Result:

Differential Expression Analysis

To identify the genes that are differentially expressed between the test and healthy control groups, the DESeq2 tool was downloaded in Linux and using r script analysis was performed. Annotated tabular files, as shown in Figure, were downloaded and visualized in Microsoft Excel.

Identification of DEGs

For the identification of differentially expressed genes, we use the DESeq2 tool. The DESeq2 tool was employed to identify the genes that exhibit differential expression between the three datasets each having 2 control, 2 infected and 2 infected with mutant. Genes were considered differentially expressed if they exhibited a log2 fold change greater than 2 or less than -2, and a p value less than 0.05 compared to expressed genes. DEGs could be either upregulated or downregulated. Annotated tabular files, as shown in Table 3, were downloaded and visualized in excel.

A total of DEGs were identified for zika virus infected cell lines samples from the dataset of them, yields a total of 706 up-regulated genes and 504 down-regulated genes in mdDCs cell line. Besides, there were 2777 up-regulated and 829 down-regulated genes in human brain organoid cells, while 189 genes were found to be upregulated and 71 down-regulated in HEK-293 cells.

Table: Top20 Upregulated Genes from All Three Cell Lines on The Bases of Log2fold Change Value

ensemble_id	symbol	log2FoldChange	Pvalue	Padj
ENSG00000176165	FOXG1	-12.69062432	1.20E-135	8.65E-134
ENSG00000186960	LINC01551	-12.01850943	1.65E-151	1.42E-149
ENSG00000257126	FOXG1-AS1	-11.23173685	6.20E-28	8.54E-27
ENSG00000231764	DLX6-AS1	-10.97849498	0	0
ENSG00000178522	AMBN	-10.81145765	7.25E-26	9.43E-25
ENSG00000219438	TAFA5	-10.56098988	1.36E-13	1.09E-12
ENSG00000289413	NOVEL GENE	-10.55644741	7.83E-59	2.13E-57

ENSG00000207513	RNU1-3	-9.639857544	4.46E-06	2.17E-05
ENSG00000105880	DLX5	-9.325650026	4.39E-140	3.39E-138
ENSG00000144355	DLX1	-9.194453316	0	0
ENSG00000257522	NOVEL GENE	-9.084816058	2.59E-84	1.02E-82
ENSG00000006377	DLX6	-8.914316574	1.46E-195	1.75E-193
ENSG00000229370	NOVEL GENE	-8.895351722	4.63E-09	2.84E-08
ENSG00000257056	LINCO2282	-8.689745851	2.73E-113	1.54E-111
ENSG00000206852	RNU6-895P	-8.618939062	4.02E-08	2.31E-07
ENSG00000236141	NOVEL GENE	-8.513619484	3.31E-08	1.91E-07
ENSG00000257748	LINCO2281	-8.512515529	5.03E-08	2.87E-07
ENSG00000289297	NOVEL GENE	-8.313750155	1.39E-07	7.65E-07
ENSG00000186895	FGF3	-8.285326938	6.65E-07	3.48E-06
ENSG00000115844	DLX2	-8.2759164	9.12E-244	1.52E-241

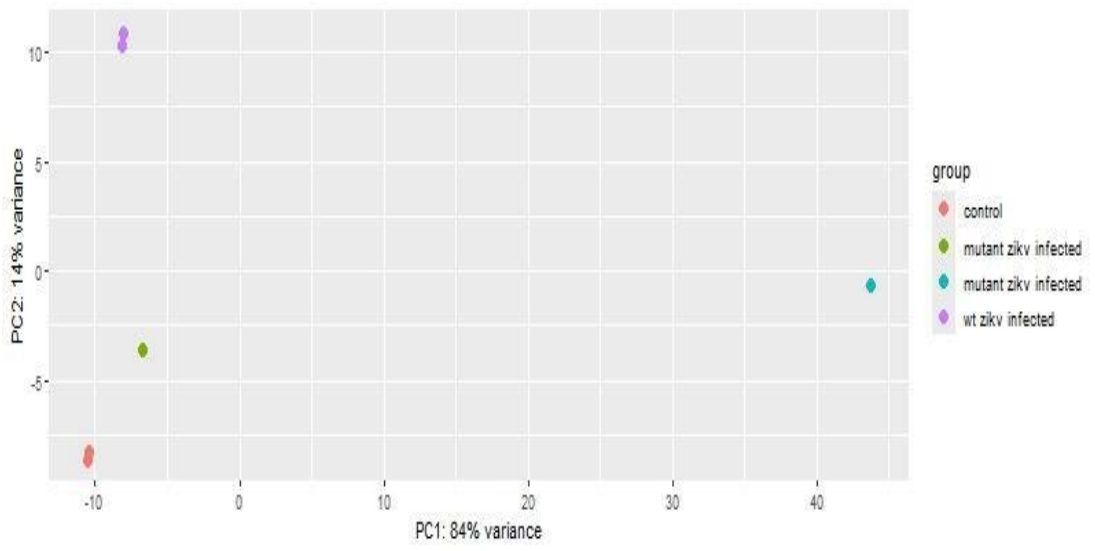
Table: Top20 Downregulated Genes from All Three Cell Lines on The Bases of Log2fold Change Value

ensemble_id	symbol	log2FoldChange	Pvalue	padj
ENSG00000186827	TNFRSF4	29.0802645	0.014080544	0.040803551
ENSG00000182393	IFNL1	14.72369741	1.10E-35	2.60E-34
ENSG00000232596	LINCO1646	14.37370529	0.015513488	0.044417346
ENSG00000142583	SLC2A5	13.85175004	0.000362088	0.001422139
ENSG00000162496	DHRS3	13.72907258	9.87E-148	8.19E-146
ENSG00000169245	CXCL10	13.51584398	2.42E-30	4.56E-29
ENSG00000171729	TMEM51	13.09588449	2.65E-18	2.63E-17
ENSG00000171855	IFNB1	12.87055193	1.30E-27	2.18E-26
ENSG00000162571	LAPTM5	12.79443514	0.002074624	0.007220777
ENSG00000287586	NOVEL GENE	12.7725275	0.013204093	0.038576487
ENSG00000174950	CDI64L2	12.63676951	2.54E-05	0.000114369

