RISING RESEARCHERS

Journal of Scientific Research Writing

SPRING 2025



STUDENT RESEARCH JOURNAL

Introduction of Rising Researchers **3**
Research Goals 3
Students Have Mastered The Following Skills And Tasks 4
Academic Team 4
Journal Editor 5
Rising Researchers Students **6**
Neel Anand 6
Havish Reddy Apparacheruvu 12
Tanmayi Chennapragada 18
Jai Degala 23
Devraj Dutta 37
Janani Venkat Easwar 53
Alexandria E. Feit 59
Rohan Gazula 72
Kamya Kainth 91
Gowtham Kantamneni 116
Suhruth Koneru 122
Kaavya Kotamraju 137
Rohan Kuruganthy 144
Rithvik Lal 160
Lauren Lee 172
Samreen Mehta 178
Siddharth Nanda 205
Thien Nguyen 223
Somya Nog 228
Esha Nuttaki 244
Ankan Pal 258
Eshika Paruchuri 264
Karthik Peddu 284
Nikhil Pesati 299
Sahaana Rangaraj 315
Maya Sammanna 334
Dev Sathyan 357
Sana Sidhu 362
Kavya Singh 367

Surya Srinivasan 376
Ibrahim Syed 391
Katelyn Szeto 410
Ria Tandra 428
Diana Wang 439

Introduction of Rising Researchers

This journal edition highlights students who participated in the Rising Researchers virtual research intensive. The course was designed to provide students with an introduction to the scientific method and an opportunity to sharpen their research skills. The core concepts learned include project design, manuscript writing, data analysis, and an introduction to public health and medicine.

Through lectures, hands-on experiments, public electronic health records databases, machine learning algorithms, statistical skills and tools, class and small group discussions, and oral presentations, students explored the computational world with emphasis on human diseases like cardiovascular disease and health equity issues.

The hands-on patient cohort selection and machine learning modeling emphasized the proper use of the scientific method to answer a research question, develop a hypothesis, carry out computational tasks, make observations, analyze, interpret and communicate results.

Research Goals

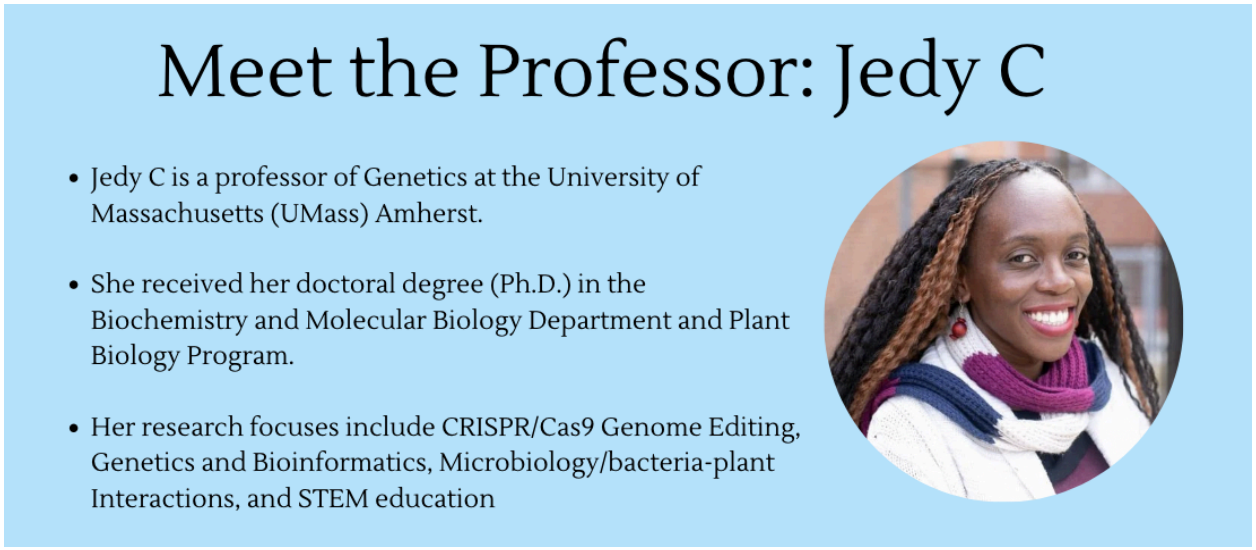
- Describe and demonstrate the ability to use the scientific method
- Independently construct a research question and design a novel research project
- Demonstrate competency in utilizing research tools and resources and interpreting data to support the research question
- Demonstrate competency to draw reasonable conclusions based on research findings
- Development of a research strategy, including the definition and justification of patient

inclusion criteria, as well as study hypotheses
- Investigation of existing literature to build reference and resource lists
- Demonstrate the ability to communicate research and results through a manuscript and poster presentation
- Creation of research report for publication submission
- Final research poster presentation and publication in academic research journal

Students Have Mastered The Following Skills And Tasks

- Describe and demonstrate the ability to use the scientific method
- Demonstrate an understanding of research design, including patient selection criteria, generation and justification of hypotheses, and choosing appropriate hypothesis testing tools
- Demonstrate ability to record and interpret data from experiments
- Demonstrate ability to run multiple experiments to evaluate different hypotheses
- Demonstrate the ability to communicate research and results

Academic Team



Meet the Professor: Loren B

- Loren B is an MD/PhD candidate at the University of North Carolina at Chapel Hill
- Her research focus is on health behaviors and disparities.
- Prior to medical school, Loren was a research assistant at Duke University, where she wrote and published scientific manuscripts and led qualitative and quantitative research studies.



Journal Editor

Meet the Editor: Kavya S

- Kavya Singh is a rising senior at Providence High School in Charlotte, North Carolina.
- Her research interests include pediatrics and neurological disorders.
- Prior to Rising Researchers, Kavya completed a month-long internship in India where she got to interview medical professionals as well as see how medicine is practiced in 3rd world countries.



Rising Researchers Students

The following students successfully participated in the 2024/2025 Rising Researchers Course.

Neel Anand

Using Bioinformatics to Investigate Potential Molecular Biomarkers that link Alzheimer's Disease with Type 2 Diabetes

ABSTRACT

Background: Athletic stress, characterized by high-intensity physical and physiological demands, leads to significant cardiovascular adaptations in athletes. These adaptations are crucial for meeting the increased demands placed on the heart during training and competition. Despite the benefits, there is a notable risk of cardiac events, such as cardiac arrest, particularly among athletes. This underscores the importance of understanding how athletic stress impacts cardiovascular function and the potential long-term implications.

Objective: To investigate the effects of athletic stress on cardiovascular function, comparing key cardiovascular parameters between athletes and non-athletes, and to highlight the importance of awareness and monitoring of athletes' cardiovascular health.

Methods:A scoping literature review was conducted using PubMed and Google Scholar. Key search terms included "athletic stress," "cardiovascular system," "exercise stress impact on heart," and "cardiovascular assessment in athletes." Inclusion criteria focused on peer-reviewed studies published in English from January 2000 to July 2024, involving human subjects and examining the cardiovascular impact of athletic stress. Exclusion criteria included studies published before 2000, non-human subjects, and articles not in English.

Results: Athletes exhibited significant cardiovascular adaptations, including a lower resting heart rate (54 bpm) compared to non-athletes (70 bpm), lower blood pressure (112/68 mmHg vs. 124/76 mmHg), higher cardiac output (6.0 L/min vs. 4.8 L/min), and greater left ventricular mass (150 grams vs. 120 grams). These findings demonstrate enhanced cardiovascular efficiency and structural adaptations in response to athletic stress.

Discussion: The results confirm that athletic stress leads to positive cardiovascular adaptations, such as improved myocardial efficiency and greater cardiovascular endurance. However, the magnitude of left ventricular hypertrophy observed was higher than expected, highlighting the need for continued monitoring of athletes' cardiovascular health. The findings suggest that cardiovascular adaptations to athletic stress are beneficial but also emphasize the importance of interdisciplinary approaches, including sports medicine, preventative care, and rehabilitation, to ensure long-term cardiovascular health in both athletes and the general population.

Keywords: athletic stress, cardiovascular system, cardiovascular adaptation to exercise, exercise stress impact on heart, sports and heart disease, and cardiovascular assessment in athletes, and Boolean operators

INTRODUCTION

High intensity workouts put strain on an athlete's body due to physical and physiological stress during training, practice, and games. This stress may be connected to the fact that one in 50,000 athletes experience cardiac arrest [1]. Intense physical activity may also lead to arrhythmias. Currently, the issue is the lack of awareness for athlete's wellness when it comes to sports. The complexity of these health issues leads to minimal post look out after treatment to these athletes knowing they are stereotyped to be "healthy" all the time. In this paper I aim to provide an understanding of how stress upon athletes relates to cardiovascular function in order to identify how to promote cardiac health among athletes.

METHODS

I searched PubMed and Google Scholar using the key terms: "athletic stress," "cardiovascular system," "cardiovascular adaptation to exercise," "exercise stress impact on heart," "sports and heart disease," and "cardiovascular assessment in athletes" combined with Boolean operators to refine the search terms. Articles published before 2000, not published in English, or not conducted in humans were excluded. Other literature reviews were also excluded. Inclusion criteria required that the articles address how athletic stress impacted the cardiovascular system. My search strategy refined the articles from 1,500 to 100 articles.

RESULTS

Cardiovascular Responses to Athletic Stress

The resting heart rate (HR) was found to be lower in athletes compared to non-athletes. The mean resting HR for athletes was 54 bpm (beats per minute), while non-athletes averaged a total of 70 bpm [2].

Table 1. Average resting heart rate among athletes compared to non-athletes

Study Type	**Average Resting HR (bpm)**
Athletes	54
Non-Athletes	70

Blood pressure (BP) is a measurement that shows athletes generally had lower systolic and diastolic numbers when being compared to non-athletes. The average systolic BP for athletes was 112 mmHg, while non-athletes had an average systolic BP of 124 mmHg [2]. The average diastolic BP for athletes was 68 mmHg in comparison to 76 mmHg in non-athletes [2].

Table 2. Systolic and diastolic blood pressure of athletes compared to non-athletes

Study Type	**Systolic BP (mmHg)**	**Diastolic BP (mmHg)**
Athletes	112	68
Non-Athletes	124	76

Cardiac output (CO) is measured in liters per minute where it can be seen it was higher in athletes. The average cardiac output for athletes was 6.0 L/min, whereas non-athletes averaged 4.8 L/min [2, 3].

Table 3. Cardiac output of athletes compared to non-athletes

Study Type	Cardiac Output (L/min)
Athletes	6.0
Non-Athletes	4.0

Left ventricular mass is measured in grams of the left ventricular. Based on prior data it shows an increased left ventricular blood flow with an average LVM of 150 grams compared to 120 grams in non-athletes.

Table 4. Left ventricular mass of athletes compared to non-athletes

Study Type	Left Ventricular Mass (grams)
Athletes	150
Non-Athletes	120

Athlete's cardiovascular adaptations included improved stroke volume (cardiac output), enhanced myocardial efficiency (heart rate), and greater overall cardiovascular endurance (blood pressure) [6].

DISCUSSION

An important correlation that needs to be noted from the results is that "athletic stress" does improve cardiovascular function in order to fulfill the needs of an athlete. This cardiovascular adaptation can be seen by the vital signs between athletes and non-athletes. This is exhibited by the lower resting heart rate of 54 bpm compared to the non-athlete who shows 70 bpm [3,4]. In athletes due to the physical stress it causes more oxygen to be delivered which decreases the amount of pumps needed per minute. Additionally seeing they also have a lower blood pressure both systolically and diastolically for similar reasons. As based on this adaptation it strengthens the myocardium muscles. This strengthening efficiently pumps the blood with less effort on the arteries which can also improve the endothelial function. Upon vitals that increased was the higher cardiac output (6.0 L/min vs. 4.8 L/min), and increased LVM (150 grams vs. 120 grams) [6,7]. This is due to knowing when pumping the amount of blood per

beat is lower in athletes, this means the amount of blood being pumped per beat increases. This can be shown for the substantial increase in cardiac output. The Left ventricular mass is a subunit of that in which it only tracks one sector of the heart [4]. By this alone it should show a positive correlation by an increase in cardiac output would show an increase in left ventricular mass. This overall shows the impact of how exercise can provide an impact upon cardiovascular health.

Most of the results were predicted to occur due to prior studies and research that consistently showed that exercise/athletic activity provides an improvement on cardiovascular efficiency and health [2, 5-11]. However, the magnitude of these changes, particularly the increased left ventricular mass, was surprising. When realizing that athletes are prone to left ventricular hypertrophy, it is supported based on the exponential increase of left ventricular mass. This shows even upon adaptation an athlete's heart still needs to be looked out based on knowing they have an equal chance of having problems just like a sedimentary individual.

Based on this study, cardiovascular adaptation doesn't just need to be focused on in sports medicine as it can be interdisciplinary with other subjects like preventative medicine and rehabilitation. This study shows the importance of athletic activity to cardiovascular health.

References

1. Mayo Clinic Staff. (2024 3). Sudden death in young people: Heart problems often blamed Review of Sudden death in young people: Heart problems often blamed]. Mayo Clinic; MFMER. https://www.mayoclinic.org/diseases-conditions/sudden-cardiac-arrest/in-depth/sudden-death/art-20047571
2. Athletes and Maximum Heart Rates: What to Know. (2021, February 17). UPMC HealthBeat. https://share.upmc.com/2021/02/athlete-heart-rate/
3. Caselli, Stefano, et al. "Upper Normal Values of Blood Pressure Response to Exercise in Olympic Athletes." American Heart Journal, vol. 177, 1 July 2016, pp. 120–128, www.sciencedirect.com/science/article/abs/pii/S0002870316300448, https://doi.org/10.1016/j.ahj.2016.04.020.

4. Berge, H. M., et al. “Blood Pressure and Hypertension in Athletes: A Systematic Review.” British Journal of Sports Medicine, vol. 49, no. 11, 1 June 2015, pp. 716–723, pubmed.ncbi.nlm.nih.gov/25631543/, https://doi.org/10.1136/bjsports-2014-093976.
5. Lovic, D., Narayan, P., Pittaras, A., Faselis, C., Doumas, M., & Kokkinos, P. (2017). Left ventricular hypertrophy in athletes and hypertensive patients. Journal of Clinical Hypertension, 19(4), 413–417. https://doi.org/10.1111/jch.12977
6. Romero, S. A., Minson, C. T., & Halliwill, J. R. (2017). The cardiovascular system after exercise. Journal of Applied Physiology, 122(4), 925–932. https://doi.org/10.1152/japplphysiol.00802.2016
7. Lorell, B. H., & Carabello, B. A. (2000). Left ventricular hypertrophy. Circulation, 102(4), 470–479. https://doi.org/10.1161/01.cir.102.4.470
8. Caselli, S., Cicconetti, M., Niederseer, D., Schmied, C., Jost, C. A., & Pelliccia, A. (2022). Left ventricular hypertrophy in athletes, a case-control analysis of interindividual variability. International Journal of Cardiology, 348, 157–162. https://doi.org/10.1016/j.ijcard.2021.12.009
9. Lavie, C. J., Arena, R., Swift, D. L., Johannsen, N. M., Sui, X., Lee, D., Earnest, C. P., Church, T. S., O’Keefe, J. H., Milani, R. V., & Blair, S. N. (2015). Exercise and the cardiovascular system. Circulation Research, 117(2), 207–219. https://doi.org/10.1161/circresaha.117.305205
10. Giese, E. A., O&Apos;CONNOR, F. G., Brennan, F. H., Jr, Depenbrock, P. J., & Oriscello, R. G. (2007, April 1). The Athletic preparticipation evaluation: cardiovascular assessment. AAFP. https://www.aafp.org/pubs/afp/issues/2007/0401/p1008.html
11. Atchley, A. E., & Douglas, P. S. (2007). Left Ventricular hypertrophy in Athletes: morphologic features and clinical correlates. Cardiology Clinics, 25(3), 371–382. https://doi.org/10.1016/j.ccl.2007.06.009
12. Qiu, Y., Fernández-García, B., Lehmann, H. I., Li, G., Kroemer, G., López-Otín, C., & Xiao, J. (2023). Exercise sustains the hallmarks of health. Journal of Sport and Health Science/Journal of Sport and Health Science, 12(1), 8–35. https://doi.org/10.1016/j.jshs.2022.10.003
13. Predel, H. (2014). Marathon run: cardiovascular adaptation and cardiovascular risk. European Heart Journal, 35(44), 3091–3098. https://doi.org/10.1093/eurheartj/eht50.

The Effect Of Therapeutic Drugs on Tau Proteins

Havish Reddy Apparacheruvu

ABSTRACT

Background: This research delves into the role of therapeutic agents' influence on tau protein levels in one's brain, ultimately exemplifying their ability to impact cognitive health.

Objective: Through a literature review, this research helps organize information presented in different research studies regarding therapeutic agents and their ability to affect tau protein levels.

Methods: Sources for this literature review were found through screening original research articles found through PubMed analyzing the effect of therapeutic agents on tau protein levels and human cognitive development.

Conclusion: Therapeutic agents have a paramount effect on tau levels in the brain, yet these effects vary depending on the agent being used. This research helped depict the potential of certain therapeutic agents in the neuroscience field and how impactful they could be when faced with neurodegenerative diseases such as Alzheimer's or Parkinson's disease.

Keywords: tau proteins, neurodegenerative diseases, and therapeutic agents

INTRODUCTION

Neurodegenerative diseases are proven to be more prevalent in the elderly [1]. Most neurodegenerative diseases lead to dementia, which is estimated to affect 150 million people worldwide by 2050 and currently costs around $10 trillion to the healthcare system [1]. With that being said, our future relies on therapeutic agents to fight these neurodegenerative diseases, with the first therapeutic agents approved by the FDA in 2003 [2]. A key pathophysiological mechanism identified with neurodegenerative diseases is excess tau levels, leading to tangles as well as amyloid plaques combining together to ultimately interfere with connectivity between neurons [3].

Some of the therapeutic agents that were identified throughout this literature review were posited to help participants with Alzheimer's disease in slowing the process in which tau proteins tangle and interfere with

communication linkage. As therapeutic agents are still being developed, their refined strategies are proving to make a difference in the field of neurodegenerative diseases [1].

Therapeutic agents may also be found in everyday consumables such as seeds and vegetables like cauliflower and lettuce [4]. Not only do these help patients with symptoms of neurodegenerative disease, but they are also known to fight them, as it is proven that cauliflower and lettuce contain a type of therapeutic agent named sulforaphane, which ultimately helps break down tau protein tangles as well as plaque created by beta amyloid fragments, both of which play a huge role in dementia.

All in all, therapeutic agents are extremely essential in our war against neurodegenerative diseases; therefore, it is important to research in this area due to the amount of potential that it may bring to the field of neuroscience.

METHODS

The review used PubMed to identify studies for inclusion. Keywords included tau proteins, neurodegenerative diseases, and therapeutic agents, with a filter option to specifically look for clinical trials.

Inclusion and exclusion criteria

Inclusion criteria for clinical trials included at least 100 participants, publication year after 2015, and quantitative result reporting.

RESULTS

In this review article, 5 articles were included that demonstrated the influence of therapeutic agents in targeting tau proteins [5-9]. Five drugs were reviewed in these articles collectively, with Suvorexant proving to have a paramount effect on tau protein deceleration while other agents such as MAPTRx were proven to reduce inflammation and neurogranin levels in patients with mild Alzheimer's disease [5,6]. Some drugs were found to also have other benefits, such as Donanemab, which reduces not only tau levels but also amyloid levels, presenting a unique finding in the field of neuroscience [7]. Although many of the drugs tested in the articles were proven to have some effect on Alzheimer's, others, such as somorinemab,

were proven to not have an effect on Alzheimer's disease, as they had little to no influence on tau protein levels throughout the 73-week time period they were monitored [9].

Study	**Focus**	**Information**	**Relevance**
"Tau-targeting antisense oligonucleotide MAPTRx in mild Alzheimer's disease: a phase 1b, randomized, placebo-controlled trial" [5]	The differing effect of MAPTRx versus a placebo on tau proteins in Alzheimer's disease.	Inflammation decreased within those dosed with MAPTRx, whereas those dosed with placebo remained the same.	This clinical trial suggests that MAPTRx may play a role in reducing inflammation for patients with Alzheimer's. Neurograining levels were reduced within those dosed with MAPTRx. Ultimately, MAPTRx helped alleviate symptoms for those diagnosed with Alzheimer's.

“Suvorexant Acutely Decreases Tau Phosphorylation and Aβ in the Human CNS” [6]	Therapeutic agents aimed at decelerating the progress of tau protein blockades.	The ratio of phosphorylated to unphosphorylated tau-threonine-181 was decreased 10%–15% in participants treated with suvorexant compared to placebo.	Given the role of tau protein blockades in the progression of Alzheimer’s disease, by decelerating this process, the drug poses a possibility to slow the progression of the disease.
“Exploratory Tau Biomarker Results From a Multiple Ascending-Dose Study of BIIB080 in Alzheimer Disease: A Randomized Clinical Trial” [8]	Varying dosages of BIIB080 may differentially influence the reduction of tau protein levels within patients.	The treatment BIIB080 was analyzed to lower tau protein buildup. Higher doses were proven to have greater efficacy with decreasing tau levels.	The relevance of this research is established through the slowing of tau accumulation through BIIB080, with higher doses showing increased efficacy.
“Safety and Efficacy of Semorinemab in Individuals With Prodromal to Mild Alzheimer Disease: A Randomized Clinical Trial” [9]	Analyze the effect of somorinemab on tau proteins to see if it may stop the accumulation and slow mild Alzheimer’s disease.	Semorinemab was deemed safe as well as observed to not have influenced tau protein levels; therefore, it didn’t succeed in slowing the effect of Alzheimer’s disease.	The relevance of this research is established through semorinemab being proven to not have an effect on the reduction of tau proteins; therefore, no effect on AD as well.

DISCUSSION

All in all, these articles exemplify the effect of therapeutic agents on tau protein levels, providing insights into Alzheimer's disease management. Most of these drugs influence the body through the reduction of tau proteins.

For example, suvorexant lowers phosphorylated tau threonine 181 levels [6]. Previously, suvorexant was used to treat patients with insomnia by selectively blocking the binding between receptors and neuropeptides [6]. This same clinical drug used for insomnia has helped block the bindings between tau proteins to prevent tangles that promote Alzheimer's disease. BIIB080 was also found to reduce Alzheimer's disease growth through the slowing of tau protein growth [8]. Different doses of BIIB080 reduced different amounts of CSF t-tau and p-tau181 [8]. It is likely that the reduction of CSF t-tau and p-tau181 may result in slowed disease progression in Alzheimer's disease. One of the more unique drugs being researched is donanemab; donanemab specializes in the reduction of tau levels as well as amyloid levels, and patients dosed with donanemab, who had complete plaque clearance within a 24-week time period, were proven to have a greater reduction in tau levels [7]. Interestingly, though MAPTRx has little to no effect on tau protein levels, its ability to alleviate symptoms for patients makes it all the more useful [5]. MAPTRx is able to execute this task through the reduction of inflammation, which can improve cognitive health [5]. MAPTRx also increases neurogranin levels, which, when reduced, play a role in cognitive decline [5]. Unfortunately, semorinemab has little to no influence on Alzheimer's disease, as the clinical trial showed zero improvement in cognitive health for patients with Alzheimer's disease [9]. Though semorinemab was proven to be safe for dosage, its lack of effectiveness in the context of Alzheimer's research suggests a need for further development [9].

The information stated above supports the hypothesis that various therapeutic agents may improve symptoms of neurodegenerative diseases such as Alzheimer's disease in the future. Future research should focus on continuing to develop, test, and implement therapeutic agents to improve patient care and quality of life

References

1. Temple S. Advancing cell therapy for neurodegenerative diseases. Cell Stem Cell. 2023;30(5):512-529. doi:10.1016/j.stem.2023.03.017
2. Athar T, Al Balushi K, Khan SA. Recent advances on drug development and emerging therapeutic agents for Alzheimer's disease. Mol Biol Rep. 2021;48(7):5629-5645. doi:10.1007/s11033-021-06512-9
3. Breijyeh Z, Karaman R. Comprehensive Review on Alzheimer's Disease: Causes and Treatment. Molecules. 2020;25(24):5789. Published 2020 Dec 8. doi:10.3390/molecules25245789
4. Schepici G, Bramanti P, Mazzon E. Efficacy of Sulforaphane in Neurodegenerative Diseases. Int J Mol Sci. 2020;21(22):8637. Published 2020 Nov 16. doi:10.3390/ijms21228637
5. Mummery CJ, Börjesson-Hanson A, Blackburn DJ, et al. Tau-targeting antisense oligonucleotide MAPTRx in mild Alzheimer's disease: a phase 1b, randomized, placebo-controlled trial [published correction appears in Nat Med. 2024 Jan;30(1):304. doi: 10.1038/s41591-023-02639-3]. Nat Med. 2023;29(6):1437-1447. doi:10.1038/s41591-023-02326-3
6. Lucey BP, Liu H, Toedebusch CD, et al. Suvorexant Acutely Decreases Tau Phosphorylation and Aβ in the Human CNS. Ann Neurol. 2023;94(1):27-40. doi:10.1002/ana.26641
7. Shcherbinin S, Evans CD, Lu M, et al. Association of Amyloid Reduction After Donanemab Treatment With Tau Pathology and Clinical Outcomes: The TRAILBLAZER-ALZ Randomized Clinical Trial. JAMA Neurol. 2022;79(10):1015-1024. doi:10.1001/jamaneurol.2022.2793
8. Edwards AL, Collins JA, Junge C, et al. Exploratory Tau Biomarker Results From a Multiple Ascending-Dose Study of BIIB080 in Alzheimer Disease: A Randomized Clinical Trial. JAMA Neurol. 2023;80(12):1344-1352. doi:10.1001/jamaneurol.2023.3861
9. Teng E, Manser PT, Pickthorn K, et al. Safety and Efficacy of Semorinemab in Individuals With Prodromal to Mild Alzheimer Disease: A Randomized Clinical Trial. JAMA Neurol. 2022;79(8):758-767. doi:10.1001/jamaneurol.2022.1375.

Tanmayi Chennapragada

Neurotransmitter Regulation Under Academic Stress

ABSTRACT

Background: Epinephrine and other neurotransmitters are released when the body faces a threat. However, less information is known regarding the nervous system's response during academic stress. Given that long or potent stressors can lead to mentally disabling states and disturbance in cognitive activities, this question is essential to explore for therapeutic interventions.
Objective: This review addresses the following research question: How does the nervous system regulate the release of transmitters like epinephrine in response to academic stress?
Methods: The research articles used in this review were identified by searching PubMed and Google Scholar with keywords including "neurological response to academics," "pressure," and "transmitter release."
Results: The nervous system controls the excretion of epinephrine, among other chemicals, through brain signals, the release of hormones, and the work of receptors.
Conclusion: Complex systems in the nervous system regulate neurotransmitters like epinephrine through the involvement of stress pathways, including but not limited to the hypothalamic-pituitary-adrenal (HPA) axis and adrenergic receptors. This is how the body copes with stress, whether it's increased focus and resilience or beating the challenge at hand, like academic pressures. This review suggests new avenues for researchers to explore in therapeutic intervention against the determinants of chronic stress.

Keywords: neurological response to academics, pressure, and transmitter release

INTRODUCTION

Stress is a crucial part of the academic journey, affecting both the body and the mind. For example, according to studies, stress can alter eating habits: some people eat more while others eat less when they're under stress [1]. Another manner stress can be assessed is stress reappraisal. This concept can aid in helping people remain calm and be more academically successful

by balancing hormones like cortisol, which play an important part in the process [2]. Another big aspect in the managing of stress is the nervous system, which releases neurotransmitters like epinephrine. This helps the body respond to stress and other stimuli. Students can explore methods to handle stress and succeed in life by learning how stress can affect both the body and mind [3].

METHODS

This review considered studies that evaluated the effects of academic stress on mental health, eating habits, and academic performance. Articles had to be about stress-induced eating behaviors like becoming a stress-eater versus stress-undereater, stress reappraisal techniques, or how stress influences brain function. Most studies selected included participants of varying ages, genders, and ethnicities. Results had to exclude any studies published before 2016, samples solely comprising clinical populations, and case studies not involving the collection of experimental data.

The studies were identified from databases like PubMed, JSTOR, and Google Scholar, utilizing keywords like "academic stress," "stress physiology," and "brain function." Eight articles were selected to review how academic stress may affect eating habits, and how neurotransmitters like epinephrine play a role in academic pressure and performance.

RESULTS

Key Findings

All the studies bring into perspective how neurotransmitter systems control complex functions of both the brain and the body, bringing awareness regarding stress, addiction, and heart rate control. Bringing attention to guanfacine and α2A-adrenergic receptor agonists highlights the ability of this agent to decrease stress reactivity and improve cognitive control for smoking cessation – a role it can play in addiction and symptoms of withdrawal [4]. Adding on, corticotropin-releasing factor studies have said that serotonin release has been limited in the stratum and lateral septum by this peptide. The role of CRF is significant when it comes to stress regulation and pharmacologically addressing depression and anxiety [5].

Modification of the actions of neurotransmitters with lower basal levels of dopamine and norepinephrine in the striatum is seen with the exercise

stimulus in endurance training. This further proves the brain's plasticity [6]. α2A and α2C adrenergic receptors have been further looked into regarding their part in the regulation of norepinephrine release because if these receptors were absent, cardiovascular dysfunctions like cardiac hypertrophy will develop [7].

Comparative Analysis

All of the articles show the regulation of neurotransmitters like epinephrine in different places and with shared functions but also differing priorities [4-8]. It's specified by these studies that neurotransmitters like serotonin, dopamine, and norepinephrine are key to brain and bodily mechanisms, especially when under stress, exercise, and pharmacological interference [4-8]. By looking into the stress pathway and studying CRF and serotonin and guanfacine research [4,5]. or through incorporating the adrenergic receptor regulation into cardiovascular health and neuronal modifications to endurance training [6,7].

The most apparent differences concern the neurotransmitter systems involved in targeted applications. In this case, the studies on CRF focus on serotonin's role in stress-induced psychiatric disorders, while adrenergic receptor research underlines the role of norepinephrine in cardiovascular functions [5,7]. While the guanfacine study bridges stress and addiction by targeting prefrontal cognitive control, the endurance training study links neurotransmitter adaptations to physical activity and behavioral resilience [4,6]. Most of these studies methodologically employ animal models and receptor-specific manipulations.

This compilation of research studies highlights the regulating of neurotransmitters, going into molecular realization as well as translational implementation into the mental, physical, and cardiovascular aspects of health. This provides an improved comprehension of neurotransmitter functions to explain the interactions between stress, neural plasticity, and the regulation of systemic physiology. The connection between these different domains shows how if one area is affected, the other interconnected parts of human health may be affected as well. Overall, these studies focus on various possible paths for recent treatments, addressing the causes of dysregulation as well as relieving symptoms.

DISCUSSION

This review addresses the following question: Under conditions of academic stress, what modulates the nervous system to secrete hormones like epinephrine? The studies included in this review demonstrate how neurotransmitter release is regulated during stress through the hypothalamic-pituitary-adrenal activation axis, adrenergic receptors, and stress-related peptides like CRF. The ability of guanfacine to reduce stress-related neurotransmitter output further strengthens the assurance of pharmacological modulation in the treatment of stress. These discoveries give a basis for the understanding of how academic stress may lead to the activation of the same physiological processes as in other stressors.

New research also suggests that there may be shared mechanisms between neuronal function seen in exercise and academic stress. In a study looking into the regulation of neurotransmitters during exercise, lower levels of dopamine and norepinephrine at baseline following the training effectively allows α2C adrenergic receptors in the body to regulate and produce epinephrine rapid release responses during high activity and stressors [6]. The neuronal function of adrenergic receptors in activity indicates that the same mechanism could occur during academic stress [7]. It is believed that such systems assist in the processes of concentration and energy spending during times of academic stress like when taking an exam [8].

It's important to bring awareness to these mechanisms to create interventions that will improve the ability to cope with stress. Academic stress in the long run may lead to changes in neurotransmitter levels and have adverse effects on mental health. As such, it's important to focus on more research on ways of managing stress accordingly. Future research should focus on the effects of different stressors on other neurotransmitter levels including glutamate and GABA, which are the CNS's main excitatory and inhibitory neurotransmitters, respectively.

References

1. Emond, M., Ten Eycke, K., Kosmerly, S., Robinson, A. L., Stillar, A., & Van Blyderveen, S. (2016). The effect of academic stress and attachment stress on stress-eaters and stress-undereaters. Appetite, 100, 210–215. https://doi.org/10.1016/j.appet.2016.01.035

2. Jamieson, J. P., Black, A. E., Pelaia, L. E., Gravelding, H., Gordils, J., & Reis, H. T. (2022). Reappraising stress arousal improves affective, neuroendocrine, and academic performance outcomes in community college classrooms. Journal of experimental psychology. General, 151(1), 197–212. https://doi.org/10.1037/xge0000893
3. World Health Organization. (2022). Determinants of brain health. In Optimizing brain health across the life course: WHO position paper (pp. 27–50). World Health Organization. http://www.jstor.org/stable/resrep44218.9
4. McKee, S. A., Potenza, M. N., Kober, H., Sofuoglu, M., Arnsten, A. F., Picciotto, M. R., Weinberger, A. H., Ashare, R., & Sinha, R. (2015). A translational investigation targeting stress-reactivity and prefrontal cognitive control with guanfacine for smoking cessation. Journal of psychopharmacology (Oxford, England), 29(3), 300–311. https://doi.org/10.1177/0269881114562091
5. Regulation of Serotonin Release in the Lateral Septum and Striatum by Corticotropin-Releasing Factor Michelle L. Price, Irwin Lucki Journal of Neuroscience 15 April 2001, 21 (8) 2833-2841; DOI: 10.1523/JNEUROSCI.21-08-02833.2001
6. MEEUSEN, R., SMOLDERS, I., SARRE, S., DE MEIRLEIR, K., KEIZER, H., SERNEELS, M., EBINGER, G. and MICHOTTE, Y. (1997), Endurance training effects on neurotransmitter release in rat striatum: an in vivo microdialysis study. Acta Physiologica Scandinavica, 159: 335-341. https://doi.org/10.1046/j.1365-201X.1997.00118.x
7. Hein, L., Altman, J. & Kobilka, B. Two functionally distinct α2-adrenergic receptors regulate sympathetic neurotransmission. Nature 402, 181–184 (1999). https://doi.org/10.1038/46040
8. Goldstein, D.S. Neurotransmitters and stress. Biofeedback and Self-Regulation 15, 243–271 (1990). https://doi.org/10.1007/BF01011108

Jai Degala

Understanding the Causes of Stress-Induced Cardiomyopathy Using Bioinformatics

ABSTRACT

Background: Stress-induced cardiomyopathy is a heart disease that is caused by stress. The main target group for this disease is post-menopausal women. It causes the left ventricle to balloon, which mimics symptoms of a heart attack including chest pain, shortness of breath, and fatigue. The problem with this disease is that we don't know much about the causes on a genetic level. Therefore, the goal of this research is to find genes that are highly expressed in Stress-induced cardiomyopathy.

Methods:In this study, GSE95368 human sample dataset from the NCBI GEO2R bioinformatics database was collected and analyzed with GEO2R to identify differentially expressed genes (DEGs). The samples were grouped into three; those infected with stress-induced cardiomyopathy, with Acute myocardial infarction, and from healthy human samples. The top 30 DEGs were then analyzed using GO and KEGG pathway enrichment analysis from SR Plot bioinformatics tool to determine their potential functions.

Results: We identified 30 top DEGs based on p-value and fold change split into 15 upregulated and 15 downregulated genes. The GO analysis showed high enrichment in the biological processes ontology with maintenance of location and plasma lipoprotein particle organization being prominent functions. The GO showed high enrichment in molecular functions with heparin and glycosaminoglycan binding as the highest functions. The KEGG results showed that cholesterol metabolism, coagulation, and complement cascade pathways were the most important pathways. Common genes found in both results were the APOE, F2, and APOB genes.

Conclusion:The study revealed that the top three genes related to Stress-Induced Cardiomyopathy are APOE, F2, and APOB. Further observations of these genes in a laboratory setting can be used to determine potential early diagnosis of Stress-Induced Cardiomyopathy.

Keywords: NCBI GEO2R, Stress-induced cardiomyopathy, SR Plot, KEGG, Gene Ontology, APOE, F2, and APOB

INTRODUCTION

Stress-induced cardiomyopathy is a heart disease that is caused by stress. It happens when someone is stressed and their left ventricle begins to weaken, which makes it harder for the heart to pump blood (1). The effects are very similar to a heart attack including chest pain, shortness of breath, and fatigue (2). However, unlike a heart attack instead of clogged arteries, the left ventricle will balloon out. This disease is important to learn about because it is potentially life-threatening and unknown due to its rarity.

The main problem of Stress-Induced Cardiomyopathy is understanding the causes of how it can affect certain people (1). We want to understand better what is specifically responsible for causing the disease to know who is at higher risk of getting it. The scientific question being investigated is 'What types of genes are responsible for causing Stress-Induced Cardiomyopathy? For example, are these genes involved in stress?'

So far it is known that Stress-Induced Cardiomyopathy is an emotionally triggered heart disease. It is also known that family history and catecholamine (stress hormone) signals can play a factor in the disease (3). The main demographic for this disease is postmenopausal women (4). Most cases of Stress-Induced Cardiomyopathy are typically benign, but there have been cases of death. However, a patient with Stress-Induced Cardiomyopathy will most likely return to their normal self within about a few weeks or months with the help of beta-blockers to reduce stress hormones (5).

In bioinformatics, it is known that genes that are responsible for stress are involved in causing this disease (5). The challenge with Stress-Induced Cardiomyopathy is finding out the specific genes that are responsible for causing the disease, or how they can be triggered. Also, there is not a lot of bioinformatics research on the disease due to its rarity and unknown causes.

The goal of this research is to understand which genes are important in causing Stress-Induced Cardiomyopathy and how they are activated. I hypothesize that the genes that are important in causing Stress-Induced Cardiomyopathy are commonly activated when someone feels stress.

This research is important because it can help other people understand what Stress-Induced Cardiomyopathy is, and how it can be prevented or diagnosed before any damage is to occur.

METHODS

This scoping literature search reviewed studies examining CVD's relation to educational attainment. The search engines used were PubMed and Google Scholar. The key search terms used included CVD, education, social determinants of health, and impact. The articles were included if they reported the level of education as a variable, studied mediation factors, and if they provided helpful information on how levels of education relate to CVD risk in general. The articles were excluded if they were not original research from credible, peer reviewed sources.

RESULTS

Data Collection and Analysis of GEO2R Data

NCBI is the National Center for Biotechnology Information that contains a great amount of biological information (6). A research paper, found on NCBI, focused on Stress-Induced Cardiomyopathy and its comparisons to healthy patients and patients with Acute Myocardial Infarction. This research paper included genes that were able to be analyzed by GEO2R. GEO2R is a tool within NCBI that is used to help visualize genetic information (6). The dataset collected was GSE95368 and it was analyzed in GEO2R by first defining these three groups: Stress-Induced Cardiomyopathy, Acute Myocardial Infarction, and Healthy (Figure 1).

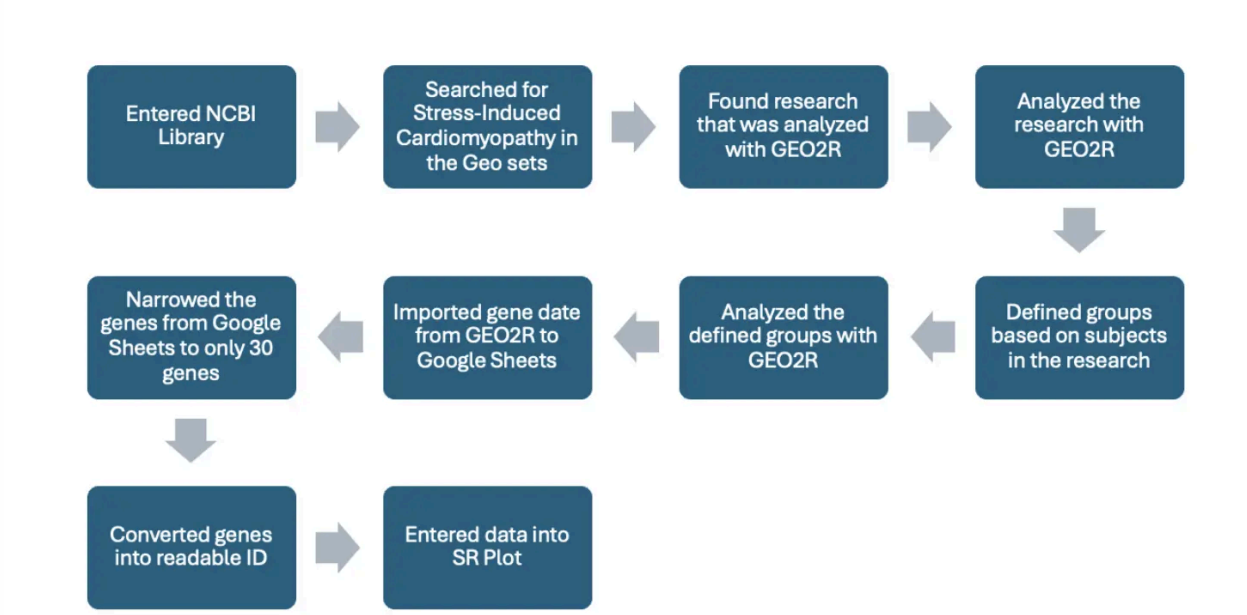


Figure 1. Research Methodology: The steps and bioinformatics used in this study.

Identification of the Top Differentially Expressed Genes

From the GEO2R results, 1225 genes were observed to be differentially expressed. To identify the 30 most significant differentially expressed genes, statistical analysis was applied. This process identified different pValues and Fold changes to prioritize the most important genes based on their differential expression across samples. The pValues were narrowed down by finding genes that had a pValue of less than 0.05. The fold changes were narrowed by finding the top 15 highest fold changes and the bottom 15 lowest fold changes.

Data Analysis Using SRPlot, KEGG, and GO Bioinformatics Tools

Then SRPlot, KEGG, and GO bioinformatics tools and databases were utilized to analyze the potential functions of these top genes. SR Plot is a bioinformatics tool that is used to create graphs to analyze gene pathways (7). SR Plot divides these graphs into two different results, KEGG and GO. KEGG is a bioinformatics tool that collects and organizes information on genes and their processes (8). GO is another bioinformatics tool that organizes genes by their function (9). Altogether, these bioinformatics tools

helped uncover the potential roles of the genes in their functions and the different pathways that they might be involved in.

RESULTS

Identification of Differentially Expressed Genes

The first bioinformatics tool that was used to identify differentially expressed genes (DEGs)was from GEO2R on NCBI. From the first two results, some genes were expressed differently in healthy patients and those with Stress-Induced Cardiomyopathy (Figures 1 and 2).

