Myriam Yasmina Soumahoro

Host transcriptional response to Influenza A infection revealed by RNA-seq analysis

Abstract

Influenza A virus infection induces rapid transcriptional changes in host cells, largely driven by innate antiviral immune responses. In this study, we used RNA sequencing (RNA-seq) to characterize host gene expression changes following Influenza A infection. RNA-seq data from infected and mock-treated samples were processed using a reproducible pipeline including quality control, transcript quantification with a combined host–virus reference, and gene-level differential expression analysis with DESeq2. Global expression profiling revealed a clear separation between infected and control samples. Differential expression analysis identified strong induction of interferon-stimulated genes (ISGs), including OAS1, OAS3, STAT1, IRF7, and IFIT family members. These results highlight the central role of interferon-mediated pathways in the early host response to Influenza A infection and provide a basis for further investigation of adaptive immune responses.

Introduction

Influenza A virus is a major human respiratory pathogen that triggers rapid activation of host immune defenses upon infection. Early antiviral responses are primarily mediated by the innate immune system, particularly through interferon signaling pathways that induce the expression of interferon-stimulated genes (ISGs). These genes play a critical role in limiting viral replication and shaping downstream immune responses.

RNA sequencing (RNA-seq) enables genome-wide characterization of host transcriptional responses to viral infection. In this study, we applied an RNA-seq analysis framework to investigate host gene expression changes induced by Influenza A infection. Using a combined host–virus reference for transcript quantification followed by host-specific differential expression analysis, we aimed to identify global transcriptional patterns and key antiviral genes activated during infection.

Materials and Methods

Dataset and experimental design

RNA sequencing data analyzed in this study were obtained from a publicly available dataset associated with Ashraf *et al.* (2020) (GEO accession: GSE155241), which investigated host transcriptional responses to Influenza A virus infection. Raw sequencing data were retrieved as

SRA run files (SRR) using the SRA Run Selector and converted to FASTQ format for downstream analysis. When necessary, FASTQ files were re-downloaded from the European Nucleotide Archive (ENA) to ensure data integrity.

The experimental design included mock-treated and Influenza A virus–infected samples, with multiple biological replicates per condition. Sample metadata were used to guide all downstream analyses. Viral transcripts were quantified at the transcript level but excluded from host gene-level differential expression analyses.

RNA-seq processing and quality control

Raw sequencing reads (FASTQ files) were subjected to standard quality control procedures to assess sequencing quality and data integrity. Quality metrics were evaluated to ensure that all samples met acceptable thresholds for downstream transcript quantification. No samples were excluded based on quality control results.

Reference construction and transcript quantification

To enable accurate quantification of both host and viral transcripts in infected samples, a combined reference transcriptome was constructed. This reference included the human transcriptome (GENCODE v49) together with the Influenza A virus genome (A/WSN/33 strain). Transcript-level quantification was performed using **Salmon**, allowing sequencing reads to be mapped simultaneously to host and viral transcripts.

This combined-reference strategy ensured proper read assignment in infected samples while preserving the ability to isolate host-specific expression profiles for downstream analyses.

Gene-level expression quantification

Transcript-level abundance estimates generated by Salmon were imported into R using the **tximport** package. Transcript-to-gene mappings were derived from GENCODE v49 annotations. Gene-level count matrices were constructed by summarizing transcript-level estimates while accounting for transcript length and library size, enabling consistent gene-level expression estimates across samples.

Differential expression analysis

Differential gene expression analysis between Influenza A–infected and mock-treated samples was performed using **DESeq2**. Gene-level counts were modeled using a negative binomial distribution, and normalization was applied to account for differences in sequencing depth across samples. Genes with low expression across all samples were filtered prior to analysis. Statistical significance was assessed using adjusted p-values to control for multiple testing.

Data transformation and visualization

For visualization and exploratory analyses, variance-stabilizing transformation (VST) was applied to gene expression counts. Principal component analysis (PCA) was used to assess global transcriptional patterns and sample clustering. Differential expression results were visualized using volcano plots to highlight genes with significant expression changes. In addition, expression patterns of selected interferon-stimulated genes (ISGs) were examined using heatmaps based on scaled VST values.

