

Research Article

The Differentially Expressed Genes and Biomarker Identification for Dengue Disease Using Transcriptome Data Analysis

Sunil Krishnan G, Amit Joshi and Vikas Kaushik*

Department of Bioinformatics, Lovely Professional University, Punjab, India

*Address Correspondence to: Vikas Kaushik, Department of Bioinformatics, Lovely Professional University, Punjab, India, E-mail: vikas.14664@lpu.co.in

Received: May 24, 2021; Accepted: June 07, 2021; Published: June 14, 2021

Copyright © 2021 Sunil Krishnan G. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

This bioinformatics and biostatistics study was designed to recognize and examine the differentially expressed genes (DEGs) linked with dengue virus infection in Homo sapiens. Thirty nine transcriptome profile datasets were analyzed by linear models for microarray analysis based on the R package of the biostatistics test for the identification of significantly expressed genes associated with the disease. The Benjamini and Hochberg (BH) standard operating procedure assessed DEGs had the least false discovery rate and chosen for further bioinformatics gene analysis. The large gene dataset was investigated for systematically extracting the biological significance of DEGs. Four clusters of DEGs were distinguished from the dataset and found the extracellular calcium sensing receptor gene expressing CASR protein was the most connecting human protein in the disease progression and discovered this protein as a potential biomarker for acute dengue fever.

Keywords: Dengue virus; Differentially expressed genes; Protein protein interaction; Bioinformatics2

Introduction

Female *Aedes aegypti* is a vector of the dengue viral disease spread. The virus is from the Flaviviridae family that infects the disease among human beings globally becomes a serious public concern [1]. Four major serotypes of DENV (1, 2, 3, and 4) share 65% genome similarity [2]. These serotypes were recognized as sources for mild to fatal dengue fevers [3,4]. The RNA genome encodes poly proteome for virus structure and reproduction [5,6]. The real time qPCR used for the detection and quantification of viremia best choice because the direct culture methods are difficult due to the low multiplication of clinical virus samples in cell culture. The inconsistency of plaque assay made quantification in viral diagnosis [7]. The DENV NS1 and IgM/IgG based diagnosis were used in the early stage of the disease but the results may vary with DENV serotypes and secondary infections [8]. A host dependent biomarker may be a solution for the consistent detection of disease progression

[9]. Several microarray studies acknowledged differentially expressed genes (DEGs) from multiple sample profiles [10,11]. Consistently DEGs identified from various previous studies [12-16] were used to make out a potential biomarker for the DENV disease. Meta-analysis approaches are common practice to discover novel DEG signatures for superior biomarkers and synthetic/biotherapeutics [17,18]. Here in this study, the gene expression patterns explored from the Gene Expression Omnibus (GEO) database. DEGs were compared and analyzed by Bioconductor supported Lima R package. Functionally related genes acknowledged by an integrated bioinformatics tool. DEGs corresponding PPI interaction network constructed. Detailed analysis of the data sets identified important biomarkers for therapeutic and diagnostic for dengue virus.

Materials and Methods

Retrieval of DENV microarray gene expression profile datasets

The microarray gene expression profile GEO dataset was retrieved from the GEO database [19]. The selected GSE17924 dataset contained 48 samples of host genome wide expression profiling during dengue disease [20]. The sample transcriptome dataset was used for the gene expression analysis.

Differentially expressed genes comparison analysis

From the GEO series, 14 DENV (1, 2, 3, and 4) samples cross compared and identified significant differentially expressed genes across experimental conditions using Bioconductor supported Lima R package. The Benjamini and Hochberg (BH) standard operating procedure was used for reduced the false P value discovery rate [19]. The median

centered distribution values of samples were selected for optimum cross comparability and differentially expressed gene identification.

Gene functional classification and identification of functionally related gene

The differentially expressed genes of DENV classified and functionally related gene groups using DAVID annotation and visualization tool retrieved from <https://david.ncifcrf.gov/tools.jsp>.

Protein interaction network analysis

The selected DEGs translating protein datasets analyzed and visualized the interaction networks and performed gene set enrichment analysis done by Gene Ontology (GO) and KEGG by using STRING. The resource availed form online at <https://string-db.org/>.

