SOMYA NOG

POMPERAUG REGIONAL HIGH SCHOOL

SOUTHBURY, CONNECTICUT

✉ snog6127@gmail.com

Using Bioinformatics to investigate the possibility of developing a vaccine to prevent the CNS-related symptoms of West Nile Virus

Somya Nog

Southbury, Connecticut

ABSTRACT

Environmental phenomena such as climate change and the changes in weather patterns they bring continue to transform the epidemiological landscape of vector-borne illnesses such as West Nile virus (WNV). WNV is a vector-borne illness primarily transmitted by Culex mosquitoes. While most people infected with the virus don't show any symptoms, the incidence of severe cases of WNV that impact the central nervous system (CNS) or brain has increased, warranting the need for a vaccine to respond to these effects. While development for an equine or horse vaccine for WNV has been successful, and there have been efforts to synthesize a human vaccine, unfortunately, there is still a lack of a conclusive and effective clinical way to manage WNV symptoms, which, in the most fatal cases, include impaired motor functioning and other neurocognitive ramifications. Thus, to address this problem of a lack of clear and effective treatment for WNV, this study employs various bioinformatics tools and databases (NCBI, GEO2R, SR Plot, GO and KEGG) to pinpoint genes that are differentially expressed in response to the WNV infection and the molecular pathways they belong to. As such, the goal of this research is to identify key viral genes and their associated potential biological processes that can be utilized to develop a mRNA vaccine for WNV. This study identified CXCL11/10, CCL8, OAS1, ISG15, and BST2 as key viral genes and that they are associated with immune responses such as those initiating chemokine and cytokine activity, which can therefore serve as potential biomarkers for further advancement in vaccine research.

INTRODUCTION

West Nile Virus (WNV), the most widespread manifestation of the class of over 70 positive-strand RNA viruses known as flaviviruses, is a vector-borne illness primarily transmitted by Culex mosquitoes (Pacenti et. al, 2015). In humans, a bite by the infected mosquito presents asymptomatically in a vast majority of cases, but up to 20% of patients have reported headaches, nausea, body aches, febrile illness, and assorted rashes (Qian et. al, 2013). In especially severe cases, the infection may spread to the central nervous system (CNS), most commonly localizing to the spinal cord, peripheral motor neurons, and dorsal root ganglia, where it can impart long-lasting neurocognitive consequences. These can include inflammation (encephalitis / meningitis), motor weakness, memory loss, depressive disorders, and more as the virus persists for up to 4 months post-bite (Fulton et. al, 2020). In fatal cases, the virus has been found to target primarily the cerebellum, which organizes fine motor coordination, and the substantia nigra, which is responsible for initiating voluntary movement (Guarner et. al, 2004). As such, the serious health impacts of WNV highlight the dire need for ongoing research to develop an effective vaccine to prevent this debilitating illness.

Some of the largest factors shaping current spread of WVN or epidemiology of WNV are those related to climate such as rainfall levels, humidity, winds, and ambient temperature (Paz, 2015). Even though extreme temperatures can reduce the survival of young or larval Culex mosquitoes, the virus replicates faster at warmer temperatures. There is a connection between heatwaves and the severity of outbreaks in humans (Kunkel et. al, 2006). As precipitation is concerned, heavy rainfall is known to facilitate mosquito larval survival and development by contributing to higher standing water levels (Chevalier et. al, 2014). Indeed, these factors have begun to play a greater role in the spread of the mosquitoes that transmit WNV, in part due to environmental phenomena including global warming. For example, in Europe and Western Asia, more frequent and intense heat waves led to WNV outbreaks in new areas because there were more mosquitoes, the vectors that carry the WNV. Another example is in the eastern USA where higher-than-average rainfall was soon followed by human WNV outbreaks. In Australia, cases of WNV-induced encephalitis, inflammation of the brain, hit a record high after extensive flooding (Paz, 2015).

Despite the rapidly increasing incidence of WNV infections in humans, a definitive cure such as a vaccine has yet to be developed and distributed (Gould et. al, 2023). Vaccines are important because they train the immune system to recognize and fight specific disease causing agents or factors, and therefore provide immunity and prevent serious illness. Interestingly, there have been novel discoveries that have led to the development of a WNV horse or equine vaccine. This vaccine uses microdoses of the killed virus injected intramuscularly to limit and/or fully prevent adverse WNV-related reactions in horses by 94-96% (Ng et. al, 2003). Several factors have slowed the development of a WNV vaccine. These include the higher chance of severe

neurological disease in horses compared to humans, the economic impact of infections in horses, and the lack of other ways to control the virus in horses, as humans can more easily avoid mosquito bites than horses can (Ulbert, 2019). Nonetheless, in humans, clinical trials have tested various classes of vaccines such as recombinant subunit, inactivated whole-virus, first and second generation DNA, and live attenuated chimeric (weakened viruses containing different parts). Even though vaccines such as the live attenuated chimeric showed promising immunogenicity in these trials, factors such as being unable to measure infection prevention in the face of short-lived viremia (presence of viruses in the blood), along with the sheer scale at which trials should be conducted to demonstrate efficacy, have prevented further clinical development (Gould et. al, 2023).

