Biostatistical Analysis of Microarray Data to Decipher Viral Pathogenesis

Usha Adiga[1], Banubadi Anil Kishore[1], B. Venkat Ramesh[1], Alfred J. Augustine[2] and Sampara Vasishta[1,*]

[1]*Department of Biochemistry, Apollo Institute of Medical Sciences and Research, Murukambattu - 517127, Chittoor, Andhra Pradesh, India*

[2]*Department of Surgery, Apollo Institute of Medical Sciences and Research, Murukambattu - 517127, Chittoor, Andhra Pradesh, India*

Abstract: *Background*: Zika virus, Kunjin virus, Yellow Fever virus, & Sindbis virus belong to *Flaviviridae* family and are involved in derailing various biological pathways which are not yet elucidated.

Aim: Understanding the gene as well as miRNA interplay which plays a vital role in pathogenesis in the diagnosis and prognosis of the disease is of utmost significance.

Materials and Methods: By leveraging microarray data from the Gene Expression Omnibus GSE232504 dataset, we meticulously examined the differentially expressed genes & micro RNAs (miRNAs) induced by viral infections.

Results: Our analysis revealed 60 statistically significant and differentially expressed genes (DEGs) out of a total of 18,725, with *SESN2* (SESTRIN 2) and *GADD45A* (Growth Arrest and DNA Damage-Inducible Alpha) standing out as highly significant players in the host cell response to these viruses. hsa-miR-148b-3p, hsa-miR-148a-3p, hsa-miR-607 & hsa-miR-5582-3p were the highly expressed micro RNAs (miRNAs). Through functional enrichment analyses, we unveiled significant pathways, including Type 1 Diabetes Mellitus and NF-kappa B Signaling, shedding light on the potential mechanisms underlying these virus-host cell interactions. Furthermore, our PPI (protein-protein interaction) network analysis highlighted key hub genes, while our exploration of miRNA-gene targeting relationships offered valuable insights into post-transcriptional regulation.

Conclusion: This study provides a robust foundation for understanding the molecular intricacies of virus-host cell interactions, offering potential targets for further experimental validation and paving the way for innovative therapeutic approaches in combatting viral infections and associated diseases.

Keywords: Viral infection, Diagnosis, Treatment, statistical modeling, data visualization.

INTRODUCTION

Flaviviruses are a group of viruses characterized by an 11 kb, positive-sense, and single-stranded RNA genome [1]. These viruses have a global distribution, posing a risk of infection to over 50% of the world's population (Pierson & Diamond, 2020). While many infections caused by flaviviruses are asymptomatic, the most common clinical presentation resembles a flu-like illness. However, it is crucial to note that flaviviruses have the potential to cause a range of severe ailments, including jaundice, encephalitis, and hemorrhagic fever, albeit in a small proportion of cases [2]. Among the notable members of the genus are Zika virus (ZIKV), Yellow Fever virus (YFV), Dengue virus (DENV), Japanese Encephalitis virus (JEV), and West Nile virus (WNV) [2,3]. These viruses are transmitted through mosquito vectors, which initially infect human skin cells, including dendritic cells. Subsequently, infected cells travel to lymphoid tissues, facilitating the dissemination of the virus throughout the body, including the central nervous system (CNS). Flaviviruses exhibit a remarkable ability to manipulate various cellular processes for their advantage upon infection. For instance, DENV has been reported to enhance fatty acid biosynthesis by employing fatty acid synthase (FASN) at viral replication sites and enhancing FASN activity [4]. This alteration changes the lipid composition of virus-infected cells, altering the properties of the membrane and enhancing the formation of viral replication compartments [5]. Furthermore, *Flaviviridae* employ strategies to counter host cell defense mechanisms. They can inhibit interferon signaling [6] and prevent the formation of stress granules [7], allowing them to overcome the translation block of viral RNA.

Current research has highlighted how flaviviruses can significantly impact the transcriptome of host cells, influencing gene expression and pre-mRNA splicing [8]. DENV, for example, affects the splicing of Spermidine/Spermine N1-Acetyltransferase 1 (SAT1), leading to the augmented addition of the fourth exon in SAT1 mRNA. This results in the degradation of SAT1 mRNA and subsequently reduces the levels of the antiviral protein SAT1 [9].

*Address correspondence to this author at the Department of Biochemistry, Apollo Institute of Medical Sciences and Research, Murukambattu - 517127, Chittoor, Andhra Pradesh, India; E-mail: vasishta_s@aimsrchittoor.edu.in

E-ISSN: 1929-6029/24

It is worth noting that most studies on transcriptomic profiling following flavivirus infection have focused on a single viral class, and investigational conditions across diverse studies have varied widely. This variability has made it challenging to compare the deviations in the host cell transcriptome caused by related flaviviruses accurately. To address this, our work aims to provide insights into the diverse molecular effects of flavivirus infections on host cells by investigating variations in the coding transcriptome and proteome upon infection with *Flaviviridae* (Zika, Kunjin, and Yellow Fever viruses).

The objective of the study was to utilize microarray data deposited by Brand *et al.* 2023 to find differentially expressed genes (DEGs) linked with the complex interplay between viruses (Zika, Kunjin, or Yellow Fever virus) and host cells (human brain-derived U87 cells) using bioinformatics tools. The research also sought to investigate gene enrichment, gene ontology (GO), enhancements in Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, protein-protein interaction (PPI) networks, and cooperative miRNA-target quality networks, all in an effort to gain a deeper understanding of the fundamental molecular processes involved in the interactions between the virus and host cells.

METHODS

Extraction of Data of Microarray

The public and functional genomics database called the Gene Expression Omnibus contains chips, high-throughput microarrays, and gene expression data. In this study, we retrieved the GSE232504 microarray dataset from https://www.ncbi.nlm.nih.gov/geo/ for analysis. This dataset was originally deposited by [8].

Processing of Data

To process raw data and identify DEGs, we employed statistical software R 4.0.1 (available at https://www.r-project.org/) in conjunction with Bioconductor (http://bioconductor.org/biocLite.R). The Limma software was employed to normalize the information in sets, a crucial step that involves fitting a gene linear model to assess differential expression potential. Limma's robust tools facilitated data reading, standardization, and analysis. Subsequently, another instance of the Limma set was utilized to screen genes based on varying expression levels, using a fold-change and p – value < 0.05. The differentially expressed genes (DEGs) were visualized on a volcano plot using the ggplot 2 package, and significant DEGs were further gathered using pheatmap.