In the volcano plot (Figure 2a) the red dots represent upregulated genes, meaning that they are much more expressed. The blue dots represent downregulated genes, meaning that they are down-expressed. The black dots represent genes that represent the same amount of expression in a healthy person and one with Stress-Induced Cardiomyopathy (Figure 2a).

In our study, the total number of genes that were differentially expressed was 1225 different genes (Figure 2b). The Venn diagram shows that between healthy, acute stress-induced cardiomyopathy and acute myocardial infarction, both diseases do not share many of the same genes, but there is an overlap of 44 between the Acute Stress-Induced Cardiomyopathy vs Healthy group and the Acute Myocardial Infarction vs Healthy group (Figure 2b).

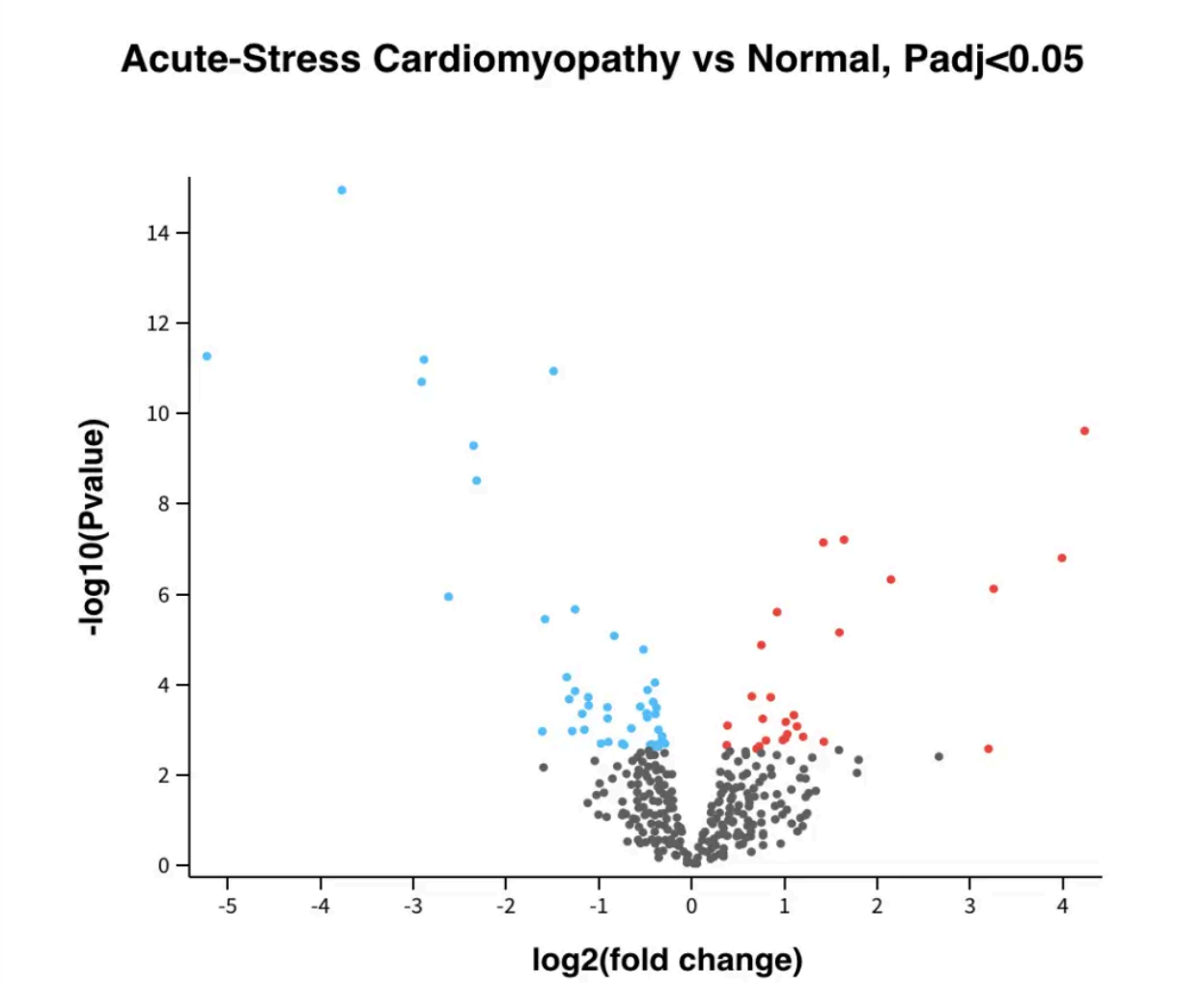


Figure 2a. The volcano plot shows how the genes are being expressed differently between Stress-Induced Cardiomyopathy and healthy individuals. The red dots represent upregulated genes, the blue dots represent down-regulated genes, and the black dots represent an equal amount of expression.

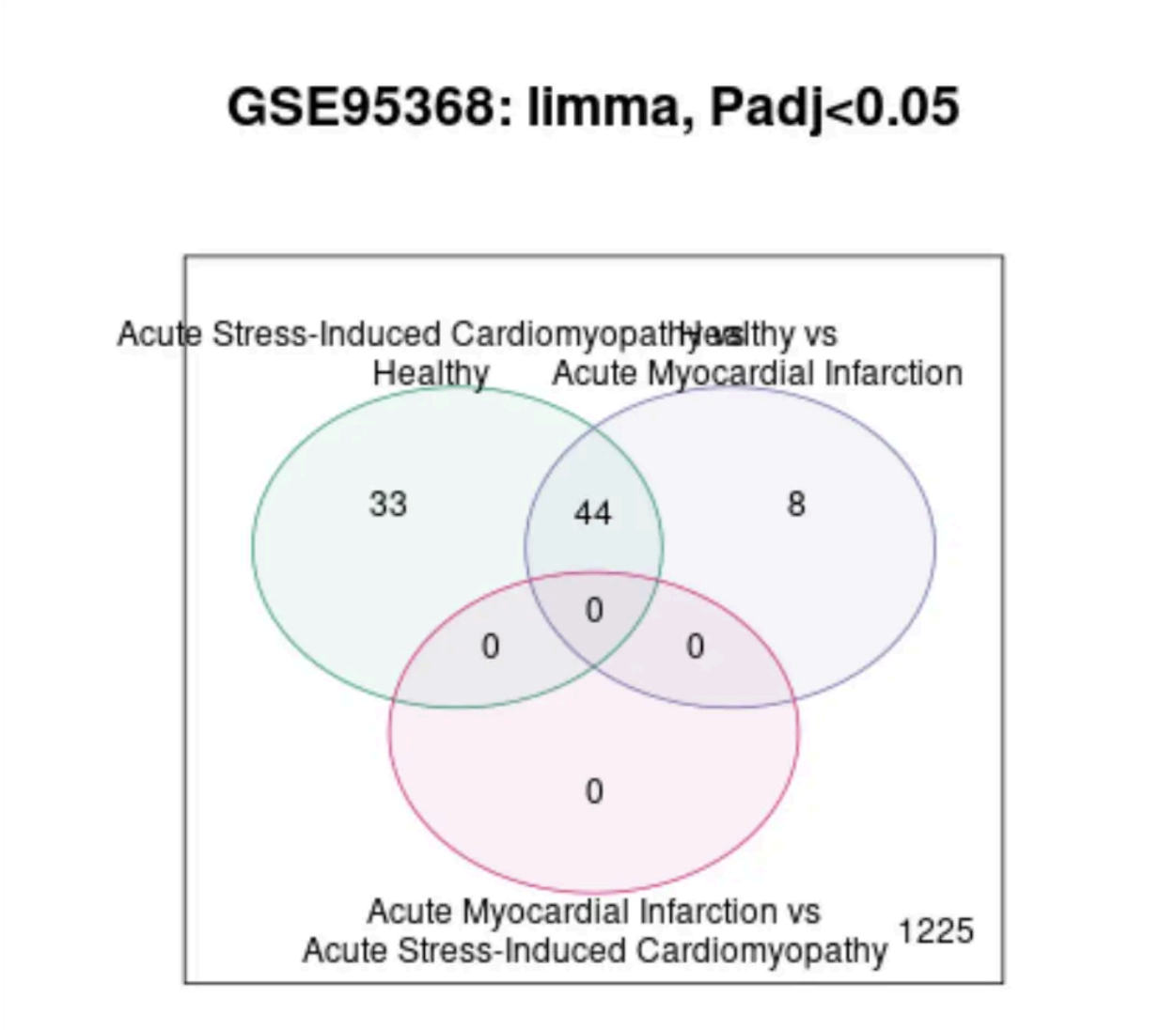


Figure 2b. The Venn diagram shows the genes that are shared between Stress-Induced Cardiomyopathy, Acute Myocardial Infarction, and healthy patients. Out of 1225 genes that are expressed there is an overlap of genes within the Stress-Induced Cardiomyopathy vs Healthy group and the Healthy vs Acute Myocardial Infarction group.

Identification of Top 30 Statistically Significant Differentially Expressed Genes (DEGs)

The genes that were downloaded from the study were narrowed down based on the p-value and fold change. First, they were selected by finding genes that had a p-value lower than 0.05. Then they were narrowed further by finding the lowest and highest fold changes Genes with a log2 fold change (|FC|) greater than 1 were classified as differentially expressed. In the end,

the genes were narrowed to only 30. These were based on the top 15 highest logFC and the bottom 15 lowest logFC. There were 15 upregulated genes and 15 downregulated genes after the statistical analysis (Top 30 Genes).

Potential Functions and Enrichment of the Identified Genes and Pathways

In the KEGG results from SR Plot, the complement and coagulation cascade pathway, and cholesterol metabolism were both identified as significant in the results (Figure 3b).. The complement and coagulation cascade pathway and cholesterol metabolism had the highest enrichment scores. From the KEGG results, one gene that stood out were the ApoE and ApoB-100 genes that were found in the complement and coagulation cascades pathway (Figure 3a). Some other genes that stood out were Serpin, F2, PAI, PCI, and C3 which were found in the cholesterol metabolism pathway (Figure 3b).

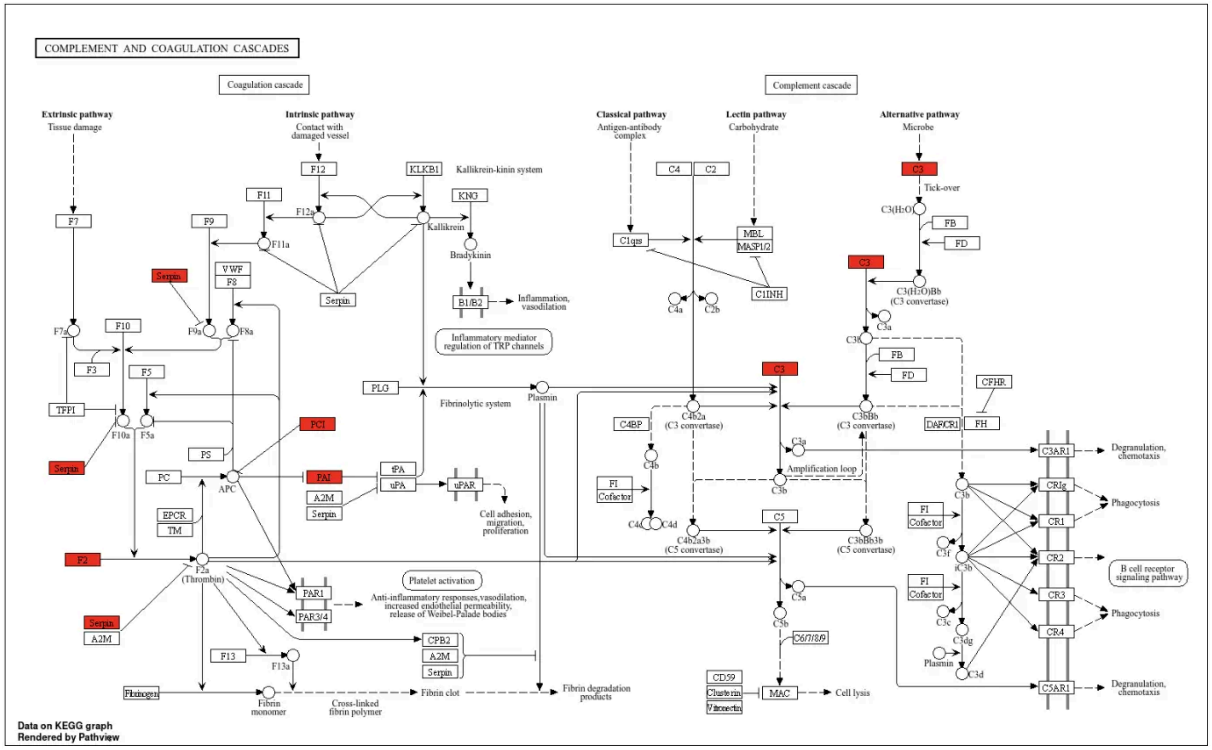


Figure 3a. This image represents the complement and coagulation cascade pathway and the genes significantly involved in the KEGG result in red. Significant genes include Serpin, C3, PAI, and F2.

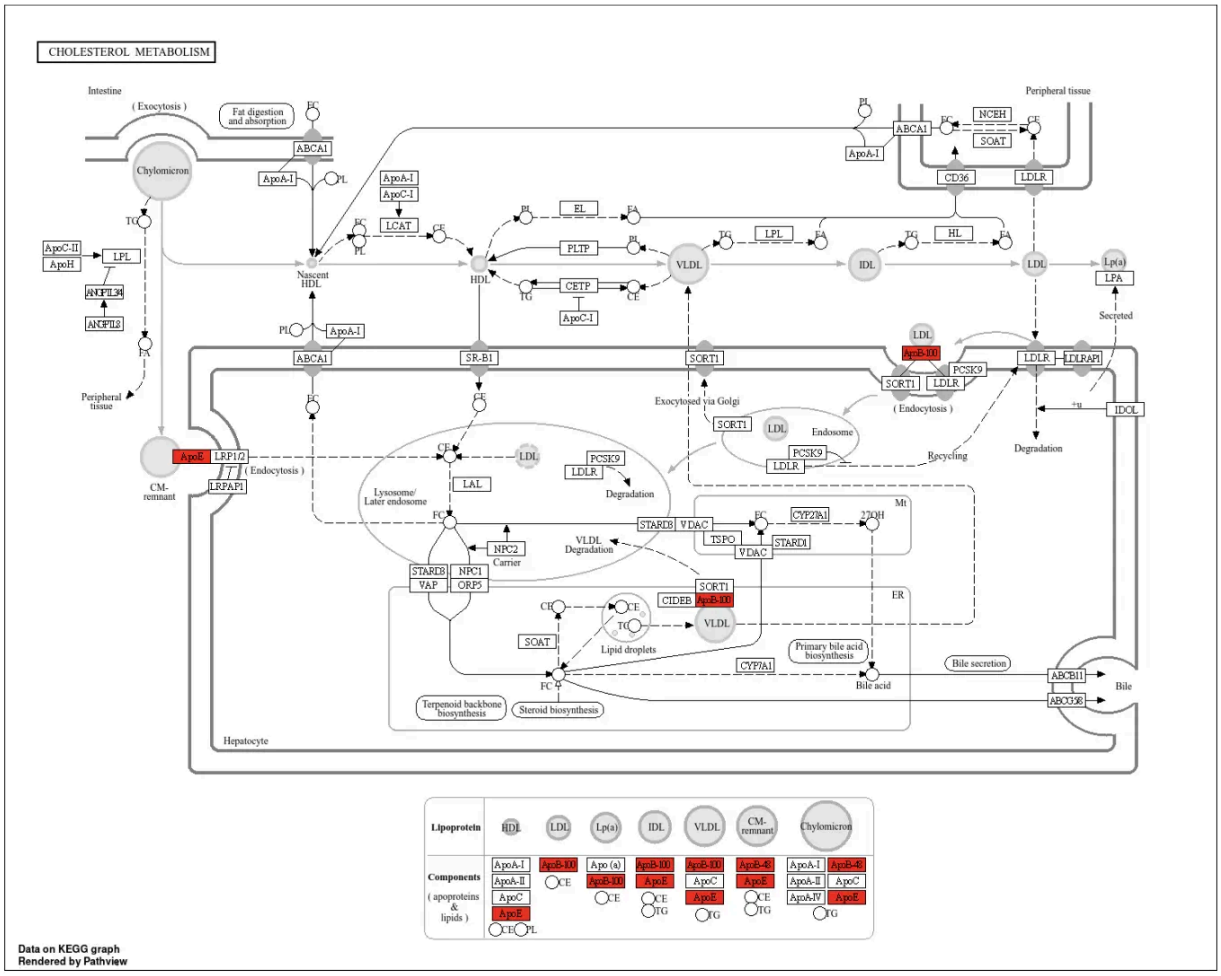


Figure 3b. This image represents the Cholesterol metabolism pathway and the genes significantly involved in the KEGG result in red. Significant genes include ApoB-100, ApoE, and ApoB-48

The Gene Ontology (GO) is separated into three ontologies: biological processes, molecular functions, and cellular components. The results showed that all of the ontologies were relatively equal in enrichment. However, some pathways with a higher enrichment score were maintenance of location and plasma lipoprotein particle organization, which is found in the biological processes ontology; as well as heparin and glycosaminoglycan binding, found in molecular functions (Figure 4a).

From the GO results, some genes that stood out in the multicellular function were the SFRP1, APOE, SERPINA5, SAA1, F2, and APOB genes. Some genes that stood out in the biological processes were the APOE, CRP, ALB, PRKACA,

F2, APOB, and C3 genes (Figure 4b). Common genes that stood out in both MP and BP ontology were the APOE, F2, and APOB genes.

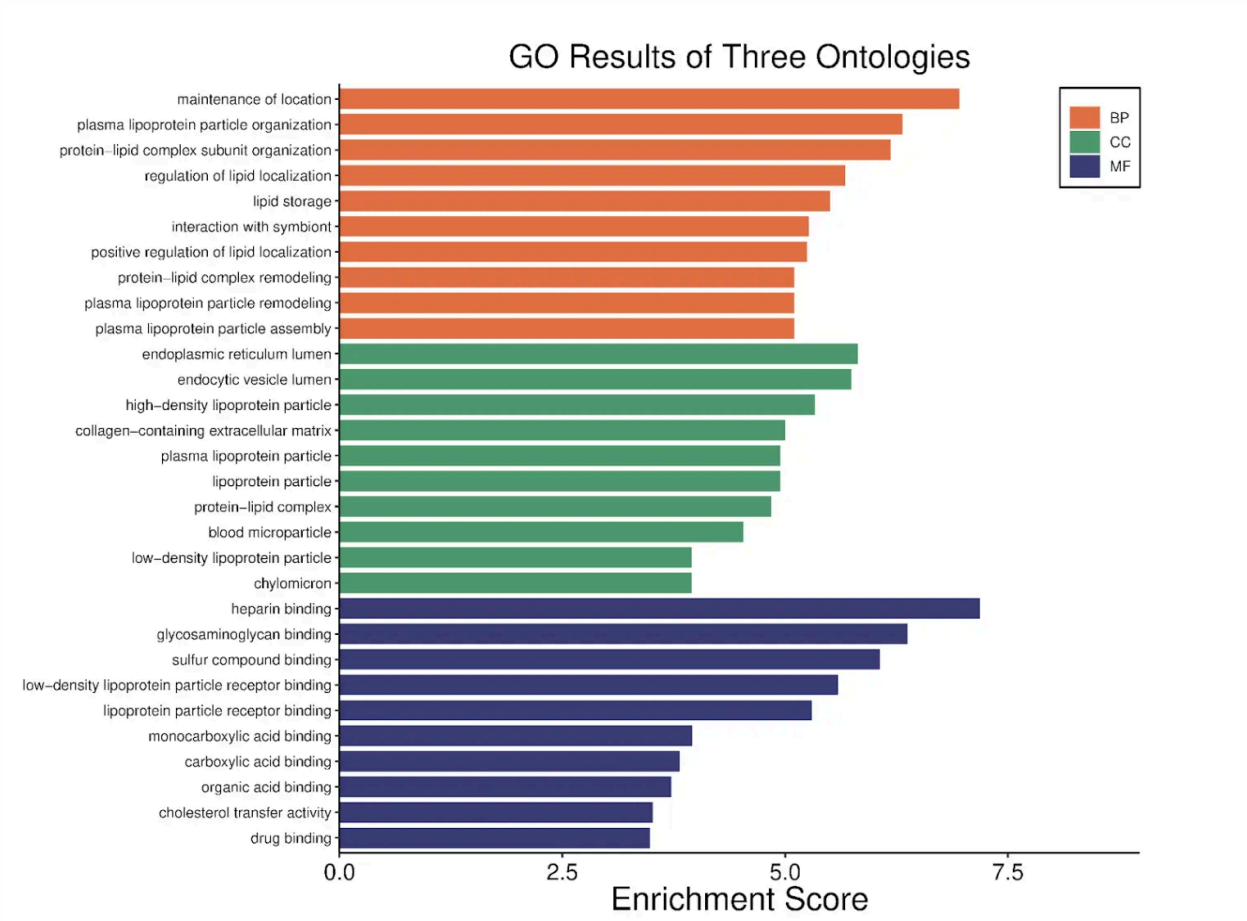


Figure 4a. This image represents the enrichment in all the ontologies from the GO results. BP stands for biological processes, CC stands for cellular components, and MF stands for molecular function. Four pathways were selected based on their high enrichment score. In BP, the maintenance of location and plasma lipoprotein particle organization had high enrichment scores. In MF, the heparin and glycosaminoglycan bindings both had high enrichment scores as well. These processes were further analyzed with the genes due to the fact they had the highest enrichment out of all the results.

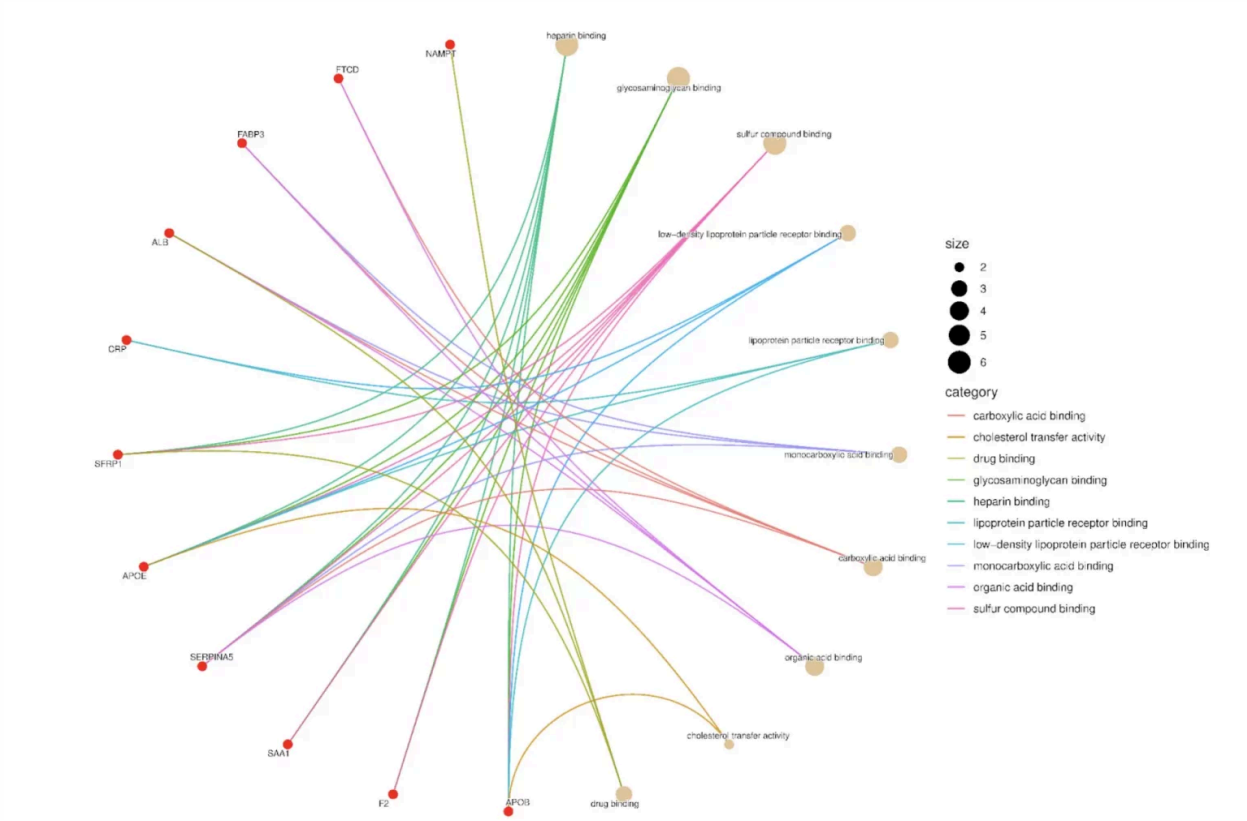


Figure 4b. This is a Cnet plot that shows which genes are related to certain pathways. We are looking for genes that are related to pathways that have a high enrichment score. APOE, F2, and APOB were genes that were related to glycosaminoglycan binding and heparin binding. These genes were specifically selected from other genes also related to glycosaminoglycan binding and heparin binding because they had similarities with the KEGG results.

DISCUSSION

Summary of Findings

The goal of this research study was to identify genes that were significant in Stress-Induced Cardiomyopathy to understand the causes of the disease better. In the study, we found that complement and coagulation cascade pathways and cholesterol metabolism pathways were found to be significant. Also, the genes were all equally enriched in all three ontologies, but we found that MF and BP had the highest enrichments (Figure 4a). The top three significant genes that were prominent in both the KEGG and GO results were the APOE, F2, and APOB genes.

Interpretation of Results

From the KEGG results, we found that coagulation complement cascades and cholesterol metabolism are very important factors. This can help show that in addition to stress, high cholesterol, and blood clotting can also be important factors in causing Stress-Induced Cardiomyopathy. The analysis also found that the highest enriched ontologies were in MF and BP, which helps us know that causes can most likely be related to biological processes and molecular functions. Out of the genes that were present in the KEGG and GO results, the APOE, F2, AND APOB genes were significant in both results. This could mean that these genes are the most significant in Stress-Induced Cardiomyopathy. We can use these genes to understand their activation, and how it can prevent Stress-Induced Cardiomyopathy.

Comparison with Previous Studies

In other studies, patients with Stress-Induced Cardiomyopathy have been found to have higher cholesterol levels (10). They believe that cholesterol could also be an influence in the disease. Genes that are related in regulating cholesterol metabolism are the APOE and APOB gene (11-12). Coagulation and complement cascades are often involved with blood clotting. In another study patients with Stress-Induced Cardiomyopathy were found to have significantly higher levels of coagulation and complements than a control group of healthy individuals (2). A gene that is associated with blood clotting is the F2 gene because it creates prothrombin, which results in blood clots (13).

Implications

Given the information on the genes involved in the cholesterol metabolism pathway and the complement and coagulation cascade pathway, we can use this information to determine if these genes are directly related to Stress-Induced Cardiomyopathy. If they are then we can observe the gene expression to see who may be at risk.

Limitations

One limitation of this study is that it may be harder to observe these identified genes in someone who has Stress-Induced Cardiomyopathy because it is such a rare disease. Given its rarity, it may be hard to find patients with the disease to experiment with. Another limitation is that the

identified genes will need to be further studied in the laboratory or clinical environment before developing a cure.

Future Direction

The identified genes can be tested in the laboratory by scientists in the laboratory or clinical trials to determine if the genes are the most significant in causing Stress-Induced Cardiomyopathy and how they can be used for early diagnosis.

References

1. Takotsubo Syndrome in Intensive Cardiac Care Unit: Challenges in Diagnosis and Management. Secondary Research. Rocco Antonio Montone, Giulia La Vecchia, Marco Giuseppe Del Buono, Antonio Abbate, Tommaso Sanna, Daniela Pedicino, Giampaolo Niccoli, Massimo Antonelli, Filippo Crea. 2022. ScienceDirect. https://www.sciencedirect.com/science/article/abs/pii/S0146280621002929.
2. Elias Rawish, Thomas Stiermaier, Francesco Santoro, Natale D. Brunetti, and Ingo Eitel. Secondary Research. 2021. Journal of Cardiovascular Development and Disease. https://www.mdpi.com/2077-0383/10/3/479.
3. Stress Cardiomyopathy: Is it limited to Takotsubo syndrome? Problems of definition. Secondary Research. Petr A. Sarapultsev, Alexey P. Sarapultsev. 2016. ScieneDirect. https://www.sciencedirect.com/science/article/abs/pii/S0167527316313882.
4. Update of Takotsubo Cardiomyopathy: Present Experience and Outlook for the Future. Secondary Research. Anastasiia V. Bairashevskaia, Sofiya Y. Belogurova, Mikhail R. Kondratiuk, Daria S. Rudnova, Susanna S. Sologova, Olga I. Tereshkina, Esma I. Avakyan. 2022. ScienceDirect. https://www.sciencedirect.com/science/article/pii/S2352906722000392
5. The Challenge of Takotsubo Syndrome: Heterogeneity of Clinical Features. Secondary Research. SchlossbauerSusanne A.,GhadriJelena-Rima,ScherffFrank,TemplinChristian. 2017. Swiss

Medical Weekly. https://smw.ch/index.php/smw/article/view/2372/3627.
6. NCBI GEO: archive for functional genomics data sets, Secondary Research, Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, Alexandra Soboleva, 2012, Nucleic Acids Research https://academic.oup.com/nar/article/41/D1/D991/1067995
7. The KEGG Database, Secondary Research, Minoru Kanehisa, Gregory bock, Jamie A. Goode, 2002, Wiley Online Library https://onlinelibrary.wiley.com/doi/abs/10.1002/0470857897.ch8
8. The what, where, how and why of gene ontology – a primer for bioinformatics, Secondary Research, Louis du Plessis, Nives Škunca, Christophe Dessimoz 2011 Briefings in Bioinformatics https://academic.oup.com/bib/article/12/6/723/221815
9. SRplot: A free online platform for data visualization and graphing, Secondary Research Doudou Tang,Mingjie Chen,Xinhua Huang,Guicheng Zhang,Lin Zeng,Guangsen Zhang,Shangjie Wu,Yewei Wang, 2023, PLOS One https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0294236
10. Liu, W., Ruiz-Velasco, A., Wang, S. et al. Metabolic stress-induced cardiomyopathy is caused by mitochondrial dysfunction due to attenuated Erk5 signaling. Nat Commun 8, 494 (2017). https://doi.org/10.1038/s41467-017-00664-8
11. HAGBERG, J. M., WILUND, K. R., & FERRELL, R. E. (2000). APO E gene and gene-environment effects on plasma lipoprotein-lipid levels. Physiological Genomics. https://doi.org/10.1152/physiolgenomics.2000.4.2.101
12. National Library of Medicine. (n.d.). F2 gene: Coagulation factor II, thrombin. MedlinePlus. Retrieved March 2, 2025, from https://medlineplus.gov/genetics/gene/f2/
13. Amanda J Whitfield, P Hugh R Barrett, Frank M van Bockxmeer, John R Burnett, Lipid Disorders and Mutations in the APOB Gene, Clinical Chemistry, Volume 50, Issue 10, 1 October 2004, Pages 1725–1732, https://doi.org/10.1373/clinchem.2004.038026

Devraj Dutta

Using Bioinformatics to Investigate the Effect of Sepsis-Causing Bacteria on Gene Expression

ABSTRACT

Sepsis is a life threatening condition caused by an organism's overreaction to foreign invaders, resulting in serious medical complications such as organ failure. While sepsis research has been progressing, there remains an abundance of complications stopping scientists from fully understanding the disease. First of all, sepsis can be caused by a variety of microbes ranging from bacteria to virus to parasites. Additionally, gene expression data in sepsis patients has been known to be difficult to extract. Although sepsis can be treated on a patient-by patient basis, creating general treatments and cures have proven difficult and complicated. This study therefore aims to study a specific branch of sepsis caused by gram-positive bacteria such as Escherichia coli and Staphylococcus aureus. In this study, gene expression data is compared between infected sepsis patients and healthy donors to see what pathways and functions are expressed differently in patients suffering from bacterial sepsis. This knowledge can lay the groundwork for future studies that attempt to find treatment methods to bacterial sepsis that take advantage of the differentially expressed genes. Our results indicate that bacterial sepsis causes changes in expression of key genes, CMA1, CRHR2 and TUBB1/2 and that these genes are enriched or associated in functions and biological pathways of negative viral regulation, peptide bonding, and cytoskeletal structure. Therefore, scientists can potentially use these genes to further understand bacterial sepsis and potentially develop targeted, life-saving treatments to improve outcomes.

INTRODUCTION

Sepsis is an extremely dangerous medical condition which results from the body's overreaction to an infection which can lead to a loss of organ function. This can lead to life threatening complications such as shock, organ failure, and death (WHO 2024). Although sepsis can be caused by a variety of pathogens, the primary cause of sepsis is due to bacteria (WHO 2024). The goal of this paper is to investigate how bacteria cause sepsis in the body with regard to their impact on gene expression. Sepsis-causing bacteria have distinct toxins that they release, which many believe to be the source

of sepsis. Some of these toxins include endotoxins and superantigens (Sizar et. al 2023). The most common cause of sepsis is due to gram-positive bacteria such as Staphylococcus aureus (S. aureus), Streptococcus pyogenes (S. pyogenes), Klebsiella spp. Escherichia coli (E. coli), and Pseudomonas aeruginosa (P. aureginosa) (Ramachandran et. al 2014). Gram-positive bacteria are known for their thicker peptidoglycan cell wall and can be found through a special dying technique which leaves gram-positive bacteria blue (Sizar et. al 2023). Patients with sepsis are seen to have high levels of Heat Shock gene expression although those seem to be from the bacteria themselves rather than the host (Shao et. al 2023). Additionally, gene expression does appear to be different in patients with sepsis, but unfortunately the specifics remain unclear. It is quite difficult to study sepsis at a populational level (WHO 2024) as gene expression varies greatly from one sepsis patient to another. Moreover, gene expression changes at a rapid rate in patients with sepsis, making it hard to diagnose anything (Maslove et. al 2014).

Therefore, to address these challenges, the goal of this research was to see if bacteria causes sepsis through changes in gene activity. Since sepsis is described as an extreme reaction, the initial hypothesis was that the bacteria triggers too much gene activity or gene expression in the host organism, leading to this overreaction. This research is important because if the gene regulation behind bacterial sepsis is understood, better treatments can be created that target the specific genes affected by sepsis, potentially saving millions of lives.

METHODS

Obtaining Biological Data sets from NCBI Bioinformatics Database

There is currently an abundance of biological data, but much of it remains unanalyzed. The National Center for Biotechnology Information (NCBI), a part of the National Institute of Health, is one of many websites storing this data (https://www.ncbi.nlm.nih.gov/)., also known as NCBI, One of the main features of NCBI is its collections of databases and analysis tools which allows researchers to obtain and analyze data (Wojcikiewicz 2016). One such database, the Gene Expression Omnibus (GEO), contains experiments on gene expression using RNA or DNA biological molecules (Barret et al, 2009). GEO is freely accessible at http://www.ncbi.nlm.nih.gov/geo/ for public use

(Barret et al, 2009). In this research, the GEO dataset identified and collected was GSE210797, which can be found on its website (GEO). GSE210797 reports the gene transcription levels through blood samples of nonviral sepsis patients, in addition to those of healthy donors (Campbell et. al 2022). Gene transcription levels are then compared with one another to figure out how bacteria affect gene expression levels in non viral sepsis.

Analyzing data with GEO2R Bioinformatics Tool

The gene expression omnibus is a database belonging to the NCBI that contains microarray data regarding gene expression which can further be analyzed by the GEO2R bioinformatics tool (Barret et al, 2009; Mohammad 2022). Instead of manually writing codes or programming, GEO2R uses artificial intelligence and machine learning algorithms using the R programming language to help users find differentially expressed genes within the data (Milian 2024). Using the dataset GSE210797, the samples were organized into two categories: infected patients and healthy donors. The data was then analyzed with GEO2R to see trends in the data. The algorithms used to generate the dataset GSE210797 dataset identified and used in this study are indicated in the supplementary results (R Script). The greatest increases and decreases in gene expression were recorded in scatter and box plots to examine any distinct clusters forming between the two groups (Figure 1).

Statistically Sorting and Selecting Differentially Expressed Genes

The information given from GEO2R collected around 15000 differentially expressed genes (DEGs) (Figure 3), which was then screened and limited to around 40 genes. First, insignificant data was removed by p-value. When setting a threshold of $p \leq 0.01$ the gene pool was narrowed to a few hundreds. Afterwards, the top 20 up and downregulated genes are selected based on the log2fold change ($|\log2(\text{fold change})|>1$), giving 40 genes to work with (Supplementary Results: Top DEGs)

Finding Biological Functions and Pathway Trends Between Differentially Expressed Genes With SR Plot

To determine which functions and biological pathways are linked to certain genes, the data was analyzed using SR Plot, another bioinformatics tool (https://www.bioinformatics.com.cn/en). Gene Ontology (GO) terms (Ashburner et al, 2000) and KEGG bioinformatics databases were used to

locate gene function and pathways. GO terms include biological process (BP), cellular component (CC), and molecular function (MF) (Ashburner et al, 2000). Using SR Plot (https://www.bioinformatics.com.cn/en), gene data from GEO2R was analyzed to see what similarities exist between the differentially expressed genes. In this present study, the top differentially expressed genes (DEGs) were chosen and analyzed using the SR Plot bioinformatics database to perform GO and KEGG pathway enrichment analysis. Figure 1 summarizes the workflow of the methods used in this research study.

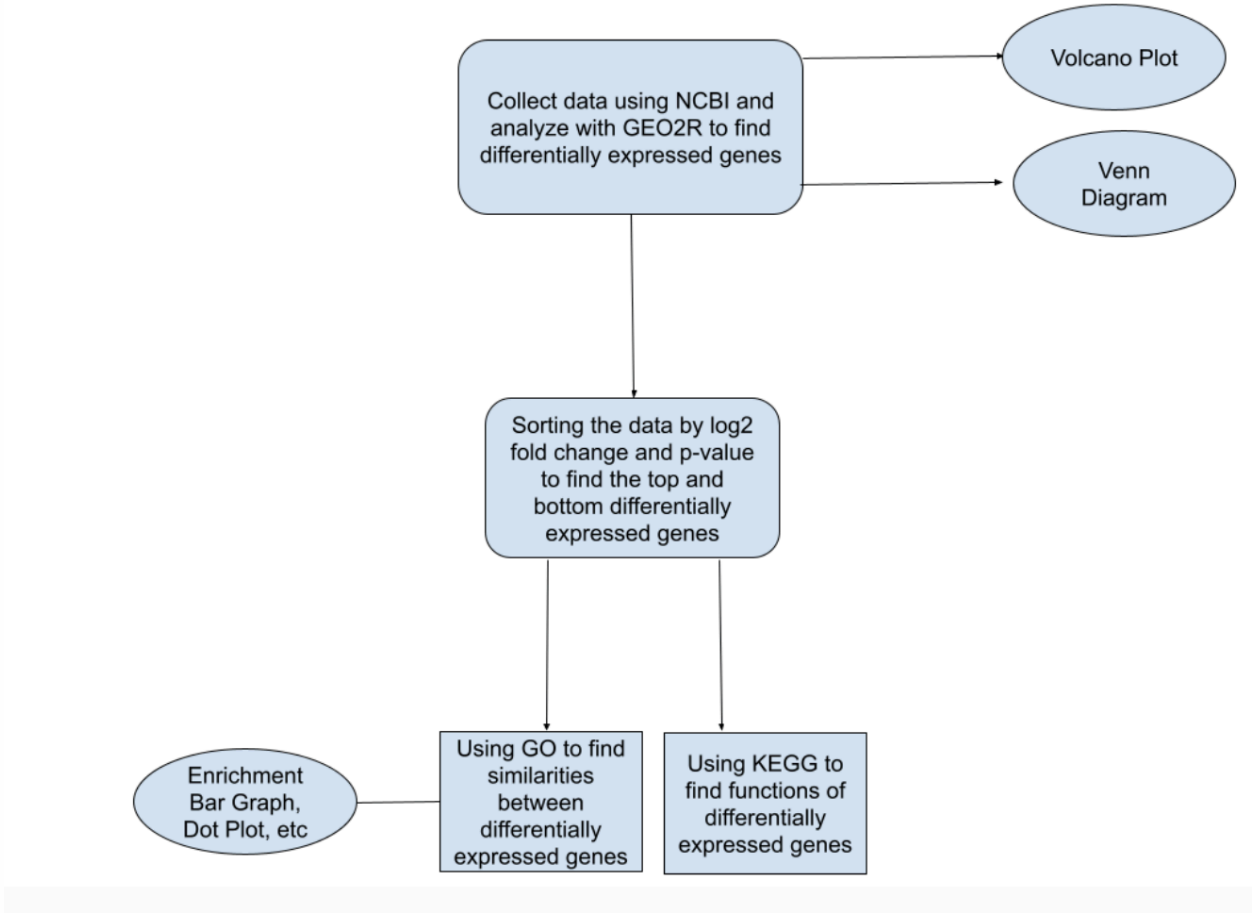


Figure 1. Research Outline: General outline of methods used throughout this research study. First, NCBI and GEO2R were used to analyze differentially expressed genes with non-viral sepsis. Afterwards, the data was sorted, mainly based on p-value ≤ 0.01, but also keeping in mind the log2 fold change change (log2(fold change)|>1), in order to find the most significant genes expressed. Afterwards, SR plot was used to conduct a gene

ontology analysis and found trends regarding biological process, molecular function, and cellular component between the differentially expressed genes. At the same time, KEGG bioinformatics database was used to find pathways shared between the differentially expressed genes.

RESULTS

Top Differentially Expressed Gene (DEG) Identification

A differentially expressed gene (DEG) represents a gene whose observed change in expression is deemed significant enough based on the gene p-value (Anjum et. al 2016). GEO2R accesses the gene expression data and finds the differentially expressed genes using a volcano plot (Figure 2) and venn diagram (Figure 3). Afterwards, the top up and down DEGs were identified using a p-value threshold of 0.01 and |log2(fold change)|>1 (Supplementary Results: Top DEGs).