Irrespective, the question still stands - can a vaccine that targets the CNS symptoms of WNV be developed? In the face of an ever-expanding epidemiological landscape that is becoming more and more difficult to predict, a vaccine would surely prove more effective in guarding against infection and potentially even death at the hand of the virus (Gould et. al, 2023).

Thus, this research work aims to understand the mechanisms behind CNS-related symptoms in WNV infection in order to assess the possibility of using this information to develop a vaccine. To achieve this goal, this research study will use different bioinformatics tools and databases to identify genes that are expressed in the WNV with the goal to identify potential viral genes that can be used to create a vaccine for WNV viral infections. These genes can then be tested in the laboratory by scientists to determine if the viral genes can be recognized by the human immune system and successfully fight the WNV infection.

MATERIALS AND METHODS

Identification of Genes Differentially Expressed in West Nile Virus

In order to identify the viral genes and the biological pathways through which West Nile Virus infections impact the CNS, various bioinformatics tools were used. Data was collected from the NCBI's Gene Expression Omnibus (GEO), an online hub of bioinformatics data (Sayers et. al, 2021). The National Center for Biotechnology Information (NCBI) is a US government-operated service that provides a platform for researchers all around the world to access and provide biological and genomic data and resources (Barrett et. al, 2009). Within GEO, the tool GEO2R was used, which applies pre-programmed algorithms (see Supplementary data) to the datasets housed within GEO to identify top differentially-expressed genes (DEGs) between user-generated groups (Kalyan et. al, 2023).

In this investigation, the microarray dataset with accession number GSE122798 was used, which housed the data collected by Maximova et. al, 2021. This dataset contains gene samples taken

from *Rhesus Macaques* infected with a 1999 strain of WNV at different time intervals after infection (n = 18), and from members of the same species that did not have the infection, referred to as the "mock infection" in the study (n = 8). Samples were taken from both the cerebellum and lumbar cord of the patients. Categorization of the samples based on the presence of infection, as well as the region the sample was taken from, revealed differences in gene expression between WNV and normal samples upon GEO2R analysis. GEO2R is a tool that links the R programming language with the Gene Expression Omnibus (GEO) database. In this present study, pre-programmed AI and ML algorithms using the R programming language were used to generate the results for the identified and utilized dataset GSE122798 (Supplementary Results: R Script for GEO2R Analysis).

The top 50 DEGs (Supplementary Results: DEG Data from GEO2R) were then screened to 50 genes using statistics; genes with $|logFC| \geq 1$ and p-value < 0.05 were considered statistically significant. Among the 50 DEGs, 25 were upregulated and 25 were downregulated. This study's methodology is summarized in Figure 1.

Functional Analysis of the Selected Top DEGs

The top 50 DEGs were then downloaded and imported into the online SR Plot bioinformatics tool to perform a GO enrichment analysis and KEGG pathway analysis (Kanehisa et al., 2012). This tool uses pre-programmed algorithms to identify the biological process, cellular compartment, and molecular function these genes participate in based on each DEG's logFC value, a measure of the difference in gene expression between the two previously defined groups (Subramanian et. al, 2005).
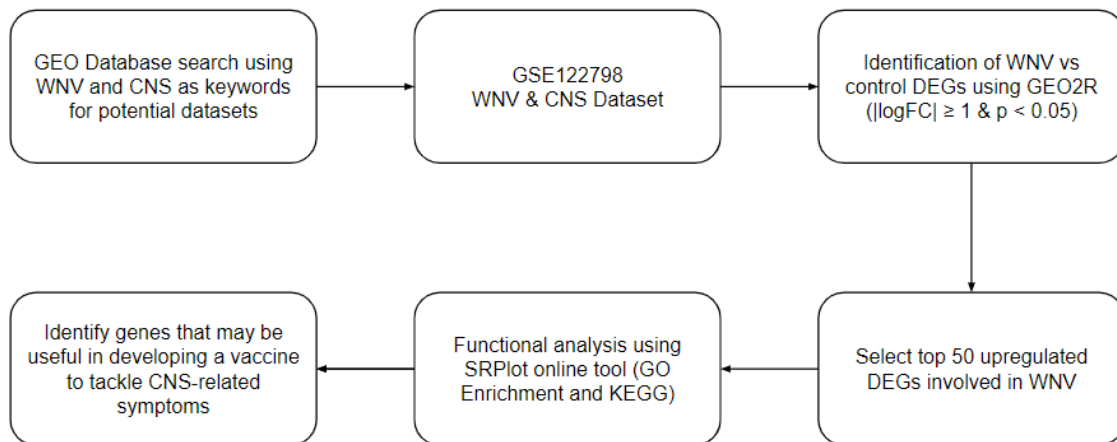

Figure 1: Summary of the methodology undertaken in this study. Overview of steps and bioinformatics tools and databases used in this study

RESULTS

Identification of top Differentially Expressed Genes (DEGs)

Differentially expressed genes, or DEGs, are genes that see a statistically significant variation in expression level as measured by mRNA or protein production in microarrays in response to changing experimental conditions such as timeframe, temperature, disease state, tissue location, and more (Porcu et. al, 2021). Here, the varying conditions used to identify DEGs are the presence of WNV and the location of the sample (cerebellum vs lumbar cord). DEGs can either be upregulated, meaning that they are expressed at a higher level in one experimental condition as compared to the other, or vice versa when they are downregulated. It is crucial to pinpoint these genes as they can reveal much about the molecular mechanisms through which diseases impact living systems (Porcu et. al, 2021).