Assessment of Function and Pathway Enrichment of Differentially Expressed Genes

We utilized the GO resource, available at http://www.geneontology.org, a community-driven bioinformatics tool. GO employs ontology to enhance our comprehension of biology, offering insights into gene and gene product functions. Additionally, we employed the KEGG resource (https://www.kegg.jp/), which serves as an information repository for understanding genomic classifications and additional biological information. It encompasses systematic, genomic, chemical, and a specialized human-specific category of health-related information. To analyze related signal pathways and biological functions, we employed GO/KEGG enrichment alongside the Cluster Profiler software package. A p - value of 0.05 or lower was considered statistically significant.

Creating a Network of Protein-Protein Interactions and Determining and Confirming Key Hub Genes

We employed the STRING database to build a protein-protein interaction network visualized in Cytoscape. Furthermore, we utilized Cytoscape's cytohubba module to identify central genes with higher scores, indicating their stronger connections within the PPI network. To confirm the significance of these genes statistically, GEO2R was used and significance was determined at a threshold of $p < 0.05$.

To examine the miRNA-gene directing relationships for overlapping differentially expressed genes and hub genes, we consulted the miRDB, DIANA, and TargetScan databases, which provide comprehensive information on empirically confirmed relationships among genes and miRNAs, including metadata, experimental methods, and conditions.

RESULTS

A Total of 1190 Genes were Modulated in Response to CDCA Treatment

Out of 18725 DEGs, 1190 genes were commonly expressed in the tissues infected with all the viruses (Figure **1**).

Each Viral Infection Displayed a Unique Profile, Highlighting Distinct Characteristics and Responses Associated with each Virus

The UMAP plot shows clear separation among various viral infections with each group represented by a different color, underscoring the distinct profiles.

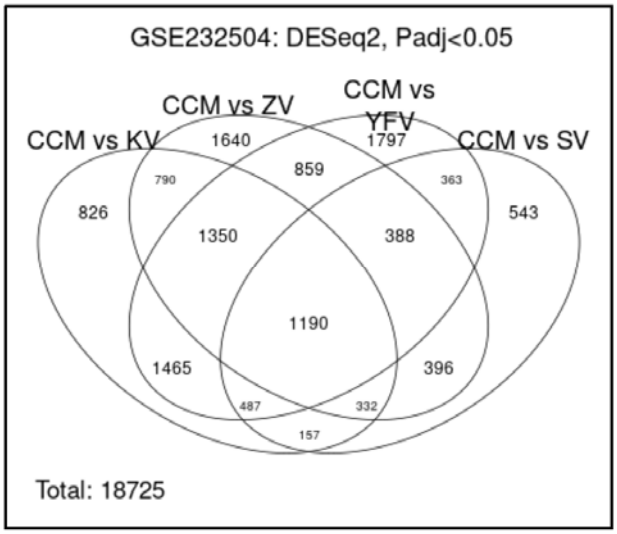

Figure 1: Venn diagram representing the common genes across all the groups. CCM – Cultured conditioned medium, ZK – Zika Virus, YFV – Yellow fever virus, SV – Sindbis virus, KV – Kunjn virus.

This UMAP (Uniform Manifold Approximation and Projection) plot showcases the clustering of different groups within the dataset GSE232504, with the number of neighbors set to seven. Each point represents a data sample, and the colors correspond to specific groups: CCM, KV, ZV, YFV, and SV. The distinct clustering patterns reflect the inherent similarity or dissimilarity among the groups, suggesting underlying biological or molecular differences captured in the dataset. For biostatistical research, such visualizations are instrumental in reducing dimensionality while preserving data structure, allowing researchers to identify patterns, relationships, and potential outliers.

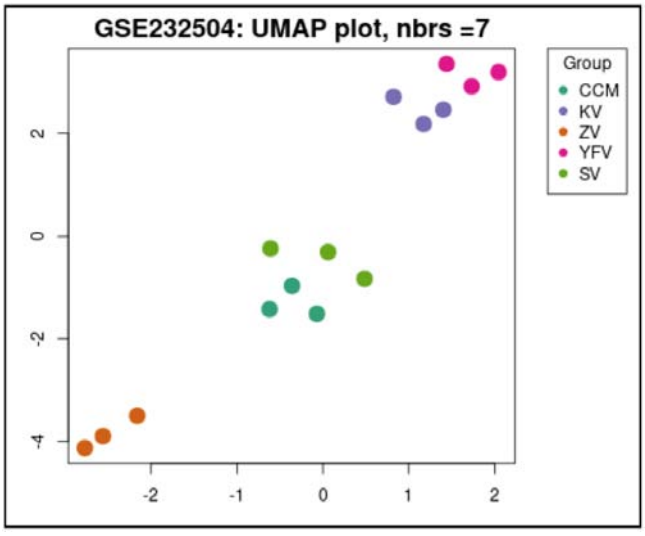

Figure 2: UMAP plot representing variations among four groups and clustering within the groups. CCM – Cultured conditioned medium, ZK – Zika Virus, YFV – Yellow fever virus, SV – Sindbis virus, KV – Kunjn virus.

The clear segregation observed here can guide downstream analyses, such as biomarker identification, differential expression studies, or pathway analyses, further enhancing our understanding of the biological phenomena represented in these groups.

Box Plots Representing the Normal Distribution of the Data

In the dataset GSE232504, normalized gene expression counts (log10-transformed) were compared across samples from five different viral infection groups: CQV, KV, ZV, YFV, and SV. The boxplot reveals that the distribution of gene expression levels is relatively consistent across all groups, as indicated by similar median values and interquartile ranges. This uniformity suggests effective normalization, minimizing technical variation between samples. No significant outliers are observed within each group, indicating that the expression levels are relatively stable within each infection type. The comparable expression levels across groups may suggest a similar transcriptional response baseline among the different infections, though further analysis would be necessary to identify specific expression differences related to each virus type.

Chenodeoxycholic Acid Induces Differential Expression of Genes in Gallbladder Cancer Cells

In the analysis of the GSE232504 dataset, KV resulted in decrease in the expression of 317 genes and increase in the expression of 25 genes (Figure **4A**). The expression of 290 genes was reduced and 138 genes was increased as a result of ZV infection (Figure **4B**). Further, 587 genes were found to be downregulated and 22 genes were upregulated in response to YFV (Figure **4C**). A total of 108 genes were downregulated and 5 genes were upregulated as a result of SV infection (Figure **4D**). This suggested that each virus has a distinct pattern in modulating the host gene expression.

