

OPEN ACCESS

EDITED BY
Lirong Wang,
University of Pittsburgh, United States

REVIEWED BY
Abhimanyu Banerjee,
Illumina, United States
Zhi-Ping Liu,
Shandong University, China
Michael Liebman,
IPQ Analytics, United States

*CORRESPONDENCE
Mostafa Rezapour
✉ mrezapou@wakehealth.edu

RECEIVED 22 March 2024
ACCEPTED 12 August 2024
PUBLISHED 30 August 2024

CITATION
Rezapour M, Niazi MKK, Lu H, Narayanan A
and Gurcan MN (2024) Machine
learning-based analysis of Ebola virus' impact
on gene expression in nonhuman primates.
Front. Artif. Intell. 7:1405332.
doi: 10.3389/frai.2024.1405332

COPYRIGHT
© 2024 Rezapour, Niazi, Lu, Narayanan and
Gurcan. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

Machine learning-based analysis of Ebola virus' impact on gene expression in nonhuman primates

Mostafa Rezapour^{1*}, Muhammad Khalid Khan Niazi¹, Hao Lu¹,
Aarthi Narayanan² and Metin Nafi Gurcan¹

¹Center for Artificial Intelligence Research, Wake Forest University School of Medicine, Winston-Salem, NC, United States, ²Department of Biology, George Mason University, Fairfax, VA, United States

Introduction: This study introduces the Supervised Magnitude-Altitude Scoring (SMAS) methodology, a novel machine learning-based approach for analyzing gene expression data from non-human primates (NHPs) infected with Ebola virus (EBOV). By focusing on host-pathogen interactions, this research aims to enhance the understanding and identification of critical biomarkers for Ebola infection.

Methods: We utilized a comprehensive dataset of NanoString gene expression profiles from Ebola-infected NHPs. The SMAS system combines gene selection based on both statistical significance and expression changes. Employing linear classifiers such as logistic regression, the method facilitates precise differentiation between RT-qPCR positive and negative NHP samples.

Results: The application of SMAS led to the identification of IFI6 and IFI27 as key biomarkers, which demonstrated perfect predictive performance with 100% accuracy and optimal Area Under the Curve (AUC) metrics in classifying various stages of Ebola infection. Additionally, genes including MX1, OAS1, and ISG15 were significantly upregulated, underscoring their vital roles in the immune response to EBOV.

Discussion: Gene Ontology (GO) analysis further elucidated the involvement of these genes in critical biological processes and immune response pathways, reinforcing their significance in Ebola pathogenesis. Our findings highlight the efficacy of the SMAS methodology in revealing complex genetic interactions and response mechanisms, which are essential for advancing the development of diagnostic tools and therapeutic strategies.

Conclusion: This study provides valuable insights into EBOV pathogenesis, demonstrating the potential of SMAS to enhance the precision of diagnostics and interventions for Ebola and other viral infections.

KEYWORDS

Ebola virus infection, gene expression profiling, biomarker discovery, machine learning in virology, transcriptomic analysis

1 Introduction

Ebola virus disease (EVD) is a severe and often fatal illness affecting humans and nonhuman primates (NHPs), with certain outbreaks, such as the 2013–2016 West Africa outbreak, resulting in high mortality rates (Speranza et al., 2018; Centers for Disease Control Prevention, 2014). The scarcity of human clinical samples for research purposes has led to the reliance on NHP models, particularly cynomolgus and rhesus macaques, for

studying EVD pathogenesis and testing potential treatments and vaccines. These models typically involve exposing animals to the Ebola virus (EBOV) through various methods, which produce symptoms and mortality patterns similar to those observed in humans, though differences in infection doses and disease progression are noted (Speranza et al., 2018).

Furthering our understanding of the immune response to EVD, particularly through transcriptomic analyses, has been a significant outcome of studies using these NHP models (Caballero et al., 2016; Versteeg et al., 2017). In this context, Speranza et al. (2018) made a notable contribution by focusing on enhancing the accuracy of NHP models to more closely reflect human EVD. Their study involved the exposure of 12 cynomolgus macaques to the EBOV/Makona strain via intranasal routes, using a target dose of 100 plaque-forming units (PFU). The administration method varied, utilizing either a pipette or a mucosal atomization device. This led to a diverse onset of symptoms and disease progression among the animals, resulting in the identification of four distinct response groups with an overall fatality rate of 83% (Speranza et al., 2018).

The animals were categorized into these groups based on the timing and nature of symptom appearance, onset of viremia, and time to death. Group 1, following a typical EVD course, showed quantifiable viremia from day 6 and had an average time to death of 10.47 days. Group 2, with a delayed onset, had detectable viremia between days 10 to 12 and an average time to death of 13.31 days. Group 3 experienced a late onset of disease, with detectable viremia emerging after day 20 and an average time to death of 21.42 days. Contrasting these, Group 4 did not develop detectable viremia during the experiment and survived until the 41-day study endpoint (Speranza et al., 2018) (see Figure 1).

In preparation for their experiment, Speranza et al. (2018) conducted pre-exposure evaluations on blood chemistry, hematology, and soluble proteins, but these did not reveal any significant physiological differences among the groups, as confirmed by principal components analysis. Employing RNA sequencing and NanoString, which focuses on 769 NHP transcripts for transcriptomic analysis, they provided an overview of the host response to EBOV exposure. Their findings revealed a uniform and predictable response to lethal EVD, irrespective of the time to onset. Remarkably, they also discovered that the expression of specific genes could predict the development of disease before the appearance of clinical signs such as fever (Speranza et al., 2018).

Despite the significant findings of Speranza et al. (2018), certain aspects of EVD pathogenesis in NHPs remain underexplored, particularly in the context of genetic markers and their predictive power. While their study laid a solid foundation for understanding the disease's progression through transcriptomic analysis, it opened the door for more targeted research in identifying specific genetic indicators of EVD. This gap in knowledge presents an opportunity for employing advanced analytical techniques, such as machine learning, to delve deeper into the genetic landscape of EVD. Our study, therefore, focuses on expanding upon these initial findings, aiming to uncover finer genetic details that could be critical in diagnosing and treating EVD more effectively. By leveraging the rich dataset provided by Speranza et al. (2018), we introduce a cutting-edge approach that promises to bring new insights into the complex interaction between the EBOV and its host at a molecular level.

Our research introduces a novel machine learning methodology aimed at prioritizing genes that are significantly associated with EVD, as determined by the Benjamin-Hochberg procedure (Benjamini and Hochberg, 1995). This innovative approach allows us to not only identify gene fingerprints uniquely related to EVD but also to differentiate between EVD-positive and EVD-negative NHPs using a single gene at a time in a supervised manner. This dual focus on gene prioritization and individual gene-based differentiation represents a significant advancement in the field, offering a more nuanced understanding of the genetic underpinnings of EVD. Our methodology's precision in identifying and analyzing individual genes holds the potential to greatly enhance the specificity of EVD diagnostics and contribute to the development of targeted therapeutic strategies. Through this work, we aim to demonstrate the power of machine learning in uncovering new insights from existing genomic data, thereby opening new avenues for research in the field of infectious diseases.

2 Materials and methods

2.1 Data

This study utilizes a dataset of NanoString gene expression profiles from NHPs infected with the EBOV, as detailed in Speranza et al. (2018). The NanoString platform used recognizes 769 specific NHP transcripts, advantageous for its rapid processing capability and lower RNA quality requirements compared to RNA-seq, while retaining efficacy for EBOV research. Normalization of the NanoString data was conducted in line with standard procedures, where background adjustments were made based on negative controls, and lane variations were accounted for using internal positive controls (Speranza et al., 2018). The most stable reference genes for normalization were identified using the NormFinder R package (Andersen et al., 2004), ensuring precise calibration for RNA input variations. Further methodological specifics are available in the referenced work by Speranza et al. (2018), which should be consulted for an in-depth understanding of the protocols and procedures applied.

2.2 Supervised Magnitude-Altitude Scoring: a machine learning approach for gene expression profiling

This section describes the Supervised Magnitude-Altitude Scoring (SMAS) method, specifically tailored to distinguish RT-qPCR positive and negative NHP samples in EBOV research. SMAS comprises a structured three-stage process:

1. In the initial stage of gene selection, our methodology rigorously identifies genes that demonstrate statistical significance between RT-qPCR positive and negative groups. To address the challenges of multiple hypothesis testing inherent in gene expression studies, we strictly apply the Benjamini-Hochberg (BH) correction (Benjamini and Hochberg, 1995). This correction is essential for controlling the false discovery rate, a critical factor in genomic data analysis where numerous tests are conducted simultaneously.

Group 1 (Genome Equivalents/ml)					Color	Description		
NHP \ Day	0	3	6	10				
1	n.d.	n.d.	8.60		n.d.	Not detectable (Negative)		
2	n.d.	n.d.	8.29	8.79		RT-qPCR Below Quantification Threshold		
12	n.d.	n.d.	7.72	9.50		RT-qPCR-positive		
						RT-qPCR-positive (GE>9: Strong Positive)		

Group 2 (Genome Equivalents/ml)							
NHP \ Day	0	3	6	10	12	14	15
3	n.d.	n.d.		9.68			
5	n.d.	n.d.	n.d.	9.21		7.21	7.62
8	n.d.	n.d.			X		
10	n.d.	n.d.		10.04			

Group 3 (Genome Equivalents/ml)					
NHP \ Day	0	3	6	10	21
4	n.d.	n.d.	n.d.		6.93
6	n.d.	n.d.			
9	n.d.	n.d.			9.81

Group 4 (Genome Equivalents/ml)									
NHP \ Day	0	3	6	10	14	21	28	35	41
7	n.d.	n.d.					n.d.	n.d.	n.d.
11	n.d.	n.d.	n.d.					n.d.	n.d.

FIGURE 1
 The tables exhibit the timeline of responses for different nonhuman primates (NHPs) to EBOV exposure, arranged in rows, over a sequence of days post-exposure, which are organized in columns. Green cells indicate time points where reverse transcription quantitative polymerase chain reaction (RT-qPCR) did not detect any viral RNA (labeled as "n.d.," not detectable), suggesting a negative result. Yellow cells denote instances where RT-qPCR detected viral RNA, but the quantity was below the quantification threshold. Orange and purple cells highlight RT-qPCR-positive results, with purple specifically denoting a high viral load, as the Genome Equivalent (GE) is >9, termed as "strong positive." The box marked with an "X" indicates that although the RT-qPCR results are positive, the corresponding NanoString information is unavailable.

We employ two-sample independent *t*-tests (Kim, 2015) with the BH correction, ensuring that identified genes are significantly different beyond random chance occurrences. Moreover, our criteria for selecting genes include a log fold change (logFC) >1. This means we only consider genes that not only pass the BH significance threshold but also exhibit substantial expression differences.

- In the second stage of our analysis, the Magnitude-Altitude Score (MAS) is calculated for each gene selected in the initial stage. The Magnitude-Altitude Scoring (MAS) formula is:

$$MAS_l = |(\log_2 FC_l)|^M |(\log_{10}(p_l^{BH}))|^A,$$

for $l = 1, 2, \dots, s$, where s is the number of rejected null hypotheses by BH adjusted method ($p_l^{BH} < \alpha$, where $\alpha = 0.05$). The hyperparameters M and A are used to achieve a balance between the adjusted p -value and the log fold change, facilitating a comprehensive analysis of gene expression changes. In traditional gene ranking methods employed by differential expression analysis tools like EdgeR (Robinson et al., 2010) [or DESeq2 (Love et al., 2014)], the focus is predominantly on p -values for ranking genes. This approach is exemplified in EdgeR's typical workflow, where genes are ranked using the "topTags" function. This function sorts genes primarily based on their p -values, which

are calculated to determine the statistical significance of differential expression between conditions. While this method effectively identifies statistically significant genes, it may not always highlight genes with the most biologically significant changes in expression. In contrast, our MAS system, with both M and A set to 1, goes beyond traditional p -value ranking. By incorporating the log fold change (logFC) into the ranking process alongside the adjusted p -values from the Benjamini-Hochberg correction, MAS makes it possible that the genes selected for further analysis show not only statistical significance but also potential biologically relevant expression changes. This balanced approach of considering both statistical and biological significance allows for a more comprehensive understanding of gene expression changes in the context of EBOV infection.

Genes are ranked according to their MAS, and the top- d genes are earmarked for predictive modeling. The number d of genes selected is capped at one-tenth of the total number of samples, in line with a widely accepted machine learning guideline that suggests having at least 10 samples for each variable. This guideline helps prevent overfitting and possibly increases the robustness of the model (Hastie et al., 2009). By adhering to this principle, our approach effectively balances the inclusion of a sufficient number of genes to capture the EBOV infection's complexity while maintaining robustness.

3. In the final stage of our methodology, the SMAS approach demonstrates its efficacy through the implementation of linear classifiers for EVD status prediction. The top- d genes, meticulously identified through MAS, serve as predictors in linear classifiers, specifically employing logistic regression or support vector classifiers with a linear kernel (James et al., 2013). This choice of linear classifiers is strategically made considering the often-limited sample size encountered in multi-omic datasets. In such scenarios, more complex models are prone to overfitting, potentially skewing predictions (Hastie et al., 2009). By opting for linear classifiers, we effectively mitigate this overfitting risk, striking a balance between model simplicity and the ability to accurately predict EVD status in unseen NHP samples.

To assess the robustness of our predictive model, we implement K -fold stratified cross-validation (Kuhn and Johnson, 2013). This validation technique is critical in our study as it ensures a balanced representation of both positive and negative samples in the training and testing datasets. Such a balanced approach is pivotal in enhancing the reliability and generalizability of our model. By incorporating a cross-validation strategy, we are not only able to gauge the model's performance on unseen data but also affirm its stability and predictive power across different subsets of the data.

Integrating the top- d genes from the SMAS method into linear classifiers showcases the utility of our approach in practical applications. By selecting genes that are both statistically significant and biologically relevant and deploying them in a well-validated predictive model, the SMAS method presents a potent tool in the field of infectious disease research, particularly for EVD analysis. This stage, therefore, underscores the effectiveness of the SMAS method in translating complex genomic data into actionable insights, paving the way for advancements in diagnostic and therapeutic strategies.

Through the SMAS method's comprehensive stages, we chose to set both M and A to 1 in the MAS formula, as opposed to the traditional ranking method where $M = 0$ and $A = 1$, which reduces the analysis to the adjusted p -values from the Benjamini-Hochberg correction. This decision was based on the following rationale:

1. Enhanced discrimination between groups: The MAS scoring with $M = A = 1$ takes into account both the statistical significance and the magnitude of change in gene expression. This dual consideration allows for a more nuanced differentiation between the RT-qPCR positive and negative groups. In contrast, the traditional ranking method primarily focuses on statistical significance, potentially overlooking genes with substantial biological changes.
2. Biological relevance of log fold change (logFC): By setting a threshold for logFC (e.g., $\log_{FC} > 1$), we increase the possibility that genes with biologically significant expression changes are considered. This threshold is crucial as it highlights genes that are not just statistically significant but also have substantial changes in expression levels, which is more likely to be biologically relevant in the context of EBOV response in NHPs.