GEO2R was implemented to highlight the top DEGs in the data set with accession number GSE122798 when divided into the groups containing data from patients with and without WNV infection. The top DEGs as calculated by the GEO2R analysis were identified, with the red dots signifying upregulated genes, blue representing downregulated genes, and black correlating to statistically insignificant ($p \geq 0.05$) genes in the volcano plot (Figure 2). The top DEG was ISG15, which was consistently upregulated in the infected samples as displayed in Figure 3. GEO2R analysis with 4 experimental groups was also conducted to visualize the overlap in gene expressions between WNV infection in different locations in a venn diagram (Figure 4). From here, the top 50 DEGs were downloaded (Supplementary Data).
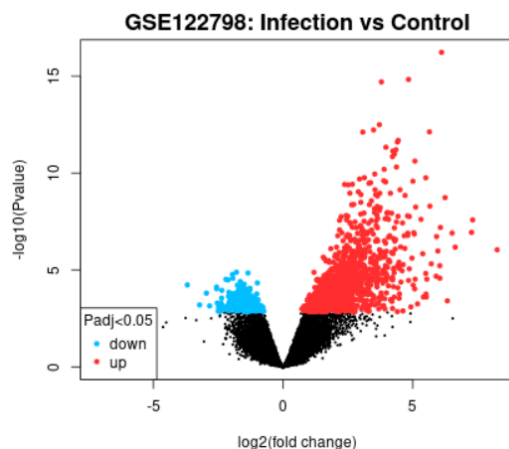
Figure 2a: Volcano plot illustrating the scope of differential expression in samples taken from patients with WNV infection vs healthy patients.
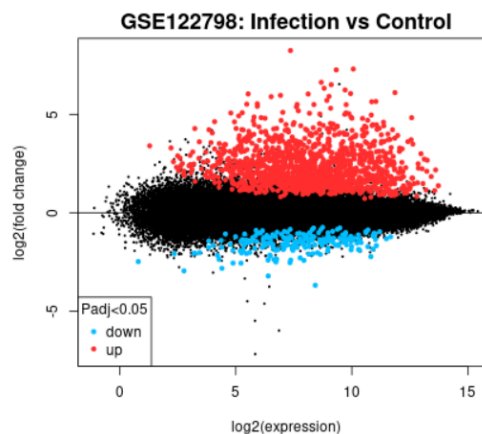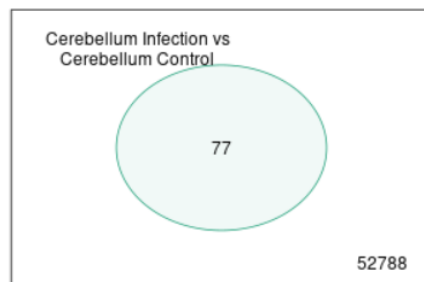
Figure 2b: Mean difference plot illustrating the scope of differential expression in samples taken from patients with WNV infection vs healthy patients.
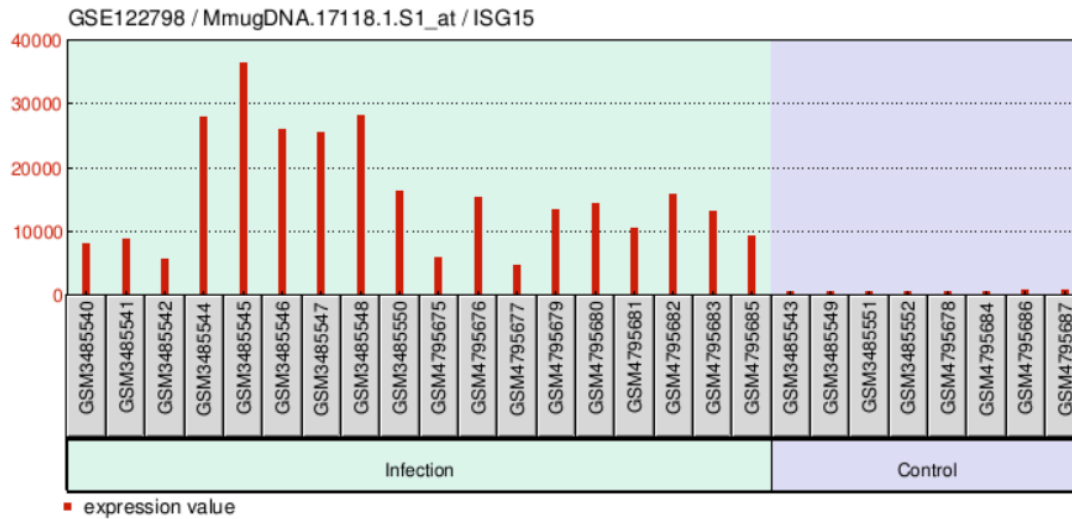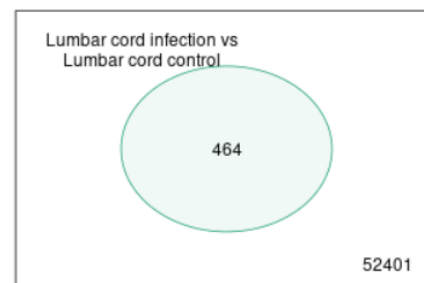
Figure 3: A bar plot-style depiction of the different levels at which the top DEG, ISG15, was expressed amongst the infection (n = 18) and control (n = 8) samples.

