Virus Research 284 (2020) 197986

Contents lists available at ScienceDirect

Virus Research

journal homepage: www.elsevier.com/locate/virusres

An *in silico* integrative protocol for identifying key genes and pathways useful to understand emerging virus disease pathogenesis

Gabriel Augusto Pires de Souza[a,b], Ezequiel Aparecido Salvador[a], Fernanda Roza de Oliveira[a], Luiz Cosme Cotta Malaquias[a], Jonatas Santos Abrahão[b], Luiz Felipe Leomil Coelho[a,*]

[a] Laboratório de Vacinas, Instituto de Ciências Biomédicas, Departamento de Microbiologia e Imunologia, Universidade Federal de Alfenas, Alfenas, Minas Gerais, Brazil
[b] Laboratório de Vírus, Instituto de Ciências Biológicas, Departamento de Microbiologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

ARTICLE INFO

Keywords:
Integrative bioinformatics
Emerging virus disease
Pathogenesis

ABSTRACT

The pathogenesis of an emerging virus disease is a difficult task due to lack of scientific data about the emerging virus during outbreak threats. Several biological aspects should be studied faster, such as virus replication and dissemination, immune responses to this emerging virus on susceptible host and specially the virus pathogenesis. Integrative *in silico* transcriptome analysis is a promising approach for understanding biological events in complex diseases. In this study, we propose an *in silico* protocol for identifying key genes and pathways useful to understand emerging virus disease pathogenesis. To validate our protocol, the emerging arbovirus *Zika virus* (ZIKV) was chosen as a target micro-organism. First, an integrative transcriptome data from neural cells infected with ZIKV was used to identify shared differentially expressed genes (DEGs). The DEGs were used to identify the potential candidate genes and pathways in ZIKV pathogenesis through gene enrichment analysis and protein-protein interaction network construction. Thirty DEGs (24 upregulated and 6 downregulated) were identified in all ZIKV-infected cells, primarily associated with endoplasmic reticulum stress and DNA replication pathways. Some of these genes and pathways had biological functions linked to neurogenesis and/or apoptosis, confirming the potential of this protocol to find key genes and pathways involved on disease pathogenesis. Moreover, the proposed *in silico* protocol performed anintegrated analysis that is able to predict and identify putative biomarkers from different transcriptome data. These biomarkers could be useful to understand virus disease pathogenesis and also help the identification of candidate antiviral drugs.

1. Introduction

Outbreaks of emerging and/or re-emerging viral infections are common threats to human health. The emerging viral pathogens are defined as novel viruses that have been recently introduced in a population. In the last years, several medical important infectious outbreaks have occurred worldwide, including West Nile virus, Chikungunya virus, Zika virus, SARS, MERS, influenza and nCov-2019 (Ellwanger and Chies, 2016; Olival et al., 2017). Nevertheless, the knowledge and understanding of virus pathogenesis is a difficult task due to the lack of scientific data about the emerging virus during outbreak threats. Several biological aspects should be studied faster, such as virus replication and dissemination, immune responses to this emerging virus on a susceptible host and especially virus pathogenesis. A rapid and regular focus on basic research about virus pathogenesis is imperative and could help researches to better understand the pathogenesis of these viruses and also to develop therapeutics and vaccines and to control the disease (Afrough et al., 2019).

Changes in gene expression is a hallmark of virus infection and the identification of the altered Gene expression profiling is very important to understand the virus pathogenesis and also to the development of new antivirals drugs and vaccines for the emerging/re-emerging viruses. Recent developments in high-throughput sequencing technologies result in an intense accumulation of *omics* data from cells, tissues and patients infected with different viruses. Integrative *in silico* transcriptome analysis (ITA) is a promising approach for understanding biological events in complex diseases. ITA also contribute to understandings of disease pathogenesis and to find new drugs to treat the disease. For example, ITA was used to found genes with significant associations with Alzheimer's disease (Jin et al., 2018), preeclampsia (Moslehi et al., 2013), HIV (Liao et al., 2017), lung cancer (Cinegaglia et al., 2016) and hepatocellular carcinoma (Hoshida et al., 2009).

* Corresponding author at: Laboratório de Vacinas, Instituto de Ciências Biomédicas, Universidade Federal de Alfenas, Rua Gabriel Monteiro, 700 Centro, CEP: 37130-000, Alfenas, Minas Gerais, Brazil.
E-mail address: luiz.coelho@unifal-mg.edu.br (L.F. Leomil Coelho).

https://doi.org/10.1016/j.virusres.2020.197986
Received 2 March 2020; Received in revised form 21 April 2020; Accepted 21 April 2020
Available online 24 April 2020

In this study, we propose an *in silico* protocol for identifying key genes and pathways useful to understand emerging virus disease pathogenesis. To validate our protocol, the emerging arbovirus *Zika virus* (ZIKV) was chosen as a target micro-organism. ZIKV is an enveloped virus with approximately 50 nm in diameter with a single-stranded positive RNA genome. It belongs to the family *Flaviviridae* and genus *Flavivirus*. ZIKV was first isolated in the Zika Forest, Uganda, in 1947 (Shi and Gao, 2017; Yun and Lee, 2017). The virus is transmitted to humans by hematophagous mosquitoes, mainly those belonging to the *Aedes* genus (Petersen et al., 2016). The virus circulation was reportedly restricted in endemic areas of Africa and Asia for several decades. Individuals infected by ZIKV in these areas exhibited a self-limited febrile disease. ZIKV started to gain attention from international public health communities in 2013 when an outbreak occurred in French Polynesia (Musso et al., 2015). In Brazil, the high rate of ZIKV infection was established as one of the leading causes of an increase in cases of newborns with microcephaly in the Brazilian Northeast. The severity of the reported neurological deficiency in the ZIKV cases led the World Health Organization (WHO) to propose measures to intensify ZIKV prevention. They also suggested the surveillance and investigation of the correlation between it and microcephaly in neonates and Guillain-Barré syndrome generally in adults (Larocca et al., 2016; Laura et al., 2018).

Our proposed integrative transcriptome analysis protocol using data from neural cells infected with ZIKV was able to identify shared differentially expressed genes (DEGs). The DEGs were used to identify the potential candidate genes and pathways in ZIKV pathogenesis through gene enrichment analysis and protein-protein interaction network construction. Some of the genes, pathways and biological functions were linked to neurogenesis and/or apoptosis, confirming the potential of this protocol to find key genes and pathways involved in disease pathogenesis. Moreover, the proposed *in silico* protocol performed an integrated analysis that is able to predict and identify putative biomarkers from different transcriptome data. These biomarkers could be useful to understand virus disease pathogenesis and also help the identification of candidate antiviral drugs.

2. Materials and methods

2.1. Transcriptome data of neural cells infected with Zika virus

Gene expression profiling data from ZIKV-infected human neural cells were retrieved from RNA-sequencing data from the *Gene Expression Omnibus* platform (https://www.ncbi.nlm.nih.gov/gds). The platform was extensively searched for datasets of any neural-origin cell type that was infected with ZIKV. Information about the number of samples used in each experimental condition, cell type, ZIKV strain and time of infection were collected in an Excel spreadsheet. The number of each dataset was recorded, and the raw sequence data were accessed using the online Galaxy platform (https://usegalaxy.org/). After the sequencing data was uploaded on the Galaxy platform, the "Tuxedo suite" protocol was applied to identify differentially expressed genes among uninfected and infected cells (Amrit and Ghazi, 2017). This protocol uses a set of tools to analyze the quality and variety of RNA-sequencing data, including short-readout mapping, splicing junction identification and differential expression detection of transcripts and isoforms, as well as tools for data visualization and control metrics of sequencing quality. After the protocol was completed, a list of all DEGs from each study with log fold change greater than 1.5 and p value $\leq$ 0.05 was collected in an Excel spreadsheet. Only datasets that were performed with at least two replicates for each experimental sample and that had a minimum of 100 DEGs were used for the subsequent analysis. Therefore, once a successful Galaxy run is completed we used the final DEG list identified from each study to perform an integrative analysis. This integrative analysis aims find genes and pathways which were coincident among all DEG lists and therefore more likely to be involved in the etiology.

