Dysregulation of gene expressions upon Zika virus infection

Yonghua Zhuang
BIOS-7659/CPBS-7659

December 15, 2016

1 Introduction

Zika virus (ZIKV) is an emerging mosquito-borne flavivirus and is transmitted by daytime-active aedes mosquitoes [1]. ZIKV was first identified in Zika forest, Uganda in 1947 and related to the dengue, yellow fever, Japanese encephalitis, and West Nile viruses. ZIKV infection often causes no or only mild symptoms, similar to a mild form of dengue fever. By far, ZIKV infection are not able to be prevented by medications or vaccines [2]. In 2015, there was a significant increase in reports of ZIKV infection in the Americas. Brazil is the most affected country, with preliminary estimates of 440,000 to 1.3 million cases of autochthonous ZIKV infection reported through December 2015 [1]. Several recent reports from the Ministry of Health of Brazil indicates a possible association between ZIKV infection in pregnancy and fetal malformations [1]. As of February 4, 2016, the ZIKV epidemic has continued to spread in many countries, and evidence of local ZIKV infection cases has been reported from 31 countries within the past 2 months, and 36 countries in the past 9 months [3]. Although ZIKV infection has been linked to microcephaly among newborn infants recently, how ZIKV infection triggers cell death (especially apoptosis) in the brain is still largely unknown.

In this project I conducted a global transcriptome analysis of RNA sequencing data to identify and compare cellular genes and signaling pathways that are disregulated in the mouse brain following infection with ZIKV. These diregulated genes and signaling pathways might contribute to ZIKV pathogenesis within the CNS and serve as potential therapeutic targets for treatment of ZIKV associated brain abnormalities.

2 Methods

2.1 Viral Inoculation of Embryonic Brains

Pregnant C57BL/6J mice with E14.5 embryos were treated with ketamine hydrochloride and xylazine to induce anesthesia. 1 µl 1.7 x 10^6 TCID50/ml ZIKV virus (Mexican isolate MEX 1-44) was injected into the lateral ventricles of E14.5 embryo brains. Control media was used as a sham injection (not infected control). Injected embryos were placed back to pregnant dams and allowed to develop after surgery for 3 days. The study includes two not infected and two ZIKV infected mouse embryonic brains (E17.5). The experiments were conducted at University of Georgia by Qiang Shao et al. [4].

2.2 RNA Isolation and High Throughput Sequencing

Total RNAs were extracted from two not infected and two ZIKV infected E17.5 brains using Trizol reagent (Life Technologies). Genomic DNA and ribosomal RNA was removed with Turbo DNA-free kit and RiboMinus Eukaryote Kit (Life Technologies). The resulting RNA fractions were subjected to strand-specific library preparation using NEBNext Ultra Directional RNA Library Prep Kit for Illumina (New England Biolabs). Sequencing was performed on Nextseq500 (Illumina) [4].

2.3 RNA-seq Data Analysis

The sequencing read data were aligned to mouse GRCm38 (mm10) reference genome using Bowtie2 and the resulatant alignment data in BAM file format were released on PubMed website (SRR4445640 - SRR4445643) [4]. The BAM files were downloaded and used to generate count matrices. Of note, the details of read quality and alignment information were not avaiiable. The raw count matrix contains many rows with only zeros, which have no or nearly no information about the amount of gene expression. The count data were filtered using with the following rule: removing rows that have no counts, or only a single count across all samples.

There are several packages available for RNA seq analysis. I chose two of most popular packages, DESeq2 [5] and edgeR [6] to identify differentially expressed genes between mock control and ZIKV infected mouse brain. For principal component analysis, count data were applied with a "regularized log" (rlog) transformation method provided in the DESeq2 package. The rlog transformation behaves similarly to a log2 transformation for genes with high counts, but shrinks together the values for different samples for genes with low counts. Thus, it avoids a commonly observed issue of the standard logarithm transformation, the spreading apart of data for genes with low counts [5]. The package DESeq2 estimates dispersion and

logarithmic fold changes by incorporating data-driven prior distributions. Size factors were determined by using the "median ratio method" (default setting) in DESeq2 to control for differences in the sequencing depth of the samples. DESeq2 tests differential expression on raw count data by use of negative binomial generalized linear models. In edgeR analysis, size factors were estimated using the "trimmed mean of M values (TMM)" method (default setting) on log2 scaled count data. TMM method is simple but robust way to estimate the ratio of RNA production using a weighted trimmed mean of the log expression ratios [6]. The Fisher's Exact Test in edgeR package was used to identify differential expression. Since the edgeR was known to be anti-conservative for lowly expressed genes and might have higher type-I error than DESeq2 [7], I focused on reporting the results derived from DESeq2.

The gene ontology (GO) term enrichment analysis were performed using GOstats package [8] to detect the signaling pathways that are dysregulated in the mouse brain following infection with ZIKV. Function and pathway analyses based on the association of GO terms to genes in a selected gene list are useful bioinformatic tools and the GOstats package has been widely used to perform such computations [8]. The significance level was set at 0.05 in this study. Benjamini-Hochberg (BH) adjustment was used to correct for multiple testing.

