that are secreted by the host cells the virus transforms the infected host cell into factories that produce new viral particles (Assou et al. 2023). As infection progresses, the infected cells undergo numerous changes in various pathways. One of these changes is the occurrence of a cytokine storm —a hyperinflammatory state characterized by the excessive release of pro-inflammatory cytokines —which leads to severe symptoms.

The mortality is primarily linked to acute respiratory distress syndrome (ARDS) a severe complication that arises from an uncontrolled immune response. and the long-term effects of infection are still not known. However, COVID-19 is not limited to respiratory involvement—Coronavirus disease 19 (COVID-19) is a multi-organ disease caused by infection of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) (Nakayama et al. 2023). Although SARS-CoV-2 primarily infects the respiratory tract infecting epithelial cells in the lungs, patients with COVID-19 present with a wide range of disease indications, including the gastrointestinal, cardiovascular and neurological systems, likely through the angiotensin-converting enzyme 2 (ACE2) receptor, which is highly expressed in many tissues (Rabaan et al. 2023).

1.2 RNA Sequencing and Its Relevance to Viral Infection Studies

High-throughput next-generation sequencing has revolutionized analysis of both the genome and transcriptome in biologic research. RNA-Seq has become a routinely and extensively applied approach for transcriptome profiling that relies on high-throughput sequencing (HTS) technologies, which provides a far more profound and precise measurement at the transcript level than microarray and other traditional gene expression analysis methods (Deshpande et al. 2023).

RNA-seq allows complete annotation of structures of transcripts (5′, 3′ ends, as well as splice junctions), quantification of expressions of transcripts, measurement of extent of alternative splicing, and allele-specific expression, typical RNA-Seq workflow involves a complex process of extracting RNA from fresh or formalin-fixed paraffin-embedded tissue, conversion of RNA to complimentary DNA (cDNA), library preparation, sequencing (Li, Varghese, and Ressom 2024)

First, reads are mapped to the genome or transcriptome. Second, mapped reads for each sample are assembled into gene-level, exon-level or transcript-level expression summaries, depending on the aims of the experiment. Next, the summarized data are normalized in concert with the statistical testing of DE, leading to a ranked list of genes with associated P-values and fold changes. Finally, biological insight from these lists can be gained by performing systems biology approaches (Jiang et al. 2024)

RNA-Seq has been instrumental in understanding the molecular mechanisms and pathogenesis of COVID-19. By analyzing the transcriptomic changes in COVID-19 patients, researchers have identified key genes, pathways, and potential biomarkers associated with the disease's progression and severity. This approach has provided insights into the host immune response, the impact of SARS-CoV-2 on different cell types, and the identification of potential therapeutic targets (Erb et al. 2022).

1.3 Research Questions and Objectives.

This project aims to investigate the transcriptional changes induced by SARS-CoV-2 infection across different organoid models, addressing the following key research questions:

1. **How does SARS-CoV-2 infection alter gene expression in different organoids?**
   This question focuses on uncovering the tissue-specific transcriptional responses to SARS-CoV-2 infection. By leveraging RNA-Seq data from multiple organoid models, the study will identify differentially expressed genes (DEGs) in infected versus non-infected organoids. The findings will provide insights into the molecular mechanisms underlying tissue-specific responses to the virus.
2. **What are the common and unique gene expression signatures of SARS-CoV-2 across different human tissues?**
   This question investigates the shared and distinct molecular responses to SARS-CoV-2 infection across various human tissues. By comparing gene expression profiles among different

organoid types, the study seeks to identify common pathways involved in the viral response, as well as unique tissue-specific transcriptional signatures. These differences may offer explanations for the varying susceptibility and outcomes of SARS-CoV-2 infection in different organs.

2 Methodology

2.1 Dataset Description

To investigate the host response to SARS-CoV-2 infection, we selected bulk RNA-seq samples derived from human pluripotent stem cell (hPSC)-derived organoids, including lung AWOs organoids (AWOs), and cardiomyocytes (CMs) that were infected with SARS-CoV-2.

The Bulk RNA-Sequencing data used in this study was generated previously and is described exhaustively in (Tang et al. 2023). RNA-Seq samples were selected for both mock-treated wild-type (WT), included as controls and SARS-CoV-2-infected cells as the experimental group. WT samples served as the baseline behavior of cells in response to infections, which is crucial for understanding viral mechanisms and developing potential treatments by comparing the natural response of these cells to viral infections, such as SARS-CoV-2, without the influence of genetic manipulations.

The tissues in this study were generated from human pluripotent stem cells (hPSCs) Cell line of “H1”, a well-known human embryonic stem cell line often used in scientific research because of its ability to differentiate into various cell types including AWOs and CMs.

Specifically, the H1 stem cells were differentiated to form organoid structures. Organoids are complex 3D multicellular constructs that can be derived from either induced pluripotent stem cells (iPSCs)or adult stem cells and grown in a supportive extracellular matrix (*e.g.* Matrigel) to mimic in particular basement membrane components. Cells grown in this 3D culture system have been found to resemble more closely the *in vivo* environment, where cells spontaneously self-organize into hierarchies of stem/progenitor cells and create a continuum of more differentiated and functional cell types. (Ekanger et al. 2022)

Several research groups have been successful in the development of adult stem cell derived human AWOs and lung organoids. These organoids mimic the physiology of AWOs tissue, allowing to study respiratory infections (like SARS-CoV-2) and other organs’ responses in a more human-like environment (“Organoids as a Novel Tool in Modelling Infectious Diseases” 2023).

The RNA-Seq data with number (SRA) were retrieved from the NCBI Sequence Read Archive (SRA) database, under identifiers:

SRA: SRP375109
BioProject: PRJNA837900
GEO: GSE202963

The specific characteristics and metadata of the selected samples are detailed in the following table:

Tissue Type	Source Name	Cell Line	Cell Type	Genotype	Treatment	Sample Count
Airway	H1	H1	Airway Organoids	WT	Mock	5
Airway	H1	H1	Airway Organoids	WT	SARS-CoV-2 Infection	5
Cardio	H1	H1	Cardiomyocytes	WT	Mock	5
Cardio	H1	H1	Cardiomyocytes	WT	SARS-CoV-2 Infection	5

- Organisms: Homo sapiens (human).
- Sample Types: Organoids representing different human tissues (e.g., lung, heart, kidney) infected with SARS-CoV-2 and corresponding non-infected controls.
- RNA Sequencing Platform: Illumina NovaSeq 6000, providing high-resolution transcriptomic data.
- Infection Time Point: The dataset includes RNA-Seq data from organoids harvested at 48 hours post-infection with SARS-CoV-2.

2.2 Data Retrieval and Prepossessing

2.2.1 Retrieval of RNA-Seq Data from NCBI SRA

The Paired-End Bulk RNA-Sequencing data was retrieved using a Bash script that employs SRA toolkit specifically using the prefetch: 3.0.3. function.

The Bash script automates the download and preprocessing of RNA sequencing data for the project. It reads a list of Sequence Read Archive (SRA) accession numbers, provided in a text file, and systematically downloads each dataset from the NCBI SRA database.

Raw Paired-End RNA-Seq data stored in the FASTQ format were processed to generate the forward and reverse reads of paired-end RNA-Seq samples, as required by certain analysis tools. Using Fastq-dump command within the bash script, with the --split-3 parameter, the spots are split into *(biological)* reads, for each read : 4 lines of FASTQ or 2 lines are written. For spots having 2 reads, the reads are written into the *_1.fastq and *_2.fastq files.

The resulting FASTQ files are stored in a designated directory, ready for quality assessment and downstream analysis.

“ .fastq” file format is used for storing sequence reads generated from NGS (next-generation sequencing) instruments. as it not only contains the sequence data similar to FASTA file format but also includes quality information. Each record in a FASTQ file (sequence read) is structured with four lines; viewing the first 20 rows of SRR19196375_1.fastq.

1. **Identifier Line (Header)**
    - Begins with `@`, followed by a unique identifier for the sequence.
    - `SRR19196375.1` is the identifier for this particular read (in this case, from an NCBI sequence run).
    - The string `A00814:269:HTWLTDMXX:2:1101:1832:1000` contains sequencing instrument information, including the flow cell ID, lane, tile, and position on the flow cell.
    - `length=51` indicates the length of the read (51 bases in this example).
2. **Sequence Line**
    - This line contains the nucleotide sequence for the read.
    - In this sequence, `N` represents an unknown or ambiguous base.
3. **Separator Line**
   `+SRR19196375.1 A00814:269:HTWLTDMXX:2:1101:1832:1000 length=51`

- Begins with a `+` symbol and may repeat the header information. It serves as a separator between the sequence and quality score lines.

4. **Quality Score Line**
   `#FFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFFF`

- The fourth line is made of ASCII 33-126 symbols where each character represents a quality score for a corresponding base in the sequence line.
- Each symbol corresponds to an ASCII value, which represents the base quality score (e.g., `F` is a high-quality score)., representing the quality of the sequence reported in the second line. The range of the quality depends by the technology and by the chemistry of the sequencing
- `#` and `F` characters reflect Phred quality scores, with higher scores indicating more reliable base calling. Quality scores help indicate the confidence level in identifying each base.

to verify the completeness of the downloaded sequencing data, ensuring there were no errors or interruptions during the download and conversion process, (Shen, Sipos, and Zhao 2024) with its “stats” function was utilized to understand key statistics on the FASTQ files and ensure data integrity prior to downstream analyses. The results displayed in the output .txt file provided key metrics such as the number of sequences (reads) in each FASTQ file. For instance, in the SRR19196360 sample, both files in the pair (R1 and R2) contain 14,675,807 sequences, with an average sequence length of 51 bases.

This step confirms consistency across paired-end reads, as the number of sequences in paired files must match since they originate from the same template.

```
  file                format  type    num_seqs        sum_len  min_len  avg_len  max_len
SRR19196360_1.fastq  FASTQ   DNA   14,675,807    748,466,157       51       51       51
```

```
SRR19196361_1.fastq  FASTQ   DNA   16,950,233    864,461,883       51       51       51
SRR19196362_1.fastq  FASTQ   DNA   21,230,628  1,082,762,028       51       51       51
SRR19196372_1.fastq  FASTQ   DNA   32,294,781  1,647,033,831       51       51       51
SRR19196373_1.fastq  FASTQ   DNA   47,483,910  2,421,679,410       51       51       51
SRR19196374_1.fastq  FASTQ   DNA   25,668,372  1,309,086,972       51       51       51
SRR19196375_1.fastq  FASTQ   DNA   79,960,920  4,078,006,920       51       51       51
```

According to general recommendations for gene expression using RNA-Seq, Paired-end sequencing enhances splice junction identification by providing two reads from each DNA fragment, which helps in accurately mapping the reads to the reference genome. This dual-read approach allows for better resolution of complex transcript structures.

Short read, like 50 bp, can effectively capture the overall expression levels of genes, especially in well-annotated genomes. They can provide a general overview of which genes are being expressed and at what levels. However, the ability to resolve splice junctions and alternative splicing events may be compromised with shorter reads. Longer reads are generally better for identifying these features, as they can span across exons and introns more effectively, providing clearer insights into transcript diversity and complexity. especially in well-annotated genomes, number of reads per sample should exceed 5 million reads (Zhao 2014).

2.2.2 Quality Control of raw reads

2.2.2.i Importance of Quality Control in RNA-Sequencing Experiments.

The sequencing process using NGS methods is complex and can introduce various errors. Starting from a possible poor quality of starting material, improper library preparation (e.g., inefficient adapter ligation or PCR amplification bias), even incorrect base calls and technical sequencing errors.

Thus, raw RNA-Seq data may have quality issues, which can significantly distort analytical results and lead to erroneous conclusions. Therefore, the raw data must be subjected to vigorous quality control (QC) procedures before downstream analysis for successful RNA-seq experiments as the QC process improves the reproducibility of the biological results (Sheng et al. 2016).

A comprehensive framework for conducting QC for RNA-seq would be examining four critical perspectives: (1) RNA quality, (2) raw read data (FASTQ), (3) alignment and (4) gene quantification. While the QC processes for RNA-seq share similarities with those used for DNA sequencing data, there are several unique characteristics inherent to RNA-seq data that necessitate tailored approaches to ensure the integrity and reliability of the results (Zhou et al. 2018).

For sRNA-seq, the standard read length is 50 nucleotides (single-end 50 cycles), and as the majority of the sRNA-seq is <50 nucleotides, this increases the likelihood of sequencing of the attached adapter sequence. Thus, adapter trimming is required for sRNA-seq data analysis (Hu et al. 2024).

To correctly evaluate the quality of the data and the results, a multi-perspective QC strategy needs to be applied, which extends throughout the full data processing course. Specifically, strong emphasis is given to focusing raw data QC on the initial stage of high-throughput sequencing technology, for example, QC on the raw data do not guarantee a good alignment rate, and QC on the alignment data can detect library construction issues, but does not identify cross-sample contamination (Sheng et al. 2016).

2.2.2.ii Initial Quality assessment and RNA data specifications. Tools used (FastQC, MultiQC)

To understand what information is stored in all FASTQ files of the samples, FastQC tool was used to examine quality metrics for data.

FastQC measures several quality metrics about sequenced reads at the raw data level in the FASTQ file, providing information about read length, average quality score at each sequenced base, GC content, presence of any overrepresented sequences (k-mers), the quality score distribution across reads, per base sequence content (%A/T/G/C), adapter contamination and overrepresented sequences.

The output from FastQC, after analysing a FASTQ file of sequence reads, is an HTML file that may be viewed in browser. To facilitate a comprehensive quality assessment across multiple samples, we employed MultiQC version 0.4 (Ewels et al. 2016) to aggregate and summarize the individual FastQC reports. This tool generates an HTML report that visually represents key quality metrics across all samples, alongside tab-delimited files containing detailed FastQC statistics.

2.2.2.iii RNA-Sequencing library specifications for FastQC reports.

The interpretation of the quality plots can vary depending on the nature and context of sequencing data. Given the unique characteristics of RNA-Seq libraries, it is essential to interpret FastQC metrics in context.

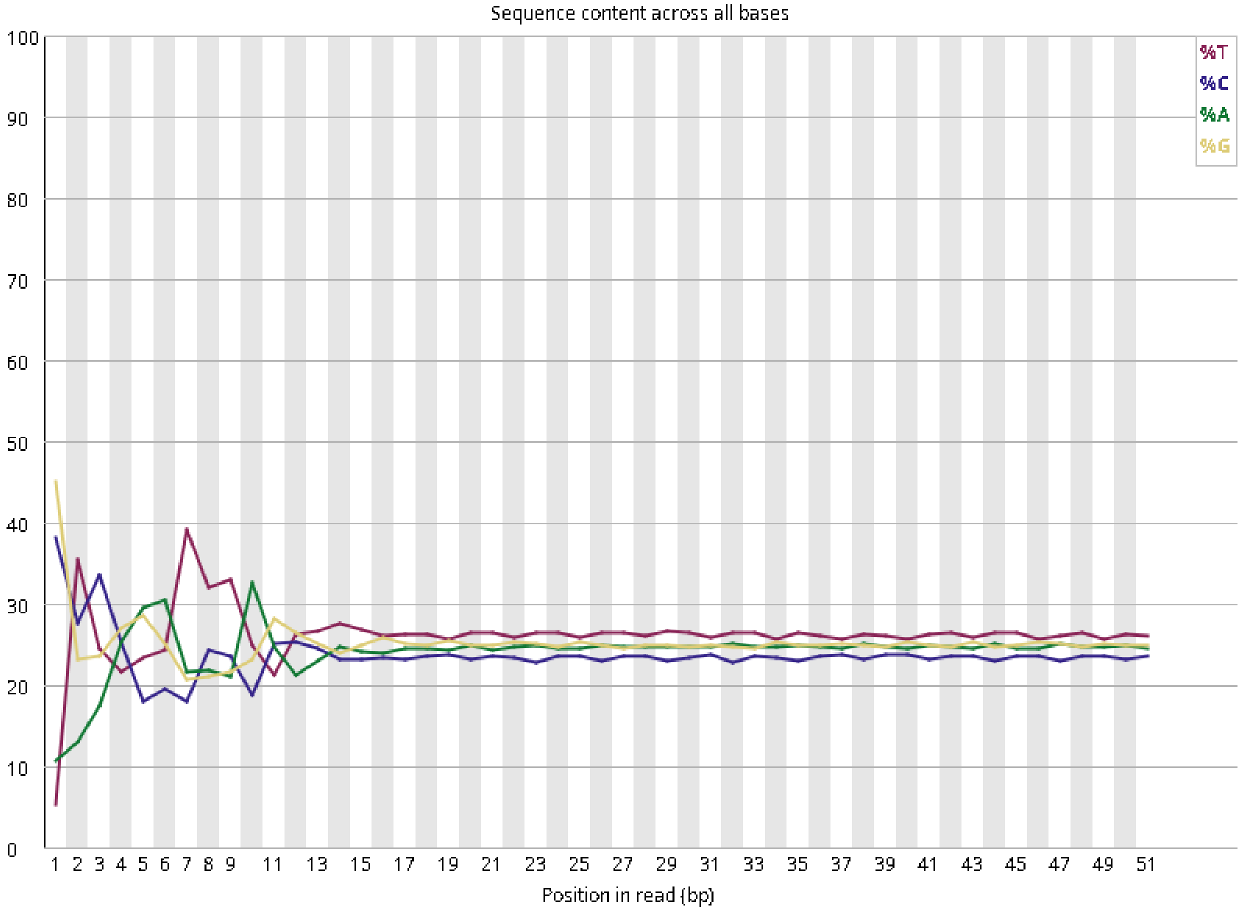
Libraries flagged for issues such as sequence duplication or GC bias may still be suitable for analysis. For example:

- High levels of duplication are often observed in RNA-Seq due to the amplification of highly expressed genes.
- Overrepresented sequences are expected for highly expressed genes such as ribosomal or housekeeping genes. While flagged by FastQC, this is an inherent feature of RNA-Seq and not a quality issue.