The Gene Expression Pattern was Found to be Unique to each Viral Infection

A significant variation in the gene expression profiles was observed between control group when compared to any other viral infection (Figure **5A-5B**).

Majority of the Genes were Found to be Significant

A significant proportion of the genes were found to be modulated in response to viral infections ($p<0.05$) (Figure **6**).

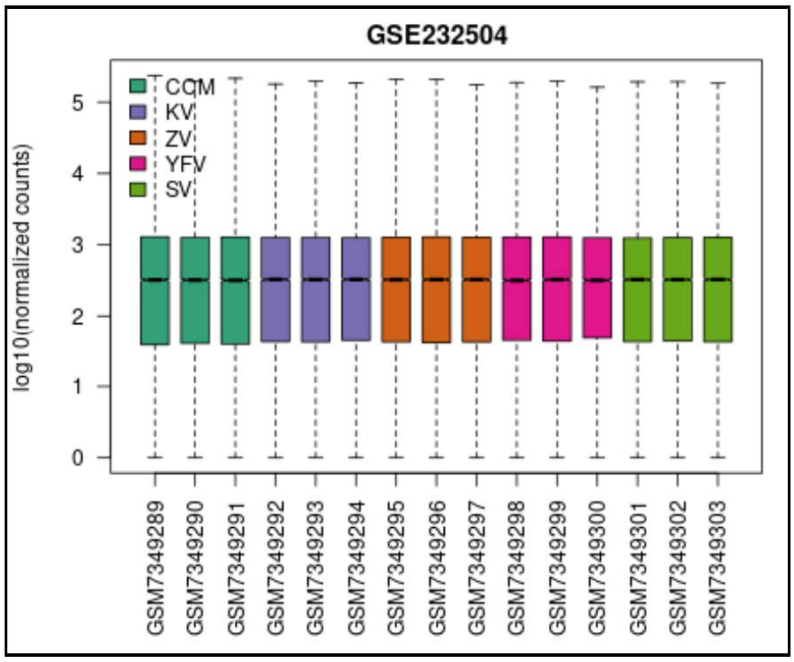

Figure 3: Boxplot of normalized gene expression counts (in Log_{10} scale) for multiple samples associated with the dataset GSE232504. Each color-coded box represents one of four distinct viral infection groups: CQV, KV, ZV, YFV, and SV, as indicated by the legend.

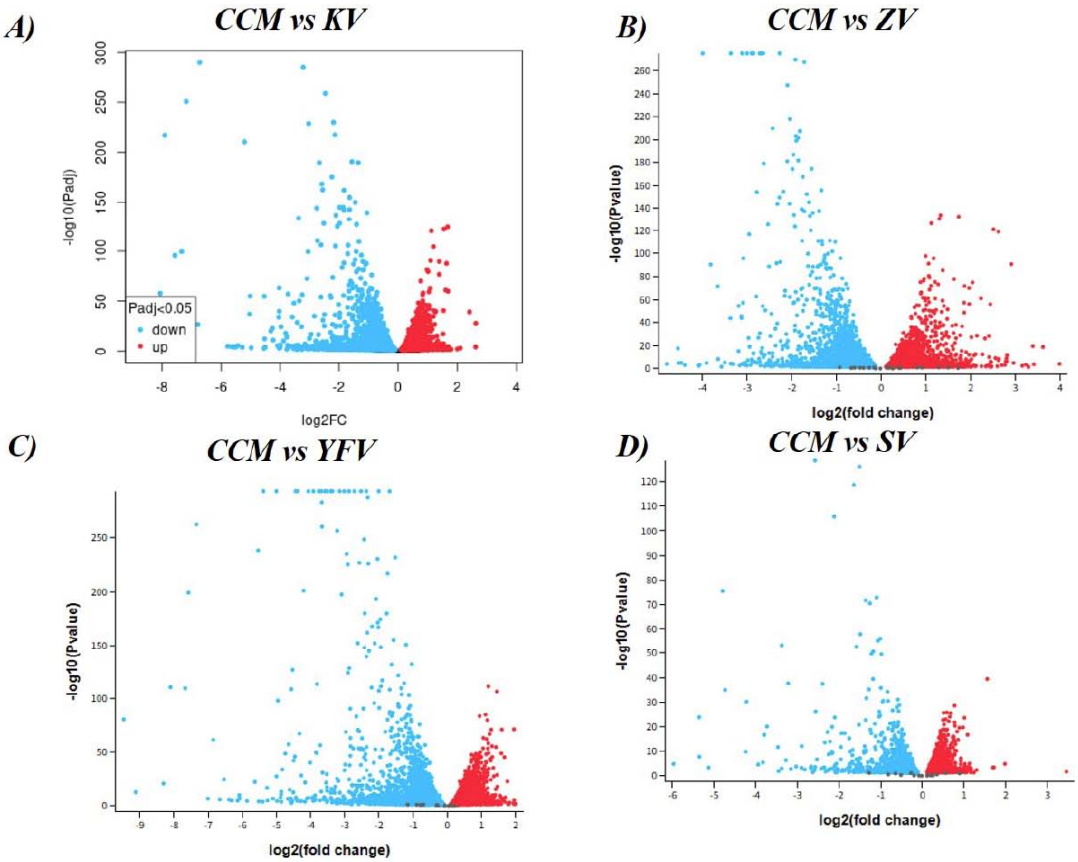

Figure 4: Differential expression of the genes in response to various infections. The x - axis (Log_2 Fold Change) signifies the degree of gene expression change, positive values signifying upregulation (red) and negative values representing downregulation (blue). The y - axis ($-Log_{10}$ p - value) represents the statistical significance of differential expression, where smaller p - value indicate greater significance. Each data point on the plot corresponds to a gene, positioned based on its log_2 fold change on x - axis and negative logarithm of its p - value on y - axis.