3 Result

3.1 Exploratory analysis and visualization

Prior to principal components analysis (PCA) and studying the relationship of samples, data transformation was required on the count data since PCA works best for data that generally has the same range of variance at different ranges of the mean values. A simple and often used approach is to take the logarithm of the normalized count values plus a small pseudocount, where the genes with the very lowest counts often tend to dominate the results. Another approach is to take rlog transformation to stabilize the variance across the mean. As shown in Figure 1, the genes with low counts (bottom left-hand corner) seem to be excessively variable on the ordinary logarithmic scale (Figure 1A), while the rlog transformation suppress variation for the low count genes for which the data provide little information about differential expression (Figure 1B).

PCA analysis was performed using the rlog-transformed values to visualize sample-to-sample distances. From the PCA plot (Figure 2), the differences between not infected control and ZIKV infected samples are large, though there were variations within each group.

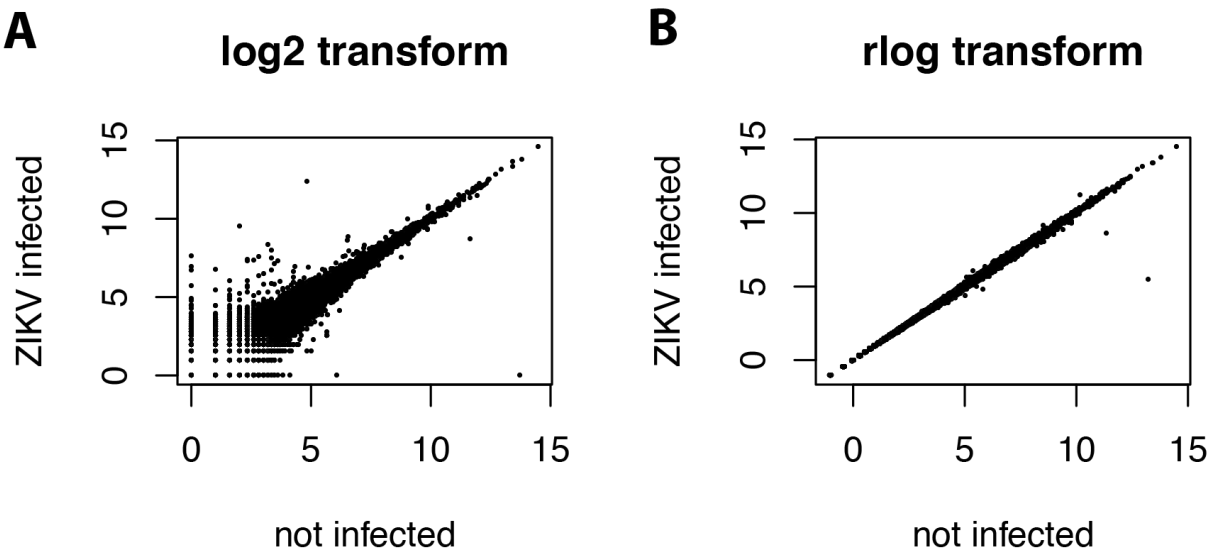

Figure 1: Scatterplot of transformed counts from two samples (not infected vs ZIKV infected). Shown are scatterplots using the log2 transform of normalized counts (left side) and using the rlog (right side).