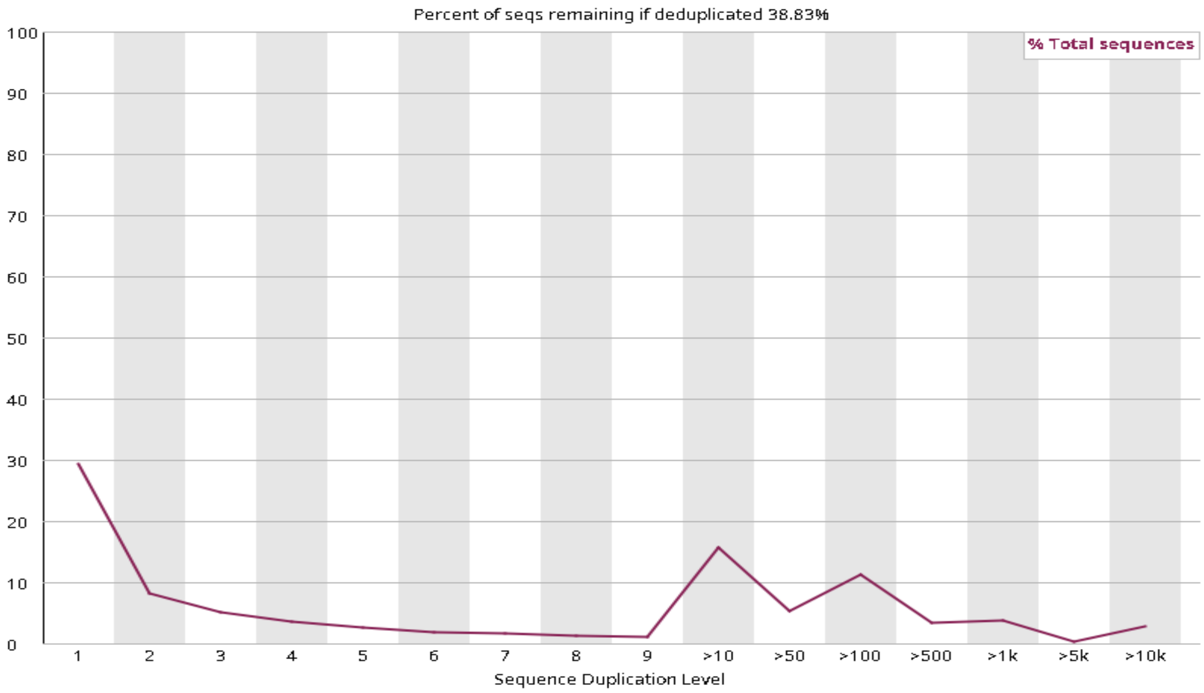
The key metric to monitor is the graph representing the average quality scores, along with the distribution of scores at each base across the length of the reads. Additionally, the Adapter Content plot is crucial, as libraries containing small RNA fragments tend to retain sequencing adapters more frequently, resulting in elevated adapter content in FastQC assessments.

2.2.2.iv Interpretation of Specific FastQC Plots

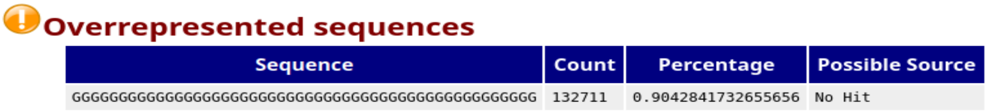
The plots below provide examples from a FastQC HTML report, highlighting key metrics and unique characteristics specific to RNA-Seq libraries:



*2.1 The **"Per base sequence content"** plot displays the proportion of each DNA base at each position across the sequence reads. In RNA-seq data, this plot typically results in a FAIL due to the non-random nature of hexamer priming during library preparation, leading to an enrichment of specific bases in the initial nucleotides (first 10-12 bases). While this bias cannot be corrected through processing, it does not adversely affect the measurement of gene expression.*

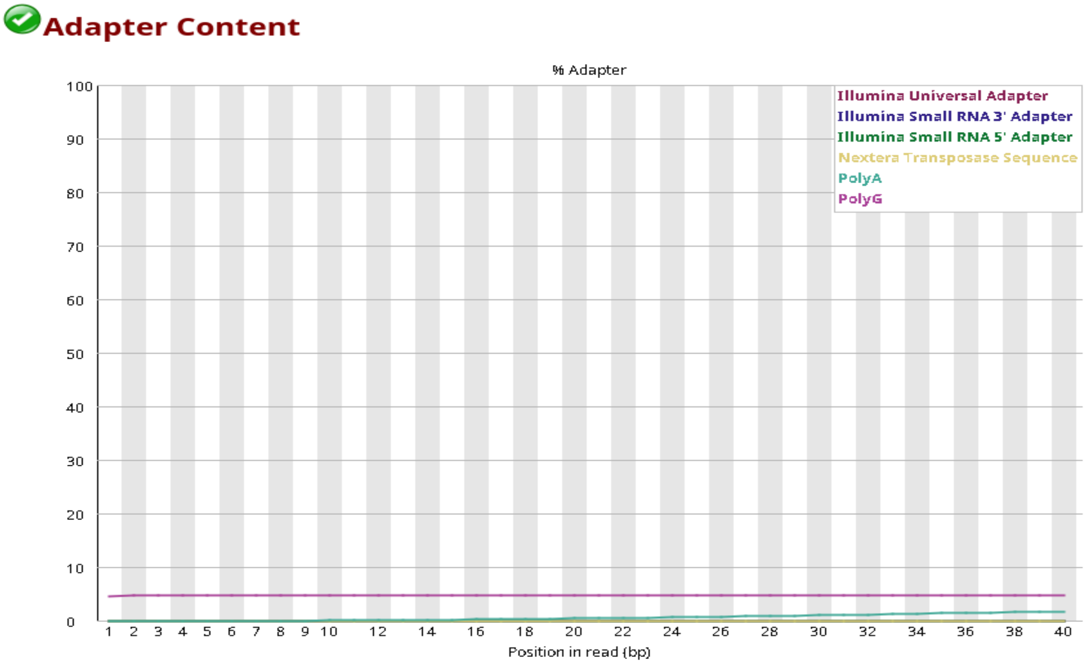


2.2 The ***Duplicate Sequences*** *plot illustrates the distribution of duplicate reads in RNA-Seq libraries, highlighting the prevalence of over-sequenced transcripts, particularly in high-expression regions. The Overrepresented Sequences plot reveals the presence of identical sequences, common in small RNA libraries, which can arise from unintentional sequencing of short RNA molecules or repetitive regions.*



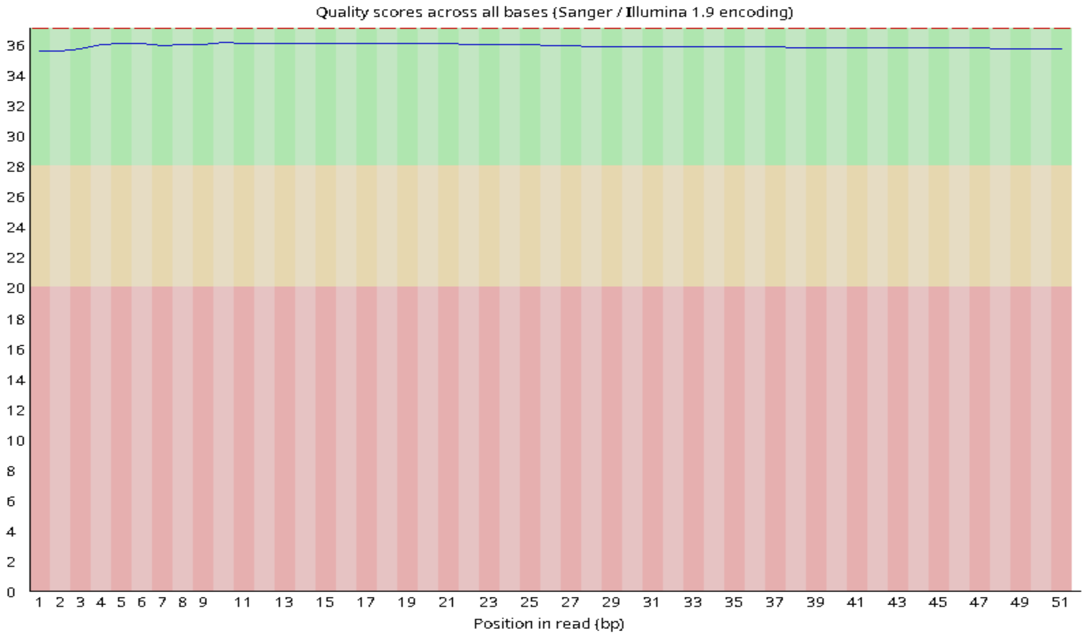
Overrepresented sequences

Sequence	Count	Percentage	Possible Source
GGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGGG	132711	0.9042841732655656	No Hit



Adapter Content

2.3 The Overrepresented Sequences Table and Adapter Plot in the FastQC report highlight the presence of identical sequences and potential adapter contamination in the sequencing data. The table highlights sequences that are significantly more abundant than expected, often indicating overrepresented short RNA molecules such as tRNAs or degraded fragments. This can occur due to non-random fragmentation and inadequate library preparation that fails to remove smaller fragments. Meanwhile, the Adapter Plot provides insight into the frequency of adapter sequences, which can also contribute to the observed overrepresentation, particularly in cases where library preparation does not effectively eliminate these contaminants.



2.4 The Per Base Quality Plot in the FastQC report displays the average quality scores across all bases in the sequencing reads. X-axis represents a different read position, with the y-axis indicating the Phred quality score. A higher score reflects better quality, while a drop in quality at specific positions may suggest issues such as sequencing errors or declining read quality towards the end of reads. This plot is crucial for assessing the overall reliability of the sequencing data and identifying any potential regions that may require further investigation or filtering before downstream analysis.

2.2.2.v Observations from specific FastQC plots for the infected samples:

When Analyzing RNA-seq data from SARS-CoV-2 infected samples, FastQC may show some distinct patterns compared to mock (uninfected) samples due to the presence of viral RNA (Westermann and Vogel 2018) for example:

- Overrepresented sequences in infected samples will likely include viral RNA fragments, especially highly expressed viral genes (e.g., N, S, M, and ORF1ab genes).
- Infected samples are likely to show higher levels of sequence duplication if viral transcripts are highly expressed, as the same viral reads could appear frequently.
- Infected samples with a high viral load might show a wider distribution of sequence quality scores, as viral sequences can differ in quality compared to host reads, depending on sequencing depth and library composition.

2.2.2.vi Steps for assessing and improving data quality:

After the evaluation of the MultiQC report, failed sampled in Adapters Plot and Overrepresented Sequences Table were filtered for further processing to completely remove ploy A, Ploy G and TruSeq Adapter and contaminant sequences identified. Steps taken:

- Blast Overrepresented Sequences: Overrepresented sequences were extracted directly from the MultiQC report and submitted to the NCBI BLAST website (Basic Local Alignment Search Tool) using **Blastn** to check their origin and remove reads that match contaminant RNA fragments like tRNA and rRNA, and confirm the viral sequences flagged as Overrepresented in infected samples. Ribosomal RNA and tRNA are considered an internal contamination since they are from the target sequencing species

- Trimming with Trimmomatic: Trimming of specified Poly G, Poly A, and contaminant sequences was conducted using the **Trimmomatic** tool, which is widely used for read preprocessing in high-throughput sequencing (Bolger, Lohse, and Usadel 2014). The trimming step targeted adapter sequences, and specific contaminant sequences such as Poly G and Poly A that were identified during the FastQC evaluation. Custom parameters were set in Trimmomatic to effectively remove these undesired sequences while retaining the maximum usable portion of the reads.

After trimming, FastqC was performed on the trimmed FASTQ files, to confirm and evaluate the step of quality control the quality control. Key metrics, including the removal of overrepresented sequences and improved adapter content scores, were inspected to confirm that the trimming had resolved previously identified issues.

2.2.3 Alignment to reference genome

To use RNA-seq data to compare expression between samples, it is necessary to turn millions of short reads into a quantification of expression. The first step in this procedure is the Read Mapping or Alignment to determine where in the genome the reads originated from.

Once the reads are in trimmed FastQ format, the RNA-seq reads were aligned to a combined reference genome comprising both the human genome (**GRCh38_113**) and the viral genome and the SARS-CoV-2 genome (**NC_045512.2**).

The task of mapping is to find the unique location where a short read is identical to the reference. However, in reality the reference is never a perfect representation of the actual biological source of RNA being sequenced. In addition to sample-specific attributes such as SNPs and indels (insertions or deletions), there is also the consideration that the reads arise from a spliced transcriptome rather than a genome.

Furthermore, short reads can sometimes align perfectly to multiple locations and can contain sequencing errors that have to be accounted for. Therefore, the real task is to find the location where each short read best matches the reference, while allowing for errors and structural variation (Chung et al. 2021).

Mapping RNA-seq data requires splicing-aware mappers to align transcriptomic reads to a reference genome accurately. As these reads originate from mRNA, it is expected that transcriptomic reads will cross exon/intron boundaries when aligned to the reference genome. Counting the number of times, a read mapped to a specific gene gives us information about how "high" or "low" a gene was being expressed.

The STAR (Spliced Transcripts Alignment to a Reference) aligner was employed, along with Bash scripting to capture both viral and human transcripts simultaneously (Dobin et al. 2013).

The alignment process consists of two steps:

1. Indexing the reference genome, Indexing allows the aligner to quickly find potential alignment sites for query sequences in a genome, which accelerates the search for potential alignment sites during mapping. Indexing the reference only has to be run once.

2. Aligning the reads to the reference genome. Following indexing, each read from the trimmed FastQ files is aligned to the reference genome. This process determines the genomic coordinates corresponding to each mRNA fragment, enabling the quantification of gene expression levels.

2.2.3.i Downloading reference genome (FASTA) and GTF annotation files.

The reads for this study were aligned to the Ensembl release GRCh38.p14 (GCA_000001405.29) human reference genome. https://ftp.ensembl.org/pub/release-113/gtf/homo_sapiens/ combined with the reference assembly for the Wuhan-Hu-1 isolate which has been imported from ENA (ASM985889v3, GCA_009858895.3, MN908947.3). http://ftp.ensemblgenomes.org/pub/viruses/fasta/sars_cov_2/dna/.

2.2.3.ii Building the genome index

The initial step in read alignment involves constructing a genome index. An index serves as a genome's "table of contents," allowing the aligner to pinpoint relevant genomic regions quickly, significantly reducing computational time. This step only needs to be performed once for a given reference genome.

To build the genome index, it is required to select compatible reference genome sequences (FASTA) and gene annotation files (GTF or GFF3) which indicates the locations of all genes within the reference genome (provided as a FASTA file), i.e. the same version number and from the same source (Ensembl, UCSC or NCBI). the FASTA and corresponding annotation files (GTF files) were downloaded from the Ensembl website to maintain compatibility across versions, ensuring consistency in chromosome names across the reference genome and annotation files.

To build the genome index, we employed the STAR aligner (Spliced Transcripts Alignment to a Reference) in `--genomeGenerate` mode and supply it with the human reference genome sequence (FASTA) and corresponding annotation (GTF) file that were sourced from Ensembl (release 75).

the output from successfully running the command was as the following:

```
STAR version: 2.7.11b   compiled: 2024-07-03T14:39:20+0000
:/opt/conda/conda-bld/star_1720017372352/work/source
Nov 21 00:47:45 ..... started STAR run
Nov 21 00:47:45 ... starting to generate Genome files
Nov 21 00:48:24 ..... processing annotations GTF
Nov 21 00:48:52 ... starting to sort Suffix Array. This may take a long
time...
Nov 21 00:49:03 ... sorting Suffix Array chunks and saving them to
disk...
Nov 21 01:25:55 ... loading chunks from disk, packing SA...
Nov 21 01:27:34 ... finished generating suffix array
Nov 21 01:27:34 ... generating Suffix Array index
Nov 21 01:31:43 ... completed Suffix Array index
Nov 21 01:31:44 ..... inserting junctions into the genome indices
```

2.2.3.iii Mapping RNA-Seq reads to the reference genome using STAR

The mapping process aligns each read from the trimmed FASTQ file (line 3) to the reference genome. This involves identifying the location within the reference genome that corresponds to the mRNA fragment represented by each read. The alignment reveals which gene was transcribed to produce the original mRNA and allows for quantification of gene expression levels by counting the number of reads aligned to each gene (Dobin and Gingeras 2015) .

Key considerations during alignment included:

- Read Type: Single-end or paired-end sequencing data.
- Strandedness: Whether the library was stranded and, if so, whether the standard dUTP method was used (detected automatically by STAR).

Additionally, the aligner requires the GTF that was downloaded earlier and corresponds precisely to the reference genome version. According to the NCBI SRA platform from which the data was retrieved:

“The RNA-Seq dataset (for this project BioProject: PRJNA837900 *was generated on the Illumina NovaSeq 6000 platform using a paired-end strategy. libraries were prepared with the TruSeq Stranded mRNA Library Kit (Illumina), ensuring high-quality, reverse-stranded data. Paired-end sequencing enabled accurate alignment and detection of complex transcript structures, while stranded preparation preserved transcript directionality for precise mapping.”*

The following output is displayed during the mapping process when running STAR.

```
   STAR version: 2.7.11b   compiled: 2024-07-03T14:39:20+0000
:/opt/conda/conda-bld/star_1720017372352/work/source
Nov 24 21:08:47 ..... started STAR run
Nov 24 21:08:47 ..... loading genome
Nov 24 21:09:14 ..... started mapping
Nov 24 21:10:45 ..... finished mapping
Nov 24 21:10:49 ..... started sorting BAM
Nov 24 21:10:57 ..... finished successfully
```

In the defined output directory, the following files are output from the paired-end reads mapped to the reference genome. including the prefix added by the code for every specific sample:

```
SRR19196375_Cardio_CoV_Aligned.sortedByCoord.out.bam
SRR19196375_Cardio_CoV_Aligned.toTranscriptome.out.bam
SRR19196375_Cardio_CoV_Log.final.out
SRR19196375_Cardio_CoV_Log.out
SRR19196375_Cardio_CoV_Log.progress.out
SRR19196375_Cardio_CoV_ReadsPerGene.out.tab
SRR19196375_Cardio_CoV_SJ.out.tab
SRR19196375_Cardio_CoV_STARgenome
```

The following table outlines key output files generated during the STAR mapping process, along with their respective descriptions.