3. Improved predictive performance in supervised learning: In supervised machine learning, the balance between statistical significance and expression magnitude is vital for model accuracy. The MAS scoring with $M = A = 1$, combined with the logFC threshold, tends to select genes that provide a clearer separation between positive and negative samples. This improved separation enhances the predictive performance of the model, particularly when using linear classifiers like logistic regression or support vector classifiers with a linear kernel.

In summary, adopting MAS scoring not only aligns with the principles of robust machine learning but also ensures that the genes selected for the predictive model are the most informative for distinguishing EVD status in NHP samples.

2.3 Differential expression analysis and classification using Supervised Magnitude-Altitude Scoring

In our exploration of the EBOV's interactions with its host, we have developed a methodical approach to analyze the dynamics of viral infection. This approach is segmented into two primary objectives:

- **Objective 1:** a comprehensive differential expression analysis and
- **Objective 2:** the development of a bi-class classification model.

Figure 2 illustrates our structured approach to studying EBOV-host interactions, breaking down into two primary objectives. The first main objective consists of four subobjectives, each dedicated to a detailed differential expression analysis. In these subobjectives, we focus on genes that are significant according to the Benjamini-Hochberg method, exhibit a log fold change (logFC) > 1 , and are further prioritized using MAS method to delineate the gene expression changes at different stages of EBOV infection in NHPs. The collective findings from these subobjectives aim to provide a comprehensive understanding of the host's genetic response. The second main objective leverages the key genes identified in Objective 1 to develop a bi-class classification model, utilizing these genes as predictors to classify NHP samples as either positive or negative for EBOV infection.

2.3.1 Objective 1. Comprehensive differential expression analysis using MAS scoring

Objective 1 systematically dissects the impact of the EBOV on gene expression in NHPs, utilizing the MAS system. This objective is divided into four key sub-objectives, each targeting a specific aspect of the host-pathogen interaction, thus providing a multi-layered understanding of the EBOV's genetic impact. Objective 1.1 focuses on extreme cases of infection, comparing the most intense positive and negative responses. This specific analysis is critical for identifying genetic markers that undergo significant alterations in severe infection scenarios, revealing key genes that are most responsive or vulnerable during heightened viral replication.

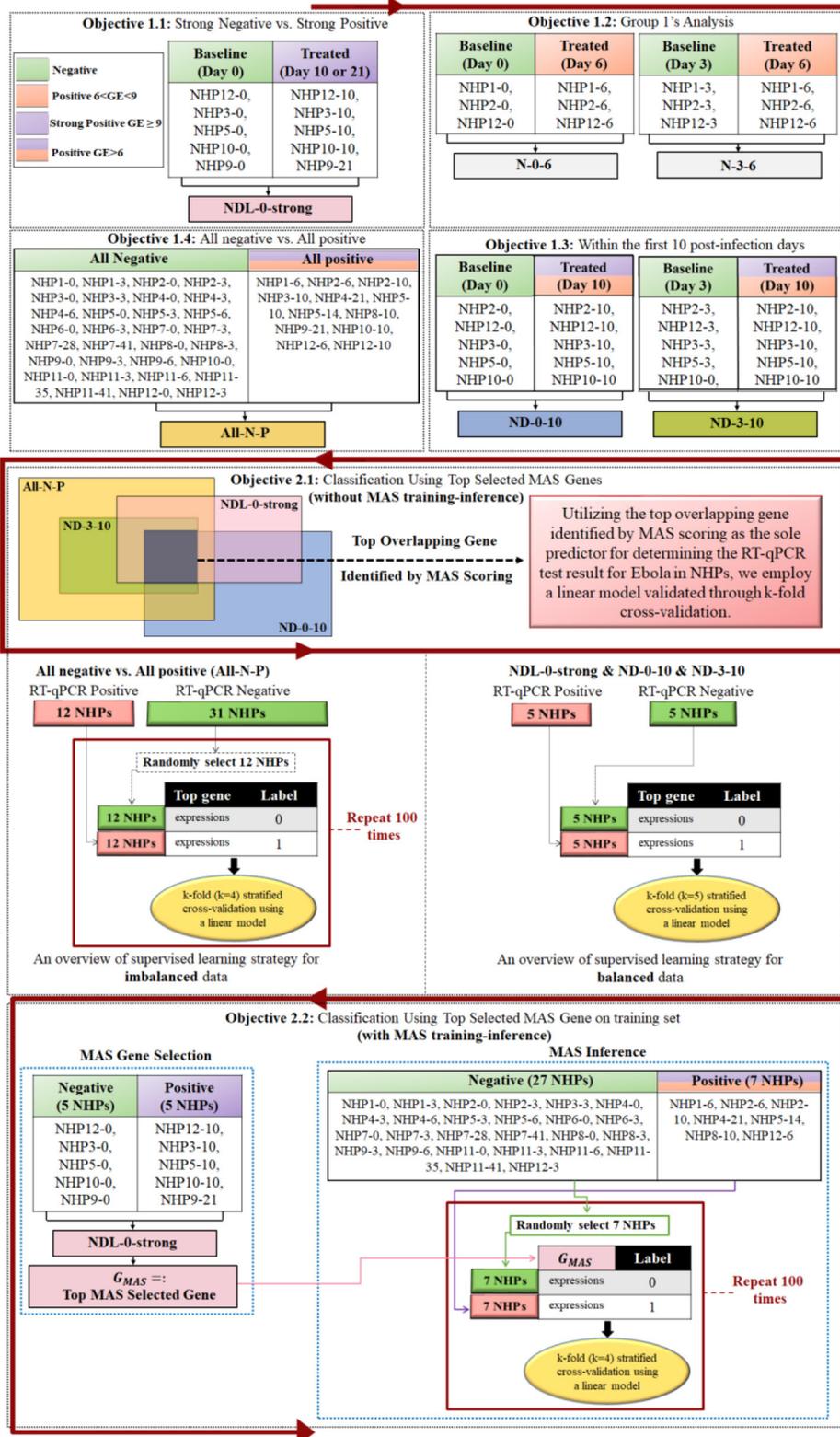


FIGURE 2 This figure encapsulates the structured approach of Objective 1, which is divided into four subobjectives, each conducting differential expression analyses on NHP samples at various stages of EBOV infection. Genes are selected based on Benjamini–Hochberg significance with $\log_{FC} > 1$ and prioritized using the MAS method. The naming convention “NHPm–n” refers to NanoString data related to nonhuman primate number “m” on day “n” post-infection, as indicated in Figure 1. The integrated results of these analyses are then applied in Objective 2 to build a bi-class classification model for predicting EBOV infection status. In Objective 1.4, since NanoString data for NHP8-12 was not available, but NHP8-10 data was available with three RT-qPCR replicates that were positive, we used NHP8-10 as the positive sample.

Objective 1.2 then shifts focus to the early stages of infection, examining gene expression dynamics from initial exposure to the peak of viral response in Group 1. This objective is instrumental in understanding the progression of early-response genes, providing crucial insights for early detection and intervention.

Objectives 1.3 and 1.4 further expand the analysis, each adding a vital layer of understanding. Objective 1.3 explores gene expression in typical and delayed response scenarios, highlighting the genetic basis for variability in host response timing. This objective is particularly important for understanding genes involved in delayed viral response or clearance, offering valuable information on how different hosts react to the virus over time. Finally, Objective 1.4 provides a broad overview, comparing all negative and positive infection statuses. This comprehensive comparison captures the overall genetic signature of EBOV infection, identifying consistent genetic markers across various infection stages. Together, these objectives offer a multi-layered and detailed understanding of the host's genetic response to the EBOV, ensuring a thorough exploration of pathogenesis and host response at the genetic level. This structured approach is crucial for revealing the complex genetic landscape shaped by EBOV infection, contributing significantly to developing more precise diagnostic tools and therapeutic strategies.

2.3.1.1 Objective 1.1. Strong negative vs. strong positive analysis (NDL-0-strong)

This objective engages in a detailed analysis to explore the multifaceted impact of the EBOV on gene expression in NHPs, employing the MAS system. The focus is to dissect the genetic response in NHPs across varying degrees of EBOV infection intensity, thereby illuminating the broad spectrum of host-pathogen interactions.

Within this comprehensive framework, a specialized subset of analysis, termed as “Strong Negative vs. Strong Positive Analysis (NDL-0-strong),” is conducted. This segment specifically aims to distinguish the gene expression contrasts between the most extreme cases of EBOV infection. By comparing the gene expression profiles of strong positive RT-qPCR samples (indicated by a high Genome Equivalent, GE >9, and marked as purple in Figure 1) against those of strong negative samples (from day 0), this analysis provides a deep dive into the genetic shifts occurring under intense viral influence.

The procedure involves carefully selecting samples from Groups 1, 2, and 3 (see Figure 1) that exhibit strong positive RT-qPCR results. These samples are then methodically contrasted with the baseline samples taken on day 0, focusing particularly on NHPs that later exhibit strong positive outcomes. This approach allows for a targeted analysis of significant gene expression changes associated with high GE values. For clarity and ease of reference, the set of all genes selected using the MAS scoring with a log fold change (logFC) >1 in this phase is collectively referred to as “NDL-0-strong.”

The rationale behind this intense focus lies in the value of contrasting extreme cases: strong positive (high GE) vs. strong negative (day 0) samples. Such an approach is instrumental in unearthing insights into the genetic markers most drastically altered in the wake of severe EBOV infection. This comparison not only aids in understanding the genetic extremes induced by the virus but also plays a crucial role in highlighting

the key genes that are most responsive or vulnerable during heightened viral replication phases. Through this nuanced analysis, the study endeavors to unravel the complex genetic landscape shaped by EVD, thereby contributing significantly to the broader understanding of the virus's genetic impact on its hosts (see Figure 2).

2.3.1.2 Objective 1.2. Differential expression in group 1 (N-0-6 and N-3-6)

Objective 1.2 delves into the specific gene expression dynamics of Group 1 (see Figure 1), which is characterized by a typical EBOV infection progression. This analysis focuses on understanding the transition from the initial exposure to the peak viral response, primarily observed on day 6 post-exposure.

The method involves applying the MAS within Group 1 to distinguish significant gene expression changes between day 6 (when RT-qPCR is positive) and days 0 and 3 (when RT-qPCR is negative). The genes with a log fold change (logFC) >1, when comparing day 6 against day 0, are categorized as “N-0-6,” while those compared against day 3 are labeled “N-3-6” (see Figure 2).

This approach is critical for identifying genes that respond early or undergo significant alterations during the acute phase of the infection. By examining these key time points, we can pinpoint the genetic markers that are pivotal in the initial stages of viral replication and host response.

2.3.1.3 Objective 1.3. Analysis within normal and delayed groups (ND-0-10 and ND-3-10)

Objective 1.3 aims to unravel the gene expression patterns in scenarios of typical and delayed EBOV response (Groups 1 and 2 in Figure 1), particularly noticeable by day 10 post-infection. This objective targets an understanding of the genetic underpinnings that might influence the timing of viral detection and clearance.

The procedure employs MAS scoring to analyze gene expression in normal and delayed response groups with positive RT-qPCR results on day 10. The set of all genes selected with a log fold change >1 when comparing day 10 against day 0 is referred to as “ND-0-10,” and the gene set from the comparison of day 10 against day 3 is termed “ND-3-10” (see Figure 2).

This analysis is vital in exploring genes that could be responsible for a delayed response to the EBOV, thereby providing insights into the genetic basis for variability in host response timing. It helps to understand how different gene expression patterns contribute to the delayed manifestation of the virus.

2.3.1.4 Objective 1.4. Comparison of all negative vs. all positive samples (All-N-P)

Objective 1.4 encompasses a broader perspective, aiming to distinguish the overarching gene expression signatures that separate overall negative from positive EBOV infection statuses. This comprehensive comparison seeks to capture the complete genetic landscape associated with EBOV presence or absence.

Using MAS scoring, this analysis compares the collective gene expression of all negative samples (indicated as green in Figure 1) against all positive samples (denoted as orange and purple). This broad comparison is instrumental in identifying markers consistently expressed across various stages of infection, categorized collectively as “All-N-P.”

The rationale behind this objective is to provide a holistic view of the gene expression changes associated with EBOV infection. It is designed to capture the overall genetic signature indicative of infection, thereby differentiating between infected and uninfected states in a comprehensive manner (see [Figure 2](#)).

Objectives 1.1 to 1.4 provide a layered and detailed understanding of the host's genetic response to the EBOV. These objectives collectively aim to identify gene fingerprints that can effectively differentiate between NHPs testing positive and those testing negative for RT-qPCR across varying infection scenarios and response patterns. This structured approach ensures a multifaceted analysis, contributing significantly to the broader understanding of EBOV pathogenesis and host response at the genetic level.

2.3.2 Objective 2. Application of MAS insights for EBOV classification

The primary goal of our paper is to identify gene signatures associated with EBOV infection by employing differential expression analysis through the MAS method. This unsupervised machine learning approach, which forms the crux of Objective 1, operates without traditional training methods and is instrumental in uncovering significant gene expressions linked to the EBOV. Objective 2 builds on this foundation, extending our analysis from gene identification to practical application in EBOV research. This objective bifurcates into two distinct yet interconnected components.

In Objective 2.1, our focus shifts to utilizing linear classifiers to demonstrate the efficacy of genes selected through the MAS method. Here, we illustrate how these MAS-selected genes can effectively differentiate between RT-qPCR negative and positive samples of EBOV in nonhuman primates. This stage is pivotal in showcasing how the insights gained from unsupervised machine learning can be transitioned into practical, diagnostic applications.

In Objective 2.2, we demonstrate the MAS method's capacity for generalization by selecting a gene from one independent dataset and then applying it for supervised classification on a different dataset. This approach not only tests the selected gene's predictive power but also showcases the adaptability and robustness of the MAS method across diverse contexts. By employing the MAS-identified gene in a separate dataset to accurately classify EBOV infection status, we provide compelling evidence of the MAS's generalizability. The ability of the MAS method to extend its application from gene selection in one scenario to effective prediction in another exemplifies its utility and versatility in infectious disease research, especially in studies of complex viral infections like Ebola.

2.3.2.1 Objective 2.1. Classification using top selected MAS genes

Objective 2.1 capitalizes on the insights gained from the differential expression analysis in Objective 1 to develop a binary classification model. This model distinguishes between positive and negative EBOV infection statuses in NHPs based on their gene expression profiles. Utilizing logistic regression, the model integrates the most significant genes identified across all sub-objectives of Objective 1 as predictive variables. To validate

and ensure the reliability of this model, we employ a k-fold stratified cross-validation approach, which rigorously assesses its performance and generalizability.