Figure 4a: A venn diagram illustrating the quantities of genes expressed at similar levels between the infection (n = 9) and control (n = 4) groups collected from the cerebellum.

Figure 4b: A venn diagram illustrating the quantities of genes expressed at similar levels between the infection (n = 9) and control (n = 4) groups collected from the lumbar cord.

Functional analysis of the top DEGs

Once the top DEGs are identified, it is then useful to understand how exactly the upregulation of these genes impacts the cell, or what role each plays in responding to the infection. Through the SR Plot online tool, GO enrichment pathways and KEGG pathways were performed on the top 50 DEGs by logFC measurement, with the defense response, extracellular space, and chemokine activity being the most enriched biological process, cellular component, and molecular function

enriched amongst the DEGs (Figure 5a), multiple DEGs being involved in multiple biological processes (Figure 5b), and the most common KEGG pathways including cytokine receptor binding, chemokine receptor binding, and GPCR binding as well (Figure 5c).
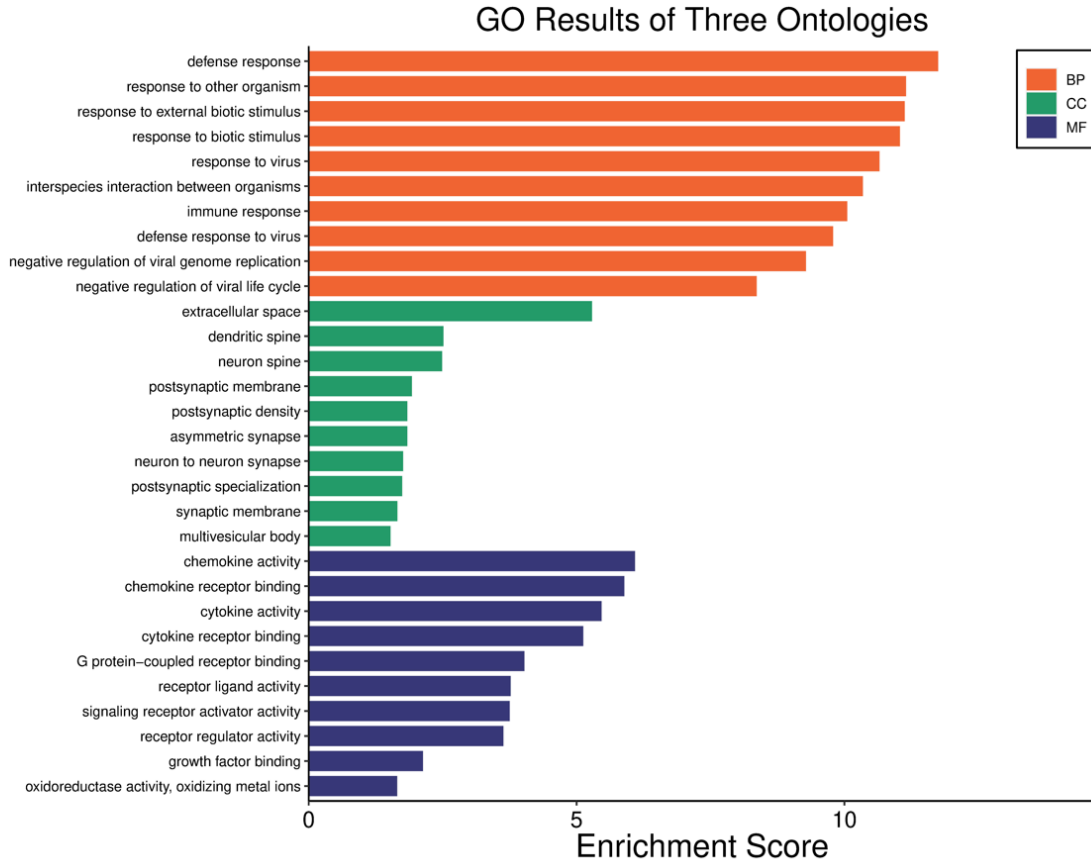
Figure 5a: The GO pathway enrichment results analyzing the level of representation of various biological processes (BP), cellular components (CC), and molecular functions (MF) within the top DEGs.
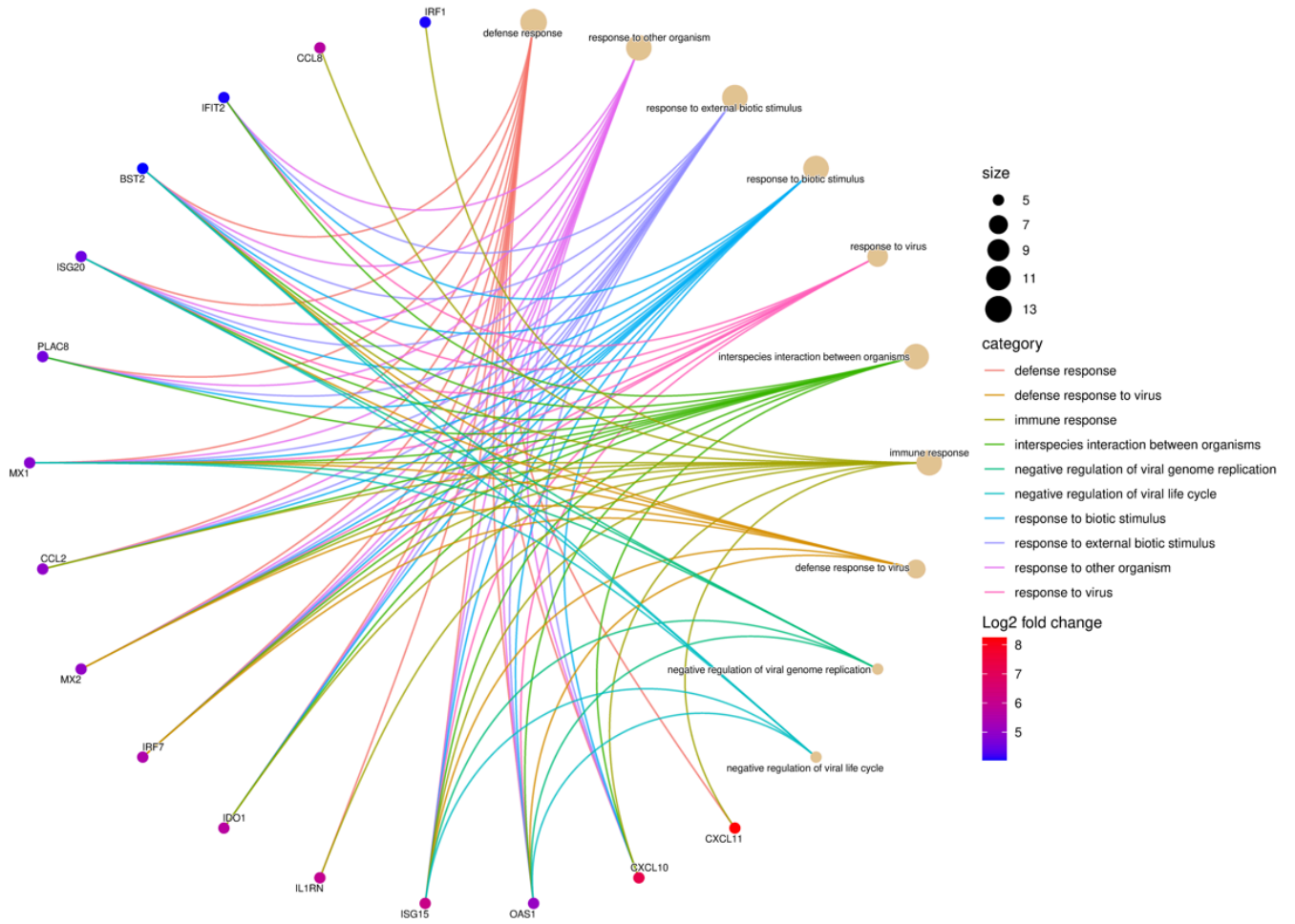
Figure 5b: Another graphic depicting the multiple roles the top 16 DEGs play in the infection process. The dots, or nodes, represent genes / biological functions, with each colored connection representing a biological function and the color of the dot representing the level at which each gene is upregulated, where red is the most significant.
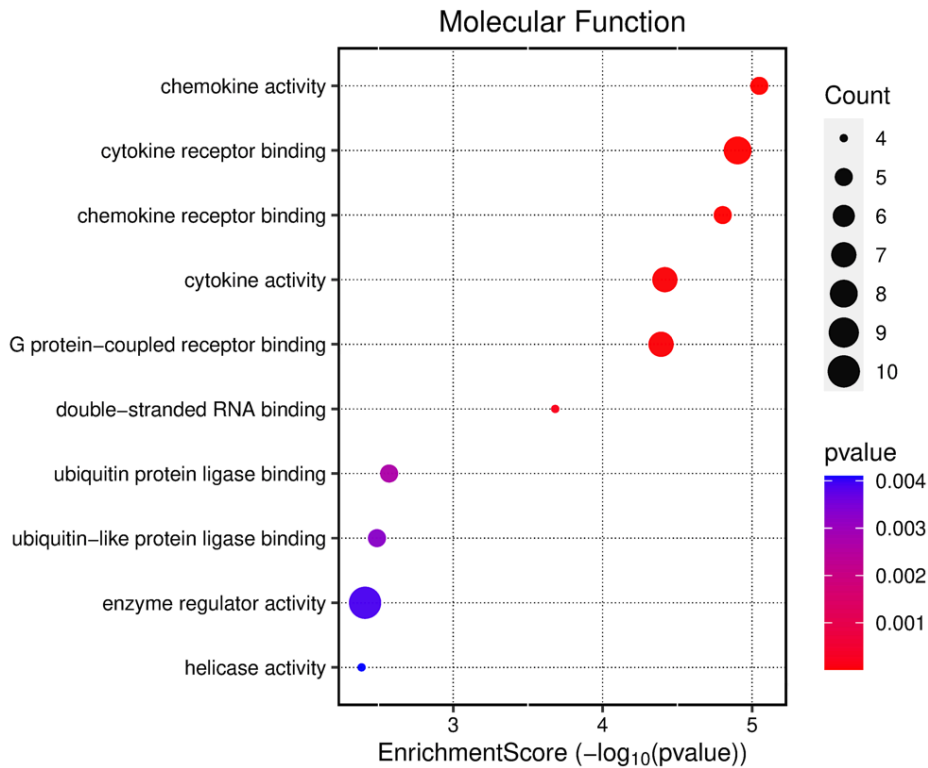
Figure 5c: A KEGG pathway enrichment dot plot highlighting the levels at which each molecular function is upregulated in the WNV infection as well as how many genes are correlated with each function.