File	Description
_ReadsPerGene.out.tab.	contains the number of reads that were mapped to each gene in the transcriptome.
_Aligned.sortedByCoord.out.bam	Alignment in BAM format (sorted by coordinate)
_Log.final.out	Alignment summary statistics such as uniquely mapped reads, percent mapping, number of unmapped reads, etc.
_Log.out	Alignment log for commands and parameters (useful in troubleshooting)
_Log.progress.out	Alignment progress report (e.g. number of reads processed during particular span of time, mapped and unmapped reads, etc.)
_sampleSJ.out.tab	Filtered splice junctions found during the mapping stage

2.2.3.iv Quality assessment of aligned reads

STAR can align spliced sequences of any length with moderate error rates when using optimized parameters, providing scalability for emerging sequencing technologies. However, QC on the alignment result file—Sequencing Alignment Map (SAM) or its equivalent compressed format of the Binary Alignment Map (BAM)—can yield additional insight into the quality of the sample and capture bad quality samples not detectable by raw data QC.

For example, it can capture efficiency and contamination of RNA from an unwanted source (other than an adapter sequence) which cannot be easily detected during raw data QC.

The command flagstat from SAMtools can also produce a quick summary of mapped, unmapped, discordantly mapped and properly paired reads from the BAM files (Danecek et al. 2021). These assessments include multiple aspects:

1. Reads:

- Number of total reads and mapped reads;
- Number of reads mapped to each specific genomic region (such as CDS and exon), which is defined in the user-specified gene model (GTF or GFF) file;
- Number of reads mapped outside the genomic regions specified in the gene model (GFF/GTF) file.

2. Coverage (gene is called “expressed” when 50% of its sequence are mapped by reads):

- Number of expressed gene and its proportion out of all genes;
- Coverage of each gene and the overall coverage distribution;
- Distribution of mapped reads.

3. Mapping:

- Gene coverage bias: average mapping coverage of each base position over the genes (scale all of the transcripts into 100 bp windows);
- Strand specificity: reads mapped to positive/negative strands, respectively;
- Library complexity: number of reads with varied mapping starting point.

4. Pair-ended read mapping:

- Number of paired mapped reads;
- Number of discordantly mapped pairs;
- Insert size distribution of mapped read pairs.

The following table summarizes the alignment statistics extracted from the alignment_stats.txt file:

```
cat  alignment_stats.txt
96736489 + 0 in total (QC-passed reads + QC-failed reads)
87725114 + 0 primary
9011375 + 0 secondary
0 + 0 supplementary
0 + 0 duplicates
0 + 0 primary duplicates
93711980 + 0 mapped (96.87% : N/A)
84700605 + 0 primary mapped (96.55% : N/A)
87725114 + 0 paired in sequencing
43862557 + 0 read1
43862557 + 0 read2
84120286 + 0 properly paired (95.89% : N/A)
84120286 + 0 with itself and mate mapped
580319 + 0 singletons (0.66% : N/A)
0 + 0 with mate mapped to a different chr
0 + 0 with mate mapped to a different chr (mapQ>=5)
```

Additionally, the alignment summary file _Log.final.out provides a comprehensive overview of the final mapping statistics. It includes detailed metrics such as the total number of input reads, average read lengths, counts and percentages of uniquely mapped reads, spliced reads, chimeric reads, and the number of unmapped reads.
Following is an example of a component from _Log.final.out file:

```
Started job on |  Dec 16 17:44:29
                       Started mapping on |  Dec 16 17:47:18 Fin-
                             ished on |  Dec 16 17:51:00
    Mapping speed, Million of reads per hour |  274.87
                     Number of input reads |  16950233
                     Average input read length |  102
                             UNIQUE READS:
              Uniquely mapped reads number |  15055480
                  Uniquely mapped reads % |  88.82%
              Average mapped length |  95.36
                 Number of splices: Total |  4688986
                 Number of splices: Annotated (sjdb) |4669106
                 Number of splices: GT/AG |  4651066
                 Number of splices: GC/AG |  32513
                 Number of splices: AT/AC |  3514
           Number of splices: Non-canonical |  1893 Mis-
                match rate per base, % |  0.20%
```

```
                    Deletion rate per base |  0.00% Dele-
                         tion average length |  1.57
                      Insertion rate per base |  0.00%
                      Insertion average length |  1.30
                       MULTI-MAPPING READS:
         Number of reads mapped to multiple loci |  1692962
            % of reads mapped to multiple loci |  9.99%
            Number of reads mapped to too many loci |  138313
            % of reads mapped to too many loci |  0.82%
                      UNMAPPED READS:
   Number of reads unmapped: too many mismatches |  776
      % of reads unmapped: too many mismatches |  0.00%
      Number of reads unmapped: too short |  55891
           % of reads unmapped: too short |  0.33% Number
          of reads unmapped: other |  6811
             % of reads unmapped: other |  0.04%
             CHIMERIC READS:
              Number of chimeric reads |  0
               % of chimeric reads |  0.00%
```

Key alignment metrics and summary:

- **Number of Input Reads**: Reflects the total reads processed, serving as a baseline for assessing sequencing completeness. Discrepancies may indicate sequencing or data transfer issues.
- **Average Read Length**: Confirms consistency with experimental design. Deviations may signal incomplete sequencing or library preparation problems.
- **Uniquely Mapped Reads (Counts/Percentages)**: Indicates reads aligned to a single genomic location. High percentages suggest quality alignment; low percentages may indicate contamination or sequencing issues.
- **Spliced Reads**: Represents reads spanning exon-exon junctions, crucial for validating RNA-Seq data and ensuring proper transcript assembly.
- **Chimeric Reads**: Align to multiple regions, linked to structural variations, fusion transcripts, or artifacts, aiding in detecting genomic rearrangements.
- **Unmapped Reads**: Includes reads failing alignment due to mismatches, insufficient matching bases, or lack of suitable alignment, highlighting potential contamination or reference genome issues.

The quality summary indicates an average input of `16950233` reads per sample with an average read length of **102 bases**. The proportion of uniquely mapped reads ranged from **88.82%**, while multi-mapped reads accounted for **9.99%.** Chimeric reads were absent (**0.00%**), ensuring the specificity of alignments. Spliced reads, totaling **6,355,032**, included 4669106 **annotated splices**, with the majority being canonical GT/AG splices (**32513**) and minimal non-canonical splices (**1893**). These metrics validate the alignment process, confirm transcriptome coverage, and demonstrate the suitability of the data for downstream analyses.

2.2.4 Gene Count Generation and Data Normalization

Once it is determined that the alignment step was successful, the next step is to count how many times each "feature" is observed in each sample in order to enumerate the number of reads associated with the genes (Jiang et al. 2024).

The goal of this step is to produce a count table using FeatureCounts to collect the raw gene count information in order to test for differential expression (Liao, Smyth, and Shi 2014); RNA aligners, e.g. STAR, will output multiple best alignments of a single read by default in the SAM/BAM file.

Therefore, counting reads directly from the SAM/BAM file will inflate the aligned read number because of this multiple alignment. To solve this, some tools, such as FeatureCounts, count a read by a fraction, so if a read has two best alignments, 0.5 will be added to both of the loci's read counts.

FeatureCounts tool will require the alignment file (BAM), and the associated gene annotation file (GTF). The main output of FeatureCounts is a table with the counts, i.e. the number of reads (or fragments in the case of paired-end reads) mapped to each gene (in rows, with their ID in the first column) in the provided annotation. FeatureCount generates also the feature length output datasets. The following is a summary of gene counts for the sample `SRR19196360` generated with FeatureCounts:

```
cat
SRR19196360_Cardio_Mock_Aligned.sortedByCoord.out_gene_counts.txt.summary
Status  /home/RNA-Sequencing-project/fastq-
files/mapped_try/SRR19196360_Cardio_Mock_Aligned.sortedByCoord.out.bam
Assigned    11393926
Unassigned_Unmapped 266615
Unassigned_Read_Type    0
Unassigned_Singleton    0
Unassigned_MappingQuality   0
Unassigned_Chimera  0
Unassigned_FragmentLength   0
Unassigned_Duplicate    0
```

```
Unassigned_MultiMapping 2840758
Unassigned_Secondary    0
Unassigned_NonSplit 0
Unassigned_NoFeatures   3015747
Unassigned_Overlapping_Length   0
Unassigned_Ambiguity    452107
```

a total of 11,393,926 reads were successfully assigned to genes. However, there are notable numbers of unassigned reads, including 266,615 that were unmapped and 2,840,758 that mapped to multiple locations. Additionally, 3,015,747 reads did not overlap with any known features in the reference genome. Other categories of unassigned reads, such as those related to read type, mapping quality, and duplicates, showed no occurrences. Overall, a significant number of reads were assigned and is considered sufficient for differential analysis testing.

2.3 Differential Gene Expression Analysis

The goal of a DE (differential expression) analysis is to highlight genes that have changed significantly in abundance across experimental conditions. In general, this means taking a table of summarized count data for each library and performing statistical testing between samples (Babu and Nobel 2022).

In this study, raw count data generated during RNA-seq processing is prepared for differential gene expression analysis using DESeq2. The count-based statistical methods, such as DESeq2 (Love, Huber, and Anders 2014), expect input data as obtained, e.g., from RNA-seq or another high-throughput sequencing experiment, in the form of a matrix of un-normalized counts where rows represent genes and columns represent samples. Each cell contains the count of sequencing reads mapped to.

After generating the gene counts using featureCounts function (Liao, Smyth, and Shi 2013) in the Rsubread package, the matrix of read counts can be directly provided from the "counts.txt" element in the list output. The count matrix and column data were typically loaded to R using base R function read.delim.

2.3.1.i Pre-filtering of the count matrix.

To refine the dataset for accurate differential expression analysis, several pre-filtering steps were applied to the count matrix, reducing noise and improving the relevance of the dataset for downstream analysis. These steps included:

1. Exclusion of Viral and Non-Relevant Genes

The initial filtering step removed viral genes and any genes with zero counts across all samples. Genes with no expression in any condition (mock or infected) lack variability, which is crucial for statistical comparisons in differential expression analysis. By excluding these genes, the focus remains on those that provide relevant insights into the biological response to SARS-CoV-2 infection.

2. Filtering Based on Expression Levels

Genes with very low expression levels were excluded, as they are unlikely to contribute meaningful biological information, particularly in bulk RNA-Seq datasets. To achieve this, genes with fewer than 10 counts in at least five samples of the same condition were filtered out. This step prioritized genes with consistent expression patterns, increasing the reliability and interpretability of downstream results.

3. Removal of Genes with Consistently Low Expression

Genes exhibiting consistently low expression across all samples were removed. These genes provide minimal statistical power and are unlikely to reveal significant patterns of differential expression.

After applying this filter, 12,586 genes remained from the original 56,637, ensuring the dataset was focused on genes with the potential to yield meaningful insights.

4. Verification and Alignment of Metadata

The count matrix was aligned with metadata detailing experimental conditions, including **Condition** (Mock vs. SARS-CoV-2 Infection) and **Tissue** (AWOs vs. CMs). Ensuring proper alignment between the count matrix and metadata is required for accurately modeling differential expression.

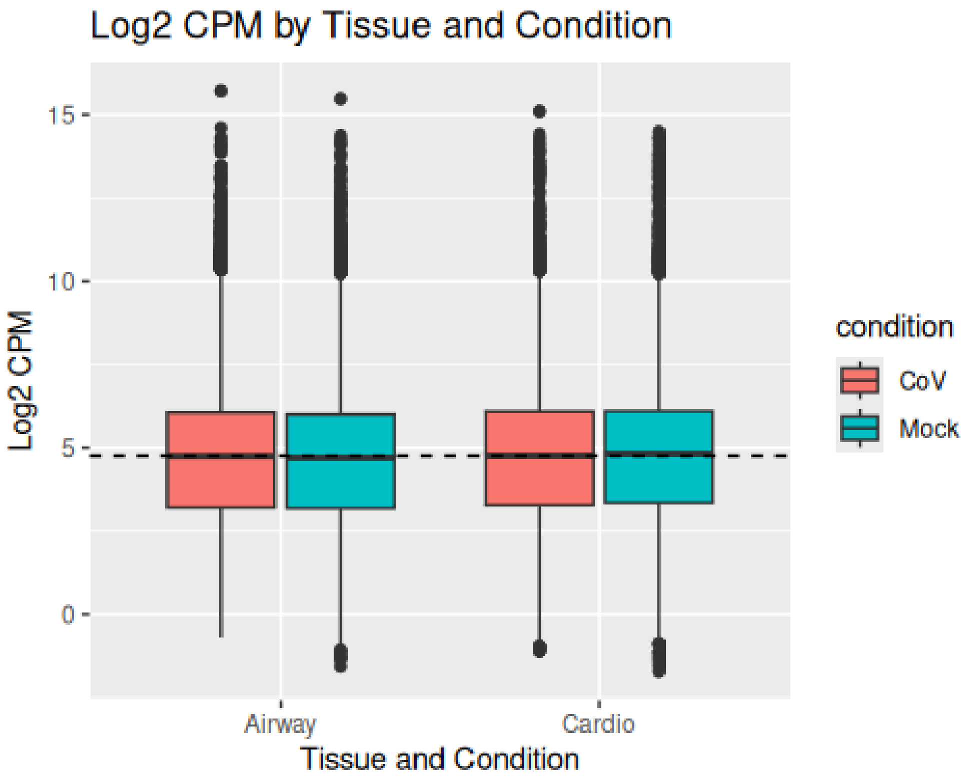
2.3.1.ii Creation of the DESeqDataSet Object

Using the filtered count matrix and metadata, a DESeqDataSet object was constructed to serve as the foundation for differential expression analysis.

2.3.2 Overview of exploratory data analysis (EDA)

2.3.2.i Visualize Library Size

EDA was performed to assess the RNA-Seq datasets from AWOs and CMs tissues, to visualize whether there is a difference in the distribution of raw counts after filtering, the count per million (CPM) values for each gene were calculated using the `cpm` function (Shreffler and Huecker 2025). These values were log-transformed to normalize the data and reduce the influence of extreme values. To visually assess the distribution of expression levels, the log2-transformed CPM values were organized and plotted as boxplots grouped by tissue type (`AWOs` and `CMs`) and experimental condition (`Mock` and `CoV`).

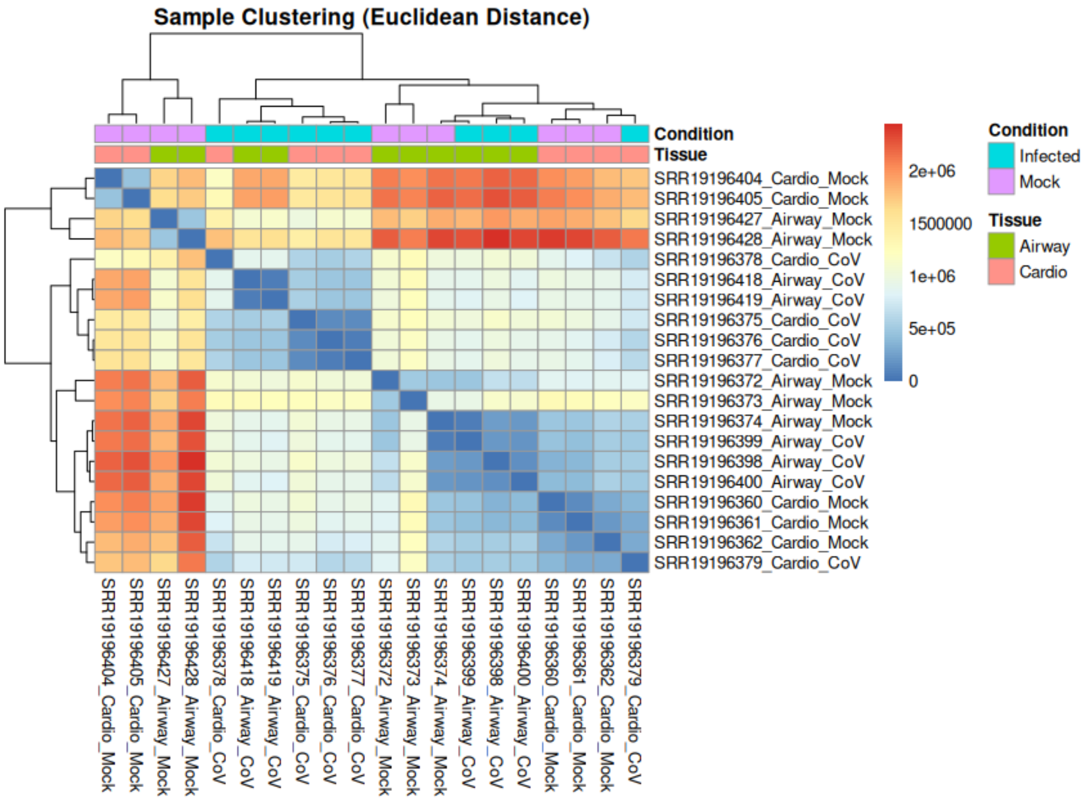


2.5 Boxplot showing the distribution of log2-transformed counts per million (CPM) values across tissues (AWOs and CMs) and experimental conditions (SARS-CoV-2 infected [CoV] and Mock). The plot highlights the overall expression levels for each group, with comparable median values and interquartile ranges, indicating consistent sequencing depth and data quality between conditions and tissue types. Points outside the whiskers represent genes with exceptionally high or low expression in the corresponding group. differences in median and spread between CoV and Mock can indicate how much infection alters gene expression.

2.3.2.ii Heatmap Visualizations

Sample Distance Heatmap

The distance heatmap calculates the Euclidean distance between the expression values for each individual sample. This is a way to identify which samples are most closely related in terms of their gene expression patterns.