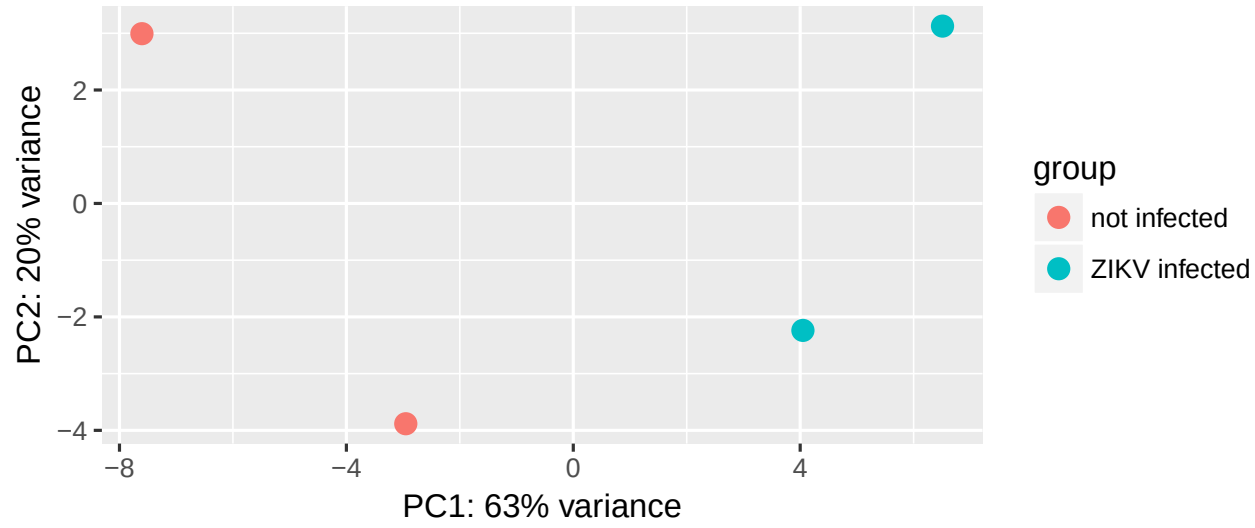

Figure 2: Principal components analysis (PCA) plot. The samples (not infected and ZIKV infected) are projected onto the 2D plane such that they spread out in the two directions that explain most of the differences. The x-axis is the direction that separates the data points the most. The values of the samples in this direction are written PC1. The y-axis is a direction that separates the data the second most. The values of the samples in this direction are written PC2. The percent of the total variance that is contained in the direction is printed in the axis label.

3.2 Differential gene expression analysis

3.2.1 Significant differentially expressed genes identified by DESeq2

Next I performed the differential expression analysis on the raw counts (untransformed) with DESeq2 method. The valcano plot (Figure 3) depicts the distributions of adjusted P values and fold changes on the tested genes. As shown in Figure 3, many genes are up-regulated upon ZIKV virus infection while little genes were down-regulated in the ZIKV infected brain. Specially, 61 significant differentially expressed genes (padj ≤ 0.05) between not infected control and ZIKV infected samples were discovered with DESeq2 method.

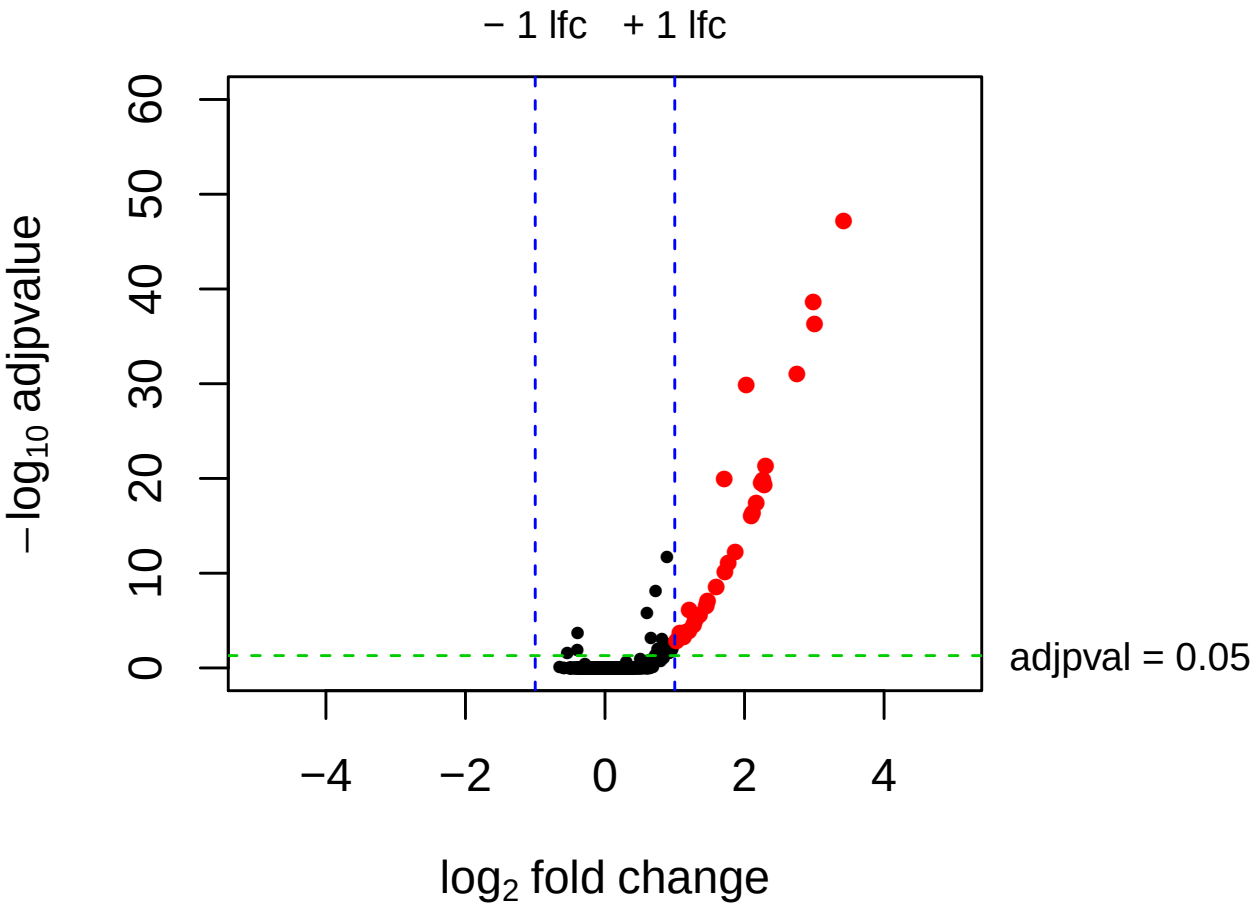

Figure 3: Volcano plots depicts the distributions of $log_2 fold\ change$ values and $-log_{10}(adjP)$. Each red dot represents the genes with adjP values ≤ 0.05 and fold change ≥ 2 or fold change < 0.5. Each black dot represents other genes either with adjP value > 0.05 or $0.5 <$ fold change < 2. The two blue dotted lines stand for $log_2 fold\ change = \pm 1$. The green dotted line is $adjP = 0.05$.