2.2. Network construction

All networks presented in this work were built using Gephi version 0.9.2 (Bastian and Heymann, 2009). The cells and genes were listed in a comma-separated values (.csv) spreadsheet for each graph, and this file was imported into the software. Another. csv spreadsheet with the connections between the cells and genes was also imported to generate the network graph. In all networks, the node diameter is directly proportional to the edge degree. The layout was generated using algorithms based on the force of attraction and repulsion of the nodes (Force Atlas 2). The nodes were submitted to local rearrangement for better visualization of the connections between nodes.

2.3. Pathway and Gene Ontology (GO) enrichment analyses of DEGs

The Enrichr platform (http://amp.pharm.mssm.edu/Enrichr/) was used to study pathway enrichment analyses of DEGs (Kuleshov et al., 2016). The Enrichr online tool for GO (http://www.geneontology.org/; Harris et al., 2004) was used to determine the function of DEGs. Data from cellular components, biological processes and molecular functions were recorded from each set of genes. An adjusted $p \leq 0.05$ was considered statistically significant for all analyses. The lists of significant GO terms were submitted to the Reduce + Visualize Gene Ontology tool (REVIGO; Supek et al., 2011) and GO terms were visualized in an interactive graph tool. Results were exported into Cytoscape software to create graph-based visualization of the identified terms for each GO category.

2.4. Protein-protein interaction (PPI) network construction

The candidate DEGs were searched in the Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database, version 11 (stringdb.org). A combined score > 0.4 was used as the criteria to establish the PPI network. The isolated nodes were deleted from the network, and the final network was downloaded as a simple tabular text file. This network file was imported into Cytoscape software (version 3.7.1) for visualization.

3. Results

3.1. Differential expressed genes on ZIKV infected cells

As illustrated in Fig. 1, a six-step immunoinformatic approach was used to identify identifying key genes and pathways useful to understand emerging virus disease pathogenesis. To validate our protocol, the emerging arbovirus ZIKV was chosen as a target micro-organism. An extensive search using public databases and indexed publications was performed to find transcriptomic studies from ZIKV-infected neural cells. The search resulted in the selection of seven studies with transcriptome data from different human neural cell lines or cerebral organoids (CO) infected at different multiplicity of infection (MOI) and different times of infection. Experimental parameters, including ZIKV strains, MOI, number of replicates and next-generation sequencing (NGS) platform, used in each study are described in Table 1. The predominant cell type used for ZIKV infection was human cortical neural progenitors (HCNP). However, transcriptome data from other ZIKV-infected cells were also identified including microglia (MCG), glioblastoma stem cells (GSC), human neural crest cells (HNCC), human peripheral neurons (PN) and CO. Different ZIKV strains (MR766, Mex-1, PRVABC59 and Dakar 41,519) and a prolonged time of infection (69.8 ± 29.96 h) were used by these studies.

To analyze changes in cellular gene expression after ZIKV infection, only datasets that had at least duplicates of the experimental conditions were used. Therefore, the dataset from ZIKV-infected microglia (Tiwari et al., 2017) was excluded from the subsequent analysis. The Tuxedo suite protocol was performed using the transcriptome, and the

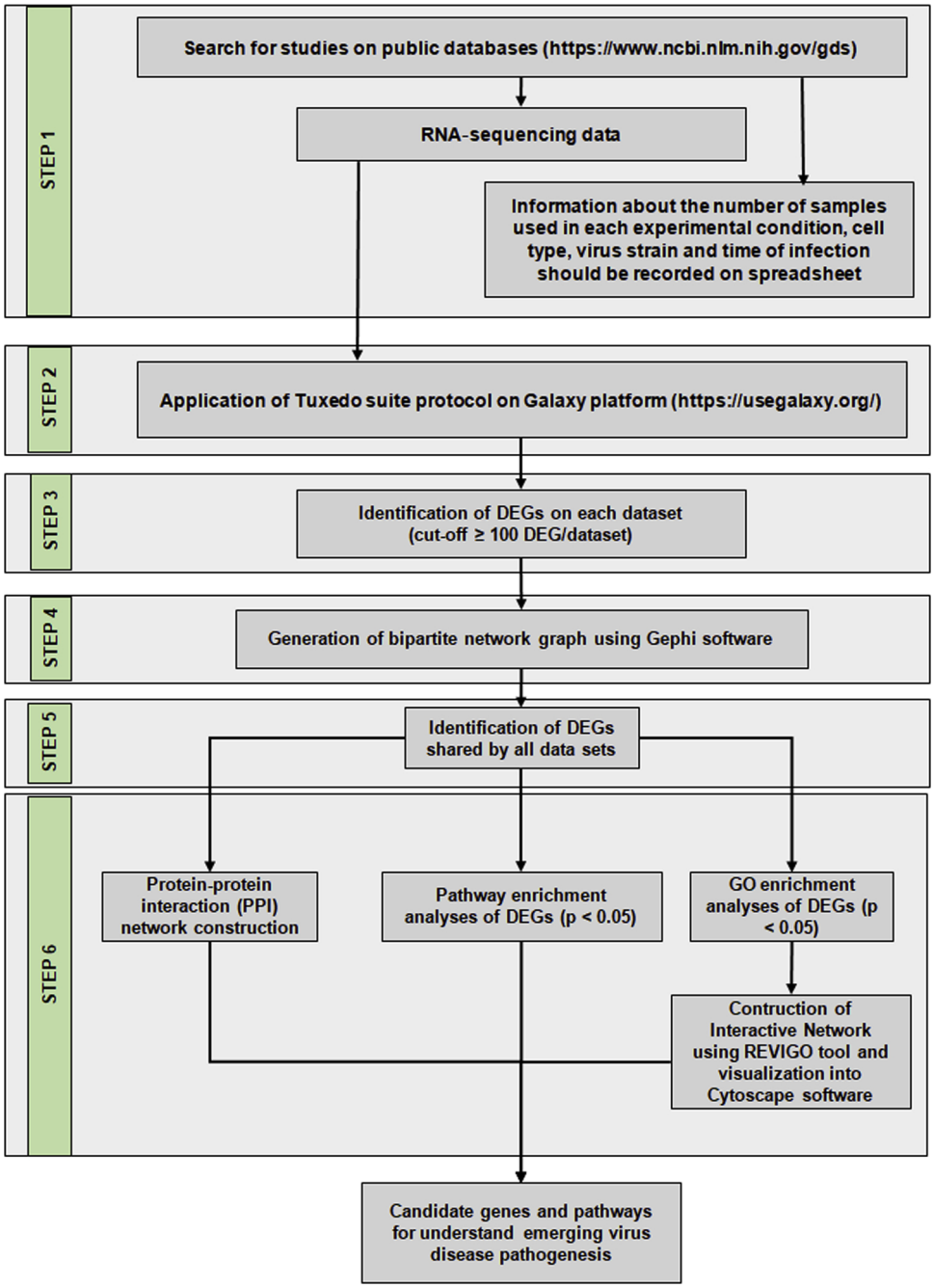


Fig. 1. Schematic representation of entire *in silico* approach for identification of key genes and pathways useful to understand emerging virus disease pathogenesis.

statistically significant DEGs were recorded. Applying the Tuxedo protocol on datasets generated two sets of results. The first one comprised the results from HCNP and GSC datasets there was a high number of DEGs in these cells after ZIKV infection. The second group was related to HNCC, PN and CO datasets, and it was characterized by a low number of DEGs after ZIKV infection (Supplementary Table 1).