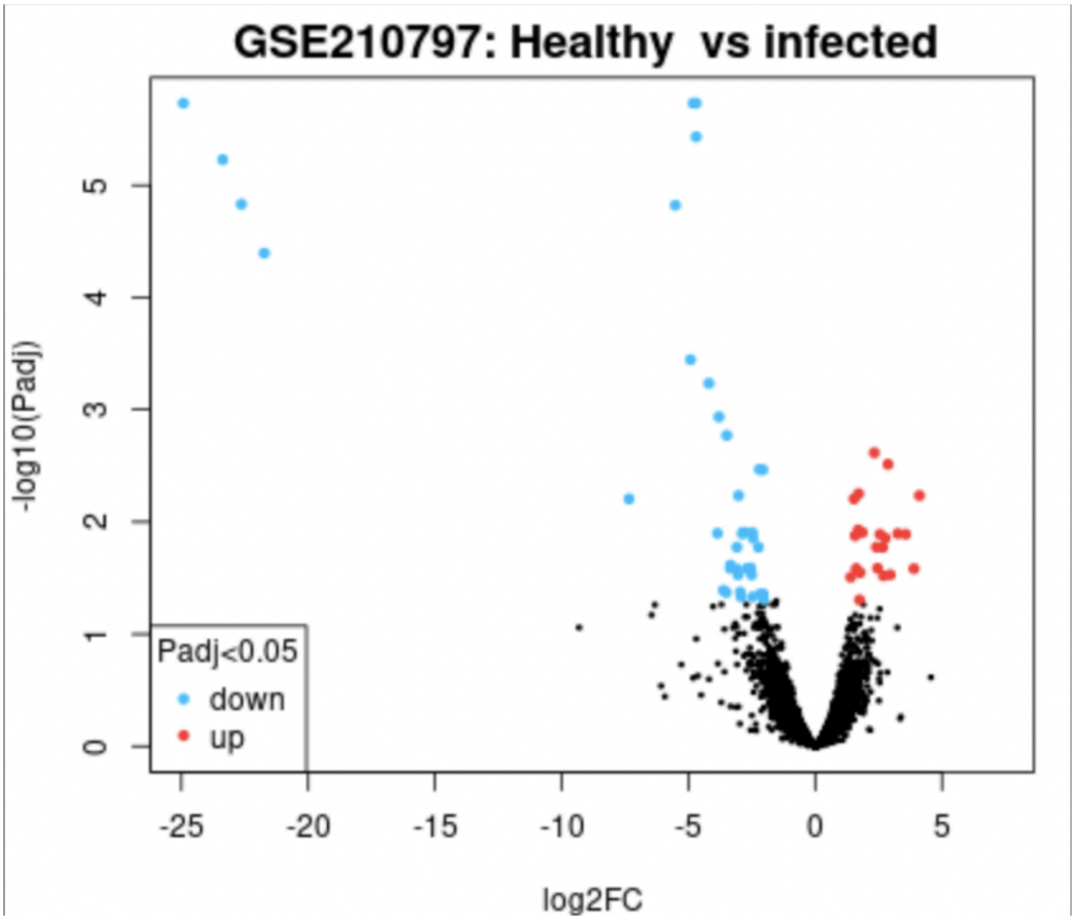


Figure 2. Volcano Plot of Healthy vs Infected Patients: GEO2R was used to analyze the genes affected by non viral sepsis. In the red are the genes

whose expression increased in patients with non-viral sepsis, and in the blue are in genes whose expression is decreased in patients with non-viral sepsis. Black dots show genes that had no difference in expression between patients and healthy individuals. As shown above, there are around 60 genes that are either up or downregulated.

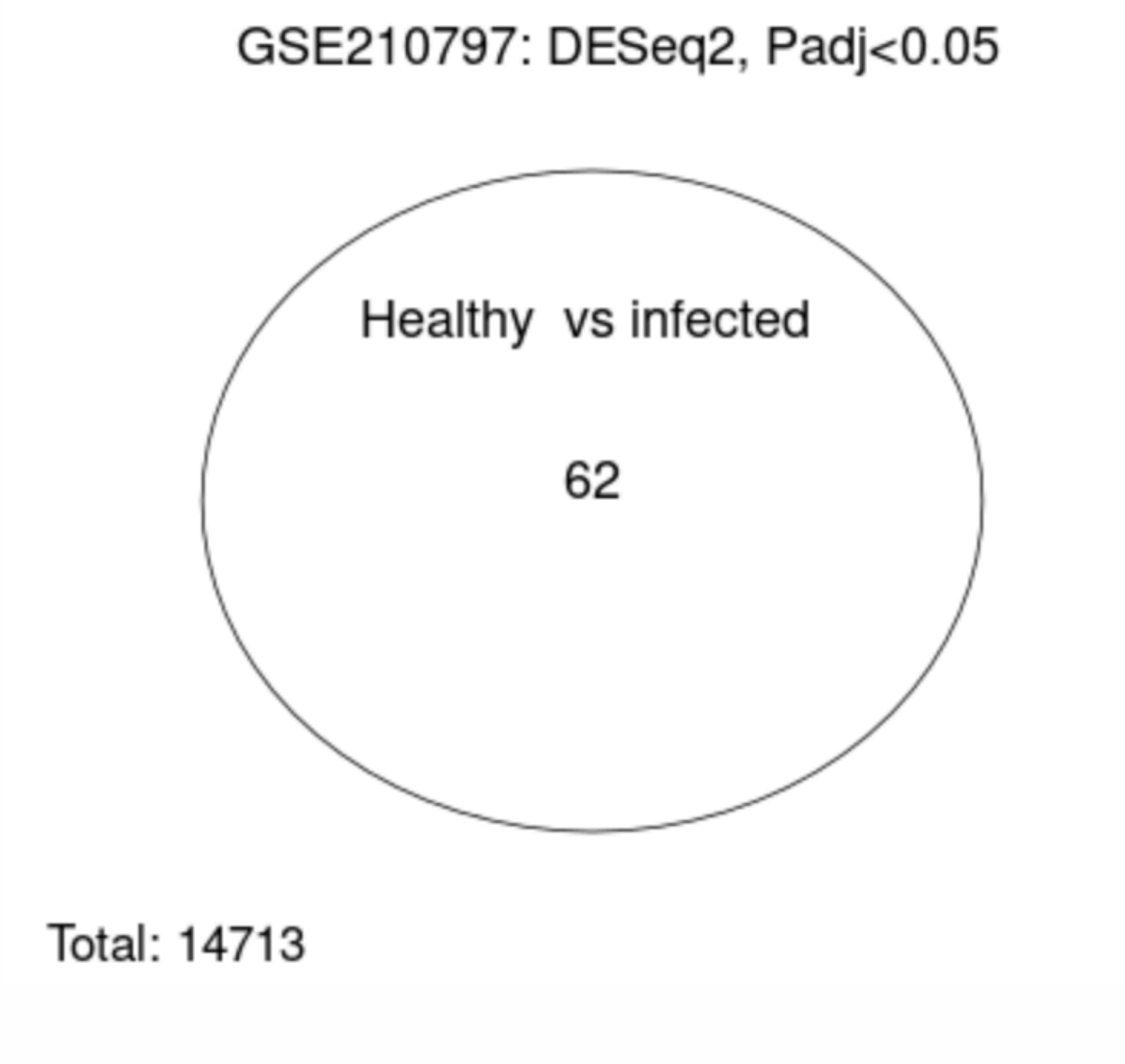


Figure 3. Venn Diagram of Healthy vs Infected Patients: The total number of genes analyzed by GEO2R was 14713. But between the healthy and infected patients, the number of genes whose expression is different between infected and healthy patients was 62 significantly affected genes in patients with non viral sepsis.

SR Plot was further used in order to visualize the differences in gene expression in terms of log fold change for the top DEGs (Figure 4). The results of the top DEGs highlight the top genes with the most significant increase and decrease in expression, measured by log2 fold change. Results here show that the genes TMSB4Y, EIFIAY, RPS4YI, and UTY exhibit the highest levels of gene expression (Figure 4).

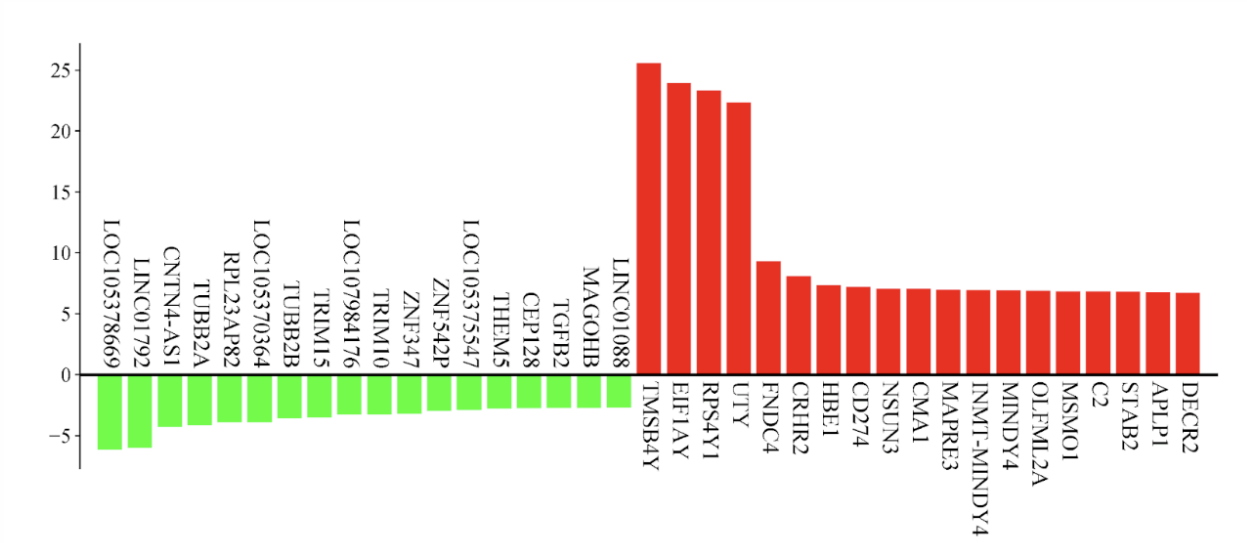


Figure 4. Up and Downregulated Genes: This figure shows the top genes with the greatest increase in expression and genes with reduced expression according to log2 fold change. In the red are the genes whose expression increased in patients with non-viral sepsis, and in the green are the genes whose expression is decreased in patients with non-viral sepsis.The main genes to focus on here would be TMSB4Y, EIFIAY, RPS4YI, and UTY because they have the greatest gene expression.

Identification of Potential Functions of Top Genes using Gene Ontology Bioinformatics Database

After identifying the top DEGs (Supplementary Results: Top DEGs), the data was analyzed using the gene ontology (GO) bioinformatics database on SR Plot bioinformatics tool (https://www.bioinformatics.com.cn/enin) to find the specific functions or pathways that are enriched or associated with the DEGs as well as the most important genes to analyze. First, the data was sorted by Gene Ontology (GO) terms (Ashburner et al., 2000) bioinformatics database to see whether bacterial sepsis targeted biological process (BP),

cellular component (CC), and/or molecular function (MF) (Ashburner et al., 2000).Although all three categories had somewhat similar enrichment scores (Figure 4), the main focus seemed to be on BP and MF. Afterwards, the functions and genes regulated were studied in regards to BP and MF using dot plots and CNet charts (Figure 5, 6 and 7).

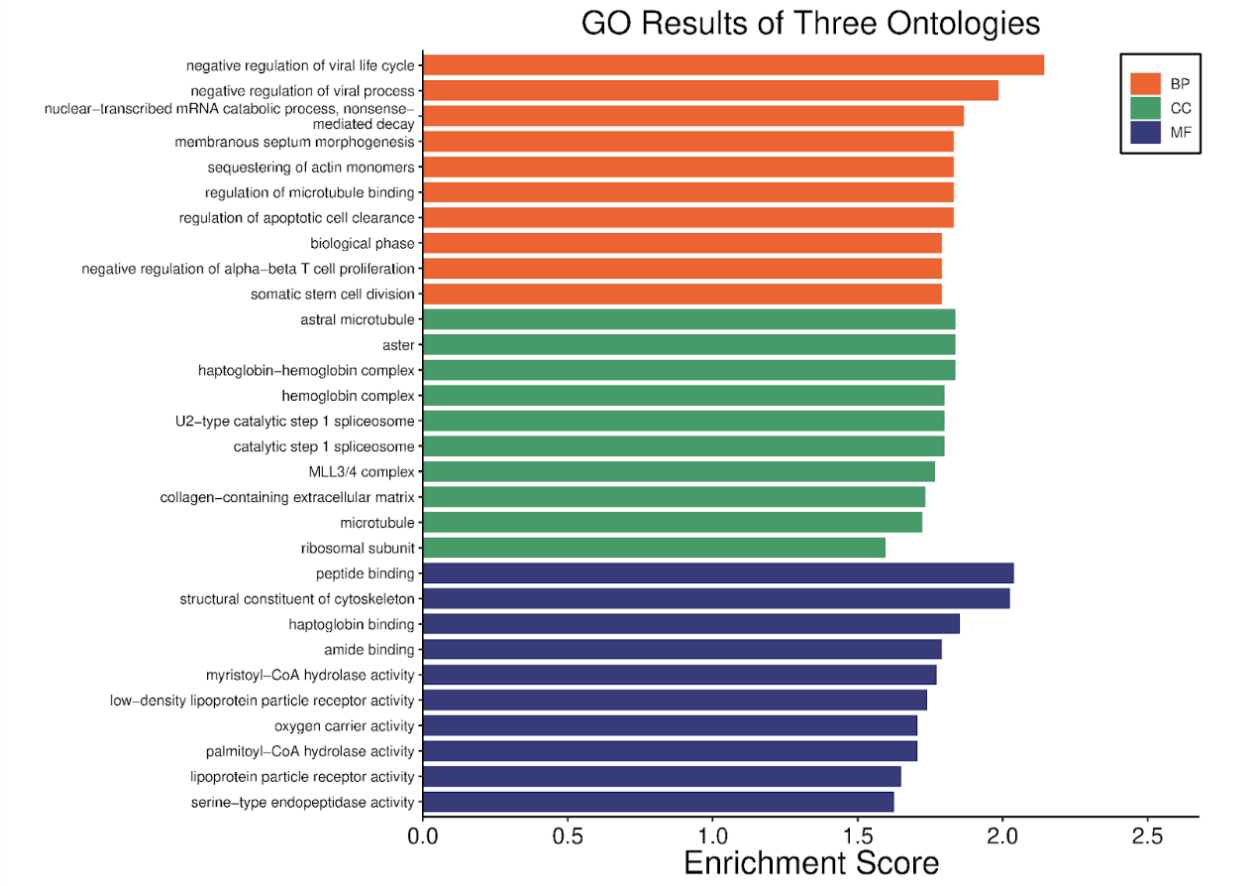


Figure 4. Gene Enrichment Chart: This figure shows the enrichment scores of the top and bottom regulated genes. In the orange are the biological processes results, in the green are the cellular component results, and in the blue are the molecular function results. The above figure shows that sepsis causing bacteria has effects mainly on biological processes and molecular function with the main areas of enrichment being negative regulation of viruses, peptide binding, and cytoskeletal structure.

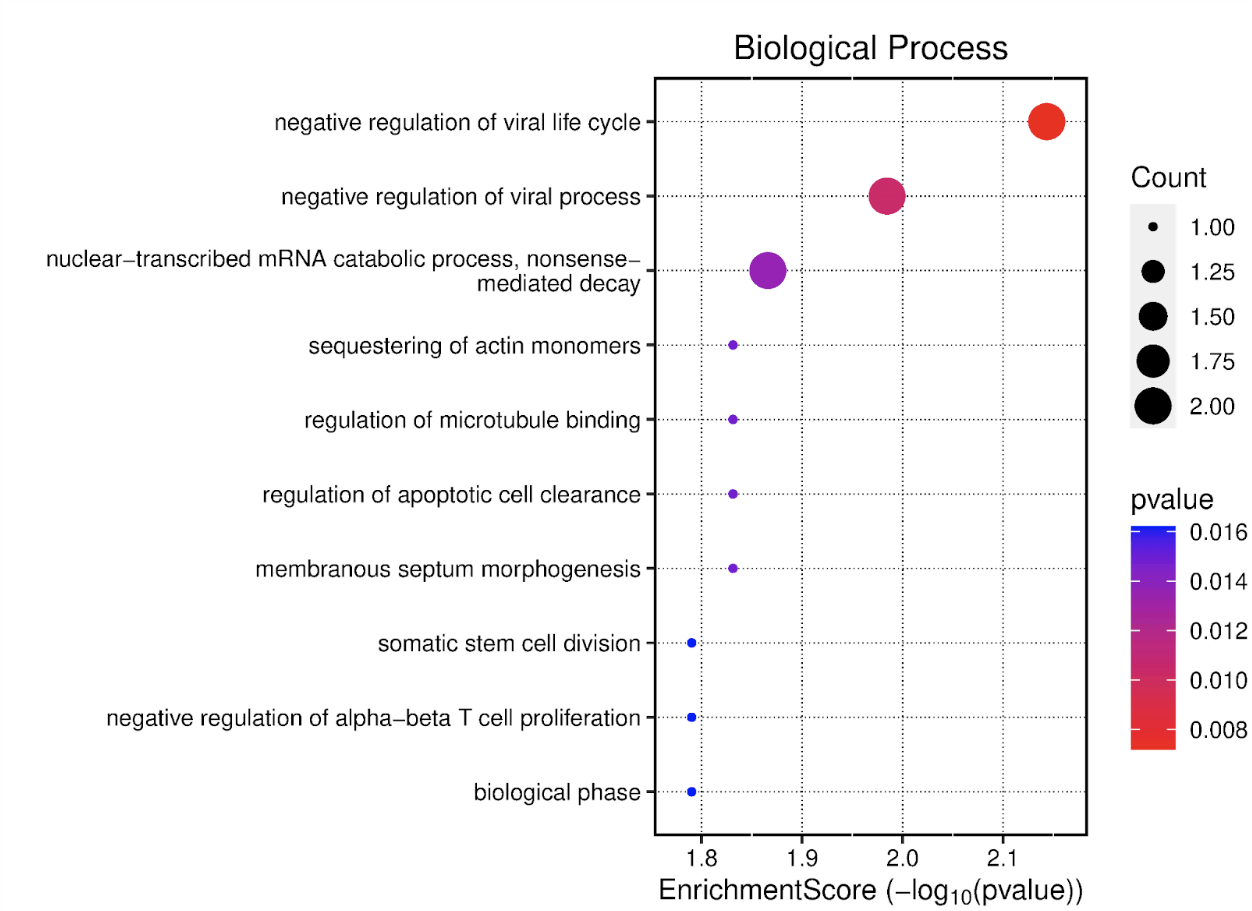


Figure 5. Biological Process Dot Plot: This figure shows the highest enrichment and most expressed biological processes in humans with bacterial sepsis. It seems that the most important processes here relate to the negative regulation of viruses who have higher counts and enrichment scores. The negative regulation of the viral life cycle should be highlighted as it also has a lower p-value than the other processes, making that data the most reliable.

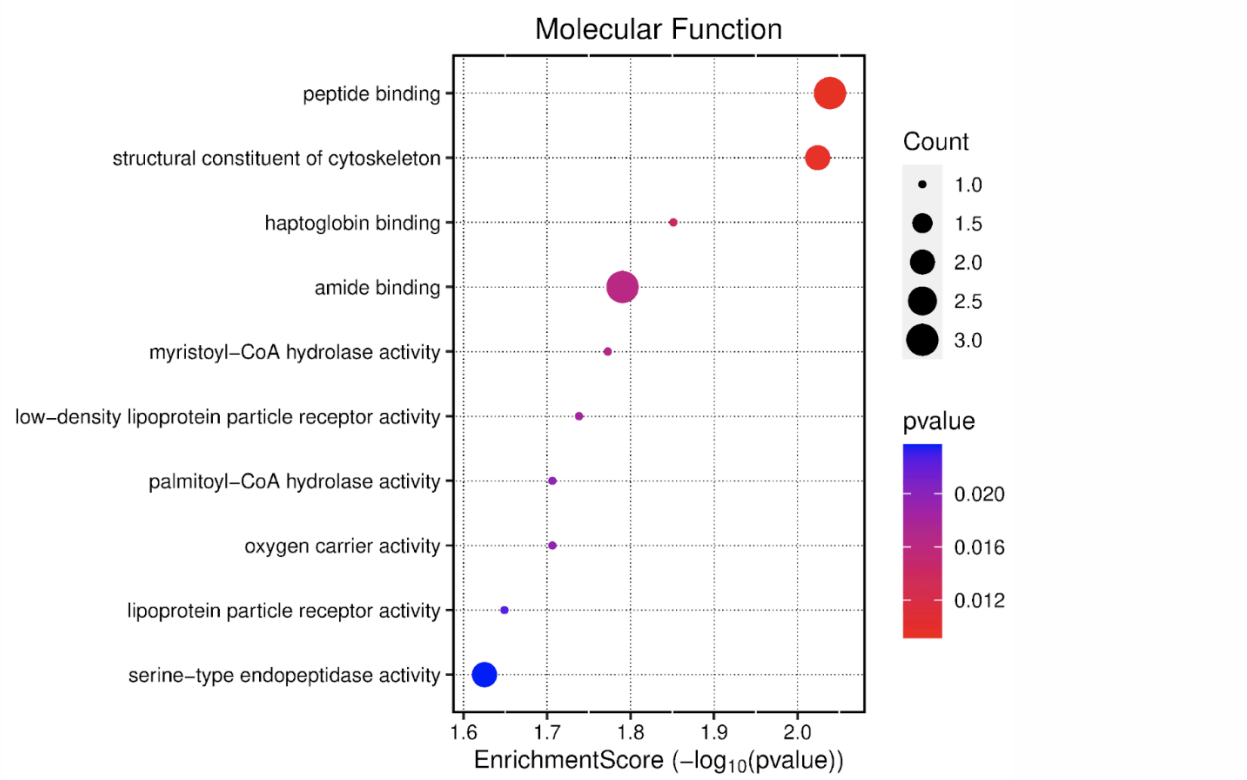


Figure 6. Molecular Function Dot Plot: This figure shows the highest enrichment and most expressed functions in humans with bacterial sepsis. Dots that are more red are most statistically significant, and higher enrichment scores indicate bigger changes in expression. It seems that the most important functions here are peptide binding and cytoskeleton structure as those processes have high enrichment and count values, as well as having lower p-values.

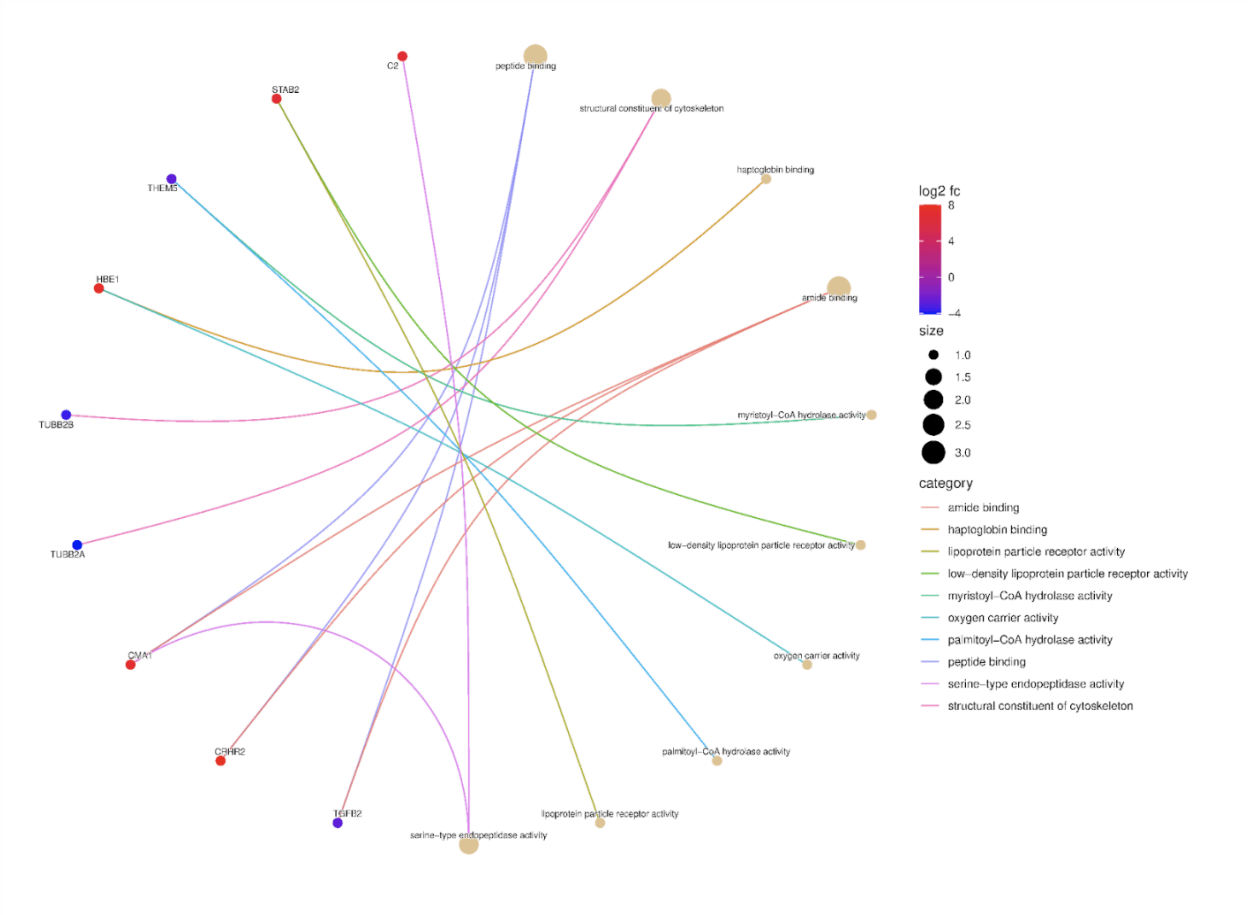


Figure 7. MF CNet Plot: This figure shows the correlation between differentially expressed genes and molecular function in an organism. Genes in red were upregulated while the genes in blue were downregulated. In this case, the important genes are C2 (linked to serine-type endopeptidase activity) , STAB2 (linked to lipoprotein receptor activity), HBE1 (linked to oxygen carrying and haptoglobin binding), CMA1 (linked to serine-type endopeptidase activity, peptide binding, and amide binding), and CRHR2 (linked to peptide binding and amide binding) as they are all significantly upregulated. Additionally, TUBB2A and TUBB2B are both significantly downregulated genes which are both connected to cytoskeletal structure.

Table 1. Summary of Key Genes Identified for Further Research

Gene Name	**Associated Functions and Biological Pathway**
CMA1	Peptide Binding – Produces peptides

	that are used for gland secretion, blood pressure monitoring, and the degradation of the extracellular matrix.
CRHR2	Peptide Binding – Enables hormone activity and acts in many cell signaling pathways
TUBB2A	Cytoskeletal Structure – Encodes beta tubulin
TUBB2B	Cytoskeletal Structure – Encodes beta tubulin

Table 1. Summary of Key Identified Genes: This table highlights and summarizes the genes whose regulation was deemed the most important in non viral sepsis. A focus here was put on molecular function as the highest regulated genes found with regard to biological processes did not align with the top biological processes (see supplementary data).

DISCUSSION

Sepsis is a highly dangerous disease which is caused from the overreaction of the body to foreign pathogens and diseases (WHO 2024). One of these causes is gram-positive bacteria such as E. Coli and S. Aureus (Ramachandran 2014). There existed many unknowns on how these bacteria actually caused sepsis in humans through changes in gene expression. Using NCBI, GEO2R, SR Plot, GO and KEGG bioinformatics databases and tools (Figure 1), 62 genes were identified that were differentially expressed between healthy and infected individuals (Figures 2, 3 and 4), which were then studied to find the general areas and specific genes that were most affected by these bacteria (Figures 5, 6 and 7). In the end, the data suggested that these bacteria had the biggest effect on the biological processes and molecular functions of patients (Figures 4, 5 and 6). The specific processes that were the most affected were those involving the negative regulation of viruses (Figure 7). Meanwhile, the molecular functions impacted were peptide binding and cytoskeletal structure (Figure 6 and 7). When looking for the most impacted genes, attention should be drawn to CMA1 and CRHR2 (Figure 7 and Table 1) as they are genes which are greatly upregulated in bacterial sepsis patients and affect peptide binding, as well as

TUBB2A and TUBB2B (Figure 7 and Table 1) due to their large down regulation in bacterial sepsis patients and correlation to cytoskeletal structure.

These results indicate that bacterial sepsis correlates to human gene expression, specifically in areas such as virus regulation, peptide binding, and cytoskeletal structure (Table 1). In the end, the hypothesis was partially proven as correlation between sepsis and gene regulation was shown. However, it failed to predict the actual genes and processes that were affected.

In terms of negative viral regulation, the data collected here matches those found in other studies. A literary review found a correlation between sepsis patients and an upregulation in pathogen recognition (Russell 2011). This aligns with the definition of negative viral regulation which is a process that stops the growth and reproduction of viruses within a host organism (GO) (Figure 4, 5 and 6).

Although there were not many previous connections found between sepsis and peptide bonding, there have been correlations drawn between sepsis and CR genes such as CRHR2. In a study done on mice, stress responses to sepsis have been linked to the release of the CRH hormone, suggesting an increased expression of genes such as CRHR2 (Gonzalez et al. 2008).

There has also been research relating sepsis to tubulin, which is produced by TUBB2A and TUBB2B and has major impacts on cytoskeletal structure (NIH 2019). Studies have analyzed HDAC6, a key gene in combating sepsis. One key aspect of HDAC6 is its activation of tubulin producing genes like TUBB2A/B (Zhang et al 2021). Further studies could observe if this activation of tubulin producing genes is in response to the down regulation of these genes by sepsis observed above.

The results from this research can be used as an effective treatment against sepsis as it gives us ideas on how to combat sepsis. Personalized medicine can be used in order to counteract some of the molecular function effects of sepsis. For example, treatments which take the role of HDAC6 and other tubulin producing factors can help combat against the down regulation of cytoskeletal structure genes. Additionally, treatments which can stop the

expression of the up-regulated genes such as CRHR2 and CMA1 (Table 1) act as another way to fight sepsis. Additionally, these findings could help vaccine and antibiotic makers target specific genes to fight off bacterial sepsis.

In the future, studies will need to be done observing bacterial sepsis on a populational level as well as laboratory work on the genes such as CRHR2, CMA1, and TUBB2A/B (Table 1) in order to study sepsis's specific effect on these genes.

Conclusion

This bioinformatics study using NCBI, GEO2R, SR Plot, GO and KEGG (Figure 1) gives us more clues on how bacterial sepsis works and its impact on gene expression. The areas most affected by bacterial sepsis are the biological process of negative viral regulation and the molecular functions of peptide binding and cytoskeletal structure. In addition to those functions, the genes to be studied are CRHR2, CMA1, TUBB2A, AND TUBB2B (Table 1). These genes can be used by scientists in future studies to understand how their gene expressions contribute to bacterial sepsis and how researchers can potentially develop better treatments that target these specific genes.

References

Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A., & Rai, A. (2016, April 1). Identification of differentially expressed genes in RNA-seq data of Arabidopsis thaliana: A compound distribution approach. Journal of computational biology : a journal of computational molecular cell biology.

Campbell, & Rondina. (2022, December 1). Series GSE210797. National Center for Biotechnology Information.

EMBL. (n.d.). Gene ontology and go annotations. QuickGO.

Wojtkiewicz, & P. Ciborowski. (2016). National Center for Biotechnology Information.

Maslove, D. M., & Wong, H. R. (2014, February 15). Gene expression profiling in sepsis: Timing, tissue, and translational considerations. Trends in molecular medicine.

Milian, R. G. (2024, May 2). Bioinformatics Thursdays – using GEO2R to quickly reanalyze public expression datasets: RNA-seq and microarrays (in-person). Yale School of Medicine.

Mohammad Yaseen Sofi, & Khalid Z. Masoodi. (2022). Gene expression omnibus. Gene Expression Omnibus – an overview | ScienceDirect Topics.

Ramachandran, G. (2013, November 5). Gram-positive and gram-negative bacterial toxins in sepsis: A brief review. PubMed Central (PMC).

Russell, J. A. (2011). Gene expression in human sepsis: What have we learned? Critical Care, 15(1), 121.

Shao, Q., Chen, D., Chen, S., Ru, X., & Ye, Q. (2023, November 27). escherichia coli infection sepsis: An analysis of specifically expressed genes and clinical indicators. Diagnostics (Basel, Switzerland).

Sizar, O., Leslie, S., & Unakal, C. G. (2023, May 30). Gram-positive bacteria. StatPearls [Internet].

SRplot. (n.d.). SRplot. https://www.bioinformatics.com.cn/en

Tanya Barrett, Dennis B. Troup, Stephen E. Wilhite, Pierre Ledoux, Dmitry Rudnev, Carlos Evangelista, Irene F. Kim, Alexandra Soboleva, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Rolf N. Muertter, Ron Edgar, NCBI GEO: archive for high-throughput functional genomic data, Nucleic Acids Research, Volume 37, Issue suppl_1, 1 January 2009, Pages D885–D890

U.S. National Library of Medicine. (n.d.). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/

U.S. National Library of Medicine. (n.d.). TUBB2B gene: Medlineplus genetics. MedlinePlus.

World Health Organization. (2024, May 3). Sepsis. World Health Organization. https://www.who.int/news-room/fact-sheets/detail/sepsis

Xavier F. Gonzales, Aniket Deshmukh, Mark Pulse, Khaisha Johnson, & Harlan P. Jone. (2007, December 31). Stress-induced differences in primary and secondary resistance against bacterial sepsis corresponds with diverse corticotropin releasing hormone receptor expression by pulmonary cd11c+ MHC II+ and CD11C− MHC II+ apcs. Brain, Behavior, and Immunity.

Zhang, Q., Wang, Y., Qu, D., Yu, J., & Yang, J. (2021, February 25). Role of HDAC6 inhibition in sepsis-induced acute respiratory distress syndrome (review). Experimental and Therapeutic Medicine.

Janani Venkat Easwar

The Impact Of Psychosocial Stress in the Workplace on Risk of Cardiovascular Disease

ABSTRACT

Background: Cardiovascular disease (CVD) is the global leading cause of death. Intervention trials on risk factors such as psychosocial work stress can help address the issue. This necessitates a further holistic understanding of how psychosocial stress affects risk of CVD across genders, races, countries, and occupations.

Methods: This study was a scoping review that utilized PubMed. The terms searched were "workplace stress and cardiovascular health." The main inclusion criteria included studies being done since 2010 on adults, and the main exclusion criteria was studies being from before 2010, studies being done on children, or literature reviews.

Results: Main results from the studies reviewed included a positive correlation between occupational stress in men and obesity/hypertension, a positive correlation between workplace bullying/violence and risk of CVD, and a positive correlation between workplace stress in women and risk of CVD. There was also data displaying a negative correlation between increased psychosocial resources and risk of CVD (especially myocardial infarction), and data displaying how a higher work support score protected against risk of CVD.

Discussion: The key findings indicate a clear positive correlation between psychosocial work stress and risk of CVD. Therefore, implications of work stress should be taken into account by those experiencing it and policymakers who could help prevent excessive occupational stress. While further research is needed to draw more accurate conclusions about this pattern across the entire globe, policymakers can take active steps towards mitigating job stress right now.

Keywords:workplace stress and cardiovascular health

INTRODUCTION

Cardiovascular disease (CVD) is the leading cause of death globally [1]. Though there have been advancements in treatments, it still takes a huge toll on the global population [1]. Sudden cardiac death from cardiac arrest accounts for an estimated 15%-20% of all international deaths [2]. One in 5 men and 1 in 17 women will have some form of coronary heart disease by the age of 60 [3]. Fortunately, intervention trials have shown that lowering risk factors reduces the subsequent rate of CVD [3]. One of such risk factors is psychosocial work stress, which has been associated with a moderately elevated risk of coronary heart disease and stroke [4]. Some governments have started taking steps towards fighting psychosocial stress [4].

However, studies still indicate a serious lack of psychosocial risk assessments, especially in smaller companies [5]. Psychosocial stress is still very much prevalent in different countries and professions [6]. This necessitates a more holistic and worldwide view of how psychosocial stress in the workplace can negatively impact cardiovascular health across countries, races, genders, and occupations.

Psychosocial factors in the workplace include the treatment of workers, organizational climate and culture, work roles, interpersonal relationships, and design and content of tasks [7].

The aim of this scoping review is to analyze and draw holistic connections between psychosocial stress in the workplace and risk of CVD encompassing different countries, races, genders, and occupations. This study will make an impact because it will provide governments of different countries with information about the negative consequences of psychosocial work stress and assist them in making policies to mitigate this type of stress and provide more psychosocial resources.

METHODS

This study is a scoping review searching PubMed using the search terms "workplace stress and cardiovascular health."

Inclusion criteria included the study being conducted only on working adults (aged 18 or older) and being published since 2010. Exclusion criteria eliminated the studies not in English, studies done on children, studies done before 2010, and literature reviews.

RESULTS

The five studies utilized in this scoping review span many fields of work, both genders, and six countries.

Results from a study conducted in Egypt on male bus drivers stated that the mean age of the bus drivers was 37.4 ± 9.0 years [8]. About 61.5% were overweight or obese, and one third experienced high blood pressure [8]. Only 16.7% of bus drivers reported that they were handling the stress well [8]. Occupational stress was severely associated with age, and the percentage with a low education level (77.5%) encountered problems significantly more than those who were more highly educated [8]. A significant association was also found between professional driving duration and age ratio (PDAR) and occupational stress [8]. The prevalence of high blood pressure among bus drivers who encountered stress problems was 4 times higher than the corresponding figure among those who handled stress moderately well [8]. After adjustment for age, it was concluded that only bus drivers who had stress problems were more likely to be obese or overweight, and the prevalence of high blood pressure was strongly connected to occupational stress [8].

A study analyzing workplace violence and bullying through a multi-cohort study in Sweden and Denmark had the following results: the mean of participant age was 43 years old, and 53% were women [9]. There were 78955 total participants with self-reported data on workplace bullying and the prevalence of bullying ranged from 8% to 13% among those participants [9]. Data on workplace violence was received from 79044 participants and ranged from 7% to 17% among those participants [9]. Social workers, personal and protective service workers, healthcare professionals, and teaching professionals had the highest exposure to workplace violence [9]. After adjustment, workplace bullying was associated with a 59% higher risk of incident CVD [9]. There was no evidence that age, gender, or smoking status impacted the results [9]. Workplace violence was also associated with a higher risk of CVD [9].

A Polish study on women undergoing stress in intellectual work both utilized women who had CVD and didn’t have CVD [10]. It concluded that women who were stressed by a health hazard or risk of accident at work had

increased risk of CVD by 35.0%, women who were stressed about changes in the workplace had increased risk by 23.2%, women who were stressed by having to complete tasks without resources had increased risk by 27.3%, and women who were stressed because of competition had increased risk by 25.2% [10].

A multi-cohort study in Denmark, Sweden, and Finland on psychosocial resources had the following results: classes with high horizontal resources (social support and culture of collaboration) and either intermediate or high vertical resources (procedural justice and leadership quality) were at less risk for developing CVD, especially myocardial infarction, corresponding to 3.4 and 2.2 fewer incident CVD cases per person-year respectively [11]. There were not significant differences among age, gender, or educational levels [11].

Finally, an Italian study conducted research on job-induced stress and overall wellbeing of logistics and distribution personnel [12]. It was 57% men and the mean age was 43.27 years [12]. 59% were classified as overweight and 24.6% suffered from hypertension [12]. The study concluded that a higher work support score protected against CVD risk [12]. Additionally, a high score in work control and work support showed a positive effect on wellbeing [12].

DISCUSSION

Based on the reviewed papers, there is a clear correlation between increased psychosocial stress in the workplace and increased risk of CVD. For example, male bus drivers with occupational stress problems were more likely to be obese and have hypertension, increasing their risk for CVD [8]. Data on people facing workplace violence and workplace bullying indicated a 59% higher risk of incident CVD [9]. Several different types of occupational stress on women increased their risk of CVD [10]. Similarly, data indicated that workplaces with high levels of psychosocial resources had less risk for CVD, especially myocardial infarction [11]. A higher work support score also protected against CVD risk [12]. All of these results indicate a clear positive correlation between the amount of psychosocial stress and risk for developing CVD.

The results of this study build on existing studies by adding a layer of universality to the findings (by utilizing data from six different countries). These results should therefore be taken into account by workplace officials or leadership, those seeking out new occupations and considering the implications of those occupations for their health, and those already in stressful occupations.

It is beyond the scope of this study to create a definitive confirmation of the results at a worldwide level, though it certainly contributes to a holistic view of the situation. Further research is needed to establish if these trends can be generalized to other populations. Policymakers in Egypt, Sweden, Denmark, Finland, Italy, and Poland should reconsider working hours, rules for employers, and the funding currently devoted to psychosocial resources for both male and female employees across many types of occupations. Avenues for further research include extending the research across more countries and continents, ethnicities, and occupations.

References

1. Kaminsky L. A., German C., Imboden M., Ozemek C., Peterman J. E., Brubaker P. H. The importance of healthy lifestyle behaviors in the prevention of cardiovascular disease. Progress in Cardiovascular Diseases. 2022; 70: 8-15.
2. Hayashi M., Shimizu W., & Albert C. M. (2015). The spectrum of epidemiology underlying sudden cardiac death. Circulation Research. 2015; 116: 1887-1906.
3. Castelli W. P. Epidemiology of coronary heart disease: the Framingham study. The American Journal of Medicine. 1984; 76: 4-12.
4. Kivimäki M., Kawachi I. Work Stress as a Risk Factor for Cardiovascular Disease. Current Cardiology Reports. 2015; 17: 630.
5. Beck D., Lenhardt U. Consideration of psychosocial factors in workplace risk assessments: findings from a company survey in Germany. International Archives of Occupational and Environmental Health. 2019; 92: 435–451.
6. Onigbogi C. B., Banerjee S. Prevalence of Psychosocial Stress and Its Risk Factors among Health-care Workers in Nigeria: A Systematic Review and Meta-Analysis. Nigerian Medical Journal: Journal of the Nigeria Medical Association. 2019; 60: 238–244.

7. Rugulies R. What is a psychosocial work environment? Scandinavian Journal of Work, Environment & Health. 2019; 45: 1–6.
8. Mohsen A. & Hakim S. Workplace stress and its relation to cardiovascular disease risk factors among bus drivers in Egypt. Eastern Mediterranean Health Journal. 2019; 25: 878–886.
9. Xu T., Magnusson Hanson L. L., Lange T., Starkopf L., Westerlund H., Madsen I. E. H., Rugulies R., Pentti J., Stenholm S., Vahtera J., Hansen Å. M., Virtanen M., Kivimäki M. & Rod N. H. Workplace bullying and workplace violence as risk factors for cardiovascular disease: a multi-cohort study. European Heart Journal. 2019; 40: 1124–1134.
10. Raczkiewicz D., Bojar I., Wdowiak A., Rzeźnicki A. & Krakowiak J. Stress at intellectual work and cardiovascular diseases in women at non-mobility working age. Annals of Agricultural and Environmental Medicine. 2019; 26: 456–461.
11. Xu T., Rugulies R., Vahtera J., Pentti J., Mathisen J., Lange T., Clark A. J., Magnusson Hanson L. L., Westerlund H., Ervasti J., Virtanen M., Kivimäki M. & Rod N. H. Workplace psychosocial resources and risk of cardiovascular disease among employees: a multi-cohort study of 135 669 participants. Scandinavian Journal of Work, Environment & Health. 2022; 48: 621–631.
12. Fruscione S., Malta G., Verso M. G., Calascibetta A., Martorana D. & Cannizzaro E. Correlation among job-induced stress, overall well-being, and cardiovascular risk in Italian workers of logistics and distribution. Frontiers in Public Health. 2024; 12: 1358212.