ENSG00000197110	IFNL3	12.45534834	6.31E-26	9.78E-25
ENSG00000162511	TTLL10	12.34897813	0.003448682	0.011516909
ENSG00000183709	IFNL2	12.28193285	3.57E-25	5.32E-24
ENSG00000130766	SESN2	12.27949811	7.20E-291	1.53E-288
ENSG00000196581	AJAPI	12.19227033	1.36E-18	1.36E-17
ENSG00000142765	SYTL1	12.13376992	1.62E-31	2.47E-30
ENSG00000271503	CCL5	12.10304745	3.97E-115	7.70E-113

PC analysis (PCA) and Volcano Plots

The Volcano Plot tool in r studio was used to create volcano charts after giving the relevant inputs. (FDR, p-value, logFC, etc). It is generally used to find genes or areas that are differentially expressed or methylated between two or more experimental settings. The upper right corner of the plot contains genes that are thought to be significantly differentially expressed or methylated between the experimental conditions because of their low p-value and high fold change .



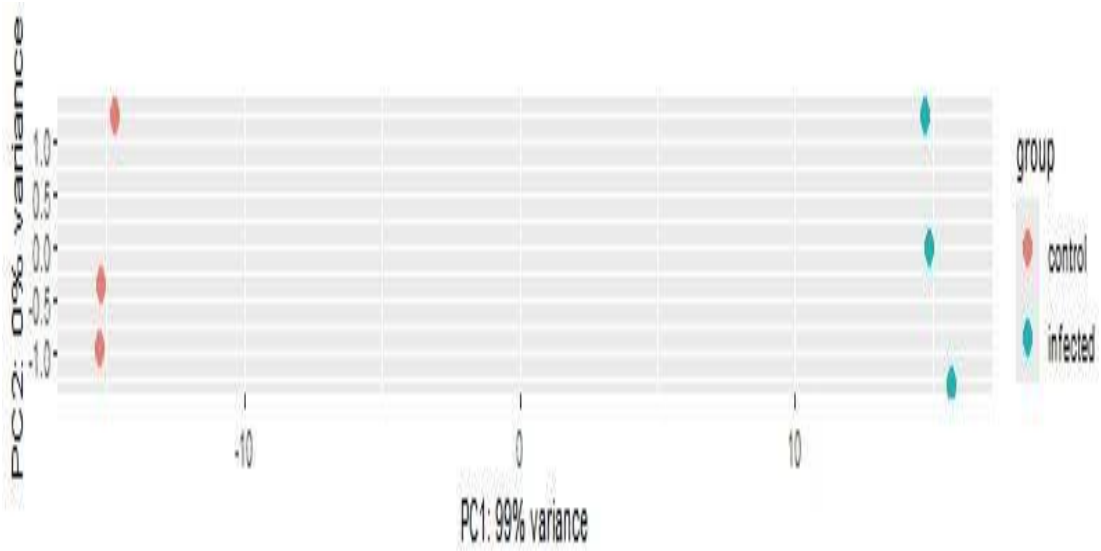
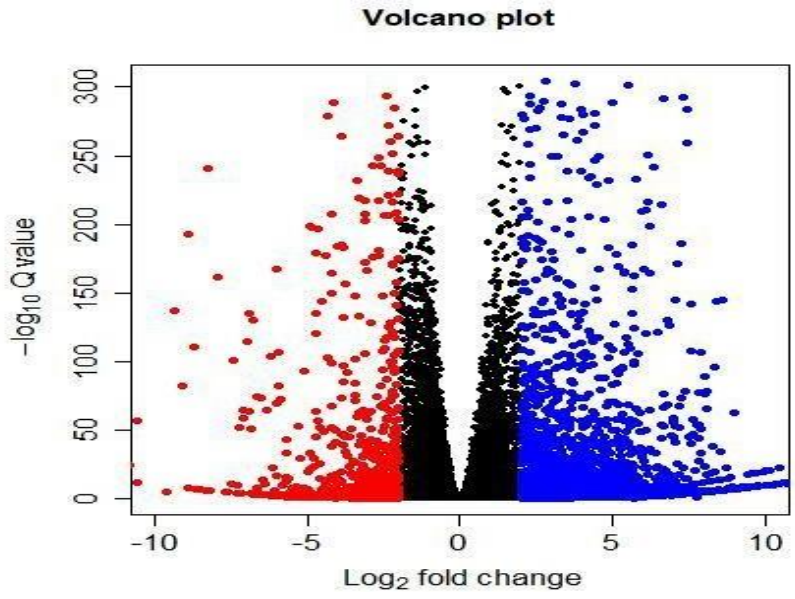
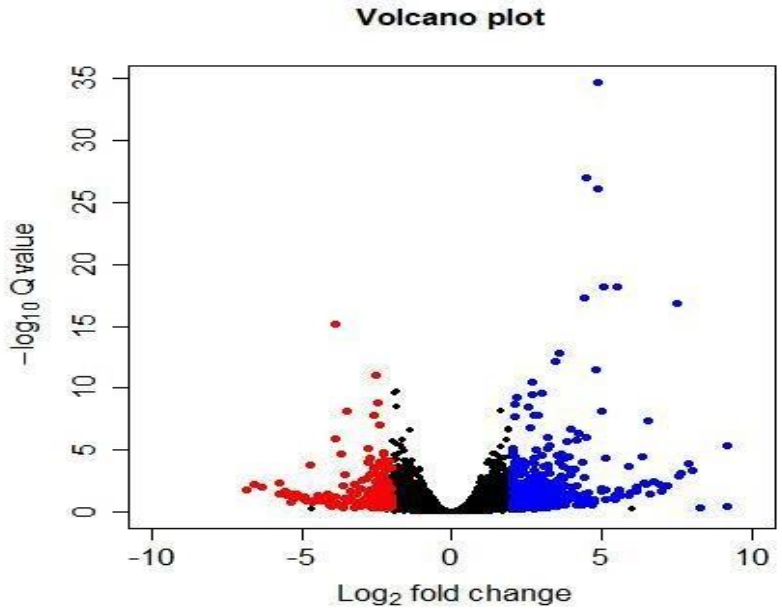