The core rationale of Objective 2.1 is to transform the complex gene expression data into a practical diagnostic tool. Leveraging the key genes identified through the MAS method, the model aims to provide a robust and precise means of classifying EBOV infection status. This application of genomic insights into a functional diagnostic model not only underscores the utility of our analytical approach but also has the potential to inform targeted therapeutic interventions.

A significant aspect of Objective 2.1, as illustrated in [Figure 2](#), involves addressing the challenge of imbalanced datasets, a common issue in biomedical research. For instance, in the All-N-P group analysis, we encounter an imbalance with 12 positive samples and 31 negative samples. To tackle this, we implement a random sampling strategy: selecting 12 negative samples to pair with the 12 positive samples, thereby creating a balanced dataset of 24 samples. This process is repeated 100 times, each with a different subset of negative samples, ensuring a comprehensive evaluation. The performance of the model across these iterations is assessed through metrics such as the average Area Under the Curve (AUC), accuracy, precision, recall, and F1-score. This approach not only enhances the model's robustness but also provides a more accurate reflection of its predictive capability.

In summary, Objective 2.1, in conjunction with Objective 1, forms a holistic analysis of the EBOV's genetic interactions with its host. While Objective 1 establishes the foundation by pinpointing critical genes and their expression patterns in response to EBOV infection, Objective 2.1 translates these findings into a pragmatic classification tool. This synergistic blend of in-depth genomic analysis and practical application is pivotal in advancing the field of infectious disease research, particularly in the context of EBOV diagnostics and therapy.

2.3.2.2 Objective 2.2. Implementing MAS on an independent set for subsequent supervised classification

Initially, MAS is applied to a balanced dataset, characterized by strong positive contrasts against strong negatives, acting as a basis for gene selection. The top MAS-selected gene, named G_{MAS} , is then identified. G_{MAS} is subsequently used as the sole predictor in a linear classifier to differentiate RT-qPCR positive and negative NHP samples in a separate dataset, which remains unexposed during the MAS top gene selection process.

As shown in [Figure 2](#), the necessity for strategic MAS application arises from the limited sample size. MAS is employed on the ND-0-strong subset to isolate the most significant gene (top-MAS selected gene) for further examination. This identified gene, G_{MAS} , becomes central to subsequent classification tasks across different sample sets. These tasks are thoroughly validated through balanced stratified cross-validation techniques as shown in [Figure 2](#).

In essence, Objective 2 is where the theoretical and analytical advancements of our study converge into practical applications. It showcases the transition from identifying gene signatures associated with EBOV infection using an unsupervised machine learning approach, to applying these insights in a supervised setting, thus advancing the field of infectious disease

research and offering new avenues for EBOV diagnostics and treatment strategies.

2.4 Comprehensive Gene Ontology analysis of identified Benjamini-Hochberg significant genes through SMAS

Following the identification of genes with statistical significance using the Supervised Magnitude-Altitude Scoring (SMAS) method and subsequent adjustments via the Benjamini-Hochberg correction, we undertook a comprehensive Gene Ontology (GO) analysis. This analysis played a crucial role in categorizing the identified genes into groups associated with biological processes (BPs), cellular components (CCs), and molecular functions (MFs). For this detailed classification, we utilized the clusterProfiler (Wu et al., 2021) and org.Hs.eg.db (Carlson et al., 2019) packages within R, enabling efficient mapping of our significant genes to specific GO terms.

This GO analysis was instrumental in illuminating the functional characteristics of these genes within the cellular context, particularly in their response to Ebola virus exposure in nonhuman primates. By mapping the significant genes to specific GO terms, we could highlight the functional disruptions and activations within the host cells. This approach helped identify critical biological processes, cellular components, and molecular functions that are likely impacted by the Ebola virus, providing insights into the pathogen-host interactions at a molecular level.

3 Results

In this section, we present the results of the methodology outlined in Section 2.

3.1 Differential expression analysis and classification using Supervised Magnitude-Altitude Scoring

3.1.1 Objective 1. Comprehensive differential expression analysis using MAS scoring

For objectives with at least 10 NHPs, we conducted supervised analysis to predict whether an NHP is RT-qPCR positive or negative.

3.1.1.1 Objective 1.1. Strong negative vs. strong positive analysis (NDL-0-strong)

Figure 3 displays the results of analysis for Objective 1.1, where we compare the NHPs with strong positive RT-qPCR results against their status on day 0, when RT-qPCR results were strongly negative. Figure 3A depicts the volcano plot and highlights the top MAS-selected genes with a log fold change (logFC) >1. Figure 3B presents a three-dimensional visualization of all strongly positive and negative NHPs, utilizing only the top three selected MAS genes: IFI27, IFI6, and HP. Given that we have more than 10 NHPs for this objective, we implemented the Supervised Magnitude-Altitude

Scoring (SMAS) approach (Objective 2) using logistic regression. This was carried out with a five-fold stratified cross-validation, specifically focusing on the top selected gene, IFI27. The Receiver Operating Characteristic (ROC) curves for each of the five folds, along with their mean, are depicted in Figure 3C. This figure also includes the AUC and accuracy metrics for each fold, providing a comprehensive evaluation of the model's predictive performance. It is important to note that we opted for a single gene and a linear model to ensure the model's strong generalizability, as the MAS-selected gene IFI27 is exceptionally suitable for this purpose.

3.1.1.2 Objective 1.2. Differential expression in group 1 (N-0-6 and N-3-6)

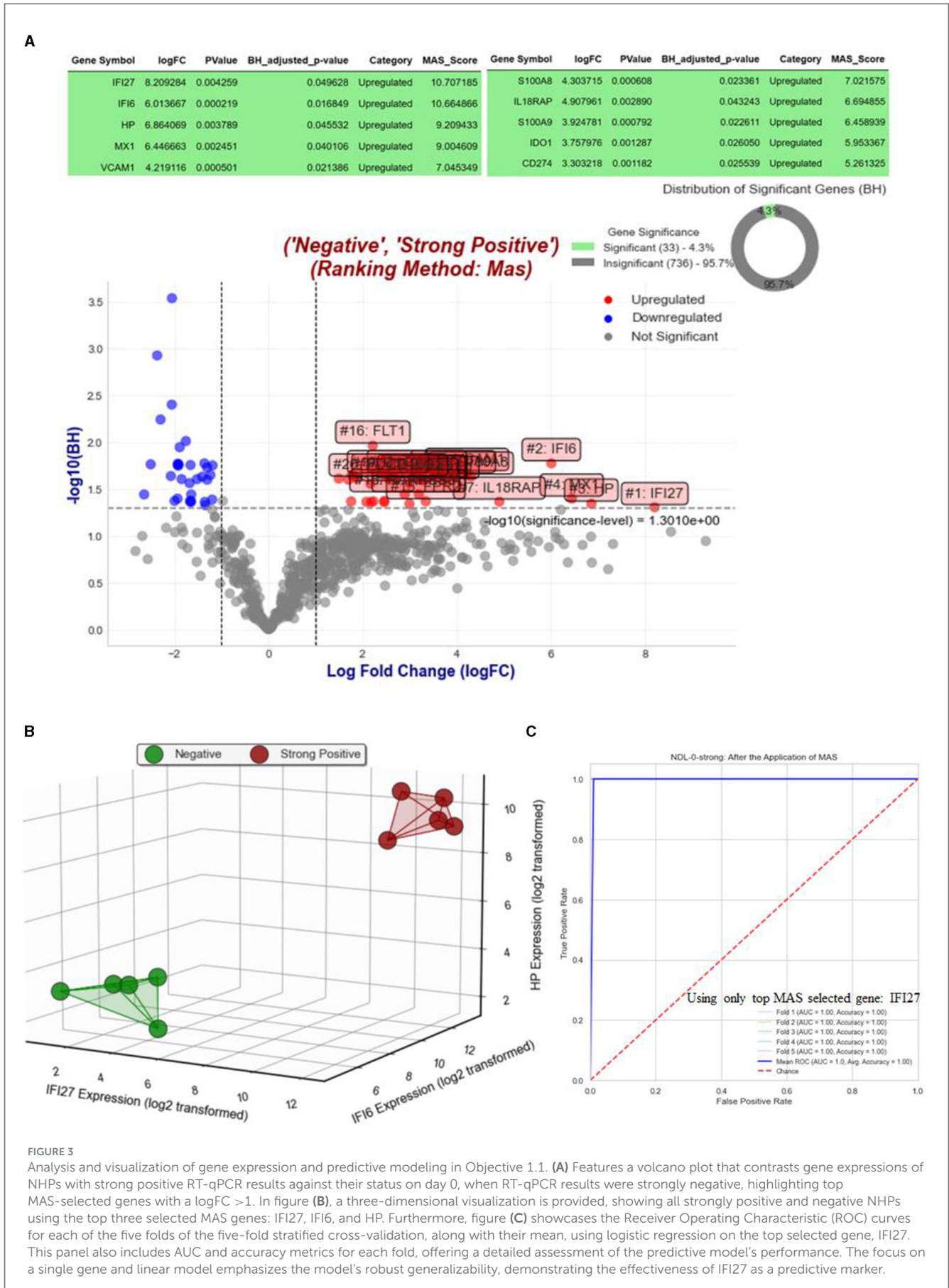
Figure 4 illustrates the results of Objective 1.2. Figures 4A, B display the volcano plots and Benjamini-Hochberg (BH) significant genes when comparing NHPs from day 6 to day 0 and day 3, respectively. IFI27, IL6, and THBD are the only BH-significant genes with a log fold change (logFC) >1 when comparing NHPs on day 6 vs. day 0. However, when comparing day 6 to day 3, only IFI27 remains significant. Notably, for the purpose of distinguishing NHPs on day 3 from day 0, we also performed MAS for Day 3 vs. Day 0, but no gene emerged as BH significant (see Figure 4C). Nevertheless, when relaxing the Benjamini-Hochberg criteria and considering only raw *p*-values, FLT1 emerged as the top significant gene with a logFC >1 based on MAS (see Figure 4D). Figure 4E illustrates the expression of IFI27 (selected from Objectives 1.1 and 1.2) and IFI6 (selected from Objective 1.1), along with FLT1, which distinguishes NHPs on day 0 from day 3. Since we have a total of 9 NHPs across all three post-infection days, we did not perform a supervised learning analysis.

3.1.1.3 Objective 1.3. Analysis within normal and delayed groups (ND-0-10 and ND-3-10)

Figure 5 presents the results corresponding to Objective 1.3. Figure 5A displays a Venn diagram that illustrates the BH significant genes with a logFC >1 when comparing NHPs on day 10 vs. day 0 and day 10 vs. day 3, highlighting both unique and common genes prioritized by MAS. ISG15, IFI6, and IFI44 emerge as the top three selected genes significantly expressed in both comparisons. Notably, the absence of unique significant genes in the day 10 vs. day 3 comparison suggests an early expression of these genes. Figures 5B, C depict the unique and common BH significant genes across Objectives 1.1, 1.2, and 1.3. From Figure 5B, the presence of IFI27 as the sole BH significant gene with a logFC >1 in the ND-0-10, N-0-6, and NDL-0-strong comparisons is noteworthy. Figure 5C indicates that IFI6, HP, and IFI27 are the top three BH-significant genes with a logFC >1, common among ND-0-10, ND-3-10, and NDL-0-strong. Utilizing IFI6, IFI27, or ISG15 (only one gene) as the sole predictor in SMAS, Figure 5D displays the ROC AUC for classifying NHPs with negative RT-qPCR on day 0 vs. positive ones on day 10. Figure 5E shows a similar analysis for classifying samples as positive on day 10 and negative on day 3.

3.1.1.4 Objective 1.4. Comparison of all negative vs. all positive samples (All-N-P)

Figure 6 presents the results related to objective 1.4. Figure 6A displays a volcano plot that emphasizes the BH significant genes with a log fold change (logFC) >1, ranked by MAS, as we contrasted



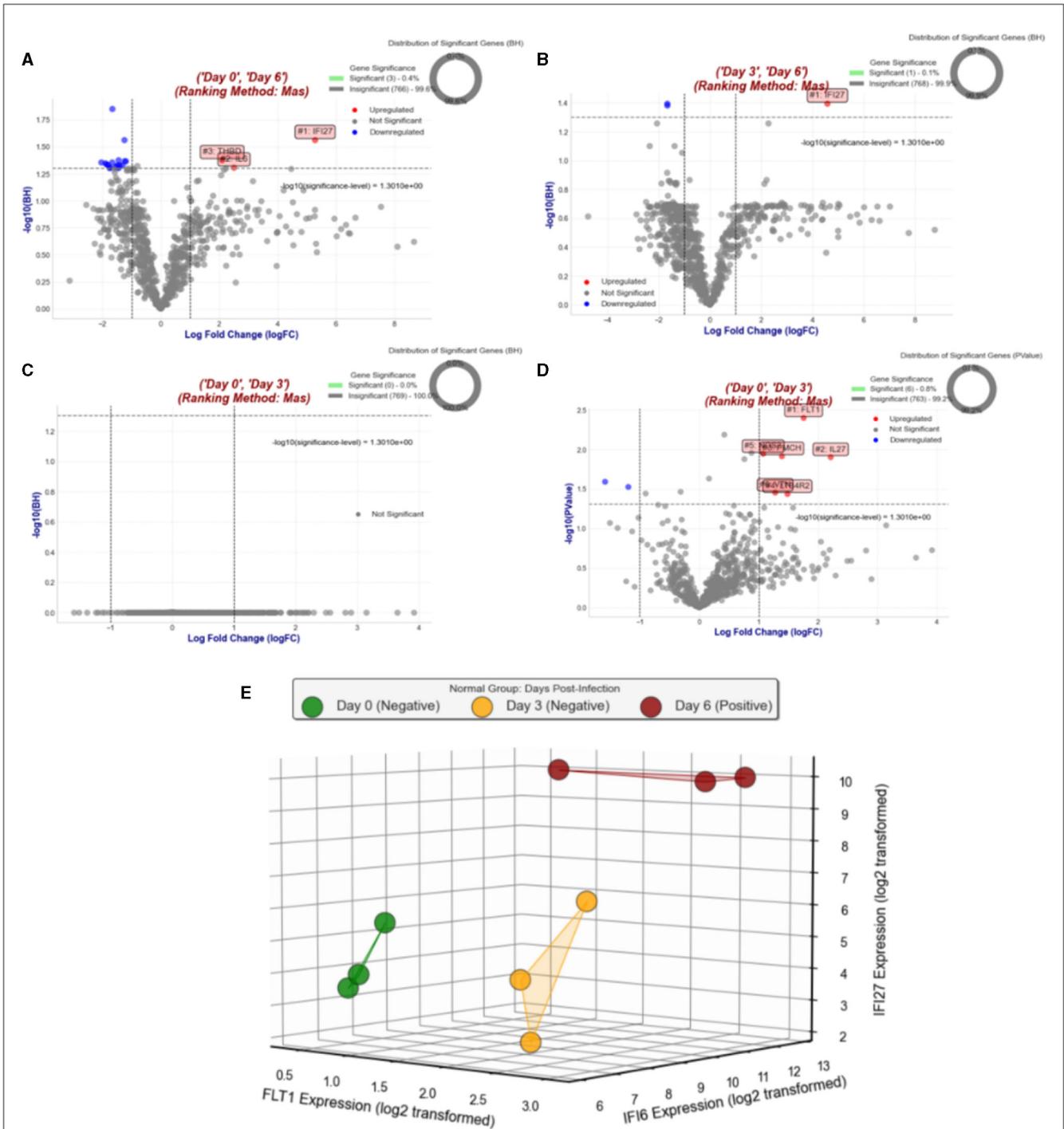


FIGURE 4 Comprehensive overview of gene expression analyses in Objective 1.2. (A) Displays a volcano plot comparing gene expressions of NHPs on day 6 vs. day 0, highlighting significant changes. (B) Shows a similar plot for day 6 vs. day 3, pinpointing early infection response genes. (C) Illustrates the lack of BH significant genes when comparing day 3 to day 0. (D) Reveals FLT1 as a top significant gene with relaxed BH criteria. Finally, (E) compares the expression patterns of IFI27 (from Objectives 1.1 and 1.2), IFI6 (from Objective 1.1), and FLT1, emphasizing their roles across different infection stages.

all positive NHPs against all negative NHPs for the RT-qPCR test. Figure 6B illustrates the ROC AUC for five-fold stratified cross-validation using only the top gene, OAS1, as the predictor for the Simplified MAS (SMAS). Figure 6C depicts the visualization of all samples using the first three principal components (PC1, PC2, and PC3) as axes. Meanwhile, Figure 6D shows the 3D visualization of all samples using the top three MAS-selected genes.