To construct a final list with the DEGs in HCNP after ZIKV infection, the lists of downregulated and upregulated genes obtained in each dataset (GSE78711, GSE80434 and GSE93385) were compared to use only the genes that were considered differentially expressed in at least two studies. For ZIKV-infected CO, only the dataset derived from 3-day infection was used, because the dataset from 5-day infected CO had a low number of DEGs (≤ 100). The final list with the identified DEGs (n = 11,049) was used to build a bipartite network graph to better represent the effect of ZIKV infection on each cell type (Table 2). The network graph for upregulated genes was composed of 5311 nodes, with 5306 genes and 5 cells all connected by 6424 edges with the same weight (w) = [1]. A network graph for downregulated genes was composed of 3688 nodes, with 3683 genes and 5 cells all connected by 4611 edges with the same weight (w) = [1]. The network of upregulated and downregulated genes showed a very similar pattern (Fig. 2A). Network analysis revealed a marked change in GSC and HCNP gene expression once these cells exhibited a high number of DEGs. There were also changes in HNCC, PN and CO gene expression, but the number of DEGs in these cells after ZIKV infection was relatively low. Furthermore, the network analysis revealed a high number of DEGs shared by GSC and HCNPC, but a low number of common DEGs among all cells. Data analysis also demonstrated that the majority of upregulated DEGs were expressed in one cell type (4338/ 81.71 %). Only 851 genes (16.02 %) were upregulated in two cell types, 94 genes (1.77 %)

Table 1
Summary of datasets and sample details.

Study	GEO Acces	Nº of samples		Cell Type	ZIKV Strain	MOI	Time of infection	NGS Platform
		Control Cells	Infected cells					
Tang et al. (2017)	GSE78711	3	3	Human cortical neural progenitors	MR766	< 0.1	56 h	GPL18573 IlluminaNextSeq 500 GPL15520 IlluminaMiSeq
Zhang et al. (2016)	GSE80434	4	4 2	Human cortical neural progenitors	MR766 FSS13025	0.02 0.04	64 h 64 h	GPL11154 IlluminaHiSeq 2000 GPL15520 IlluminaMiSeq GPL18573 IlluminaNextSeq 500
McGrath et al (2017)	GSE93385	9	9	Human cortical neural progenitors	Mex1 – 7	10	5 days	GPL18460 IlluminaHiSeq 1500
Tiwari et al. (2017)	GSE93870	1	1	Microglia	MR766	1	24 h	GPL18573 IlluminaNextSeq 500
Zhu et al. (2017)	GSE102924	3	3	Glioblastoma stem cells	Dakar 41,519	5	36 – 48 h	GPL21290 IlluminaHiSeq 3000
Oh et al. (2018)	GSE87750	3	4	Human neural crest cells	PRVABC59	0.4	65 h	GPL16791 IlluminaHiSeq 2500 GPL21290 IlluminaHiSeq 3000
	GSE87750	2	4	Human Peripheral neurons	PRVABC59	0.4	65 h	GPL16791 IlluminaHiSeq 2500 GPL21290 IlluminaHiSeq 3000
Watanabe et al. (2017)	SE104279	3	3	Cerebral organoids	PRVABC59	0.3	3 or 5 days	GPL21290 IlluminaHiSeq 3000

Table 2
Number of Differentially expressed genes used to construct interaction networks among ZIKV infected cells.

Cell type	Up	Dow	Total
Human cortical neural progenitors	3586	1821	5407
Glioblastoma stem cells	2487	2537	5024
Human neural crest cells	188	137	325
Peripheral neurons	99	44	143
Cerebral organoids	76	74	150

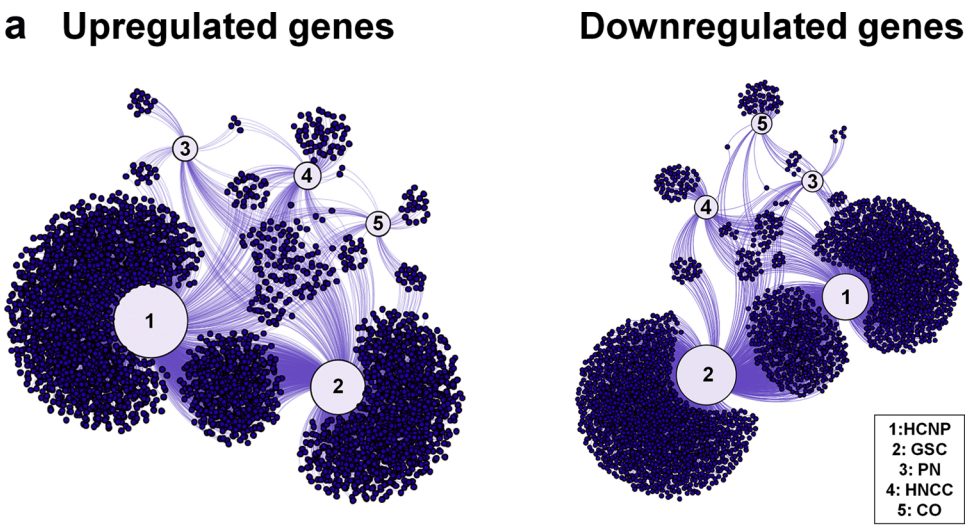
in three, 23 genes (0.43 %) in four and 3 genes (0.05 %) in all cell types. Regarding downregulated DEGs, 3207 genes (82.59 %) were expressed in only one cell type, 630 genes (16.22 %) in two, 40 genes (1.03 %) in three and 6 genes (0.15 %) in four (Fig. 2B).

3.2. Protein-Protein interaction network using the DEGs shared by all cells types after ZIKV infection

The results showed that the differential effect of ZIKV on cell transcriptome could be related to the differentiation status of the cell. Whereas the CO is composed of several cell types, using DEGs from this type of tissue can induce an error in the integrative analysis of the transcriptome of ZIKV-infected cells. Therefore, a new analysis was performed using only the DEGs identified in HCNP, GSC, HNCC and PN cells after ZIKV infection. Using this approach, we identified DEGs expressed in all cells (Fig. 3A; Table 3; Supplementary Table 2). The PPI network was constructed using the DEGs shared by all cells. The PPI network of upregulated DEGs was composed of 14 nodes and 28 edges, and the PPI network of downregulated genes was composed of 4 nodes and 6 edges. DEGs that encode proteins without any interaction were excluded from the PPI networks (upregulated: CLK1, EIF1B, GOT1, MXD1, SEC11C, SEC24D, SLC25A25, STMN4, TSPYL2 and TUFT1; downregulated: PRTG and IGFBPL1). To investigate the expression level of the target genes of ZIKV infected host cells, the RNA-seq transcriptomic expression data sets were analyzed and the fold change values were calculated comparing the mock and ZIKV treated cells. The DEGs shared by all cells were highly differentially modulated compared with the control condition (Fig. 3B).

3.3. Pathway and Gene Ontology enrichment analysis using the DEGs shared by all cells types after ZIKV infection

Pathway enrichment analysis using the upregulated DEGs indicated that they were involved in protein processing in the endoplasmic reticulum (ER), and downregulated DEGs were involved in DNA replication and cell cycle (Fig. 4A). Using the upregulated DEGs shared by all cells, it was possible to identify several biological processes related to the response to ER stress (GO:1905898, GO:1902235, GO:0034976, GO:1902043, GO:0036498, GO:1900102 and GO:0036499; Fig. 4B). There were also biological processes related to type I interferon signaling pathway (GO:0060337 and GO:0071357) and positive regulation of transcription from RNA polymerase II promoter (GO:0045944). Molecular functions related to upregulated DEGs were core promoter binding (GO:0000987, GO:0000978, GO:0001046, GO:0001047 and GO:1990837) and regulation of transcription (GO:0044212, GO:0000976 and GO:0000982). For downregulated DEGs, DNA replication (GO:0006260) and mitotic cell cycle phase transition were the identified biological process (GO:0000082, GO:0044843, GO:0044772). For downregulated DEGs, the identified molecular functions were related to DNA helicase activity (GO:0003678 and GO:0004003). The upregulated DEG list did not produce any significant cellular component ($p \geq 0.05$). However, the downregulated DEG list indicated chromosome telomeric region (GO:0000781), nuclear chromosome-telomeric region (GO:0000784) and nuclear chromosome part (GO:0044454) as cellular components related to these genes (Fig. 4B).