Alexandria E. Feit

Investigating the Difference of Gene Expression of MRSA Using Bioinformatics Tools

ABSTRACT

Background: Methicillin-resistant Staphylococcus aureus (MRSA) is a contagious staph bacterium causing skin infections that can spread to vital organs and is resistant to many antibiotics, including vancomycin. Genetic factors have been linked to susceptibility to persistent MRSA infections, providing potential insights for better treatments. This study aims to identify immunity-related genes that may contribute to resistance or susceptibility to MRSA.

Methods: Bioinformatics tools and databases were used to analyze MRSA data. Data were collected from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO), which included three groups: mock (healthy mice), mock with linezolid (a treatment), and infected with linezolid. Differentially expressed genes (DEGs) were identified by removing data with a p-value higher than 2.20e-12. The top 50 DEGs were analyzed using ShinyGO, which incorporated KEGG, Gene Ontology (GO), and SRplot tools to determine key pathways and associated genes.

Results: From 20,859 genes, the analysis identified the top 50 DEGs with significant differential expression across groups. ShinyGO revealed two key pathways: the Cytosolic DNA-Sensing Pathway and the Toll-Like Receptor Signaling Pathway. Key genes identified within these pathways included TREX1, IRF7, MLKL, CXCL10, CD14, IP-10, and MIG. These genes are associated with immune responses and regulation, offering insight into MRSA immunity mechanisms.

Conclusion: The study identified key pathways and genes that potentially contribute to MRSA immunity, consistent with previous findings. These genes can be tested in laboratory settings to explore potential antibiotics capable of targeting and weakening MRSA resistance.

Keywords: Methicillin Resistant Staphylococcus Aureus infection, MRSA, NCBI, GEO, KEGG, GO, TREX1, IRF7, MLKL, CXCL10, CD14, IP-10, MIG, Cytosolic DNA-Sensing Pathway, and Toll-Like Receptor Signaling Pathway

INTRODUCTION

Methicillin Resistant Staphylococcus Aureus infection, also known as MRSA, is a very contagious type of staph bacteria that causes skin infection and is resistant to many antibiotics. The bacteria becomes dangerous when it enters the skin through a cut. Once the skin is infected, the infection can spread to the bloodstream, lungs, heart, bones, and joints (1).

MRSA is resistant to many common antibiotics and now it's becoming resistant to its usual treatment vancomycin. Doctors are working on better ways to treat it, but MRSA keeps evolving quickly (1). Doctors are starting to think of new ways and methods to get rid of MRSA and have to keep track of it and study the differences (2)(3).

The MRSA ability to develop antibiotic resistance either happens naturally or the bacteria is able to create defense mechanisms. The antibiotics that do work, don't work everytime. People will need multiple treatments to try to kill the disease (1). For example, MRSA is resistant to a type of antibiotic related to penicillin, cephalosporins, and carbapenems (4). Several studies have shown that genetic factors or genes make some people more prone to persistent MRSA infections, potentially paving the way for improved treatments (5).

Therefore the goal of this study is to identify genes that are related to immunity and therefore potentially identify what makes it immune. Results from this study can potentially identify genes related to MRSA's immune system and serve as potential targets for treatment. We hypothesize that there will be differences between the expression of genes among the different mice samples; mock, mock and linezolid, infected and linezolid in during MRSA.

To conduct this study, several bioinformatics tools and databases were used. First, the National Center for Biotechnology Information (NCBI), Gene Expression Omnibus (GEO) was used to collect data on MRSA and identify the most important or top significant genes that are differentially expressed in MRSA. NCBI is a division of the National Library of Medicine (6). Gene Expression Omnibus (GEO) repository is a bioinformatics tool that archives and freely distributes gene expression experiment results conducted from the microarray method. Data in GEO represent original research deposited by the scientific community, often in compliance with grant or journal

directives that require data to be made publicly available in a MIAME-supportive database (6).

The next bioinformatics tool used was the Kyoto Encyclopedia of Genes and Genomes (KEGG). KEGG is a database that helps to understand the basic principles, as well as practical utilities, of the relations between gene and genomic information and higher order functional information of genes (7). In this study, KEGG was used to determine the potential functions of the identified significant genes from the GEO bioinformatics.

Along with KEGG, the Gene Ontology (GO) bioinformatics database was used to help determine the potential functions of the identified genes. GO is the representation and processing of information about gene products and functions (8).

SRplot is a no-code bioinformatics platform that helps to run the KEGG and GO analysis without the need for computer programming (9). SRPlot is developed for scientists and biologists and supports a wide variety of graphs commonly used in biomedical and bioinformatics publications, which are integrated into a user-friendly graphic user interface (9).

Using all these different bioinformatics tools, this study on MRSA and gene expression can help provide information on what makes the bacteria so immune and identify potential immune related genes that can be used to treat the disease in the future.

METHODS

To conduct this research, the National Center for Biotechnology Information (NCBI) using Gene Expression Omnibus (GEO) bioinformatics tool was used. NCBI-GEO2R is an archive that stores and freely shares research conducted by scientists worldwide using the microarray method. (5). In the NCBI GEO search bar, “MRSA microarray.” was typed and several results were generated. One research was selected with the title, Linezolid Exerts Greater Bacterial Clearance but No Modification of Host Lung Gene Expression Profiling: a Mouse MRSA Pneumonia Model. The researchers in this study had a goal to examine gene expression differences between healthy and infected genes in mice (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE46071).

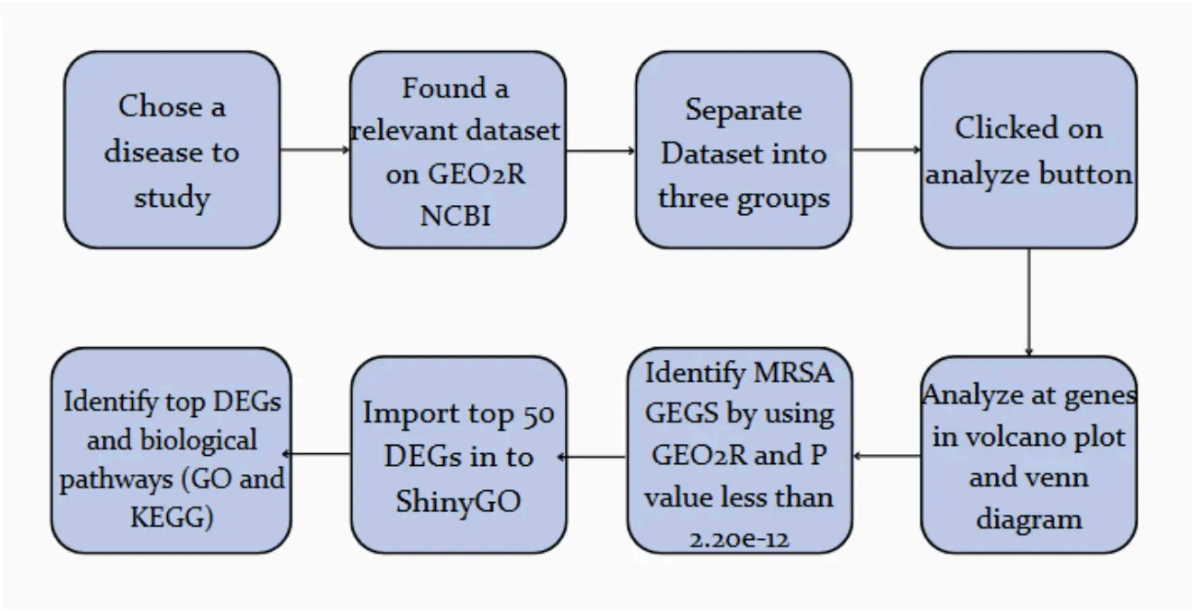


Figure 1: Methods flow chart. This flow chart shows the steps and bioinformatics tools of database Linezolid Exerts Greater Bacterial Clearance but No Modification of Host Lung Gene Expression Profiling: a Mouse MRSA Pneumonia Model.

Data Collection

Next, the provided microarray data was analyzed using the GEO2R bioinformatics tool by using a provided list of mice test subjects. The dataset I used was GSE46071. This data set was with male mice being genes being tested against healthy, healthy and with an antibiotic, and then infection and antibiotic. I divided the mice samples into three groups: mock, mock and linezolid, and infected and linezolid. Mock were the genes in the healthy mice, linezolid is a type of vaccine, and infected are the genes present in the mice infected with MRSA.

Data Analysis of GEO2R Results

To analyze this data, I used the no-code GEO2R 'Analyze' tool which uses pre-programmed coding using R programming language (R-Script). Multiple graphs were generated and selected graph results were analyzed. In the volcano plot, I analyzed how the genes were expressed between the different groups of samples. In the Venn diagram results, I evaluated the number of expressed genes and how many genes overlapped between the different mice sample groups.

Identification of Top 50 Differentially Expressed Genes

To identify the most important expressed genes, I used statistical analysis. I downloaded the top differentially expressed genes (DEGs) data into a Google Sheet and selected the top 50 DEGs by removing any data with a p-value higher than 2.20e-12. Top 50 DEGs).

Data Analysis using SR Plot, KEGG and GO Bioinformatics Tools and Databases

SRPlot, KEGG and GO bioinformatics tools and databases were used to find out what these top genes potentially do in the bacteria and what makes them have immunity against many antibiotics. The species in ShinyGo (ShinyGO 0.81) was set to mouse (Mouse genes GRCm39) because our data was originally collected from mice samples. The IDs of the top 50 differentially expressed genes (DEGs) were analyzed in ShinyGO bioinformatics tool and several results were generated as Gene Ontology results, and the KEGG using no-code R programming language.

RESULTS

Identification of Differentially Expressed Genes

The GEO2R bioinformatics tool was utilized to identify differentially expressed genes (DEGs). The results were visualized using a volcano plot and a Venn diagram, as shown in Figure 2. In the volcano plot comparing "infected and linezolid" versus "mock," the genes were observed to be differentially expressed. A significant portion of the genes was highly expressed, while a notable number displayed lower expression levels. In the plot, red dots represent highly expressed (upregulated) genes, blue dots indicate less expressed (downregulated) genes, and black dots signify no significant differences between the two groups.

The Venn diagram revealed a total of 20,859 genes, with 771 genes specific to "infected and linezolid versus mock," 552 genes specific to "mock and linezolid versus infected and linezolid," and 3,515 genes overlapping between the two comparisons.

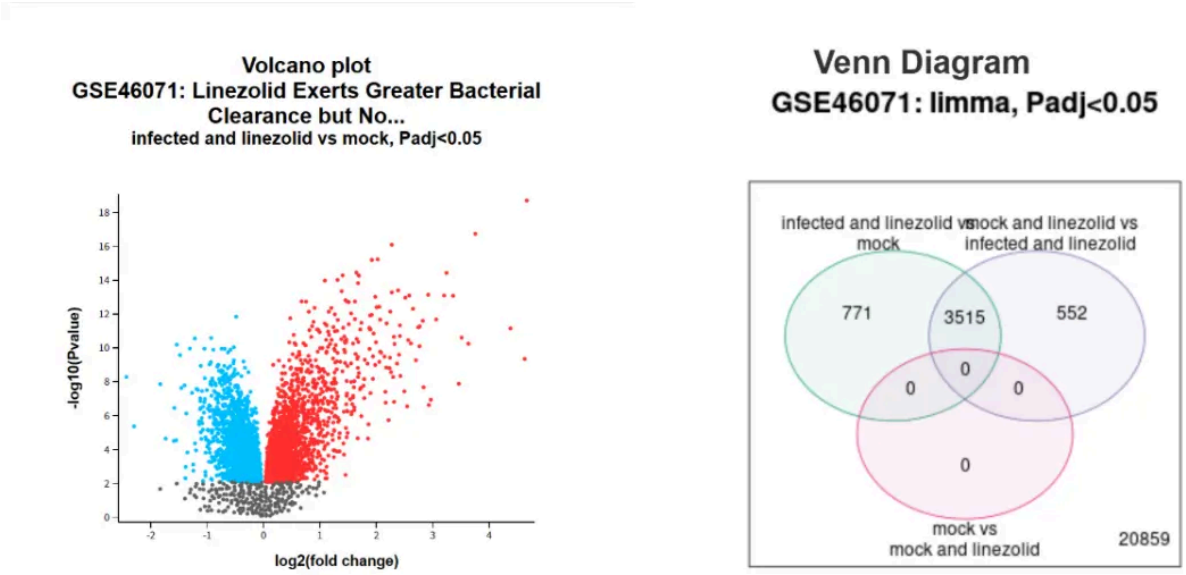


Figure 2: A and B: Data from GEO2R results A. volcano plot and B. venn diagram. In the volcano plot Red dots mean highly expressed or up-regulated. Blue dots means less express or down-regulated. And black dots mean they're no differences between the 2 groups. Most were highly expressive, but there was still a good amount of less expressive genes. In the venn diagram, there are a total of 20859 genes. 771 of the genes are in infected and linezolid vs mock, 552 genes are in mock and linezolid vs infected and linezolid, and 3515 genes overlapped in both.

dentification of 50 Statistically Significant Differentially Expressed Genes (DEGs)

To identify my top 50 Differentially Expressed Genes (DEGs) from the total of 20859 genes, DEGs with a p-value greater than 2.20e-12 were excluded (Top 50 DEGs).

Potential Functions and Enrichment of the Identified Genes

Through ShinyGO I identified 2 pathways, Cytosolic DNA-Sensing Pathway and Toll-Like Receptor Signaling Pathway, from my KEGG Results that had with key genes of TREX1, IRF7, MLKL, CXCL10, CD14, IP-10, and MIG **(Figure 3)**.

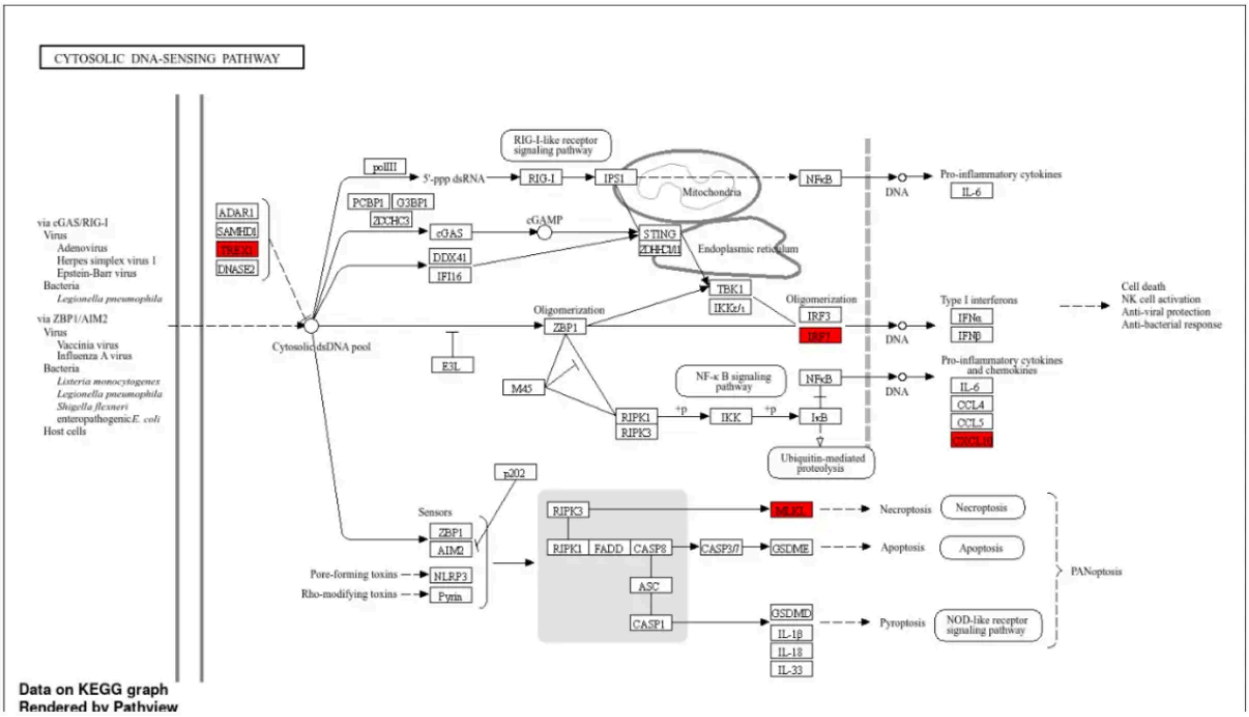


Figure 3: Key Pathway identified to be associated with the top 50 DEGs. Cytosolic DNA-Sensing Pathway with key genes highlighted in red TREX1, IRF7, MLKL, CXCL10.

Also the GO results key pathways identified were also Cytosolic DNA-Sensing Pathway and Toll-Like Receptor Signaling Pathway **(Figure 4).**

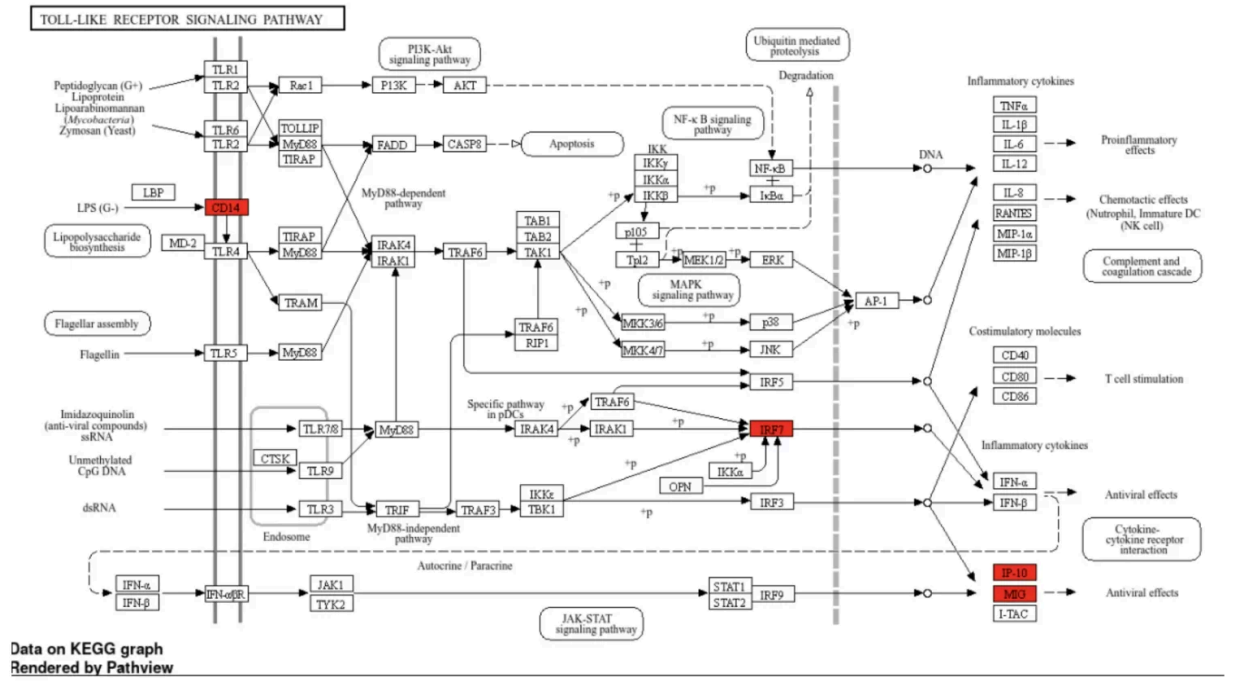


Figure 4: Second Key Pathway identified to be associated with the top 50 DEGs Toll-Like Receptor Signaling Pathway with key genes highlighted in red; CD14, IRF7, IP-10, and MIG

In summary, the key genes in the KEGG results were TREX1, IRF7, MLKL, CXCL10, CD14, IRF7, IP-10, and MIG and key pathways from GO results is Cytosolic DNA-sensing pathway an

DISCUSSION

Some studies have shown that genetic factors make some people more prone to persistent MRSA infections, potentially paving the way for improved treatments. For example, one study found that a mutation in DNMT3A, an enzyme involved in regulating gene expression through DNA methylation, helps fight MRSA infections. It does this by boosting gene regulation and reducing levels of the anti-inflammatory cytokine IL-10 (5).

The goal of this study was to identify genes in MRSA that are related to immunity **(Figure 3 and 4)** and the differentially expressed genes in the groups as mock, mock and linezolid, and infected and linezolid **(Figure 2).** My hypothesis was that the genes were differentially expressed among the different groups. Most of the genes were highly expressed making it react to change easily and its key genes and pathways helped it to be immune.

The findings of the key genes and pathways from this study **(Summary Table 1)** are supported with many other studies stating MRSA immunity (1). First, results identified TREX1 **(Figure 3)**. Previous studies have shown that indeed TREX1 helps the MRSA immune system not attack its own DNA (9). Next, this study identified IRF7, which is an important transcriptional regulator of cellular responses in MRSA (10). MLKL helps the MRSA fight off infections by killing infected cells (11) CXCL10 sends signals to certain cells, telling them where to go and what to do to fight off infections or diseases (in this case vaccines and antibiotics) (12). CD14 is an important molecule that helps activate the body's first line of defense against infections that is found either attached to the surface of immune cells or floating freely in the body (13). IRF7 is an important transcriptional regulator of cellular responses in MRSA (10). IP-10 is a special protein that gets released when the body is fighting off infections (14). Mig is a special protein that is released when the MRSA gets a signal from interferon-gamma and acts like a "signal" that attracts activated T cells (15).

In the real world, we can use this information on creating vaccines that can target MRSA immunity. Below is a summary of the key genes and pathways identified in this study that are linked to MRSA and with a potential for targets in vaccine development.

Table 1. Genes and Pathways Identified in this Study

Gene	**Pathways**	**Connection to MRSA**
TREX1	Cytosolic DNA-Sensing Pathway	TREX1 helps the MRSA immune system not attack its own DNA (10).
IRF7	Cytosolic DNA-Sensing Pathway	IRF7 is an important transcriptional regulator of cellular responses in MRSA (11).
MLKL	Cytosolic DNA-Sensing Pathway	MLKL helps the MRSA fight off infections by killing infected cells

		(12)
CXCL10	Cytosolic DNA-Sensing Pathway	CXCL10 sends signals to certain cells, telling them where to go and what to do to fight off infections or diseases (in this case vaccines and antibiotics) (13).
CD14	Toll-Like Receptor Signaling Pathway	CD14 is an important molecule that helps activate the body's first line of defense against infections that is found either attached to the surface of immune cells or floating freely in the body (14).
IRF7	Toll-Like Receptor Signaling Pathway	IRF7 is an important transcriptional regulator of cellular responses in MRSA (11).
IP-10	Toll-Like Receptor Signaling Pathway	IP-10 is a special protein that gets released when the body is fighting off infections (15).
MIG	Toll-Like Receptor Signaling Pathway	Mig is a special protein that is released when the MRSA gets a signal from interferon-gamma and acts like a "signal" that attracts activated T cells (16).

Limitations

One limitation is that the identified genes will need to be further studied by scientists , since we used bioinformatics datasets from microarray experiments conducted by other researchers. These identified genes can be

tested in the laboratory by scientists to find what antibiotics can weaken the gene.

References

1. "MRSA Infection – Symptoms & Causes." Mayoclinic, Mayo Foundation for Medical Education and Research, 8 Nov. 2022, www.mayoclinic.org/diseases-conditions/mrsa/symptoms-causes/syc-20375336.
2. Rasmussen RV, Fowler VG Jr, Skov R, Bruun NE. Future challenges and treatment of Staphylococcus aureus bacteremia with emphasis on MRSA. Future Microbiol. 2011 Jan;6(1):43-56. doi: 10.2217/fmb.10.155. Erratum in: Future Microbiol. 2011 Feb;6(2):250. PMID: 21162635; PMCID: PMC3031962.
3. Mason A, Foster D, Bradley P, Golubchik T, Doumith M, Gordon NC, Pichon B, Iqbal Z, Staves P, Crook D, Walker AS, Kearns A, Peto T. Accuracy of Different Bioinformatics Methods in Detecting Antibiotic Resistance and Virulence Factors from Staphylococcus aureus Whole-Genome Sequences. J Clin Microbiol. 2018 Aug 27;56(9):e01815-17. doi: 10.1128/JCM.01815-17. PMID: 29925638; PMCID: PMC6113501.
4. "MRSA (Methicillin-resistant Staphylococcus aureus)." Cleveland Clinic, my.clevelandclinic.org/health/diseases/11633-methicillin-resistant-staphylococcus-aureus-mrsa. Accessed 7 May 2024.
5. F. Mba Medie, B.K. Sharma-Kuinkel, F. Ruffin, L.C. Chan, M. Rossetti, Y. Chang, L.P. Park, A.S. Bayer, S.G. Filler, R. Ahn, E.F. Reed, D. Gjertson, M.R. Yeaman, V.G. Fowler, , Genetic variation of DNA methyltransferase-3A contributes to protection against persistent MRSA bacteremia in patients, Proc. Natl. Acad. Sci. U.S.A. 116 (40) 20087-20096,https://doi.org/10.1073/pnas.1909849116 (2019).
6. Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, Alexandra Soboleva, NCBI GEO: archive for functional genomics data sets—update, Nucleic Acids Research, Volume 41, Issue D1, 1 January 2013, Pages D991–D995, https://doi.org/10.1093/nar/gks1193

7. Minoru Kanehisa, Susumu Goto, Shuichi Kawashima, Akihiro Nakaya, The KEGG databases at GenomeNet, Nucleic Acids Research, Volume 30, Issue 1, 1 January 2002, Pages 42–46, https://doi.org/10.1093/nar/30.1.42
8. Smith B, Williams J, Schulze-Kremer S. The ontology of the gene ontology. AMIA Annu Symp Proc. 2003;2003:609-13. PMID: 14728245; PMCID: PMC1480173.
9. Turner NA, Sharma-Kuinkel BK, Maskarinec SA, Eichenberger EM, Shah PP, Carugati M, Holland TL, Fowler VG Jr. Methicillin-resistant Staphylococcus aureus: an overview of basic and clinical research. Nat Rev Microbiol. 2019 Apr;17(4):203-218. doi: 10.1038/s41579-018-0147-4. PMID: 30737488; PMCID: PMC6939889.
10. Yan N. Immune Diseases Associated with TREX1 and STING Dysfunction. J Interferon Cytokine Res. 2017 May;37(5):198-206. doi: 10.1089/jir.2016.0086. PMID: 28475463; PMCID: PMC5439420.
11. Qing F, Liu Z. Interferon regulatory factor 7 in inflammation, cancer and infection. Front Immunol. 2023 May 12;14:1190841. doi: 10.3389/fimmu.2023.1190841. PMID: 37251373; PMCID: PMC10213216.
12. Tovey Crutchfield EC, Garnish SE, Hildebrand JM. The Role of the Key Effector of Necroptotic Cell Death, MLKL, in Mouse Models of Disease. Biomolecules. 2021 May 28;11(6):803. doi: 10.3390/biom11060803. PMID: 34071602; PMCID: PMC8227991.
13. Liu M, Guo S, Stiles JK. The emerging role of CXCL10 in cancer (Review). Oncol Lett. 2011 Jul;2(4):583-589. doi: 10.3892/ol.2011.300. Epub 2011 May 9. PMID: 22848232; PMCID: PMC3406435.
14. Y. Tesfaigzi, M. Daheshia, CD14, Editor(s): Geoffrey J. Laurent, Steven D. Shapiro, Encyclopedia of Respiratory Medicine, Academic Press, 2006, Pages 343-347, ISBN 9780123708793, https://doi.org/10.1016/B0-12-370879-6/00063-6.
15. Sasya Madhurantakam, Zachary J Lee, Aliya Naqvi, Shalini Prasad, Importance of IP-10 as a biomarker of host immune response: Critical perspective as a target for biosensing, Current Research in Biotechnology, Volume 5, 2023, 100130, ISSN 2590-2628, https://doi.org/10.1016/j.crbiot.2023.100130.
16. Sgadari C, Farber JM, Angiolillo AL, Liao F, Teruya-Feldstein J, Burd PR, Yao L, Gupta G, Kanegane C, Tosato G. Mig, the monokine induced

by interferon-gamma, promotes tumor necrosis in vivo. Blood. 1997 Apr 15;89(8):2635-43. PMID: 9108380.

Rohan Gazula

Exploring the Genetics of Glaucoma through Bioinformatics Research

ABSTRACT

Background: Glaucoma is the leading cause of irreversible blindness worldwide, affecting roughly 80 million people. Anyone can be at risk of glaucoma, although it is more common among African-American and Latino populations. Glaucoma is characterized as the deterioration of the optic nerve, which commonly occurs as a result of excessive intraocular eye pressure (IOP). Glaucoma can result in loss of vision and in some cases blindness. The problems in this disease lie in the fact that it is not very well understood, despite its prevalence. There are multiple complex molecular mechanisms at play in the behavior of this disease. The exact mechanisms that lead to the deterioration of the optic nerve are not fully understood, and while elevated IOP is certainly a factor, there could exist several other influences such as blood flow and autoimmune responses.

Therefore, the goal of this research study is to uncover potential genes or genetic components that are associated with the behavior of glaucoma, ultimately to provide genetic information and biomarkers that can potentially assist in the understanding of this disease.

Methods: This study used several bioinformatics tools—namely, Gene Expression Omnibus (GEO), GEO2R, and SR Plot. The GEO dataset used to collect data was GSE218153, a study that specifically analyzed human trabecular meshwork cells cultured in a lab. As a control, the researchers used a gene called normal TIPARP in the control group and increased TIPARP in the experimental group using a special virus, lentivirus-GFP-TIPARP. In our study, we used GEO2R to categorize and analyze the groups into a control group with elevated eye pressure and a group with upregulated TIPARP with lowered eye pressure. Genes expressed differently (DEGs) between these two groups were identified using GEO2R. Then SR Plot was used to conduct enrichment analysis of the top 30 DEGs and was used for GO and KEGG pathway analysis.

Results: 5248 genes were initially identified across the control and experimental groups (Figure 2B), but after considering fold change and p

value, 30 top DEGs were identified (Supplementary Results 2). Enriched gene ontology terms include “Response to lipopolysaccharide,” “Response to molecules of bacterial origin,” “cellular response to lipopolysaccharide,” and “cellular response to molecules of bacterial origin”. Enriched KEGG Pathways include “IL-17 signaling pathway,” Viral protein interaction with cytokine and cytokine receptor,” cytokine-cytokine receptor interaction,” and “Influenza A”. “Response to lipopolysaccharide,” “cellular response to lipopolysaccharide,” “Viral protein interaction with cytokine and cytokine receptor,” and “cytokine-cytokine receptor interaction” summarize the findings related to the DEGs along with the top common genes CXCL11, CXCL10, RSAD2, and CSF2.

Conclusion: The gene function “response to lipopolysaccharide” and biological pathway “Viral protein interaction with cytokine and cytokine receptor,” along with 4 top common genes: CXCL11, CXCL10, RSAD2, CSF2 are all in some way related to inflammation, which reveals an important insight into glaucoma. Glaucoma-related inflammation is now a potential field of further study as it relates to many of the common genes, functions, and biological pathways. Future study in this direction could accelerate the development of a cure for glaucoma.

INTRODUCTION

As the learning cause of irreversible blindness, glaucoma affects roughly 80 million people worldwide (1, 2). In spite of its prevalence, this disease is problematic in that it is not very well understood; there are multiple complex molecular mechanisms involved in its pathology (2). This research study attempts to address this important issue by revealing findings surrounding the genetic mechanisms of glaucoma. Specifically, this research study attempts to answer the question, “what genes or genetic components are associated with the behavior of glaucoma?”; in other words, “what genetic information and biomarkers can assist in the understanding of this disease?” GEO (http://www.ncbi.nlm.nih.gov/geo) provides publicly available datasets, containing the gene expression of several diseases (3). Using a GEO Dataset containing glaucoma, along with several bioinformatics tools, this study will provide sufficient insights into the precise genes associated with glaucoma.

Glaucoma comprises a family of conditions that are marked by damage to the optic nerve as a result of several risk factors, commonly as a result of high pressure in the eyes (1). However, as the molecular pathways that lead to this deterioration of the optic nerve are rather unclear, the experimental findings of this study are essential to providing a better understanding of this complex disease (2). It is known that if left untreated, in some cases glaucoma can lead to irreversible blindness due to excessive deterioration of the optic nerve (1). It is also known that there are several types of glaucoma, including primary open-angle glaucoma, angle-closure glaucoma, normal-tension glaucoma, secondary glaucoma, or congenital glaucoma (1).

In regards to bioinformatics study, it is also known that risk factors of glaucoma include elevated eye pressure, older age (>40), race/ethnicity, other medical conditions (diabetes, hypertension, CVD), and eye injuries (2). Thus, genetics associated with these trends are pathways to explore in the understanding of this disease. It is further known that an effective treatment plan and an early detection of glaucoma can help mitigate vision loss, which suggests the importance of bioinformatics efforts in the study of this disease (1). However, the exact mechanisms that lead to the deterioration of the optic nerve are not fully understood (2). Although elevated eye pressure is a factor which contributes to damage to the optic nerve (1) there could be several other risk factors including blood flow and autoimmune responses that may also play a role in damaging the optic nerve.

The goal of this research regarding gene expression is to uncover notable upregulated or downregulated genes associated with the glaucoma pathology, that is, the typical behavior of the condition of glaucoma. Given the association between elevated Intraocular pressure (IOP) and glaucoma, the hypothesis is that, there certainly will exist differences in the expression of genes between controls and patients, with some genes being upregulated and downregulated in glaucoma patients, especially those related to IOP. In order to accomplish this goal, the bioinformatics tools of Gene Expression Omnibus (GEO), GEO2R, and SR Plot were used. GEO (http://www.ncbi.nlm.nih.gov/geo) is a publicly available database within the National Center for Biotechnology Information (3). Specifically, GEO provides publicly available datasets, containing the gene expression of several diseases (3). GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r/) is a program designed to compare gene expression between different groups,

yielding fold-change and P-values (4). SR Plot is a program which assists finding gene functions and biological mechanisms (5). SR Plot also assists in visualizing this data and exploring this data through different graphs (5).

The importance of this research lies in the fact that as a disease, glaucoma is highly prevalent among the population, serving as the leading cause of irreversible blindness world-wide (1). Furthermore, as it concerns disease whose genetic mechanisms are relatively less-understood, my research is thus critical to finding a treatment for this both prevalent and enigmatic disease. A better understanding of the genetic mechanisms behind glaucoma can ultimately make a large impact, as it could reveal underlying biological mechanisms and facilitate the development of new treatments.

METHODS

To conduct this bioinformatics study, different methods were used and are summarized in the flowchart below (Figure 1). This study utilized several bioinformatics tools, including the Gene Expression Omnibus (GEO), GEO2R, and SR Plot.

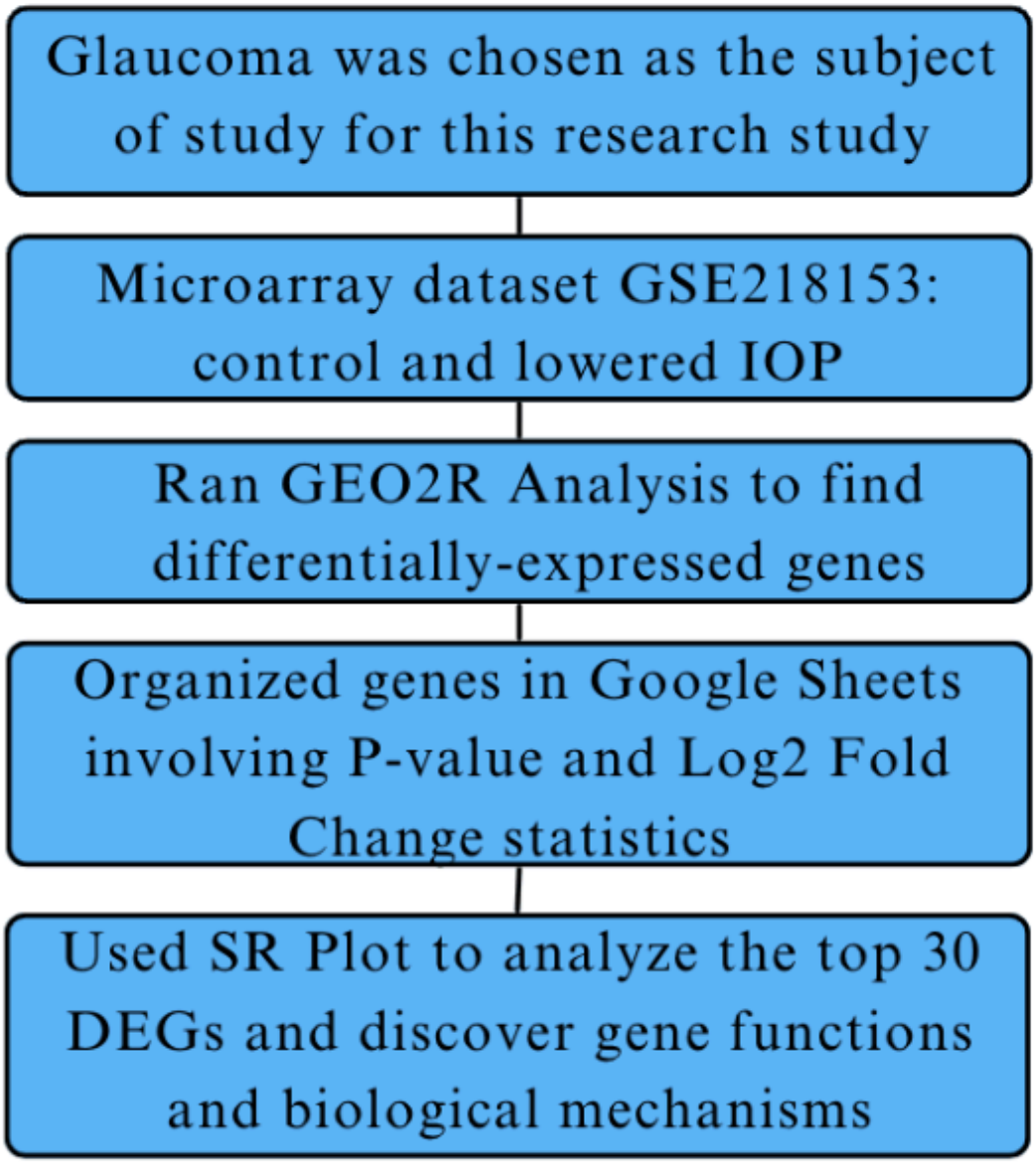


Figure 1: Flowchart of steps taken to conduct this bioinformatics study. Summary of the various steps and bioinformatics tools used in this tidy including the Gene Expression Omnibus (GEO), GEO2R, and SR Plot.

Determining the Dataset for Data Collection

The study dataset used for data collection was identified by consulting GEO, a publicly available database containing several pathologies' gene expression, to obtain the glaucoma dataset (3). After searching through the

GEO database, the dataset of study "TIPARP is involved in the regulation of intraocular pressure" was selected as an adequate pool of data to analyze, with the accession number GSE218153. This study, which explores the effect of the TIPARP gene, was conducted by analyzing human trabecular meshwork cells cultured in a lab, with normal TIPARP in the control group and increased TIPARP in the experimental group using a special virus, lentivirus-GFP-TIPARP (6). The study then explored the difference between the control group and the upregulated TIPARP group, implementing gene expression analysis using RNA sequencing (6).

Identifying the most Differentially-Expressed Genes using GEO2R Bioinformatics Tool

Differentially expressed genes are genes that are prevalent to different extents among distinct groups (7).

To identify the most Differentially-Expressed Genes among individuals with elevated and normal eye pressure, the dataset GSE218153 was analyzed using GEO2R, an online bioinformatics tool for the purpose of evaluating different sample groups (4). Before analyzing the groups through this method, the data was first deliberately sorted into two groups, a control group and TIPARP-upregulated group. Then, data was analyzed using the GEO2R user interface, an AI and ML pre-programmed interface that uses the R programming language. This GEO2R analysis generated several results of differentially expressed genes (DEGs).

The DEGs or differentially expressed genes were then statistically narrowed down to the top 30 DEGs using the P-value and log2FC. All genes with a P-value greater than 0.05 were eliminated, leaving genes with sufficient authenticity. Finally, the top 30 DEGs were determined by sorting the log2FC to those with the highest magnitudes specifically, the 15 most positive and the 15 most negative. These 30 genes were either the most comparatively upregulated in the control or the experimental group. After identifying the 30 most expressed DEGs on both ends of the spectrum, namely, comparative upregulation in the control vs comparative upregulation in the experimental groups these genes were isolated on a separate Google Sheet, along with their corresponding log2FC values.

Determining Functions and Biological Pathways of top DEGs using SR Plot

Next, the SR Plot bioinformatics tool was used to determine the functions of these specific genes and the biological pathways that they contribute to (5). After navigating to bioinformatics.com.cn/en, the term "enrichment," was searched. Finally, after clicking on "GO Pathway Enrichment Analysis," the list of 30 genes and their FC values, was copied and pasted into the bioinformatics tool. This data was then downloaded, which included the functions of the 30 DEGs in the Gene Ontology and the biological pathways of the 30 DEGs in the KEGG enrichment analysis.

RESULTS

Differentially Expressed Genes and GEO2R Analysis

GEO2R analysis of the dataset GSE218153 initially resulted in 5,248 genes across the control and experimental groups (Figure 2B). However, after considering the statistics of fold change and p-value, the list was refined to the top 30 differentially expressed genes (DEGs) (Top DEGs).

The red dots in the volcano plot diagram indicate genes that are upregulated in the higher IOP, control group. On the contrary, the blue dots in the volcano plot indicate the genes that are downregulated in the higher IOP, control group. Finally, the black dots indicate genes that are not statistically significant enough—in other words, the genes with too high of a p-value. The venn diagram shows that out of a total of 16277 genes, 5248 genes were differentially expressed across the control and experimental group. These GEO2R results were generated using R programming which is already abstracted and pre-programmed using AI and ML algorithms (R Script Used to Generate GEO2R Results).