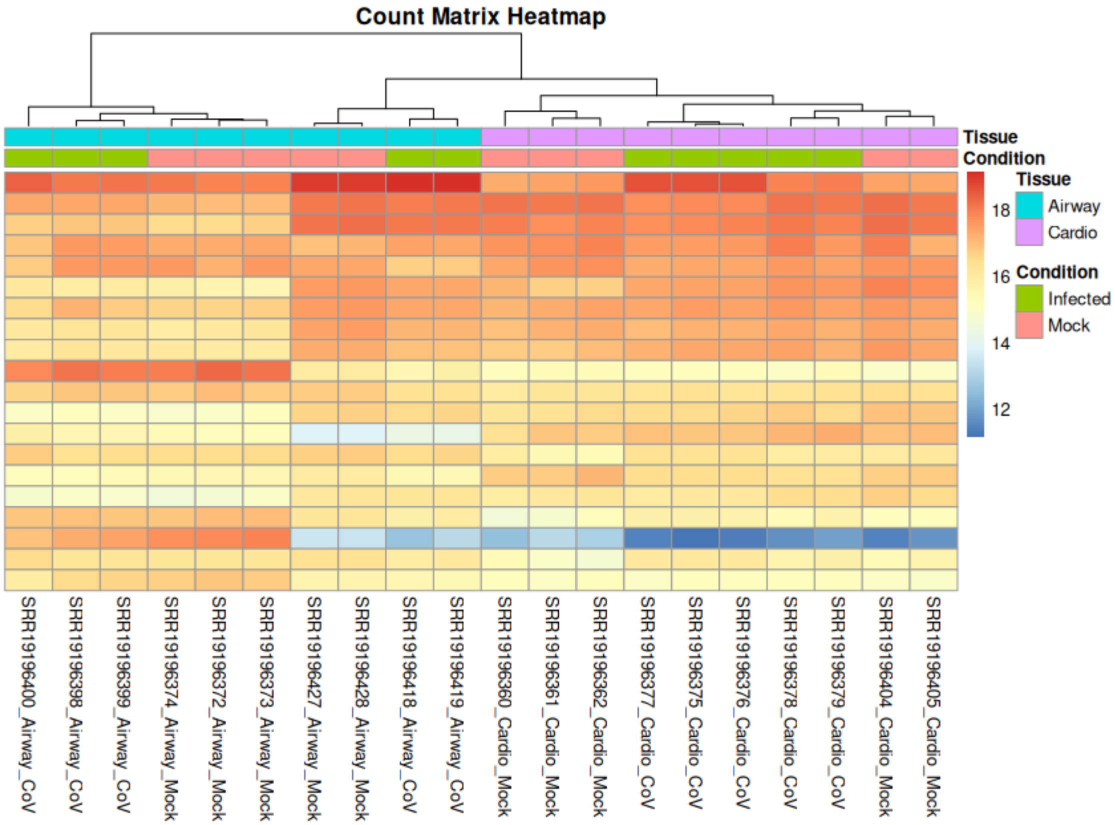


2.6 Sample Distance Heatmap Using Euclidean Distance, The heatmap illustrates pairwise Euclidean distances between samples based on their gene expression profiles. Samples are hierarchically clustered along both axes, allowing visualization of grouping patterns according to tissue types (AWOs or CMs) and conditions (Mock or Infected). The color gradient from blue (shorter distances, higher similarity) to red (longer distances, lower similarity) highlights the degree of similarity between samples. This clustering analysis provides a valuable overview of data structure, helping to identify biologically meaningful groupings, detect potential outliers, and ensure consistency across replicates within experimental groups.

We observe that all the Infected samples cluster together and all the Mock samples cluster together. Within each Tissue type, these clustering patterns make sense, suggesting that the experiment was successful.

Count Matrix Heatmap

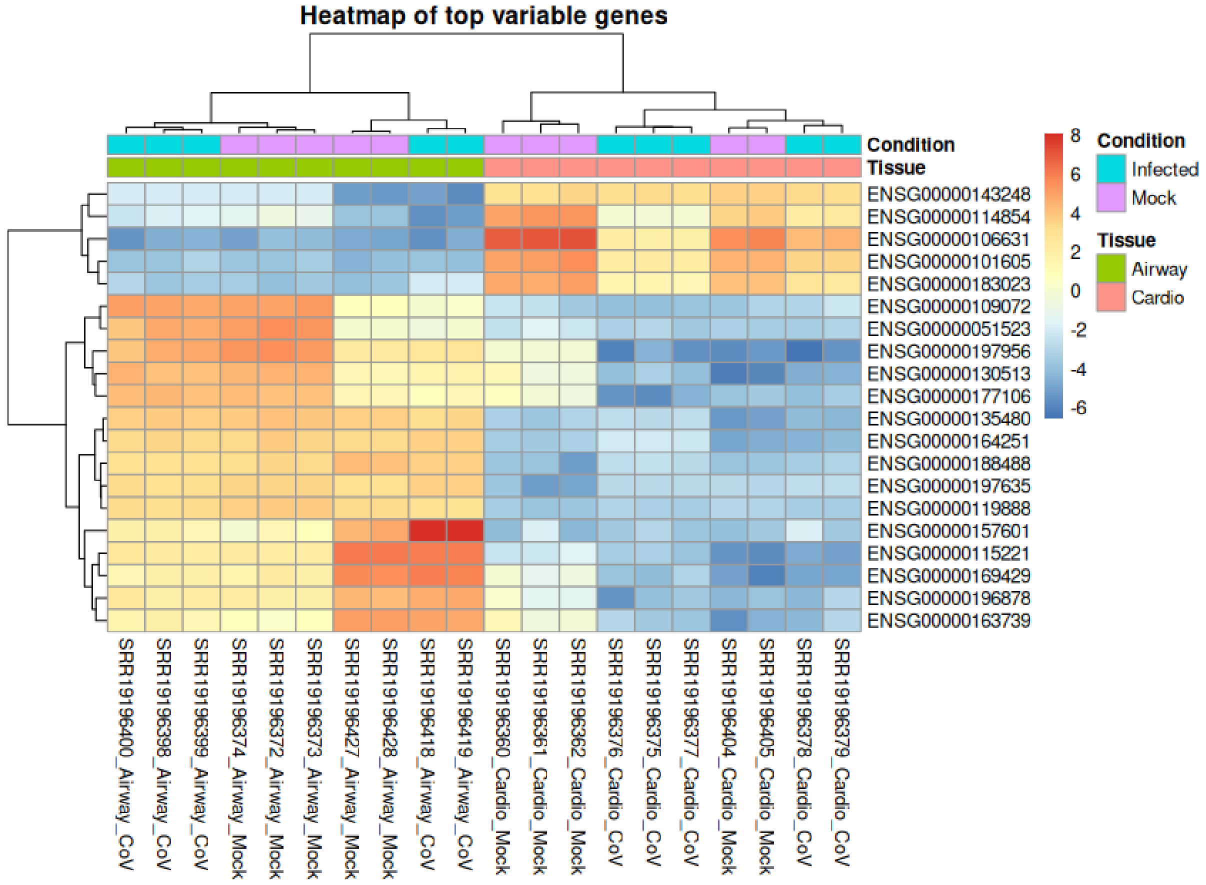
A count matrix heatmap provides a global view of raw or normalized gene expression counts across all genes and samples. It helps detect overall trends, technical artifacts, or batch effects, and ensures normalization methods yield balanced results. This heatmap is particularly valuable for inspecting data quality before downstream analyses.



2.7 Count Matrix Heatmap illustrating global gene expression counts across samples grouped by tissue type (AWOs vs. CMs) and condition (Infected vs. Mock). Rows represent genes, while columns correspond to samples. The color scale indicates expression levels, with blue representing lower expression and red representing higher expression. Hierarchical clustering of both samples and genes reveals trends in expression patterns and potential group-specific variations.

Top Variable Genes Heatmap

The following heatmap highlights the most variable genes across samples, selected based on their variance. It clusters both genes and samples to reveal patterns of expression that differ significantly across conditions or tissues. It is ideal for exploring genes driving biological differences and identifying co-expression or unique expression patterns.

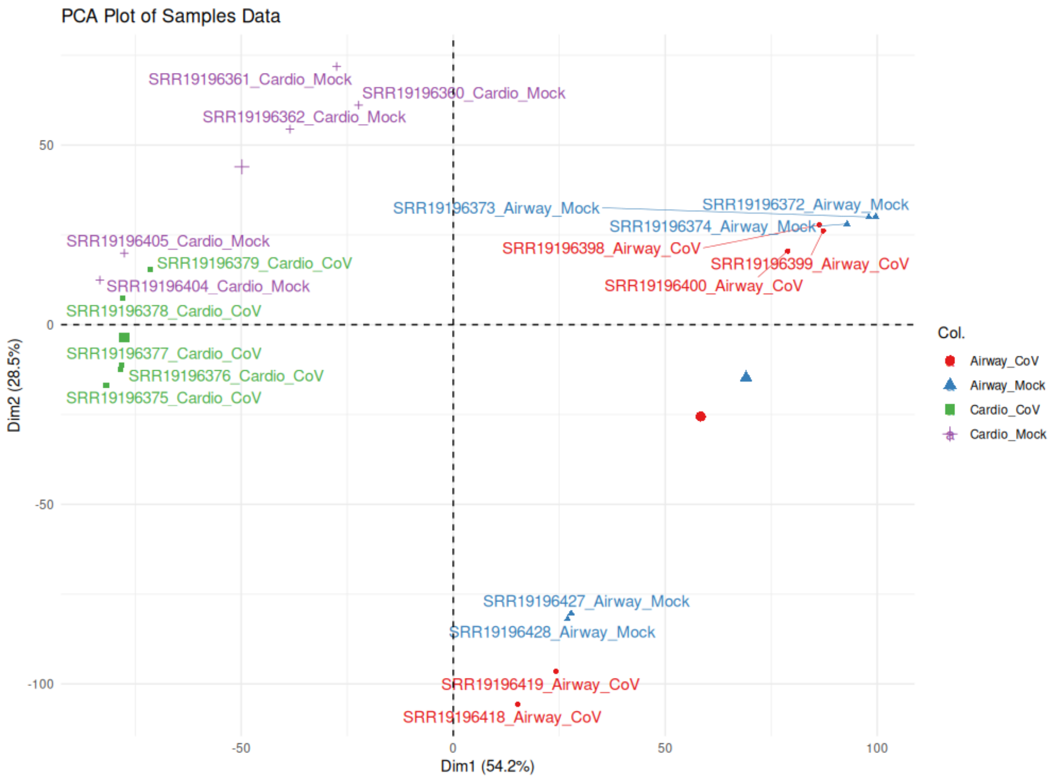


2.8 Heatmap for identifying genes driving biological differences and co-expression patterns, highlighting the most variable genes across samples, selected based on variance. Samples are clustered based on expression similarity. Genes that are associated with each cluster on the right-hand side. the genes are simply chosen based on variability across samples.

2.3.2.iii PCA Plot

Principle component analysis (PCA) allows for the unbiased identification of batch effects, which then need to be taken into account in the experimental design, and potentially reveals samples to exclude, if only one or two are "grouping" the wrong way.

In the context of differential expression analysis, we expect that most of the variability in our data is explained by the conditions we perturbed (i.e Infected vs Mock, AWOs vs CMs). If this is not the case, and distinct samples are grouped together, this indicates that a differential expression analysis is likely to be unreliable, since this analysis specifically looks at the *variance* in the expression of genes in the different libraries.



2.9 The PCA plot is shown in Figure. The horizontal axis (Dim1) is the one that captures the most variance or separation (54.2%) in samples. Dim2 on the vertical axis captures the second most variance or separation (28.5%). We see that along Dim1, the AWOs and CMs samples are clearly separated. Along Dim2, while the AWOs samples cluster closely, we see that some Airway_ mock and Airway_Cov samples are off by themselves (away from the other 6 samples in this group). This could indicate some underlying biology of AWOs Tissue or maybe it's caused by some technical factor.

2.3.3 Differential Expression Analysis

To identify tissue-specific gene expression changes in response to infection, differential expression analysis was conducted using the DESeq2 package to identify genes significantly altered between mock and infected conditions for both AWOs and CMs tissue samples.

A two-phase differential expression analysis was used. This methodological approach combined an initial comprehensive exploration of gene expression changes with subsequent refinement to focus on high-confidence differentially expressed genes (DEGs).

Initial DE analysis with consistent thresholds across tissues was generated to identify genes significantly altered between mock and infected conditions for both AWOs and CMs tissue samples.

DGE using RNA-Seq

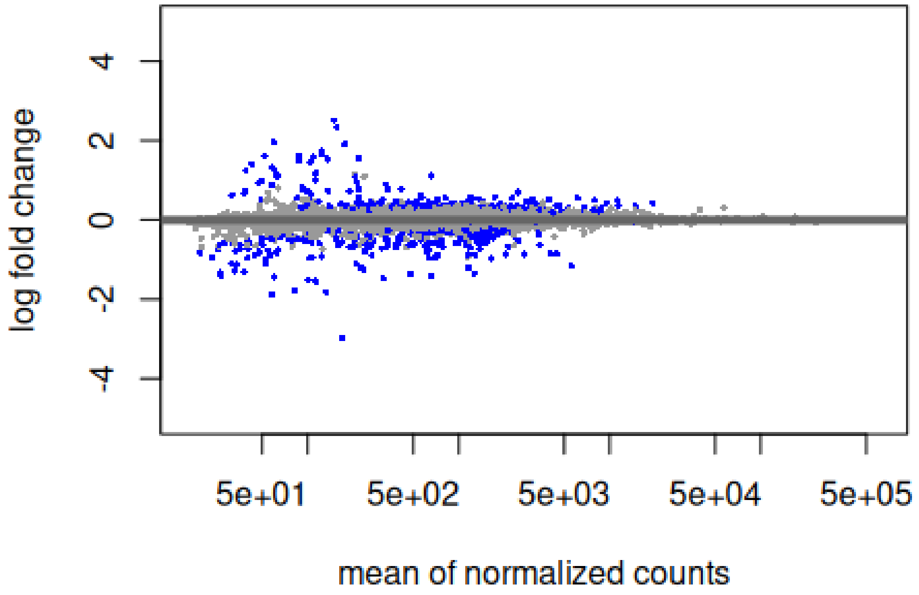
The initial phase of analysis aimed to provide a comprehensive view of differential gene expression in AWOs and CMs tissues under infection. Using an adjusted p-value threshold of <0.1 and no fold-change cutoff (LFC=0), This ensured that all genes with statistically significant differences in expression were identified regardless of the magnitude of change.

For the AWOs tissue, 263 upregulated and 210 downregulated genes were identified, reflecting localized immune responses to infection. In contrast, the CMs tissue exhibited a broader transcriptional response, with 3067 upregulated and 4130 downregulated genes, indicative of systemic physiological adjustments. This comprehensive overview formed the foundation for the subsequent focused analysis.

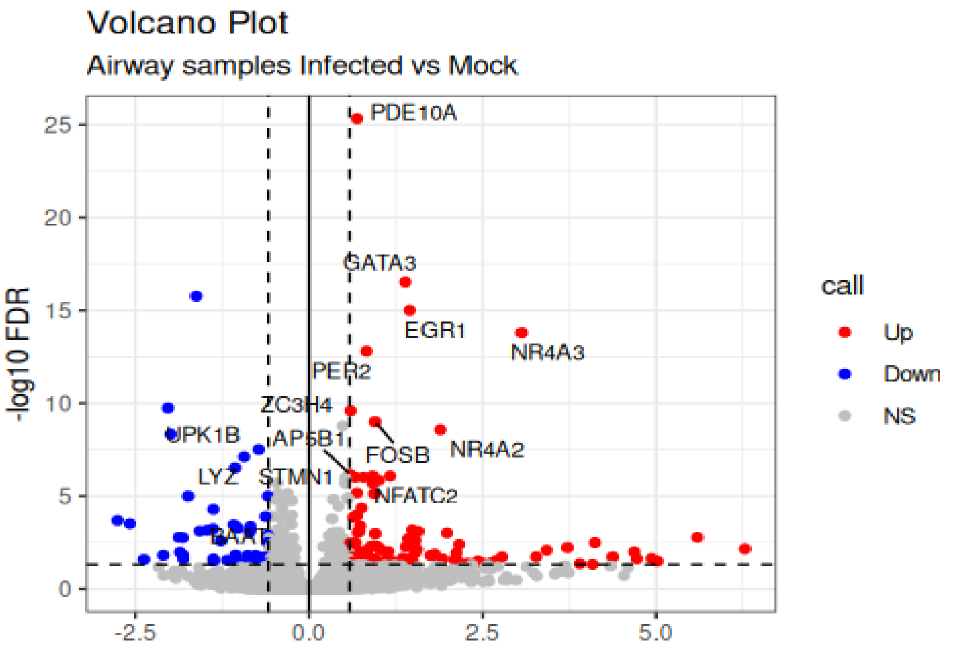
To visually summarize the results, several plots depicting the initial broader datasets were generated, highlighting the distribution of genes based on their log2 fold-change and statistical significance. These include volcano plots, MA plots, and heatmaps.

2.3.3.i AWOs Tissue:

The MA plot (Figure 2.10) for AWOs depicts the relationship between the log2 fold-change and the mean expression level (baseMean), providing a comprehensive view of transcriptional changes across the full range of expression levels. In parallel, the volcano plot (Figure 2.11) for the same dataset illustrates the distribution of genes based on their log2 fold-change and adjusted p-value, highlighting the genes that are significantly upregulated or downregulated in response to infection. Finally, the heatmap (Figure 1C) summarizes the expression profiles of the top differentially expressed genes in AWOs tissue, highlighting distinct clusters of upregulated and downregulated genes between mock and infected conditions.



2.10 MA-plot showcasing the distribution of estimated coefficients across all genes. The y-axis represents "M" (minus) - the log of the ratio, while the x-axis denotes "A" (average). Each point signifies an individual gene, with axes showing overall expression level and magnitude of difference. Significant genes are highlighted, with a fanning effect visible at low expression levels due to high relative fold-change. This plot is also known as a mean-difference plot or Bland-Altman plot.