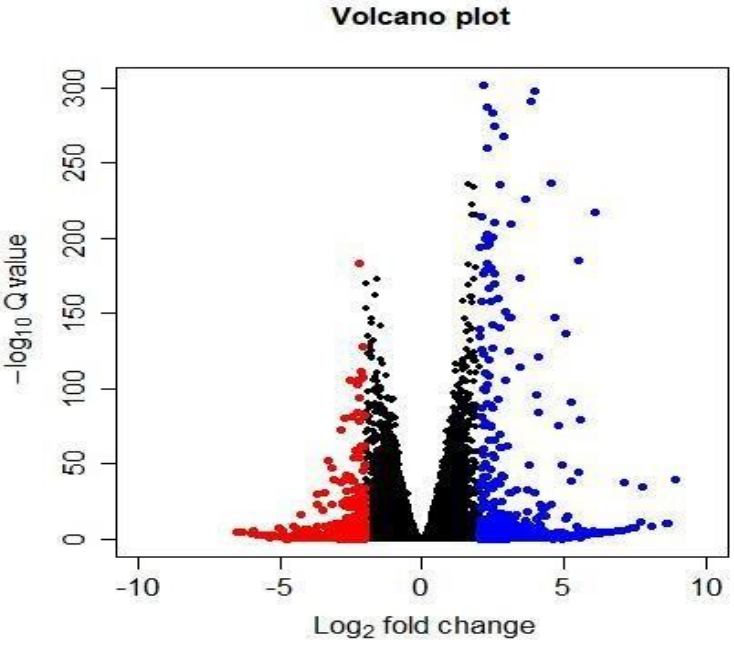
Figure : PCA plot of differentially expression analysis of control vs infected inall three cell lines



(A)



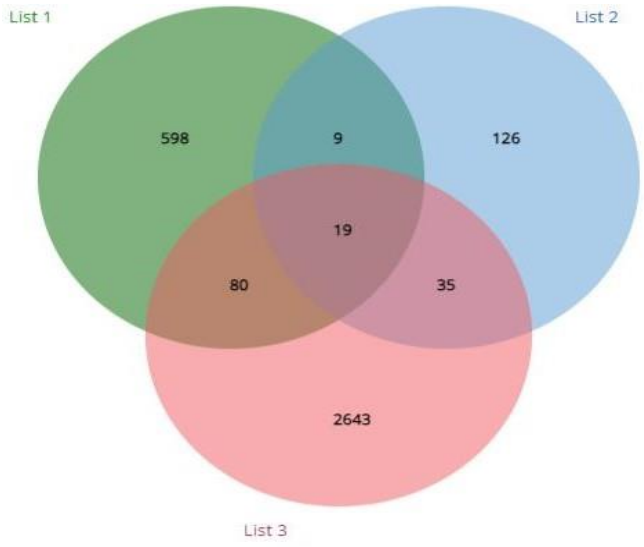
(B)



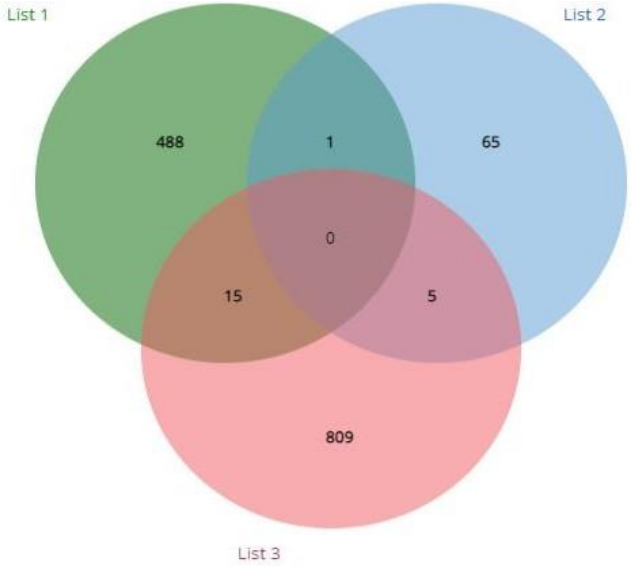
(C)

Figure : volcano plot representing dys-regulated genes in A) human brain organoid cell lines b) dendritic cell lines c)human kidney cell lines .

Comman degs :



(A) Upregulated



(B) Downregulated

Figure : Pie chart of DEGs between three datasets

Table : Common Degs From Three Cell Lines And Their Descreption

Ensemble Gene ID	Symbol	Chr	Position	Description
ENSG00000137959	IFI44L	1	78.619902	interferon induced protein 44 like
ENSG00000162654	GBP4	1	89.181144	guanylate binding protein 4
ENSG00000154451	GBP5	1	89.256189	guanylate binding protein 5

ENSG00000134321	RSAD2	2	6.865557	radical S-adenosyl methionine domain containing 2
ENSG00000078098	FAP	2	162.170684	fibroblast activation protein alpha
ENSG00000144802	NFKBIZ	3	101.827991	NFKB inhibitor zeta
ENSG00000169245	CXCL10	4	76.021118	C-X-C motif chemokine ligand 10
ENSG00000169248	CXCL11	4	76.033682	C-X-C motif chemokine ligand 11
ENSG00000138646	HERC5	4	88.457119	HECT and RLD domain containing E3 ubiquitin protein ligase 5
ENSG00000171855	IFNB1	9	21.077104	interferon beta 1
ENSG00000119922	IFIT2	10	89.283694	interferon induced protein with tetratricopeptide repeats 2
ENSG00000119917	IFIT3	10	89.327307	interferon induced protein with tetratricopeptide repeats 3
ENSG00000185745	IFIT1	10	89.392546	interferon induced protein with tetratricopeptide repeats 1
ENSG00000167550	RHEBL1	12	49.064676	RHEB like 1
ENSG00000135114	OASL	12	121.017763	2'-5'-oligoadenylate synthetase like
ENSG00000172183	ISG20	15	88.63567	interferon stimulated exonuclease gene 20
ENSG00000271503	CCL5	17	35.871491	C-C motif chemokine ligand 5
ENSG00000182393	IFNL1	19	39.296407	interferon lambda 1
ENSG00000105559	PLEKHA4	19	48.837097	pleckstrin homology domain containing A4