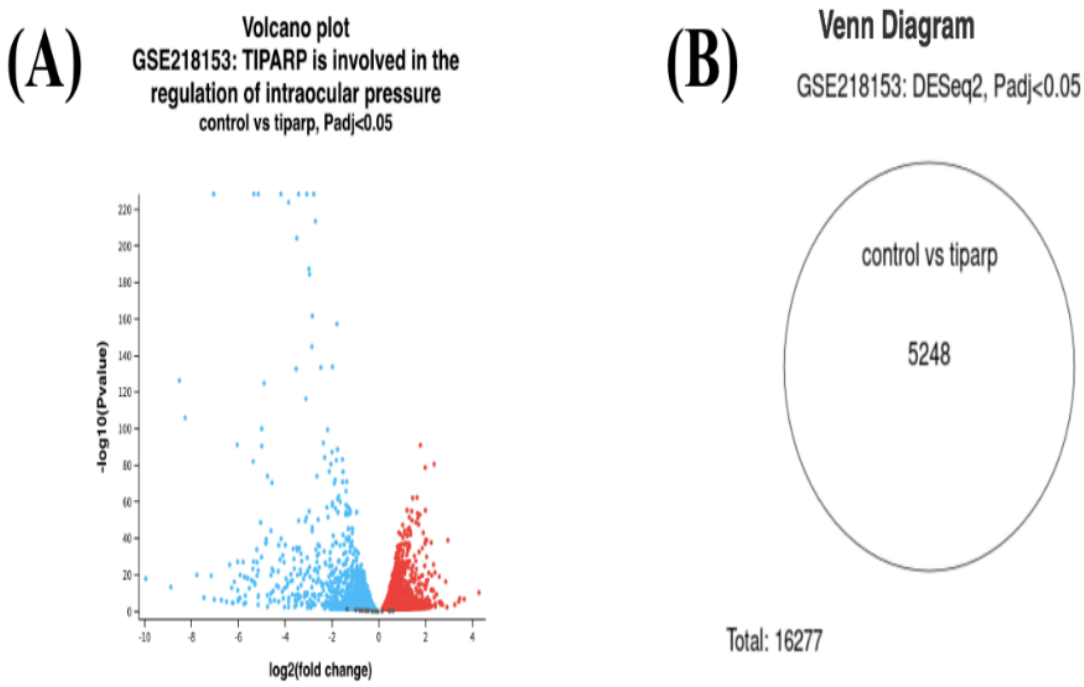


Figure 2. Differentially expressed genes from GEO2R analysis. Volcano Plot (A) shows DEGs across control and TIPARP upregulated groups. Venn Diagram (B) shows the DEGs in the control versus the TIPARP upregulated groups.

30 Top DEGs total were identified; 15 upregulated and 15 downregulated in control

Top DEGs

GO Enrichment Analysis with SR Plot

Gene ontology analysis using SR Plot revealed several enriched terms, including “Response to lipopolysaccharide,” “Response to molecule of bacterial origin,” “Cellular response to lipopolysaccharide,” and “Cellular response to molecule of bacterial origin.” These terms suggest a significant involvement of these genes in immune response mechanisms.

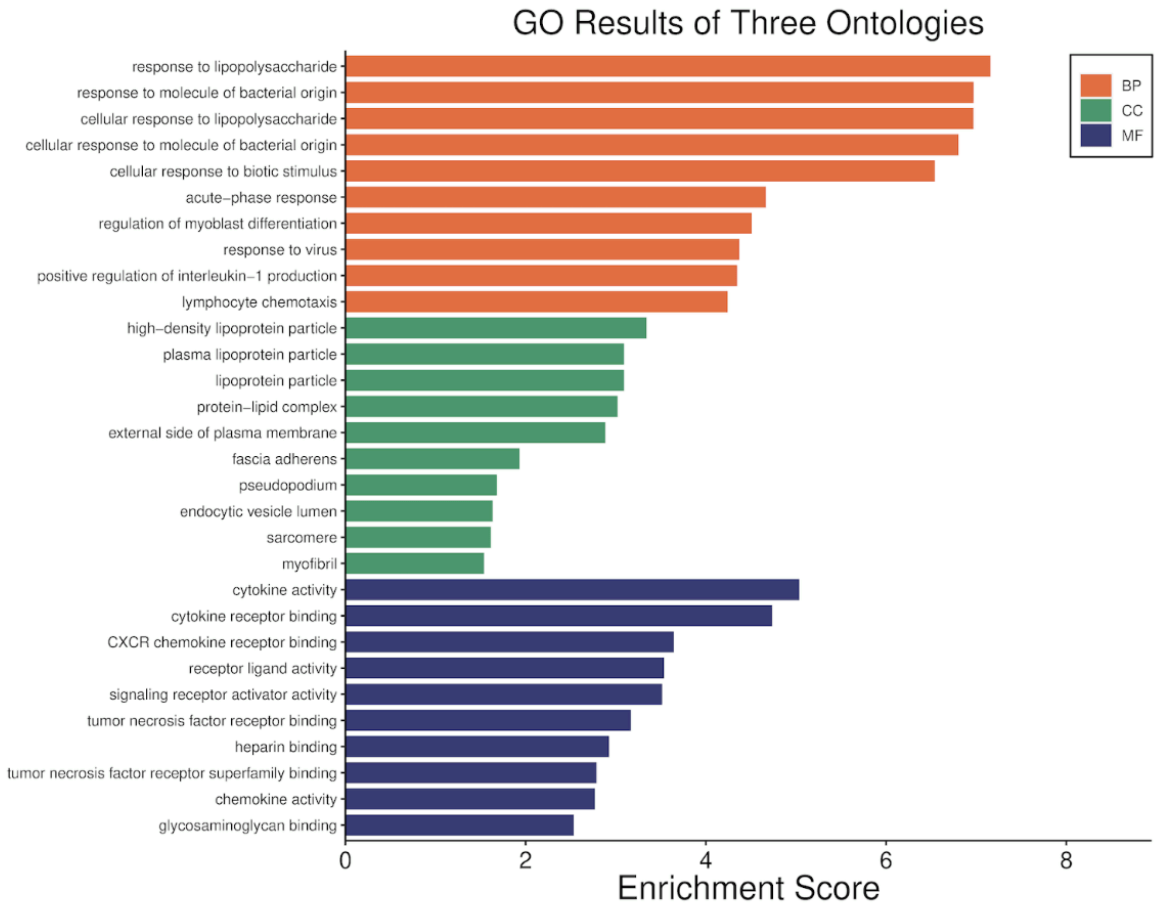


Figure 3. GO Results of three ontologies. The orange bars designate biological processes, the green bars designate cellular components, and the purple bars designate molecular functions.

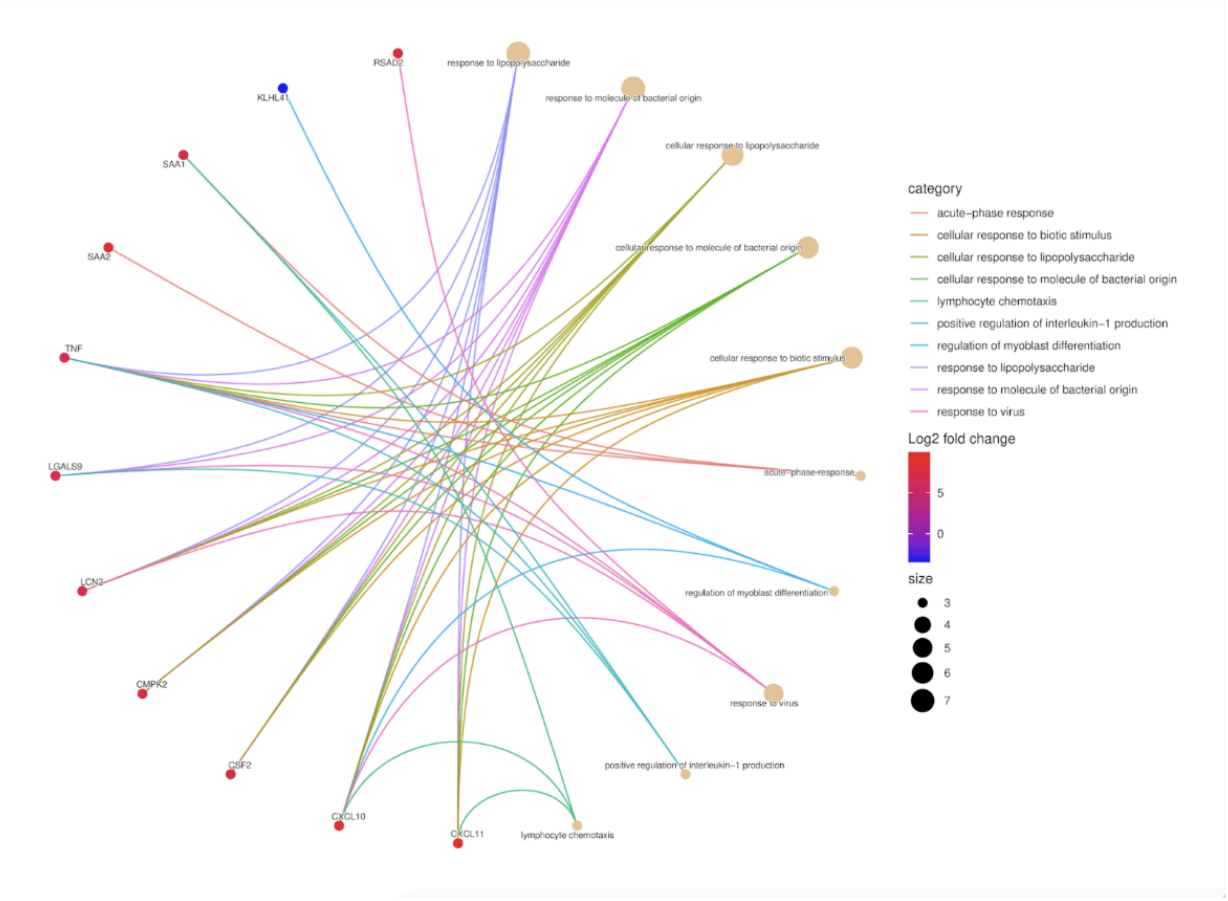


Figure 4. GO Enrichment Analysis – Biological Processes Cnet Plot. The circles on the outside of the web represent genes, and the more enriched genes are CXCL10, CXCL11, CSF2, CMPK2, LCN2, LGALS9, TNF, SAA2, SAA1, and RSAD2. These genes link to the biological processes response to lipopolysaccharide, response to molecule of bacterial origin, cellular response to lipopolysaccharide, cellular response to biotic stimulus, acute-phase response, regulation of myoblast differentiation, response to virus, positive resolution of interleukin-1 productions, and lymphocyte chemotaxis. Genes with a higher Log2 fold change are more red while genes with a lower Log 2 fold change are more blue.

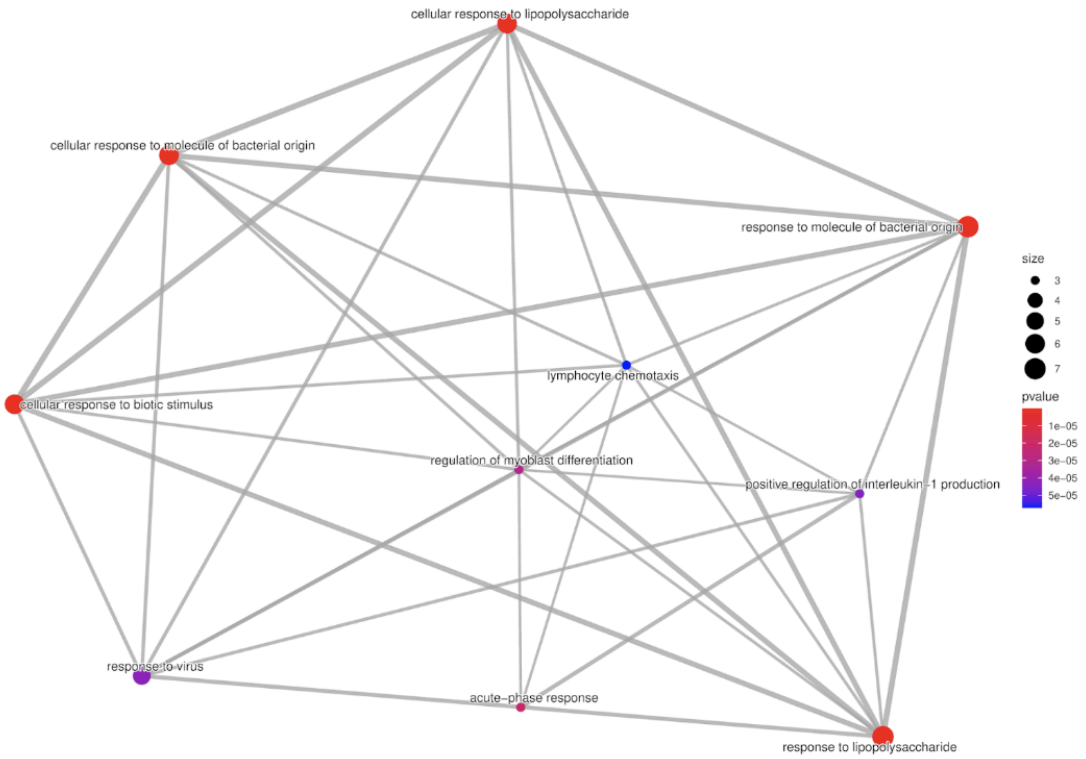


Figure 5. GO Enrichment Analysis – Biological Processes Emap Plot. The circles connecting the lines represent biological processes, and the circles with a lower p-value are redder and more significant. The more significant processes are response to lipopolysaccharide, response to molecule of bacterial origin, cellular response to lipopolysaccharide, cellular response to biotic stimulus, and acute-phase response. The circles with a higher p-value are more blue and less significant.

KEGG Pathway Enrichment Analysis with SR Plot

In addition to gene ontology, pathway enrichment analysis using the KEGG database identified several relevant pathways. These include the "IL-17 signaling pathway," "Viral protein interaction with cytokine and cytokine receptor," "Cytokine-cytokine receptor interaction," and "Influenza A" pathways. These pathways highlight the potential roles of the DEGs in inflammatory responses and interactions with viral proteins.

Enriched Pathways: IL-17 signaling pathway, Viral protein interaction with cytokine and cytokine receptor, cytokine-cytokine receptor interaction, Influenza A

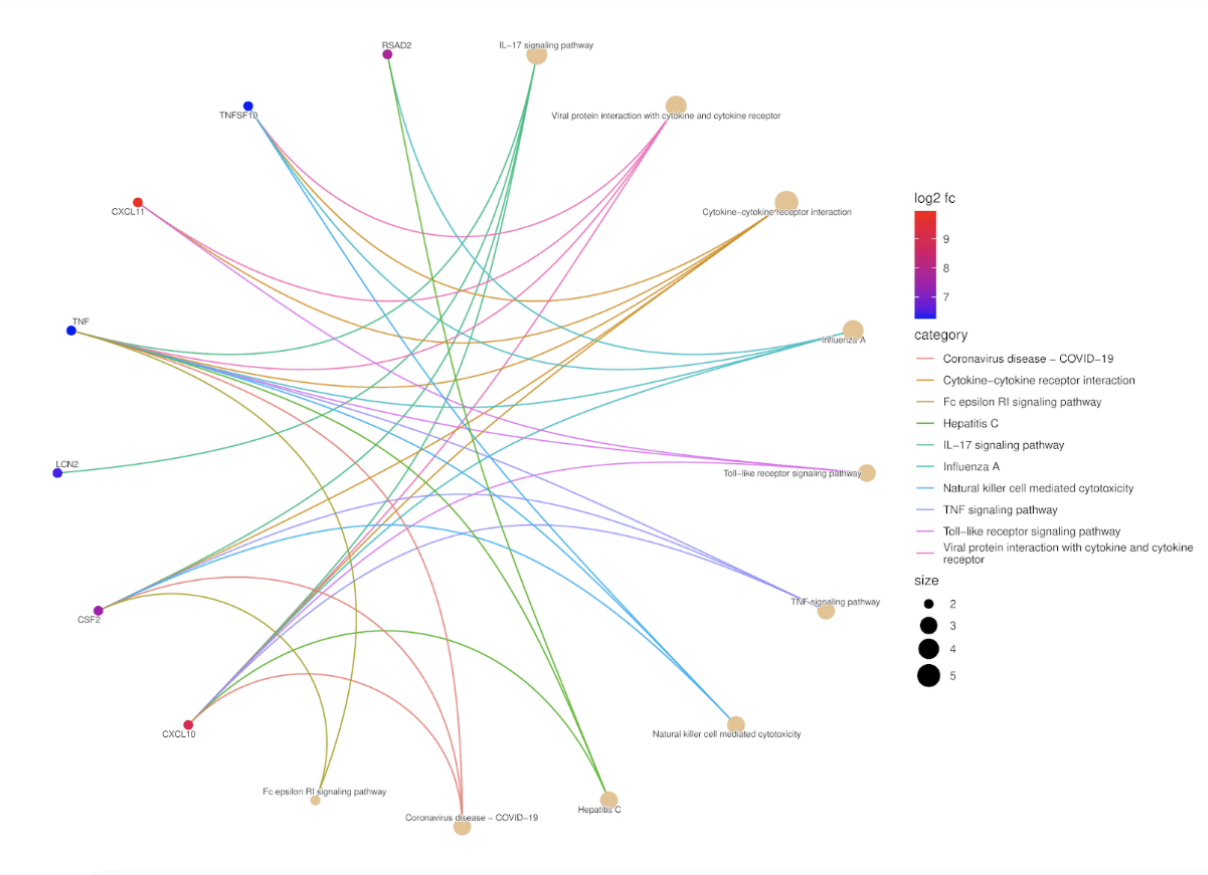


Figure 6. Kegg Pathway Analysis – Cnet Plot. The circles on the outside of the web represent genes, and genes with a higher Log2 fold change are redder and more significant. The most significant genes are CXCL10 and CXCL11 which are associated with the pathways of viral protein interaction with cytokine and cytokine receptor, cytokine-cytokine receptor interaction, toll-like receptor signaling pathway, coronavirus disease – COVID-19, Hepatitis C, TNF signaling pathway, Influenza A, and IL-17 signaling pathway. Genes with a lower Log 2 fold change are more blue and less significant.

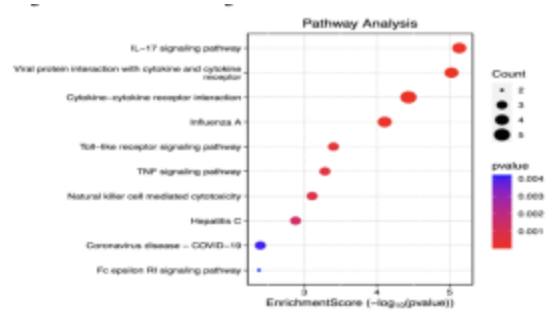


Figure 7. Kegg Pathway Analysis – Enrichment Score Dotplot. The circles on the graph represent biological pathways, and the circles with a lower p-value are more red and further to the right (more significant) while those with a higher p-value are more blue and further to the left.

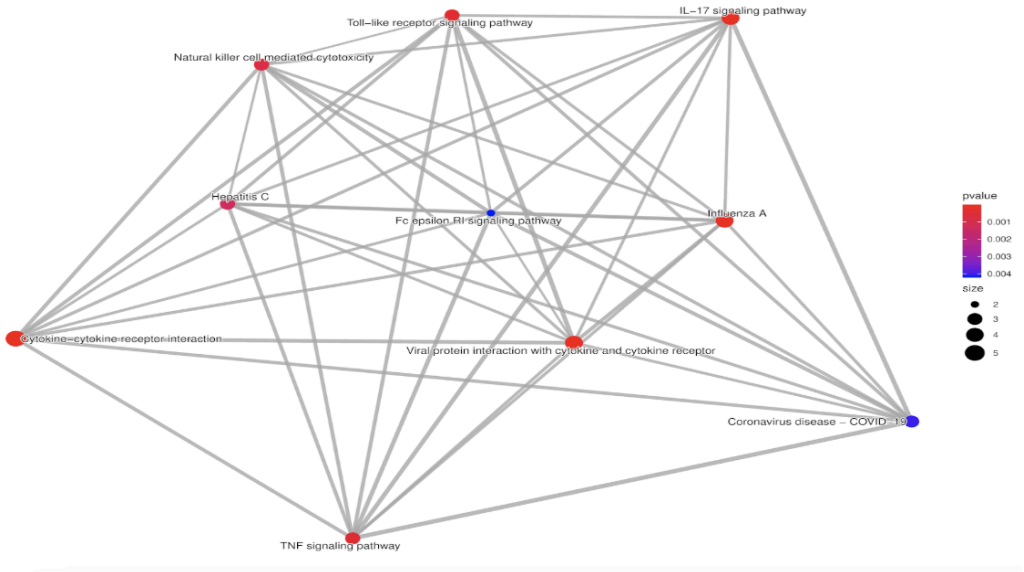


Figure 8. Kegg Pathway Analysis – Emap Plot. The circles connecting the lines represent biological pathways, and the circles with a lower p-value are redder and more significant. The most significant pathways are cytokine-cytokine receptor interaction, IL-17 signaling pathway, toll-like receptor signaling pathway, natural killer cell mediated cytotoxicity, Hepatitis C, Influenza A, Viral protein interaction with cytokine and cytokine receptor, and TNF signaling pathway. The circles with a higher p-value are more blue and less significant.

Key findings related to the top genes, functions and biological pathways that are potentially linked to Glaucoma disease are summarized in Table 1. The

key gene functions were related “Response to lipopolysaccharide,” “Cellular response to lipopolysaccharide,” “Viral protein interaction with cytokine and cytokine receptor,” and “Cytokine-cytokine receptor interaction.” Among the top common genes identified were CXCL11, CXCL10, RSAD2, and CSF2, which are known to play critical roles in immune response and inflammation. These results provide a deeper understanding of the molecular mechanisms at play and identify potential targets for further research and therapeutic intervention.

Table 1. Summary Table. Table shows the most significant gene functions and biological pathways, along with most significant common top genes

Gene Functions	**Biological Pathways**	**Common Top Genes**
Response to lipopolysaccharide	IL-17 signaling pathway	CXCL11
Response to molecule of bacterial origin	Viral protein interaction with cytokine and cytokine receptor	CXCL10
cellular response to lipopolysaccharide	cytokine-cytokine receptor interaction	RSAD2
cellular response to molecule of bacterial origin	Influenza A	CSF2

Supplementary Materials

Supplementary Materials 1. R Script used to generate DEG results using GEO2R Bioinformatics tool:R Script Used to Generate GEO2R Results

Supplementary Materials 2. Statistically significant top 30 DEGs:
Top DEGs

DISCUSSION

These results provide insight into the overall goal of this research study which asked, “which genes or genetic components are associated with the behavior of glaucoma and how can they assist in the understanding of this disease?” The gene enrichment analysis revealed enriched gene functions,

biological processes, and top common genes that could play significant roles in the behavior of the glaucoma disease.

The results in Figure 2 reveal both the most significant and most differentially expressed genes in a volcano plot. Figures 3-5 reveal many of the gene functions of the top 30 most differentially expressed genes, and these findings are relevant to the overall research question, as they could serve as potential biomarkers to assist in the understanding and treatment development of glaucoma. Among the gene functions, the biological processes (BP) category was the most enriched, suggesting its importance to the study. Additionally, figures 6-8 suggest several biological pathways surrounding the top 30 most differentially expressed genes, these findings again contributing to the research question by providing potential biomarkers of glaucoma and thus avenues to study. Figure 9 finally displays the most statistically significant gene functions, biological pathways, and top common genes. The top common genes were CXCL11, CXCL10, RSAD2, and CSF2.

One of the most significant gene functions was the "response to lipopolysaccharide". According to existing scientific literature in this field, lipopolysaccharide (LPS) can potentially result in an acute inflammatory response, releasing a large quantity of inflammatory cytokines in many different types of cells (8). This inflammatory response may relate to inflammation in the eye, which could damage the trabecular meshwork and result in increased IOP, affecting glaucoma (6). The gene functions of "response to molecule of bacterial origin" and "viral protein interaction with cytokine and cytokine receptor" also may contribute to inflammation in the eye in a similar fashion, according to existing scientific literature (6). Furthermore, among the most significant common genes—CXCL11, CXCL10, RSAD2, and CSF2—while there has yet to be extensive study concerning these genes' relation to glaucoma, all of these are related to inflammatory or neuroinflammatory processes (9, 10, 11, 12). CXCL11 is upregulated in cancer-associated fibroblasts, which result from inflammation (9). CXCL10 has been associated with myeloproliferative neoplasms in mouse models, which is excessive growth of blood cells (10). RSAD2 is upregulated among systemic lupus erythematosus (SLE) patients, which involves inflammatory processes (11). CSF2 is associated with the immune inflammation that is associated with Moyamoya Disease (12).

As shown in the results, many of the upregulated genes among those with higher IOP are associated with inflammation and inflammatory processes. The notion of inflammation in the human body as it relates to glaucoma is significant in the real world as it offers a genuine field for further study. A better understanding of the glaucoma-related inflammation could expedite the development of a glaucoma medicine or vaccine. More specifically, the genes CXCL11, CXCL10, RSAD2, and CSF2 can serve as potential biomarkers for glaucoma, not only serving as avenues of future study but also helping to detect glaucoma earlier in patients, which increases the chance of successful treatment (1, 12).

One limitation in this study is the limited sample size, as this research study only considered a single GEO bioinformatics dataset from microarray experiments. This limited pool of data could influence the findings of this research study, so the genes identified in this study may need to be studied further before finalizing the results. The identified genes, gene functions, and biological pathways will thus certainly need to be further studied in either the laboratory or clinical environment in the future, before the development of treatment for glaucoma.

Based on the findings of this study, several genes, gene functions, and biological processes can be associated with the behavior of glaucoma, and understanding this information as it relates to genomics can certainly be studied further to improve treatment steps for patients of a disease such as glaucoma (13). Specifically, the genes CXCL11, CXCL10, RSAD2, and CSF2 and their relation to inflammatory processes surrounding glaucoma are avenues for future study. If tested in the laboratory by scientists or through clinical trials, these biomarkers could provide insights into the intricate behavior of glaucoma. However, one possible opportunity for future study would be further analyzing the neurological aspects of glaucoma, in addition to the inflammation-related aspects. Especially considering the role of the optic nerve in glaucoma and its proximity to the brain (14), the connection between glaucoma and neurological diseases is another field to be studied. Understanding if other processes are at work alongside glaucoma, such as a correlation with Alzheimer’s and Parkinson’s, serves as a potential area of bioinformatics study and could advance the development of more effective treatment plans for glaucoma.

Conclusion

In conclusion, 5248 total DEGs were identified across two groups. After narrowing these DEGs down to 30, the most significant gene functions notably included "response to lipopolysaccharide" which is related to the significant biological pathway "Viral protein interaction with cytokine and cytokine receptor" (8). Additionally, 4 top common genes were identified and upregulated among patients with higher IOP: CXCL11, CXCL10, RSAD2, CSF2. All of these genes are in some way related to inflammation, which supports the hypothesis of potential upregulated genes. Furthermore, these findings answer the research goal of improving understanding of this disease, by suggesting a possible opportunity for further research on glaucoma-related inflammation.

References

1. Aung, T., & Khor, C. C. (2016). Glaucoma genetics: Recent advances and future directions. The Asia-Pacific Journal of Ophthalmology, 5(4), 256-259. https://journals.lww.com/apjoo/fulltext/2016/07000/Glaucoma_Genetics__Recent_Advances_and_Future.7.aspx
2. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Fonzar, J., LePage, D., & Soboleva, A. (2011). NCBI GEO: Archive for functional genomics data sets–10 years on. Nucleic Acids Research, 39(Database issue), D1005-D1010. https://doi.org/10.1093/nar/gkq1184
3. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Fonzar, J., LePage, D., & Soboleva, A. (2013). NCBI GEO: Archive for functional genomics data sets–update. Nucleic Acids Research, 41(Database issue), D991-D995. https://doi.org/10.1093/nar/gks1193
4. Belmonte, M., Foster, A., & Jones, H. (2024). Increased CXCL10 (IP-10) is associated with advanced myeloproliferative neoplasms and its loss dampens erythrocytosis in mouse models. Experimental Hematology, 135, 104246. https://doi.org/10.1016/j.exphem.2024.104246
5. Gamm, D. M., & Albert, D. M. (2024). Optic nerve. Encyclopedia Britannica. https://www.britannica.com/science/optic-nerve

6. Gene Ontology Consortium. (2024). GO enrichment analysis. Gene Ontology Resource. https://geneontology.org/docs/go-enrichment-analysis/
7. KEGG: Kyoto Encyclopedia of Genes and Genomes. (2024). Kyoto University Bioinformatics Center. https://www.genome.jp/kegg/kegg1a.html
8. Li, H., Li, X., & Yang, Z. (2024). GM-CSF promotes the development of dysfunctional vascular networks in Moyamoya disease. Neuroscience Bulletin, 40(4), 451-465. https://doi.org/10.1007/s12264-023-01158-y
9. Liu, G., Wang, Z., & Xu, L. (2022). The LINC00152/miR-205-5p/CXCL11 axis in hepatocellular carcinoma cancer-associated fibroblasts affects cancer cell phenotypes and tumor growth. Cellular Oncology, 45(6), 1435-1449. https://doi.org/10.1007/s13402-022-00730-4
10. National Center for Biotechnology Information. (2024). Programs. NCBI. https://www.ncbi.nlm.nih.gov/home/about/programs/
11. "Genomics in Health Care." (2023). National Geographic.
12. Rao, K. N., Nagireddy, S., & Chakrabarti, S. (2011). Complex genetic mechanisms in glaucoma: An overview. Indian Journal of Ophthalmology, 59(Suppl1), S31-S42. https://doi.org/10.4103/0301-4738.73685
13. Selvan, H., Patel, H., & Williams, K. (2022). Juvenile-onset open-angle glaucoma – A clinical and genetic update. Survey of Ophthalmology, 67(4), 1099-1117. https://doi.org/10.1016/j.survophthal.2021.09.001
14. Tang, D., Xie, Z., & Zhou, X. (2023). SRplot: A free online platform for data visualization and graphing. PLOS One, 18(11), e0294236. https://doi.org/10.1371/journal.pone.0294236
15. Tucureanu, M. M., Enciu, A. M., & Popescu, B. O. (2017). Lipopolysaccharide-induced inflammation in monocytes/macrophages is blocked by liposomal delivery of Gi-protein inhibitors. International Journal of Nanomedicine, 13, 63-76. https://doi.org/10.2147/IJN.S150918
16. Wiggs, J. L. (2007). Genetic etiologies of glaucoma. Archives of Ophthalmology, 125(1), 30-37. https://doi.org/10.1001/archopht.125.1.30

17. Wiggs, J. L., & Pasquale, L. R. (2017). Genetics of glaucoma. Human Molecular Genetics, 26(R1), R21-R27. https://doi.org/10.1093/hmg/ddx183
18. Xin, Y., Lu, X., & Huang, J. (2023). Interferon-α regulates abnormally increased expression of RSAD2 in Th17 and Tfh cells in systemic lupus erythematosus patients. European Journal of Immunology, 53(8), e2350420. https://doi.org/10.1002/eji.202350420
19. Zhang, Y., Song, M., & Lei, Y. (2022). TIPARP is involved in the regulation of intraocular pressure. Communications Biology, 5(1), 1386. https://doi.org/10.1038/s42003-022-03882-1
20. "How to Employ Statistical Approaches to Identify Differentially Expressed Genes (DEG)." (2023). Novogene. https://www.novogene.com/us-en/resources/blog/how-to-employ-statistical-approaches-to-identify-differentially-expressed-genes-deg/

Kamya Kainth

Using Bioinformatics Tools to Analyze the Differentially Expressed Genes of Treated Triple-Negative Breast Cancer Cells

ABSTRACT

Background: Breast cancer is a global disease that affects millions of women (those assigned female at birth) around the world. It is caused by the build-up of cancerous cells in the breast, which can be due to a number of factors, including radiation, genetic mutations, and oestrogens. Specifically, the cause of Triple-Negative Breast Cancer (TBNC), a subtype of breast cancer with a low survival rate, remains unclear to many scientists. To fill this research gap, this study further explored the effects of optimal medication trials on differentially expressed genes of Triple-Negative Breast Cancer cells.

Methods: This study used various bioinformatics tools and databases, namely, NCBI GEO, SR Plot, GO, and KEGG. First, the GEO bioinformatics database was used to find and collect data from GEO Dataset ID GSE264630, which focuses on paclitaxel-resistant triple negative breast cancer cells treated with various medications including DMSO and LLY283. The dataset was then divided into three groups based on these different treatment plans: a control group, the DMSO group, and the LLY283 group. A GEO2R analysis of this dataset resulted in thousands of differentially expressed genes (DEGs) that were narrowed down to the top 40 DEGs based on the lowest p-values. Finally, to determine the functions and associated biological pathways of the top 40 DEGs, SR Plot bioinformatics tool was used to perform gene ontology (GO) and KEGG pathway analyses.

Results: GEO2R analysis of the GEO dataset resulted in a total of 26,567 genes. Among these genes, 42 were differentially expressed (DEGs) between all the groups. The 40 top DEGs were narrowed down using the lowest p-values. Some enriched gene ontology terms after the GO analysis included eicosanoid receptor activity, G protein-coupled peptide receptor activity, peptide receptor activity, and protein-ADP ribosylase activity. Enriched KEGG pathways included neuroactive ligand-receptor interaction, alpha linoleic acid metabolism, and linoleic acid metabolism. Top common genes included LTB4R, LTB4R2, PARP12, PLAG12, and HTT.

Conclusion: Looking specifically at the molecular function of the top genes, such as LTB4R, LTB4R2, PARP12, and PLAG12, all play different roles

in cancer. Some act as tumor suppressors while others act as inhibitors of biomarkers in cancer cells. Expression of these genes in the TBNC samples can reveal the effectiveness of the specified treatments.

INTRODUCTION

Breast cancer is the leading type of cancer that impacts 12% of women in the United States. Around 2.089 million women were diagnosed with breast cancer in 2018 and cases continue to rise (Smolarz, Beata, et al., 2022). At its core, breast cancer forms from a collection of cells that turn cancerous and build up to form tumors or masses. Some risk factors include oestrogens, late menopause, and high concentrations of endogenous oestradiol (Key, Timothy J, et al., 2001). Approximately 90% of breast cancer cases in women are caused by gene alterations obtained during their lifetimes. Among many gene mutations, key alterations that form in cancer cells are in the tumor suppressor genes (BRCA genes) and proto-oncogenes (Advanced Journal of Chemistry). Breast cancer is subcategorized into three main types based on the molecular markers for estrogen or progesterone receptors and human epidermal growth factor 2 (ERBB2). This includes HER2 negative, HER2 positive, and triple-negative, which indicates the lack of all three standard molecular markers in the tumor cells. This research study focuses on the latter, triple-negative blood cancer (TBNC) (Breast Cancer Treatment 2), which constitutes fifteen to twenty percent of breast cancers. Those diagnosed with TNBC typically live for around one year. (Smolarz, Beata, et al., 2022).

In the past, researchers have used GEO databases as a way to find key biomarkers for breast cancer. In one particular study, ten Hub genes were identified (PBK, CCNA2, CDCA8, MELK, NUSAP1, BIRC5, CCNB2, HMMR, MAD2L1, and PRC1) that are overexpressed in cancerous tissue, specifically from an individual with triple-negative breast cancer (Yang, Kaidi, et al., 2019). The critical roles these differentially expressed genes play can allow for earlier detection of breast cancer.

One challenge of Triple-Negative Breast Cancer is that it remains extremely hard to treat due to the lack of estrogen receptors and HER2 amplification targets. Therefore, scientists are currently working on creating personalized treatments for patients by focusing on the biomarkers present rather than

the three molecular markers it lacks (Leon-Ferre, Roberto A, et al., 2023). Some limitations the science community still faces is creating targeted therapies without the target receptors, like estrogen, progesterone, and HER, present. Triple-negative blood cancer or TBNC is highly heterogeneous, meaning that it has multiple subtypes with multiple molecular profiles. The heterogeneity of this cancer requires deep research to formulate more helpful treatments for patients (Obidiro, Onyinyechi, et al., 2023). Some promising approaches include targeting cancer energy metabolism and targeting protein to protein signaling. These approaches can be used by researchers to find novel therapeutics (Chapdelaine et al., 2023). This research aims to bridge this gap by using bioinformatics to contribute to finding more novel therapeutics.

The goal of this research is to identify differentially expressed genes within a GEO dataset and determine the potential functions and biological pathways they are associated with as potential targets for novel therapeutics. It is hypothesized that there will be a significant difference between gene expression in the control group, DMSO group, and the LLY283 group because various treatments have been added to the samples. For more context, DMSO or dimethyl sulfoxide is a pharmaceutical solvent for cell therapies (Hoang, Ba X et al., 2023). LLY283 is a drug used for antitumor purposes as a inhibitor of protein arginine methyltransferase 5 (PRMT5) (Feustel, et al., 2022). The dataset used in this study focuses on the role of PRMT5 inhibitors and other anticancer agents on paclitaxel-resistant triple negative breast cancer cells. The information can be used to further compare the effects of various medications on triple-negative breast cancer cells.

As mentioned previously, triple-negative breast cancer is a difficult type of cancer to cure. With 13 in every 100,000 women in the United States being diagnosed (Triple-Negative Breast Cancer: Facts and Stats), it is important that this specific type of breast cancer is further explored. Currently, there are not many personalized medication plans that work on triple-negative breast cancer. For that reason, this research will contribute to important differentially expressed genes that can point towards optimal medication to target the cancer cells effectively.

METHODS

Data Collection

The overall workflow of the research is summarized in Figure 1. First, using the NCBI GEO dataset, we identified a fitting dataset on triple-negative breast cancer. NCBI is a website created by researchers and programmers to organize public databases of genes, transcripts, and proteins (http://ncbi.nlm.nih.gov). It provides standard reference sequences and descriptions of different genes (Pruitt, Kim D, et al., 2024). The data set GSE264630 was chosen, whose primary research specifically focused on paclitaxel-resistant triple breast cancer. Paclitaxel is a well known anticancer agent that inhibits tumor growth through the blockage of cell division in the G2 and M phase (Xu, Xiangwei, et al., 2022). Epigenetic chemical probe screen data uncovered that these resistant cancer cells are vulnerable to inhibition of protein arginine methyltransferases (PRMTs) through cell division stabilizing. PRMT5 is known to promote tumor cell growth early on in the diagnosis (Shailesh, Harshita et al., 2018). For this dataset, researchers conducted RNA-sequencing on the paclitaxel-resistant cells with PRMT5 inhibitors including LLY283 and GSK591.

Defining Groups

In order to analyze this dataset, the GEO2R bioinformatics tool was used. This bioinformatics tool allows individuals to compare two or more groups in the Gene Expression Omnibus (GEO) series in order to identify patterns in expressed genes and produce analytical graphs (Clough et al., 2016). Groups were assigned based on the treatment type applied to the cancerous cell. Of the samples given, 2 samples were defined as controls, 10 samples were defined as the LLY283 group, and 8 samples were defined as the DMSO group. To reiterate, DMSO is an anticancer drug that inhibits tumor proliferation (Villarroel, Ariana, et al., 2020). Since it is not one of the treatments that act as PRMT5 inhibitors, it will act as a second control for parts of the analysis. LLY283, an inhibitor for the PRMT5 protein, is the treatment for the experimental group. It was decided not to include GSK591 in the analysis as it had less correlation with the other data and thus could obscure trends. Once the groups were defined, they were analyzed through GEO2R. This resulted in statistical analysis compiled into graphs and tables as well as a list of the top differentially expressed genes within the dataset. These results were generated using the R programming language with pre-programmed AI and ML algorithms (R-script of GEO Dataset).

Narrowing down the Top Differentially Expressed Genes

It is crucial to narrow down the list of top DEGs before putting it into another bioinformatics tool, such as SR plot (www.bioinformatics.com.cn/srplot). To do this, the raw list of genes in the dataset and their accompanying information was downloaded into a Google Sheet. The data came with associated p-values, which are important because they describe the probability of the results happening by chance. Therefore, for data to be more significant, the p-value needs to be low. The list of DEGs was narrowed down to the 40 genes that had the lowest p-values (ranging from 2.37E-21 to 2.95E-08) (DEGs for Breast Cancer Study).

Using SR plot Bioinformatics Tool for Gene Ontology Analysis and KEGG Pathway

Once the list was narrowed down to the top 40 (DEGs for Breast Cancer Study), it was inputted into a bioinformatics tool called SR plot. SR plot is a free online bioinformatics platform used for data collection and visualization pertaining to enrichment and functional analysis (Tang, Doudou, et al., 2023). In SR plot, the data was analyzed through the GEO pathway enrichment analysis for Gene Ontology Analysis and KEGG pathway. Gene Ontology is a bioinformatics tool used to categorize genes into different groups based on their description/function ("Gene Ontology Overview."). The three independent branches are descriptions of the genes' molecular function (MF), their location in cellular components (CC), and biological processes (BP) they contribute to (Dameron, O., Bettembourg, C., Meur, N. L.). Similarly, the KEGG pathway is a bioinformatics collection of databases with genomes, biological pathways, diseases, drugs, and substances from the Kyoto Encyclopedia of Genes/Genomes that depict diagrams of processes like metabolic and signaling pathways (Kegg Pathway Map).

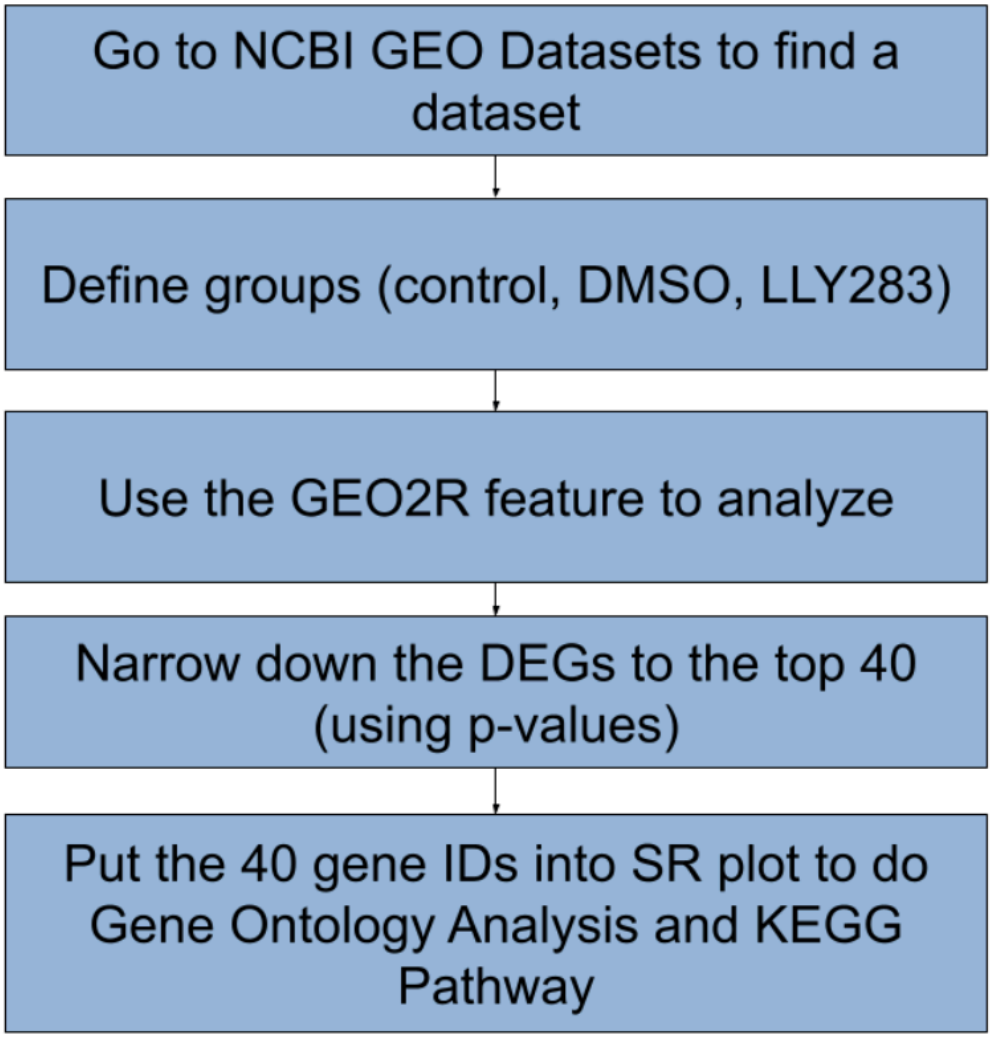


Figure 1. Summary of method steps used in this study. Various bioinformatics databases and tools were used in this study. NCBI GEO was used to collect data and identify differentially expressed genes. SR Plot was used to determine the potential gene functions and associated biological pathways of the key top DEGs.