3.1.2 Objective 2. Application of MAS insights for EBOV classification

3.1.2.1 Objective 2.1. Classification using top selected MAS genes

Figure 7 demonstrates the results related to Objective 2. Since it was challenging to display all common and unique selected MAS genes for All-N-P, NDL-0-strong, ND-0-10, and

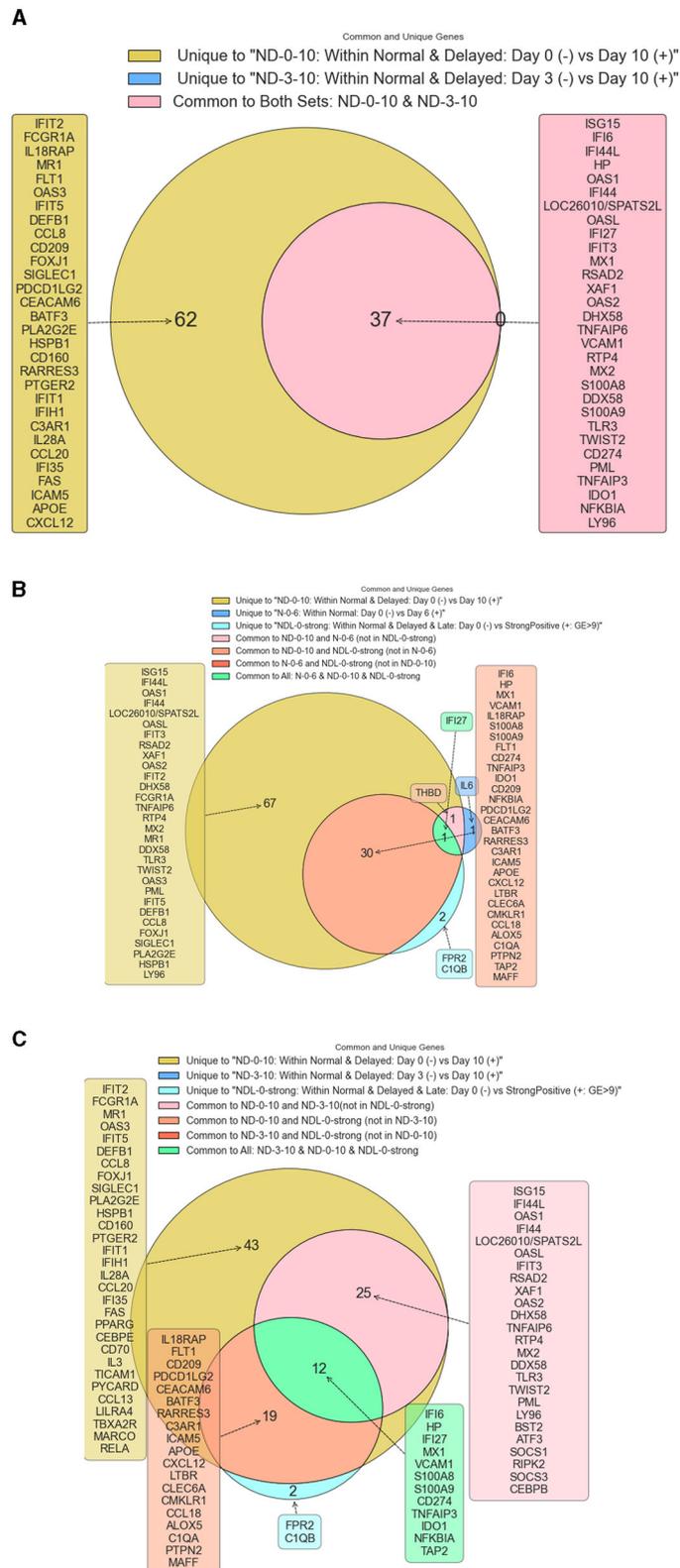


FIGURE 5
(Continued)

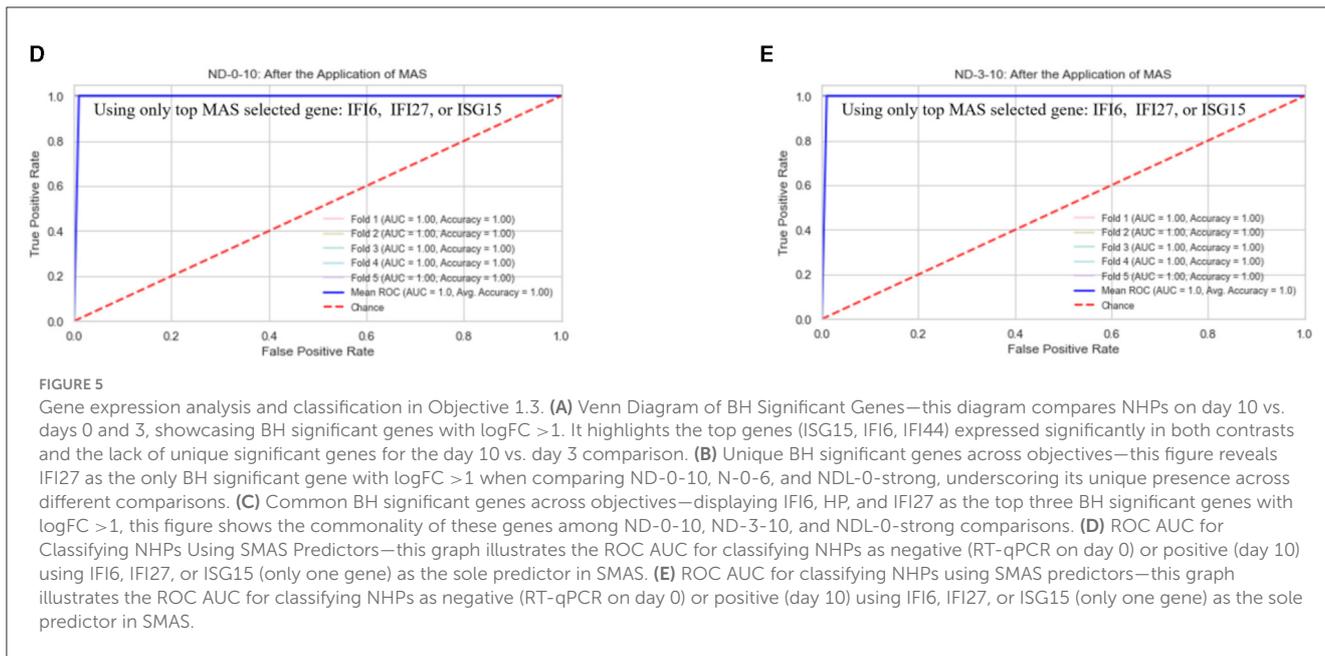


FIGURE 5

Gene expression analysis and classification in Objective 1.3. (A) Venn Diagram of BH Significant Genes—this diagram compares NHPs on day 10 vs. days 0 and 3, showcasing BH significant genes with $\log_{FC} > 1$. It highlights the top genes (ISG15, IFI6, IFI44) expressed significantly in both contrasts and the lack of unique significant genes for the day 10 vs. day 3 comparison. (B) Unique BH significant genes across objectives—this figure reveals IFI27 as the only BH significant gene with $\log_{FC} > 1$ when comparing ND-0-10, N-0-6, and NDL-0-strong, underscoring its unique presence across different comparisons. (C) Common BH significant genes across objectives—displaying IFI6, HP, and IFI27 as the top three BH significant genes with $\log_{FC} > 1$, this figure shows the commonality of these genes among ND-0-10, ND-3-10, and NDL-0-strong comparisons. (D) ROC AUC for Classifying NHPs Using SMAS Predictors—this graph illustrates the ROC AUC for classifying NHPs as negative (RT-qPCR on day 0) or positive (day 10) using IFI6, IFI27, or ISG15 (only one gene) as the sole predictor in SMAS. (E) ROC AUC for classifying NHPs using SMAS predictors—this graph illustrates the ROC AUC for classifying NHPs as negative (RT-qPCR on day 0) or positive (day 10) using IFI6, IFI27, or ISG15 (only one gene) as the sole predictor in SMAS.

ND-3-10, Figure 7A only shows the common and unique genes for the first three groups. However, the top-3 genes for common genes among all four groups (All-N-P, NDL-0-strong, ND-0-10, and ND-3-10) are IFI6, IFI27, and MX1, respectively. Figures 7B, C illustrate the expression of these top three selected genes through a 3D visualization and heat map clustering, respectively. Figure 7C reveals the hierarchical clustering where the “Ward” method is employed (Ward, 1963). This agglomerative clustering approach starts with each sample as a separate cluster and iteratively merges them into larger ones. The Ward method aims to minimize the total within-cluster variance, thereby tending to create more evenly sized, spherical clusters. Using these three genes, the method effectively separates the positive and negative groups from each other.

When we repeat this process using the traditional gene ranking commonly used (genes with $\log_{FC} > 1$ and then prioritizing based on BH adjusted p -values, so that the gene with the minimum BH adjusted p -value and $\log_{FC} > 1$ becomes the top gene in the traditional method), it turns out that the top gene is FLT1. Our aim is to pinpoint a gene capable of differentiating between positive and negative RT-qPCR NHP samples. When focusing on the extreme cases within NDL-0-strong, FLT1 emerges as the top-ranked gene using the traditional ranking method, yet it falls to rank 16 when assessed with the MAS score. While FLT1 proves to be a reliable gene marker within the context of extreme cases in NDL-0-strong, its effectiveness wanes in other scenarios. Table 1, which demonstrates the performance of logistic regression via four-fold cross-validation for All-N-P, indicates that FLT1 is not the most suitable gene for differentiating between negative and positive samples. Using IFI6 as the only predictor and logistic regression as the modeling approach, coupled with five-fold stratified cross-validation, we attained 100% accuracy and an AUC of 100% for objectives 1.1, 1.3 and 1.4 using balanced data.

3.1.2.2 Objective 2.2. Implementing MAS on an independent set for subsequent supervised classification

To clearly illustrate the differences between traditional ranking methods and the MAS approach, Figure 8 compares the top 20 genes selected by each method on ND-0-strong. Following the strategy outlined in Section 2 under Objective 2.2, the gene identified as IFI27 emerges as the top selection, G_{MAS} , from the MAS process. Utilizing IFI27 as the primary predictor in logistic regression, we conducted four-fold cross validation with all negative and positive samples, excluding those from ND-0-strong, as shown in Figure 2 (Objective 2.2). This process was also replicated using the traditional ranking method for comparison. Table 2 presents the performance of the logistic regression model, comparing the results obtained using the top-selected genes from both MAS and the traditional method.

3.2 Comprehensive Gene Ontology analysis of identified Benjamini-Hochberg significant genes through SMAS

Figure 9 illustrates the five most significant GO terms identified from the Gene Ontology analysis of upregulated genes through the SMAS method, following the Benjamini-Hochberg correction for multiple testing. Figure 10 presents the top 10 Gene Ontology (GO) terms for six crucial genes identified in our study: IFI27, IFI6, MX1, HP, OAS1, and ISG15. Each panel displays the GO terms ranked by their significance, providing insights into the primary biological functions influenced by each gene.