Results

DENV microarray gene expression profile data

The input of the ‘‘Dengue’’ query keyword has resulted in (n=15830) GEO Dataset. The search was then filtered by ‘expression profiling by array’ and ‘Top organism’- ‘Homo sapiens’ resulted (n=39) GEO dataset. GEO accession number GSE17924 selected. The selected dataset contained (n=48) samples of host genome wide expression profiling during dengue disease and filtered (n=14) samples for our study.

Identification of differentially expressed genes

The GEO samples (GSM) data identified from GEO2R through the GEO series accession number. The identified GSM samples are grouped into four according to the DENV serotypes. The GEO sample are GSM447796, GSM447797, GSM447815, GSM447822 (DENV 1), GSM447781, GSM447791, GSM447804, GSM447819

(DENV2), GSM447783, GSM447784, GSM447785, GSM447786 (DENV 3) and GSM447807, GSM447823 (DENV 4). The BH procedure narrowed false positives results [12]. The selected samples were found suitable for comparative analysis as per the determined calculated distribution values and visualized as a Boxplot in Figure 1A. This analysis compared the four groups of DENV serotypes samples data and identified the top 250 DEGs based on the lowest P value. Genes with the smallest P value are found the most significant in studies [19]. DENV 2 vs. 4 cross comparisons predicted the highest number of significantly expressed genes. The entire up and down regulated gene volcano plots are visualized in Figure 1B. Visualization for the expression density of samples in Figure 1C. The Venn diagram visualization of significantly expressed genes across DENV serotypes in Figure 1D. The mean variance relationship of expression data visualized in Figure 1E. The details of top up (n=5) and down (n=5) regulated DEGs are explained in Table I.

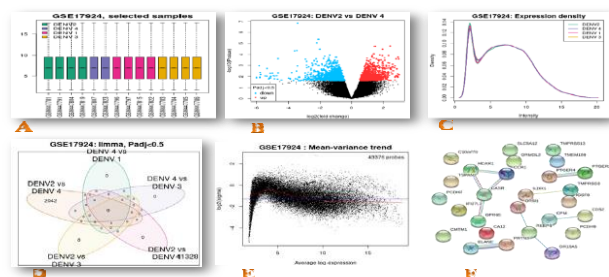


Figure 1: Visualization differentially expressed genes and Protein interaction (A) Boxplot of GSM sample’s distribution values (B) Volcano plot of up-regulated (red colour) and down-regulated (blue colour) DENV genes (C) Expression density of samples (D) Venn Diagram of significantly expressed genes across the DENV serotypes (E) Expression data Mean-variance relationship (F) Protein-Protein interaction of selected proteins.

Table 1: Details of top 5 Up and down-regulated differentially expressed genes.

Top 5 Down-regulated DEGs details			Top 5 Up-regulated DEGs details		
DEGs symbol	HGNC approved DEGs name	Log2 FC value for selected DEGs	DEGs symbol	HGNC approved DEGs name	Log2 FC value for selected DEGs
USHBP1	USH1 protein network component harmonin binding protein 1	-6.009	SYCP2	synaptonemal complex protein 2	2.893
WWTR1	WW domain containing transcription regulator1	-5.831	PLS3	plastin 3	2.903
ZBED9	Z inc finger BED-type containing 9	-5.792	PTGIS	prostaglandin I2 (prostacyclin) synthase	2.913
WNT5A	Wntfamily member 5A	-5.774	HCAR1	hydroxycarboxylic acid receptor 1	2.963
TRABD2B	TraB domain containing 2B	-5.739	OCLN	occludin	2.986