RESULTS

Differences and similarities between the top DEGs from the GEO2R Bioinformatics Tool on NCBI

In this study, GEO2R was implemented to highlight the top differentially expressed genes. Differentially expressed genes (DEGs) are genes that show a statistically significant change in expression levels, typically measured by

mRNA or protein production in microarrays, in response to varying experimental conditions such as time, temperature, disease state, tissue location, and other factors. R-script programming pre-programmed by AI and ML was used to identify top DEGs using GEO2R bioinformatics tools. The R-script used in this study is linked here: R-script of GEO Dataset

Results in **Figure 2** depicts volcano plots showing relationships between the top differentially expressed genes across the groups. Each dot represents a gene that is expressed differently between treatment and control groups. The red dots indicate genes that are upregulated and the blue dots indicate genes that are downregulated when compared with the control. Gray dots are genes that were expressed the same in both groups.

Between the control group and DMSO, 2,080 DEGs were upregulated and 3,257 DEGs were downregulated from a total of 5,337 DEGs **(Figure 2a).** Between the DMSO and LLY283 group, 161 DEGs were upregulated and 391 DEGs were downregulated from a total of 552 DEGs **(Figure 2b)**. Between LLY283 and control, 31 DEGs were upregulated and 3 DEGs were downregulated from a total of 34 DEGs **(Figure 2c).** The overlapping Venn Diagram **(Figure 3)** shows common significant genes amongst the different groups with 42 common genes among all the three groups (control, DMSO, and LLY283). There are a total of 26,567 DEGs in total.

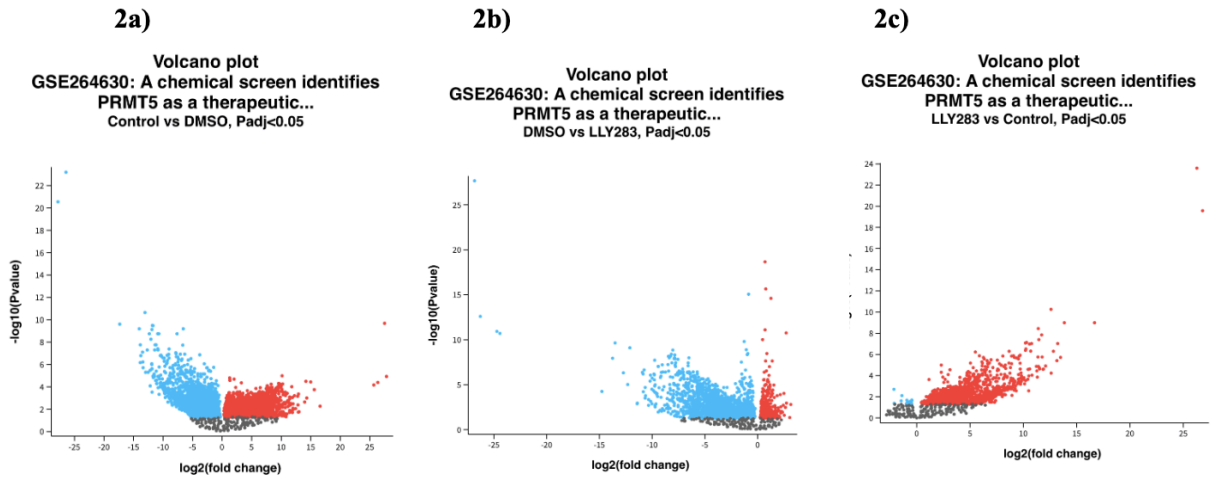


Figure 2. Volcano plots compare the expression of genes between two specific treatment/control groups. Red signifies upregulation and blue signifies downregulation.

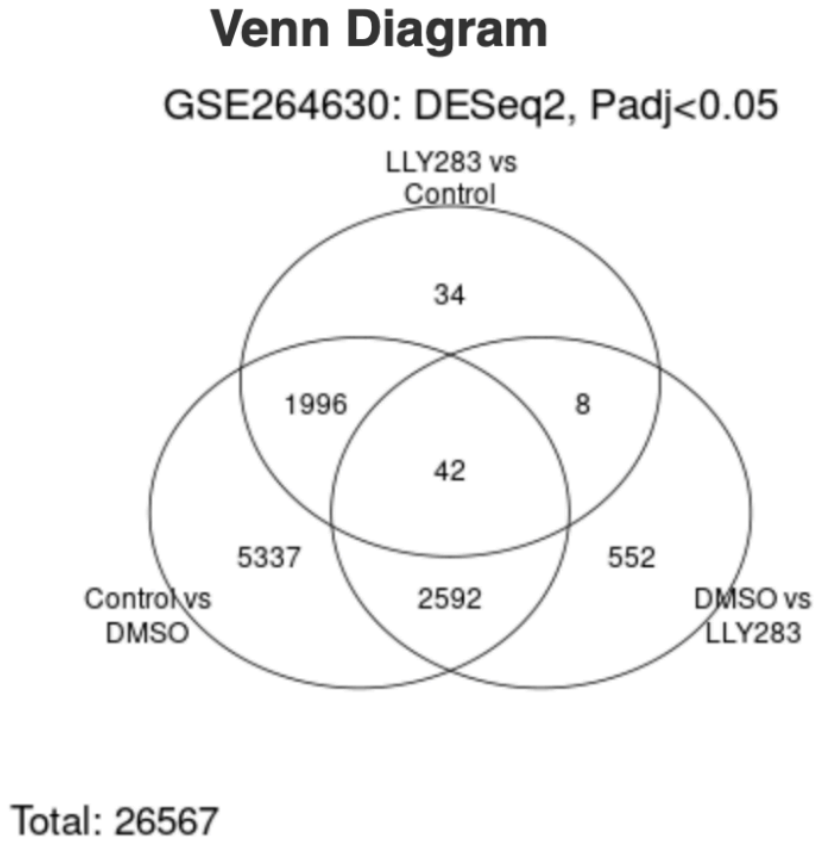


Figure 3. The Venn Diagram describes the common significant genes between the different treatment groups.

Top 40 DEGs analyzed using SR plot Gene Ontology Analysis

The GO results reveal increased enrichment of biological processes, specifically the positive regulation of calcium-mediated signaling, and molecular function, specifically icosanoid receptor activity (Figure 4). Since the molecular function enrichment score is the only one above 3, it will be focused on in this study. Figure 5 connects significant MF genes with related functions. Gene PARP12 connects to protein- ADP ribosylase activity. Gene PLA2G12A is associated with calcium-dependent phospholipase A2 activity. Gene GSTM4 is related to carbon-sulfur lyase activity, oligopeptide binding, and glutathione binding. Gene HTT is associated with profilin binding and

dynactin binding. Gene LTB4R is connected to icosanoid activity, G protein-coupled peptide receptor activity, and peptide receptor activity. Gene LTB4R2 is connected to G protein-coupled peptide receptor activity and peptide receptor activity.

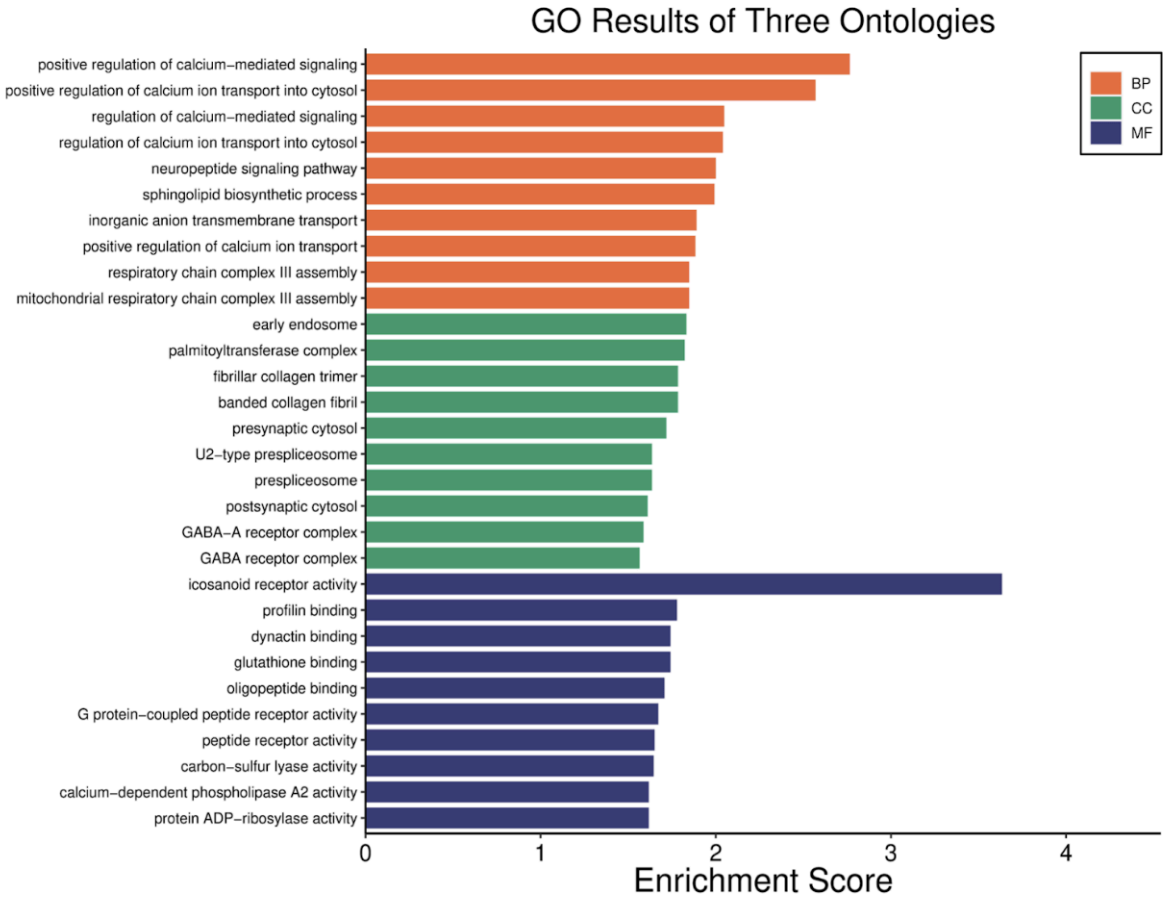


Figure 4. Biological process (BP), cellular component (CC), and molecular function (MF) are the three ontologies on this graph. It shows that the top 40 DEGs are the most enriched in BP and MF . The enrichment score threshold was 3 or above.

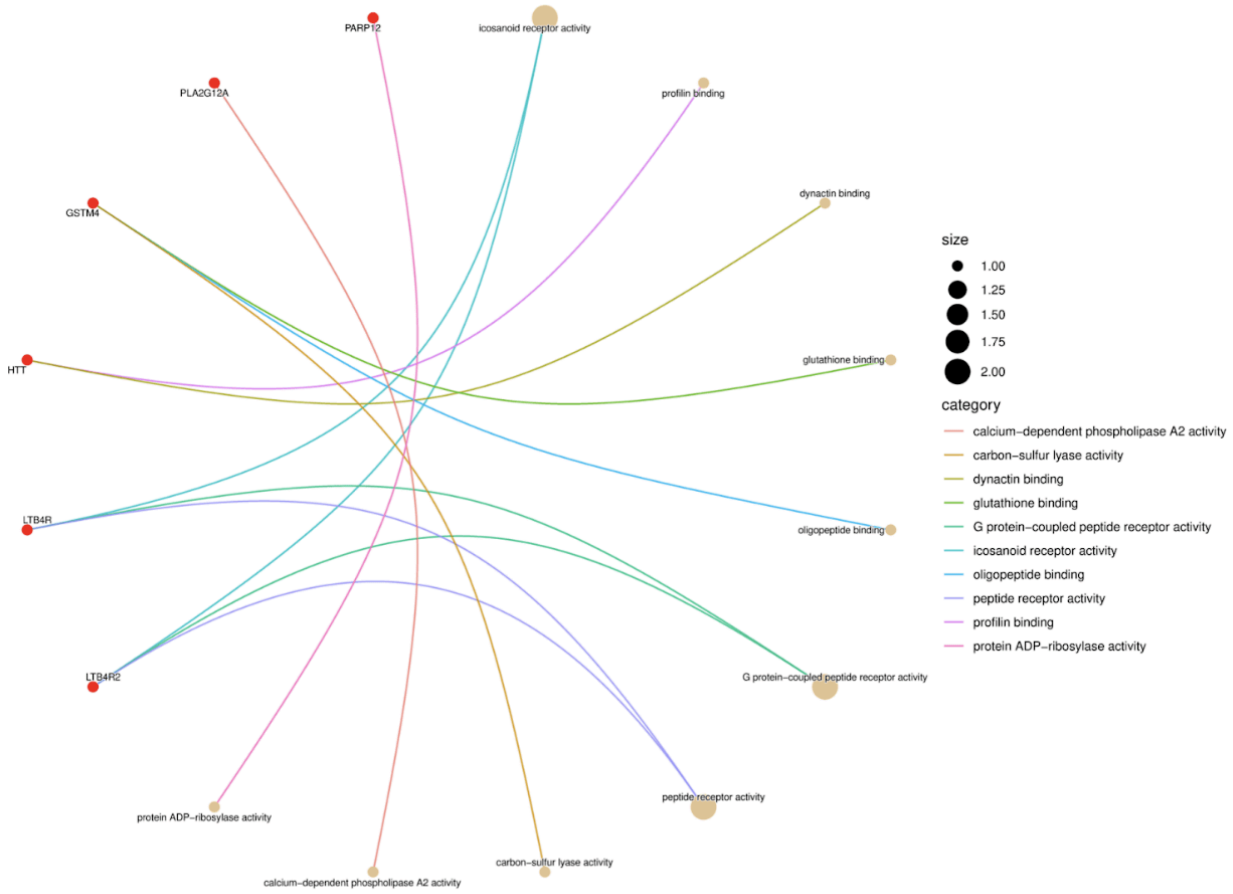


Figure 5. Cnet plot that connects key genes and their relation to significant activities and functions. This plot focuses specifically on molecular function. Gene PARP12: protein- ADP ribosylase activity. Gene PLA2G12A: calcium-dependent phospholipase A2 activity. Gene GSTM4: carbon-sulfur lyase activity, oligopeptide binding, and glutathione binding. Gene HTT: profilin binding and dynactin binding. Gene LTB4R: icosanoid activity, G protein-coupled peptide receptor activity, and peptide receptor activity. Gene LTB4R2: G protein-coupled peptide receptor activity and peptide receptor activity.

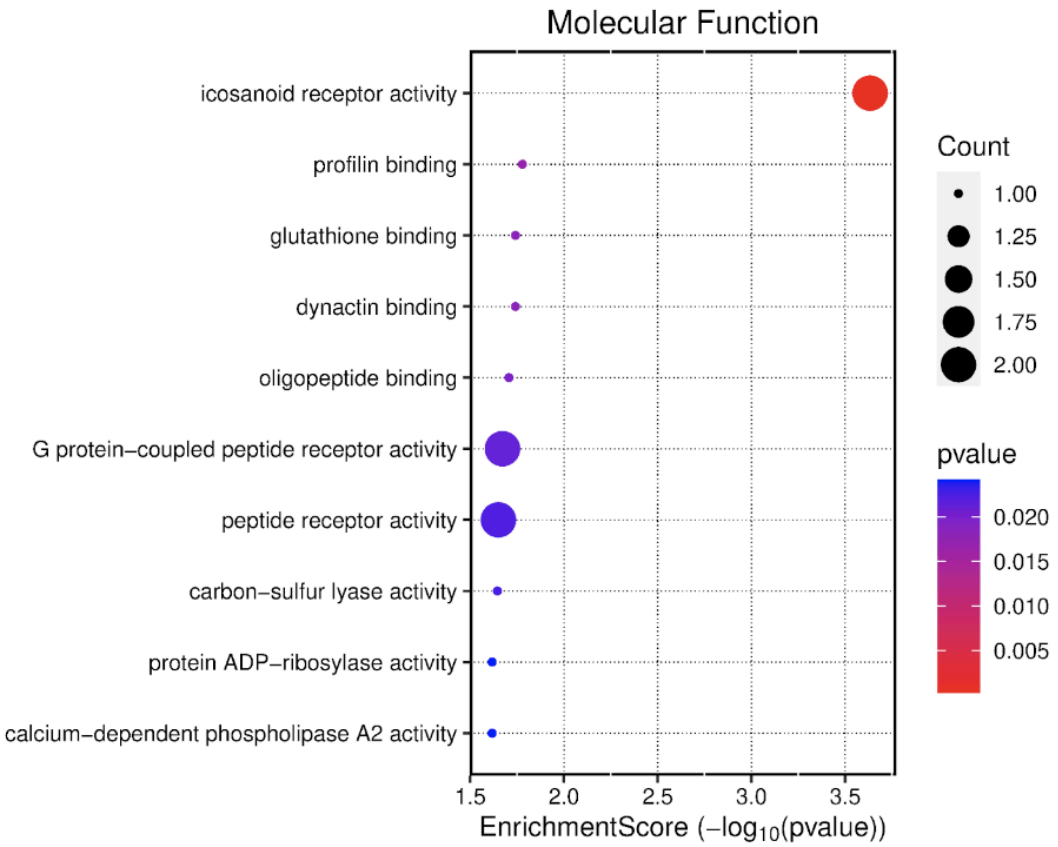


Figure 6. In this dot plot on molecular function ontology, blue/purple dots indicate insignificant data due to the high p-value, while big, red dots are considered the most significant. Icosanoid receptor activity is the only significant molecular function that is enriched in the top 40 DEGs.

KEGG Pathway of Top 40 DEGs

Figure 7 shows that the most enriched pathways are neuroactive ligand-receptor interaction, alpha-linolenic acid metabolism, and linoleic acid metabolism. Among the three pathways, neuroactive ligand-receptor interaction is the most enriched with an enrichment score of around 2 **(Figure 9)**. As seen in **Figure 8**, the main genes involved in the neuroactive ligand-receptor interaction are LTB4R and GABR.

Specific Gene Analysis through GEO2R

Figures 10, 11, and 12 are all detailed charts on the expression of one particular gene in the different experimental groups. PLA2G12A is expressed lower in the control group as well as the LLY283 group and the highest in the DMSO group. The highest expression value is around 27 in the DMSO group

and the lowest expression value is 5 in the LLY283 group **(Figure 10).** From the lowest expression of GSTM4 to the highest expression of GSTM4, the groups are control, LLY283, and DMSO. The highest expression value is around 24.5 in the DMSO group and the lowest expression value is 5 in the control group and LLY283 group **(Figure 11)**. **Figure 12** describes the expression of both LTB4R and LTB4R2. In both cases, DMSO samples have the lowest expression of the gene and LLY283 groups have the highest. The highest expression value is 12 and the lowest is 1 **(Figure 12a).**

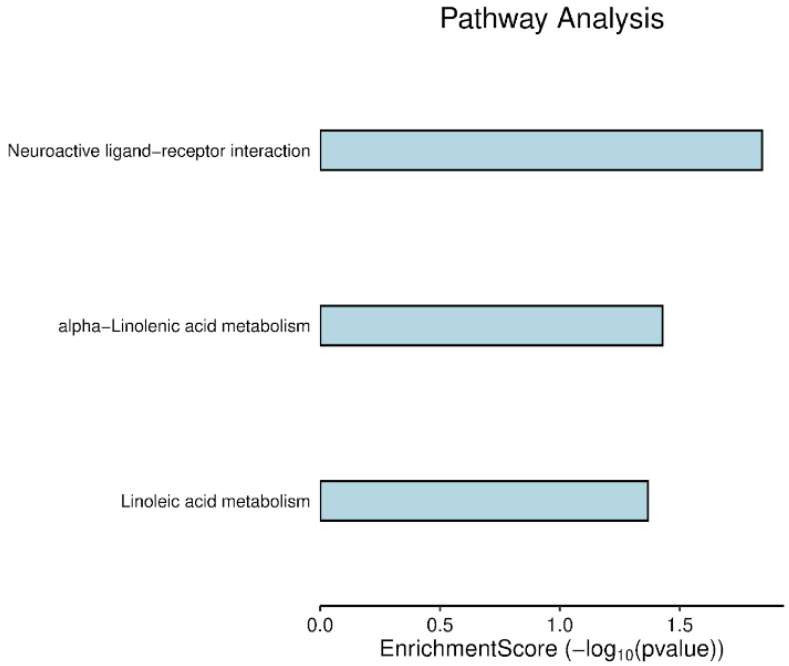


Figure 7. The top 40 DEGs are most enriched in neuroactive ligand-receptor interaction.

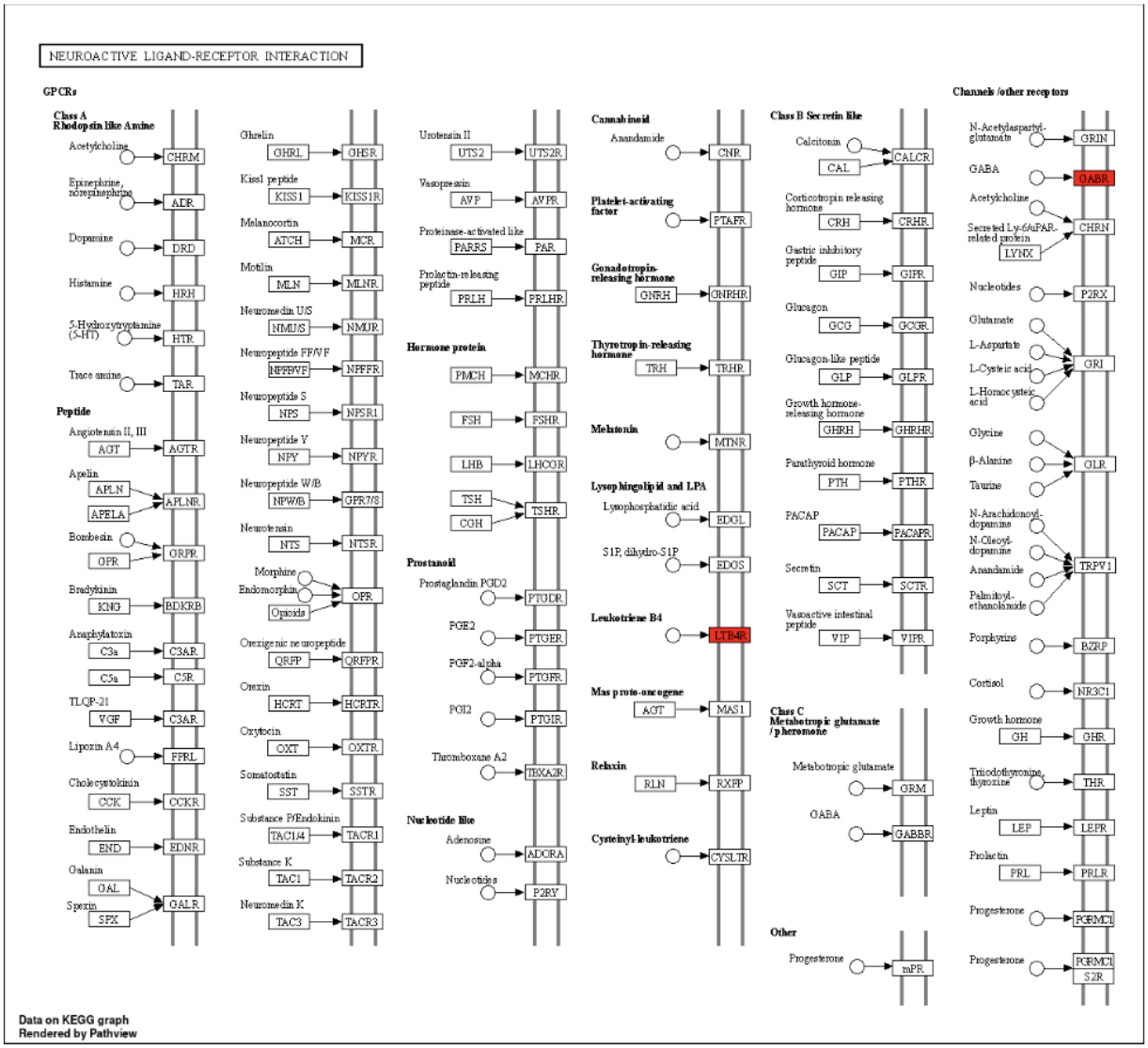


Figure 8. Pathway enrichment diagram of neuroactive ligand-receptor interaction dealing with specific genes like LTB4R and GABR.

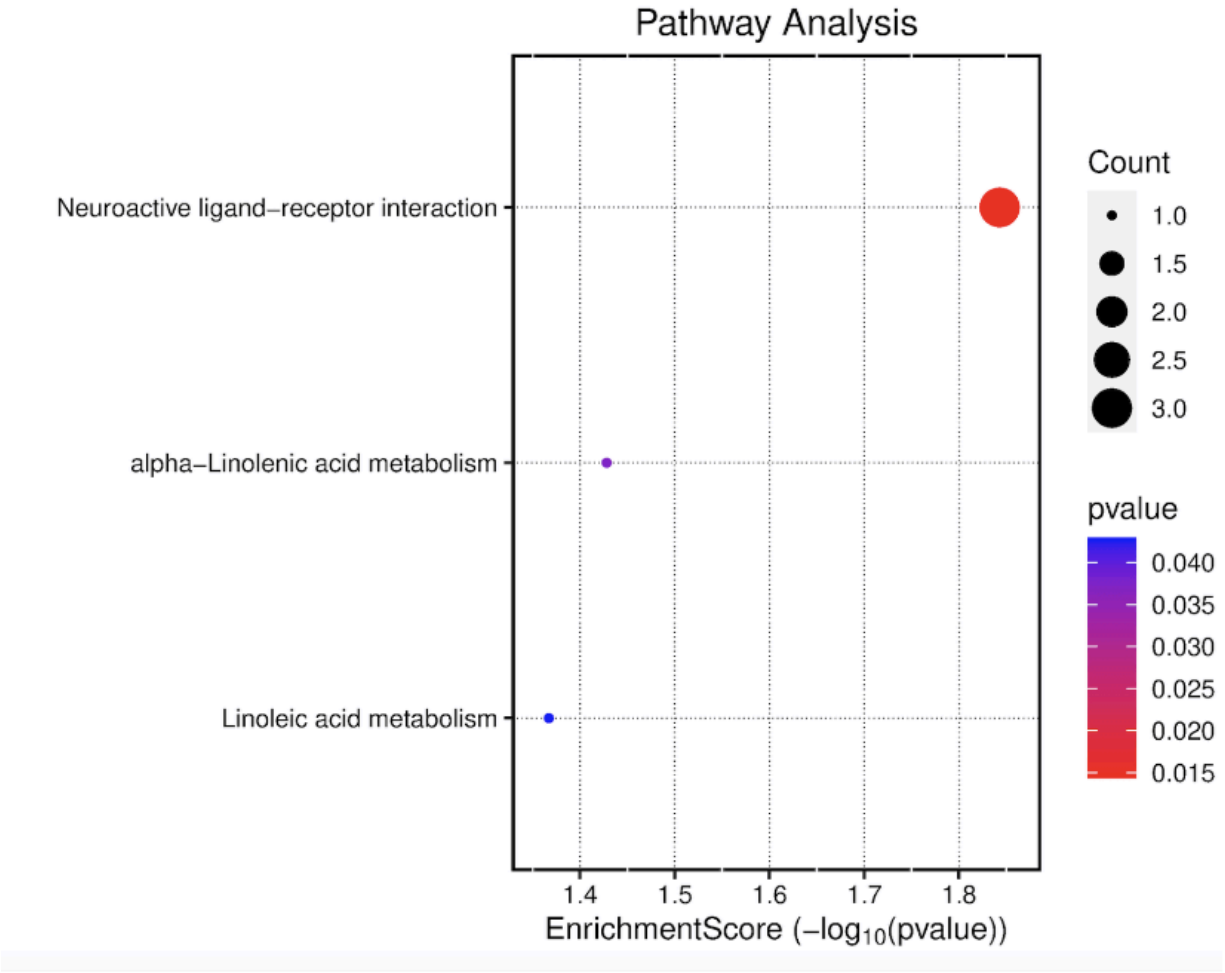


Figure 9: Neuroactive ligand-receptor interaction most highly enriched in the top 40 DEGs. The red dots are considered significant while purple/blue are insignificant.

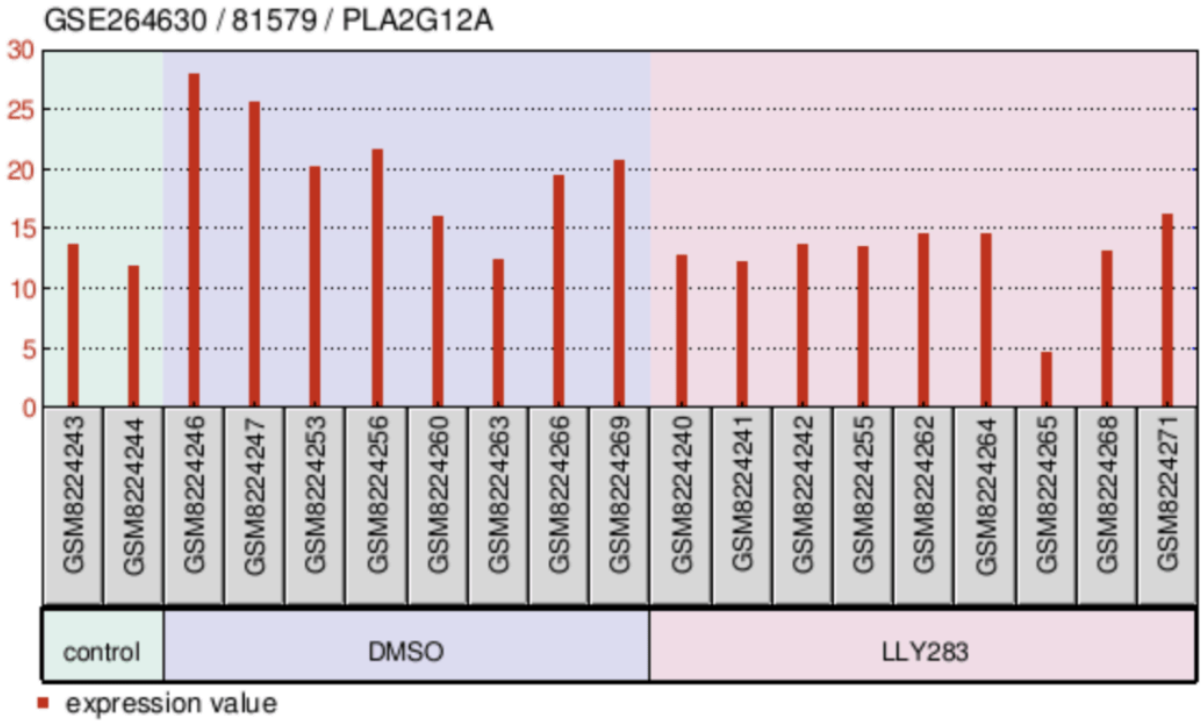


Figure 10. Expression levels of gene PLA2G12A in the samples of the GEO Dataset.

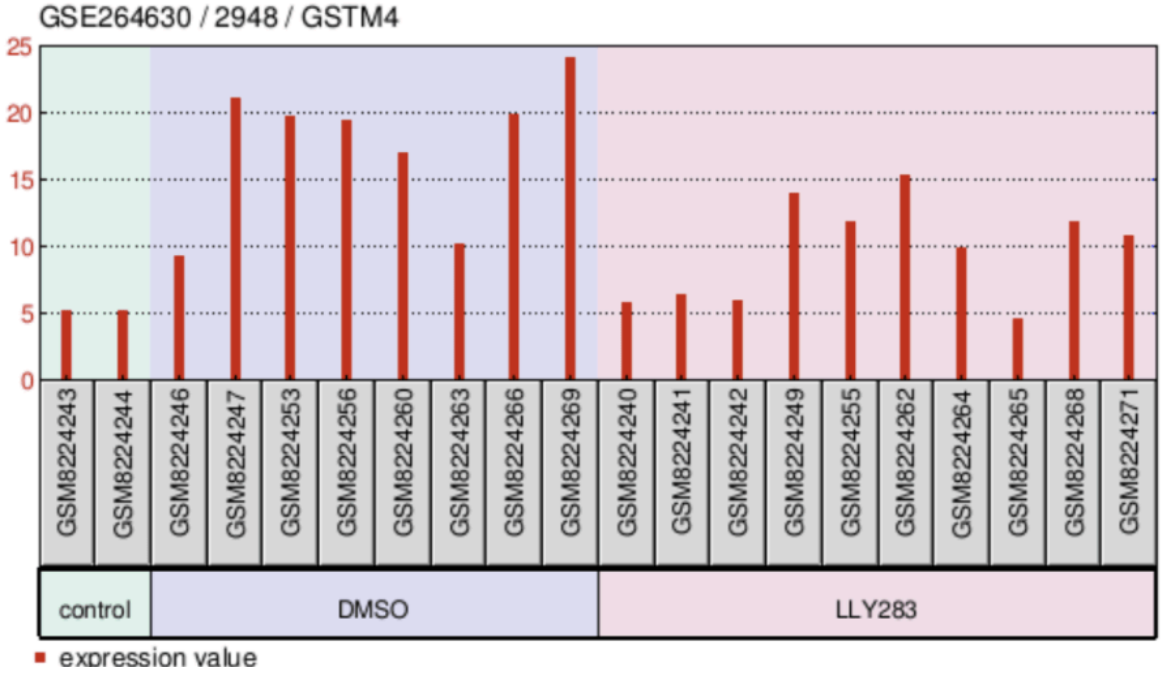


Figure 11. Expression levels of gene GSMT4 in the samples of the GEO Dataset.

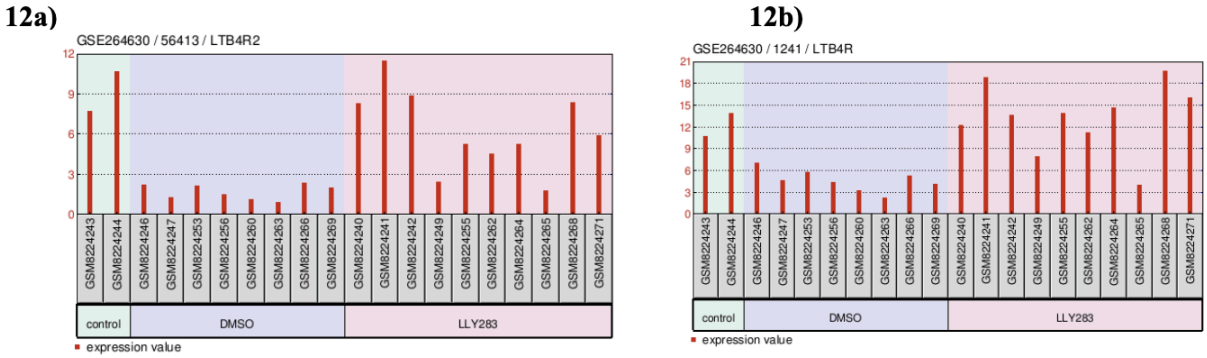


Figure 12. Expression levels of gene LTB4R and LTB4R2 in the samples of the GEO Dataset.

Supplementary Materials
R Script: R script Used to Generate GEO2R Results
Top DEGs: DEGs for Breast Cancer Study

DISCUSSION

The main goal of this study is to identify top differentially expressed genes that are enriched in the samples of the GEO dataset, GSE264630. Through GEO2R analysis and SR plot gene enrichment analysis, we narrowed down the top 40 DEGs to less than 10 genes that were significant in the samples. Gene ontology revealed the enrichment of molecular function, specifically with the genes LTB4R, LTB4R2, PARP12, GSTM4, PLA2G12A, and HTT. Further research into these genes and their related pathways allowed for a further refinement of top genes.

PARP12 is a tumor suppressor gene that interacts with FHL2, which is one of its functional partners. Through a study performed by Shao et.al, PARP12 deficiency correlates to a decrease in FHL2 levels, which promotes the migration and infiltration of hepatocellular carcinoma (Shao, Changjuan, et al., 2018). It is related to protein-ADP ribosylation activity as a member of the ADP-ribosyl transferase family. PARP12 works by blocking mRNA translation, typically during DNA correction, in order to suppress mutations. Therefore, it can also work as a cellular defense against viral infections (Welsby, lain, et al., 2014). Current research articles find no direct linkage between PARP12 and a specific type of cancer (Scarpa, Emanuele S., et al., 2013). However, as the data from **Figure 5** suggests, PARP12 does carry a linkage and correlation to breast cancer.

PLA2G12A is another prominent gene in this study. Although there are mixed conclusions from other studies on this gene, some articles correlate this gene with protection against diet-induced obesity and insulin resistance (Wu, Min, et al., 2024). In addition, the connected calcium-dependent phospholipase A2 activity has functions in relation to neurotransmission and inflammatory signaling (Wang, Shaowei et al., 2022). However, on a less specific level, phospholipase A2 activity, which involves hydrolyzing of phospholipids and releasing of fatty acids and lysophospholipids, plays a role in colorectal cancer. Lower expression of PLA2G12A in this study led to tumor growth and recurrence (Parisi, Eva et al., 2023). **Figure 10** shows the lower expression of the gene in the control group, which is the untreated cancer cell, and the higher expression of the gene in the treatment groups. Since deficiency means greater tumor growth, the graph shows the treatment plans working, especially DMSO as the anticancer agent. Compared to LLY283 samples, DMSO samples had drastically higher levels of PLA2G12A, meaning less tumor growth overall.

Past studies indicate a strong association between GSMT4 and a type of bone cancer called Ewing's sarcoma. This gene creates proteins that play an important role in drug resistance. Lowered amounts of GSTM4 result in an increased susceptibility of the cancer cells to anticancer agents (Luo, W et al., 2009). **Figure 5** shows an association of carbon-sulfur lyase activity, oligopeptide binding, and glutathione binding to GSTM4. Carbon-sulfur lyase activity, which is defined as the act of an enzyme that breaks a carbon-sulfur bond, does not have much relation to activities of GSTM4 in regards to cancer ("Carbon-Sulfur Lyases-Mesh-NCBI"). Oligopeptide binding is the binding of the oligopeptide ligand to αvβ3 integrin receptor. This receptor is overexpressed in tumor cells, so the binding of a ligand to this receptor can increase the effectiveness of treatment plans and can also reduce possible side effects ("Oligopeptide."). Lastly, glutathione is involved with controlling cell differentiation, proliferation, and apoptosis. Mutations or an increase in glutathione expression is linked to tumor growth (Kennedy, Luke et al., 2020). **Figure 11** shows the approximate levels of GSTM4 genes in the different types of sample in the GEO Dataset. Based on these results, the control has lower levels, LLY283 has medium levels, and DMSO has high levels of GSTM4. This contradicts studies that state that lower levels of GSTM4 typically correlates with effective treatment plans (Luo, W et al.,

2009). If this were the case, then the treatment groups, LLY283 and DMSO would have the lower levels of GSTM4. However, it must be noted that the correlation between GSTM4 and cancer was mentioned with Ewing's sarcoma cells and not breast cancer cells (Luo, W et al., 2009).

Figure 5 indicates the importance of the HTT gene, but there is no direct linkage between this gene and cancer. HTT is a disease connected to Huntington's disease, which is a neurodegenerative disorder ("HTT Huntington"). However, a study describes a possible unexpected association of HTT and mutant HTT to tumor growth (Thion, Morgane Sonia, et al., 2018). Interestingly, the mutated HTT gene in Huntington's disease patients creates molecules toxic to cancer cells and thus leads to 80% less chance of cancer than the normal population (Shanley, Mathew, 2021). Profilin binding has a role in cell functions and the cell cycle. It is said to be dysregulated in types of cancer, making it a key biomarker (Thione et al., 2018). Dynactin binding regulates oncogene expression (Izdebska, Magdalena et al., 2020). However, there is no indication of whether the HTT gene is mutated or not in the GEO Dataset and therefore, there is very little clear association between HTT and cancer cells treated with different anticancer agents.

Lastly, LTB4R and LTB4R2 carry important significance in the samples and therefore breast cancer cells. LTB4R and LTB4R2 are leukotrienes B4 receptors 1 and 2 that promote metastasis or cell growth and are involved in tumor development. An earlier study found that TNBC cells have a lower survival rate in LTB4R expression groups (Kalinkin, A.l., et al., 2019). LTB4R and LTB4R2 associated pathways include eicosanoid activity, G protein-coupled peptide receptor activity, and peptide receptor activity. Eicosanoid is especially significant due to its mention in **Figure 4** and **Figure 6** in addition to **Figure 5**. Eicosanoid is a signal molecule that consists of many essential elements including leukotriene, which is a primary part of LTB4R. It plays a role in cell differentiation and cell growth, two big parts of cancer ("Eicosanoid."). As eicosanoids are elevated in patients with tumors, eicosanoid receptors are appealing therapeutics to work as anticancer agents (Johnson, Amber M., et al., 2020). Like eicosanoids, g-protein coupled receptors (GPCRs), which is a type of peptide receptor, play a role in cell proliferation and tumor metastasis. Overexpression of GPCRs leads to rapid tumor growth (Dorsam, et al., 2007). Interestingly, the control group and LLY283 group had drastically higher expressions of LTB4R

and LTB4R2 compared to the DMSO group **(Figure 12).** Moreover, since higher LTB4R/2 expression indicates greater tumor development, the DMSO group seems to be the optimal treatment plan to most effectively combat this form of cancer.

Table 1. Top differentially expressed genes in the samples that are highly enriched and significant.

Gene	**Function**	**Connection**
LTB4R/LTB4R2	Promotes cell survival and proliferation (icosanoid activity)	Major role in tumor development
PARP12	Tumor suppressor gene to fight infections	Protein-ADP riboslyase activity correlated directly with mitosis in cancer
PLA2G12A	Phospholipase A2 Activity	Prognostic Biomarker seen in Colorectal Cancer

A lot of the information gathered from this study supports existing knowledge found in primary research articles. In a study of PLAG12A in colorectal cancer, lower expression of the gene led to greater tumor growth of cancer cells (Parisi, Eva et al., 2023). In the GEO Dataset data analyzed in this study, Figure 10 shows the low expression of PLAG12A in the control group. This corroborates the existing study as the control includes uncured TNBC cells, meaning greater tumor growth. This correlation between the two studies indicates the relevance of PLAG12A in not only colorectal cancer, but triple-negative breast cancer as well.