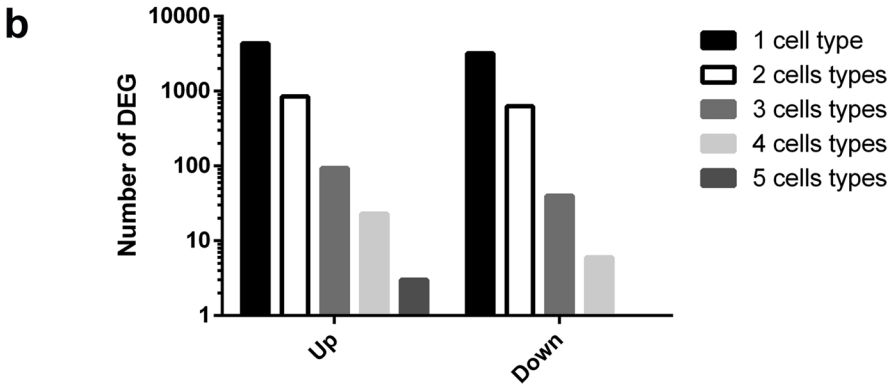


Fig. 2. The network of up- and downregulated genes after ZIKV infection of different cells types. (A) The bipartite network graph shows a spatially connected network among differentially expressed genes (DEGs) and cells types after ZIKV infection. Each node represents a gene or cell type. The larger nodes represent cell types and the connected smaller nodes represent DEG. The layout was generated using a force-based algorithm followed by manual rearrangement to better visualize the connections. A total of 11,049 DEGs (6436 upregulated and 4613 downregulated) and five cell types are represented. (B) Number of DEGs expressed in ZIKV-infected cells.

3.4. Search for published studies on Pubmed

To verify if the identified candidate genes and pathways had a correlation with ZIKV pathogenesis, a bibliographic survey was conducted in the PubMed (https://www.ncbi.nlm.nih.gov/pubmed). The search strategy for original articles was composed of two components: a) the gene, pathway or molecular function name and b) *Zika virus*. To identify relevant indexed studies the search filter was structured by combining the descriptors selected by boolean operators AND and NOR, as well as the Title / Abstract [TIAB] algorithm. Reviews are excluded from this analysis. Among the upregulated DEGs, it was identified 26 published studies that investigated the correlation of 14 upregulated DEGs with ZIKV pathogenesis (Supplementary Table 3). ATF3 and STAT1 genes exhibit the highest number of hits. For downregulated DEGs, it was identified 8 published studies, being the CENPF the most frequent gene (Fig. 5A). Regarding the pathways and biological processes, all of them had at least one published study showing the modulation of these pathways or biological processes after ZIKV infection (Fig. 5B). In a special way, 7 of 15 identified published studies showed a correlation of the inhibition of the cell cycle pathway with ZIKV pathogenesis. The ER stress has the highest number of hits among the significant biological process related to DEGs.

4. Discussion

Outbreaks of emerging and/or re-emerging viral infections are common threats to human health. When a new virus was discovery and associated with human disease, it is necessary to understand virus pathogenesis as soon as possible to develop countermeasures to the disease such as vaccines and antiviral drugs (Afrough et al., 2019). The *omics* technologies, especially the transcriptome studies could generate thousands of information about the gene regulation on cells after virus infection and therefore, it can help researches to gain insights about the pathogen-host interaction (Berkhout and Coombs, 2013; Jean Beltran et al., 2017). In this way, we propose a novel and friendly *in silico* integrative protocol for identifying key genes and pathways useful to understand emerging virus disease pathogenesis. The main advantage of this protocol is the use of galaxy platform to identify DEGs from different transcriptome studies. Galaxy is a web open-source, web-based platform for intensive biomedical research. Galaxy is used by several researches, because this platform has thousands of different tools for many different scientific fields. It can allow the analysis of large sequencing datasets by researchers without programming skills (Afgan et al., 2018). In this way, our proposed *in silico* protocol could permit the researcher DEGs from different datasets and also integrate the data. Data integration is vital to connect all the different data sources to exploit the value of insights because it reduces data complexity and also increases the value of data through unified systems. Once all the DEGs are available in a single place in real-time, researches would be able to use this protocol to integrate different transcriptome data and also perform the identification of candidate genes and pathways involved in virus pathogenesis.

To validate our protocol, the emerging arbovirus ZIKV was chosen as a target micro-organism. Since ZIKV is considered to be a risk factor for microcephaly development in humans, several studies attempted to understand the effect of ZIKV infection on different cell types. It is