3.2.2 Method comparison: DESeq2 vs. edgeR

For method comparison, I also performed differential gene expression analysis with another popular method, edgeR. 102 differentially expressed genes were identified with edgeR method (FDR ≤ 0.05). As shown in venn diagram (Figure 4), 59 out of 61 differentially expressed genes identified with DESeq2 method are also significant in edgeR analysis.

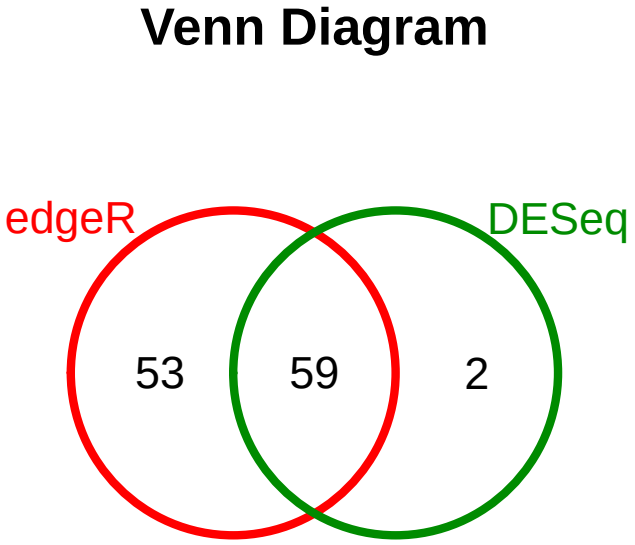

Figure 4: Heatmap of relative rlog-transformed values across samples. Infection statuses are shown with colored bars at the top of the heatmap. Blocks of genes that covary across samples. In the heatmap, a set of genes for which the ZIKV-infected samples have higher gene expression.

3.2.3 Top significant differentially expressed genes

Due to the consistency between two tested methods and the potential highter Type I error in edgeR analysis, I focused on analyzing the top significant differentially expressed genes identified in DESeq2 method.

Table 1 lists the 14 significant differentially expressed genes with $\geq$ 2 fold changes, which include Gls, Parp9, Tmprss2, Oas1h, Oasl2, Art3, Gbp5, 3110062M04Rik, Zfp619, H2-T22, Sult6b1, Slfn5os, Kdm7a, Ifi27 and Tbxas1 genes. Both 2'-5' oligoadenylate synthetase 1H (Oas1h) and 2'-5' oligoadenylate synthetase-like 2(Oasl2) genes are associated with oligoadenylates. ADP-ribosyltransferase 3 (Art3) gene encodes an arginine-specific ADP-ribosyltransferase and plays a role in immune response. The interferon alpha inducible protein 27 (IFI27) and guanylate-binding protein (Gbp5) involves in interferon pathway and apoptosis reaction. The glutaminase (GLS) gene encodes an enzyme that generates glutamate from glutamine and is related to glutaminase toxicity. The lysine demethylase 7A(KDM7A) is involved in histone modifications.

Table 1: Summary of significant differentially expressed genes

gene	log2FoldChange	lfcSE	stat	pvalue	padj
Gls	1.706	0.168	10.178	2.492E-24	3.702E-21
Parp9	2.248	0.221	10.152	3.253E-24	4.296E-21
Tmprss2	2.262	0.226	10.018	1.269E-23	1.371E-20
Oas1h	2.093	0.226	9.279	1.711E-20	1.564E-17
Oasl2	2.074	0.226	9.195	3.742E-20	3.177E-17
Art3	1.849	0.226	8.186	2.701E-16	2.140E-13
Gbp5	1.706	0.226	7.558	4.095E-14	2.704E-11
3110062M04Rik	1.462	0.223	6.551	5.724E-11	3.240E-08
Zfp619	1.440	0.226	6.377	1.802E-10	9.737E-08
H2-T22	1.203	0.195	6.174	6.639E-10	3.430E-07
Sult6b1	1.347	0.224	6.001	1.957E-09	9.303E-07
Slfn5os	1.260	0.226	5.582	2.376E-08	1.046E-05
Kdm7a	1.194	0.225	5.307	1.111E-07	4.718E-05
Ifi27	1.062	0.205	5.184	2.166E-07	8.581E-05
Tbxas1	1.121	0.225	4.988	6.114E-07	2.271E-04

Next I performed the gene cluster analysis on these 14 significant differentially expressed genes with the rlog transformed counts. Heat map (Figure 5) shows that all these 14 genes were up-regulated in ZIKV infected brain. In addition, it might form 4 major clusters of tested genes: Ifi27, Tmprss2, Oas1h, and Oasl2; H2-T22, 3110062M04Rik, Gls and Kdm7a; Sult6b1, Zfp619, Tbxas1 and Slfn5os; Art3, Parp9 and Gbp5.