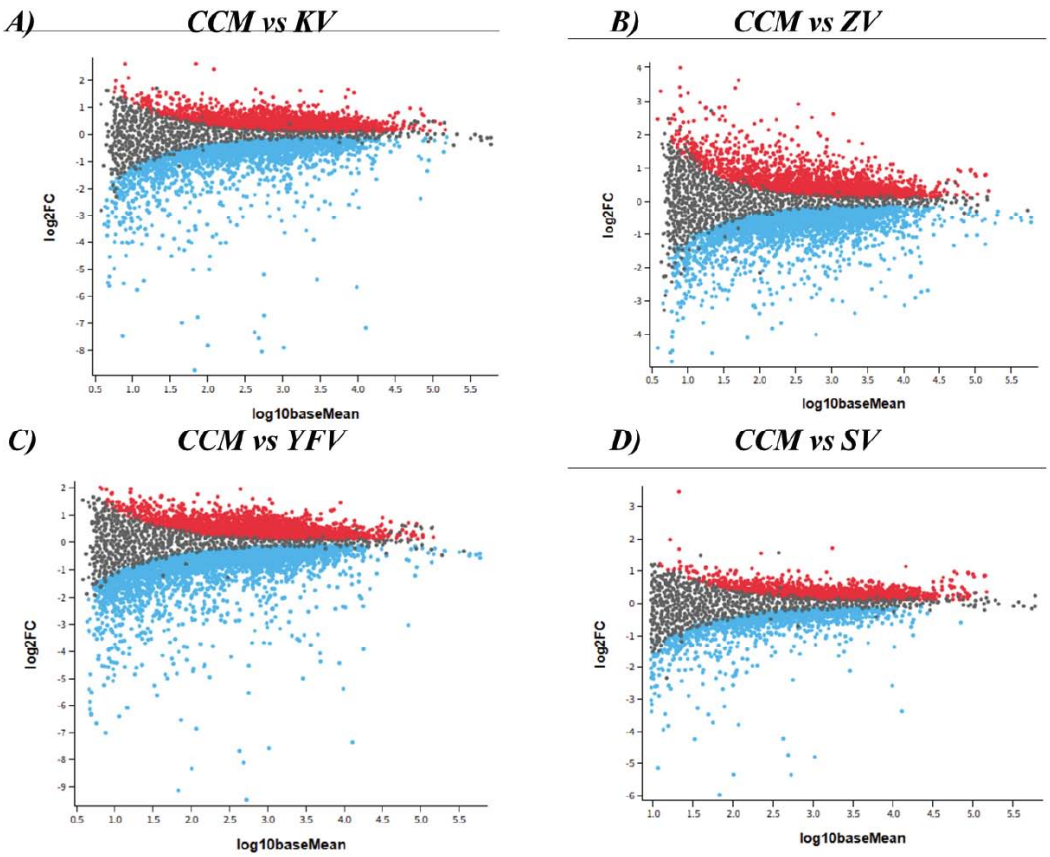

Figure 5: MA (mean-difference) plots comparing gene expression between different conditions, with respect to control and various other groups (KV, ZV, YFV, and SV). The four panels (**A**, **B**, **C**, and **D**) show the $\log_2$ fold change ($\log_2$FC) of each gene on the y-axis, plotted against the $\log_{10}$ transformed mean expression level ($\log_{10}$ base Mean) on the x-axis.

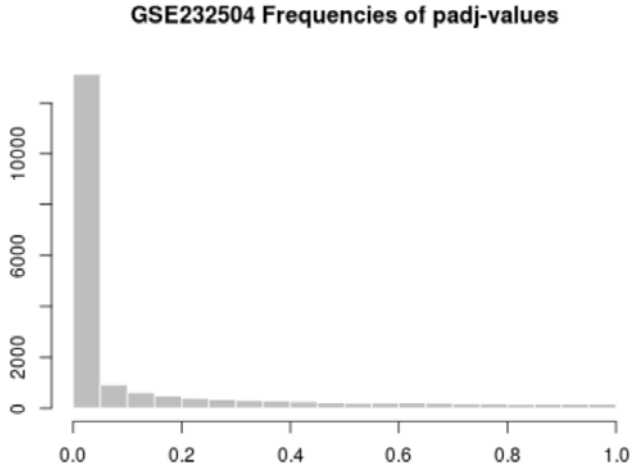

Figure 6: Viral infection has significantly modulated the gene expression patterns. The histogram represents p-value against the number of genes.

Top 10 differentially expressed genes are represented in Table **1**. The table contains information about several genes or proteins, including their symbols, descriptions, statistical significance (p - value), and interactions with specific microRNAs (miRNAs). hsa-miR-148b-3p, hsa-miR-148a-3p, hsa-miR-5582-3p and hsa-miR-607 were the highly expressed miRNAs. Each row in the table represents a unique gene or protein, and the associated miRNAs are listed alongside. The p - value of 0 for each entry suggests that these interactions are highly statistically significant. This information is crucial in understanding the governing associations between miRNAs and their mark genes or proteins. It indicates that these miRNAs likely perform a significant role in the post-transcriptional regulation of the mentioned genes or proteins. Further analysis and experimental validation may be required to uncover the specific biological implications of these interactions in cellular processes and diseases.

Our gene ontology analysis revealed that top 10 genes are associated with molecular fucntions such as leucine binding, cytokine activity, protein heterodimerization activity etc. (Figure **7A**). Cellular component analysis showed cytoplasmic lumen, organelle outer lumen, HFE transferring receptor complex and lumen of various organelles (Figure **7B**). The key biological processes associated with these 10 genes are regulation of VEGF, T cell proliferation and unsaturated fatty acid biosynthesis (Figure **7C**).

Table 1: Enrichment of Top 10 Genes and their Associated miRNAs

Symbol	Description	p - value	miRNA
SESN2	Sestrin 2	0	hsa-miR-605-3p hsa-miR-148b-3p hsa-miR-148a-3p hsa-miR-891a-3p
GADD45A	Growth Arrest And DNA Damage Inducible Alpha	0	hsa-miR-659-3p hsa-miR-607 hsa-miR-5582-3p hsa-miR-3163
CTH	Cystathionine Gamma-Lyase	0	hsa-miR-27b-3p hsa-miR-9985 hsa-miR-582-5p hsa-miR-27a-3p
F3	Coagulation Factor III, Tissue Factor	0	hsa-let-7a-3p hsa-miR-5011-5p hsa-let-7b-3p hsa-miR-98-3p
HSPA6	Heat Shock Protein Family A (Hsp70) Member 6	0	hsa-miR-6779-3p hsa-miR-6515-5p hsa-miR-1343-3p hsa-miR-6721-5p
PTGS2	Prostaglandin-Endoperoxide Synthase 2	0	hsa-miR-95-5p hsa-miR-5692a hsa-miR-26a-5p hsa-miR-1297
ATF3	Activating Transcription Factor 3	0	hsa-miR-600 hsa-miR-3200-5p hsa-miR-6823-5p hsa-miR-4251
RSAD2	Radical S-Adenosyl Methionine Domain Containing 2	0	hsa-miR-141-3p hsa-miR-590-5p hsa-miR-21-5p hsa-miR-200a-3p
IL1A	Interleukin 1 Alpha	0	hsa-miR-543 hsa-miR-323b-5p hsa-miR-3662 hsa-miR-410-5p
IL1B	Interleukin 1 Beta	0	hsa-miR-495-3p hsa-miR-5688 hsa-miR-1291 hsa-miR-5692a

Top 10 Modulated Genes Influence Immune Pathways

Protein – protein interaction analysis showed that the top 10 genes are modulate inflammatory response mediated by interleukins (Figure **8**).