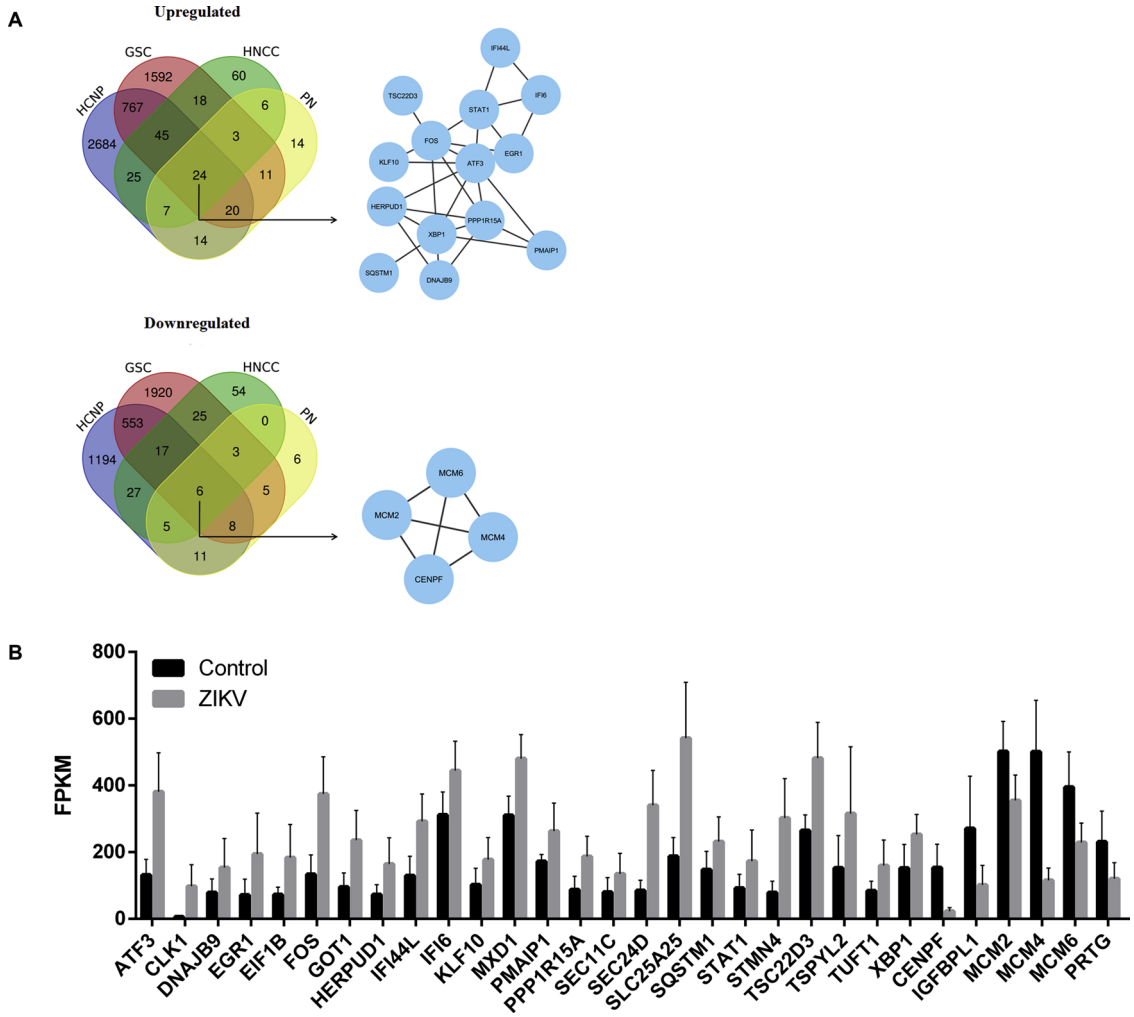


Fig. 3. Integrative transcriptome analysis of ZIKV-infected neural cells. (A) The Venn diagram shows the degree of overlap of the upregulated and downregulated genes in ZIKV-infected cells. The Venn diagrams were constructed using the DEGs identified in each cell type using the Calculate and draw custom Venn diagrams tool available at http://bioinformatics.psb.ugent.be/webtools/Venn/. The protein-protein interaction (PPI) networks for the common upregulated and downregulates genes were constructed using the STRING online database. DEG-encoded proteins without any interaction were excluded from the network. The network file was visualized with Cytoscape software. Abbreviations – HCNP: human cortical neural progenitors; GSC: glioblastoma stem cells, HNCC: human neural crest cells; PN: human peripheral neurons. (B) Expression levels (FPKM) of DEGs on ZIKV infected cells.

Table 3
Number of Differentially expressed genes indentified by integrative analysis.

Group	Up	Dow	Total
All cells	24	6	30
Undifferentiated cells (HCNP and GSC)	856	584	1440
Differentiated cells (HNCC and PN)	40	14	54

crucial to characterize the gene expression profile in ZIKV-infected cells, including neural-origin cells. Therefore, some studies published the transcriptome of ZIKV-infected cells to identify genes that could be involved in the development of neurological disorders induced by ZIKV. A search of the literature identified some transcriptome studies from ZIKV-infected neural cells, particularly neural progenitor cells (Table 1). This finding may be associated with the damage that ZIKV infection induces in this cell type during cell differentiation and in the development of the cerebral compartment.

Many studies presented evidence of disrupted brain development after infection of neural stem cells and neuroprogenitors by ZIKV infectious particles (Russso and Beltrão-Braga, 2017). However, other studies also showed the effect of ZIKV infection on GSC, HNCC, PN and CO transcriptomes (Table 1). In these studies, ZIKV can infect the target cells, but the outcomes are different. Data from many studies showed more severe cell death and higher viral load in undifferentiated cells. ZIKV infection and apoptosis rates are significantly higher in GSC compared to proliferating tumor or differentiated cells (Zhu et al., 2017). In the same context, ZIKV infection and cell death occur preferentially in neural progenitors present in CO (Watanabe et al., 2017). Zhang et al. (2016) reported higher viral load and cell death in ZIKV-infected HCNP. Regarding the effect of ZIKV infection on differentiated cells, the virus replicates approximately 10-fold more in HNCC compared to PN. There is also a significantly higher rate of apoptosis in HNCC compared to PN (Oh et al., 2018).

Applying the Tuxedo protocol on the identified datasets allowed the identification of DEGs in these cells after ZIKV infection (Table 2). However, the integrative analysis of the transcriptomes showed a differential effect on gene expression depending on the infected cell type. HCNP and GSC showed more DEGs when compared to other cells (Fig. 2A). Additionally, as HCNP and GSC shared many DEGs, data suggest a similar effect of ZIKV infection on the transcriptome of these cells. The data presented by Kaid et al. (2018) showed novel *in vitro* and *in vivo* evidence about the use of a Brazilian ZIKV strain as an oncolytic therapy to treat aggressive human embryonal tumors of the central nervous system. Therefore, the most prominent oncolytic effects of

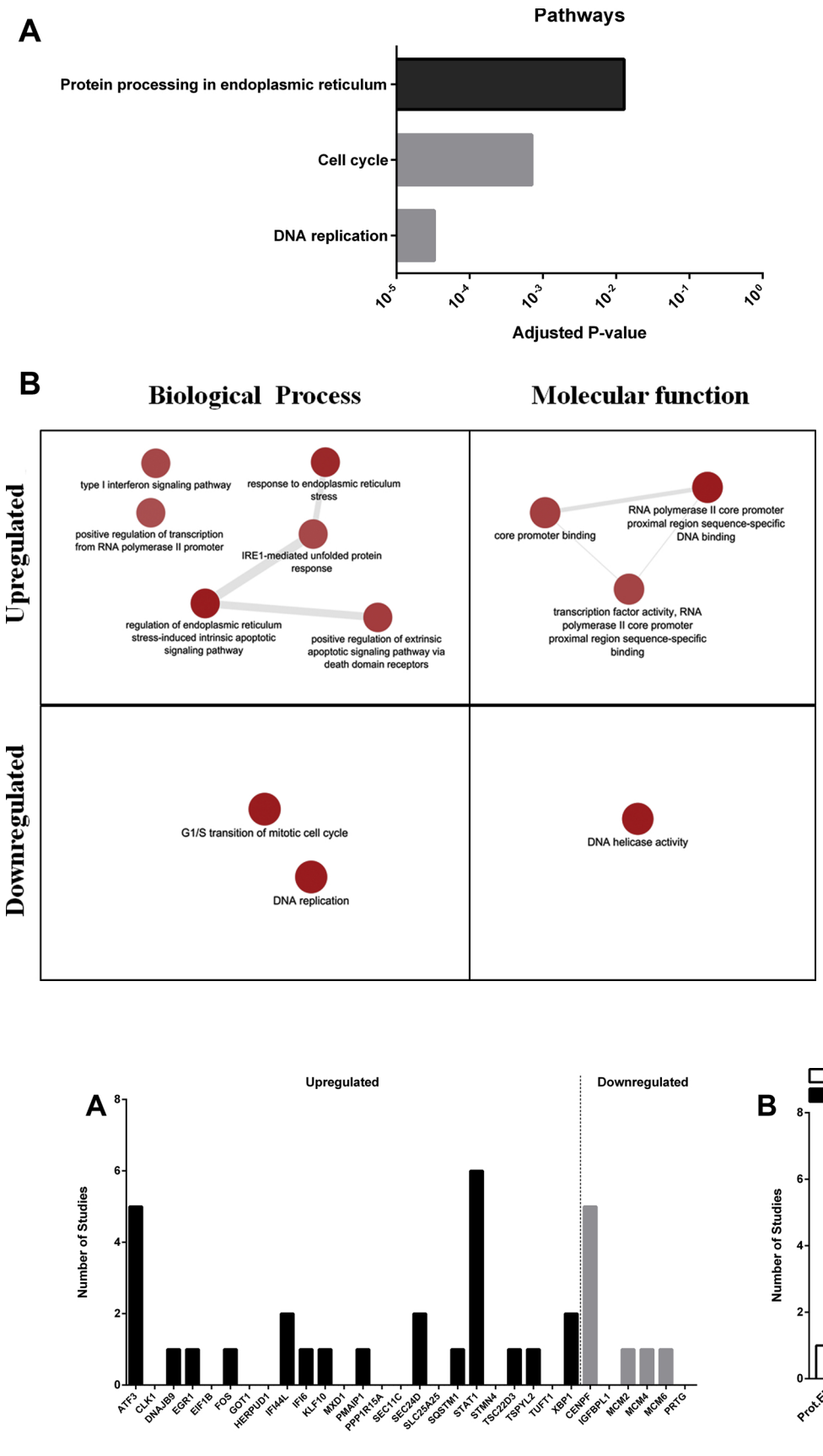


Fig. 4. Enriched Kyoto Encyclopedia of Genes and Genome (KEGG) pathways and Gene Ontology (GO) terms identified in integrative analysis of ZIKV-infected cells. A) DEG-encoded proteins identified from the protein-protein interaction (PPI) network in each cell type after ZIKV infection were used to identify significant pathways (adjusted $p < 0.05$) using the Enrichr platform (http://amp.pharm.mssm.edu/Enrichr/). B) Visualization of the biological process and molecular function GO annotations in ZIKV-infected cells using REVIGO. Bubble color indicates the p-value; bubble size indicates the frequency of the GO term. Highly similar GO terms are linked by edges in the graph, where the line width indicates the degree of similarity.