3.3 Gene Ontology enrichment analysis

To further study a functional profile of differentially expressed gene set and better understand the underlying biological processes, Gene Ontology (GO) enrichment analysis was performed. The results revealed gene sets that are associated with the interferon response and immune response as well as histone modifications (Table 2).

4 Discussion

This report analyzed gene expression in the mouse brain following infection with ZIKV virus with two popular packages, DESeq2 and edgeR. We found 59 significant differentially expressed gene in both methods and most of them are up-regulated during ZIKV infection. The majority of these up-regulated genes are associated with immune response, interferon

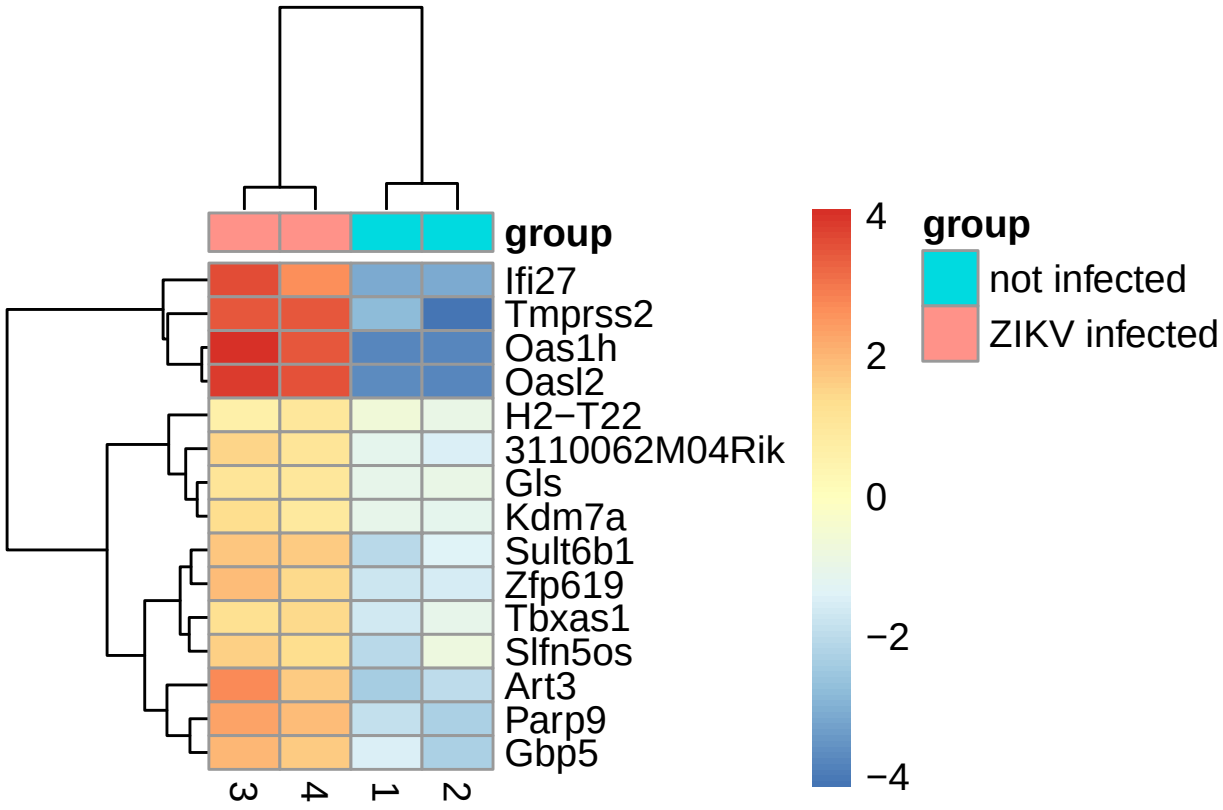

Figure 5: Heatmap of relative rlog-transformed values across samples. Infection statuses are shown with colored bars at the top of the heatmap. Blocks of genes covary across samples. In the heatmap, a set of genes for which the ZIKV-infected samples have higher gene expression.

response, glutamate toxicity and histone modification, which supports Qiang Shao et al. recent finding that ZIKV induces dysregulation of genes involved in immune response [4]. Of note, Qiang Shao et al. claimed to identify more than 200 significant differentially expressed genes. The huge discrepancy in the number of significant differentially expressed gene might be partially explained by the lack of multiple testing correction in Qiang Shao's study. Due to the absence of method details in Qiang Shao's analysis, we could not perform further comparisons.