SUPPLEMENTARY RESULTS

R Script for GEO2R Analysis

DEG Data from GEO2R

DISCUSSION

West Nile Virus (WNV) can cause very serious and sometimes deadly illnesses including those that impact the central nervous system (CNS) or brain.In the United States since 1999, it has led to over 1,500 deaths, and in Europe in 2018, it caused 181 deaths (Ulbert, 2019). In the past twenty five years, scientists have created several potential vaccines to protect humans from West Nile Virus (WNV). Some of these have been tested in clinical trials, but none have been approved for use yet (Ulbert, 2019; CDC WNV Treatment and Prevention). Currently, the main way WNV is prevented is through programs that control mosquitoes (Petersen et al, 2013). These

programs are cheaper than making vaccines, because vaccine development needs a lot of time researching, clinical trials, and money for testing. However, mosquito control only works for a short time. Vaccines would be a better way to protect people from WNV in the long run because they provide lasting protection.

Therefore, the goal of this research study was to understand how West Nile Virus (WNV) affects the central nervous system (CNS) to explore the possibility of developing a vaccine. To achieve this goal, this research used various bioinformatics tools and databases to identify genes that are expressed by WNV which could be used to create a vaccine. Results from this study can then be used in future studies by scientists interested in creating a vaccine through testing these genes in the lab to see if the human immune system can recognize them and effectively fight the WNV infection.

There are different types of vaccines which include live weakened (attenuated), inactivated, subunit, toxoid, mRNA, and viral vector vaccines. These various types all aim to train the immune system to recognize and fight specific disease causing agents (pathogens), and thereby provide immunity and prevent serious illness. Generally, vaccines work by introducing antigens, which are weakened, inactive, or parts of a disease causing agent (pathogen), to stimulate the immune system without causing disease. This exposure prompts the production of antibodies and memory cells, which remember the antigen and enable a quick response to future infections.

Because WNV is an mRNA virus, this research study focussed on the identification of West Nile viral genes that can be used to develop an mRNA vaccine to combat this disease. To make mRNA vaccines, scientists take a piece of DNA with the instructions for making a part of the virus (the antigen) and use it to create mRNA in a lab (Rosa et al, 2021). The mRNA is introduced into the cells of an organism such as a human and once inside the cells, the cell reads the mRNA and makes the virus part, which helps the immune system learn to fight the virus. This mRNA is flagged by the cell's receptors, which then prompts the immune system to prepare to defend the body (Rosa et al, 2021).