Results

Global transcriptional changes induced by Influenza A infection

To assess overall transcriptional differences between mock-treated and Influenza A virus–infected samples, principal component analysis (PCA) was performed on variance-stabilized gene expression data. PCA revealed a clear separation between mock and infected samples along the first principal component (PC1), which explained 98% of the total variance (Figure 1). This strong separation indicates that Influenza A infection is the dominant source of transcriptional variation in the dataset.

The second principal component (PC2) accounted for 1% of the variance and captured minor variability within each condition. Replicates clustered consistently according to infection status, indicating a robust and reproducible host transcriptional response to Influenza A infection.

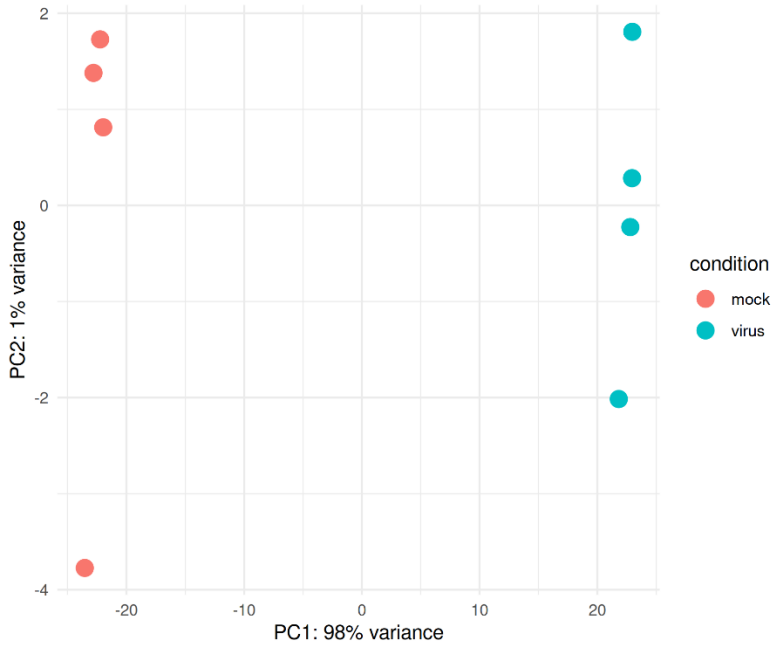

***Figure 1*: Principal component analysis (PCA) of host gene expression profiles**.

Differential gene expression following Influenza A infection

Differential expression analysis was performed using DESeq2 to identify host genes significantly affected by Influenza A infection. The volcano plot illustrates widespread transcriptional remodeling in infected samples compared to mock-treated controls (Figure 2). Many genes displayed significant changes in expression, with both upregulated and downregulated gene sets observed.

Genes with the strongest induction exhibited large positive log2 fold changes and high statistical significance, indicating a pronounced activation of infection-associated transcriptional programs. Conversely, a subset of genes showed significant downregulation in infected samples, reflecting repression of host processes during viral infection.