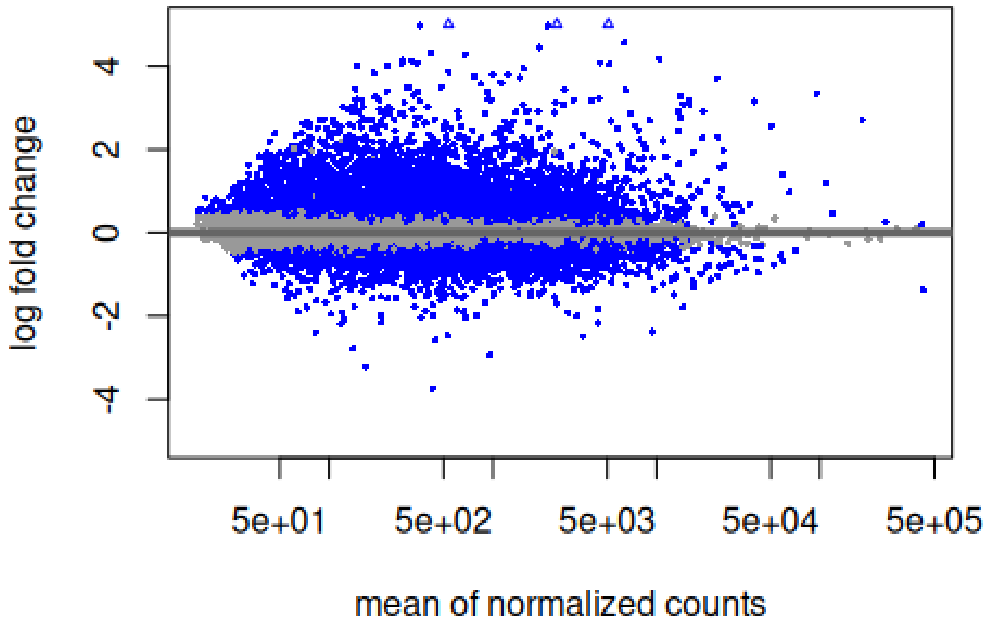
2.11 The plot displays the log2 fold change (x-axis) against the -log10 false discovery rate (FDR) (y-axis) for all genes analyzed. Genes significantly upregulated in infected airway samples are represented in red (Up), while significantly downregulated genes are shown in blue (Down). Non-significant genes (NS) are depicted in gray. Selected key DEGs, including EGR1, NR4A3, FOSB, *and* PDE10A, *are labeled for emphasis. Vertical dashed lines indicate the log2 fold change cutoff, and the horizontal dashed line represents the FDR threshold for significance.*

The AWOs tissue is the primary site of infection, where smaller changes in gene expression can have significant biological impacts. Using a broad threshold allowed for the detection of subtle yet potentially meaningful changes in response to infection.

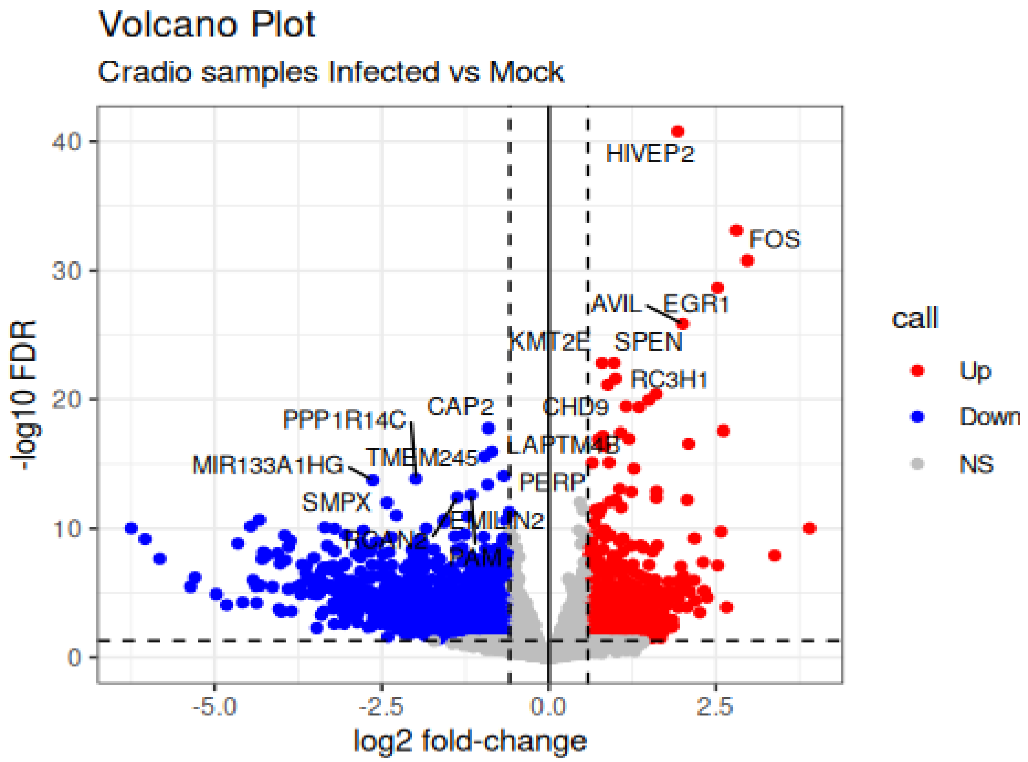
2.3.3.ii Cardiac Tissue:

Similarly, for CMs tissue, the MA plot (Figure 2B) highlights the broader range of expression changes, particularly among highly expressed genes. Complementing these visualizations, the volcano plot (Figure 2A) showcases the distribution of genes with a higher number of significant genes compared to AWOs tissue, reflecting the dynamic transcriptional profile observed in CMs tissue. The corresponding, the heatmap (Figure 2C) captures the differential expression patterns of the top genes, demonstrating clear separation between mock and infected samples.

.



2.12 The MA plot visualizes the relationship between the mean of normalized counts (x-axis, shown on a logarithmic scale) and the log2 fold change (y-axis) for each gene. Blue points represent statistically significant differentially expressed genes (adjusted p-value < 0.1), with positive log2 fold changes indicating upregulated genes and negative values indicating downregulated genes. Black points represent non-significant genes (adjusted p-value ≥ 0.1). Open triangles denote genes with log2 fold changes exceeding the y-axis limits, highlighting extreme expression changes. The plot illustrates global expression trends and highlights genes with substantial changes, aiding in the identification of candidates for further analysis.



Cardiac tissue showed broader changes due to its involvement in systemic responses to infection, such as stress, inflammation, and metabolic shifts. The larger number of DEGs reflects the complexity and diversity of pathways activated or suppressed in the heart under infectious stress.

*2.13 The volcano plot displays the log2 fold change (x-axis) against the -log10 FDR (y-axis) for each gene. Red points represent significantly upregulated genes, and blue points represent significantly downregulated genes (adjusted p-value < 0.1). Gray points denote genes that are not statistically significant (adjusted p-value ≥ 0.1). Notable upregulated genes such as **FOS**, **EGR1**, and **HIVEP2** and downregulated genes like **PPP1R14C**, **SMPX**, and **TMEM245** are labeled. The vertical dashed lines mark the threshold log2 fold change values, and the horizontal dashed line indicates the FDR significance threshold, highlighting genes with substantial expression changes between the two conditions.*

Subsequently, stringent filtering criteria were applied to narrow down the lists of differentially expressed genes (DEGs) for pathway and enrichment analysis. Due to inherent differences in tissue-specific transcriptional dynamics, sequencing depth, and baseline expression levels, it was necessary to apply slightly differing thresholds for each tissue during data filtering to ensure biologically meaningful and robust results.

For the AWOs tissue, the broader dataset presented a constrained transcriptional profile, with few genes meeting stringent thresholds due to low overall differential expression levels.

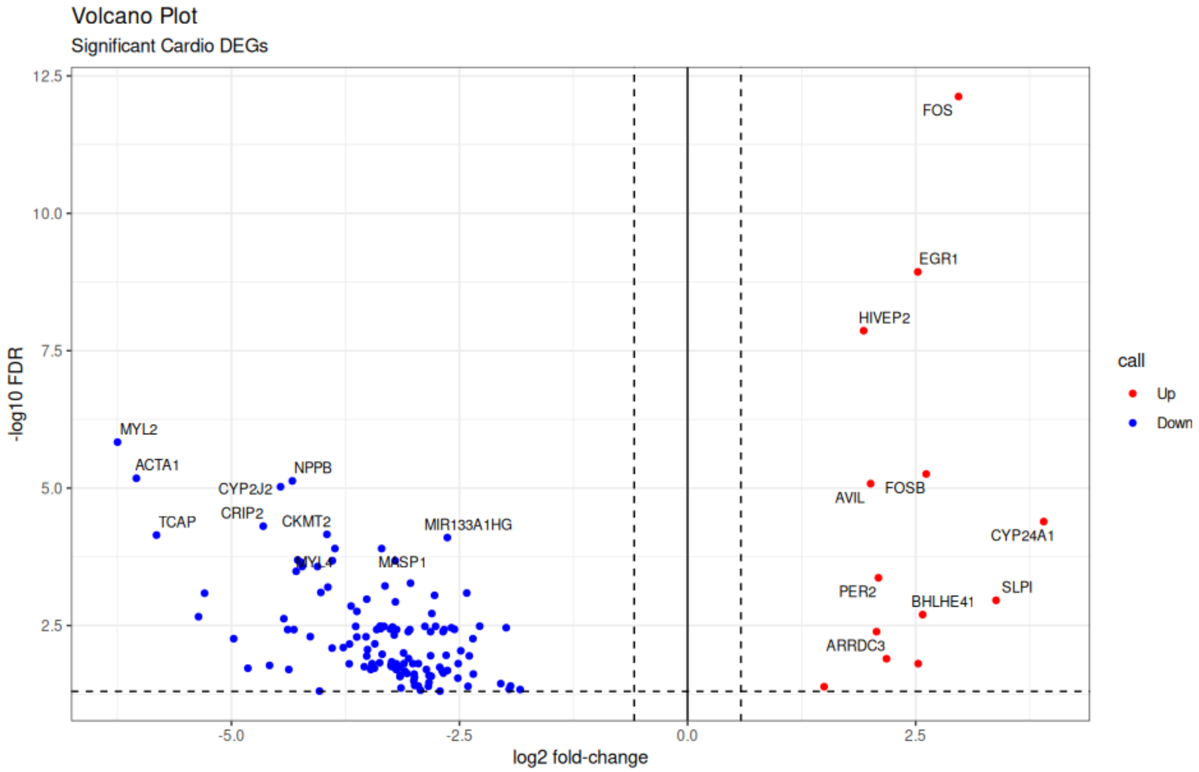
Therefore, differential expression analysis was initially performed without a fold-change threshold, identifying all genes with significant expression differences (LFC = 0) and genes with a **padj < 0.05** will be retained after the multiple-testing adjustment. Post-analysis filtering was applied to retain only those genes with substantial changes (|LFC|>1) and sufficient expression (baseMean>20). This approach was chosen to capture small but potentially meaningful changes, reflecting the tissue's critical role in the immune response to infection.

In contrast, for the CMs dataset, a moderately stringent filtering criteria was applied fold-change threshold (|LFC|>1) incorporated directly during statistical testing, a false discovery rate (FDR) cutoff of 0.05, and a minimum baseMean of 20. The CMs tissue exhibited dynamic transcriptional responses with a relatively larger number of significantly expressed genes, necessitating an explicit LFC threshold to focus on genes with substantial changes in expression. This allowed for the retention of biologically relevant genes while reducing potential false positives and reflected the focus on major pathways like muscle contraction and tissue remodeling, which are biologically central in the context of infection-induced stress.

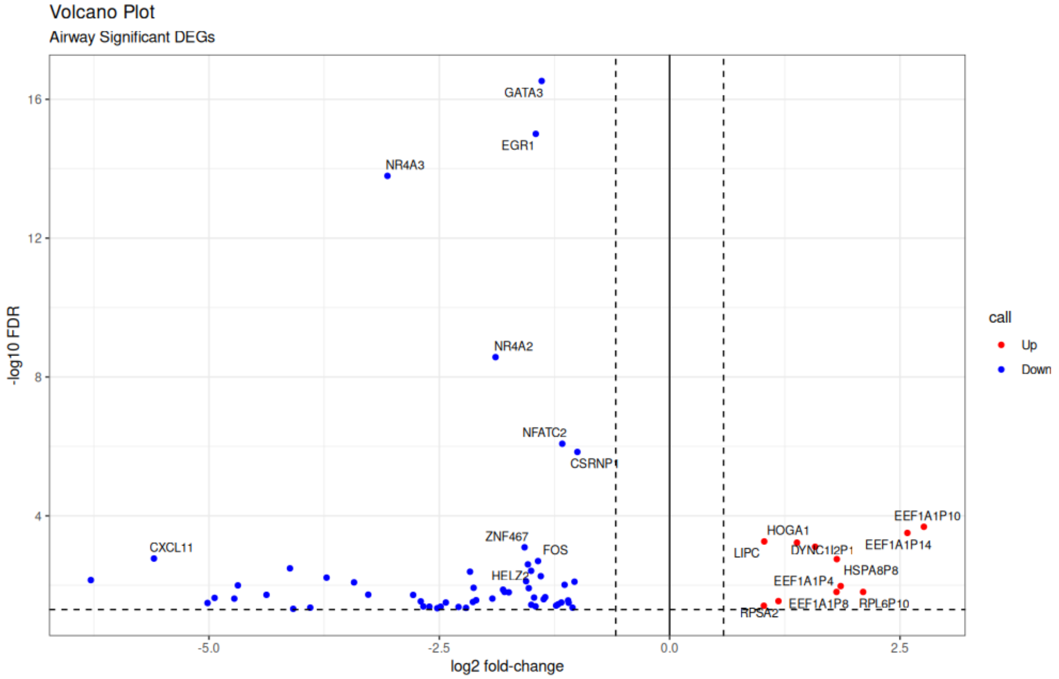
This strategy of adapting thresholds ensures that the analysis captures meaningful biological insights while accounting for tissue-specific variability and maximizing the number of shared significant genes between the two tissues (Love, Huber, and Anders 2014).

Volcano plots, figures (2.14, 2.15) were regenerated for the two datasets to visualize the distribution and significance of differentially expressed genes in the AWOs and CMs tissues, highlighting the refined set of significant genes identified through the second, more stringent analytical approach.

DGE using RNA-Seq

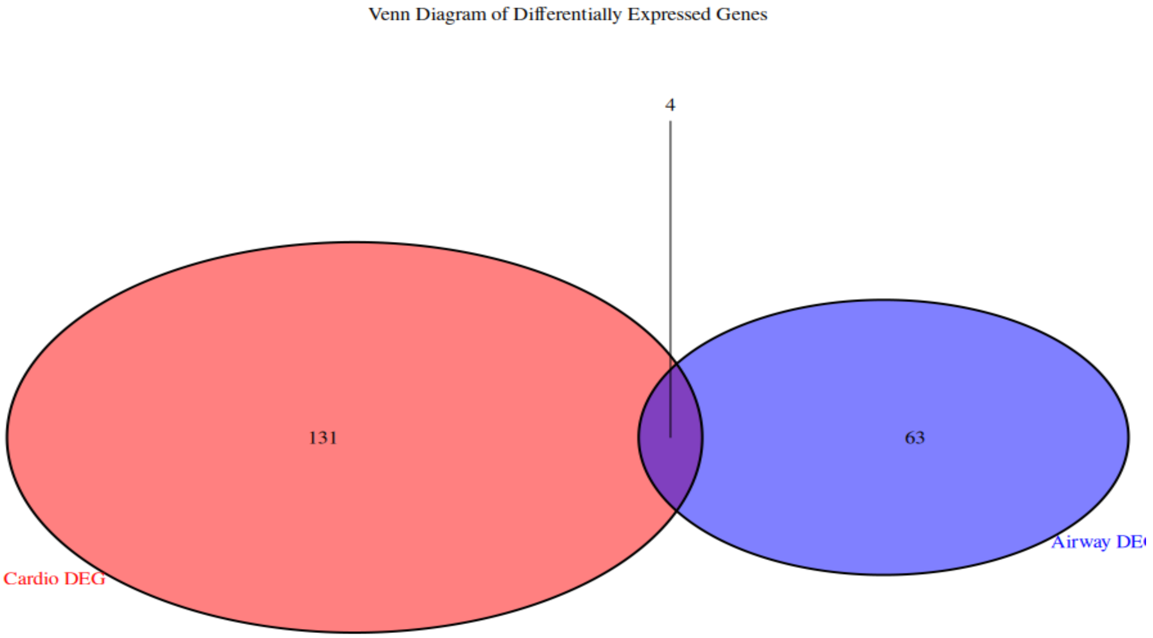


*2.14 Volcano plot of differential gene expression in CMs samples. Significantly upregulated (red) and downregulated (blue) genes are plotted based on log2 fold change (x-axis) and -log10 FDR (y-axis). Notable genes include **FOS**, **EGR1**, **HIVEP2** (upregulated), and **ACTA1**, **NPPB**, **MYL2** (downregulated). Thresholds are marked by dashed lines.*



*2.15 Volcano plot of differential gene expression in AWOs samples. Significantly upregulated (red) and downregulated (blue) genes are plotted based on log2 fold change (x-axis) and -log10 FDR (y-axis). Notable genes include **GATA3**, **EGR1** (upregulated), and **LIPC**, **HOGA1** (downregulated). Thresholds are marked by dashed lines.*

A Venn diagram was utilized to identify common differentially expressed genes (DEGs) between the AWOs and CMs datasets, following the application of stringent filtering criteria. The overlap between the two datasets resulted in the identification of **four common genes**, which likely represent core transcriptional responses conserved across both tissues. These genes were prioritized for downstream analysis, including pathway and enrichment studies, to explore shared mechanisms and tissue-specific functional differences during SARS-CoV-2 infection.



2.16 ***Venn diagram of differentially expressed genes (DEGs) in cardiac and airway samples.***
The diagram shows 131 unique DEGs in CMs samples, 63 unique DEGs in AWOs samples, and 4 overlapping DEGs shared between the two conditions.

2.3.4 Enrichment Analysis- Gene Ontology (GO) and pathway analysis.

creating lists of DE genes is not the final step of the analysis; further biological insight into an experimental system can be gained by looking at the expression changes of sets of genes. To explore the functional relevance of differentially expressed genes (DEGs) and determine whether a specific category of terms is over-represented in this analysis, Gene Ontology (GO) enrichment analysis was performed using the clusterProfiler package (Yu et al. 2012). The analysis focused on Biological Processes (BP), utilizing Entrez gene identifiers annotated with the org.Hs.eg.db database.

To determine whether a specific category of terms is over-represented in this analysis, and identify tissue-specific biological processes and pathways impacted by infection, a GO-term and pathway enrichment analysis of all differentially expressed genes was performed to identify enriched GO-terms and pathways in this list.

Gene Ontology (GO) and pathway enrichment analyses provide essential insights into the biological functions and pathways represented by differentially expressed genes (DEGs). The Gene Ontology (GO) provides a framework and set of concepts for describing the functions of gene products from all organisms (Thomas 2017).

Functional enrichment analysis was performed using both overlapping and unique DEGs from each tissue identified through the **venn diagram comparison** to identify enriched GO-terms and pathways in this analysis

For this study, **ClusterProfiler** was employed to investigate both Gene Ontology (GO) terms and KEGG pathways. ClusterProfiler, integrated with the org.Hs.eg.db database, was particularly employed for its advanced visualization capabilities and ability to identify enriched GO terms across the Biological Process (BP), Molecular Function (MF), and Cellular Component (CC) categories. Additionally, it was used to map DEGs to KEGG pathways, facilitating a comprehensive understanding of the biological pathways implicated in the study.