The results presented in the KEGG pathway analysis report (Table **2**) indicate the enrichment of specific biological pathways within a dataset of genes or proteins. Notably, the Type 1 Diabetes Mellitus pathway showed significant enrichment, with 2 out of 38 genes associated with this autoimmune disease, and a low false discovery rate (FDR), suggesting a reliable finding. Conversely, the Tuberculosis pathway exhibited enrichment, but with a somewhat higher FDR, warranting cautious interpretation. The NF-kappa B Signaling Pathway was highly enriched, with a strength value of 1.77 and an impressively low FDR of 0.0019, indicating a robust association with this cellular

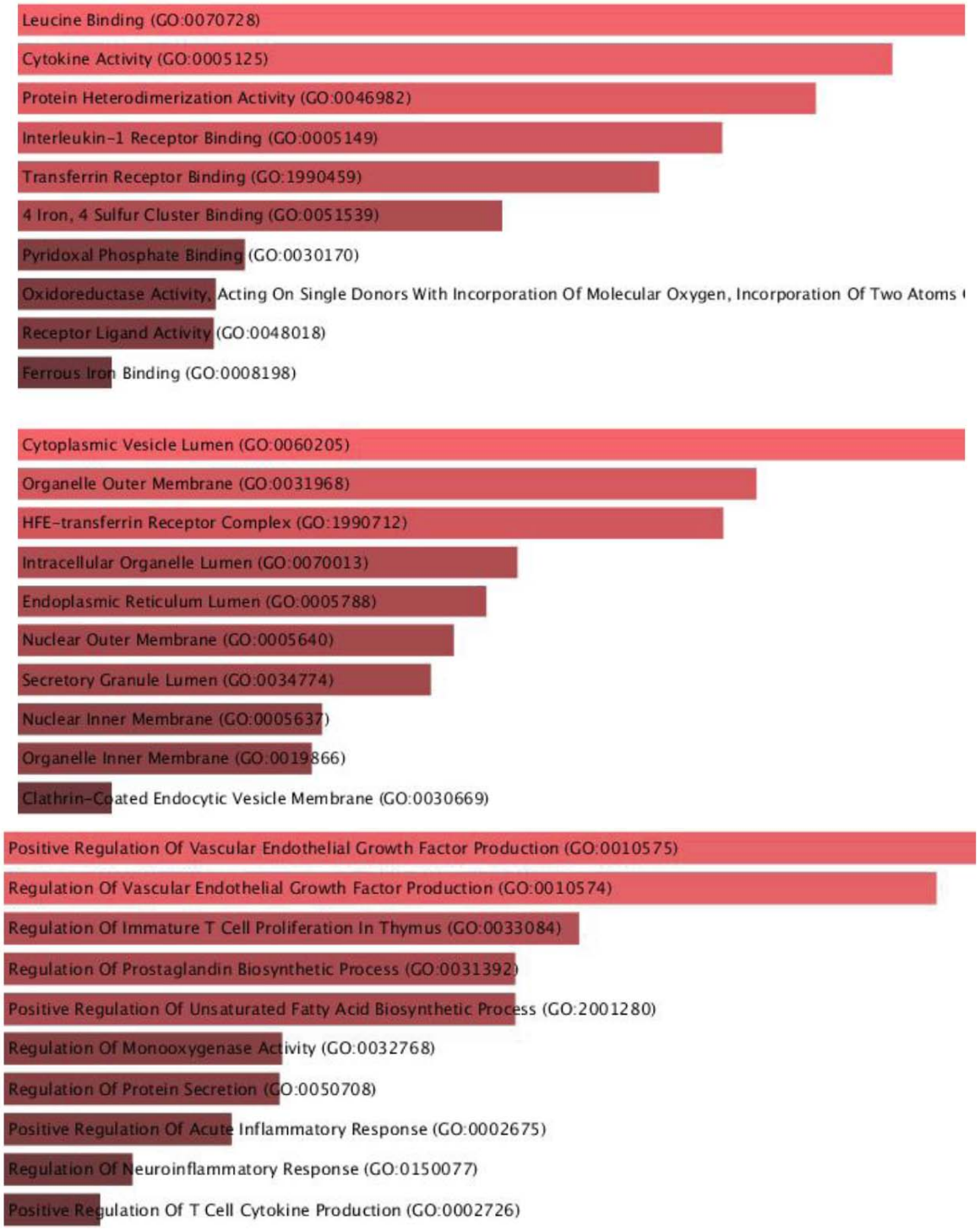

Figure 7: GO analysis of top 10 modulated genes. **A**) Molecular function, **B**) Cellular component analysis and **C**) Biological processes.

process. Furthermore, Alzheimer's Disease and the AGE-RAGE Signaling Pathway in Diabetic Complications both displayed enrichment, with moderate FDR values of 0.0172 and an exceptionally low 0.0019, respectively. These findings offer insights into potential biological processes associated with the dataset and can guide further research, with the reliability of each enrichment assessed by the FDR values.

DISCUSSION

Virus-host interactions are crucial aspects of virology and are essential for understanding how

viruses infect, replicate within, and interact with their host organisms.

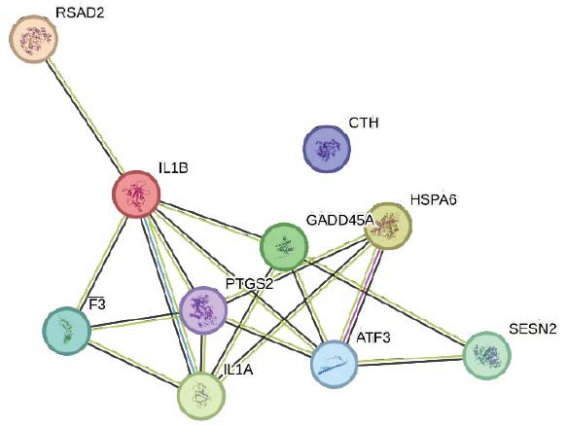

Figure 8: Protein-protein interactions of top 10 genes.