Fig. 5. Search hits for terms related to DEGs, pathways, gene ontology terms and ZIKV identified in PubMed. (A) Search for upregulated and downregulated DEGs. (B) Search for pathways and gene ontology terms. Legend: *Prot.ER*: protein processing in the endoplasmic reticulum, *DNA rep.*: DNA replication; *ER stress*: endoplasmic reticulum stress; *IFN-I sig.*: type I interferon signaling pathway; *Mit. Cell.*: mitotic cell cycle phase.

ZIKV infection was observed in tumors generated by cell lines with highly similar characteristics to early neural stem and neuroprogenitor cells. The few identified DEGs shared by all cell types could be attributed to the differentiation status of these cells. HCNP and GSC are undifferentiated, while HNCC and PN are differentiated cells. Therefore, the number of shared DEGs between cells under the same differentiation status should be greater than the DEG shared by undifferentiated and differentiated cells (Table 3). There were few DEGs shared by the four cell types. Consequently, few pathways, molecular functions and biological processes were enriched using this set of genes (24 upregulated and 6 downregulated genes).

The searches for published studies that investigate the expression of DEGs on ZIKV infected cells also confirm the potential of this *in silico* protocol to gain insights about virus pathogenesis. Eighteen of thirty identified DEGs (60 %) had at least one published study (Fig. 5A). Some of these DEGs have a direct correlation with antiviral response, neurogenesis and apoptosis. STAT1 and ATF3 were the upregulated DEGs with the highest number of hits. The increase of STAT1 expression on infected cells is a common antiviral response on virus-infected cells. Like other flaviviruses, ZIKV can antagonize the type I interferon pathway through interference with STAT-1 signaling. This immune evasion could be responsible to increase viral replication and apoptosis on infected cells of neural or non-neural origin (Chen et al., 2018). The ATF3 is a member of the activation transcription factor/cAMP-responsive element-binding protein family of transcription factors in the mammalians. Changes in ATF3 expression was observed on ZIKV infected cells (Moni and Lio, 2017; Zanini et al., 2018). This transcription marker is activated on injured peripheral neurons and therefore it is linked to neural cell survival (Lindå et al., 2011; Mahar and Cavall, 2018). CENPF is the downregulated DEG with the highest number of hits. The CENPF encoded protein associates with the centromere-kinetochore complex and its nuclear localization suggests that it play a role in chromosome segregation during mitosis (Varis et al., 2006). Several studies had demonstrated the decrease of CENPF on ZIKV neural infected cells, showing the correlation of CENPF downregulation with microcephaly (Dang et al., 2019; Moni and Lio, 2017; Paul et al., 2018; Zhang et al., 2016b).

Protein processing in the ER was the only significant pathway identified using the upregulated DEGs. This data is in agreement with several molecular functions related to the response to ER stress. Our literature search was able to identify several published studies that investigate the role of ER stress on ZIKV pathogenesis (Fig. 5B). The ZIKV can induce the expression of ER stress sensors in mice and human neural cells (Singh et al., 2018; Tan et al., 2018). Like other flaviviruses, ZIKV depends on the ER for its translation, replication and packaging. The biological events related to ZIKV replication in ER membranes induce modifications that could trigger ER stress. Therefore, the unfolded protein response should be activated to reduce stress on ZIKV-infected neural cells. This event could be crucial to the microcephaly development in newborns from ZIKV-infected mothers (Tan et al., 2018). Enrichment analysis using the downregulated DEGs identified cell cycle and DNA replication as important affected pathways with deregulated biological processes in all ZIKV-infected cells. Hammack et al. (2019), showed that ZIKV can impair the cell cycle by inducing DNA double-strand breaks in the host genome. This mechanism could be especially important for HNCP, because the growth arrest on these cells could impair brain development. Our data strengthen the idea that cell cycle arrest and DNA replication could be an important mechanism in all neural ZIKV-infected cells, independent of the differentiation status but with a more detrimental effect on the undifferentiated cells. Among the six downregulated DEGs identified by our analysis, four are related to the regulation of early events of DNA replication (MCM2, MCM4 and MCM6) on neocortex development. The proteins encoded by these genes are part of the minichromosomal maintenance complex (MMC). This complex consists of six subunits (MCM2 through MCM7) with ATP-dependent helicase activity. The main function of this complex is to attach to double-stranded DNA and induce the binding of multiple replication factors to the replisome. Defective DNA replication is involved in the pathophysiology of some cortical malformations like microcephaly. Therefore, decreased MCM2, MCM4 and MCM6 expression could disrupt the MMC in ZIKV-infected cell and thus affect the cell cycle. Some works demonstrated a reduction of proliferation and apoptosis in ZIKV-infected neural progenitor cells (Li et al., 2016; Tang et al., 2017). These data in connection with the downregulation of CENPF strengths the importance of maintaining the cell cycle to inhibit the deleterious effects caused by the multiplication of ZIKV in neural progenitor cells.

Using ZIKV as a model to validate our protocol we were able to perform an integrative transcriptome analysis of different ZIKV-infected cells. This analysis allowed us the identification of gene expression profiles in each cell type. The results showed that each cell type exhibited a particular response to ZIKV infection. Indeed, most identified DEGs were present in only one cell type. The common DEGs are related to two main critical pathways (DNA replication and ER stress) that could be involved in defects in brain development in ZIKV-infected newborns. Therefore, these DEGs can be subsequently used as potential targets for the development of new drugs that allow effective treatment of the disease. In conclusion, the proposed *in silico* protocol performed an integrated analysis that is able to predict and identify putative biomarkers from different transcriptome data. These biomarkers could be useful to understand virus disease pathogenesis and also help the identification of candidate antiviral drugs.

Funding

This work was supported in part by the Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) (grants APQ-01165–16 and PPM-00399–18). G.A.P.S. and E.A.S. received fellowships from FAPEMIG. F.R.O. received fellowships from Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq - PIBIC). L.F.L.C. and J.S.A. received PQ fellowship from CNPq. This study was financed in part by the Coordenação deAperfeiçoamento de Pessoal de Nível Superior, Brasil (CAPES). Finance Code 001.

CRediT authorship contribution statement

G.A.P.S. and L.F.L.C. conceived and planned the experiments. G.A.P.S., E.A.S., F.R.O. carried out the experiments. G.A.P.S., E.A.S., F.R.O., L.C.C.M., J.S.A., L.F.L.C. contributed to the interpretation of the results. G.A.P.S. and L.F.L.C took the lead in writing the manuscript. All authors provided critical feedback and helped shape the research, analysis and manuscript

Declaration of Competing Interest

The authors declare that they have no conflict of interest.

Appendix A. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.virusres.2020.197986.