The activation of IFN and immune response following viral infection has been seen following infection with a variety of viruses, and it is well demonstrated that IFN signaling and immune reaction play protective roles following virus infection [9, 10]. Glutamate is the key excitatory neurotransmitter in the neurological system. As shown in previous work [10, 11], an excess of glutamate may result in apoptosis of neurons through glutamate excitotoxicity and contribute to pathology in ZIKV infections. Taken together, aggressive cellular response including immune response, interferon response and so on could limit ZIKV growth

Table 2: Summary of Gene Ontology enrichment analysis

GOBPID	Pvalue	OddsRatio	Term
GO:0034341	0.00	65.98	response to interferon-gamma
GO:0035574	0.00	1068.46	histone H4-K20 demethylation
GO:0045087	0.00	13.07	innate immune response
GO:0006543	0.00	534.19	glutamine catabolic process
GO:0006537	0.01	267.06	glutamate biosynthetic process
GO:0071557	0.01	267.06	histone H3-K27 demethylation
GO:0030644	0.01	213.63	cellular chloride ion homeostasis
GO:0046598	0.01	213.63	positive regulation of viral entry into host cell
GO:0060330	0.01	213.63	regulation of response to interferon-gamma
GO:0055064	0.01	178.01	chloride ion homeostasis
GO:0016540	0.01	152.57	protein autoprocessing
GO:0033169	0.01	152.57	histone H3-K9 demethylation
GO:0070544	0.01	152.57	histone H3-K36 demethylation

but also cause brain damage by triggering neuronal death. The identified disregulated genes and signaling pathways might serve as potential therapeutic targets for treatment of ZIKV associated brain abnormalities.

There were several limitations in this study. One limitation was that the sample size was too small. There were only two uninfected and two infected samples. To get more convincing data, more samples should be required. Another limitation was that there were large variation within each group, as shown in PCA plot (Figure 2). The large variation within group could be due to many factors, such as large intrinsic gene variation between mouse or not well controlled experimental procedures. The differential quality of individual samples might also contribute to the variation within and between sample groups. Without the details of experiment procedures and RNA-sequencing and alignment, it is hard to completely reveal potential issues in this study. The third limitation was that the samples were only collected at one time point: 3 days after ZIKV infection. However, massive neuron death were observed 10 days after ZIKV infection and ZIKV kept growing from the initial infection to the end of experiment (10 days after infection). It is highly possible that many gene expression changes might not occur yet after 3-day infection. In addition, real-time PCR was not done to verify the key findings derived from RNA-seq analysis. Given these three limitations and the concern of lack of data pre-processing details (such as quality score), more evidences and further analysis are needed before drawing a final conclusion.

References

[1] Jernej Mlakar, Misa Korva, Nataša Tul, Mara Popović, Mateja Poljšak-Prijatelj, Jerica Mraz, Marko Kolenc, Katarina Resman Rus, Tina Vesnaver Vipotnik, Vesna Fabjan Vodušek, Alenka Vizjak, Jože Pižem, Miroslav Petrovec, and Tatjana Avšič Županc. Zika virus associated with microcephaly. *N Engl J Med*, 374(10):951–8, Mar 2016.

[2] Enny S Paixão, Florisneide Barreto, Maria da Glória Teixeira, Maria da Conceição N Costa, and Laura C Rodrigues. History, epidemiology, and clinical manifestations of zika: A systematic review. *Am J Public Health*, 106(4):606–12, Apr 2016.

[3] Hui-Lan Chen and Ren-Bin Tang. Why zika virus infection has become a public health concern? *J Chin Med Assoc*, 79(4):174–8, Apr 2016.

[4] Qiang Shao, Stephanie Herrlinger, Si-Lu Yang, Fan Lai, Julie M Moore, Melinda A Brindley, and Jian-Fu Chen. Zika virus infection disrupts neurovascular development and results in postnatal microcephaly with brain damage. *Development*, 143(22):4127–4136, Nov 2016.

[5] Michael I Love, Wolfgang Huber, and Simon Anders. Moderated estimation of fold change and dispersion for rna-seq data with deseq2. *Genome Biology*, 15:550, 2014.

[6] Mark D Robinson, Davis J McCarthy, and Gordon K Smyth. edger: a bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1):139–40, Jan 2010.

[7] Simon Anders and Wolfgang Huber. Differential expression analysis for sequence count data. *Genome Biol*, 11(10):R106, 2010.

[8] S Falcon and R Gentleman. Using gostats to test gene lists for go term association. *Bioinformatics*, 23(2):257–8, Jan 2007.

[9] Richard E Randall and Stephen Goodbourn. Interferons and viruses: an interplay between induction, signalling, antiviral responses and virus countermeasures. *J Gen Virol*, 89(Pt 1):1–47, Jan 2008.