Functional and pathway enrichment analyses of DEGs

Further, the common DEGs were analyzed for the potential biological processes, cellular components, and molecular functions that were the gene ontology terms and the pathways using the online DAVID tool for performing the Gene Ontology (GO) analysis. The GO terms identified were utilized to annotate activities belonging to the biological processes (BP), cellular component (CC), or molecular function categories (MF). A total of 19 common DEGs were analyzed using the DAVID server to gain insights into the biological processes, molecular functions, and cellular components. The genes were submitted as Official Gene Symbols. The graph was then made using srplot tool (figure)

- **In the Biological Process (BP) term, the results demonstrated that the DEGs were mainly involved in:**

1. Negative regulation of viral genome replication
2. Biological process involves in interspecies interaction between organism

3. Defense response to virus
4. Defense response to symbiont
5. Innate immune response

- **In the cellular component (CC) term, the results demonstrated that the DEGs were mainly involved in:**

1. Host cellular component
2. Host cell

- **In the Molecular Function (MF) term, the results demonstrated that the DEGs were mainly involved in:**

1. CXCR Chemokine Receptor Binding
2. ISG 15 Transferase Activity
3. Cytokine Receptor Binding
4. Signaling Receptor Activator Activity

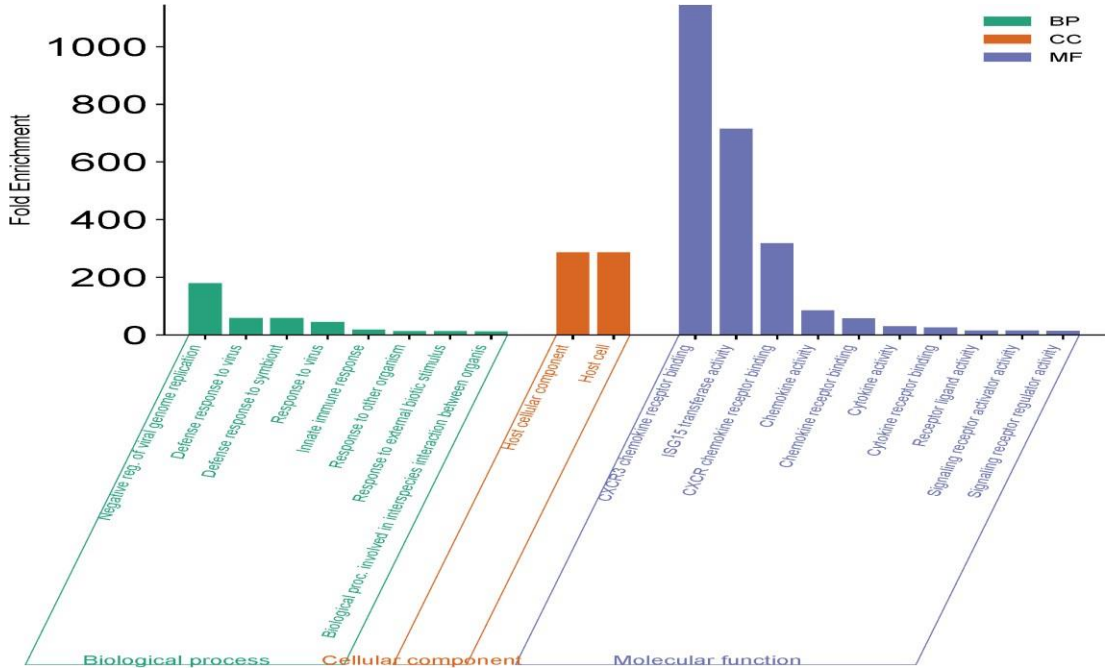


Figure: Functional Enrichment Analysis of common DEGs.

In the Pathway Analysis, the term DEGs were involved mainly in :