Gene functional classification and identification results

The systematically organized genes were useful to inter-

pret the biological importance of the expressed gene [12]. Through medium stringency, the genes were classified from the expressed gene list. Functional annotation cluster-

ing tool based on kappa statistics to quantitatively measure and identify functionally related genes are involved in the similar biological mechanism associated with a set of similar annotation terms [13]. This tool helped to reduce the redundancy and identify similar annotations of the DEG dataset. The kappa similarity and classification parameters predicted higher quality of functional classification [8]. From the DEGs dataset 34 genes are clustered into four big gene functional groups. The enrichment score determined the importance of the gene group from the gene list. Three gene groups were selected for further analysis based on the highest enrichment scored (>1). The first group of genes (CMTM1, IFI27L2, REEP5, C10orf76, TMEM199, ORM-DL2) had an enrichment score of 1.44. The second group of genes (SLC5A12, PCDH9, CaSR, PCDH7, GPR65, HCAR1, IGSF8, CA12, CPM, CCR1, OR5I1, CD52, PTGER4, OR10A5, PTGER2, ILDR1, TSPAN2) and the third group (TMPRSS3, TMPRSS13, PRTN3, ELANE) had enrichment score 1.3 and 1.26 respectively.

Network status of protein interaction and functional enrichments

The PPI network analysis shows 10 edges (interaction) jointly contribute to shared functional associations in the 27 network nodes (proteins). The predicted average node degree (0.741) and local clustering coefficient (0.494). Also found PPI enrichment p value (0.000685). The predicted PPI network view was visualized in Figure 1F, the edge line colored differently. These coloured lines correspond to the types of functional associations between proteins. Figures 2 and 3 visualized protein interaction clusters. The CaSR protein was identified as the highest interacting with Glycosphingolipid psychosine (GPR65), Ahydroxy carboxylic acid receptor 1 (HCAR1), and C-C chemokine receptor type 1 (CCR1) proteins. The CaSR plays a key role in the production of parathyroid hormone (PTH) and GPR65 has a role in immune responses. HCAR1 mediates its anti lipolytic effect and CCR1 responsible for affecting stem cell proliferation. The functional enrichment analysis predicted biological processes (n=10), molecular functions (n=8), and cellular components (n=4), GO terms significantly enriched in the predicted network. The down regulated CaSR is a proliferation marker in colorectal cancer [21], prostate cancer [22], and breast cancer [23]. Here we hypothesized from this study that CaSR protein was a potential biomarker in the DENV disease.

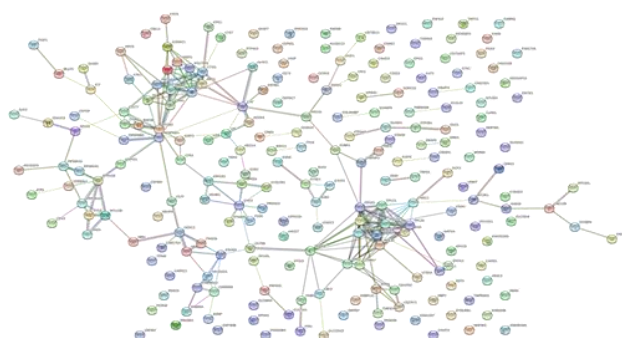


Figure 2: Visualization of protein interaction clusters.



Figure 3: Graphical abstract

Highlights

- Transcriptome profile data retrieval and analysis
- Identification of differentially expressed genes (DEGs) associated with dengue virus infection.
- Significant differentially expressed genes were statistically analyzed by the lima R package.
- Functionally related gene classification and identification.
- Gene Ontology (GO) gene set enrichment analysis of selected DEGs.
- Protein interaction network analysis and biomarker identification.

Discussion

Despite the reality, although dengue can be a significant subtropical illness, very little understood about the pathophysiology, attributable to the complicated cell activities that take place in sick. Numerous transcripts, as well as host genetic mechanisms, are elevated after dengue fever, according to transcriptional microarray data. Additional quantitative studies differentiated among host genetic networks step in establishing an intrinsic communication mechanism (Nuclear factor driven transcripts as well as the Interferon network) but those engaged in viral replication (NF- κ B driven transcripts or the Interferon route) (ubiquitin dependent proteasome) [24]. GEO2R tool was very effective for microarray profile analyzing datasets in many recent studies on dengue [25]. Log₂FC values assist in determining expression levels of host genes in response to dengue infection. In humans or other vertebrate animal tissues, p53/mitochondrial driven apoptotic_pathways were triggered by the dengue fever, microarray studies assist in DEG analysis to reveal such systems [26]. We identified the top 250 differentially expressed genes based on significant P values. DENV2 and 4 had the highest significantly expressed genes in GEO2R comparison analysis. This could be a potent biomarker for the therapeutic and diagnosis of dengue viral disease. The top five upregulated genes that were found in this study are USHBP1, WWTR1, ZBED9, WNT5A, TRABD2B; and top downregulated genes were SYCP2, PLS3, PTGIS, HCAR1, OCLN. CASR a G-protein coupled receptor protein found the highest interacting