4 Discussion

In our study, the comprehensive differential expression analysis using MAS elucidated significant changes in gene expression due to

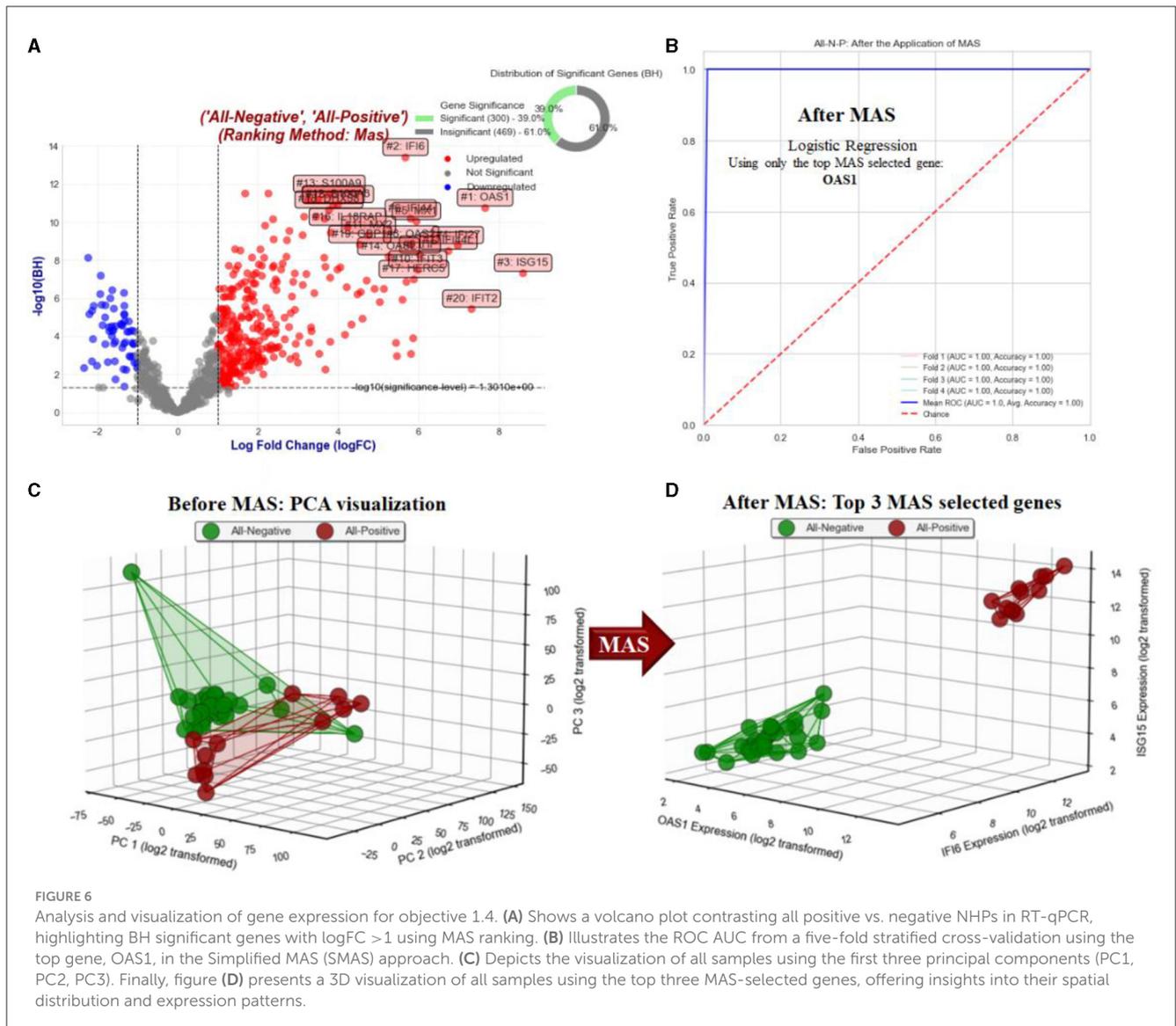


FIGURE 6
 Analysis and visualization of gene expression for objective 1.4. (A) Shows a volcano plot contrasting all positive vs. negative NHPs in RT-qPCR, highlighting BH significant genes with logFC > 1 using MAS ranking. (B) Illustrates the ROC AUC from a five-fold stratified cross-validation using the top gene, OAS1, in the Simplified MAS (SMAS) approach. (C) Depicts the visualization of all samples using the first three principal components (PC1, PC2, PC3). Finally, figure (D) presents a 3D visualization of all samples using the top three MAS-selected genes, offering insights into their spatial distribution and expression patterns.

EBOV infection in nonhuman primates (NHPs). Below, we discuss the results obtained in Section 3 for different objectives.

4.1 Objective 1. Comprehensive differential expression analysis using MAS scoring

4.1.1 Objective 1.1. Strong negative vs. strong positive analysis (NDL-0-strong)

The “Strong Negative vs. Strong Positive Analysis (NDL-0-strong)” within our study has proven to be a critical component in discerning the genetic alterations associated with extreme cases of EBOV infection. Utilizing the MAS method, we successfully identified key genes, notably IFI27, IFI6, and HP, which displayed significant expression level changes in strong positive cases as opposed to strong negatives. This analysis, enriched by three-dimensional visualization as illustrated in Figure 3B and enhanced

by logistic regression with five-fold stratified cross-validation (see Figure 3C), underscores the strength of our approach.

The precision and effectiveness of our model are most notably reflected in the ROC curves, which exhibit a remarkable 100% AUC and accuracy metrics. This exceptional performance is largely attributed to our focused reliance on a single, highly indicative gene, IFI27. Such a focused approach not only addresses the complexity of EBOV pathogenesis but also reveals a clear genetic signature. This signature is of paramount importance, presenting itself as an instrumental asset for the future development of diagnostic and therapeutic strategies targeting EBOV infections. The attainment of 100% accuracy and AUC in our analysis highlights the potential of our methodological approach in revolutionizing the understanding and management of severe viral infections.

Supporting this selection of IFI27, several studies have provided insights into its role in different viral infections. Huang et al. (2022) highlighted IFI27 as a gene associated with the progression of HIV infection, suggesting its potential as a target for immunotherapy. Similarly, Shojaei et al. (2023) emphasized the expression of

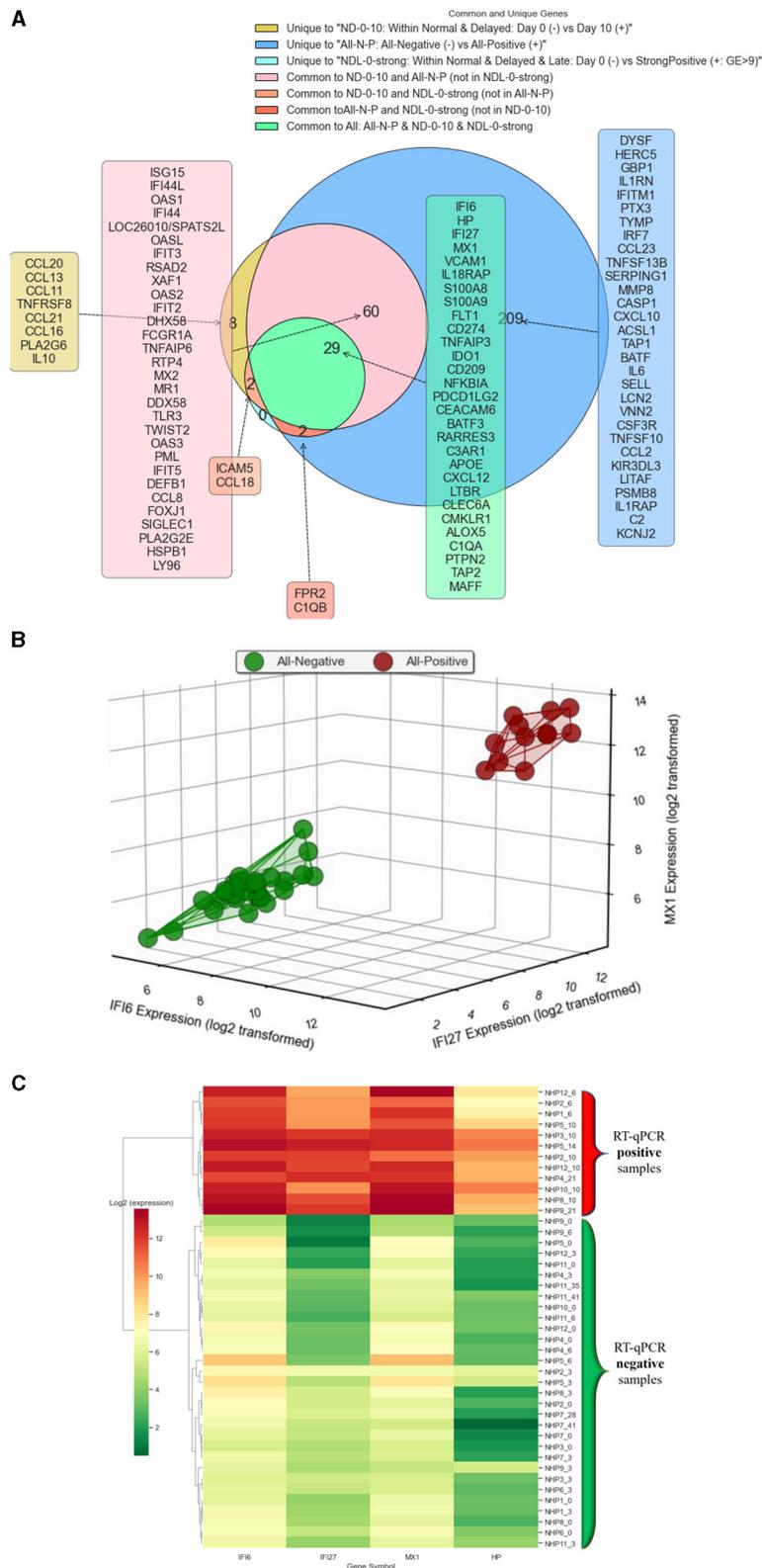


FIGURE 7 Gene expression analysis and predictive modeling in Objective 2. **(A)** The figure focuses on the selected MAS genes for All-N-P, NDL-0-strong, ND-0-10, and ND-3-10. Due to the complexity of presenting all common and unique genes across these groups, this figure specifically illustrates only those for the first three groups. However, it highlights IFI6, IFI27, and MX1 as the top three common genes shared among all four groups, underscoring their significance across various analysis categories. **(B)** The figure presents a 3D visualization of gene expression, focusing on the top three selected genes: IFI6, IFI27, and MX1. **(C)** The figure displays a heat map with hierarchical clustering of the top three selected genes: IFI6, IFI27, MX1, and HP.

TABLE 1 Comparative analysis of gene selection methods for RT-qPCR NHPs.

Ranking method	Top-selected gene	Average AUC	Average accuracy	Average precision	Average recall	Average F1-score
MAS (ours)	IFI6	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Traditional [EdgeR (Robinson et al., 2010) or DESeq2 (Love et al., 2014)]	FLT1	0.88 ± 0.07	0.82 ± 0.05	0.87 ± 0.06	0.80 ± 0.06	0.81 ± 0.06

The table highlights the efficacy of the MAS method in identifying IFI6 as a gene capable of perfectly distinguishing between all negative and positive samples, achieving 100% accuracy and AUC. This contrasts with the traditional method, which does not yield such definitive separation. Notably, due to the imbalance in the All-N-P dataset, we randomly selected 12 negative samples to form a balanced dataset comprising 24 samples (12 per class). Following this, we conducted four-fold stratified cross-validation, repeated 100 times, to calculate the average performance metrics with their standard deviations (STD) for each method.

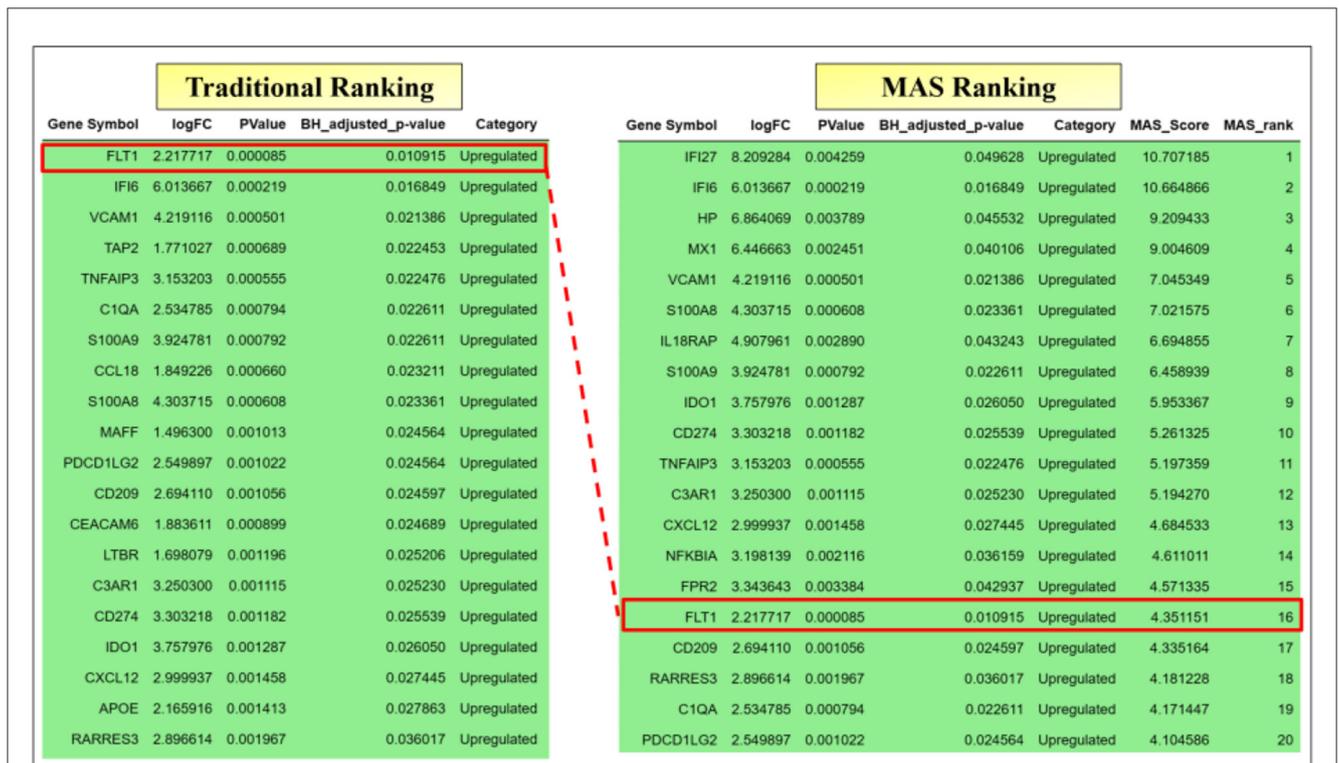


FIGURE 8 Comparative Analysis of Gene Selection between Traditional Ranking and MAS Methods for Objective 1.1. The figure displays the top 20 genes selected by the traditional ranking method (left) and the MAS method (right) for objective 1.1. It highlights differences in gene prioritization between the two methods, revealing potential insights into gene selection for distinguishing negative and positive samples in RT-qPCR NHPs.

IFI27 in the respiratory tract of COVID-19 patients, correlating its increased expression with a high viral load. Their findings also underscored the role of IFI27 as a marker of systemic host response, demonstrating its high sensitivity and specificity in predicting clinical outcomes. In the context of influenza, Tang et al. (2017) identified IFI27 as a single-gene biomarker with high predictive accuracy for distinguishing between influenza and bacterial infections. Their work showed that IFI27 is upregulated by TLR7 in plasmacytoid dendritic cells, more responsive to influenza virus than bacteria, and confirmed its expression in influenza patients through multiple patient cohorts.