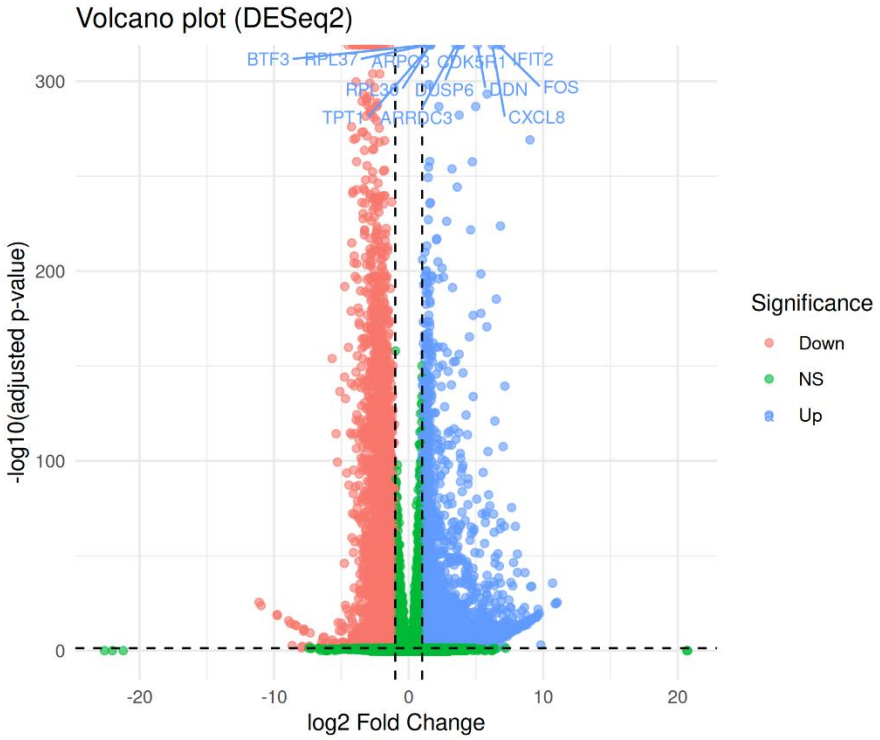

***Figure 2*: Differential gene expression analysis of host genes following Influenza A-infected and mock-treated samples.**

Induction of interferon-stimulated genes

To further characterize the nature of the host response to Influenza A infection, expression patterns of selected interferon-stimulated genes (ISGs) were examined. The heatmap in Figure 3 shows a

clear infection-dependent expression pattern, with ISGs displaying higher expression in virus-infected samples compared to mock-treated controls.

Genes such as **MX1, IFI27, IFIT1, IFIT2, IFIT3, ISG15, RSAD2, OAS1, OAS2, OAS3, STAT1, and IRF7** were consistently upregulated in infected samples. Hierarchical clustering grouped samples according to infection status, indicating coordinated induction of interferon-mediated antiviral pathways in response to Influenza A infection.