The input for these analyses consisted of significant DEGs identified through the differential expression analysis pipeline. We removed duplicated genes and the values lacked specific gene symbols from whole datasets.

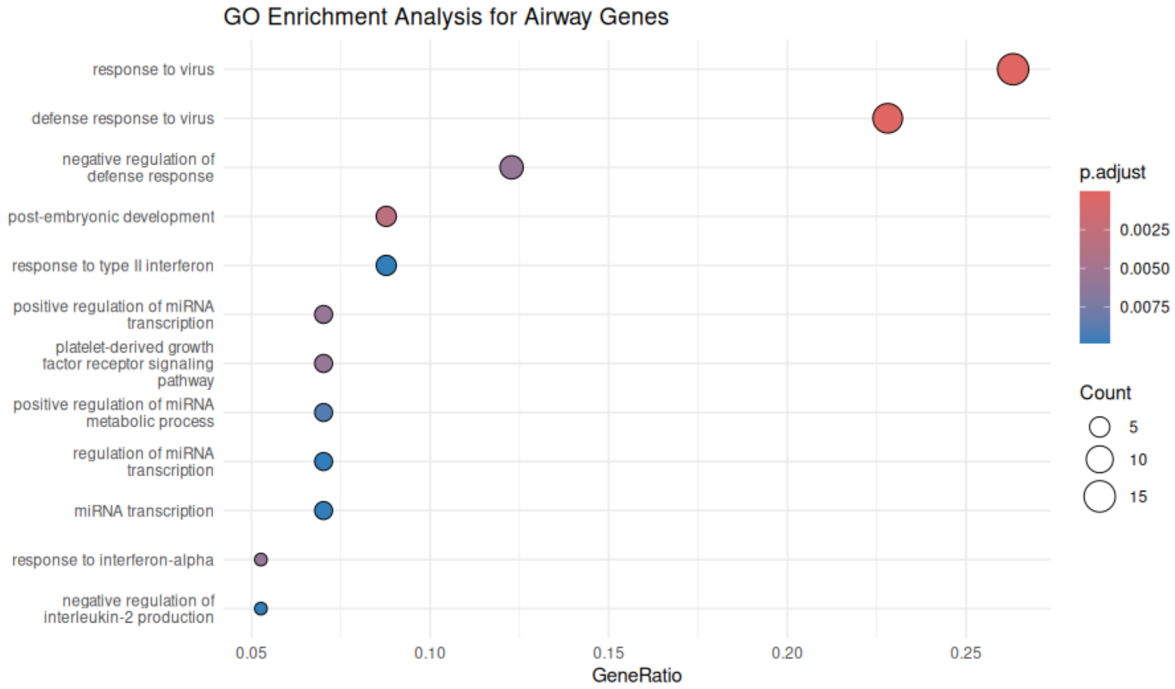
The **enrichGO** function was used to identify enriched GO terms within the Biological Process (BP) category for each tissue. GO terms were considered significant if they met a Benjamini-Hochberg adjusted p-value threshold of 0.01 and a false discovery rate (q-value) threshold of 0.05. The results were visualized using a dot plot, highlighting the top 20 enriched biological processes, ranked by significance and gene ratio.

Additionally, the **gseKEGG** function was utilized for gene set enrichment analysis (GSEA) of KEGG pathways, specifically targeting enriched pathways in the CMs and AWOs tissues. The parameters for this analysis included a p-value cutoff of 0.05, a minimum gene set size of 3, and a maximum gene set size tailored to the DEG list size (136 for CMs and 67 for AWOs datasets).

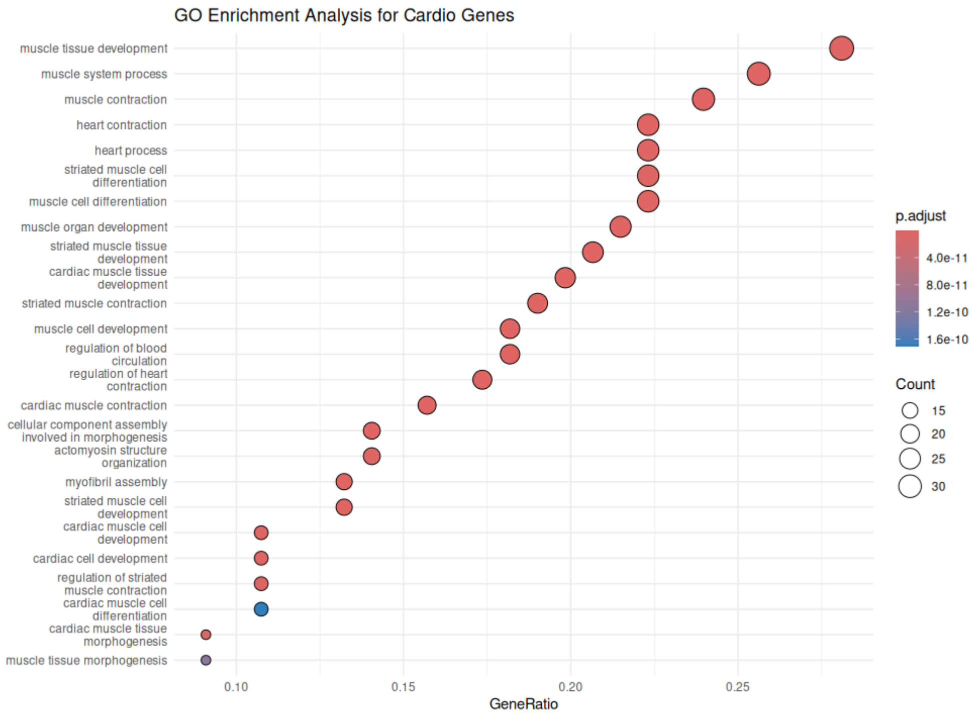
The analysis revealed ranked lists of enriched GO terms and KEGG pathways including key biological pathways potentially associated with the AWOs-specific and cardio-specific responses to infection. The significant GO terms enriched in this dataset provide insights into critical processes such as immune response regulation, cellular signaling pathways, and inflammation.

To focus on the most biologically relevant pathways, we applied a Fold Enrichment threshold of 2 to the GO enrichment results. This threshold ensures that the selected terms are at least twice as enriched in the test set compared to the background, highlighting the most significant biological processes. After filtering, 13 terms remained for the AWOs dataset and 21 terms for the CMs dataset, which were subsequently visualized to identify tissue-specific and shared pathways.

The dot plot (Figure X) for the AWOs samples and (Figure X) summarize the enriched GO terms, offering a visual overview of the functional landscape of the DEGs in both tissues.



2.17 GO enrichment analysis of differentially expressed airway genes. Key processes include antiviral responses such as "response to virus" and "defense response to virus," as well as pathways related to interferon signaling and miRNA transcription regulation. Dot size represents the gene count, and the color gradient indicates the adjusted p-value, with darker red indicating higher significance.



2.18 GO enrichment analysis of differentially expressed cardiac genes. The analysis highlights key biological processes and pathways, such as muscle contraction, heart contraction, and cardiac muscle tissue development, involved in cardiac function. The dot size represents the gene count, and the color gradient indicates the adjusted p-value, with darker red signifying higher significance.

3 Results

3.1 Differential Expression in Organoids

Following differential expression analysis using stringent filtering, genes lacking annotations were excluded. **131 significant DEGs** were retained for CMs tissue, and **63 significant DEGs** were retained for the AWOs tissue.

The results of the differential expression analysis reveal significant differences in the number and nature of differentially expressed genes (DEGs) between CMs and AWOs tissues following infection. This disparity is expected, as the baseline transcriptome of these tissues differs markedly. Cardiac tissue predominantly expresses genes essential for contraction, metabolism, and structural repair, whereas AWOs tissue is characterized by genes involved in cilia function, mucus production, and barrier integrity. The higher number of DEGs in CMs tissue may be attributed to its greater metabolic and structural complexity, which renders it more susceptible to transcriptional perturbations by viral infection. Additionally, CMs tissue exhibits heightened sensitivity to inflammation, often responding to even low levels of insult with systemic changes in gene expression.

3.1.1 DEG Overview for AWOs and Cardiac Tissues Under Infection Conditions

3.1.1.i AWOs-Specific DEGs

Differential expression analysis of AWOs tissues identified 67 differentially expressed genes (DEGs), comprising 11 upregulated genes and 56 downregulated genes. Among the upregulated genes, the top five were **NR4A3**, **EGR1**, **GATA3**, **IFIT2**, and **IRF1**, with log2 fold changes (LFC) of 3.06, 1.45, 1.39, 1.28, and 1.18, respectively. For the downregulated genes, the most significant included **ITGB1P1** (LFC = -2.58), **PCBP2P2** (LFC = -2.03), **MORF4L1P1** (LFC = -1.99), **FTH1P20** (LFC = -1.82), and **LIPC** (LFC = -1.75).

The upregulated genes predominantly represented transcription factors and immune-related regulators. These included genes such as **GATA3**, **EGR1**, and **NR4A3**, which are involved in transcriptional activation and immune signaling, as well as **IRF1**, **SOCS1**, and **HERC5**, which contribute to immune modulation and antiviral responses. A significant portion of the upregulated genes were interferon-stimulated genes (ISGs), reflecting an active antiviral response. Key ISGs included **IFIT1**, **IFIT2**, **IFIT3**, **IFI16**, and **IFI27**, which are known for inhibiting viral replication. Other ISGs such as **MX2** and **BST2** also contributed to antiviral activity through diverse mechanisms. Additionally, genes involved in cytokine and chemokine signaling, such as **CXCL11** and **TNFSF10 (TRAIL)**, were upregulated, indicative of heightened immune cell recruitment and apoptosis. Genes associated with cellular stress and apoptosis, including **TIPARP**, **TXNIP**, and **ZFP36**, were also elevated, suggesting cellular adaptation to stress and tissue remodeling. Furthermore, several upregulated genes related to chromatin organization and cell structure, such as **H2AC15**, **H2BC8**, **MDGA1**, and **PCDH17**, pointed to changes in chromatin remodeling and structural integrity.

In contrast, the downregulated genes were largely dominated by pseudogenes and genes involved in translation and protein synthesis. For example, multiple downregulated DEGs, including **EEF1A1P10**, **EEF1A1P14**, **EEF1A1P4**, and **EEF1A1P8**, were pseudogenes of **EEF1A1**, which is implicated in elongation during translation. Additionally, downregulated genes such as **LIPC** and **FTH1P20** suggested a reduction in lipid metabolism and iron homeostasis, respectively. Many of the downregulated pseudogenes, including **ITGB1P1**, **PCBP2P2**, and **MORF4L1P1**, had less clearly defined roles in AWOs responses but were markedly suppressed in this dataset.

Overall, the AWOs-specific DEGs demonstrated distinct patterns of transcriptional regulation, with upregulated genes reflecting immune activation and antiviral responses, while downregulated genes indicated a decline in translation-related processes and metabolic activity.

3.1.1.ii CMs-Specific DEGs

In the analysis of CMs tissue, a total of 142 differentially expressed genes (DEGs) were identified, consisting of 16 upregulated genes and 126 downregulated genes. Among the upregulated genes, **FOS** exhibited the highest fold change (log2FoldChange, LFC = 2.20), followed by **EGR1** (LFC = 1.90) and **HIVEP2** (LFC = 1.80). The most significant statistics for these genes include a base mean expression of 966 for **FOS**, with a p-value of 4.57e-10 and an adjusted p-value (padj) of 0.00000553, indicating strong statistical significance in its upregulation. Similarly, **EGR1** had a base mean of 3551, with a p-value of 6.19e-7 and padj of 0.00107.

Conversely, the most significantly downregulated genes included **MYL2** (LFC = -5.10), **NPPB** (LFC = -4.70), and **ACTA1** (LFC = -4.50). Notably, **NPPB** had a base mean expression of 10411, with a p-value of 5.14e-7 and padj of 0.00107, while **ACTA1** showed a base mean of 5089, with a p-value of 9.16e-8 and padj of 0.000277.

A large proportion of downregulated genes are associated with CMs structure and contractile function, reflecting potential tissue remodeling or loss of normal CMs activity during infection.

A closer examination of the downregulated DEGs revealed several notable patterns. A significant portion of these genes is linked to muscle contraction and structural integrity, with key examples including **MYL2**, **MYL3**, **MYL7**, **MYH6**, **MYH7**, and **ACTC1**, which encode essential myosin and actin isoforms. The downregulation of these structural proteins may contribute to impaired contractility in CMs tissues.

Additionally, genes that play critical roles in calcium signaling and regulation, such as **CASQ1**, **PLN**, and **SLC8A1**, were found to be downregulated. These genes are vital for calcium storage, release, and exchange within CMs tissues. The downregulation of these calcium handling genes could lead to dysregulation of intracellular calcium levels, potentially resulting in arrhythmias or heart failure.

Furthermore, downregulated genes associated with energy metabolism and mitochondrial function were identified, including **CKMT2** (LFC = -3.95; base mean = 1001; p-value = 6.02e-6; padj = 0.00662), **ATP5F1D**, and **NDUFB7**, alongside **ALDOA** and **ECH1**, which are involved in glycolysis and fatty acid metabolism. The reduction in energy metabolism-related genes suggests a shift away from efficient ATP production, which is crucial for maintaining CMs function.

The analysis also highlighted the downregulation of cytoskeletal and stress-response proteins; notable examples include **CRYAB**, **DES**, and **HSPB7**, which are crucial for maintaining cytoskeletal integrity and responding to cellular stress. Moreover, several transcription factors and signaling pathway regulators were observed to be downregulated, including **NKX2-5**, **IRX4**, and the cell cycle regulator **CDKN1A**. This suggests a broader impact on gene regulatory networks within the CMs tissue.

Immune-related genes such as **MIF**, **C7**, and **MASP1** were also found to be downregulated, indicating potential alterations in immune system modulation.

On the other hand, the upregulated DEGs predominantly featured immediate early response genes such as **FOS** (base mean = 966; p-value = 4.57e-10; padj = 0.00000553), **FOSB**, **EGR1**, and **EGR**2, which are rapidly activated in response to stress and various cellular signals. Additionally, transcriptional regulators and signal transduction genes like **HIVEP2** (LFC = 1.80) suggest a shift towards transcriptional reprogramming in response to stressors.

The upregulation of inflammatory and immune modulation genes such as **SLPI** and **NFATC2** further indicates a potential adaptive response to immune challenges. Moreover, genes involved in metabolic processes and stress responses were also upregulated; for instance, **CYP24A1** (LFC = 3.90; base mean = 428; p-value = 3.96e-6; padj = 0.00479) and **ARRDC3** play roles in regulating these processes. **AVI**L, a gene associated with actin filament dynamics, was among those that showed increased expression. Notably, **H2BC8**, a variant histone gene, may contribute to chromatin remodeling during infection.

Lastly, the upregulation of **PER2**, a core circadian rhythm gene, may indicate infection-related alterations in circadian rhythm, potentially affecting metabolic and immune processes.

3.1.1.iii Shared DEGs

Four genes (**EGR1**, **NFATC2**, **FOS**, and **H2BC8**) were identified as shared DEGs between AWOs and CMs tissues. All four genes were upregulated in both tissues but exhibited tissue-specific variations in magnitude:

- **EGR1** had a slightly higher expression in CMs tissues (*LFC = 1.90*) compared to AWOs tissues (*LFC = 1.45*).
- **FOS** showed a more pronounced upregulation in CMs tissues (*LFC = 2.20*) compared to AWOs tissues (*LFC = 1.43*).
- **H2BC8** and **NFATC2** displayed consistent upregulation across both tissues with subtle differences in fold changes.

The identification of these common DEGs suggests a conserved host response to SARS-CoV-2 infection across different tissue types. The statistical analysis of these genes indicates that while their expression is elevated in both AWOs and CMs tissues, the degree of upregulation varies, potentially reflecting distinct tissue responses to the viral infection.

The common DEGs identified between AWOs and CMs tissues ("**EGR1**," "**NFATC2**," "**FOS**", and "**H2BC8**") point to a conserved host response to SARS-CoV-2 infection. These genes are heavily involved in transcriptional regulation and stress-response pathways, suggesting a central role in modulating the cellular response to viral invasion. Specifically:

EGR1 (Early Growth Response 1): A transcription factor that is rapidly induced in response to cellular stress. Its role in regulating genes involved in inflammation and apoptosis highlights its importance in the initial response to SARS-CoV-2 infection. Dysregulation of **EGR1** could exacerbate inflammatory responses, contributing to tissue damage.

FOS (Fos Proto-Oncogene, AP-1 Transcription Factor): As part of the AP-1 complex, **FOS** regulates processes such as cell proliferation and differentiation. Its activation may reflect the cellular stress induced by SARS-CoV-2 and the need for reparative mechanisms in damaged tissues.

NFATC2 (Nuclear Factor of Activated T-Cells 2): This gene's involvement in immune regulation suggests its role in coordinating the antiviral response. It may also contribute to inflammatory cascades, potentially linking to severe manifestations such as cytokine storms observed in some COVID-19 patients.

H2BC8 (**H2B** Clustered Histone 8): As a histone protein, **H2BC8'**s role in nucleosome assembly and chromatin organization underscores its involvement in regulating gene expression during infection. Its association with the innate immune response in mucosa may reflect the host's attempt to counteract viral replication.

The DEG profiles highlighted distinct transcriptional responses in AWOs and CMs tissues to infection. While AWOs tissues predominantly exhibited immune-related and transcriptional gene changes, CMs tissues were characterized by extensive downregulation of genes related to structural and functional maintenance.

3.2 Gene Ontology (GO)-Term Enrichment Analysis

3.2.1 Identifying biological pathways and GO terms enriched in DEGs

To investigate the pathways driving these distinctions, we performed gene set enrichment analyses[9] (GSEA) on the genes differentially expressed (DE) between SARS-CoV-2 infected and mock samples of both tissues.

3.2.1.i AWOs DEGs:

Gene Ontology (GO) enrichment analysis and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis were performed to elucidate the biological processes and pathways significantly enriched in the differentially expressed genes (DEGs) of AWOs tissues in response to infection. The results provided insights into both the upregulated and downregulated genes, reflecting complex molecular changes in AWOs tissues (Rodrigues, Costa, and Henriques 2022).