Kunjin virus, a member of the *Flaviviridae* family, primarily infects birds but can also affect humans and horses [10]. The virus enters host cells by binding its envelope protein E to cellular receptors like DC-SIGN and heparan sulfate. Once inside, it replicates in the cell's cytoplasm, hijacking host machinery for RNA replication and protein production [11]. The host's immune system responds with innate immune mechanisms, including interferon production. While Kunjin virus is generally less pathogenic in humans, severe cases can occur, particularly in individuals with weakened immune systems [12].

Zika virus, another flavivirus, is known for its association with microcephaly in newborns [13]. It enters host cells via interactions between its envelope protein E and receptors like AXL, Tyro3, and TIM-1,

Table 2: Representation of KEGG Pathway

Pathway	Description	Count in network	Strength	False discovery rate
hsa04904	Type 1 diabetes mellitus	2 of 38	2.02	0.0077
hsa05152	Tuberculosis	2 of 165	1.38	0.0397
hsa04668	TNF signaling pathway	2 of 111	1.55	0.0239
hsa05222	Small cell lung cancer	2 of 92	1.63	0.0199
hsa05323	Rheumatoid arthritis	2 of 83	1.68	0.0181
hsa05020	Prion disease	3 of 263	1.35	0.0102
hsa05133	Pertussis	2 of 73	1.73	0.0172
hsa04115	P53 signaling pathway	2 of 72	1.74	0.0172
hsa04380	Osteoclast differentiation	2 of 120	1.52	0.0265
hsa04932	NAFLD	2 of 146	1.43	0.0354
hsa04064	NF-kB pathway	2 of 101	1.77	0.0019
hsa04217	Necroptosis	2 of 147	1.43	0.0354
hsa05162	Measles	3 of137	1.63	0.0027
hsa04010	MAPK signaling pathway	4 of 286	1.44	0.0019
hsa05140	Leishmaniasis	3 of 69	1.93	0.0019
hsa05134	Legionellosis	2 of55	1.86	0.0122
hsa05164	Influenza A	2 of 163	1.56	0.0038
hsa05321	Inflammatory bowel disease	2 of 59	1.82	0.0127
hsa04657	IL-17 signaling pathway	2 of 91	1.64	0.0199
hsa04640	Hematopoietic cell lineage	2 of 90	1.64	0.0199
hsa05332	Graft-versus-host disease	2 of 36	2.04	0.0077
hsa05418	Fluid shear stress and atherosclerosis	2 of 129	1.48	0.0291
hsa04218	Cellular senescence	2 of 150	1.42	0.0354
hsa04625	C type lectin receptor signaling pathway	2 of 101	1.59	0.0209
hsa05010	Alzheimer disease	3 of 354	1.22	0.0172
hsa04933	AGE-RAGE signaling pathway in diabetic complications	3 of 96	1.79	0.0019

with a particular affinity for neural cells [14]. The virus can infect neural stem cells and developing brain cells, contributing to microcephaly. Host immune responses include interferon production and pro-inflammatory cytokines, although Zika virus can evade some immune defenses. Cross-reactivity with other flaviviruses can impact the immune response and possibly worsen outcomes in those previously exposed to related viruses [15].

Yellow Fever virus, a mosquito-borne flavivirus, enters host cells by binding envelope protein E to receptors like DC-SIGN and heparan sulfate [16]. Replication occurs in the cell's cytoplasm using the host's machinery. The virus triggers a robust immune response, activating both innate and adaptive components. This immune response is vital in controlling and clearing the infection [17]. Importantly, a highly effective Yellow Fever vaccine has been developed, providing long-lasting immunity and preventing outbreaks by priming the host's immune system to recognize and combat the virus effectively [18].

Studying virus-host interactions using bioinformatics tools is of paramount importance for several reasons;

Bioinformatics helps to dissect the molecular interactions between viruses and their host organisms at a systems level. This understanding is crucial for unraveling the mechanisms underlying viral infections, replication, and pathogenesis, which can inform the development of targeted therapies and vaccines. By identifying key host factors and pathways involved in virus-host interactions, bioinformatics can aid in the discovery of potential drug targets. This knowledge can guide the development of antiviral drugs and vaccines, offering promising strategies for the treatment and prevention of viral diseases. Bioinformatics tools can analyze viral genomes and predict how viruses may evolve over time. This is vital for tracking the emergence of new viral strains, assessing their potential for increased virulence or transmission, and developing strategies to mitigate outbreaks. Understanding individual variations in host responses to viral infections is crucial for personalized medicine. Bioinformatics can help identify genetic factors that influence susceptibility to certain viruses, enabling more tailored treatment plans and vaccination strategies.

Bioinformatics tools can pinpoint biomarkers associated with viral infections. These biomarkers can be used for early detection, monitoring of progression, and treatment efficacy assessment. Studying virus-host interactions on a large scale can help public health organizations monitor and respond to infectious disease outbreaks more effectively. This information can inform containment strategies, vaccination campaigns, and resource allocation. Bioinformatics allows for the integration of diverse data sources, including genomics, transcriptomics, proteomics, and clinical data. Integrating these data can provide a comprehensive view of virus-host interactions and their impact on health and disease. Computational models and simulations generated through bioinformatics can be used to test hypotheses and predict outcomes of virus-host interactions. This enables researchers to explore various scenarios and strategies in a cost-effective and time-efficient manner. Bioinformatics facilitates data sharing and collaboration among researchers worldwide. Open-access databases and bioinformatics tools ensure that valuable information is widely available to the scientific community, fostering collaboration and accelerating research. Bioinformatics also plays a role in the ethical analysis of virus-host interactions, helping to address issues related to data privacy, informed consent, and the responsible use of genetic information in research and healthcare.

Advanced technologies such as microarrays have transformed our capacity to investigate overall genome expression levels and identify changes in gene expression without relying on assumptions. In the past decade, these technologies have played a pivotal role in studies pinpointing irregularities in gene expression linked to mental health disorders. Beyond the conventional hypothesis-driven approach, transcriptomics research holds the potential to uncover fresh indicators associated with various mental disorders, opening the door to innovative treatment strategies and personalized medicine.