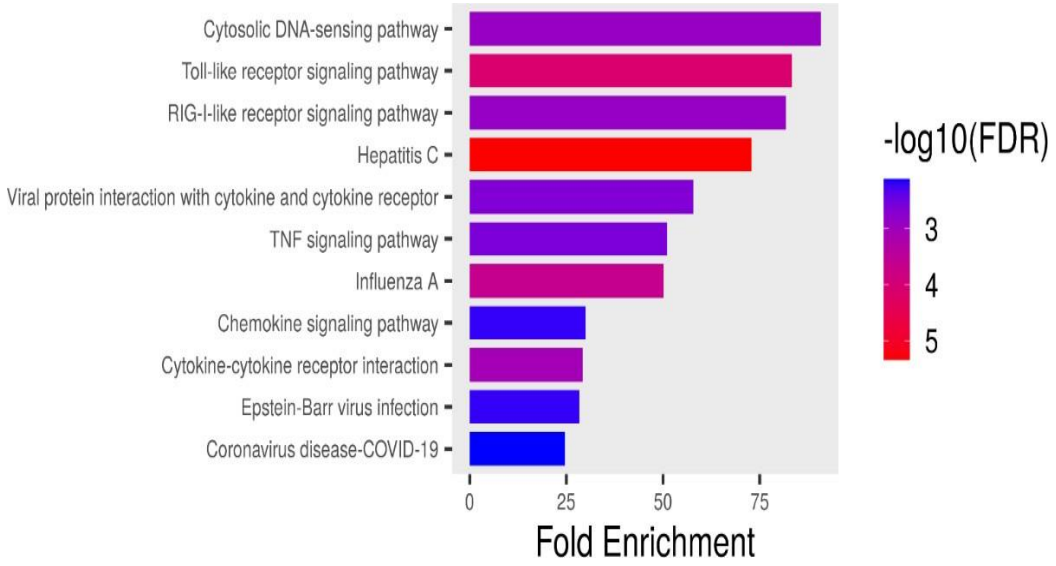


Figure : KEGG Pathway analyses of common DEGs

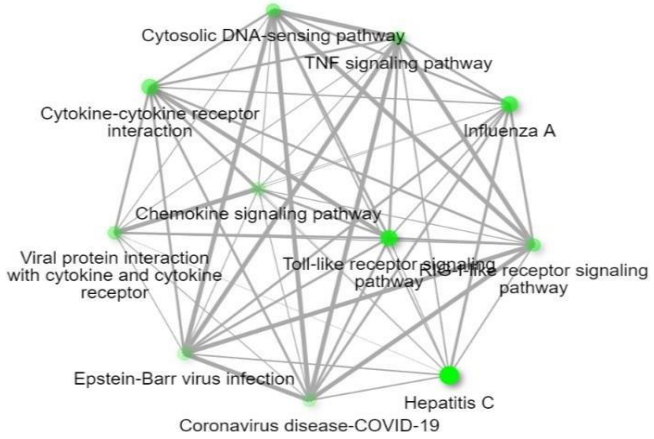


Figure: Network between pathways of top DEG's

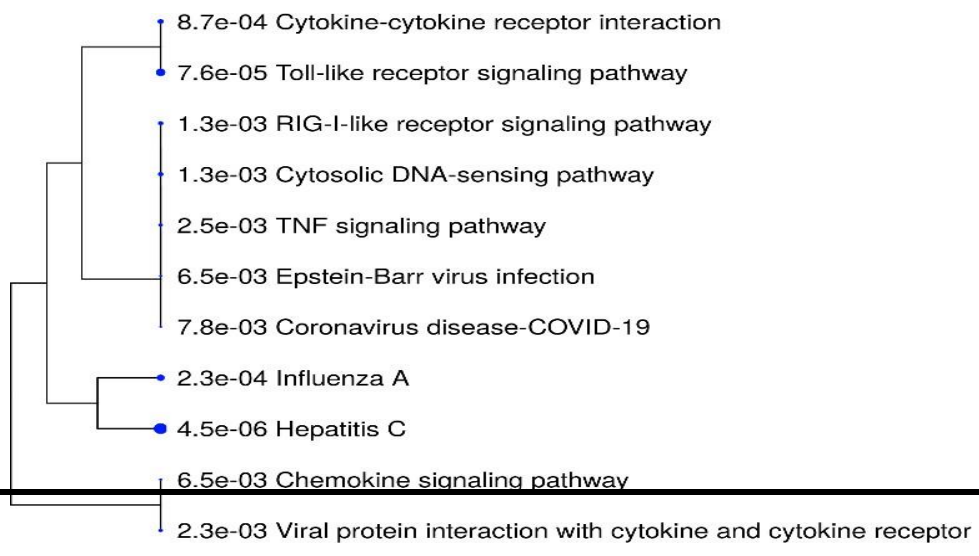


Figure : Tree analysis of pathway connected together

PPI Network Construction and Hub Gene Identification:

To find the interactive relationships between the shared DEGs, use the online search tool STRING. All 19 common DEGs were uploaded to the STRING database as Official Gene Symbols. In the study, interactions with a Confidence score of more than or equal to 0.4 were considered significant . The PPI network had nodes and edges, with an average node degree of.

The Cystoscope software was used to download and visualize the network . To identify the hub genes, the CytoHubba plugin was installed in Cytoscape. The top 10 hub genes were ranked and visualized based on the three criterion .