and enrichment analysis results also supportive to our findings.

Conclusion

We identified the top 250 differentially expressed genes based on significant P values. DENV2 and 4 had the highest significantly expressed genes in GEO2R comparison analysis. The significant DEGs (n=34) are clustered into four big gene functional groups using the DAVID bioinformatics tool and three groups contain 27 genes selected for further analysis based on the highest enrichment scored (>1). PPI network analysis shows 10 protein interactions among the nodes and selected one protein which is highly interacting with the other 3 protein in the network. CASR a G-protein coupled receptor protein found the highest interacting and enrichment analysis results also supportive of the findings. This could be a potent biomarker for the therapeutic and diagnosis of dengue viral disease.

Acknowledgments

Authors SKG, AJ, and, VK are grateful to, Bioinformatics division of Lovely professional university, Jalandhar, Punjab, India for providing a computational and bioinformatics environment for this computational research.

References

1. H. Caraballo, K. King, Emergency department management of mosquito-borne illness: malaria, dengue, and West Nile virus. *Emerg Med Pract*, 6 (2014),5:1-23.
2. B.W. Johnson, B.J. Russell, R. S. Lanciotti, Serotype specific detection of dengue viruses in a fourplex real-time reverse transcriptase PCR assay. *J Clin Microbiol*, 43(2005),10:4977-83.
3. D. J. Gubler, Dengue and dengue hemorrhagic fever. *Clin Microbiol Rev*, 11(1998), 3:480-96.
4. F.P. Pinheiro, S.J. Corber, Global situation of dengue and dengue haemorrhagic fever, and its emergence in the Americas. *World Health Stat Q*. 50(1997), 161-9.
5. G.S. Krishnan, A. Joshi, N. Akhtar, V. Kaushik, Immunoinformatics designed T cell multi epitope dengue peptide vaccine derived from non structural proteome. *Microb Pathog*, 150 (2021):104728
6. R.J. Kuhn, W. Zhang, M.G. Rossmann, S. V. Pletnev, J. Corver, et al. Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell*. 8(2002), 5:717-25
7. M.M. Choy, B.R. Ellis, E.M. Ellis, D.J. Gubler, Comparison of the mosquito inoculation technique and quantitative real time polymerase chain reaction to measure dengue virus concentration. *Am J Trop Med Hyg*. 89(2013), 5:1001-5.
8. J.G. Low, E.E. Ooi, S.G. Vasudevan, Current status of dengue therapeutics research and development. *J Infect Dis*. 215(2017),S96-S102.
9. P.Y. Shu, L.K. Chen, S.F. Chang, Y.Y. Yueh, L. Chow, et al. Comparison of capture immunoglobulin M (IgM) and IgG enzyme-linked immunosorbent assay (ELISA) and nonstructural protein NS1 serotype-specific IgG ELISA for differentiation of primary and secondary dengue virus infections. *Clin Diagn Lab Immunol*. 10(2003), 4:622-30.
10. Y. Xu, C. Qiao, S. He S, C. Lu, S. Dong, et al. Identification of functional genes in pterygium based on bioinformatics analysis. *Biomed Res Int*. 20(2020):2383516.
11. L. Gu, J. Ni, S. Sheng, K. Zhao, C. Sun, J. Wang. Microarray analysis of long non-coding RNA expression profiles in Marfan syndrome. *Exp Ther Med*. 20(2020),4:3615-3624.
12. S.B. Halstead, S. Mahalingam, M.A. Marovich, S. Ubol, D.M. Mosser. Intrinsic antibody-dependent enhancement of microbial infection in macrophages: disease regulation by immune complexes. *Lancet Infect Dis*. 10(2010):712-22.
13. Y. Qi, Y. Li, L. Zhang, J. Huang, MicroRNA expression profiling and bioinformatic analysis of dengue virus infected peripheral blood mononuclear cells. *Mol Med Rep*. 7(2013),3:791-8.
14. P. Sun P, J. García, G. Comach, M.T. Vahey, Z. Wang, B.M. Forshey, et al. Sequential waves of gene expression in patients with clinically defined dengue illnesses reveal subtle disease phases and predict disease severity. *PLoS Negl Trop Dis*. 11(2013) :7(7):e2298.
15. P.A. Tambyah, C.S. Ching, S. Sepramaniam, J.M. Ali, A. Armugam, et al. microRNA expression in blood of dengue patients. *Ann Clin Biochem*. 2016 Jul;53(Pt 4):466-76.
16. P. Becquart, N. Wauquier, D. Nkoghe, A. Nd-joyi-Mbiguino, C. Padilla, et al. Acute dengue virus 2 infection in Gabonese patients is associated with an early innate immune response, including strong interferon alpha production. *BMC Infect Dis*. 10(2010),10:356.
17. D. Toro-Domínguez D, J. A. Villatoro-García, J. Martorell-Marugán, Y. Román-Montoya Y, M.E. Alarcón-Riquelme, et al. A survey of gene expression meta-analysis: Methods and applications. *Brief Bioinform*. 22(2021),2:1694-1705.
18. M. F. D. de Souza, A. F. da Silva Filho, A.P. de Barros Albuquerque, M. W. L. Quirino, de Souza Albuquerque MS, et al. Overexpression of UDP-Glucose 4-Epimerase Is associated with differentiation grade of gastric cancer. *Dis Markers* 20(2019), 6325326.
19. T. Barrett, S. E. Wilhite, P. Ledoux P, C. Evangelista, I.F. Kim, et al. NCBI GEO: Archive for functional genomics data sets update. *Nucleic Acids Res*. (2013),D991-5
20. S. Devignot, C. Sapet, V. Duong, A. Bergon, P. Rihet, et al. Genome-wide expression profiling deciphers host responses altered during dengue shock syndrome