The study found that LTB4R, LTB4R2, PARP12, and PLA2G12A are top differentially expressed genes that play a significant role in TNBC. Evaluating expressions of these genes in cancer cells can provide researchers with accurate biomarkers to assess the stage and intensity of the cancer. Inhibitors of these genes, like oligopeptide binding related to GSTM4, can be explored as potential novel therapeutics to combat the growth of tumors in the breast.

A potential limitation to the information presented in this study is attributed to the strict use of bioinformatic datasets from other researchers through RNA-sequencing methods. The top identified genes will need to be further analyzed for understanding their intricate roles in cancer. Another key limitation was the amount of samples in the groups of the GEO Dataset. There were only two control samples given as opposed to 8 samples of the DMSO treatment and 10 of the LLY283 treatment. The imbalance of samples could potentially play a role in skewing the results.

In the future, the four main genes, LTB4R, LTB4R2, PARP12, and PLAG12A, can be tested in the lab on sample cancer cells with different controls to further identify the roles they play. In addition, more research needs to be done on HTT and its unexpected relationship to breast cancer. The HTT gene found in the samples needs to be tested to see whether it is referring to the mutant HTT gene or the wild type one in order to understand its relevance to the negative effects of cancer.

Conclusion

Over 26,000 genes from the GEO Dataset samples were narrowed down to the top 40 differentially expressed genes. These top 40 were further refined through gene ontology and KEGG pathway to four main genes: LTB4R, LTB4R2, PARP12, and PLAG12A. These significant genes play important roles in cell differentiation, mitosis, and cell proliferation of tumor cells.

The research conducted to identify the difference in expression of genes between the cancer cells and the treated cancer cells was depicted through GEO expression results of the aforementioned four genes. These results showed the differences in amounts of each gene in the samples per group, indicating the effectiveness of each treatment plan. The study's hypothesis was that there will be a difference between gene expression in the control group, DMSO group, and the LLY283 group, which will allow us to compare the effects of different treatments on triple-negative breast cancer cells. As hypothesized, the results showed clear distinction between gene expression levels based on cancer cell growth in comparison to cancer cells receiving treatment. Further research studies should focus on testing these top four genes in the lab and investigating the exact relationship between the HTT gene, a gene primarily involved in Hunington's disease and breast cancer.

References

1. Advanced Journal of Chemistry, Section B: Natural Products and Medical Chemistry – Editorial Board, www.ajchem-b.com/journal/editorial.board. Accessed 23 July 2024.
2. "Carbon-Sulfur Lyases – Mesh – NCBI." National Center for Biotechnology Information, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/mesh?Db=mesh&Cmd=DetailsSearch&Term=%22Carbon-Sulfur%2BLyases%22%5BMeSH%2BTerms%5D#:~:text=Enzymes%20that%20catalyze%20the%20cleavage,other%20than%20hydrolysis%20or%20oxidation. Accessed 5 Aug. 2024.
3. Chapdelaine, Abygail G., and Gongqin Sun. "Challenges and opportunities in developing targeted therapies for triple negative breast cancer." Biomolecules, vol. 13, no. 8, 1 Aug. 2023, p. 1207, https://doi.org/10.3390/biom13081207.
4. Clough, Emily, and Tanya Barrett. "The Gene Expression Omnibus Database." SpringerLink, Springer New York, 2016, link.springer.com/protocol/10.1007/978-1-4939-3578-9_5.
5. Dameron, O., Bettembourg, C., & Meur, N. L. (n.d.). Measuring the evolution of ontology complexity: The gene ontology case study. PLOS ONE. https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0075993
6. Dorsam, Robert T., and J. Silvio Gutkind. "G-protein-coupled receptors and cancer." Nature Reviews Cancer, vol. 7, no. 2, Feb. 2007, pp. 79–94, https://doi.org/10.1038/nrc2069.
7. "Eicosanoid." Eicosanoid – an Overview | ScienceDirect Topics, www.sciencedirect.com/topics/neuroscience/eicosanoid#:~:text=Eicosanoids%20have%20important%20significant%20role,et%20al.%2C%202021). Accessed 5 Aug. 2024.
8. Feustel, Kavanya, and Gerald S. Falchook. "Protein arginine methyltransferase 5 (PRMT5) inhibitors in oncology clinical trials: A Review." Journal of Immunotherapy and Precision Oncology, vol. 5, no. 3, 22 June 2022, pp. 58–67, https://doi.org/10.36401/jipo-22-1.
9. "Gene Ontology Overview." Gene Ontology Resource, 11 July 2024, geneontology.org/docs/ontology-documentation/#:~:text=The%20Gene%20Ontology%20(GO)%20describes,catalysis%E2%80%9D%20or%20%E2%80%9Ctransport%E2%80%9D.

10. Hoang, Ba X et al. “The Rationality of Implementation of Dimethyl Sulfoxide as Differentiation-inducing Agent in Cancer Therapy.” Cancer diagnosis & prognosis vol. 3,1 1-8. 3 Jan. 2023, doi:10.21873/cdp.10172
11. “HTT Huntingtin [Homo Sapiens (Human)] – Gene – NCBI.” National Center for Biotechnology Information, U.S. National Library of Medicine, www.ncbi.nlm.nih.gov/gene/3064#:~:text=Studies%20indicate%20that%20mutant%20Huntingtin,poor%20prognosis%20and%20metastasis%20development. Accessed 5 Aug. 2024.
12. Izdebska, Magdalena et al. “Involvement of Actin and Actin-Binding Proteins in Carcinogenesis.” Cells vol. 9,10 2245. 6 Oct. 2020, doi:10.3390/cells9102245
13. Johnson, Amber M., et al. “Eicosanoids in cancer: New roles in immunoregulation.” Frontiers in Pharmacology, vol. 11, 18 Nov. 2020, https://doi.org/10.3389/fphar.2020.595498.
14. Kalinkin, A.I., et al. “Leukotriene B4 receptors abnormal gene expression is associated with either shorter or longer survival in breast cancer patients depending on the intrinsic tumour molecular subtype.” Annals of Oncology, vol. 30, Nov. 2019, p. vii17, https://doi.org/10.1093/annonc/mdz413.061.
15. Kegg Pathway Map (HELP), www.genome.jp/kegg/document/help_pathway.html#:~:text=The%20KEGG%20PATHWAY%20database%20is,pathways%20associated%20with%20human%20diseases. Accessed 5 Aug. 2024.
16. Kennedy, Luke et al. “Role of Glutathione in Cancer: From Mechanisms to Therapies.” Biomolecules vol. 10,10 1429. 9 Oct. 2020, doi:10.3390/biom10101429
17. Key, Timothy J, et al. “Epidemiology of Breast Cancer.” The Lancet Oncology, vol. 2, no. 3, Mar. 2001, pp. 133–140, https://doi.org/10.1016/s1470-2045(00)00254-0.
18. Leon-Ferre, Roberto A, and Matthew P Goetz. “Advances in systemic therapies for triple negative breast cancer.” BMJ, 30 May 2023, https://doi.org/10.1136/bmj-2022-071674.
19. Luo, W et al. “GSTM4 is a microsatellite-containing EWS/FLI target involved in Ewing’s sarcoma oncogenesis and therapeutic resistance.” Oncogenevol. 28,46 (2009): 4126-32. doi:10.1038/onc.2009.262

20. Obidiro, Onyinyechi, et al. “Triple negative breast cancer treatment options and limitations: Future outlook.” Pharmaceutics, vol. 15, no. 7, 23 June 2023, p. 1796, https://doi.org/10.3390/pharmaceutics15071796.
21. “Oligopeptide.” Oligopeptide – an Overview | ScienceDirect Topics, www.sciencedirect.com/topics/chemistry/oligopeptide#:~:text=The%20RGD%20oligopeptide%20(Arg%2DGly,and%20metastasis%20of%20solid%20tumors%2C. Accessed 5 Aug. 2024.
22. Parisi, Eva et al. “PLA2G12A as a Novel Biomarker for Colorectal Cancer with Prognostic Relevance.” International journal of molecular sciences vol. 24,13 10889. 30 Jun. 2023, doi:10.3390/ijms241310889
23. Polyak, Kornelia. “Breast cancer: Origins and evolution.” Journal of Clinical Investigation, vol. 117, no. 11, 1 Nov. 2007, pp. 3155–3163, https://doi.org/10.1172/jci33295.
24. Pruitt, Kim D, et al.. “Introducing RefSeq and LocusLink: Curated Human Genome Resources at the NCBI.” Redirecting, doi.org/10.1016/S0168-9525(99)01882-X. Accessed 11 July 2024.
25. Scarpa, Emanuele S., et al. “A role of intracellular mono-adp-ribosylation in cancer biology.” The FEBS Journal, vol. 280, no. 15, 10 May 2013, pp. 3551–3562, https://doi.org/10.1111/febs.12290.
26. Shailesh, Harshita et al. “Protein arginine methyltransferase 5 (PRMT5) dysregulation in cancer.” Oncotarget vol. 9,94 36705-36718. 30 Nov. 2018, doi:10.18632/oncotarget.26404
27. Shanley, Mathew. “Why Is Cancer Uncommon in Huntington’s Disease Patients?” HCP Live, HCP Live, 4 Feb. 2021, www.hcplive.com/view/cancer-uncommon-in-hd.
28. Shao, Changjuan, et al. “Parp12 (ARTD12) suppresses hepatocellular carcinoma metastasis through interacting with FHL2 and regulating its stability.” Cell Death & Disease, vol. 9, no. 9, 28 Aug. 2018, https://doi.org/10.1038/s41419-018-0906-1.
29. Smolarz, Beata, et al. “Breast cancer—epidemiology, classification, pathogenesis and treatment (review of literature).” Cancers, vol. 14, no. 10, 23 May 2022, p. 2569, https://doi.org/10.3390/cancers14102569.
30. Tang, Doudou, et al. “SRplot: A Free Online Platform for Data Visualization and Graphing.” PloS One, U.S. National Library of Medicine, 9 Nov. 2023,

www.ncbi.nlm.nih.gov/pmc/articles/PMC10635526/#:~:text=SRplot%3A%20A%20free%20online%20platform,The%20site%20is%20secure.

31. Thion, Morgane Sonia, and Sandrine Humbert. “Cancer: From Wild-Type to Mutant Huntingtin.” Journal of Huntington’s disease vol. 7,3 (2018): 201-208. doi:10.3233/JHD-180290
32. “Triple-Negative Breast Cancer: Facts and Stats.” Medical News Today, MediLexicon International, www.medicalnewstoday.com/articles/facts-stats-triple-negative-breast-cancer#:~:text=TNBC%20accounts%20for%20about%2010,females%20in%20the%20United%20States. Accessed 23 July 2024.
33. Villarroel, Ariana, et al. “DMSO inhibits human cancer cells and downregulates the expression of cdk2 and cyclin A.” The FASEB Journal, vol. 34, no. S1, Apr. 2020, pp. 1–1, https://doi.org/10.1096/fasebj.2020.34.s1.03158.
34. Waks, Adrienne G., and Eric P. Winer. “Breast cancer treatment.” JAMA, vol. 321, no. 3, 22 Jan. 2019, p. 288, https://doi.org/10.1001/jama.2018.19323.
35. Wang, Shaowei et al. “Calcium-dependent cytosolic phospholipase A2 activation is implicated in neuroinflammation and oxidative stress associated with ApoE4.” Molecular neurodegeneration vol. 17,1 42. 15 Jun. 2022, doi:10.1186/s13024-022-00549-5
36. Wang, Ya;Wang. “Profilin 1 Protein and Its Implications for Cancers.” Cancer Network, MJH Life Sciences, www.cancernetwork.com/view/profilin-1-protein-and-its-implications-for-cancers. Accessed 5 Aug. 2024.
37. Welsby, Iain, et al. “Parp12, an interferon-stimulated gene involved in the control of protein translation and inflammation.” Journal of Biological Chemistry, vol. 289, no. 38, Sept. 2014, pp. 26642–26657, https://doi.org/10.1074/jbc.m114.589515.
38. Wu, Min, et al. “pla2g12a protects against diet-induced obesity and insulin resistance by enhancing energy expenditure and clearance of circulating triglycerides.” The FASEB Journal, vol. 38, no. 9, 4 May 2024, https://doi.org/10.1096/fj.202302075r.
39. Xu, Xiangwei, et al. “Synthesis of paclitaxel derivatives for remote loading into liposomes and improved therapeutic effect.” Molecules, vol. 27, no. 22, 17 Nov. 2022, p. 7967, https://doi.org/10.3390/molecules27227967.

40. Yang, Kaidi, et al. “Identification of key pathways and hub genes in basal-like breast cancer using bioinformatics analysis.” OncoTargets and Therapy, Volume 12, Feb. 2019, pp. 1319–1331, https://doi.org/10.2147/ott.s158619.

Gowtham Kantamneni

Determining the best methods of Health Education for Cardiovascular Health Disease

ABSTRACT

Background: Coronary heart disease (CHD) has increased over the last twenty years. Previous studies show that health education may have a role in exacerbating this epidemic. This study aims to find the best ways to help alleviate the inequality regarding CHD between educated and less educated individuals.

Methods: Google Scholar was searched using the key search terms coronary disease and education to identify three articles to be reviewed. They focused on education and CHD, studies written in English, and published within the last 15 years and excluded studies with a population of children under the age of 18 or study design of a literature review.

Results: Nearly 42% of patients with CHD reported gaps in their cardiology care, primarily at the transition age of 19.9 years, while only 8% of patients with severe CHD experienced such gaps. Individuals with less exposure to health education had higher total cholesterol, systolic blood pressure, and smoking intensity, leading to a 6.8% increased cumulative risk of CHD in men and 6.2% in women by age 75, with 9334 CHD or stroke events occurring during an 11.5-year follow-up.

Discussion: Health education decisively plays a key role in the onset of CHD. Teaching self-care to patients, particularly in young adults is a potential method to incorporate to help lessen the inequalities elicited by education.

INTRODUCTION

Coronary heart disease (CHD) has become the single most common cause of death globally [1]. The average lifespan of an individual with CHD is 62 years of age [2]. Although this study confirmed that leading a healthy life is mostly achievable until the late stages of life due to advances in healthcare, we cannot overlook the overall impact of health education on lifespan which is still a primary concern [2]. In just the ten years from 2000 to 2010, there has been a 55% increase in adults with CHD, further emphasizing the urgency of the issue [3].

A lifestyle consisting of frequent exercise, a healthy diet, and free of smoking is pivotal for a heart that is devoid of CHD [4]. Unfortunately, adults who have less health education are less likely to follow the habits listed above, and as such, are more likely to expose themselves to a higher risk of developing CHD [5]. Studying the causes of the differences between uneducated individuals can help clarify what makes them more exposed to this condition and help decrease the prevalence of CHD as a whole by lessening its impact on one of its largest targets. The purpose of this paper is to determine methods for alleviating the inequality elicited by differences in health education between individuals by measuring their quantitative effectiveness on patients.

METHODS

I used Google Scholar to identify research papers using the key terms coronary disease and education. I chose an inclusion criteria of studies written in English and published in the past 15 years. Studies with participants under the age of 18 were excluded as were other literature reviews.

RESULTS

Importance of Self-Care

Patients with heart failure have shown a great increase in their self-care behavior from being educated and supported by a nurse in the hospital and home setting [8]. Self-care behavior was increased in patients from both the intervention and control groups within 1 month of discharge, but the increase in the intervention group was significantly more after 1 month [8]. Self-care behavior in both groups decreased over the next 8 months, but the increase from baseline remained statistically significant in the intervention group only, not in the control group [8]. The intervention had no significant effects on resource use [8].

At Risk Uneducated Age Groups

Of patients with CHD, 42% reported gaps in their cardiology care [9]. This number was also much lower at just 8% for patients with severe CHD [9]. The most telling result was that the typical age for gaps in care occurred during the transition from adolescence to adulthood with a mean age of 19.9 [9].

Most pivotal risk factors by education level

At baseline, less educated individuals had higher mean levels of total cholesterol and systolic blood pressure, were more likely to be current smokers, under antihypertensive treatment, and obese (p-values < 0.05) [10]. Additionally, less educated women had lower HDL cholesterol and a higher prevalence of diabetes, with smoking intensity increasing as education decreased, particularly among women [10]. During a median follow-up of 11.5 years, 9334 incident coronary heart disease (CHD) or stroke events and 7020 non-CVD deaths occurred [10]. Low education was significantly associated with an increased cumulative risk of CHD (p-value < 0.0001 for both genders) and stroke (p-values 0.003 in men, 0.001 in women), with an absolute difference in CVD risk by age 75 of 6.8% in men and 6.2% in women [10]. Additive interaction analysis revealed that smoking increased the cumulative risk of CVD by 3.1% in less-educated men and 1.5% in less-educated women, while obesity was associated with a lower cumulative risk in less-educated men [10].

The cumulative risk of CHD or stroke, above all else, is greater the lower your education level, holding all extraneous variables equivalent [10].

DISCUSSION

Relation between CHD and Education

A study examining health education's relationship with CHD confirmed that education impacts CHD levels even amongst individuals with similar habits and lifestyles [10]. As such, there is a need to quell disparities brought about by education.

Methods to consider to aid in the combat of CHD

With regards to specific methods of helping uneducated individuals with CHD, focusing on teaching self care to patients has been shown to decrease hospital visits as well as prolonged increase in self-care habits [8]. In essence, taking the time to teach patients how to properly care for themselves increases their likelihood of following these habits and leading a healthier life after discharge. The teaching resonated with patients, improving their self-care habits for many months more than the control groups. As teaching self-care led to a smaller readmission rate to

hospitals, teaching self-care can save time and money for hospitals, while allowing patients to maintain a healthier life.

In a study which looked at the prevalence in "gaps" in cardiology care, which are essentially points at which patients are not administering proper care for their respective conditions, gaps were substantially more prevalent in young adults with a mean age of 20 [9]. This means that young adults with CHD are the least likely to properly care for themselves and should be targeted by educational interventions. Although these ages may have a smaller prevalence of CHD than older individuals, lessons learned at this age have been shown to hold true throughout much of adulthood, meaning addressing these ages could not only benefit the most disadvantaged age group, but also improve their CHD care in the future. Based on the findings, we can conclude that the best way of alleviating health education shortcomings is to teach self-care to patients who are transitioning from adolescence to adulthood, around the age of 20.

References

1. Anderson, L., Brown, J. P., Clark, A. M., Dalal, H., Rossau, H. K. K., Bridges, C., &
2. Taylor, R. S. (2017). Patient education in the management of coronary heart disease.
3. Cochrane Database of Systematic Reviews, (6).
4. Castelli, W. P. (1984). Epidemiology of coronary heart disease: the Framingham study.
5. The American journal of medicine, 76(2), 4-12.
6. Marelli, A. J., Ionescu-Ittu, R., Mackie, A. S., Guo, L., Dendukuri, N., & Kaouache, M.
7. (2014). Lifetime prevalence of congenital heart disease in the general population from
8. 2000 to 2010. Circulation, 130(9), 749-756.
9. Puddu, P. E., & Menotti, A. (2015). The impact of basic lifestyle behaviour on health:
10. How to lower the risk of coronary heart disease, other cardiovascular diseases, cancer and
11. all-cause mortality. E-Journal of Cardiology Practice, 13(32).
12. Jacobsen, B. K., & Thelle, D. S. (1988). Risk factors for coronary heart disease and level

13. of education: The Tromsø Heart Study. American Journal of Epidemiology, 127(5), 923- 932.
14. Bambra, C. (2016). Health divides: where you live can kill you. Policy Press.
15. Nietzel, M. T. (2024b, February 20). Percentage of U.S. adults with college degree or postsecondary credential reaches new high, according to Lumina report. Forbes.
16. Jaarsma, T., Halfens, R., Huijer Abu-Saad, H., Dracup, K., Gorgels, T., Van Ree, J., &Stappers, J. (1999). Effects of education and support on self-care and resource utilizationin patients with heart failure. European heart journal, 20(9), 673-682.
17. Gurvitz, M., Valente, A. M., Broberg, C., Cook, S., Stout, K., Kay, J., … & Alliance for Adult Research in Congenital Cardiology (AARCC) and Adult Congenital Heart Association. (2013). Prevalence and predictors of gaps in care among adult congenital heart disease patients: HEART-ACHD (The Health, Education, and Access Research Trial). Journal of the American College of Cardiology, 61(21), 2180-2184.
18. Veronesi, G., Tunstall-Pedoe, H., Ferrario, M. M., Kee, F., Kuulasmaa, K., Chambless, L. E., … & MORGAM Project. (2017). Combined effect of educational status and
19. cardiovascular risk factors on the incidence of coronary heart disease and stroke in
20. European cohorts: implications for prevention. European journal of preventive
21. cardiology, 24(4), 437-445.
22. Silventoinen, K., Pankow, J., Jousilahti, P., Hu, G., & Tuomilehto, J. (2005). Educational
23. inequalities in the metabolic syndrome and coronary heart disease among middle-aged
24. men and women. International Journal of Epidemiology, 34(2), 327-334.
25. Healthy habits begin in childhood – Nebraska Department of Education. (n.d.).
26. https://www.education.ne.gov/ns/cacfp/foodnutrition/healthy-habits-begin-in-childhood/

Suhruth Koneru

Using Bioinformatics to Study the Effects of COVID-19 on the Heart through Myocarditis

ABSTRACT

This research is addressing the problem with most COVID-19 vaccines. The problem with most COVID-19 vaccines is that after the patient gets the vaccine, their chances of getting a heart condition called myocarditis (the inflammation of the heart) is higher. The goal of this research is to understand the effects of COVID-19 on the heart through myocarditis as well as creating a vaccine which will not only be effective like the COVID-19 vaccine right now but also a much lower risk of myocarditis. There are many different ways to create vaccines, however, one of the most common ways is when the pathogen is isolated and grown until it loses its effectiveness in causing diseases. The weakest pathogens are then selected to put into the vaccine. Different bioinformatics tools and databases were used to collect, perform bioinformatics research experiments, analyze and interpret results. Using NCBI and GEO2R bioinformatics tool and database, we found 30 differentially expressed genes (DEGs) in the dataset GSE235433 with 15 genes being upregulated and 15 genes downregulated. The GO and KEGG pathway analysis showed that the upregulated DEGs were mainly involved in the signal transductor. From the DEGs we identified key genes with DAVID. The key DEGs were mainly related to 3 gene ontology terms.

INTRODUCTION

SARS-CoV-2 is a virus which causes the disease COVID-19. COVID-19 is very dangerous to the heart as it increases the chances of myocarditis substantially. Myocarditis is the inflammation of the heart and it can cause complications such as a heart attack, heart failure, and stroke (Myocarditis: Symptoms and Causes; Lasica et al, 2023). Myocarditis is a serious disease with high death and illness rates, especially in young people. A big challenge is that it is difficult to diagnose myocarditis accurately and quickly because there are many different causes and related diseases that can lead to it (Lasica et al, 2023).

Myocarditis is usually caused by an infection, but it can also result from drug hypersensitivity, radiation, metabolic disorders, collagen diseases, sarcoidosis, Kawasaki disease, and exposure to excessive heat or chemicals (Lasica et al, 2023).

It is difficult to know exactly how common myocarditis is because many people with the condition aren't diagnosed. However, it's estimated that between 10.2 and 105.6 out of every 100,000 people worldwide have it, with about 1.8 million new cases each year (Lasica et al, 2023).

During the COVID-19 pandemic, up to 28% of patients experienced heart damage, shown by increased levels of a heart-specific protein called troponin (Guo et al, 2020). A study found that 54% of patients who had COVID-19 showed signs of myocarditis (heart inflammation) on a cardiac MRI (Altay, 2022).

The scientific question being investigated in this research study is how SARS-CoV-2 is causing myocarditis in terms of gene expression and what can we do to prevent this from causing life threatening complications?

Almost everyone knows about COVID-19, however, most people only know about its effects on the lungs and the respiratory system (Galiatsatos et al., 2022), the effects COVID-19 has on the heart can be even more life threatening (Liu et al., 2020). The goal is to find how COVID-19 affects the heart and the effects of SARS-CoV-2 on the chances of myocarditis.

It is hypothesized that genes related to the heart will be differentially expressed in samples with the COVID-19 infection compared to samples without the infection. Therefore, the goals of this research are to first identify genes that are differentially expressed between samples that are COVID-19 infected and healthy samples. Then to identify genes related to the heart that resulting from SARS-CoV-2 that can be used to potentially create vaccines which do not cause an increase in the chances of myocarditis.

Throughout this study, many tools were used such as GEO (https://www.ncbi.nlm.nih.gov/gds/?term=), GEO2R, NCBI, and DAVID (https://david.ncifcrf.gov/tools.jsp). This research is important because

using these DEGs and the analysis of them through Gene Ontology, we can create a vaccine to help protect people from COVID-19 as well as reduce the chances of myocarditis

METHODS

Using NCBI to attain biological datasets:

The National Center for Biotechnology Information (NCBI) is the home to many biological datasets (https://www.ncbi.nlm.nih.gov/). NCBI allows us to analyze these datasets that we wouldn't have been able to analyze before using its different features and tools (https://www.ncbi.nlm.nih.gov/geo/geo2r/).

In this study, NCBI was used and filtered out only GEO datasets so that when the dataset was found, the dataset would be able to be analyzed with GEO2R. Within GEO is the GEO2R that is a no-coding bioinformatics platform with different datasets of experiments based on RNA and genes (http://www.ncbi.nlm.nih.gov/geo/geo2r (Davis and Meltzer, 2007).

The GEO dataset we used was GSE235433. This dataset can be found on the NCBI website by searching up the exact dataset ID in the search bar. (GSE235433)

Analyzing the Dataset with GEO2R and DAVID:

Once I found the dataset that pertained to my research topic, I then split the dataset into 9 different groups after which I clicked analyze with GEO2R. After analyzing the dataset with GEO2R, I was able to obtain differentially expressed genes (DEGs) and many graphs.

To select the top differentially expressed genes (DEGs),statistics were used via p value and log2 | fold change (FC) |. The P-value < 0.05 was considered to have statistical significance and to achieve significant differentially expressed genes. Log FC was used by selecting genes with a positive log FC and negative FC values (Supplementary Data: Gene IDs Differentially Expressed Genes).

Identifying functions and biological pathways where genes are enriched using DAVID:

In order to predict and determine functions of genes and to identify where in the biological pathway certain genes are involved in, Gene Ontology (GO) functional terms (Ashburner et al, 2000) and KEGG databases are used (Kanehisa et al., 2012). Specifically, the GO terms for gene functions include biological process (BP), cellular component (CC), and molecular function (MF) ((Ashburner et al., 2000).

In this research study, I selected the top DEGs and analyzed them with the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics database (https://david.ncifcrf.gov/) (Huang da et al., 2009) to perform the gene ontology and KEGG Pathway enrichment analysis (Kanehisa et al., 2012).

The main flow of steps in this research work is shown in the chart below. First, the research topic was found. Next, the geodataset related to the topic of research was found and then the data was analyzed with GEO2R. Then, based on the p-values specific DEGs were worked with and then DAVID was used to find the highly enriched terms through GO enrichment analysis.

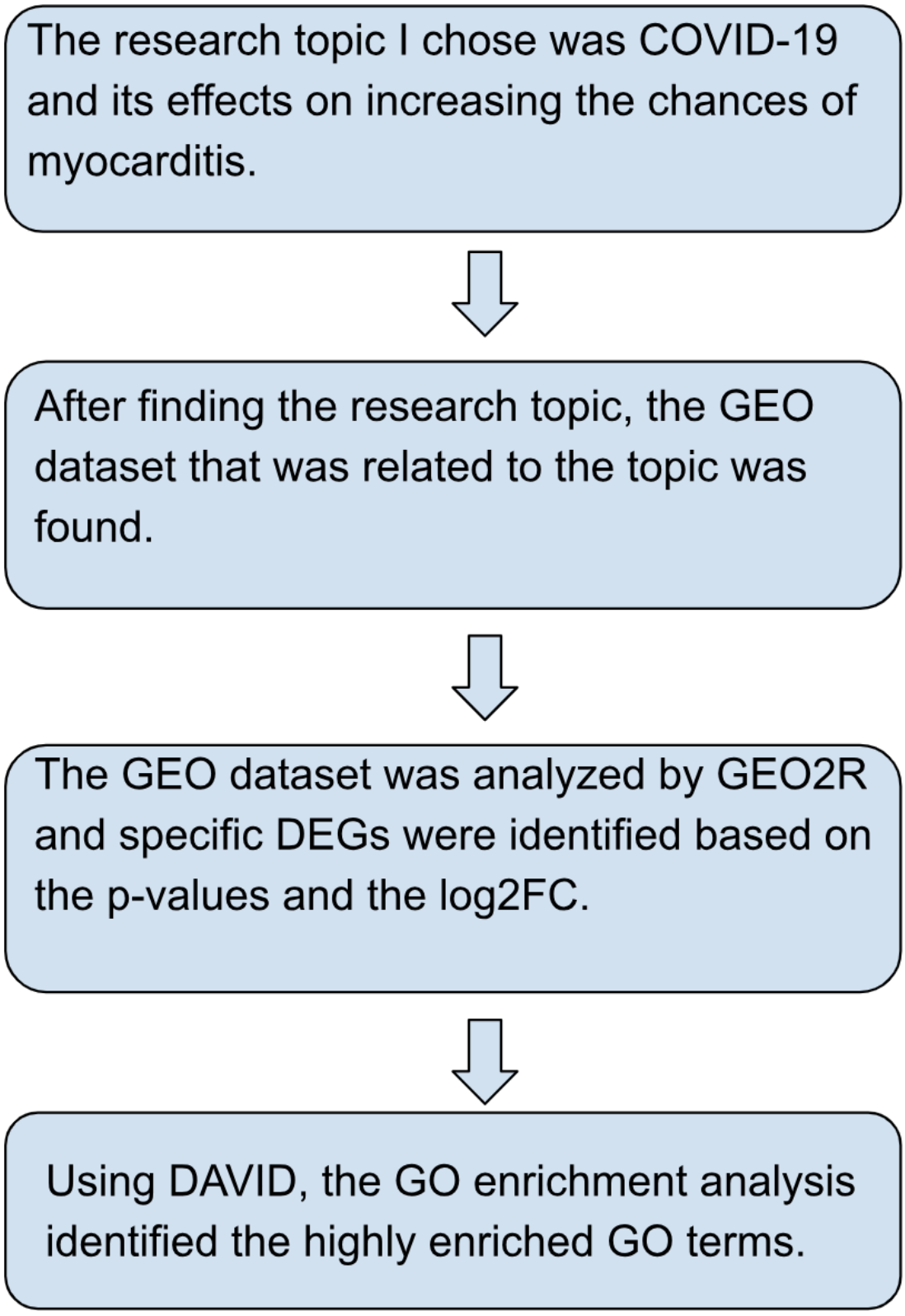

The research topic I chose was COVID-19 and its effects on increasing the chances of myocarditis.
After finding the research topic, the GEO dataset that was related to the topic was found.
The GEO dataset was analyzed by GEO2R and specific DEGs were identified based on the p-values and the log2FC.
Using DAVID, the GO enrichment analysis identified the highly enriched GO terms.

Figure 1: Summary of steps used in this research work. Overview of the steps and methods used in this bioinformatics research work.

RESULTS

GEO2R Analysis:

When finding the DEGs that would like to be worked with, only the top DEGs that are statistically significant are selected. To do this, the top 15 upregulated (high enrichment) and top 15 down regulated (low enrichment) DEGs must be determined based on their p-values and log2 Fold change (Supplementary results: Gene IDs). These gene IDs and p-values are obtained from the GEO2R analysis of the GEO dataset that is being worked with (GSE235433). After the GEO2R analysis is performed, graphs are created based on how the data is grouped.

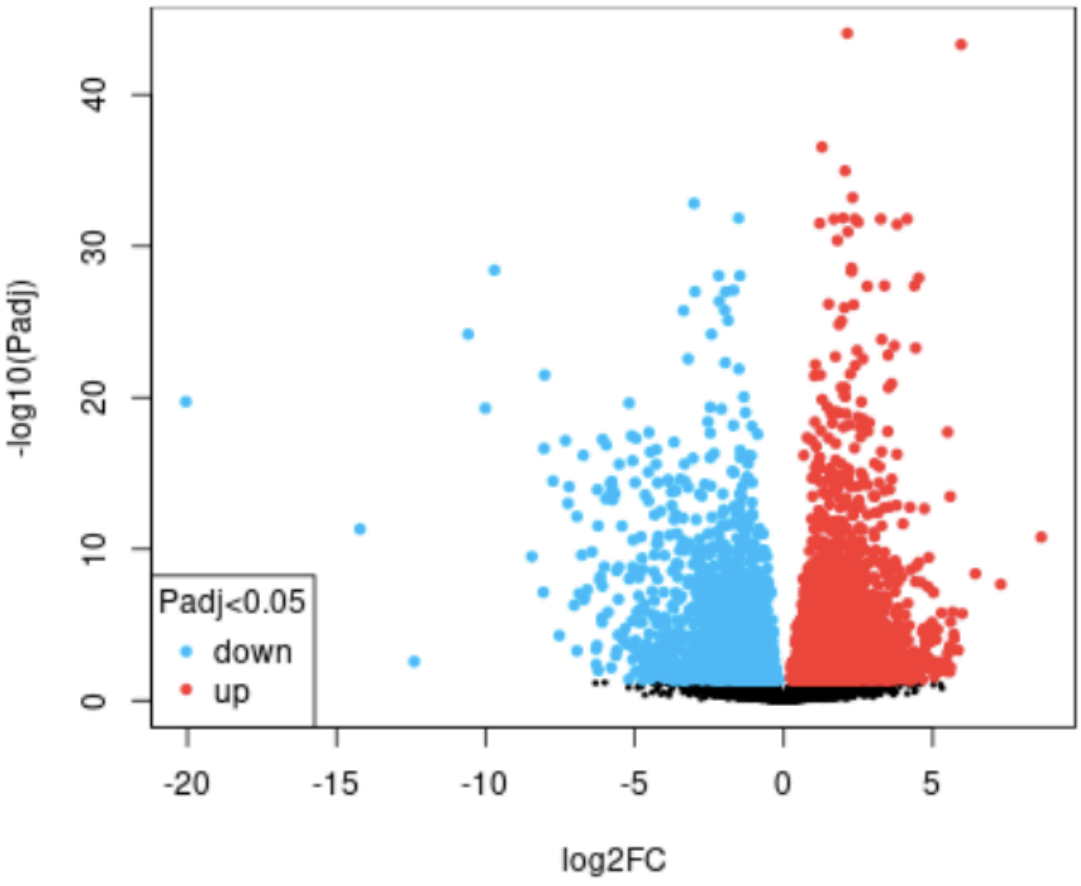


Figure 2a): The upregulated and downregulated genes in the infected timepoint 2 vs the infected timepoint 8 for the virus SARS-CoV-2. The blue dots represent the down regulated genes, while the red dots represent the up regulated genes. The black dots, however, represent genes that do not show any differences in expression between the control and infected samples.

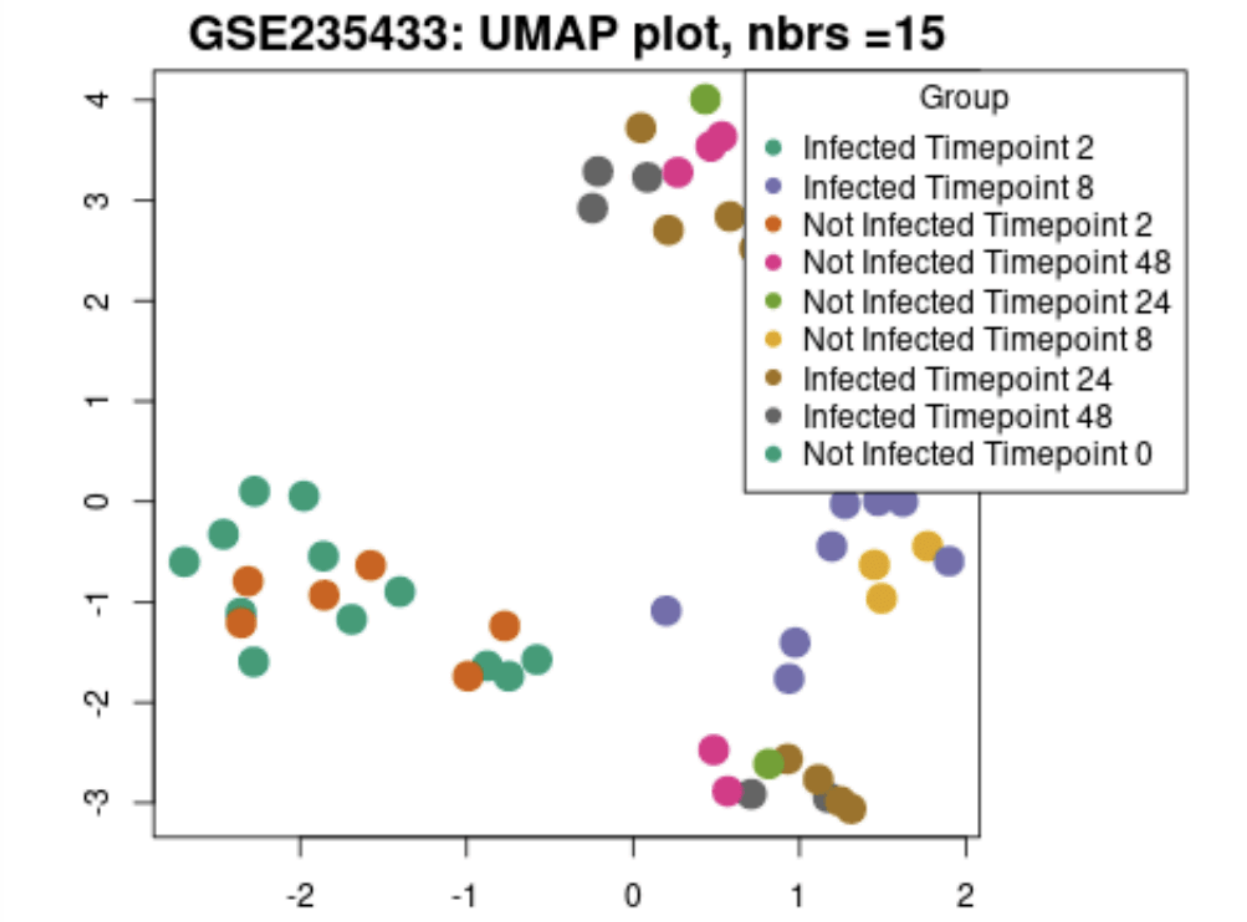


Figure 2b: UMAP Scatter plot containing genes of all 9 different groups. The closer each gene is to another, the more they are associated with each other.

Gene Ontology Enrichment Analysis:

The top 30 DEGs were found by selecting DEGs with less than 0.05 as the p-value and the top 15 DEGs with a positive LogFC value and top 15 DEGs with a negative LogFC value (Supplementary Results: Gene IDs). Using

these 30 top DEGs, the gene IDs of these DEGs can be pasted into DAVID (https://david.ncifcrf.gov/tools.jsp) and then the gene ontology can be observed and it can be seen which terms are more enriched and which aren't. Gene Ontology helps describe the biological domain as well as combine the gene and gene product attributes.

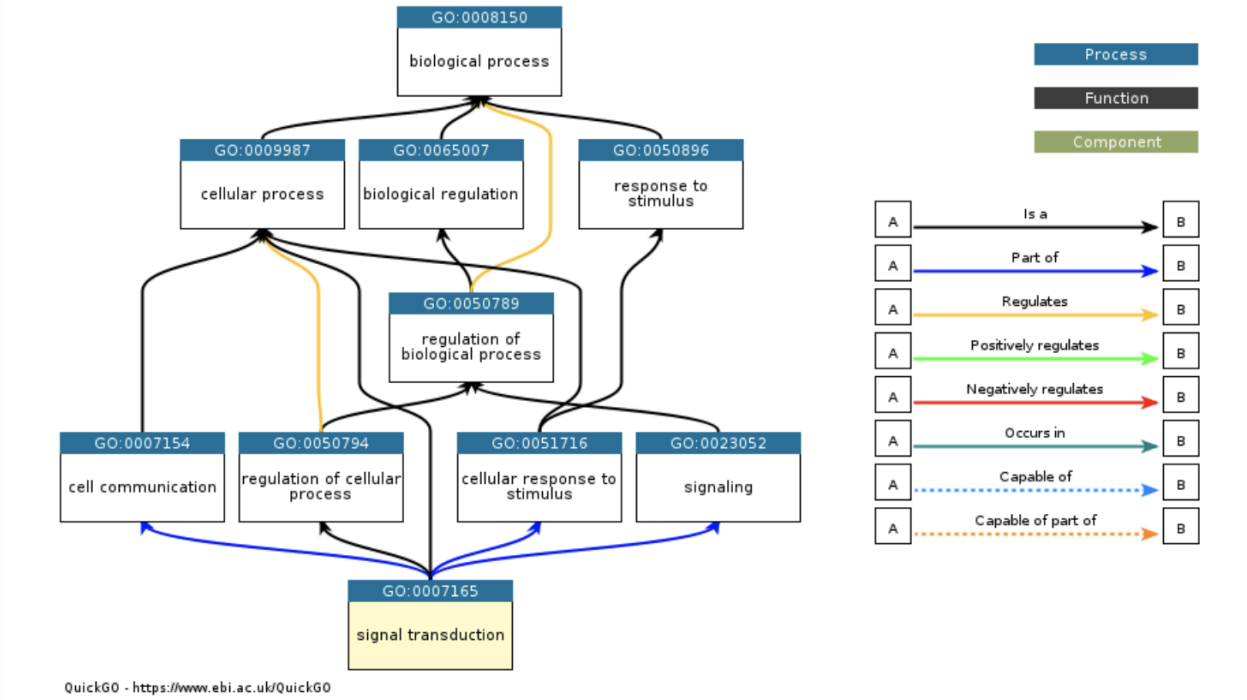


Figure 3a: The ancestry of the signal transduction GO term. The legend on the right of the chart shows what each arrow means and how each arrow relates to the flow of the ancestry chart. This chart shows the enrichment of the signal transduction term. The signal transduction is the pathway in which the COVID-19 virus travels through the cell (Rex et al., 2021).