References

Afgan, E., Baker, D., Batut, B., Van Den Beek, M., Bouvier, D., Ech, M., Chilton, J., Clements, D., Coraor, N., Grüning, B.A., Guerler, A., Hillman-Jackson, J., Hiltemann, S., Jalili, V., Rasche, H., Soranzo, N., Goecks, J., Taylor, J., Nekrutenko, A., Blankenberg, D., 2018. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Res. 46, W537–W544. https://doi.org/10.1093/nar/gky379.

Afrough, B., Dowall, S., Hewson, R., 2019. Emerging viruses and current strategies for vaccine intervention. Clin. Exp. Immunol. 196, 157–166. https://doi.org/10.1111/cei.13295.

Amrit, F.R.G., Ghazi, A., 2017. Transcriptomic analysis of C. Elegans RNA sequencing data through the Tuxedo suite on the galaxy project. J. Vis. Exp. 122, 1–12. https://doi.org/10.3791/55473.

Bastian, M., Heymann, S., 2009. Gephi : an open source software for exploring and manipulating networks. Fond. Maison des Sci. l'Homme.

Berkhout, B., Coombs, K.M., 2013. Quantitative omics and its application to study virus-host interactions-a new frontier. Front. Microbiol. 4 (2012–2013). https://doi.org/10.3389/fmicb.2013.00031.

Chen, Jian, Yang, Y.F., Yang, Y., Zou, P., Chen, Jun, He, Y., Shui, S.L., Cui, Y.R., Bai, R., Liang, Y.J., Hu, Y., Jiang, B., Lu, L., Zhang, X., Liu, J., Xu, J., 2018. AXL promotes Zika virus infection in astrocytes by antagonizing type i interferon signalling. Nat. Microbiol. 3, 302–309. https://doi.org/10.1038/s41564-017-0092-4.

Cinegaglia, N.C., Andrade, S.C.S., Tokar, T., Pinheiro, M., Severino, F.E., Oliveira, R.A., Hasimoto, E.N., Cataneo, D.C., Cataneo, A.J.M., Defaveri, J., Souza, C.P., Márques, M.M.C., Carvalho, R.F., Coutinho, L.L., Gross, J.L., Rogatto, S.R., Lam, W.L., Jurisica, I., Reis, P.P., 2016. Integrative transcriptome analysis identifies deregulated microRNA-transcription factor networks in lung adenocarcinoma. Oncotarget 7, 28920–28934. https://doi.org/10.18632/oncotarget.8713.

Dang, J.W., Tiwari, S.K., Qin, Y., Rana, T.M., 2019. Genome-wide integrative analysis of zika-virus-Infected neuronal stem cells reveals roles for MicroRNAs in cell cycle and stemness. Cell Rep. 27, 3618–3628.e5. https://doi.org/10.1016/j.celrep.2019.05.059.

Ellwanger, J.H., Chies, J.A.B., 2016. Emergent diseases in emergent countries: we must study viral ecology to prevent new epidemics. Braz. J. Infect. Dis. 20, 403–404. https://doi.org/10.1016/j.bjid.2016.02.003.

Hammack, Christy, Ogden, S.C., Madden Jr., J.C., Medina, A., Xu, C., Philips, E., Son, Y., Cone, A., Giovinazzi, S., Didier, R.A., Gilbert, D., 2019. Zika virus infection induces DNA damage response in human neural progenitors that enhaces viral replication. J. Virol. https://doi.org/10.1128/JVI.00638-19.

Harris, M.A., et al., 2004. The Gene Ontology (GO) database and informatics resource. Nucleic Acids Res. 32, 258–261. https://doi.org/10.1093/nar/gkh036.

Hoshida, Y., Nijman, S.M.B., Kobayashi, M., Chan, J.A., Brunet, J., Chiang, D.Y., Villanueva, A., Newell, P., Hashimoto, M., Watanabe, G., Gabriel, S., Friedman, S.L., 2009. Integrative transcriptome analysis reveals common molecular subclasses of human hepatocellular carcinoma. Cancer Res. 69, 7385–7392. https://doi.org/10.1158/0008-5472.CAN-09-1089.Integrative.

Jean Beltran, P.M., Federspiel, J.D., Sheng, X., Cristea, I.M., 2017. Proteomics and integrative omic approaches for understanding host–pathogen interactions and infectious diseases. Mol. Syst. Biol. 13, 922. https://doi.org/10.15252/msb.20167062.

Jin, C., Kang, H., Ryu, J.R., Kim, S., Zhang, Y., Lee, Y., Kim, Y., Han, K., 2018. Integrative brain transcriptome analysis reveals region-specific and broad molecular changes in Shank3-Overexpressing mice. Front. Mol. Neurosci. 11, 1–14. https://doi.org/10.3389/fnmol.2018.00250.

Kaid, C., Goulart, E., Caires-j, L.C., Ventini, D.C., Astray, R.M., Assoni, A.F., 2018. Zika virus selectively kills aggressive human embryonal CNS tumor cells in vitro and in vivo. Cancer Res. 26, 3363–3375. https://doi.org/10.1158/0008-5472.CAN-17-3201.

Kuleshov, M.V., Jones, M.R., Rouillard, A.D., Fernandez, N.F., Duan, Q., Wang, Z., Koplev, S., Jenkins, S.L., Jagodnik, K.M., Lachmann, A., Mcdermott, M.G., Monteiro, C.D., Gundersen, W., Ma, A., 2016. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 44, 90–97. https://doi.org/10.1093/nar/gkw377.

Larocca, R.A., Abbink, P., Peron, J.P.S., Zanotto, P.M.D.A., Iampietro, M.J., Badamchi-zadeh, A., Boyd, M., Ng, D., Kirilova, M., Nityanandam, R., Mercado, N.B., Li, Z., Moseley, E.T., Bricault, C.A., Borducchi, E.N., Giglio, P.B., Jetton, D., Neubauer, G., Nkolola, J.P., Maxfield, L.F., Barrera, R.A.D.La, Jarman, R.G., Eckels, K.H., Michael, N.L., Thomas, S.J., Barouch, D.H., Badamchi-zadeh, A., Boyd, M., Ng, D., Kirilova, M., Nityanandam, R., Mercado, N.B., Maxfield, L.F., Barrera, R.A.D.La, Jarman, R.G., Eckels, K.H., Neubauer, G., Joseph, P., Michael, N.L., Thomas, S.J., Barouch, D.H., 2016. Vaccine protection agains Zika virus from Brazil. Nature 000, 1–19. https://doi.org/10.1038/nature18952.

Laura, A., Victoria, T., María, L., Morales, A., Groisman, B., Baricalla, A., Fabbri, C., Sinchi, A., Alonso, A., Gonzalez, C., Ledesma, B., Masi, P., Silva, M., Israilev, A., Rocha, M., Quaglia, M., Paz, M., Rosa, B., Pablo, L., 2018. Surveillance of microcephaly and selected brain anomalies in Argentina : relationship with Zika virus and other congenital infections. Birth Defects Res. 110, 1016–1026. https://doi.org/10.1002/bdr2.1347.

Li, H., Saucedo-cuevas, L., Regla-nava, J.A., Chai, G., Tang, W., Terskikh, A.V., Shresta, S., Gleeson, J.G., 2016. Zika Virus Infects Neural Progenitors in the Adult Mouse Brain and Alters Proliferation 19. pp. 593–598.

Liao, Q., Wang, J., Pei, Z., Xu, J., Zhang, X., 2017. Identification of miRNA-mRNA crosstalk in CD4 + T cells during HIV-1 infection by integrating transcriptome analyses. J. Transl. Med. 15, 1–11. https://doi.org/10.1186/s12967-017-1130-y.