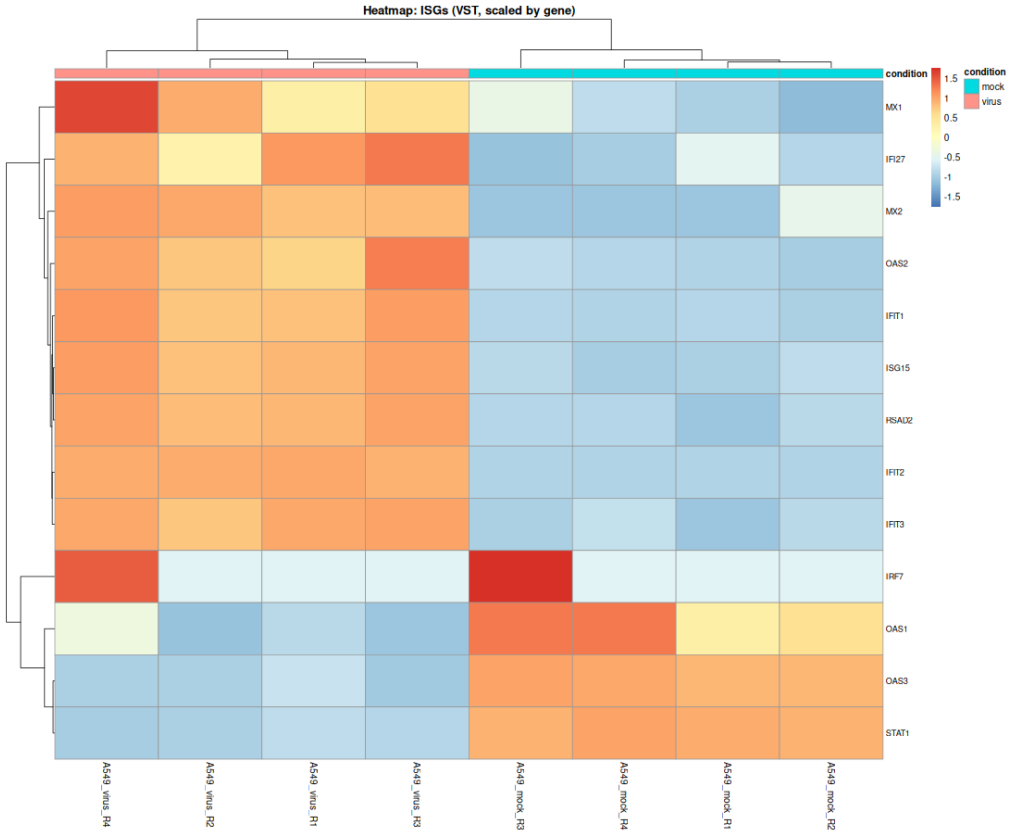

***Figure 3*: Expression patterns of interferon-stimulated genes in response to Influenza A infection.**

Discussion

In this study, we used RNA sequencing to characterize host transcriptional responses to Influenza A infection. Global expression analysis revealed a strong infection-driven effect, with infected samples clearly separating from mock-treated controls. Differential expression analysis further demonstrated extensive transcriptional remodeling, highlighting robust induction of interferon-stimulated genes (ISGs) consistent with activation of innate antiviral defense mechanisms.

The coordinated upregulation of genes such as OAS1, OAS3, STAT1, IRF7, and IFIT family members reflects a canonical interferon-mediated response to viral infection. These genes are well known for their roles in viral RNA sensing, inhibition of viral replication, and amplification of

antiviral signaling pathways. The consistency of ISG induction across samples supports the robustness of the observed host response and confirms the biological relevance of the analysis.

From a methodological perspective, the use of a combined host–virus reference for transcript quantification was a deliberate and conservative choice. By allowing simultaneous quantification of host and viral transcripts, this approach minimizes the risk of read misassignment in infected samples while preserving the ability to focus on host-specific transcriptional changes during downstream analysis. Viral transcripts were subsequently excluded at the gene-level summarization stage, enabling a clean and interpretable host differential expression analysis.

This project was conducted prior to any single-cell RNA-seq analyses and therefore reflects a bulk RNA-seq perspective of host responses. As such, the results represent averaged transcriptional signals across heterogeneous cell populations. While this approach does not resolve cell-type–specific responses, it provides a robust global overview of infection-induced transcriptional programs and serves as a necessary foundation for more granular analyses.

Building on these results, a follow-up project focusing on T-cell responses to Influenza infection was initiated. However, this project was intentionally discontinued due to insufficient annotation of experimental conditions, including incomplete information on stimulation context and sample grouping. Rather than proceeding with uncertain biological assumptions, the analysis was halted to preserve interpretability and scientific rigor. This decision underscores the importance of well-defined experimental design and metadata quality in transcriptomic studies.

Overall, this RNA-seq analysis demonstrates the power of bulk transcriptomics to capture global host responses to viral infection and highlights interferon signaling as a dominant early response to Influenza A. Importantly, it also illustrates how bulk RNA-seq analyses can inform the design and feasibility of subsequent, more targeted studies. These results provided both biological insight and methodological grounding that later informed the choice and interpretation of single-cell RNA-seq approaches in downstream projects.

Conclusion

This study provides a transcriptomic overview of host responses to Influenza A infection using an RNA-seq approach. Global expression analyses revealed that infection is the primary driver of transcriptional variation, with strong and consistent induction of interferon-stimulated genes across infected samples. These results highlight the central role of interferon-mediated pathways in the early antiviral response to Influenza A.

Methodologically, this work demonstrates the value of bulk RNA-seq for capturing global host transcriptional programs and for guiding subsequent analytical decisions. The use of a combined host–virus reference ensured accurate read assignment in infected samples while preserving interpretability of host-specific differential expression analyses. Importantly, this project

established a rigorous and reproducible analytical framework that informed later exploration of more complex transcriptomic approaches.

Overall, this analysis serves as a foundational study for understanding host–virus interactions at the transcriptional level and illustrates how careful interpretation of bulk RNA-seq data can inform the design, feasibility, and scope of downstream immune-focused investigations.

Reference

Ashraf U. et al. (2020). Influenza virus infection induces widespread alterations of host cell splicing. GEO accession: GSE155241.