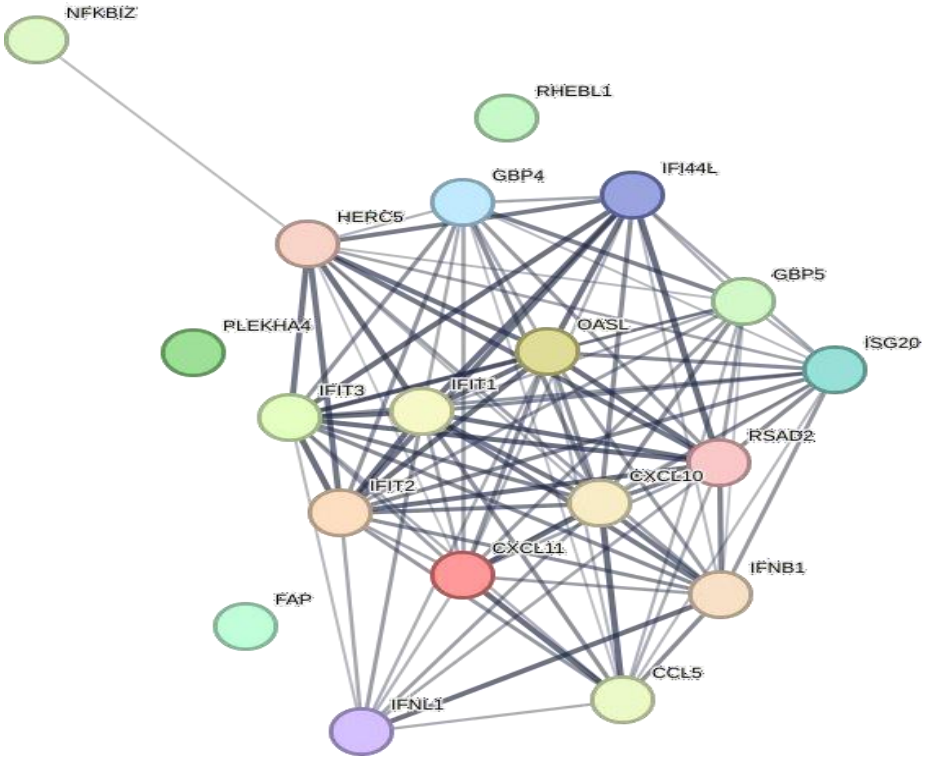


Figure : PPI network obtained from STRING Database

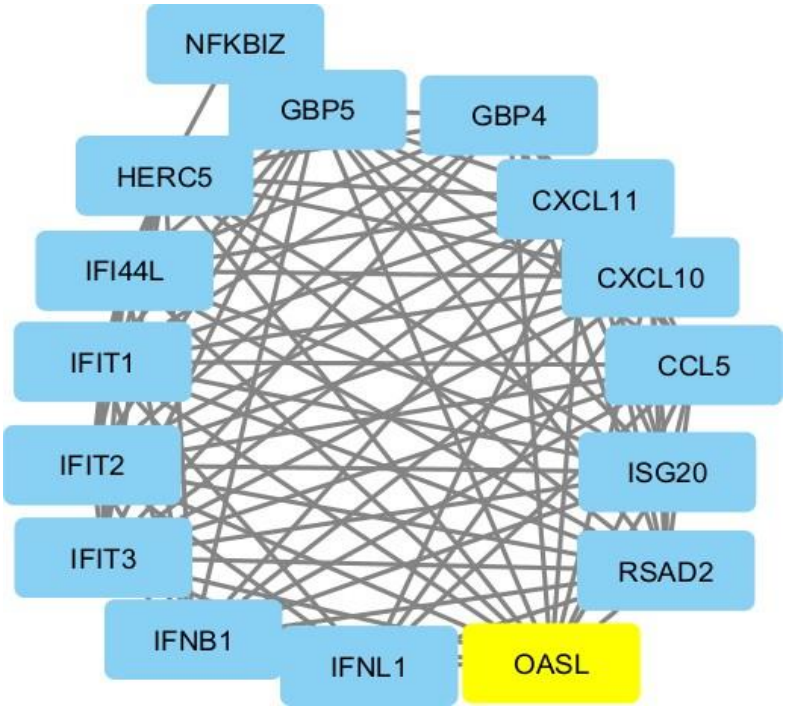
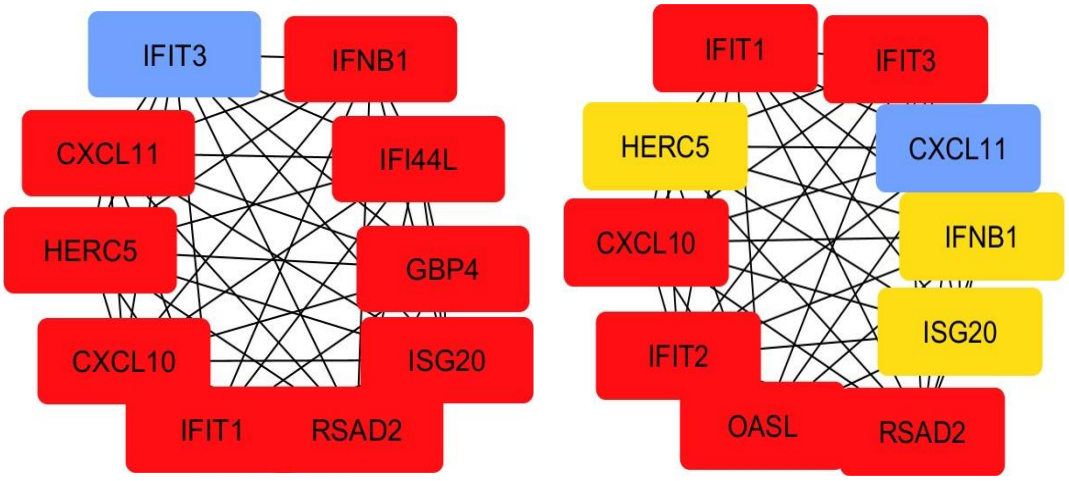
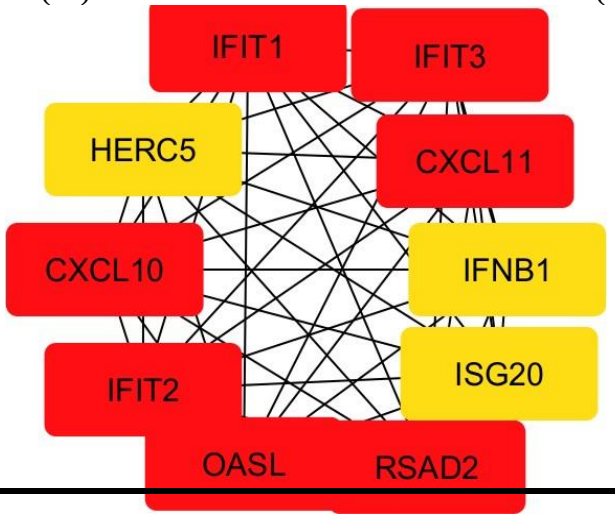


Figure : PPI Network as visualized in Cystoscope



(A) (B)



(C)

Figure: Top 10 Hub Genes ranked based on a) Degree, b) Eccentricity c)closeness.

Table: list of Hub gene with full name and their respective role in the biologicalsystem

Gene Symbol	Gene name	Summary
IFIT1	interferon induced protein with tetratricopeptide repeats 1	This gene encodes a protein containing tetratricopeptide repeats that was originally identified as induced upon treatment with interferon. The encoded protein may inhibit viral replication and translational initiation. This gene is located in a cluster on chromosome 10 with five other closely related genes. There is a pseudogene for this gene on chromosome 13. Alternatively spliced transcript variants encoding multiple isoforms have been observed

CXCL10	C-X-C motif chemokine ligand 10	This antimicrobial gene encodes a chemokine of the CXC subfamily and ligand for the receptor CXCR3. Binding of this protein to CXCR3 results in pleiotropic effects, including stimulation of monocytes, natural killer and T-cell migration, and modulation of adhesion molecule expression. This gene may also be a key regulator of the 'cytokine storm' immune response to SARS-CoV-2 infection
CXCL11	C-X-C motif chemokine ligand 11	Chemokines also play fundamental roles in the development, homeostasis, and function of the immune system, and they have effects on cells of the central nervous system as well as on endothelial cells involved in angiogenesis or angiostasis. Chemokines are divided into 2 major subfamilies, CXC and CC. This antimicrobial gene is a CXC member of the chemokine superfamily. Its encoded protein induces a chemotactic response in activated T-cells and is the dominant ligand for CXC receptor-3 IFN-gamma is a potent inducer of transcription of this gene.
IFIT3	interferon induced protein with tetratricopeptide repeats 3	Enables identical protein binding activity. Involved in negative regulation of apoptotic process; negative regulation of cell population proliferation; and response to virus.
RSAD2	radical S-adenosyl methionine domain containing 2	The protein encoded by this gene is an interferon-inducible antiviral protein that belongs to the S-adenosyl-L-methionine (SAM) superfamily of enzymes. The protein plays a role in cellular antiviral response and innate immune signalling. Antiviral effects result from inhibition of viral RNA replication, interference in the secretory pathway, binding to viral proteins and dysregulation of cellular lipid metabolism. The protein has been found to inhibit both DNA and RNA