Protein-protein interactions (PPIs) are pivotal in predicting the functions of target proteins and evaluating the drug-like qualities of molecules. PPIs govern various biological processes, including metabolism, development, and interactions between cells. Information concerning these interactions is essential in pinpointing potential therapeutic targets. Research has revealed that proteins with numerous connections, known as hubs, encompass enzyme families, transcription factors, and intrinsically disordered proteins. PPIs involve intricate processes and a wide-reaching regulatory scope, underscoring

the necessity of understanding diverse interactions and their impacts. In-silico methods are frequently employed to scrutinize PPIs. For instance, STRING uses functional associations to link proteins contributing to particular biological functions. According to the analysis from the STRING database presented in Figure **8**, there is no noteworthy gene co-expression, co-occurrence, or gene fusions associated with the examined gene [19].

miRNAs, capable of targeting over 60% of human genes, play a pivotal and diverse role in gene regulation [20]. Prior research has established their importance in kidney development, structure, function, and the regulation of electrolytes, fluid, blood pressure and acid-base balance. MiRNAs are also implicated in pathological processes. Importantly, miRNA levels in serum and urine remain stable despite storage conditions, making them valuable diagnostic and prognostic markers [20-24].

This study demonstrates the application of biostatistical tools in virology to analyze transcriptomic data, focusing on Zika virus, Kunjin virus, Yellow Fever virus, and Sindbis virus of the Flaviviridae family. By examining microarray data from the GSE232504 dataset, the analysis identified differentially expressed genes (DEGs) and miRNAs critical to the viruses' pathogenic mechanisms. Statistical methods enabled the discovery of key DEGs like SESN2 and GADD45A, and miRNAs such as hsa-miR-148b-3p and hsa-miR-148a-3p. Functional enrichment and pathway analyses, supported by PPI network mapping, elucidated the involvement of significant pathways like NF-kappa B signaling in viral pathogenesis. This integrative approach underscores the value of biostatistics in decoding complex virus-host interactions, identifying potential biomarkers, and guiding therapeutic innovations.

The application of biostatistical techniques to transcriptomic data in this study highlights their broader significance in modeling complex biological systems, particularly in virology. By leveraging data reduction and clustering methods such as UMAP, combined with differential expression analysis, the study successfully identified key genes and miRNAs involved in the host response to Zika virus, Kunjin virus, Yellow Fever virus, and Sindbis virus. These viruses, belonging to the Flaviviridae family, disrupt multiple pathways critical to maintaining cellular homeostasis. Biostatistical modeling provides a robust framework to unravel these complex interactions, enabling researchers to identify significant molecular players like SESN2, GADD45A, and miRNAs such as hsa-miR-148b-3p and hsa-miR-148a-3p, which could serve as potential diagnostic or therapeutic targets.

A key strength of these techniques lies in their ability to handle high-dimensional data, such as the 18,725 genes analyzed in this study. By identifying only 60 statistically significant DEGs, the methods not only reduced data complexity but also emphasized biologically relevant findings. Similarly, the integration of pathway enrichment analyses and PPI network construction provided deeper insights into the biological systems affected by these viruses. For instance, the identification of pathways like NF-kappa B signaling and Type 1 Diabetes Mellitus underscores the potential systemic implications of these viral infections, further linking molecular disruptions to clinical outcomes. This systems biology approach offers a holistic view, connecting gene-level changes to pathway-level effects and organism-wide responses.

The miRNA-gene interaction analysis adds another layer of complexity to the study, shedding light on post-transcriptional regulatory mechanisms. MiRNAs are known to fine-tune gene expression and play critical roles in viral pathogenesis. The identification of miRNAs such as hsa-miR-607 and hsa-miR-5582-3p highlights their potential as biomarkers or therapeutic targets. Biostatistical modeling facilitates this understanding by providing tools to predict and validate miRNA-gene interactions, creating a roadmap for further experimental studies.

Beyond this study, the implications of these biostatistical techniques extend to a wide range of biological systems. They enable the modeling of dynamic, non-linear relationships within and between molecular pathways, making them invaluable for understanding complex diseases, multi-organ interactions, and host-pathogen dynamics. These methods can also be applied to study other emerging viral infections, adapt to new datasets, and integrate multi-omics data for a more comprehensive understanding. As the field of virology evolves, the ability to model and predict biological responses using biostatistical frameworks will become increasingly critical, driving advances in precision medicine and public health strategies.

In conclusion, the biostatistical approaches used in this study not only provided valuable insights into virus-

host interactions but also demonstrated their potential for broader applications in understanding and modeling complex biological systems. These techniques serve as a bridge between data and biology, paving the way for translational research that could significantly impact diagnostic, prognostic, and therapeutic paradigms.

In summary, the study of virus-host interactions through bioinformatics tools has far-reaching implications for human health, disease prevention, and our understanding of the complex interplay between viruses and their hosts. It empowers researchers to develop innovative solutions and strategies to combat infectious diseases and protect public health.

LIMITATIONS OF THE STUDY

As this is a bioinformatics study validation, in the clinical cohorts is warranted. However, the study utilized the dataset involving human derived cell lines and hence, the findings are of significant relevance.

CONCLUSION

In this extensive bioinformatic analysis of the GSE232504 dataset, 18,725 DEGs were found, with 60 genes demonstrating significant statistical differences. The results were visualized using a volcano plot, where genes with higher log2 fold changes were considered upregulated, and those with lower negative log p -value were considered statistically significant. Notably, *SESN2* and *GADD45A* out as highly significant genes. hsa-miR-148b-3p,hsa-miR-148a-3p, hsa-miR-891a-3p and hsa-miR-605-3p were the top expressed miRNAs

ACKNOWLEDGEMENTS

The authors thank the Apollo Institute of Medical Sciences and Research (AIMSR) for providing the infrastructure.

FUNDING SOURCES

None

DECLARATION OF THE INTEREST

The authors declare no potential conflicts of interest

DECLARATION OF GENERATIVE AI AND AI-ASSISTED TECHNOLOGIES IN THE WRITING PROCESS

No AI or AI-assisted technologies were used for drafting this manuscript.