Using NCBI's GEO2R bioinformatics tool and database, this research study identified 50 West Nile viral differentially expressed genes (DEGs). Amongst these top 50 DEGs (Supplementary Results: DEG Data from GEO2R; Figure 2a), many genes coding for ubiquitins and interferon-induced proteins were noted (Table 1), which align with the current correlation between these 2 types of proteins and immune response to the WNV infection in the CNS. While the specific mechanism through which ubiquitins, which aid in protein degradation, contribute to the signaling antagonism undertaken by flaviviruses such as WNV in causing encephalitis is unknown (Hay-McCullagh and Morris, 2021), it is thought that interferons play a crucial role in some of the first immune pathways activated upon initial viral infection, such as the control of viral replication (Walker et. al, 2021). In fact, the top DEG, ISG15 (see Figure 3), codes for a

ubiquitin-like protein that serves a chief role in the upregulation of interferon-stimulated genes associated with host response to the WNV infection (Perng and Lenschow, 2018). And, the most versatile DEGs, which, according to Figure 5b, include BST2 and OAS1, have similar functions, inhibiting the release of enveloped virions from infected cells (Zhao et. al, 2022) and activating RNAses that degrade viral RNA (Jiang et. al, 2023).

Furthermore, the results from Figure 5c indicate that the top DEGs are most likely associated with cytokine receptor binding, chemokine receptor binding, and G protein-coupled receptor (GPCR) binding. The upregulation of genes coding for ligands that bind to chemokines, such as some of the top DEGs CXCL11 and CXCL 10 (Supplementary Results: DEG Data from GEO2R) is essential in recruiting other immune cells such as neutrophils, T cells, dendritic cells, and macrophages to the site of the infection to begin the immune response to WNV (Benzarti et. al, 2023). On the other hand, cytokines, whose ligands include those coded by genes such as the 3rd top DEG, CCL8, aid in long-term immunity and viral clearance by promoting the differentiation of B and T cells (Benzarti et. al, 2023). In both protein signaling pathways, GPCRs are commonly utilized as receptors, aiding in the activation of antiviral cells (Benzarti et. al, 2023).

In all, the initiation of the WNV infection primarily causes upregulation of genes instrumental in the immune response to the virus. The top DEGs identified in this study, including CXCL11/10, CCL8, OAS1, ISG15, and BST2 (all of whose functions are summarized in Table 1), can be utilized by scientists to develop a mRNA vaccine that targets the CNS-related symptoms of WNV to help boost the production of ubiquitins and interferon-induced proteins as well as initiate chemokine and cytokine activities. Such a combination could potentially not only build up immunity to WNV, but also aid in defending against the virus should one's immunity ever be compromised (Gould et. al, 2023).

Table 1: A summary of the identified top DEGs and their associated functions

GENE ID	FUNCTION
CXCL11/10	Ligand that activates chemokines, which recruit immune cells such as macrophages, neutrophils, T cells, and dendritic cells to the infection site
CCL8	Ligand that activates cytokines, which are part of the long-term immune response and effort to pursue viral clearance
OAS1	Activate RNAses that degrade viral RNA
ISG15	Codes for a ubiquitin that stimulates interferon-induced genes associated with host immune response, mainly through stabilization of cytokines

BST2	Inhibits the release of enveloped virions from foreign cells

As the primary bioinformatics datasets used in this study were taken from microarray experiments conducted by other researchers, the identified DEGs will need to be further experimented upon and understood in clinical environments in order to develop a vaccine. Additionally, while GEO2R, the tool that provided the algorithm identifying the top DEGs, is a novel service, its in-built analysis software could not be used to analyze every single dataset the database stores, indicating a need for further data collection in order to generate new datasets. Furthermore, not every DEG generated by GEO2R had an associated gene ID, which was required for SR Plot analysis.

In any case, the identified genes and molecular pathways should be further studied in laboratory environments or clinical settings to determine the effectiveness of the immune response in vaccine format. However, considering that the cerebellum and lumbar cord have very different functions and mechanisms through which they operate, specialized research focusing on WNV infection's impacts on each structure would be recommended. It may also be interesting for similar experiments to be conducted with WNV infections at different environmental temperatures, moisture levels, or other climate-related variables in order to best understand how to adapt vaccine development for the virus's spreading epidemiological landscape.

Conclusion

In summary, the goal of this work was to identify genes of interest that can be utilized to develop a vaccine targeting the CNS-related symptoms of West Nile Virus (WNV) infections. To do this, the NCBI's GEO2R bioinformatics tool and database was employed. The results were then inputted into a GO Enrichment and KEGG pathway bioinformatics tool called SR Plot (Kanehisa et al., 2012) in order to functionally characterize the top differentially expressed genes. These bioinformatics tools highlighted five genes involved in chemokine and cytokine signaling pathways, especially those coding for proteins acting as ligands for the receptors that initiate these immune mechanisms. The human response to these genes, namely, CXCL11/10, CCL8, OAS1, ISG15, and BST2 (Table 1) can be tested further, for the development of a vaccine which could potentially both prevent the development of the West Nile Virus into the CNS in the most fatal cases and provide additional immune protection.

REFERENCES

Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., ... & Edgar, R. (2009). NCBI GEO: archive for high-throughput functional genomic data. Nucleic acids research, 37(suppl_1), D885-D890.

Benzarti, E., Murray, K. O., & Ronca, S. E. (2023). Interleukins, chemokines, and tumor necrosis factor superfamily ligands in the pathogenesis of West Nile virus infection. Viruses, 15(3), 806.