		viruses, including influenza virus, human immunodeficiency virus (HIV-1) and Zika virus.
HERC5	HECT and RLD domain containing E3 ubiquitin protein ligase 5	This gene is a member of the HERC family of ubiquitin ligases and encodes a protein with a HECT domain and five RCC1 repeats. Pro-inflammatory cytokines upregulate expression of this gene in endothelial cells. The protein localizes to the cytoplasm and perinuclear region and functions as an interferon-induced E3 protein ligase that mediates ISGylation of protein targets. The protein also acts as a modulator of the antiviral immune response.
ISG20	interferon stimulated exonuclease gene 20	Enables 3'-5' exonuclease activity and RNA binding activity. Involved in defence response to virus; negative regulation of viral genome replication; and nucleobase-containing compound catabolic process
IFNB1	interferon beta 1	This gene encodes a cytokine that belongs to the interferon family of signalling proteins, which are released as part of the innate immune response to pathogens. The protein encoded by this gene belongs to the type I class of interferons, which are important for defence against viral infections. In addition, type I interferons are involved in cell differentiation and anti-tumour defences. Following secretion in response to a pathogen, type I interferons bind a homologous receptor complex and induce transcription of genes such as those encoding inflammatory cytokines and chemokines. Overactivation of type I interferon secretion is linked to autoimmune diseases.
OAS1	2'-5'-oligoadenylate synthetase 1	This protein plays a key role in innate cellular antiviral response, and has been implicated in other cellular processes like cell growth and apoptosis. Alternative splicing results in multiple transcript variants with different enzymatic activities. Polymorphisms in this gene have been associated

		with susceptibility to viral infection, including SARS-CoV-2, and diabetes mellitus, type 1

Variant Calling and filtering

Before we using bcftools, we use RMDUP tools for remove PCR duplicates from our aligned file (BAM). It is important to use the tool in the data processing pipeline to ensure the accuracy and reliability of downstream analyses. For SNPs studies here we use bcftools. Bcftools can handle various types of NGS data, including single nucleotide polymorphisms (SNPs), insertions, deletions, and structural variations. Here we take the alignment file (BAM) of all three cell lines for variant calling.

To identify SNPs, variant calling was carried out for each read merging method using BCFtools. The 'mpileup' function was utilized to generate read alignments, and the 'call' function was employed to detect and analyze the variants .

. Table: bcftools mpileup file

Chr	REF	ALT	Position	QUAL
chr1	C	T	629906	228.407
chr1	A	G	818802	110.415
chr2	A	C	189675990	71.4148
chr3	G	A	117292408	132.416
chr7	T	G	79263418	105.415
chr10	C	A	132107553	100.813

chr15	T	C	68704102	123.415
chr18	C	T	75283838	72.7715
chr22	G	A	20573979	112.723

From mpileup, we use the "bcftools call" tool is used to identify and call variants from the input alignment. After this, we use VCF Filter for filtering the SNPs .

Filter based on, QUAL>50, DP> 10.

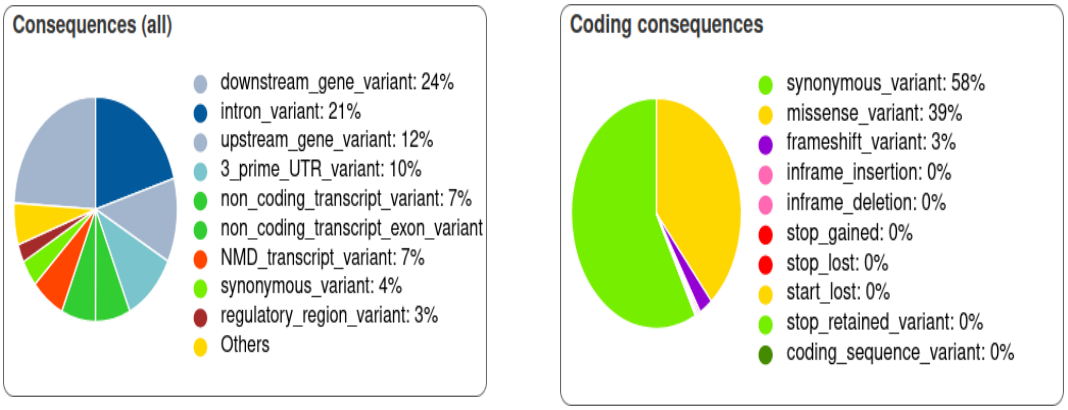
Annotation:

For SNPs annotation here we use a web-based tool called WANNOVAR. We upload the input file as a VCF file to WANNOVAR. Here we check the annotation of all three vcf file of different cell lines and selected the degs if observed and proceed for further analysis. Here we found some degs as follows having synonymous and non-synonymous SNVs out of which GBP4 was found to be the most damaged according to the master predictor.

Table : Table containing information about SNPs after annotation

Chr	Start	End	Ref	Alt	Gene	ExonicFunc	Mutation Taster_pred
1	89651028	89651028	T	C	GBP4	nonsynonymous SNV	N
1	89652027	89652027	G	T	GBP4	nonsynonymous SNV	N
1	89652761	89652761	A	G	GBP4	synonymous SNV	P
1	89654433	89654433	C	T	GBP4	synonymous SNV	P
1	89662942	89662942	G	A	GBP4	synonymous SNV	P
1	89733145	89733145	C	G	GBP5	nonsynonymous SNV	D
4	88475940	88475940	A	G	HERC5	nonsynonymous SNV	P
4	88494245	88494245	T	C	HERC5	synonymous SNV	.
10	89306703	89306703	T	A	IFIT2	synonymous SNV	.
10	89307012	89307012	C	A	IFIT2	nonsynonymous SNV	P
10	89402740	89402740	A	G	IFIT1	synonymous SNV	.
15	88639471	88639471	T	C	ISG20	synonymous SNV	.
19	48865586	48865586	T	C	PLEKHA4	nonsynonymous SNV	P

Further for more details about the SNVs another annotation tool was used VEP (Variant Effect Predictor) to predict the functional consequence of genetic variations. Pictorial Representation consequence refers to the effect of genetic variations on the structure, function, and expression of genes and their products .