3.2.1.i.a GO Enrichment Analysis for Upregulated DEGs

The upregulated DEGs were predominantly associated with immune-related processes and antiviral responses. Among the significantly enriched biological processes, the **"response to interferon-alpha"** (GO:0035455; adjusted $p = 0.0042$) and **"response to interferon-beta"** (GO:0035456; adjusted $p = 0.0097$) were highlighted, with enrichment factors of 59.4 and 31.4, respectively. These processes are supported by core genes such as **2625**, **684**, and **4600**, which are key mediators of antiviral defense. Similarly, the **"defense response to virus"** (GO:0051607; adjusted $p = 1.4 \times 10^{-9}$) and **"response to virus"** (GO:0009615; adjusted $p = 3.4 \times 10^{-10}$) exhibited strong enrichment, with high gene participation (13 and 15 genes, respectively), indicating the activation of broad antiviral programs.

Additional enriched GO terms included the **"platelet-derived growth factor receptor signaling pathway"** (GO:0048008; adjusted $p = 0.0042$), which may play a role in tissue remodeling and inflammation, and processes associated with microRNA (miRNA) regulation, such as **"positive regulation of miRNA transcription"** (GO:1902895; adjusted $p = 0.0042$) and **"miRNA transcription"**

(GO:0061614; adjusted *p* = 0.0071). These findings suggest a strong interplay between immune signaling and post-transcriptional regulation in the AWOs tissue response to infection.

3.2.1.i.b GO Enrichment Analysis for Downregulated DEGs

Downregulated DEGs were enriched in metabolic and biosynthetic processes, highlighting the suppression of key pathways in AWOs tissues during infection. Notably, terms related to **"monocarboxylic acid biosynthetic process"** (GO:0072330; adjusted *p* = 0.0228) and **"carboxylic acid biosynthetic process"** (GO:0046394; adjusted *p* = 0.0228) were identified, both with significant enrichment factors (41.9 and 28.3, respectively), indicating broad downregulation of organic acid biosynthesis. Core genes involved in these pathways included **3990** and **112817**.

Processes related to lipid metabolism were also suppressed, as evidenced by the enrichment of **"triglyceride-rich lipoprotein particle remodeling"** (GO:0034370; adjusted *p* = 0.0228) and **"very-low-density lipoprotein particle remodeling"** (GO:0034372; adjusted *p* = 0.0228), each involving key genes such as **3990**. Additionally, pathways like **"aldehyde catabolic process"** (GO:0046185; adjusted *p* = 0.0228) and **"amino sugar catabolic process"** (GO:0046348; adjusted *p* = 0.0228) were suppressed, suggesting impaired breakdown of small molecules and lipids.

Further, structural and functional processes such as **"ribosomal small subunit assembly"** (GO:0000028; adjusted *p* = 0.0228) and **"respiratory chain complex IV assembly"** (GO:0008535; adjusted *p* = 0.0253) were also enriched, with core genes like **388524** and **51241**, reflecting potential disruptions in protein synthesis and mitochondrial respiratory function.

3.2.1.i.c KEGG Pathway Enrichment Analysis.

KEGG pathway analysis of upregulated DEGs revealed modest enrichment in immune-related pathways. For example, the pathway **"Human T-cell leukemia virus 1 infection"** (hsa05166; NES = 1.79, adjusted *p* = 0.17) was associated with immune modulation and involved core genes such as **2353**, **4773**, and **7538**. Additional pathways, such as **"Th1 and Th2 cell differentiation"** (hsa04658; NES = 1.61, adjusted *p* = 0.17) and **"Kaposi sarcoma-associated herpesvirus infection"** (hsa05167; NES = 1.61, adjusted *p* = 0.17), were identified, albeit with borderline statistical significance. These results reflect immune signaling activation in AWOs tissues in response to infection.

For downregulated DEGs, significant suppression was observed in two metabolic pathways: **"Arginine and proline metabolism"** (hsa00330; NES = -1.20, adjusted *p* = 0.0008) and **"Glyoxylate and dicarboxylate metabolism"** (hsa00630; NES = -1.20, adjusted *p* = 0.0008). Both pathways had strong enrichment scores, with **3990** and **112817** identified as key genes. These results underscore the metabolic reprogramming of AWOs tissues during infection, with downregulation of amino acid and organic acid metabolism pathways.

These findings illustrate a dichotomy in the biological response of AWOs tissues to infection, with upregulated DEGs predominantly driving immune activation and viral defense, while downregulated DEGs reflect metabolic and biosynthetic suppression, potentially impacting the tissue's ability to maintain homeostasis during infection.

3.2.1.ii CMs DEGs:

Gene Ontology (GO) and KEGG pathway enrichment analyses were performed to identify biological processes and pathways enriched in the differentially expressed genes (DEGs) of CMs tissues. These analyses were conducted separately for upregulated and downregulated genes to better understand the distinct biological responses of CMs tissues to infection.

3.2.1.ii.a GO Enrichment Analysis for Upregulated DEGs

The GO analysis of upregulated DEGs revealed four significantly enriched biological processes. Among these, **"circadian regulation of gene expression"** (GO:0032922) was the most enriched, with a GeneRatio of 3/13 and a fold enrichment of 62.3. This result highlights the potential involvement of circadian rhythm-related genes in the CMs response to infection. Another enriched process was **"skeletal muscle cell differentiation"** (GO:0035914), with a GeneRatio of 3/13 and a fold enrichment of 56.6, emphasizing the role of genes associated with muscle differentiation and tissue adaptation in response to infection.

The GO term **"cellular response to calcium ion"** (GO:0071277) was also significantly enriched, with a GeneRatio of 3/13 and a fold enrichment of 50.1. This result underscores the importance of calcium signaling in CMs tissue, particularly in regulating processes critical for cellular stress responses. Finally, the term **"response to corticosterone"** (GO:0051412) was enriched, with a fold enrichment of 170.9, suggesting potential activation of stress hormone-related pathways. These findings collectively indicate that the upregulated DEGs in CMs tissues are involved in processes related to stress response, muscle differentiation, and calcium ion regulation, which may contribute to tissue remodeling during infection.

3.2.1.ii.b GO Enrichment Analysis for Downregulated DEGs

The enrichment analysis of downregulated DEGs revealed a much larger number of significantly enriched biological processes, particularly those related to CMs structure, muscle contraction, and calcium signaling. The GO term **"myofibril assembly"** (GO:0030239) was the most enriched, with 16 DEGs and a fold enrichment of 36.34. Closely related terms, such as **"striated muscle cell development"** (GO:0055002) and **"CMs myofibril assembly"** (GO:0055003), were also highly enriched, each with 16 and 6 DEGs, respectively. These terms indicate significant suppression of genes involved in the structural integrity and development of CMs muscle, suggesting compromised myofibrillar organization in response to infection.

Processes directly associated with CMs contraction were notably enriched among downregulated DEGs. For example, **"regulation of the force of heart contraction"** (GO:0002026) had a GeneRatio of 8/108 and a fold enrichment of 53.81, highlighting significant transcriptional suppression in genes responsible for modulating the contractile force of CMs tissue. Similarly, **"actin-myosin filament sliding"** (GO:0033275), with a fold enrichment of 54.65, underscores the impact of infection on the basic molecular machinery underlying CMs muscle contraction.

Calcium signaling pathways, critical for excitation-contraction coupling, were also significantly enriched among downregulated genes. The GO term **"regulation of cardio muscle**

contraction by calcium ion signaling" (GO:0010882), with a GeneRatio of 5/108 and a fold enrichment of 33.63, points to the transcriptional suppression of genes involved in calcium regulation. Similarly, **"release of sequestered calcium ion into cytosol by sarcoplasmic reticulum"** (GO:0014808) and **"regulation of release of sequestered calcium ion into cytosol by sarcoplasmic reticulum"** (GO:0010880) were enriched, with fold enrichments of 24.29 and 28.20, respectively. These results suggest that calcium homeostasis is significantly impacted in CMs tissue during infection.

Developmental and morphogenetic processes were also prominently represented. The terms **"Cardio muscle tissue morphogenesis"** (GO:0055008) and **"ventricular cardio muscle tissue development"** (GO:0003229) were enriched, with 11 and 10 DEGs, respectively, reflecting transcriptional suppression of genes involved in CMs tissue development and organization. Additionally, the term **"response to muscle stretch"** (GO:0035994), with a fold enrichment of 27.98, highlights alterations in mechanical stress responses in the CMs tissue.

3.2.1.ii.c KEGG Pathway Enrichment Analysis

The KEGG pathway enrichment analysis for upregulated and downregulated DEGs in CMs tissues provided distinct insights into the metabolic and signaling alterations induced by infection.

For the **upregulated DEGs**, two pathways were identified, though neither reached the threshold for statistical significance. The pathway **"Parathyroid hormone synthesis, secretion, and action"** (hsa04928) had a normalized enrichment score (NES) of 1.25 and included the upregulated genes **1591** and **2353**. Similarly, the pathway **"Osteoclast differentiation"** (hsa04380), which had an NES of 1.17, was associated with the genes **2353** and **2354**. While these pathways were not strongly enriched, their presence suggests potential activation of calcium-regulatory and bone-resorptive signaling mechanisms, which might reflect systemic effects of infection on CMs tissues.

In contrast, the **downregulated DEGs** revealed significant enrichment in key pathways, shedding light on suppressed biological processes. The most notable was the pathway **"Metabolic pathways"** (hsa01100), which had a set size of 15 genes, an NES of -1.72, and included genes involved in broad metabolic processes, such as **1158**, **1152**, **27124**, and **4713**. This result highlights a general downregulation of metabolic activity in CMs tissue, likely reflecting impaired energy production and utilization. Another enriched pathway was **"Regulation of actin cytoskeleton"** (hsa04810), which had an NES of -1.67 and included genes such as **4629**, **730**, and **4633**. This finding underscores the disruption of actin cytoskeletal organization, which is critical for maintaining cellular structure and contractile function in CMs tissues.

Additionally, the pathway **"Motor proteins"** (hsa04814), with an NES of -1.54, included 13 genes such as **7134**, **4625**, and **7137**, reflecting suppression of genes essential for intracellular transport and contractile activity. Another noteworthy pathway, **"Prion disease"** (hsa05020), with an NES of -1.51, highlighted three downregulated genes (**513**, **730**) potentially linked to neurodegenerative-like mechanisms, though the relevance of this finding in CMs tissue remains unclear.

Together, these results reveal a clear divergence between upregulated and downregulated pathways in CMs tissues during infection. The upregulated pathways suggest modest activation of calcium and signaling-related processes, while the downregulated pathways highlight significant suppression of core metabolic pathways, cytoskeletal regulation, and contractile functions, providing mechanistic insights into the transcriptional remodeling of CMs tissues in response to infection.

3.2.2 Shared Pathways and Biological Processes in AWOs and Cardiac Tissues Based on KEGG Analysis

The results of Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis reveal several biological processes and signaling pathways that are commonly perturbed in AWOs and CMs tissues in response to infection. Among the shared pathways, four—**Cytosolic DNA-Sensing Pathway**, **RIG-I-Like Receptor Signaling Pathway**, **Viral Life Cycle - HIV-1**, and **Viral Protein Interaction with Cytokine and Cytokine Receptor**—were significantly **upregulated** in both tissues, while **Ascorbate and Aldarate Metabolism** was consistently **downregulated**.

The upregulation of the **Cytosolic DNA-Sensing Pathway** and **RIG-I-Like Receptor Signaling Pathway** highlights the activation of innate immune responses as a core mechanism of antiviral defense. These pathways play critical roles in recognizing viral or bacterial nucleic acids and initiating inflammatory responses through type I interferons and pro-inflammatory cytokines. Similarly, the activation of pathways associated with **Viral Life Cycle - HIV-1** and **Viral Protein Interaction with Cytokine and Cytokine Receptor** underscores the immune system's efforts to combat infection while also reflecting the ability of viral pathogens to exploit host cellular machinery for replication and immune evasion.

In contrast, the consistent downregulation of **Ascorbate and Aldarate Metabolism** in both tissues suggests a depletion of antioxidant defenses. This pathway, which is critical for the metabolism of vitamin C and related compounds, is essential for reducing oxidative stress and maintaining tissue integrity. Its downregulation indicates a potential increase in oxidative damage during infection, contributing to tissue injury and dysfunction in both AWOs and CMs systems. Oxidative stress is initially recognized as a means of combating viruses and protecting the host, contributing to apoptosis. However, with the development of research, more and more researchers found that oxidative stress promoted viral replication, which was a common mechanism used by some specific viruses (Wu et al. 2022).

As for the GO enrichment analysis the most significant shared biological processes enriched in both AWOs and CMs tissues was the **positive regulation of miRNA transcription.** Dysregulation of miRNA expression has been linked to the pathogenesis of COVID-19, particularly in severe cases characterized by cytokine storm, acute respiratory distress syndrome (ARDS), and multi-organ damage, including myocarditis and CMs dysfunction. miRNAs can act as regulators of both host and viral gene expression, influencing critical pathways such as the cytokine response, oxidative stress, and the viral life cycle. The enrichment of this pathway in both tissues underscores the systemic role of miRNA-mediated regulation during COVID-19 (Rasizadeh et al. 2023).

These findings demonstrate the systemic nature of the infection's impact, with immune signaling and metabolic dysregulation being coordinated across multiple tissues. The KEGG analysis results emphasize the interconnectedness of these pathways, revealing shared biological processes that are integral to understanding how infections affect the body at a molecular level. This systemic perspective is critical for deciphering the interplay between immune responses, viral mechanisms, and metabolic changes across tissues, ultimately providing insights into potential therapeutic targets.

4 Discussion

COVID-19, caused by the SARS-CoV-2 virus, is a systemic disease characterized by its profound impact on multiple organ systems, particularly the respiratory and cardiovascular systems. Respiratory dysfunction is the leading cause of death in COVID-19 patients, with studies reporting that nearly 96.7% of fatal cases involve respiratory system damage, and about half exhibit multi-organ involvement, including cardiovascular complications. This highlights the systemic nature of the disease, which can trigger a cascade of inflammatory and immune responses that extend beyond the lungs. Cardiovascular dysfunction, the second most common organ system affected, has been linked to mechanisms such as myocardial inflammation, cytokine storms, and endothelial dysfunction, which may contribute to conditions such as myocarditis, heart failure, and arrhythmias.

In this study, we examined the specific and shared molecular pathways and biological processes enriched in differentially expressed genes (DEGs) in AWOs and CMs tissues to better understand the mechanisms driving severe respiratory and CMs symptoms in COVID-19.

4.1 Identification of tissue-specific responses to SARS-CoV-2 and Potential therapeutic targets.

The AWOs epithelium serves as the first line of defense against respiratory pathogens, including SARS-CoV-2, and plays a pivotal role in orchestrating host immune responses. Our differential expression analysis of AWOs tissue revealed a distinct set of DEGs that reflect the tissue's critical functions in viral recognition, epithelial barrier maintenance, and immune signaling. Among these, the robust upregulation of interferon I (IFN-I) pathway genes indicates a strong antiviral immune response. Key interferon-stimulated genes (ISGs), such as **IFIT2**, **BST2**, and **MX2**, were significantly upregulated, underscoring their critical roles in restricting viral replication and enhancing epithelial resilience. These findings are consistent with prior studies highlighting that interferon pathways, particularly type I and type III interferons (e.g., IFN-λ), are central to the AWOs epithelium's antiviral defense (Busnadiego et al. 2020)

Notably, this potent interferon-driven immune response not only limits viral spread but also minimizes ACE2 expression, thereby reducing the susceptibility of AWOs cells to further viral entry.

However, this dual role of the AWOs epithelium—mounting antiviral defenses while contributing to inflammation—may also drive pathology in severe COVID-19 cases. This variety and complexity of immune responses is caused by the virus's ability to evade or manipulate the IFN-mediated host responses, which might not be perfectly tuned to combat this novel pathogen. These observations validate the AWOs epithelium's critical function in SARS-CoV-2 infection and highlight interferon signaling pathways as promising therapeutic targets. Enhancing the activity of interferon, I (IFN-I) pathways could boost the antiviral response in organoids. This could involve the use of recombinant interferons or drugs that stimulate endogenous IFN production to improve immune defense against SARS-CoV-2 (Mihaescu et al. 2024).

Our findings also emphasize the maladaptive aspects of AWOs immune responses, particularly in severe COVID-19, where hyperinflammatory conditions exacerbate tissue damage. For example, upregulation of SOCS1 and IRF1 in the AWOs dataset aligns with prior reports of excessive interferon signaling contributing to delayed viral clearance and immune dysregulation. This suggests that while the AWOs epithelium efficiently mounts antiviral defenses, dysregulated responses may perpetuate inflammation and impair recovery (Guo et al. 2023).

The AWOs tissue findings from this study also provide valuable insights into mechanisms that may contribute to the destruction of ciliated cells and impaired mucociliary clearance observed in COVID-19. Our findings elucidate mechanisms contributing to the loss of ciliated cells and impaired mucociliary clearance in COVID-19. The enrichment of **negative regulation of cell adhesion** suggests compromised epithelial integrity, leading to cell detachment and weakened barrier function, COVID-19 provokes the destruction of ciliated cells and leads to a reduction in mucociliary clearance, which in turn promotes the accumulation of mucus and debris in the airway, providing an optimal environment for the virus to replicate and spread (Gonzalez-Rubio et al. 2023).

Additionally, **synaptic membrane adhesion** dysregulation may impair epithelial-neuronal communication, disrupting mucus clearance. The activation of **cellular response to epidermal growth factor (EGF) stimulus** reflects repair attempts, but these may be ineffective in the inflammatory environment of COVID-19 (Engler et al. 2023).

Together, these disruptions align with clinical observations of mucus accumulation and increased viral replication in the AWOs.

In contrast, **CMs tissue** DEGs revealed unique insights into myocardial involvement during SARS-CoV-2 infection, suggesting a less pronounced interferon-driven immune response compared to AWOs tissue. Instead, several key genes related to CMs function and homeostasis were found to be dysregulated (Madeddu 2020).