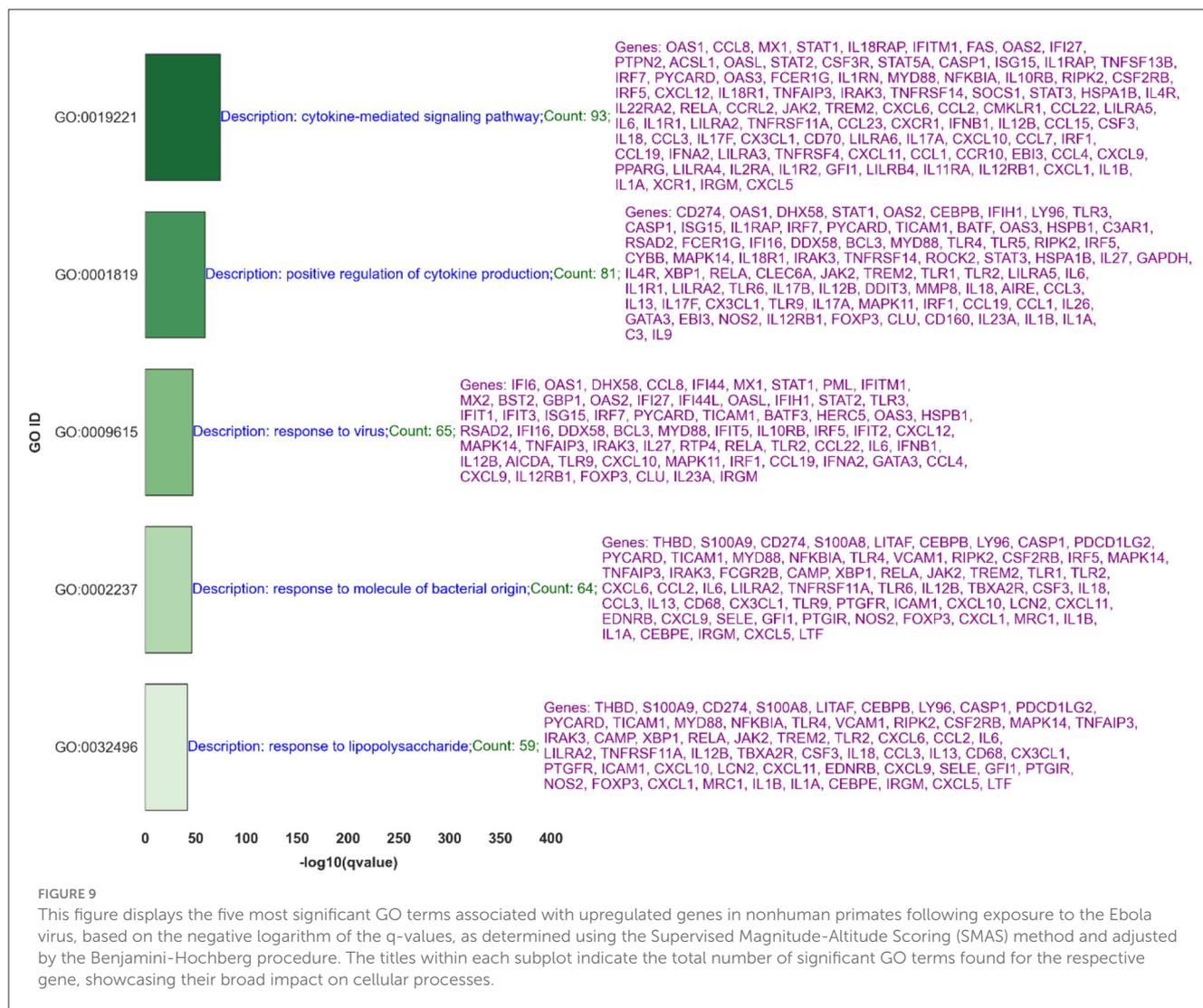
Further corroborating the relevance of IFI27, Villamayor et al. (2023) explored its role in regulating innate immune responses to viral infections. Their study highlighted a novel function of IFI27 in modulating responses triggered by cytoplasmic RNA recognition

and binding. The interaction of IFI27 with nucleic acids and the PRR retinoic acid-inducible gene I (RIG-I) was a key discovery, revealing its potential to impair RIG-I activation and thus modulate innate immune responses.

The convergence of findings from various studies establishes IFI27 as a critical gene in the context of viral infections, highlighting its significant role across different viral diseases. These insights further validate the selection of IFI27 through the MAS method, particularly within the NDJ-0-strong objective for EBOV infection in NHPs. The prominence of IFI27 in this analysis underscores its importance in deciphering the complex mechanisms of viral pathogenesis and the associated immune responses. This validation by MAS in identifying IFI27 is significant, as it indicates the gene's involvement in crucial biological pathways during severe viral infections. By focusing

TABLE 2 Performance comparison of logistic regression model using top genes selected by MAS and the traditional methods from ND-0-strong, applied to remaining samples excluding ND-0-strong.

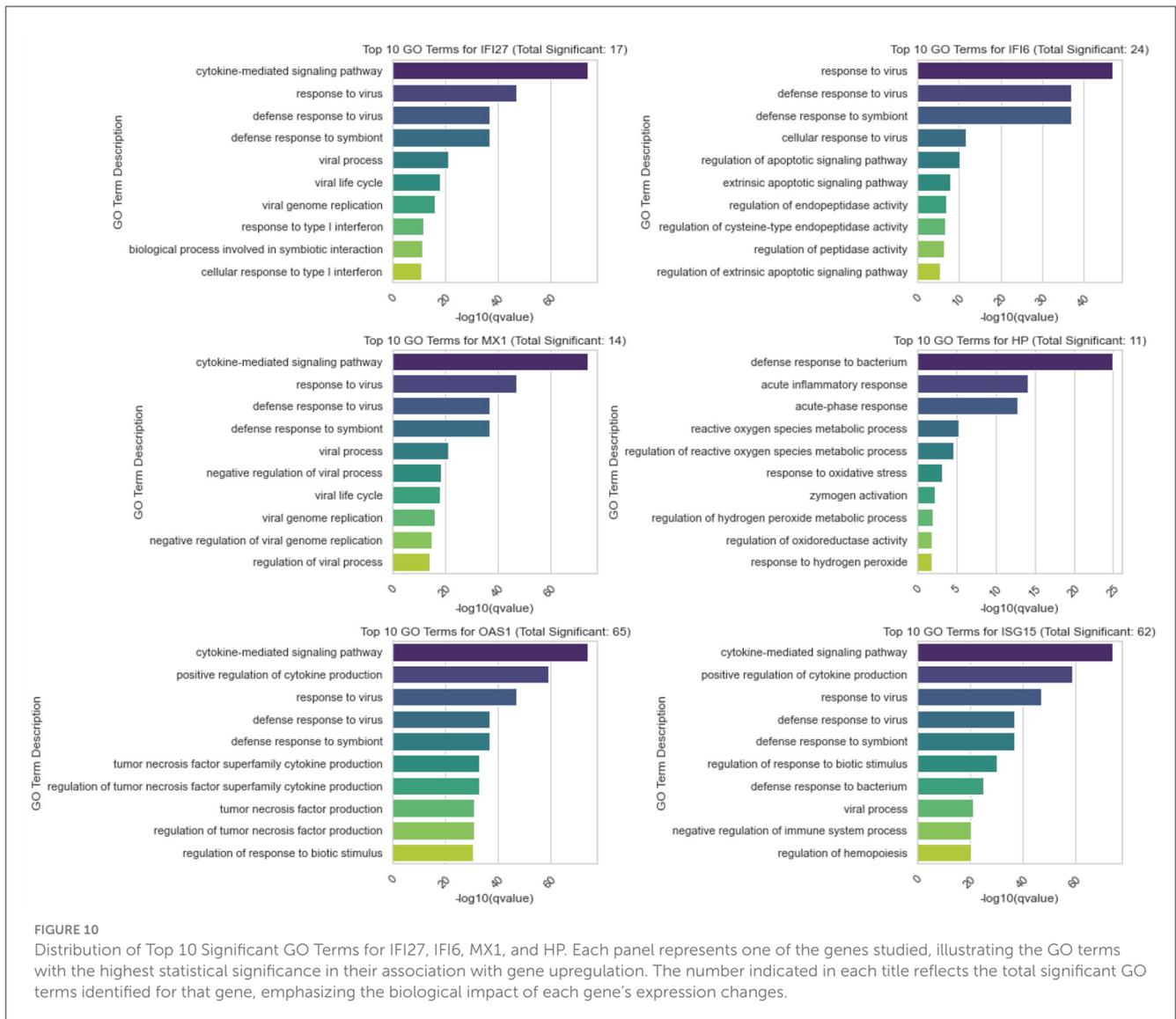
Ranking method	Top-selected gene	Average AUC	Average accuracy	Average precision	Average recall	Average F1-score
MAS (ours)	IFI27	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00	1.00 ± 0.00
Traditional [EdgeR (Robinson et al., 2010) or DESeq2 (Love et al., 2014)]	FLT1	0.72 ± 0.27	0.69 ± 0.17	0.71 ± 0.22	0.72 ± 0.14	0.67 ± 0.18



on IFI27, our study not only identifies a key genetic marker but also paves the way for a deeper understanding of the interaction networks and pathways it influences or participates in. This comprehensive approach enriches our grasp of the genetic underpinnings at play in viral infections and could potentially inform the development of targeted therapeutic interventions. The pathway analysis of IFI27, therefore, offers a more nuanced view of its role, extending beyond mere gene expression to its functional implications in the context of viral infections and immune system responses.

4.1.2 Objective 1.2. Differential expression in group 1 (N-0-6 and N-3-6)

As the positive RT-qPCR tested NHPs on day 6 post-infection are contrasted against the same NHPs on days 0 and 3 post-infection, when they were negative for RT-qPCR in Group 1, IFI27 emerges as the top MAS-selected gene (see Figure 4). Although we did not perform supervised prediction due to limited samples, Figure 4E suggests that IFI27 perfectly separates the positive and negative samples. Considering IFI27 as the top MAS-selected gene for both Objectives 1.1 (extreme cases) and 1.2 (regular



cases), it becomes evident that IFI27 is a signature marker for Ebola infection.

The findings of our study, particularly the identification of IFI27 as a key biomarker in the early stages of EBOV infection, align closely with the insights from Normandin et al.'s (2023) comprehensive analysis. Their work, which involved extensive RNA sequencing across multiple tissues in rhesus monkeys infected with EBOV (EBOV), revealed a significant correlation between the expression of IFI27 and viral RNA load. This parallel discovery in both studies underscores the crucial role of IFI27 in the host's immune response to EBOV infection. By demonstrating the early detection capability of IFI27 as a biomarker in our study and its correlation with viral load in various tissues in the study by Normandin et al., a more nuanced understanding of the molecular mechanisms of EVD pathogenesis emerges.

Blengio et al. (2023) conducted a study exploring the gene expression profiles in response to Ebola vaccination, revealing crucial insights into the immune response mechanisms. Among the key findings, IFI27 stood out, being one of the early expressed genes

whose upregulation on day 1 post-vaccination correlated with the magnitude of the antibody response observed both 21 days after the MVA-BN-Filo and 364 days after the Ad26.ZEBOV vaccinations.

Kash et al. (2017) conducted an in-depth analysis of the immune response in a patient with severe EVD, revealing critical insights into the dynamics of host gene expression in relation to the course and severity of the illness. Their study, conducted during the unprecedented 2013–2015 Ebola outbreak in Guinea, Liberia, and Sierra Leone, involved daily microarray analysis of peripheral blood samples from a patient treated at the National Institutes of Health Clinical Center. This comprehensive approach allowed them to correlate gene expression changes with various clinical parameters, including viral load, antibody responses, coagulopathy, organ dysfunction, and eventual recovery.

A key finding from Kash et al.'s (2017) research was the identification of IFI27 as a significantly expressed gene in response to EBOV replication. Their analysis revealed that IFI27, along with other type I interferon-stimulated genes (ISGs), showed marked expression changes correlating with phases of the disease.

Particularly on day 13 post-infection, IFI27, among other ISGs, was found to be highly expressed, aligning with the peak antiviral response in the patient. This aligns with our findings and those of other studies like [Normandin et al. \(2023\)](#) and [Blengio et al. \(2023\)](#), where IFI27 was identified as a critical biomarker in the context of EBOV infection and response.

As discussed in Objective 1.1, IFI27's role extends beyond being a mere biomarker. The consistent upregulation of IFI27 in Ebola-infected NHPs, as compared to their negative counterparts, underlines its significance in the pathophysiology of the infection. This pattern indicates a potential role of IFI27 in the host's response to the EBOV, possibly in the modulation of the immune system or the activation of specific cellular pathways. Given the crucial role of interferon-stimulated genes like IFI27 in the innate immune response, its significant expression in Ebola-infected NHPs could reflect an activation of defense mechanisms against the virus. This aligns with findings in other viral infections, where IFI27 is implicated in immune response modulation and viral replication control.

Therefore, the prominence of IFI27 in Ebola infection, at least in NHP models, highlights its potential as a therapeutic target or a biomarker for early diagnosis and treatment strategies. Understanding the exact mechanisms by which IFI27 influences Ebola pathogenesis could pave the way for novel approaches to manage and treat this severe viral infection. This study opens new avenues for investigating IFI27's role in EBOV infection and its potential application in clinical settings.

4.1.3 Objective 1.3. Analysis within normal and delayed groups (ND-0-10 and ND-3-10)

When we contrasted the RT-qPCR positive NHPs on day 10 against the same NHPs on days 0 and 3, when they were still negative in the RT-qPCR test, interesting results emerged. [Figure 5A](#) indicates that from day 0 to day 10, a total of 99 genes are significantly upregulated (as indicated by Benjamini-Hochberg significance) with a log fold change (logFC) >1. Furthermore, from day 3 to day 10, only 37 of these genes remained significantly upregulated. The difference in gene expression between day 3 and day 10 is particularly noteworthy. While the number of significantly upregulated genes decreases to 37, these genes (e.g., ISG15, IFI6, IFI44L, HP, OAS1, IFI44, OASL, and IFI27) likely play a crucial role in the response to the progressing infection. The reduction in the number of upregulated genes could indicate a more targeted or evolved response of the host's immune system as the infection progresses. Alternatively, it may reflect a shift in the virus-host interaction dynamics over time. This pattern of gene expression highlights the complex and evolving nature of the host response to EBOV infection. Identifying these key genes that remain upregulated over time could provide insights into critical pathways and mechanisms that the virus exploits or the host deploys in response to the infection. This understanding could be invaluable for developing targeted therapies or diagnostics for the EBOV.