- and reveals the role of innate immunity in severe dengue. *PLoS One*, 5(2010),7:e11671
21. A. Aggarwal, M. Prinz-Wohlgenannt, S. Tennakoon, J. Höbaus, C. Boudot, et al. The calcium-sensing receptor: A promising target for prevention of colorectal cancer. *Biochim Biophys Acta*. 1853(2015),9:2158-67
 22. G. Baio, G. Rescinito, F. Rosa, D. Pace, S. Boccardo, et al. Correlation between Choline Peak at MR spectroscopy and calcium sensing receptor expression level in breast cancer: A preliminary clinical study. *Mol Imaging Biol*, 17(2015), 4:548-56
 23. S. Das, P. Clézardin, S. Kamel, M. Brazier, R. Mentaverri. The CaSR in Pathogenesis of Breast Cancer: A New Target for Early Stage Bone Metastases. *Front Oncol*. 5(2020),5:10:69.
 24. J. Fink, F. Gu F, L. Ling, T. Tolfvenstam, F. Olfat F, et al. Host gene expression profiling of dengue virus infection in cell lines and patients. *PLoS Negl Trop Dis*. 1(2007), 2:e86.
 25. T. T. N. Thao, E. de Bruin, H. T. Phuong, N. H. Thao Vy, H. J. van den Ham, et al. Using NS1 flavivirus protein microarray to infer past infecting dengue virus serotype and number of past dengue virus infections in Vietnamese individuals. *J Infect Dis*. (2020), 22:jiaa018.
 26. A. M. Nasirudeen, D. X. Liu. Gene expression profiling by microarray analysis reveals an important role for caspase-1 in dengue virus-induced p53-mediated apoptosis. *J Med Virol*. 81(2009):61069-81.