Lindå, H., Sköld, M.K., Ochsmann, T., 2011. Activating transcription factor 3, a useful marker for regenerative response after nerve root injury. Front. Neurol. (MAY), 1–6. https://doi.org/10.3389/fneur.2011.00030.

Mahar, Marcus, Cavall, Valeria, 2018. Intrinsic mechanisms of neuronal axon regeneration. Nat. Rev. Neurosci. 19, 323–337. https://doi.org/10.1038/s41583-018-0001-8.

Moni, M.A., Lio, P., 2017. Genetic profiling and comorbidities of zika infection. J. Infect. Dis. 216, 703–712. https://doi.org/10.1093/infdis/jix327.

Moslehi, R., Mills, J.L., Signore, C., Kumar, A., Ambroggio, X., Dzutsev, A., 2013. Integrative transcriptome analysis reveals dysregulation of canonical cancer molecular pathways in placenta leading to preeclampsia. Sci. Rep. 3, 1–11. https://doi.org/10.1038/srep02407.

Musso, D., Cao-Lormeau, V.M., Gubler, D.J., 2015. Zika virus: following the path of dengue and chikungunya? Lancet 386, 243–244. https://doi.org/10.1016/S0140-6736(15)61273-9.

Oh, Y., Zhang, F., Wang, Y., Lee, E.M., Choi, I.Y., Lim, H., Mirakhori, F., Li, R., Huang, L., Xu, T., Wu, H., Li, C., Qin, F., Wen, Z., Wu, Q., Tang, H., Xu, Z., Song, H., Ming, G., Lee, G., Malvin, H., Orleans, N., Sciences, B., Biology, C., 2018. Zika virus directly infects peripheral neurons and induces cell death. Nat. Neurosci. 20, 1209–1212. https://doi.org/10.1038/nn.4612.Zika.

Olival, K.J., Hosseini, P.R., Zambrana-Torrelio, C., Ross, N., Bogich, T.L., Daszak, P., 2017. Host and viral traits predict zoonotic spillover from mammals. Nature 546, 646–650. https://doi.org/10.1038/nature22975.

Paul, A.M., Acharya, D., Neupane, B., Ashely Thompson, E., Gonzalez-Fernandez, G., Copeland, K.M., Garrett, M., Liu, H., Lopez, M.E., Cruz, Mde, Flynt, A., Liao, J., Guo, Y.L., Gonzalez-Fernandez, F., Vig, P.J.S., Bai, F., 2018. Congenital Zika virus infection in immunocompetent mice causes postnatal growth impediment and neurobehavioral deficits. Front. Microbiol. 9, 1–12. https://doi.org/10.3389/fmicb.2018.02028.

Petersen, E., Wilson, M.E., Touch, S., McCloskey, B., Mwaba, P., Bates, M., Dar, O., Mattes, F., Kidd, M., Ippolito, G., Azhar, E.I., Zumla, A., 2016. Rapid spread of zika virus in the Americas - implications for public health preparedness for mass gatherings at the 2016 Brazil olympic games. Int. J. Infect. Dis. 44, 11–15. https://doi.org/10.1016/j.ijid.2016.02.001.

Russso, B.R., Beltrão-Braga, P., 2017. Biochemical and Biophysical Research Communications the impact of Zika virus in the brain. Biochem. Biophys. Res. Commun. 492, 603–607. https://doi.org/10.1016/j.bbrc.2017.01.074.

Shi, Y., Gao, G.F., 2017. Structural biology of the zika virus. Trends Biochem. Sci. 1–14. https://doi.org/10.1016/j.tibs.2017.02.009. xx.

Singh, P.K., Khatri, I., Jha, A., Pretto, C.D., Spindler, K.R., Arumugaswami, V., Giri, S., Kumar, A., Bhasin, M.K., 2018. Determination of system level alterations in host transcriptome due to Zika virus (ZIKV) Infection in retinal pigment epithelium. Sci. Rep. 8, 1–16. https://doi.org/10.1038/s41598-018-29329-2.

Supek, F., Bosnjak, M., Skunca, N.T.S., 2011. REVIGO summarizes and visualizes long lists of gene ontology terms. PLoS Negl. Trop. Dis. 6, 1–9. https://doi.org/10.1371/journal.pone.0021800.

Tan, Z., Zhang, W., Sun, J., Fu, Z., Ke, X., Zheng, C., Zhang, Y., Li, P., Liu, Y., Hu, Q., Wang, H., Zheng, Z., 2018. ZIKV infection activates the IRE1-XBP1 and ATF6 pathways of unfolded protein response in neural cells. J. Neuroinflammation 1–16.

Tang, H., Hammack, C., Ogden, S.C., Wen, Z., Qian, X., 2017. Zika virus infects human cortical neural precursors and attenuates their growth. Cell Stem Cell 18, 587–590. https://doi.org/10.1016/j.stem.2016.02.016.Zika.

Tiwari, S.K., Dang, J., Qin, Y., Lichinchi, G., Bansal, V., Rana, T.M., 2017. Zika virus infection reprograms global transcription of host cells to allow sustained infection. Emerg. Microbes Infect. 6, e24–10. https://doi.org/10.1038/emi.2017.9.

Varis, A., Salmela, A.L., Kallio, M.J., 2006. Cenp-F (mitosin) is more than a mitotic marker. Chromosoma 115, 288–295. https://doi.org/10.1007/s00412-005-0046-0.

Watanabe, M., Buth, J.E., Vishlaghi, N., Taxidis, J., Khakh, B., Coppola, G., Pearson, C.A., Gong, D., Dai, X., Damoiseaux, R., Aliyari, R., Liebscher, S., Schenke-layland, K., Caneda, C., Huang, E.J., Zhang, Y., Cheng, G., Geschwind, D.H., Golshani, P., Novitch, B.G., Disabilities, D., Sciences, B., Behavior, H., Pharmacology, M., Genetics, M., 2017. Self-organized cerebral organoids with human specific features predict effective drugs to combat Zika virus infection. Cell Rep. 21, 517–532. https://doi.org/10.1016/j.celrep.2017.09.047.Self-organized.

Yun, S.-I., Lee, Y.-M., 2017. Zika virus: an emerging flavivirus. J. Microbiol. 55, 204–219. https://doi.org/10.1007/s12275-017-7063-6.

Zanini, F., Pu, S.Y., Bekerman, E., Einav, S., Quake, S.R., 2018. Single-cell transcriptional dynamics of flavivirus infection. Elife 7, 1–21. https://doi.org/10.7554/eLife.32942.

Zhang, F., Hammack, C., Ogden, S.C., Cheng, Y., Lee, E.M., Wen, Z., Qian, X., Nguyen, H.N., Li, Y., Yao, B., Xu, Miao, Xu, T., Chen, L., Wang, Z., Feng, H., Huang, W., Yoon, K., Shan, C., Huang, L., Qin, Z., Christian, K.M., Shi, P., Xu, Mingjiang, Xia, M., Zheng, W., Wu, H., Song, H., Tang, H., 2016a. Molecular signatures associated with ZIKV exposure in human cortical neural progenitors. Nucleic Acids Res. 44, 8610–8620. https://doi.org/10.1093/nar/gkw765.

Zhang, N., Zhang, N., Qin, C.F., Liu, X., Shi, L., Xu, Z., 2016b. Zika virus disrupts neural progenitor development and leads to microcephaly in mice. Cell Stem Cell 19, 120–126. https://doi.org/10.1016/j.stem.2016.04.017.

Zhu, Z., Gorman, M.J., Mckenzie, L.D., Chai, J.N., Hubert, C.G., Prager, C., Fernandez, E., Richner, J.M., Zhang, R., Shan, C., Tycksen, E., Wang, X., Shi, P.Y., Diamond, M.S., Rich, J.N., Chheda, M.G., 2017. Zika virus has oncolytic activity against glioblastoma stem cells. J. Exp. Med. 214, 2843–2857.