[10] Penny Clarke, J Smith Leser, Richard A Bowen, and Kenneth L Tyler. Virus-induced transcriptional changes in the brain include the differential expression of genes associated with interferon, apoptosis, interleukin 17 receptor a, and glutamate signaling as well as flavivirus-specific upregulation of trna synthetases. *MBio*, 5(2):e00902–14, Mar 2014.

[11] Pennelope K Blakely, Bette K Kleinschmidt-DeMasters, Kenneth L Tyler, and David N Irani. Disrupted glutamate transporter expression in the spinal cord with acute flaccid

paralysis caused by west nile virus infection. *J Neuropathol Exp Neurol*, 68(10):1061–72, Oct 2009.

Appendix: R codes

```r
## Count matrix prepartation
rm(list=ls())
setwd("/Volumes/Seagate_Backup_Plus_Drive/zika")
sampleTable <- read.table("sample_table.txt", header = T, sep = "\t")
filenames <- file.path(paste0(sampleTable$Run_s, ".bam"))
library("Rsamtools")
bamfiles <- BamFileList(filenames)
library("GenomicFeatures")
library("GenomicAlignments")
gtffile <- file.path("Mus_musculus.GRCm38.86.chr.gtf")
txdb <- makeTxDbFromGFF(gtffile, format="gtf")
genes <- exonsBy(txdb, by="gene")
genes <-renameSeqlevels(genes,c("1"="chr1", "2"="chr2", "3"="chr3",
                                "4"="chr4","5"="chr5", "6"="chr6",
                                "7"="chr7", "8"="chr8", "9"="chr9",
                                "10"="chr10", "11"="chr11",
                                "12"="chr12", "13"="chr13",
                                "14"="chr14","15"="chr15",
                                "16"="chr16", "17"="chr17",
                                "18"="chr18","19"="chr19",
                                "X"="chrX", "Y"="chrY", "MT"="chrM"))
se <- summarizeOverlaps(features=genes, reads=bamfiles, singleEnd=TRUE)
save(file="zika_objects_all.Rda", se)
load("zika_objects_all1.Rda")
head(assay(se))
se$infection_status_s
#check the millions of fragments that uniquely aligned to the genes.
round(colSums(assay(se)) / 1e6, 1 )
## Pre-filtering the dataset and rlog transformation
library(DESeq2)
dds <- DESeqDataSet(se, design = ~ infection_status_s)
nrow(dds)
#Pre-filtering the dataset
dds <- dds[rowSums(counts(dds)) > 1, ]
nrow(dds)
dds <- estimateSizeFactors(dds)
#transform to log2 scale
#sequencing depth correction is done automatically for the rlog method
rld <- rlog(dds, blind=FALSE)
head(assay(rld), 3)
#Scatterplot of transformed counts from two samples (Mock1 vs Zika1)
```

```
pdf("Scatterplot.pdf", width=6, height=3)
par( mfrow = c( 1, 2 ) )
plot(log2(counts(dds, normalized=TRUE)[,c(1,3)] + 1),
     pch=16, cex=0.3, xlab="Mock", ylab="Zika", main="log2 transform")
plot(assay(rld)[, c(1:3)],
     pch=16, cex=0.3, xlab="Mock", ylab="Zika", main="rlog transform")
dev.off()
#PCA plotting
library("pheatmap")
library("RColorBrewer")
sampleDists <- dist(t( assay(rld) ) )
sampleDistMatrix <- as.matrix(sampleDists )
rownames(sampleDistMatrix) <- paste( rld$infection_status_s, sep="-" )
pdf("PCA.pdf", width=6, height=4)
plotPCA(rld, intgroup = c("infection_status_s"))
dev.off()
#Differential expression analysis
##MA-plot
dds <- DESeq(dds)
pdf("MAplot.pdf", width=4.5, height=3)
plotMA(dds)
dev.off()
##Significant differentially expressed genes identified by DESeq2
res <- results(dds, alpha = 0.05)
mcols(res, use.names=TRUE)
summary(res)

#volcano plot
tab <- data.frame(logFC = res$log2FoldChange, negLogPval = -log10(res$
    padj))
pdf("Volcano.pdf", width=4.5, height=4)
par(mar = c(5, 4, 4, 5))
plot(tab, pch = 16, cex = 0.6, xlab = expression(log[2]~fold~change),
     ylab = expression(-log[10]~adjpvalue), xlim=c(-5, 5), ylim=c(0,
         60))
lfc = 1
pval = 0.05
signGenes = (abs(tab$logFC) > lfc & tab$negLogPval > -log10(pval))
points(tab[signGenes, ], pch = 16, cex = 0.8, col = "red")
abline(h = -log10(pval), col = "green3", lty = 2)
abline(v = c(-lfc, lfc), col = "blue", lty = 2)
mtext(paste("adjpval =", pval), side = 4, at = -log10(pval), cex = 0.8,
     line = 0.5, las = 1)
```

```
mtext(c(paste("-", lfc, "fold"), paste("+", lfc, "fold")), side = 3, at