Treatment and prevention of West Nile virus disease. (2024). Centers for Disease Control and Prevention.

Chevalier, V., Tran, A., & Durand, B. (2014). Predictive modeling of West Nile virus transmission risk in the Mediterranean Basin: how far from landing?. International Journal of Environmental Research and Public Health, 11(1), 67-90.

Fulton, C. D., Beasley, D. W., Bente, D. A., & Dineley, K. T. (2020). Long-term, West Nile virus-induced neurological changes: A comparison of patients and rodent models. Brain, Behavior, & Immunity-Health, 7, 100105.

Gould, C. V., Staples, J. E., Huang, C. Y. H., Brault, A. C., & Nett, R. J. (2023). Combating west nile virus disease—Time to revisit vaccination. New England Journal of Medicine, 388(18), 1633-1636.

Guarner, J., Shieh, W. J., Hunter, S., Paddock, C. D., Morken, T., Campbell, G. L., ... & Zaki, S. R. (2004). Clinicopathologic study and laboratory diagnosis of 23 cases with West Nile virus encephalomyelitis. Human pathology, 35(8), 983-990.

Hay-McCullough, E., & Morrison, J. (2021). Contributions of Ubiquitin and Ubiquitination to Flaviviral Antagonism of Type I IFN. Viruses, 13(5), 763.

Jiang, S., Deng, X., Luo, M., Zhou, L., Chai, J., Tian, C., ... & Luo, Z. (2023). Pan-cancer analysis identified OAS1 as a potential prognostic biomarker for multiple tumor types. Frontiers in Oncology, 13.

Kalyan, G. U., Reddy, D. P., Chandrakanth, G., Pooja, B., Anitha, V., & Vivek, D. (2023, October). Gene Association Disease Prediction by GEO2R Tool. In 2023 International Conference on Evolutionary Algorithms and Soft Computing Techniques (EASCT) (pp. 1-6). IEEE.

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), 27-30.

Kunkel, K. E., Novak, R. J., Lampman, R. L., & Gu, W. (2006). Modeling the impact of variable climatic factors on the crossover of Culex restauns and Culex pipiens (Diptera: Culicidae), vectors of West Nile virus in Illinois. American Journal of Tropical Medicine and Hygiene, 74(1), 168-173.

Maximova, O. A., Sturdevant, D. E., Kash, J. C., Kanakabandi, K., Xiao, Y., Minai, M., ... & Pletnev, A. G. (2021). Virus infection of the CNS disrupts the immune-neural-synaptic axis via induction of pleiotropic gene regulation of host responses. Elife, 10, e62273.

Ng, T., Hathaway, D., Jennings, N., Champ, D., Chiang, Y. W., & Chu, H. J. (2003). Equine vaccine for West Nile virus. Developments in biologicals, 114, 221-227.

Paz, S. (2015). Climate change impacts on West Nile virus transmission in a global context. Philosophical Transactions of the Royal Society B: Biological Sciences, 370(1665), 20130561.

Petersen, L. R., Brault, A. C., & Nasci, R. S. (2013). West Nile virus: review of the literature. Jama, 310(3), 308-315.

Porcu, E., Sadler, M. C., Lepik, K., Auwerx, C., Wood, A. R., Weihs, A., ... & Kutalik, Z. (2021). Differentially expressed genes reflect disease-induced rather than disease-causing changes in the transcriptome. Nature communications, 12(1), 5647.

Perng, Y. C., & Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. Nature Reviews Microbiology, 16(7), 423-439.

Qian, F., Chung, L., Zheng, W., Bruno, V., Alexander, R. P., Wang, Z., ... & Montgomery, R. R. (2013). Identification of genes critical for resistance to infection by West Nile virus using RNA-Seq analysis. Viruses, 5(7), 1664-1681.

Rosa, S. S., Prazeres, D. M., Azevedo, A. M., & Marques, M. P. (2021). mRNA vaccines manufacturing: Challenges and bottlenecks. Vaccine, 39(16), 2190-2200.

Sayers, E. W., Beck, J., Bolton, E. E., Bourexis, D., Brister, J. R., Canese, K., ... & Sherry, S. T. (2021). Database resources of the national center for biotechnology information. Nucleic acids research, 49(D1), D10.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceedings of the National Academy of Sciences, 102(43), 15545-15550.

Ulbert, S. (2019). West Nile virus vaccines–current situation and future directions. Human Vaccines & Immunotherapeutics, 15(10), 2337-2342.

Walker, F. C., Sridhar, P. R., & Baldridge, M. T. (2021). Differential roles of interferons in innate responses to mucosal viral infections. Trends in Immunology, 42(11), 1009-1023.

Zhao, Y., Zhao, K., Wang, S., & Du, J. (2022). Multi-functional BST2/tetherin against HIV-1, other viruses and LINE-1. Frontiers in Cellular and Infection Microbiology, 12, 979091.