SARS-CoV-2 is known to cause **myocarditis-like syndromes** and cardiovascular complications, for instance, **NPPB (B-type Natriuretic Peptide)**, a biomarker of CMs stress and dysfunction, was downregulated, suggesting direct myocardial effects or systemic factors contributing to impaired heart function (Sobreira and Nóbrega 2021). Similarly, the downregulation of **ACTA1** and **MYL2**, which encode essential structural proteins of CMs muscle, indicates potential mechanisms of CMs injury, including disrupted contractility and cardiomyocyte damage.

Another notable finding was the downregulation of **CYP2J2**, a gene involved in the metabolism of arachidonic acid to epoxyeicosatrienoic acids (EETs). EETs are known for their cardioprotective properties, particularly in reducing CMs inflammation. The reduced expression of **CYP2J2** in CMs tissue may exacerbate inflammatory processes, compounding myocardial damage during infection (Wang et al. 2022).

These observations are consistent with clinical findings of myocarditis, arrhythmias, and heart failure in COVID-19 patients, highlighting the virus's ability to induce both direct and indirect CMs damage.

The contrast between the strong interferon-driven immune response in AWOs tissue and the metabolic and structural dysregulation in CMs tissue underscores the tissue-specific responses to SARS-CoV-2 infection. These differences align with clinical observations where respiratory system dysfunction is the primary driver of mortality, while cardiovascular complications, although less frequent, are critical contributors to multi-organ failure and severe disease outcomes. The identification of these tissue-specific molecular signatures advances our understanding of the systemic nature of COVID-19 and provides a foundation for developing targeted therapeutic strategies to address organ-specific complications.

4.2 Shared molecular pathways across organoids

The shared pathways identified in this study provide a molecular basis for understanding the systemic effects of COVID-19, particularly the link between severe respiratory symptoms and cardiac complications. The upregulated immune pathways correspond to clinical observations of hyperinflammation, cytokine storm, and immune-mediated damage in ARDS and myocarditis. This finding aligns with previous reports highlighting that respiratory dysfunction, primarily due to ARDS, is the leading cause of mortality in COVID-19 patients, often exacerbated by multi-organ damage and systemic inflammation. Conversely, the downregulation of antioxidant pathways is consistent with increased oxidative stress observed in patients with severe disease. These findings align with clinical reports of myocardial injury, arrhythmias, and long-term cardiovascular complications in COVID-19 patients.

As the first line of host defense, the innate immune system plays a pivotal role in recognizing and responding to SARS-CoV-2 infection. Among the key pathways identified, the **cGAS-STING pathway**, involved in the detection of cytosolic DNA, was significantly upregulated in both airway and cardiac tissues. This pathway is a key driver of type I interferon responses and antiviral immunity, but its overactivation in the context of COVID-19 has been implicated in sustained inflammation and tissue damage, particularly in severe cases. Evidence suggests that SARS-CoV-2 can activate the cGAS-STING pathway, triggering an excessive immune response that contributes to the cytokine storm characteristic of severe COVID-19. Enhancing the activity of interferon, I (IFN-I) pathways could boost the antiviral response in organoids. This could involve the use of recombinant interferons or drugs that stimulate endogenous IFN production to improve immune defense against SARS-CoV-2.

In airway tissues, overactivation of the **cytosolic DNA-sensing pathway** likely exacerbates ARDS by promoting excessive inflammation and epithelial damage. In cardiac tissues, this same pathway has been linked to inflammatory conditions such as myocarditis and arrhythmogenic right ventricular cardiomyopathy (ARVC). Clinical reports of elevated interferon-stimulated gene (ISG) expression in COVID-19 patients further validate the systemic immune activation observed in this study. These findings underscore the need to balance immune activation to combat the virus while minimizing collateral tissue damage.

Another notable finding was the downregulation of the **ascorbate and aldarate metabolism pathway**, which plays a critical role in vitamin C metabolism and antioxidant defense. This pathway is essential for neutralizing reactive oxygen species (ROS) and mitigating oxidative stress. In both airway and cardiac tissues, its downregulation suggests a depletion of antioxidant reserves, contributing to the accumulation of oxidative damage.

In airway tissues, reduced antioxidant activity likely exacerbates epithelial damage and inflammation, worsening ARDS and impairing mucociliary clearance. In cardiac tissues, oxidative stress is a major driver of myocardial injury, arrhythmias, and long-term cardiac remodeling. These findings are consistent with clinical observations of elevated markers of oxidative stress and reduced antioxidant capacity in severe COVID-19 patients, emphasizing the critical need for therapeutic strategies aimed at restoring redox balance.

Together, these shared pathways—immune activation via cGAS-STING signaling and reduced antioxidant defenses—highlight the interconnected nature of respiratory and cardiac complications in COVID-19. These findings provide a molecular framework for understanding the systemic effects of the disease and point toward potential therapeutic targets, including modulators of the cGAS-STING pathway and strategies to enhance antioxidant defenses.

5 Conclusion

This study provides a comprehensive investigation into the molecular mechanisms underlying SARS-CoV-2 infection, with a focus on airway and cardiac tissues. RNA-Seq analysis revealed distinct tissue-specific responses that reflect the pathophysiology of COVID-19.

In airway tissues, upregulation of interferon-stimulated genes (**IFIT2**, **BST2**, and **MX2**) highlights the critical role of interferon signaling in antiviral defense. However, disruptions in pathways such as **negative regulation of cell adhesion** and **cellular response to epidermal growth**

factor stimulus suggest mechanisms driving epithelial damage, impaired mucociliary clearance, and inflammation observed in severe COVID-19.

In cardiac tissues, the downregulation of structural and functional genes (**NPPB**, **ACTA1**, **MYL2**, and **CYP2J2**) indicates myocardial injury, impaired contractility, and reduced cardioprotective mechanisms. These findings align with clinical observations of myocarditis, arrhythmias, and heart failure in COVID-19 patients, emphasizing the virus's capacity for both direct and systemic cardiac damage.

Shared pathways, including the **RIG-I-like receptor signaling pathway**, **Cytosolic DNA-sensing pathway**, and **positive regulation of miRNA transcription**, underline the systemic nature of COVID-19 and its link to immune dysregulation and severe outcomes. Therapeutic targets such as interferon pathways, **SOCS1**, and antioxidant defenses hold promise for mitigating tissue damage and systemic inflammation.

This study advances our understanding of COVID-19 pathogenesis, highlighting tissue-specific and shared molecular responses as potential therapeutic avenues to address the systemic and organ-specific complications of the disease.

6 References

Assou, Said, Engi Ahmed, Lisa Morichon, Amel Nasri, Florent Foisset, Carine Bourdais, Nathalie Gros, et al. 2023. "Human Airway Ex Vivo Models: New Tools to Study the Airway Epithelial Cell Response to SARS-CoV-2 Infection." bioRxiv. https://doi.org/10.1101/2023.04.15.536998.

Babu, Golap, and Fahim Alam Nobel. 2022. "Identification of Differentially Expressed Genes and Their Major Pathways among the Patient with COVID-19, Cystic Fibrosis, and Chronic Kidney Disease." *Informatics in Medicine Unlocked* 32 (January):101038. https://doi.org/10.1016/j.imu.2022.101038.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics (Oxford, England)* 30 (15): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Busnadiego, Idoia, Sonja Fernbach, Marie O. Pohl, Umut Karakus, Michael Huber, Alexandra Trkola, Silke Stertz, and Benjamin G. Hale. 2020. "Antiviral Activity of Type I, II, and III Interferons Counterbalances ACE2 Inducibility and Restricts SARS-CoV-2." *mBio*, September. https://doi.org/10.1128/mbio.01928-20.

Chung, Matthew, Vincent M. Bruno, David A. Rasko, Christina A. Cuomo, José F. Muñoz, Jonathan Livny, Amol C. Shetty, Anup Mahurkar, and Julie C. Dunning Hotopp. 2021. "Best Practices on the Differential Expression Analysis of Multi-Species RNA-Seq." *Genome Biology* 22 (1): 121. https://doi.org/10.1186/s13059-021-02337-8.

Danecek, Petr, James K Bonfield, Jennifer Liddle, John Marshall, Valeriu Ohan, Martin O Pollard, Andrew Whitwham, et al. 2021. “Twelve Years of SAMtools and BCFtools.” *GigaScience* 10 (2): giab008. https://doi.org/10.1093/gigascience/giab008.

Deshpande, Dhrithi, Karishma Chhugani, Yutong Chang, Aaron Karlsberg, Caitlin Loeffler, Jinyang Zhang, Agata Muszyńska, et al. 2023. “RNA-Seq Data Science: From Raw Data to Effective Interpretation.” *Frontiers in Genetics* 14:997383. https://doi.org/10.3389/fgene.2023.997383.

Dobin, Alexander, Carrie A. Davis, Felix Schlesinger, Jorg Drenkow, Chris Zaleski, Sonali Jha, Philippe Batut, Mark Chaisson, and Thomas R. Gingeras. 2013. “STAR: Ultrafast Universal RNA-Seq Aligner.” *Bioinformatics (Oxford, England)* 29 (1): 15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dobin, Alexander, and Thomas R. Gingeras. 2015. “Mapping RNA-Seq Reads with STAR.” *Current Protocols in Bioinformatics* 51 (September):11.14.1-11.14.19. https://doi.org/10.1002/0471250953.bi1114s51.

Ekanger, Camilla Tvedt, Fan Zhou, Dana Bohan, Maria Lie Lotsberg, Maria Ramnefjell, Laurence Hoareau, Gro Vatne Røsland, et al. 2022. “Human Organotypic Airway and Lung Organoid Cells of Bronchiolar and Alveolar Differentiation Are Permissive to Infection by Influenza and SARS-CoV-2 Respiratory Virus.” *Frontiers in Cellular and Infection Microbiology* 12 (March). https://doi.org/10.3389/fcimb.2022.841447.

Engler, Melanie, Dan Albers, Pascal Von Maltitz, Rüdiger Groß, Jan Münch, and Ion Cristian Cirstea. 2023. “ACE2-EGFR-MAPK Signaling Contributes to SARS-CoV-2 Infection.” *Life Science Alliance* 6 (9): e202201880. https://doi.org/10.26508/lsa.202201880.

Erb, Anna, Ulrich M. Zissler, Madlen Oelsner, Adam M. Chaker, Carsten B. Schmidt-Weber, and Constanze A. Jakwerth. 2022. “Genome-Wide Gene Expression Analysis Reveals Unique Genes Signatures of Epithelial Reorganization in Primary Airway Epithelium Induced by Type-I, -II and -III Interferons.” *Biosensors* 12 (11): 929. https://doi.org/10.3390/bios12110929.

Ewels, Philip, Måns Magnusson, Sverker Lundin, and Max Käller. 2016. “MultiQC: Summarize Analysis Results for Multiple Tools and Samples in a Single Report.” *Bioinformatics (Oxford, England)* 32 (19): 3047–48. https://doi.org/10.1093/bioinformatics/btw354.

Gonzalez-Rubio, Julian, Vu Thuy Khanh Le-Trilling, Lea Baumann, Maria Cheremkhina, Hannah Kubiza, Anja E. Luengen, Sebastian Reuter, et al. 2023. “SARS-CoV-2 Particles Promote Airway Epithelial Differentiation and Ciliation.” *Frontiers in Bioengineering and Biotechnology* 11 (November). https://doi.org/10.3389/fbioe.2023.1268782.

Guo, Tony J. F., Gurpreet K. Singhera, Janice M. Leung, and Delbert R. Dorscheid. 2023. “Airway Epithelial-Derived Immune Mediators in COVID-19.” *Viruses* 15 (8): 1655. https://doi.org/10.3390/v15081655.

Jiang, Gao, Juan-Yu Zheng, Shu-Ning Ren, Weilun Yin, Xinli Xia, Yun Li, and Hou-Ling Wang. 2024. “A Comprehensive Workflow for Optimizing RNA-Seq Data Analysis.” *BMC Genomics* 25 (1): 631. https://doi.org/10.1186/s12864-024-10414-y.

Li, James, Rency S. Varghese, and Habtom W. Ressom. 2024. “RNA-Seq Data Analysis.” *Methods in Molecular Biology (Clifton, N.J.)* 2822:263–90. https://doi.org/10.1007/978-1-0716-3918-4_18.

Liao, Yang, Gordon K. Smyth, and Wei Shi. 2014. “featureCounts: An Efficient General Purpose Program for Assigning Sequence Reads to Genomic Features.” *Bioinformatics (Oxford, England)* 30 (7): 923–30. https://doi.org/10.1093/bioinformatics/btt656.

Love, Michael I., Wolfgang Huber, and Simon Anders. 2014. “Moderated Estimation of Fold Change and Dispersion for RNA-Seq Data with DESeq2.” *Genome Biology* 15 (12): 550. https://doi.org/10.1186/s13059-014-0550-8.

Madeddu, Paolo. 2020. “Cardiovascular Complications of COVID-19: Evidence, Misconceptions, and New Opportunities.” *Vascular Biology* 2 (1): E3–6. https://doi.org/10.1530/VB-20-0008.

Mihaescu, Grigore, Mariana Carmen Chifiriuc, Roxana Filip, Coralia Bleotu, Lia Mara Ditu, Marian Constantin, Roxana-Elena Cristian, et al. 2024. “Role of Interferons in the Antiviral Battle: From Virus-Host Crosstalk to Prophylactic and Therapeutic Potential in SARS-CoV-2 Infection.” *Frontiers in Immunology* 14 (January). https://doi.org/10.3389/fimmu.2023.1273604.

Nakayama, Ryuichi, Naofumi Bunya, Takashi Tagami, Mineji Hayakawa, Kazuma Yamakawa, Akira Endo, Takayuki Ogura, et al. 2023. “Associated Organs and System with COVID-19 Death with Information of Organ Support: A Multicenter Observational Study.” *BMC Infectious Diseases* 23 (1): 814. https://doi.org/10.1186/s12879-023-08817-5.

“Organoids as a Novel Tool in Modelling Infectious Diseases.” 2023. *Seminars in Cell & Developmental Biology* 144 (July):87–96. https://doi.org/10.1016/j.semcdb.2022.09.003.

Rabaan, Ali A, Samira Smajlović, Huseyin Tombuloglu, Sabahudin Ćordić, Azra Hajdarević, Nudžejma Kudić, Abbas Al Mutai, et al. 2023. “SARS-CoV-2 Infection and Multi-Organ System Damage: A Review.” *Biomolecules & Biomedicine* 23 (1): 37–52. https://doi.org/10.17305/bjbms.2022.7762.

Rasizadeh, Reyhaneh, Parisa Shiri Aghbash, Javid Sadri Nahand, Taher Entezari-Maleki, and Hossein Bannazadeh Baghi. 2023. “SARS-CoV-2-Associated Organs Failure and Inflammation: A Focus on the Role of Cellular and Viral microRNAs.” *Virology Journal* 20 (1): 179. https://doi.org/10.1186/s12985-023-02152-6.

Rodrigues, Pedro, Rafael S. Costa, and Rui Henriques. 2022. “Enrichment Analysis on Regulatory Subspaces: A Novel Direction for the Superior Description of Cellular Responses to SARS-CoV-2.” *Computers in Biology and Medicine* 146 (July):105443. https://doi.org/10.1016/j.compbiomed.2022.105443.

Shen, Wei, Botond Sipos, and Liuyang Zhao. 2024. “SeqKit2: A Swiss Army Knife for Sequence and Alignment Processing.” *iMeta* 3 (3): e191. https://doi.org/10.1002/imt2.191.

Sheng, Quanhu, Kasey Vickers, Shilin Zhao, Jing Wang, David C. Samuels, Olivia Koues, Yu Shyr, and Yan Guo. 2016. “Multi-Perspective Quality Control of Illumina RNA Sequencing Data Analysis.” *Briefings in Functional Genomics*, September, elw035. https://doi.org/10.1093/bfgp/elw035.

Sobreira, Débora R., and Marcelo A. Nóbrega. 2021. “Regulatory Landscapes of *Nppa* and *Nppb*.” *Circulation Research* 128 (1): 130–32. https://doi.org/10.1161/CIRCRESAHA.120.318495.

Tang, Xuming, Dongxiang Xue, Tuo Zhang, Benjamin E. Nilsson-Payant, Lucia Carrau, Xiaohua Duan, Miriam Gordillo, et al. 2023. “A Multi-Organoid Platform Identifies CIART as a Key Factor for SARS-CoV-2 Infection.” *Nature Cell Biology* 25 (3): 381–89. https://doi.org/10.1038/s41556-023-01095-y.

Thomas, Paul D. 2017. “The Gene Ontology and the Meaning of Biological Function.” *Methods in Molecular Biology (Clifton, N.J.)* 1446:15–24. https://doi.org/10.1007/978-1-4939-3743-1_2.

Wang, Guyi, Bing Xiao, Jiayi Deng, Linmei Gong, Yi Li, Jinxiu Li, and Yanjun Zhong. 2022. “The Role of Cytochrome P450 Enzymes in COVID-19 Pathogenesis and Therapy.” *Frontiers in Pharmacology* 13 (February). https://doi.org/10.3389/fphar.2022.791922.

Westermann, Alexander J., and Jörg Vogel. 2018. “Host-Pathogen Transcriptomics by Dual RNA-Seq.” *Methods in Molecular Biology (Clifton, N.J.)* 1737:59–75. https://doi.org/10.1007/978-1-4939-7634-8_4.

Wu, Yaru, Min Zhang, Cui Yuan, Zhenling Ma, Wenqing Li, Yanyan Zhang, Lijuan Su, Jun Xu, and Wei Liu. 2022. “Progress of cGAS-STING Signaling in Response to SARS-CoV-2 Infection.” *Frontiers in Immunology* 13 (December). https://doi.org/10.3389/fimmu.2022.1010911.

Yu, Guangchuang, Li-Gen Wang, Yanyan Han, and Qing-Yu He. 2012. “clusterProfiler: An R Package for Comparing Biological Themes among Gene Clusters.” *Omics : A Journal of Integrative Biology* 16 (5): 284–87. https://doi.org/10.1089/omi.2011.0118.

Zhao, Shanrong. 2014. “Assessment of the Impact of Using a Reference Transcriptome in Mapping Short RNA-Seq Reads.” *PloS One* 9 (7): e101374. https://doi.org/10.1371/journal.pone.0101374.

Zhou, Qian, Xiaoquan Su, Gongchao Jing, Songlin Chen, and Kang Ning. 2018. “RNA-QC-Chain: Comprehensive and Fast Quality Control for RNA-Seq Data.” *BMC Genomics* 19 (1): 144. https://doi.org/10.1186/s12864-018-4503-6.