Category	Count
Variants processed	51882
Variants filtered out	0
Novel / existing variants	4647 (9.0) / 47235 (91.0)
Overlapped genes	18699
Overlapped transcripts	114889
Overlapped regulatory features	9245

Figure : Pictorial representation consequences of SNPs

Conclusion

In conclusion, this comprehensive research project focused on zika virus host pathogenesis in different cell lines and aimed to identify potential biomarkers and understand the underlying genetic mechanisms. Through the analysis of RNA sequencing data, several differentially expressed genes (DEGs) were identified, reflecting the dysregulation of crucial biological processes such as negative regulation of viral genome replication, innate immune response, and metabolic pathways. These findings provide valuable insights into the molecular alterations associated with zikv infection.

Pathway analysis revealed the involvement of various signalling pathways, including the cytosolic DNA-sensing pathway, Toll like receptor signalling, RIG-1 like receptor signalling, TNF signalling, chemokine signalling pathways, suggesting their potential roles in zikv pathogenesis. Furthermore, the construction of a protein-protein interaction (PPI) network led to the identification of ten hub genes (IFIT1, CXCL10, IFIT3, CXCL11, RSAD2, HERC5, ISG20, IFNB1, OAS1), which hold promise as potential biomarkers for zikv infection. These genes play critical roles in processes such as immune response, INF regulation, viral replication and neuronal damage, indicating their potential relevance to zikv infection and disease such as microcephaly. Moreover, the study investigated the impact of single nucleotide polymorphisms (SNPs) on infected cell lines. Analysis of these genetic variations highlighted potentially damaging SNPs within the GBP5 gene, as well as other genes associated with ZIKV infection. These SNPs may contribute to the upregulation of biomarkers and provide insights into the functional consequences of genetic variations in zikv infected cell lines.

The findings of this research contribute to our understanding of zikv pathogenesisby providing valuable information about dysregulated genes, disrupted pathways, and potential biomarkers. These insights could have significant implications for early detection, risk assessment, and the development of targeted interventions for zikv infection. Ultimately, this study contributes to the broader goal of improving diagnosis, treatment, and management strategies for individuals affected by this devastating form of ZIKA virus

IFNB1 and IFNL1 belong to the interferon family of cytokines, which encompasses various type I and type III interferons. These cytokines play a crucial role in orchestrating the host response to viral infections and other pathogens. They initiate signaling cascades that culminate in the induction of antiviral proteins, suppression of viral replication, modulation of immune cell activity, and establishment of an antiviral state within infected cells. Notably, ISG20 has been reported to upregulate other interferon response proteins, such as IFIT1 genes, thereby impeding translation of the alphavirus genome. This underscores the central role of interferon-stimulated gene 20 (ISG20) in controlling Zika virus infection in trophoblast cells.

Studies reveals that ISG20, functioning as a 3′-5′ exonuclease, possesses the capability to degrade Zika virus (ZIKV) RNA, thereby impeding viral replication within first-trimester placental trophoblast cells. We have identified and characterized ISG20 as a pivotal regulator in controlling ZIKV infection specifically within trophoblast cells.

Nevertheless, recent reports indicate a contrasting role for ISG15 in promoting Zika virus (ZIKV) infection. It has been found to participate in facilitating ZIKV infection by modulating the JAK/STAT and ISGylation pathways. Specifically, ISGylation refers to the process of protein conjugation with ISG15, which has been implicated in enhancing ZIKV infectivity .

GBP5 gene codes for Guanylate-binding proteins (GBPs) that belong to the GTPase superfamily of ISGs. Guanylate-binding proteins (GBPs), including GBP4, are induced by interferon and function as guanosine triphosphatases (GTPases), hydrolyzing GTP to both GDP and GMP. This family of proteins, comprising a subset of interferon-inducible GTPases, plays a crucial role in bolstering host defenses against a diverse array of intracellular pathogens. Human cells harbor seven distinct GBPs, each contributing to various cellular functions such as impeding cell spreading and proliferation, activating the inflammasome, and exhibiting antimicrobial activity against viruses, bacteria, and protozoans.

Experiments conducted on transfected HEK293T cells and THP-1 knockout cells have revealed that GBP2/5 significantly inhibit the replication of both Brazilian and

African Zika virus strains. Interestingly, both strains possess a Furin consensus sequence in their premembrane protein (prM). GBP2/5 exhibit a broad antiviral activity, restricting various viral pathogens that rely on Furin-mediated processing of their envelope glycoproteins for full infectivity. Furthermore, the knockout of GBP5 has been observed to moderately attenuate the inhibitory effect of interferon-gamma (IFN-γ) on influenza A virus, suggesting that GBP5 is among the multiple IFN-γ-inducible antiviral factors .

Using string and cystoscope ppi network and hub gene were found, Top 10 COMMON HUB genes by using degree, closeness, Eccentricity method are IFIT1, CXCL10, IFIT3, CXCL11, RSAD2, HERC5, ISG20, IFNB1, OAS1. Most of which were associated with antiviral functions, such as activation of the complement system (C3), inhibition the viral replication (IFIT1, IFIT3, ISG20, OAS1, and RSAD2), participate in viral protein interaction with cytokine and cytokine receptor (CXCL10, CXCL11), some have a key role in pathway such as TNF signaling pathway and chemokine signaling pathway (IFNB1, CXCL10 and HERC5). Information of the gene names and corresponding functions summarized in NCBI database .

The above results showed that most of the common genes among all organs were related to general antiviral responses. Besides, some of them could also be detected in other viral responses. For instance, RSAD2 was found to inhibit several viruses including influenza virus, ZIKV and HIV-1, CXCL10 gene may be a key regulator of the 'cytokine storm' immune response to SARS-CoV-2 infection, OAS1 protein plays a key role in innate cellular antiviral response, and has been implicated in other cellular processes like cell growth and apoptosis. Polymorphisms in this gene have been associated with susceptibility to viral infection, including SARS-CoV-2, and diabetes mellitus type 1.

Before variant calling using bcftools, the RMDUP tool was used to remove PCR duplicates from the aligned file (BAM). Using the RMDUP tool is crucial in the data processing pipeline to ensure accurate and reliable downstream analyses. For variant calling Bcftools is a powerful tool for handling various types of NGS data, including SNPs, insertions, deletions, and structural variations. After variant calling, VCF Filter was used for filtering SNPs based on the conditions QUAL>50 and DP>10. After