[Figure 5B](#) demonstrates the common and unique MAS-selected genes with logFC >1 among the negative RT-qPCR results on day 0 compared to the positive ones on day 10 and day 6 and the strong

positives on days 10 or 21. It turns out that the only Benjamini-Hochberg (BH) significant gene with logFC >1, common among all these three groups, is IFI27. When the N-0-6 group is replaced with ND-3-10, there are 12 BH significant genes with logFC >1, with IFI6, HP, IFI27, and MX1 as the top selected genes (see [Figure 5C](#)). Using either IFI6, IFI27, or ISG15 as the single predictor for logistic regression, we achieved 100% accuracy in separating the positive NHPs on day 10 from their counterparts on days 0 or 3 (see [Figures 5D, E](#)). This analysis provides crucial insights into the gene expression profiles in NHPs associated with EBOV infection. IFI27's consistent upregulation across different groups and time points indicates its potential as a biomarker for Ebola infection. Its presence in all examined groups underscores its importance in early virus detection. The ability to differentiate between positive and negative samples with 100% accuracy using IFI6, IFI27, or ISG15 as predictors highlights the significant alteration of these genes following infection and their potential as reliable indicators of EBOV presence in NHPs.

Building on this, several studies have explored the roles of related genes in viral infections. [Sajid et al. \(2021\)](#) focused on IFI6 in HBV, demonstrating its antiviral properties post-type I IFN-alpha stimulation. Similarly, [Qi et al. \(2015\)](#) showed that IFI6 expression increases in DENV-infected cells, influencing apoptosis-related processes. [Liu et al. \(2019\)](#) studied lncRNA-IFI6 in HCV infection, uncovering its regulatory effects on the antiviral ISG IFI6. [Park et al. \(2013\)](#) investigated IFI6's genetic variations in chronic liver disease patients with HBV.

[Morales and Lenschow \(2013\)](#) reviewed the current understanding of ISG15, examining its role in mediating protection against different viral infections and exploring the mechanisms by which it exerts antiviral activity. Their work highlights the importance of ISG15 in the immune system's response to viral threats, emphasizing its potential as a target for therapeutic interventions in viral diseases. [Jeon et al. \(2010\)](#) focused on the role of ISG15 in the immune response, particularly its involvement in the conjugation process known as ISGylation. Their findings indicate ISG15's significant role in the innate immune response. [Perng and Lenschow \(2018\)](#) focused their review on the multifaceted role of ISG15, a ubiquitin-like protein, in the host's response to viral infection. They discussed how ISG15, induced by type I interferons, not only directly inhibits viral replication but also modulates various host responses, including damage repair, immune response, and other signaling pathways. Their review highlighted the diverse and pathogen-dependent actions of ISG15, emphasizing its importance in antiviral defense. Additionally, they explored how viruses evolve strategies to evade ISG15's actions. The review also integrated new findings on individuals deficient in ISG15 and the identification of a cellular receptor for ISG15, providing deeper insights into how ISG15 influences the host response to viral infections.

These studies collectively enhance our understanding of the roles of genes like IFI6 and ISG15 in the body's response to viral infections, providing valuable context for our findings on EBOV. The mechanisms elucidated in these studies, particularly relating to IFI6's antiviral properties and ISG15's role in immune modulation and viral replication inhibition, offer insights that may be relevant in understanding the specific genetic responses we observed in NHPs infected with Ebola.

4.1.4 Objective 1.4. Comparison of all negative vs. all positive samples (All-N-P)

When contrasting all RT-qPCR positive tested NHPs against all negative ones, regardless of time, OAS1 emerged as the top MAS-selected gene, perfectly separating negative from positive samples with 100% accuracy (see Figure 6). The second and third top MAS-selected genes are IFI6 and ISG15, consistent with our findings in previous objectives.

Melchjorsen et al. (2009) investigated the expression patterns of the 2'-5' oligoadenylate synthetase (OAS) family genes, particularly Oligoadenylate synthetase 1 (OAS1) and OASL, during viral infections such as with Sendai virus and Influenza A virus. Their findings suggest that OASL behaves like an antiviral gene, providing new insights into the roles of the OAS gene family members in the immune response to viral infections. Fish and Boissinot (2016) conducted a comprehensive study on OAS1, highlighting its critical role as a first line of defense against various viral pathogens, particularly in Old World monkeys. Their research revealed that OAS1 is evolving under positive selection in these species, with most of the positively selected sites located in the RNA-binding domain (RBD), which is responsible for binding viral dsRNA. This positive selection, especially concentrated in a specific region of the RBD, suggests a sub-functionalization within this domain, potentially enhancing OAS1's ability to recognize and bind diverse viral RNAs. Additionally, Fish and Boissinot (2016) identified positively selected residues around the active site's entry, indicating an evolutionary adaptation to possibly evade viral antagonism or to produce oligoadenylates of varying lengths. These findings underscore the evolutionary pressures shaping OAS1 and its importance in the innate immune response to viral infections.

Building upon these studies, OAS1's pathway in viral infection response can be delineated. OAS1 is activated upon recognizing double-stranded RNA (dsRNA) from viruses, including Ebola, leading to the synthesis of 2'-5'-linked oligoadenylates (2-5A). This synthesis subsequently activates Ribonuclease L (RNase L), which then cleaves both viral and host RNA, effectively inhibiting viral replication. However, this activation by OAS1 is a double-edged sword, as indicated by Carey et al.'s (2019) findings show that OAS1 activation can also lead to autoactivation by host RNAs, presenting potential costs to the host. The evolutionary study by Fish and Boissinot (2016) added another layer to this pathway by suggesting that OAS1's RNA-binding domain has evolved under positive selection in some primate species, enhancing its ability to recognize and bind diverse viral RNAs. This evolutionary adaptation may be a response to the need for balancing effective viral defense with the minimization of collateral damage to the host. In the context of EBOV infection, this pathway underscores OAS1's critical role in the initial detection and response to the virus, marking it as a key player in the host's antiviral defense mechanism.

4.2 Objective 2. Application of MAS insights for EBOV classification

4.2.1 Objective 2.1. Classification using top selected MAS genes

Objective 2.1 of our study focused on identifying a definitive gene marker for distinguishing between positive and negative

NHPs in the RT-qPCR test, irrespective of the severity or level of Genome Equivalent. Our approach involved selecting the top MAS-selected gene with a logFC > 1, consistently identified across objectives 1.1, 1.2, and 1.4. It turns out that IFI6, IFI27, and MX1 are the top common genes across all objectives (Figure 7A shows the common genes across three groups). According to Figures 7B, C, using any of these genes effectively separates the positive and negative samples into distinct clusters.

To evaluate the effectiveness of the MAS ranking method, we compared it with the traditional ranking method. We found that the top MAS-selected gene, IFI6, successfully differentiated all negative from positive samples with 100% average accuracy and AUC. In contrast, the top gene selected by the traditional ranking method, FLT1, achieved an average accuracy of 82% and an average AUC of 0.88, as indicated in Table 1. Additionally, we compared the MAS and traditional ranking mechanisms within the extreme cases of Objective 1.1. As shown in Figure 8, the MAS approach effectively accounts for both biological and statistical significance, highlighting its superiority in identifying key genetic markers in EBOV infection studies.

We have discussed the pathways of IFI6 and IFI27 in previous objectives. Regarding MX1, Caballero et al. (2016) focused on the host response to EBOV infection, revealing that MX1, along with ISG15 and OAS1, is significantly upregulated in circulating immune cells during infection. Their study found that these genes, particularly MX1, are part of a strong innate immune response triggered by active virus replication. This response, which also includes other interferon-stimulated genes, contrasts with *in vitro* evidence suggesting suppression of innate immune signaling. The prominence of MX1 in the immune response to the EBOV highlights its potential role in the pathogenesis of the infection and underscores its significance as a key component of the host's defense mechanism against viral threats. Pillai et al. (2016) highlighted the critical role of MX1 in the immune response to viral infections (particularly to IAV) and underscored the complex interplay between viral defense mechanisms and subsequent bacterial complications. Fuchs et al. (2017) investigated the role of MX1 proteins in bats, particularly in relation to their innate immune defense against various viruses, including Ebola. Bats, known reservoirs for zoonotic viruses such as Ebola, typically do not exhibit clinical symptoms from these infections. They focused on cloning MX1 cDNAs from three bat families and analyzing their antiviral potential. They found that bat MX1 proteins are key factors in controlling viral replication, including Ebola, in their bat hosts, and offer insights into the coevolution of these proteins with bat-borne viruses.

4.2.2 Objective 2.2. Implementing MAS on an independent set for subsequent supervised classification

The results presented in Table 2 underscore the remarkable efficiency of the MAS method in gene selection for predictive modeling, especially when contrasted with traditional ranking methods. The perfect scores across all metrics achieved by the logistic regression model using the MAS-selected gene IFI27, including Average AUC, Accuracy, Precision, Recall, and F1-Score, firmly establish MAS as a highly effective approach

in this context. These scores not only reflect the method's impeccable accuracy in classifying RT-qPCR positive and negative samples but also its consistency, as indicated by the absence of variability in the results. This level of performance highlights the potential of MAS to reliably identify genes that are not just statistically significant but also practically relevant for diagnostic and therapeutic applications, particularly in the field of EBOV research.

In stark contrast, the traditional ranking method, as evidenced by its moderate scores using the gene *FLT1*, demonstrates a less robust performance. While it maintains a certain level of effectiveness, the considerable variability in its metrics suggests lower reliability and precision. This comparison draws attention to the limitations inherent in traditional methods, which often prioritize statistical significance over practical utility. The discrepancies in performance between the two methods may stem from their differing focuses: MAS balances statistical and biological significance, while traditional methods may disproportionately emphasize statistical metrics, potentially overlooking genes with substantial biological relevance.

A key finding from our analysis was the identification of *IFI27* as a prominent gene in both extreme and regular cases of EBOV infection. *IFI27* consistently emerged as the top MAS-selected gene, establishing itself as a robust marker for the presence of the virus. This was further corroborated by its exceptional performance in logistic regression models, where it demonstrated high accuracy in differentiating between positive and negative samples. Moreover, our study revealed the significant role of genes like *IFI6*, which, alongside *IFI27*, showed marked expression level changes in response to the infection. These genes formed a distinct pattern of expression, aligning with the progression and severity of the infection in NHPs.

We also observed that other genes, such as *MX1* and *ISG15*, were significantly upregulated, suggesting their involvement in the host's immune response to the EBOV. The pattern of upregulation in these genes provided insights into the complex interplay between the host's defense mechanisms and viral pathogenesis. In a broader classification analysis, we found that genes like *OAS1* also played a significant role in our study. *OAS1* was identified as a crucial marker, effectively separating positive from negative EBOV samples with high accuracy. This underscores its potential role in the innate immune response and its importance as part of the genetic signature of EBOV infection.

In summary, the application of SMAS has successfully highlighted six key genes, *IFI27*, *IFI6*, *MX1*, *HP*, *OAS1*, and *ISG15* as critically involved in the response to Ebola virus infection in nonhuman primates (NHPs). Note that the integration of both the magnitude of expression changes (logFC) and their statistical significance (adjusted *p*-values) in the MAS formula is designed to overcome some of the common pitfalls associated with traditional gene expression analysis methods. While large fold changes in gene expression are used as indicators of potential biological importance, our methodology critically evaluates these changes within the statistical framework provided by the Benjamini-Hochberg correction. This dual consideration ensures that the genes identified as significant are not only statistically robust but also likely to be biologically meaningful in the context of

EBOV infection. Additionally, by incorporating a comprehensive Gene Ontology (GO) analysis of the top-ranked genes, we aim to validate their functional relevance and connectivity in biological pathways directly related to viral pathogenesis. This GO analysis helps elucidate the broader biological implications of these gene expression changes, supporting their role in EBOV infection beyond mere statistical artifacts.

The results of our GO term analysis, detailed in [Figures 9, 10](#), not only corroborate the biological significance of these genes but also illuminate their interconnected roles in viral pathogenesis and host defense mechanisms. Below, we delve into the specifics of the GO terms associated with each gene, elucidating how they collectively contribute to a comprehensive understanding of their function in the context of Ebola.

- *IFI27* has been pinpointed as a central gene in the immune response to viral infection, as evidenced by its involvement in the “cytokine-mediated signaling pathway” (GO:0019221) and “response to type I interferon” (GO:0034340). These pathways are crucial for activating antiviral states in cells, a key defense mechanism against Ebola virus proliferation. The gene's significant roles in “viral process” (GO:0016032) and “viral genome replication” (GO:0019079) suggest a direct interaction with viral components, influencing the lifecycle of the Ebola virus within host cells. These interactions underscore the potential of *IFI27* to serve as a biomarker for disease progression and severity.
- The involvement of *IFI6* in “response to virus” (GO:0009615) and “defense response to virus” (GO:0051607) highlights its function in initiating and modulating antiviral responses. This gene's connection with the “regulation of apoptotic signaling pathway” (GO:2001233) and “extrinsic apoptotic signaling pathway” (GO:0097191) points to its role in controlling cell death during infection, a critical factor in limiting Ebola virus spread by preventing premature cell demise. *IFI6*'s participation in these processes enhances the host's ability to maintain cellular integrity while fighting the infection, illustrating its comprehensive role in viral defense.
- *MX1* is robustly linked to several key antiviral processes, including “defense response to virus” (GO:0051607) and “negative regulation of viral genome replication” (GO:0045071). By impacting the “viral life cycle” (GO:0019058) and actively participating in “cytokine-mediated signaling pathways” (GO:0019221), *MX1* emerges as a pivotal factor in the immune system's strategy to combat Ebola. These GO terms collectively underscore *MX1*'s role in not only halting the progression of viral infection but also in modulating the immune response to enhance efficacy and specificity.
- *HP*'s association with “acute inflammatory response” (GO:0002526) and “acute-phase response” (GO:0006953) identifies it as a key player in the initial immune response to Ebola virus infection. The gene's involvement in “response to oxidative stress” (GO:0006979) and “regulation of reactive oxygen species metabolic process” (GO:2000377) further suggests its role in managing cellular stress caused by viral invasion. These activities are crucial in maintaining cellular

homeostasis and enhancing the resilience of host cells against the destructive nature of Ebola virus replication.