REFERENCES

[1] Boytz R, Keita K, Pawlak JB, Laurent-Rolle M. Flaviviruses manipulate mitochondrial processes to evade the innate immune response. NPJ Viruses 2024; 2(1): 47. https://doi.org/10.1038/s44298-024-00057-x

[2] Pierson TC, Diamond MS. The continued threat of emerging flaviviruses. Nat Microbiol 2020; 5(6): 796-812. https://doi.org/10.1038/s41564-020-0714-0

[3] Silva NM, Santos NC, Martins IC. Dengue and Zika Viruses: Epidemiological History, Potential Therapies, and Promising Vaccines. Trop Med Infect Dis 2020; 5(4). https://doi.org/10.3390/tropicalmed5040150

[4] Hsia JZ, Liu D, Haynes L, Cruz-Cosme R, Tang Q. Lipid Droplets: Formation, Degradation, and Their Role in Cellular Responses to Flavivirus Infections. Microorganisms 2024; 12(4). https://doi.org/10.3390/microorganisms12040647

[5] Avula K, Singh B, Kumar P V, Syed GH. Role of Lipid Transfer Proteins (LTPs) in the Viral Life Cycle. Front Microbiol 2021; 12: 673509. https://doi.org/10.3389/fmicb.2021.673509

[6] Zoladek J, Nisole S. Mosquito-borne flaviviruses and type I interferon: catch me if you can! Front Microbiol 2023; 14: 1257024. https://doi.org/10.3389/fmicb.2023.1257024

[7] Bonenfant G, Williams N, Netzband R, Schwarz MC, Evans MJ, Pager CT. Zika Virus Subverts Stress Granules To Promote and Restrict Viral Gene Expression. J Virol 2019; 93(12). https://doi.org/10.1128/JVI.00520-19

[8] Brand C, Deschamps-Francoeur G, Bullard-Feibelman KM, Scott MS, Geiss BJ, Bisaillon M. Kunjin Virus, Zika Virus, and Yellow Fever Virus Infections Have Distinct Effects on the Coding Transcriptome and Proteome of Brain-Derived U87 Cells. Viruses 2023; 15(7). https://doi.org/10.3390/v15071419

[9] Pozzi B, Bragado L, Mammi P, *et al.* Dengue virus targets RBM10 deregulating host cell splicing and innate immune response. Nucleic Acids Res 2020; 48(12): 6824-6838. https://doi.org/10.1093/nar/gkaa340

[10] Du Pont KE, Sexton NR, McCullagh M, Ebel GD, Geiss BJ. A Hyperactive Kunjin Virus NS3 Helicase Mutant Demonstrates Increased Dissemination and Mortality in Mosquitoes. J Virol 2020; 94(19). https://doi.org/10.1128/JVI.01021-20

[11] Cagno V, Tseligka ED, Jones ST, Tapparel C. Heparan Sulfate Proteoglycans and Viral Attachment: True Receptors or Adaptation Bias? Viruses 2019; 11(7). https://doi.org/10.3390/v11070596

[12] Huang B, West N, Vider J, *et al.* Inflammatory responses to a pathogenic West Nile virus strain. BMC Infect Dis 2019; 19(1): 912. https://doi.org/10.1186/s12879-019-4471-8

[13] Vinhaes CL, Arriaga MB, de Almeida BL, *et al.* Newborns With Zika Virus-Associated Microcephaly Exhibit Marked Systemic Inflammatory Imbalance. J Infect Dis 2020; 222(4): 670-680. https://doi.org/10.1093/infdis/jiaa197

[14] Wang KH, Ding DC. Dual targeting of TAM receptors Tyro3, Axl, and MerTK: Role in tumors and the tumor immune microenvironment. Tzu Chi Med J 2021; 33(3). https://journals.lww.com/tcmj/fulltext/2021/33030/dual_targeting_of_tam_receptors_tyro3,_axl,_and.7.aspx

[15] Komarasamy TV, Adnan NAA, James W, Balasubramaniam VRMT. Zika Virus Neuropathogenesis: The Different Brain Cells, Host Factors and Mechanisms Involved. Front Immunol 2022; 13: 773191. https://doi.org/10.3389/fimmu.2022.773191

[16] Anwar MN, Akhtar R, Abid M, *et al.* The interactions of flaviviruses with cellular receptors: Implications for virus entry. Virology 2022; 568: 77-85. https://doi.org/10.1016/j.virol.2022.02.001

[17] Kleinert RD V, Montoya-Diaz E, Khera T, *et al.* Yellow Fever: Integrating Current Knowledge with Technological Innovations to Identify Strategies for Controlling a Re-Emerging Virus. Viruses 2019; 11(10). https://doi.org/10.3390/v11100960

[18] Gianchecchi E, Cianchi V, Torelli A, Montomoli E. Yellow Fever: Origin, Epidemiology, Preventive Strategies and Future Prospects. Vaccines 2022; 10(3). https://doi.org/10.3390/vaccines10030372

[19] Szklarczyk D, Gable AL, Lyon D, *et al.* STRING v11: Protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids Res 2019; 47(D1): D607-D613. https://doi.org/10.1093/nar/gky1131

[20] Zamzam YA, Mansour TF, Salem RM, Aziz RSA, Elsendiony SA. Serum miR-124a and miR-34a as Potential Biomarkers for Rheumatoid Arthritis. Biomed Biotechnol Res J 2024; 8(2). https://doi.org/10.4103/bbrj.bbrj_142_24

[21] Jaber SQ, Kadhim AS, Al Kateeb AI. Investigating the Expression of miR 203a 3p and Its Role in Inflammatory Response in Severe Preeclampsia of Iraqi women Patients – A Comparative Study. Biomed Biotechnol Res J 2024; 8(3). https://doi.org/10.4103/bbrj.bbrj_210_24

[22] Mishra S, Sarangi R, Das S, Mishra A. Nonalcoholic Fatty Liver Disease and MicroRNAs: A Weighty Consideration. Biomed Biotechnol Res J 2023; 7(1). https://doi.org/10.4103/bbrj.bbrj_319_22

[23] Dsouza N, Gupta DB, Chellasamy SK. Distinct Host Gene Targets and Mode of Action in the MicroRNA of Nipah Virus from Malaysia and Bangladesh: A Comparative in-silico Based Analysis. Biomed Biotechnol Res J 2023; 7(3). https://journals.lww.com/bbrj/fulltext/2023/07030/distinct_host_gene_targets_and_mode_of_action_in.4.aspx

[24] Al-Jumaili MMO. A Potential Interpretation of RNA Interference and GC Island Modification in Progressive Colorectal Cancer Suppression. Biomed Biotechnol Res J 2023; 7(3). https://journals.lww.com/bbrj/fulltext/2023/07030/a_potential_interpretation_of_rna_interference_and.16.aspx

Received on 27-10-2024 Accepted on 25-11-2024 Published on 27-12-2024

https://doi.org/10.6000/1929-6029.2024.13.37