    = c(-lfc, lfc), cex = 0.8, line = 0.5)
dev.off()

#No. of significant genes
res.05 <- results(dds, alpha=.05)
table(res.05$padj <= .05)
#Fold change
resLFC1 <- results(dds, lfcThreshold=1)
#Number of genes with Padj <= 0.05 and fold change >=2
table(resLFC1$padj <= 0.05)

##Significant differentially expressed genes identified by edgeR
library(edgeR)
edger <- DGEList(assay(se), group=colData(se)$infection_status_s)
edger <- edger[rowSums(cpm(edger)) > 1, ]
edger <- calcNormFactors(edger, method=c("TMM"))
edger <- estimateCommonDisp(edger)
edger <- estimateTagwiseDisp(edger)
res_edgeR <- exactTest(edger)
res_edgeR <- topTags(res_edgeR, n = nrow(edger$table),
                     adjust.method="BH", p.value=1)$table
resSig_edgeR <- res_edgeR[res_edgeR$FDR < 0.05, ]
dim(resSig_edgeR)
##Method comparison: DESeq2 vs. edgeR
res_DESeq <- results(dds, alpha=.05)
res_DESeq <- res_DESeq[!is.na(res_DESeq$padj),]
resSig_DESeq <- res_DESeq[res_DESeq$padj < .05, ]
source("http://faculty.ucr.edu/~tgirke/Documents/R_BioCond/My_R_Scripts
    /overLapper.R")
setlist <- list(edgeR=rownames(resSig_edgeR), DESeq=rownames(resSig_
    DESeq))
OLlist <- overLapper(setlist=setlist, sep="_", type="vennsets")
counts <- sapply(OLlist$Venn_List, length)
pdf("Vennplot.pdf", width=4.5, height=4)
vennPlot(counts=counts, mysub="")
dev.off()

#Annotation
library("genefilter")
library("AnnotationDbi")
library("org.Mm.eg.db")
res$symbol <- mapIds(org.Mm.eg.db, keys=row.names(res), column="SYMBOL"
```

```
    ,
122                       keytype="ENSEMBL", multiVals="first")
123 res$entrez <- mapIds(org.Mm.eg.db, keys=row.names(res), column="
    ENTREZID",
124                       keytype="ENSEMBL",multiVals="first")
125 nres <- res[!is.na(res$symbol),]
126 resOrdered <- nres[order(nres$padj),]
127 resOrderedDF <- as.data.frame(resOrdered)
128 write.csv(resOrderedDF, file="results.csv")
129 sig <- resOrdered[!is.na(resOrdered$padj) &
130                     resOrdered$padj<0.05 &
131                     abs(resOrdered$log2FoldChange)>=1,]
132 write.table(rownames(sig), "sig.ensem.txt", row.names = FALSE, col.
    names=FALSE, quote=FALSE)
133 sig.symbol <- sig[, c(2:6)]
134 row.names(sig.symbol) <- sig$symbol
135 sig.symbol <- data.frame(sig.symbol)
136 write.table(rownames(sig.symbol), "sig.symbol.txt", row.names = FALSE,
    col.names=FALSE)
137 library(xtable)
138 print.xtable(xtable(sig.symbol, digits=c(0, 3,3, 3,-3, -3)), type = "
    latex", file = "siggenes.tex",
139              latex.environments = "center", include.rownames = TRUE)
140 #heatmapping
141 #select genes
142 selected <- rownames(sig)
143 selectedddds <- dds[rownames(dds) %in% selected,]
144 selectedrld <- rlog(selectedddds, blind=FALSE)
145 selectedmat <- assay(selectedrld)
146 selectedmat <- selectedmat - rowMeans(selectedmat)
147 selectedmat1 <- selectedmat
148 rownames(selectedmat1) <- sig$symbol[match(rownames(selectedmat1),
    rownames(sig))]
149 df <- as.data.frame(colData(selectedrld)[,c("infection_status_s")])
150 colnames(df) <- "group"
151 pdf("Heatmap.pdf", width=4.5, height=3, onefile=FALSE)
152 pheatmap(selectedmat1, annotation_col=df)
153 dev.off()
154 #Goterm enrichment analysis
155 universe <- rownames(resOrdered)
156 library(org.Mm.eg.db)
157 genemap <- select(org.Mm.eg.db, keys = selected,
158                   columns=c("ENTREZID","SYMBOL","GENENAME"), keytype="
```

```
                    ENSEMBL")
univmap <- select(org.Mm.eg.db, keys = universe,
                  columns=c("ENTREZID","SYMBOL","GENENAME"), keytype="
                    ENSEMBL")
library(GOstats)
param<- new ("GOHyperGParams", geneIds = genemap, universeGeneIds=
    univmap, annotation="org.Mm.eg.db", ontology="BP",pvalueCutoff=0.01,
     conditional=FALSE, testDirection="over")
# run analysis
hyp<-hyperGTest(param)
# visualize
hypres <- summary(hyp)[, c(1, 2, 3, 7)]
print.xtable(xtable(hypres), type = "latex", file = "hypres.tex",
             latex.environments = "center", include.rownames = FALSE)
```