- OAS1 plays a significant role in the immune response to viral infections, particularly evident from its involvement in multiple GO terms related to immune processes. The “cytokine-mediated signaling pathway” (GO:0019221) and the “positive regulation of cytokine production” (GO:0001819) highlight its crucial function in modulating cytokine responses, which are essential for initiating and sustaining effective antiviral responses against pathogens like the Ebola virus. The gene’s participation in the “response to virus” (GO:0009615) and “defense response to virus” (GO:0051607) underscores its direct engagement in combating viral infections. These pathways are pivotal in orchestrating a coordinated immune response, enhancing the host’s ability to suppress viral replication and spread. Furthermore, OAS1’s role extends to the regulation of complex immune pathways, as shown in the “regulation of response to biotic stimulus” (GO:0002831) and involvement in tumor necrosis factor-related activities such as “tumor necrosis factor production” (GO:0032640) and its regulation (GO:0032680). These interactions suggest that OAS1 not only responds to viral infections but also influences broader immune regulatory networks that control inflammation and cellular signaling in response to pathogens. This comprehensive involvement makes OAS1 a key player in the immune defense against Ebola, contributing to both the immediate antiviral response and the regulation of systemic immune reactions that are critical during severe viral infections.
- ISG15 stands out for its extensive involvement in the immune response, particularly through its association with the “cytokine-mediated signaling pathway” (GO:0019221) and “positive regulation of cytokine production” (GO:0001819). These roles are integral to amplifying the immune system’s response to viral invasion, facilitating rapid and effective defense mechanisms. Additionally, ISG15’s involvement in the “response to virus” (GO:0009615) and “defense response to virus” (GO:0051607) highlights its direct role in antiviral defense, particularly in mediating the immune system’s reaction to viral pathogens. Moreover, ISG15’s functions in “negative regulation of immune system process” (GO:0002683) and its regulation of hemopoiesis (GO:1903706) suggest a balancing role in immune modulation, preventing overactivation that could lead to pathology. This dual role in both promoting and regulating immune responses ensures a measured and effective antiviral response, which is crucial in the management of Ebola virus infections. The involvement in “viral process” (GO:0016032) and specifically in the “defense response to bacterium” (GO:0042742) also indicates its broader antimicrobial roles, which may be critical in scenarios where secondary bacterial infections complicate viral disease courses.

Our findings highlight the power of the MAS method in unraveling the complex genetic responses to EBOV infection. By identifying key genes that are significantly altered during infection, our study not only contributes to the understanding of EBOV pathogenesis but also paves the way for developing targeted

diagnostic and therapeutic strategies. The distinct gene signatures we identified could serve as crucial tools in managing and treating EBOV infections, providing a foundation for future research in this critical area of virology. The integration of these GO terms provides a nuanced view of how each gene contributes to the host’s defense against Ebola, reflecting the complex interplay of immune response, viral adaptation, and host survival strategies. By leveraging the SMAS methodology, our study not only identifies genes with statistical significance but also elucidates their profound biological relevance to Ebola virus infection, paving the way for future research into targeted therapeutic interventions and enhanced diagnostic techniques.

5 Conclusions

This study marks a pivotal advancement in the field of virology, particularly in understanding and detecting Ebola virus (EBOV) infection in nonhuman primates. By employing the novel Supervised Magnitude-Altitude Scoring (SMAS) methodology, tailored specifically for multi-omic data with limited samples, we have successfully identified crucial biomarkers. This approach, focusing on a balanced integration of statistical and biological significance, allows for the effective use of NanoString gene expression data, which, while not exhaustive like RNA-seq, is highly targeted and informative for known genes of interest.

Notably, the SMAS methodology enhances traditional gene expression analysis by incorporating both log fold change (logFC) and adjusted *p*-values from the Benjamini-Hochberg correction into the gene selection process. This dual consideration ensures that selected genes are not only statistically significant but also biologically relevant, making them particularly suitable for predictive modeling in infectious disease research.

Key genes such as IFI6, IFI27, MX1, OAS1, and ISG15 were identified as significant, demonstrating exceptional predictive performance with perfect AUC metrics in classifying various stages of EBOV infection. These genes play critical roles in the immune response to EBOV, as confirmed by comprehensive Gene Ontology (GO) analysis. The GO analysis highlighted their involvement in vital biological processes and immune responses, such as cytokine-mediated signaling and the regulation of antiviral mechanisms, which are essential for combating EBOV proliferation.

The increased expression of these genes and their detailed functional roles elucidated through GO terms provide profound insights into the intricate dynamics of host defense mechanisms. These findings not only underscore the biological importance of the identified genes but also lay a robust groundwork for developing refined diagnostic methods and therapeutic approaches against EBOV infection.

In summary, the application of the SMAS methodology within this study has not only showcased its capability to handle complex genomic data efficiently but also its potential to significantly contribute to virological diagnostics and therapeutic research. By effectively identifying key biomarkers through a method that accommodates the constraints of limited sample sizes typically associated with multi-omic data, SMAS stands out as a highly promising tool in the arsenal against viral infections such as EBOV. This study paves the way for future research into targeted

therapeutic interventions and enhanced diagnostic techniques, which are crucial for managing and treating Ebola and potentially other viral infections.

6 Study limitations and future work

Our findings, while robust in the context of Ebola virus infection, provide foundational insights that necessitate cautious interpretation, especially when considering their applicability to human infections and other pathogens. This caution forms a cornerstone of our ongoing and future research efforts. Key points to consider include:

- **Applicability to human infection:** the results derived from nonhuman primate models serve as preliminary insights that need validation in human subjects. The transition from nonhuman primate models to human studies is a critical step that requires careful consideration and methodological adjustments to ensure relevance and accuracy.
- **Sample size and generalizability:** the study's current limitations imposed by our small sample size affect the generalizability of our findings. To address this, we plan to expand our sample pool in future studies. Increasing the number and diversity of samples will enhance the robustness and applicability of our results, allowing for a more comprehensive evaluation of our methodologies and findings.
- **Long-term expression changes and functional analyses:** our focus thus far has been on short-term gene expression changes. Recognizing the importance of understanding the prolonged effects and functional roles of these genes, future work will integrate long-term gene expression data and functional assays.
- **Expansion to other infectious diseases:** to further validate and possibly refine the specificity of our identified biomarkers, we plan to extend our methodologies to include control studies involving other infectious diseases. These comparative studies will help determine if the biomarkers identified are specific to Ebola virus infection or represent a generic response to various pathogens.

Data availability statement

Publicly available datasets were analyzed in this study. This data can be found at: <https://www.science.org/doi/10.1126/scitranslmed.aaq1016>.

References

- Andersen, C. L., Jensen, J. L., and Ørntoft, T. F. (2004). Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res.* 64, 5245–5250. doi: 10.1158/0008-5472.CAN-04-0496
- Benjamini, Y., and Hochberg, Y. (1995). Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc. B* 57, 289–300. doi: 10.1111/j.2517-6161.1995.tb02031.x

Author contributions

MR: Formal analysis, Investigation, Methodology, Software, Validation, Visualization, Writing – original draft, Writing – review & editing. MN: Conceptualization, Formal analysis, Investigation, Supervision, Writing – original draft, Writing – review & editing. HL: Formal analysis, Investigation, Methodology, Validation, Writing – review & editing, Writing – original draft. AN: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. MG: Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Effort sponsored by the U.S. Government under HDTRA 12310003, “Host signaling mechanisms contributing to endothelial damage in hemorrhagic fever virus infection,” PI: Narayanan. The US Government is authorized to reproduce and distribute reprints for Governmental purposes, notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the U.S. Government.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Blengio, F., Hocini, H., Richert, L., Lefebvre, C., Durand, M., Hejblum, B., et al. (2023). Identification of early gene expression profiles associated with long-lasting antibody responses to the Ebola vaccine Ad26. ZEBOV/MVA-BN-Filo. *Cell Rep.* 42:113101. doi: 10.1016/j.celrep.2023.113101

- Caballero, I. S., Honko, A. N., Gire, S. K., Winnicki, S. M., Melé, M., Gerhardinger, C., et al. (2016). *In vivo* Ebola virus infection leads to a strong innate response in circulating immune cells. *BMC Genomics* 17, 1–13. doi: 10.1186/s12864-016-3060-0

- Carey, C. M., Govande, A. A., Cooper, J. M., Hartley, M. K., Kranzusch, P. J., Elde, N. C., et al. (2019). Recurrent loss-of-function mutations reveal costs to OAS1 antiviral activity in primates. *Cell Host Microbe* 25, 336–343. doi: 10.1016/j.chom.2019.01.001
- Carlson, M., Falcon, S., Pages, H., and Li, N. (2019). *org. Hs. eg. db: genome wide annotation for Human*. R package version 3, 3.
- Centers for Disease Control and Prevention (2014). *Ebola outbreak in West Africa—case counts, Vol. 8*. Available at: https://www.cdc.gov/ebola/about/index.html?CDC_AAref_Val=https://www.cdc.gov/vhf/ebola/outbreaks/2014-west-africa/casecounts.html
- Fish, I., and Boissinot, S. (2016). Functional evolution of the OAS1 viral sensor: insights from old world primates. *Infect. Genet. Evol.* 44, 341–350. doi: 10.1016/j.meegid.2016.07.005
- Fuchs, J., Hölzer, M., Schilling, M., Patzina, C., Schoen, A., Hoenen, T., et al. (2017). Evolution and antiviral specificities of interferon-induced Mx proteins of bats against Ebola, influenza, and other RNA viruses. *J. Virol.* 91, 10–1128. doi: 10.1128/JVI.00361-17
- Hastie, T., Tibshirani, R., Friedman, J. H., and Friedman, J. H. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Berlin: Springer. doi: 10.1007/978-0-387-84858-7
- Huang, H., Lv, J., Huang, Y., Mo, Z., Xu, H., Huang, Y., et al. (2022). IFI27 is a potential therapeutic target for HIV infection. *Ann. Med.* 54, 314–325. doi: 10.1080/07853890.2021.1995624
- James, G., Witten, D., Hastie, T., and Tibshirani, R. (2013). *An Introduction to Statistical Learning*. Berlin: Springer. doi: 10.1007/978-1-4614-7138-7
- Jeon, Y. J., Yoo, H. M., and Chung, C. H. (2010). ISG15 and immune diseases. *Biochim. Biophys. Acta* 1802, 485–496. doi: 10.1016/j.bbadis.2010.02.006
- Kash, J. C., Walters, K.-A., Kindrachuk, J., Baxter, D., Scherler, K., Janosko, K. B., et al. (2017). Longitudinal peripheral blood transcriptional analysis of a patient with severe Ebola virus disease. *Sci. Transl. Med.* 9:eaa19321. doi: 10.1126/scitranslmed.aai9321
- Kim, T. K. (2015). T test as a parametric statistic. *Korean J. Anesthesiol.* 68, 540–546. doi: 10.4097/kjae.2015.68.6.540
- Kuhn, M., and Johnson, K. (2013). *Applied Predictive Modeling*. Berlin: Springer. doi: 10.1007/978-1-4614-6849-3
- Liu, X., Duan, X., Holmes, J. A., Li, W., Lee, S. H., Tu, Z., et al. (2019). A novel lncRNA regulates HCV infection through IFI6. *Hepatology* 69:1004. doi: 10.1002/hep.30266
- Love, M. I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* 15, 1–21. doi: 10.1186/s13059-014-0550-8
- Melchjorsen, J., Kristiansen, H., Christiansen, R., Rintahaka, J., Matikainen, S., Paludan, S., et al. (2009). Differential regulation of the OASL and OAS1 genes in response to viral infections. *J. Interferon Cytokine Res.* 29, 199–208. doi: 10.1089/jir.2008.0050
- Morales, D. J., and Lenschow, D. J. (2013). The antiviral activities of ISG15. *J. Mol. Biol.* 425, 4995–5008. doi: 10.1016/j.jmb.2013.09.041
- Normandin, E., Triana, S., Raju, S. S., Lan, T. C. T., Lagerborg, K., Rudy, M., et al. (2023). Natural history of Ebola virus disease in rhesus monkeys shows viral variant emergence dynamics and tissue-specific host responses. *Cell Genomics* 3:100440. doi: 10.1016/j.xgen.2023.100440
- Park, G.-H., Kim, K.-Y., Cho, S. W., Cheong, J. Y., Im Yu, G., Shin, D. H., et al. (2013). Association between interferon-inducible protein 6 (IFI6) polymorphisms and hepatitis B virus clearance. *Genomics Inform.* 11:15. doi: 10.5808/GI.2013.11.1.15
- Perng, Y.-C., and Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. *Nat. Rev. Microbiol.* 16, 423–439. doi: 10.1038/s41579-018-0020-5
- Pillai, P. S., Molony, R. D., Martinod, K., Dong, H., Pang, I. K., Tal, M. C., et al. (2016). Mx1 reveals innate pathways to antiviral resistance and lethal influenza disease. *Science* 352, 463–466. doi: 10.1126/science.aaf3926
- Qi, Y., Li, Y., Zhang, Y., Zhang, L., Wang, Z., Zhang, X., et al. (2015). IFI6 inhibits apoptosis via mitochondrial-dependent pathway in dengue virus 2 infected vascular endothelial cells. *PLoS ONE* 10:e0132743. doi: 10.1371/journal.pone.0132743
- Robinson, M. D., McCarthy, D. J., and Smyth, G. K. (2010). edgeR: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26, 139–140. doi: 10.1093/bioinformatics/btp616
- Sajid, M., Ullah, H., Yan, K., He, M., Feng, J., Shereen, M. A., et al. (2021). The functional and antiviral activity of interferon alpha-inducible IFI6 against hepatitis B virus replication and gene expression. *Front. Immunol.* 12:634937. doi: 10.3389/fimmu.2021.634937
- Shojaei, M., Shamshirian, A., Monkman, J., Grice, L., Tran, M., Tan, C. W., et al. (2023). IFI27 transcription is an early predictor for COVID-19 outcomes, a multi-cohort observational study. *Front. Immunol.* 13:1060438. doi: 10.3389/fimmu.2022.1060438
- Speranza, E., Bixler, S. L., Altamura, L. A., Arnold, C. E., Pratt, W. D., Taylor-Howell, C., et al. (2018). A conserved transcriptional response to intranasal Ebola virus exposure in nonhuman primates prior to onset of fever. *Sci. Transl. Med.* 10:434. doi: 10.1126/scitranslmed.aaq1016
- Tang, B. M., Shojaei, M., Parnell, G. P., Huang, S., Nalos, M., Teoh, S., et al. (2017). A novel immune biomarker IFI27 discriminates between influenza and bacteria in patients with suspected respiratory infection. *Eur. Respir. J.* 49:1602098. doi: 10.1183/13993003.02098-2016
- Versteeg, K., Menicucci, A. R., Woolsey, C., Mire, C. E., Geisbert, J. B., Cross, R. W., et al. (2017). Infection with the Makona variant results in a delayed and distinct host immune response compared to previous Ebola virus variants. *Sci. Rep.* 7:9730. doi: 10.1038/s41598-017-09963-y
- Villamayor, L., López-García, D., Rivero, V., Martínez-Sobrido, L., Nogales, A., DeDiego, M. L., et al. (2023). The IFN-stimulated gene IFI27 counteracts innate immune responses after viral infections by interfering with RIG-I signaling. *Front. Microbiol.* 14:1176177. doi: 10.3389/fmicb.2023.1176177
- Ward, J. H. Jr. (1963). Hierarchical grouping to optimize an objective function. *J. Am. Stat. Assoc.* 58, 236–244. doi: 10.1080/01621459.1963.10500845
- Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., et al. (2021). clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* 2:100141. doi: 10.1016/j.xinn.2021.100141