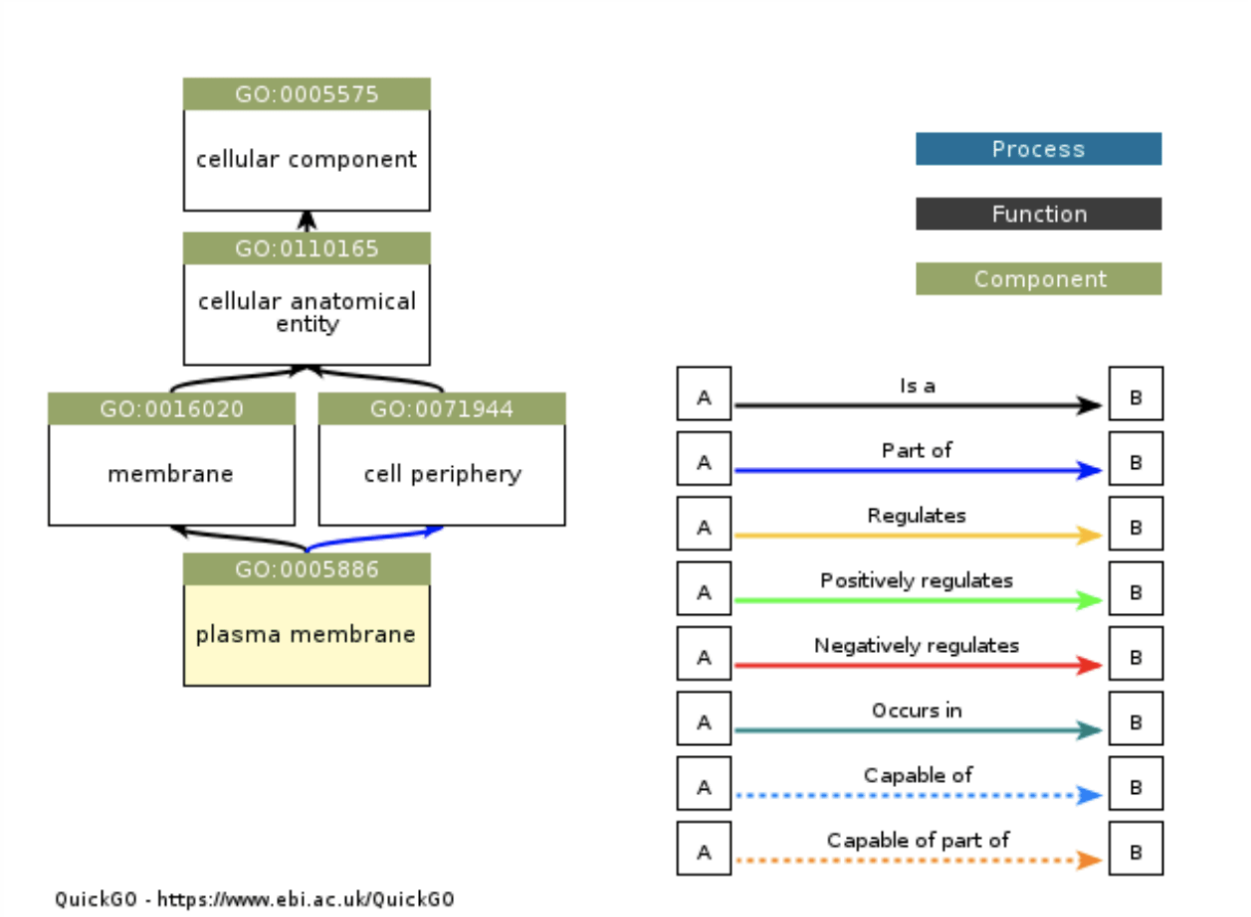


Figure 3b: The ancestry of the plasma membrane GO term. The legend on the right of the chart shows what each arrow represents and how each arrow relates to the origin of each part of the ancestry chart. This chart shows the enrichment of the plasma membrane term. The plasma membrane is involved with COVID-19 when COVID-19 stays along the plasma membrane the receptor molecules and fusion proteins bind together on the plasma membrane (Cohen et al., 2015).

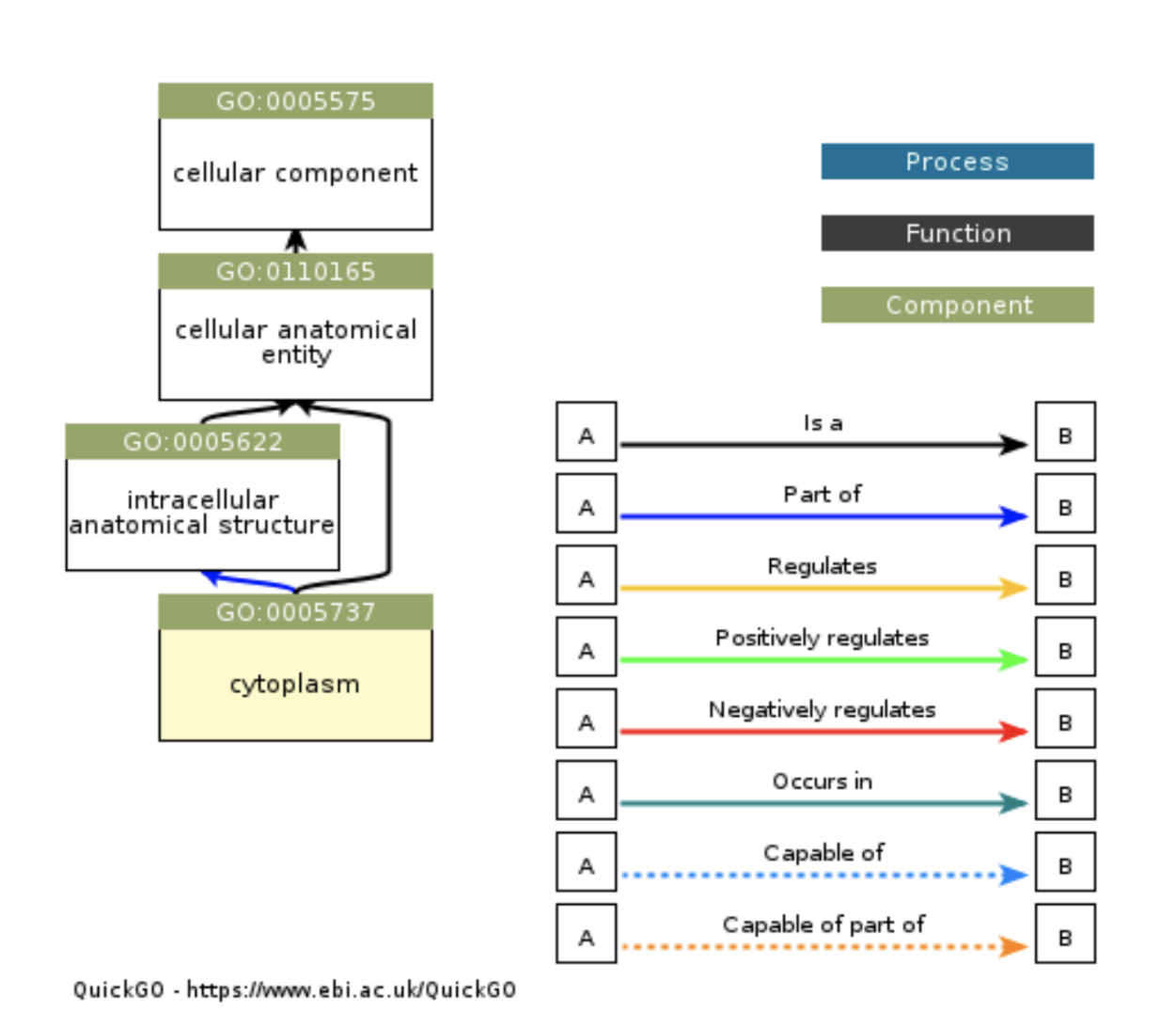


Figure 3c: The ancestry of the cytoplasm GO term. The legend on the right of the chart shows what each arrow represents and how each arrow relates to the flow of the ancestry chart. This chart shows the enrichment of the cytoplasm term. The cytoplasm is involved with COVID-19 when the COVID-19 cell replicates in the cytoplasm (Chen et al., 2022).

KEGG Pathway Enrichment Analysis:

The process for KEGG Pathway enrichment analysis is similar to the process for Gene Ontology enrichment analysis. The top DEGs are pasted into DAVID, and the KEGG Pathway can be observed and it can be seen which terms relate to the terms that are highly enriched in gene ontology and which terms don't relate to them. The results showed that when analyzing any biological pathways with KEGG none of these DEGs showed significance

involving the KEGG Pathway. There were no chart records of the KEGG Pathway involved with these DEGs.

Table 1) Summary of gene ontology terms that were highly enriched in the DEGs

Gene functions which were highly enriched	Specific gene ontology term
Signal Transduction	Biological Process
Plasma Membrane	Cellular Component
Cytoplasm	Cellular Component

The chart shown above shows the gene ontology terms that were highly enriched by performing the gene ontology enrichment analysis. Gene Ontology helps describe the biological domain as well as combine the gene and gene product attributes.

DISCUSSION

The main goal of this study was to identify and determine which differentially expressed genes in COVID-19 may be involved in the increased chances of myocarditis that the virus causes. To identify genes that are expressed, we used the Gene Expression Omnibus Bioinformatics database that is based on the R Programming language (https://www.ncbi.nlm.nih.gov/geo/). Then to identify the GO terms biological functions of the differentially expressed genes and to put these genes into corresponding biological pathways from the KEGG bioinformatics database, the Database for Annotation, Visualization and Integrated Discovery (DAVID) bioinformatics database (https://david.ncifcrf.gov/) was used (Huang da et al., 2009).

The results from performing the GEO2R analysis and selecting specific DEGs based on their p-values shows that these DEGs are highly associated with COVID-19 and its effects of causing an increase in chances of myocarditis. Using these DEGs, when we perform the gene ontology this process shows us that the terms signal transductor, plasma membrane, and cytoplasm are highly enriched which leads us to understand that COVID-19 associates with

these terms the most (Figure 3). The gene ontology process also shows that of all the GO (gene ontology) terms, the cellular component is associated the most as it has more enriched terms and higher enriched terms (Table 1). This means that in terms of how these results relate to our research question, COVID-19's association with the cellular component is one of the main causes of the increased chances of myocarditis.

This study's results are similar to others such as this article from NCBI (Heidecker et al., 2022). As said in the article most patients have an increased chance of myocarditis after getting the vaccine, which is what we are trying to overcome as we are performing this research in an attempt to decrease those chances of myocarditis with a similarly effective vaccine.

The results produced by this research have great significance in developing a better vaccine for COVID-19. As of now, the COVID-19 virus has multiple vaccines, however, with most of these vaccines comes a higher chance of getting myocarditis (Heidecker et al., 2022). Using these results, however, the current COVID-19 vaccines have a possibility of working effectively along with the advantage of having a lower chance of getting myocarditis.

One limitation this research faced, however, was that all the genes that were used as a part of this research were part of bioinformatics datasets from other research, therefore, before creating a vaccine that reduces the chances of myocarditis, all the gene ontology terms identified must be studied in a laboratory or clinical setting.

The identified gene ontology terms can be tested in the laboratory by scientists in the laboratory or clinical trials to determine if vaccines can be made for COVID-19 without the chances of myocarditis being increased for future research.

Conclusion

In conclusion, throughout this study tools like NCBI (https://www.ncbi.nlm.nih.gov/) and DAVID (https://david.ncifcrf.gov/tools.jsp) were used to study gene expression in samples that are both infected and not infected by this virus. Through analysis with the GO bioinformatics database on the 30 DEGs, the highly enriched or associated gene ontology (GO) terms of the study were able to

be understood. Results indicate that the signal transduction gene ontology term is the pathway in which the COVID-19 virus travels through the cell. Further, the results indicate that the plasma membrane is involved with COVID-19 when COVID-19 stays along the plasma membrane the receptor molecules and fusion proteins bind together on the plasma membrane. The cytoplasm is involved with COVID-19 when the COVID-19 cell replicates in the cytoplasm. The goal of this research was to understand the effects of this virus (COVID-19) on the heart through myocarditis and this study summarizes how COVID-19 potentially affects the heart through myocarditis. With these results from Table 1, researchers can potentially develop a vaccine that does not have the effects of an increased chance of myocarditis.

SUPPLEMENTARY DATA
GEO2R Analysis R-script
Identified DEGs from GEO2R

References

1. Altay S. COVID-19 myocarditis cardiac magnetic resonance findings in symptomatic patients. Acta Radiol. 2022;63:1475–1480. doi: 10.1177/02841851211046502.
2. Chen, M., Ma, Y., & Chang, W. (2022). Sars-cov-2 and the nucleus. International Journal of Biological Sciences, 18(12), 4731–4743. https://doi.org/10.7150/ijbs.72482
3. Cohen, F. S. (2016). How viruses invade cells. Biophysical Journal, 110(5), 1028–1032. https://doi.org/10.1016/j.bpj.2016.02.006
4. Covid's damage lingers in the heart | harvard medicine magazine. (n.d.). Retrieved June 20, 2024, from https://magazine.hms.harvard.edu/articles/covids-damage-lingers-heart
5. David functional annotation bioinformatics microarray analysis. (n.d.). Retrieved June 20, 2024, from https://david.ncifcrf.gov/
6. Guo T., Fan Y., Chen M., Wu X., Zhang L., He T., Wang H., Wan J., Wang X., Lu Z. Cardiovascular Implications of Fatal Outcomes of Patients with Coronavirus Disease 2019 (COVID-19) JAMA Cardiol. 2020;5:811–818. doi: 10.1001/jamacardio.2020.1017
7. Geo2r—Geo—Ncbi. (n.d.). Retrieved June 20, 2024, from https://www.ncbi.nlm.nih.gov/geo/geo2r/?acc=GSE235433

8. Heart problems after covid? Yes, it’s possible. (n.d.). Cleveland Clinic. Retrieved June 20, 2024, from https://my.clevelandclinic.org/health/articles/heart-problems-after-covid
9. Heidecker, B., Dagan, N., Balicer, R., Eriksson, U., Rosano, G., Coats, A., Tschöpe, C., Kelle, S., Poland, G. A., Frustaci, A., Klingel, K., Martin, P., Hare, J. M., Cooper, L. T., Pantazis, A., Imazio, M., Prasad, S., & Lüscher, T. F. (2022). Myocarditis following COVID-19 vaccine: Incidence, presentation, diagnosis, pathophysiology, therapy, and outcomes put into perspective. A clinical consensus document supported by the Heart Failure Association of the European Society of Cardiology (Esc) and the ESC Working Group on Myocardial and Pericardial Diseases. European Journal of Heart Failure, 10.1002/ejhf.2669. https://doi.org/10.1002/ejhf.2669
10. Lasica R, Djukanovic L, Savic L, Krljanac G, Zdravkovic M, Ristic M, Lasica A, Asanin M, Ristic A. Update on Myocarditis: From Etiology and Clinical Picture to Modern Diagnostics and Methods of Treatment. Diagnostics (Basel). 2023 Sep 28;13(19):3073. doi: 10.3390/diagnostics13193073. PMID: 37835816; PMCID: PMC10572782.
11. Liu, P. P., Blet, A., Smyth, D., & Li, H. (2020). The science underlying covid-19: Implications for the cardiovascular system. Circulation, 142(1), 68–78. https://doi.org/10.1161/CIRCULATIONAHA.120.047549
12. Redirect notice. (n.d.-a). Retrieved June 20, 2024, from https://www.google.com/url?q=https://www.frontiersin.org/journals/microbiology/articles/10.3389/fmicb.2018.01093/full%23B45&sa=D&source=docs&ust=1718745276853694&usg=AOvVaw02bNVIGEQuaTYwpspkTaoW
13. Redirect notice. (n.d.-b). Retrieved June 20, 2024, from https://www.google.com/url?q=https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8523273/&sa=D&source=docs&ust=1718745276853892&usg=AOvVaw3gvqQLyp3eL571TlAPO-DB
14. Rex, D. A. B., Dagamajalu, S., Kandasamy, R. K., Raju, R., & Prasad, T. S. K. (2021). SARS-CoV-2 signaling pathway map: A functional landscape of molecular mechanisms in COVID-19. Journal of Cell Communication and Signaling, 15(4), 601–608. https://doi.org/10.1007/s12079-021-00632-4

15. Sanbomics. (2022, January 8). Simple gene ontology and pathway enrichment from a gene list. https://www.youtube.com/watch?v=XLRA0A5qsoE
16. Severe lung infection during COVID-19 can cause damage to the heart. (2024, March 20). National Institutes of Health (NIH). https://www.nih.gov/news-events/news-releases/severe-lung-infection-during-covid-19-can-cause-damage-heart
17. Wang, S.-J., Brodie, K. C., De Pons, J. L., Demos, W. M., Gibson, A. C., Hayman, G. T., Hill, M. L., Kaldunski, M. L., Lamers, L., Laulederkind, S. J. F., Nalabolu, H. S., Thota, J., Thorat, K., Tutaj, M. A., Tutaj, M., Vedi, M., Zacher, S., Smith, J. R., Dwinell, M. R., & Kwitek, A. E. (2022). Ontological analysis of coronavirus associated human genes at the covid-19 disease portal. Genes, 13(12), 2304. https://doi.org/10.3390/genes13122304
18. Zhuang, Z., Zhong, X., Chen, Q., Chen, H., & Liu, Z. (2022). Bioinformatics and system biology approach to reveal the interaction network and the therapeutic implications for non-small cell lung cancer patients with covid-19. Frontiers in Pharmacology, 13, 857730. https://doi.org/10.3389/fphar.2022.857730
19. Vaccines. (2024, February 14). https://www.hopkinsmedicine.org/health/treatment-tests-and-therapies/vaccines

Kaavya Kotamraju

Pharmacological Enhancement of Neuroplasticity: Exploring the Effects of Psilocybin, Monoaminergic, and Dopaminergic Drugs

ABSTRACT

Neuroplasticity, defined as the ability of the brain to reconfigure and adapt, is a significant factor in learning, memory and emotion regulation. This review discusses the effects of psilocybin, monoamine and dopaminergic drugs on neuroplasticity in different brain compartments. Evidence from research shows that the capacity of psilocybin to increase plasticity in the prefrontal cortex may be useful therapeutic for mood disorders. Furthermore, monoaminergic drugs may counteract the inhibitory effect of plasticity in the motor cortex, especially in rehabilitation of older people. Dopaminergic drugs deal with reward pathways and therefore are valuable clues for recovery processes from addiction and decision-making. By synthesizing research across these domains, this review ties these findings into what is currently known regarding the therapeutic potential for drug-induced neuroplasticity and urges further work. Understanding how the drugs facilitate this adaptability could unveil opportunities for future innovative treatments within mental health, addiction, and neurorehabilitation.

INTRODUCTION

Neuroplasticity, i.e., brain ability to change and reconfigure neural networks, creates the basis for cognition, emotional cognition, and motor function [1-5]. Studies of the modulation of this adaptability by pharmacological substances, psilocybin, monoaminergic/dopaminergic substances, etc., are also emphasized recently, for the role that may be played by these substances via different neural pathways. Longitudinal studies of psilocybin have indicated that its consumption induces emotional control and cognitive flexibility by acting on variability in brain function in some regions of prefrontal cortex (PFC) and limbic structures. [1] Monoamine drugs promote plasticity in the motor cortex and there is an invaluable application of this in motor rehabilitation in the elderly. [2] Dopamine antagonists exert effects on reward channels and decision-making mechanisms and have an important role in addiction treatment. [6]

This article explores the mechanisms related to neuroplasticity that are influenced by certain drugs, focusing on their effects in the prefrontal cortex, limbic system, motor cortex, and reward pathways. The aim of this review article is to present a detailed view of how these pharmacologic agents can modulate neural networks for the treatment of a diverse array of neurological and psychiatric disorders.

METHODS

To explore how different drugs affect neuroplasticity, original research studies were reviewed from databases like PubMed and Google Scholar. I focused on studies investigating the effects of psilocybin, monoaminergic drugs, and dopaminergic agents on brain function. The search terms used were "neuroplasticity," "drug effects on the brain," and "cognitive flexibility." Primary research with a clear experimental design and measurable outcomes was prioritized, while review articles or theoretical papers were excluded. These studies utilized fMRI and EEG imaging, behavioral testing, and neurochemical analysis to assess changes in specific brain regions. By bringing these findings together, this review outlines patterns and gaps in the current knowledge on drug-induced neuroplasticity.

RESULTS

The studies consistently demonstrate that various drugs, including psilocybin, monoaminergic drugs, and dopaminergic agents, enhance neuroplasticity in distinct, overlapping brain regions [1-5]. Psychedelics such as psilocybin prominently influence the prefrontal cortex and limbic system, enhancing gamma oscillations tied to cognitive flexibility and emotional regulation [1,2]. Monoaminergic drugs show a marked effect on motor cortex plasticity, particularly in older adults [4], while dopaminergic drugs modulate reward anticipation and decision-making processes in the orbitofrontal cortex [3,5,6].

Psilocybin and monoaminergic drugs both promote neural adaptability, yet their effects are localized to different systems: psilocybin influences emotional and cognitive pathways, while monoaminergic drugs enhance motor function [1,2,4]. Cocaine withdrawal and subliminal drug cues specifically activate limbic and dopaminergic pathways, highlighting how addiction-related stimuli reshape neural circuits [3,5,6].

Author, Year	Type of Drug	Part of Brain affected	Effect on Neuroplasticity
Kalivas, et al., 2008	Dopaminergic drugs	Orbitofrontal cortex	Modulates reward anticipation and decision-making by altering reward-related neuronal activity.
Skosnik et al., 2023	Psilocybin	Prefrontal cortex	Enhances gamma oscillations, linked to neuroplasticity, and reduces depressive symptoms.
Kesar et al., 2017	Monoaminergic drugs	Motor cortex	Enhances motor cortex plasticity and motor performance, particularly in older adults during task training.
Nestler, 2013	Cocaine	Dopaminergic pathways	Induces time-dependent increase in craving, highlighting changes in neuroplasticity during withdrawal.
Regier, Paul et al., 2021	Subliminal drug cues	Amygdala	Activates limbic regions, even without

			conscious awareness, influencing neural pathways tied to craving and emotional response.

DISCUSSION

The studies in this review article relied on small sample sizes or specific populations, such as older adults or individuals with a history of addiction, limiting generalizability [2,4,5]. Additionally, certain drugs, like psilocybin, have limited longitudinal data to confirm the durability of their effects on neuroplasticity [1-2]. Studies using animal models may not fully replicate human neurophysiological responses [3,5].

This review looks into the structures of the brain on cognitive processes, motor activities, and emotional control with the aim of understanding more about the action of psychopharmaceutical drugs on neuroplasticity. It addresses the issue of how psilocybin, monoaminergic agents and dopaminergic compounds may enhance the brain plasticity and its possible uses in treating drug addiction, mental disorders, and neurorehabilitation. Based on a range of studies about these substances, the review shows the unique effects drugs can have on controlled neuroplasticity. These different approaches suggest the possibility for developing more personalized therapies in the future. In addition, it is important to understand the mechanisms of action of drugs that cause the neuroplastic changes in order to fully realize the therapeutic effects of the drugs. This understanding will facilitate the conduct of further studies that can result in better care and treatment of patients in several disciplines.

For example, psilocybin has been shown to be able to interact with serotonin receptors which leads to increased levels of neuroplasticity in the prefrontal cortex and limbic areas – regions of the brain that are important in mood and cognition [1]. This makes it plausible as a drug for treating depression and other mood disorders. On the other hand, the monoaminergic agents are noted to have a good effect on motor cortex plasticity, suggesting applications for motor rehabilitation and age-related impairments [5].

Research into dopaminergic agents has elucidated their role in modulating addiction-related neuroplasticity by influencing reward pathways, providing valuable insights into the dynamic processes underlying addiction [4,6]. Additionally, the interplay between subliminal drug cues and limbic system responses offers a novel perspective on the relationship between emotional processing and addiction [3].

In a world where the need for effective pharmacological interventions in the realm of mental health, addiction, and aging is increasingly urgent, understanding the capacity of the brain to undergo neuroplasticity holds great promise. This review synthesizes findings from diverse studies, bridging critical gaps in the literature and offering a unified framework for leveraging drug-induced neuroplasticity to support brain health.

This paper is an important contribution at the intersection of pharmacology and neuroplasticity. While previous research has focused largely on the effects of individual drugs, this review brings together evidence from a range of drug classes in an effort to provide the most comprehensive analysis yet of their influence on neuroplasticity [1–6]. Thus, the review carries valuable theoretical insights and practical implications for the fields of clinical sciences—most especially psychiatry, addiction medicine, and neurorehabilitation—in which the promise of drug-induced neuroplasticity holds the greatest hope for the enhancement of outcomes. Still, several limitations should be taken into account with the current review.

Many included studies suffer from a small sample size and an exploration of only the acute or very short-term effects; it thereby restricts the insight into the longer term. Furthermore, individual variability in response—determined by factors such as age, mental health status, and genetic predispositions—underscores the need for more tailored therapeutic strategies [2,5]. A number of important areas need to be addressed by future research in order to advance knowledge of drug-induced neuroplasticity. First, longitudinal studies are necessary to determine the long-term durability of neuroplastic changes, particularly in the therapeutic setting [1,5]. Furthermore, understanding the molecular mechanisms underlying neuroplasticity in the various brain regions is important for appreciating its role in addiction and mental health disorders [3,4]. Personalized treatment strategies will be worth investigating to understand

how individual differences, such as genetic factors, age, and pre-existing conditions, might affect the efficacies of pharmacological interventions [5]. Finally, the synergistic effects of combined interventions—combining pharmacological treatments with cognitive therapies or physical exercise—may lead to a holistic approach for enhancing neuroplasticity and improving clinical outcomes [2].

Conclusion

Research into how drugs influence neuroplasticity offers important insights into the adaptability of the brain and its possible therapeutic uses. Psilocybin, monoaminergic drugs, and dopaminergic agents have different effects on specific brain regions and hold promise for the treatment of mood disorders, enhancement of motor recovery, and improvement of addiction outcomes. Psilocybin helps in emotional processing, monoaminergic drugs support motor cortex plasticity, and dopaminergic substances help reshape reward pathways, pointing to their therapeutic potential in various conditions.

However, some points remain unclear regarding the generalization of results and long-term effects, as many studies include small samples or animal models. Further research is needed to get over such limitations and study the molecular mechanisms underlying those, focusing on personalized treatment approaches that can maximize therapeutic benefits from drug-induced neuroplasticity. By doing so, we will be in a position to go further in such aspects and develop therapies that are more potent and targeted against neurological and psychological disorders.

References

1. Skosnik, P. D., Sloshower, J., Safi-Aghdam, H., et al. (2023). Sub-acute effects of psilocybin on EEG correlates of neural plasticity in major depression: Relationship to symptoms. Journal of Psychopharmacology, 37(7), 687-697. https://doi.org/10.1177/02698811231179800
2. Kesar, T. M., Belagaje, S. R., Pergami, P., Haut, M. W., Hobbs, G., & Buetefisch, C. M. (2017). Effects of monoaminergic drugs on training-induced motor cortex plasticity in older adults. Brain Research, 1670, 106-117. https://doi.org/10.1016/j.brainres.2017.06.015

3. Tremblay, L., & Schultz, W. (2000). Reward-related neuronal activity during go-nogo task performance in primate orbitofrontal cortex. Journal of Neurophysiology, 83(4), 1864-1876. https://doi.org/10.1152/jn.2000.83.4.1864
4. Childress, A. R., Ehrman, R. N., Wang, Z., et al. (2008). Prelude to passion: Limbic activation by "unseen" drug and sexual cues. PLoS ONE, 3(1), e1506. https://doi.org/10.1371/journal.pone.0001506
5. Volkow, N. D., Chang, L., Wang, G.-J., Fowler, J. S., Ding, Y.-S., Sedler, M., et al. (2001). Cocaine decreases dopamine D2 receptors in the ventral striatum of patients with cocaine dependence. Journal of Neuroscience, 21(10), 9414–9418. https://doi.org/10.1523/JNEUROSCI.21-10-09414.2001
6. Kalivas, P. W., Volkow, N. D., & Seamans, J. K. (2008). Unmanageable motivation in addiction: A pathology in prefrontal-accumbens glutamate transmission. Neuron, 59(5), 677–686. https://doi.org/10.1016/j.neuron.2008.07.00

Rohan Kuruganthy

Investigating the Replication of Sars Cov 2 Virus Using Bioinformatics

ABSTRACT

Background: SARS-CoV-2 is the virus that causes the disease, COVID-19. As we know, COVID-19 has caused many deaths and changed the world as we know it. According to worldometers.info, Sars Cov2 killed 7,010,681 people and diagnosed 704,753,890 people. How did the virus replicate so fast?

Purpose: The purpose of this study was to identify genes that are differentially expressed and potentially linked to the rapid replication of SARS-CoV-2.

Methods: In this study, we utilized a variety of bioinformatics tools and databases to identify differentially expressed genes (DEGs) and their potential functional and biological pathways. We used NCBI – GEO2R and ShinyGO bioinformatics tools. Specifically we studied the microarray GEO Dataset GSE268196 to identify and analyze genes that are expressed differently between infected and uninfected samples using NCBI's GEO2R bioinformatics tool. Then to predict the functions or associations of these genes, we used the ShinyGO bioinformatics tool to conduct Gene Ontology (GO) enrichment analysis. Furthermore, we performed KEGG enrichment analysis to determine the biological pathways in which these DEGs are involved and their potential contribution to the rapid replication of SARS-CoV-2.

Results: GEO2R bioinformatics analysis showed that various genes were differentially expressed between infected samples and uninfected samples, and the top 39 differentially expressed genes (DEGs) were identified using statistics. Further analysis of the top 39 differentially expressed genes showed that these genes are mostly enriched in several pathways related to inflammation, including mitogen-activated protein (MAP) kinase (MARK), ERK1/2 Mitogen-activated proteins and protein kinase during infection with SARS-CoV-2.

Conclusion: This study reveals the involvement of several proteins and pathways related to inflammation; mitogen-activated protein kinase (MAPK) during infection with SARS-CoV-2. Because COVID-19 often leads to

uncontrolled inflammatory responses such as MAPK pathways, the pathways and proteins found to be enriched in our study can potentially be used in future studies to provide customized anti-inflammatory treatment to COVID-19 patients alongside the regular early use of antiviral drugs.

Keywords: GEO2R, ShinyGO, NCBI, GEO, SARS-CoV-2

INTRODUCTION

The virus that causes COVID-19 called severe acute respiratory syndrome coronavirus (SARS-CoV-2) has caused many deaths and changed the world as we know it (Hu et al, 2021). SARS-CoV-2 can be caught from the air, physical contact, and through spit (Hu et al, 2021). According to worldometers.info, Sars Cov2 killed 7,010,681 people and diagnosed 704,753,890 people. How did the virus replicate so fast?

The problem this research is addressing is the question; why does SARS-CoV-2 replicate so fast and what genes are differentially expressed during infection with SARS-CoV-2 virus? In order for viruses to infect an organism and spread, they need to replicate (V'kovski, P et al, 2021). Therefore, one way to successfully find a cure or medicine to treat viral infections is by understanding why viruses replicate.

SARS-CoV-2 replicates in the lungs and results in various clinical outcomes, ranging from cases with no symptoms, to individuals nearing death (Wiersinga et al., 2020).

Specifically SARS-CoV-2 infections can develop into very serious lung injuries, inflammation and even death (V'kovski, P et al, 2021, Cusato et al, 2023). In severe cases, an excessively active immune response leads to harmful inflammation. This inflammation affects various tissues and cell types, including those not infected by the virus, and resembles the inflammation observed in certain autoimmune diseases (Zhang et al, 2024).

The tragic COVID-19 pandemic led many researchers to carry out replication studies on SARS-CoV-2. For a virus to replicate successfully and cause disease in an organism, six main steps are involved. First the virus attaches to the organism or host, then it penetrates the organisms or host. Following

penetration into the organism, it removes its outer protein structure (uncoats) and then it replicates to spread inside the organism. Following replication, the virus assembles itself and then releases its genetic material or virion inside the organism to cause disease (V'Kovski Pet al 2021). According to previous studies, during infection with SARS-CoV-2, over 300 human proteins have been found to interact with SARS-CoV-2. Researchers have hypothesized that if these interactions between human and viral proteins were blocked, the replication of the virus can be stopped and this can stop the virus from being passed on from one person to another to cause the pandemic (V'kovski, P et al, 2021).

Therefore, the goal of this study is to identify genes that are expressed differently (DEGs) during infection with SARS-CoV-2 and the potential functions and biological pathways that may explain how SARS-Cov-2 replicates so fast. We hypothesized that the SARS-CoV-2 virus replicates by taking over the host's gene expression machinery and upregulating genes related to replication.To address the hypothesis, this research used different bioinformatics tools and databases to find the genes in SARS-CoV-2 that could be targets for antiviral medicine against COVID-19. Specifically, this study used NCBI's GEO2R bioinformatics tool and database (https://www.ncbi.nlm.nih.gov/geo/geo2r/) to find a dataset that contained SARS-Cov-2 gene expression data, and gene enrichment bioinformatics tools to determine gene functions and biological pathways associated with the identified genes.

The research is important because the genes and or biological pathways identified in this research can then be used by scientists to potentially develop additional antiviral medication or vaccines that can stop the virus from replicating and thus stop it from spreading.

METHODS

We used bioinformatics tools, NCBI's GEO (https://www.ncbi.nlm.nih.gov/geo/geo2r/) and ShinyGO (http://bioinformatics.sdstate.edu/go/). GEO was used to identify genes that are expressed differently between infected samples and control samples. Then to analyze the different genes further, ShinyGo bioinformatics tool (http://bioinformatics.sdstate.edu/go/) was used. The methodology used in this study is summarized in Figure 1.

Identification of differentially expressed genes

To identify which genes are differentially expressed during SARS-CoV-2 infection, GEO2R within the NCBI Bioinformatics database was used. Specifically, the GEO Dataset: GSE268196 was used. In this dataset, the scientists took lung cell samples from monkeys infected with Sars Cov2 on days 0, 3, 7, 14, and around a month after infection. They have 71 samples in total. These samples are divided into three groups: ALFNg (22 samples), ALL10 (24 samples), and control (24 samples). From this study, we selected a top 39 DEGs from the GEO2R bioinformatics analysis (https://www.ncbi.nlm.nih.gov/geo/geo2r/) based on p value ($p < 0.05$). The pre-programmed AI and ML algorithm with R programming language was used to generate results for our study (supplementary results: R Script used to generate GEO2R Results)

Determination of gene functions and biological pathways associated with the genes

The 39 DEGs were further analyzed for their functional enrichment to determine which biological functions or pathways they were mostly associated with or enriched in. This functional enrichment analysis was conducted using the ShinyGo bioinformatics tool (http://bioinformatics.sdstate.edu/go/).

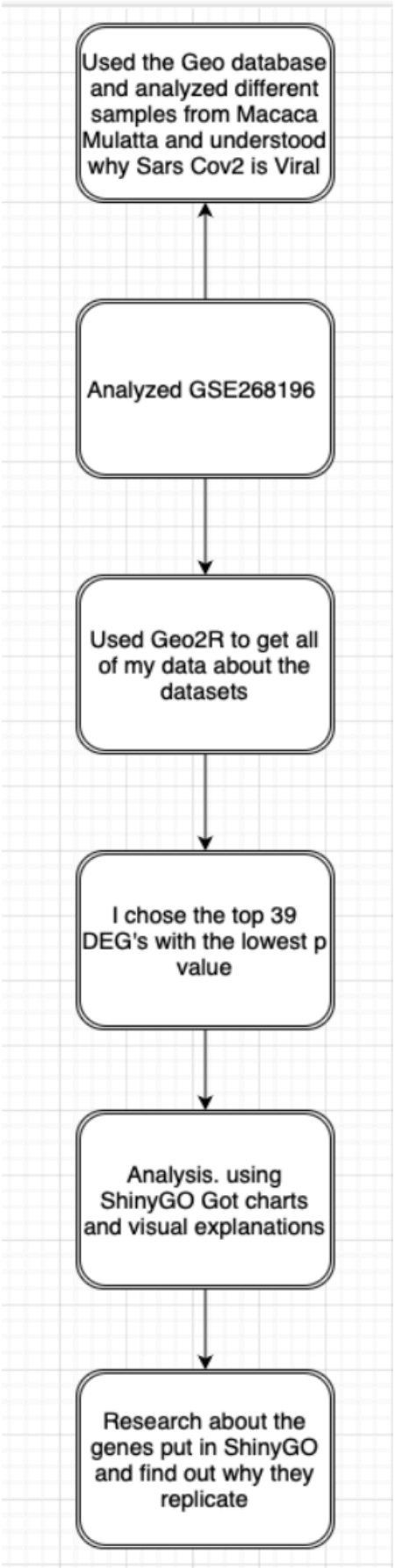

Used the Geo database and analyzed different samples from Macaca Mulatta and understood why Sars Cov2 is Viral
Analyzed GSE268196
Used Geo2R to get all of my data about the datasets
I chose the top 39 DEG's with the lowest p value
Analysis. using ShinyGO Got charts and visual explanations
Research about the genes put in ShinyGO and find out why they replicate

Figure 1: Flowchart. This flowchart summarizes the workflow of the methods used in this research study.

RESULTS

We downloaded the gene expression profile of GSE268196 from the GEO database and used the GEO2R interface connecting the R programming language with the Gene Expression Omnibus (GEO) database to generate our results (Figure 2). We got the graphs by categorizing and analyzing the gene expression profile in the previous study dataset GSE268196. The pre-programmed R Script that generated results in this study is included in the supplementary results: R Script used to generate GEO2R Results.

The GEO2R analysis of the genes expressed differently (DEGs) in the control and infected samples produced several results including a volcano plot and venn diagram as shown in Figure 2 (A, B, C and D). DEGs are genes that show big changes in their activity or gene expression levels under different conditions (Anjum et al., 2016). These conditions can be different tissues, treatments, diseases, stages of development, or environmental factors (Anjum et al., 2016). In Figure 2 the black dots represent the genes that are not expressed differently. Adding on to that, the red dots show genes that are up regulated, and blue represents genes that are down regulated.

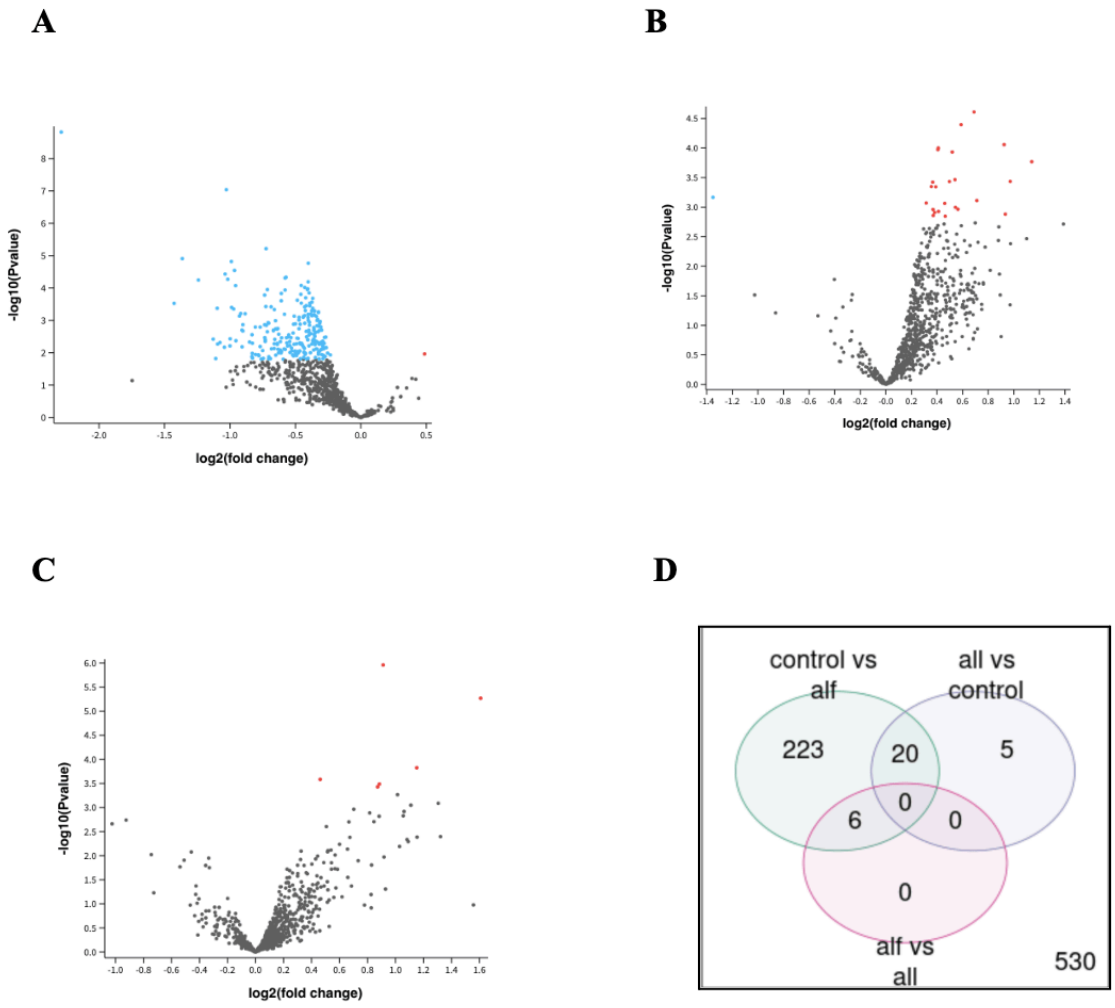


Figure 2: Genes differentially expressed between control samples and samples infected with the virus in Rhesus monkey. Monkeys infected with Sars Cov2 were divided into 3 groups: ALF, ALL, and control (no infection). **2A)** ALF vs CONTROL. **2B)** ALLvsCONTROL – **2C)** ALL vs ALF-ALL and ALF. **2D)** Venn diagram shows the total number of differentially expressed genes (DEGs) when comparing the different combinations of infected monkey samples and controls (not infected). A total of 233 DEGs were found in the ALF vs CONTROL sample groups, and 5 in the ALL vs CONTROL sample groups. While there were zero DEGs in the ALF vs ALL sample groups.

In summary, figure 2 shows that some genes were indeed expressed differently between the samples and the controls as shown by the red and blue dots (genes) and the numbers in the venn diagram.

Further Analysis of DEGs to determine Functions or Pathways the DEGs are Enriched in

The number of DEGs obtained in Figure 2 were then narrowed down by identifying the top 39 DEGs using the statistic, P-value (Top 39 DEGs). DEGs are genes that show big changes in their activity levels under different conditions (Anjum et al., 2016). These conditions can be different tissues, treatments, diseases, stages of development, or environmental factors (Anjum et al., 2016). In this study, a gene's activity was considered important if the P-value was less than 0.05. From these, the top 39 were chosen for further analysis using ShinyGo bioinformatics tool (http://bioinformatics.sdstate.edu/go/).

ShinyGo showed results of what pathways the DEGs were enriched in (Figure 3), the relationship between significantly enriched pathways (Figure 4) and correlation among significant pathways (Figure 5). Results show that the top DEG are mostly associated with or enriched in the pathways, mitogen-activated protein (MAP) kinase, ERK1/2 and Mitogen-activated protein (fold enrichment about 450).

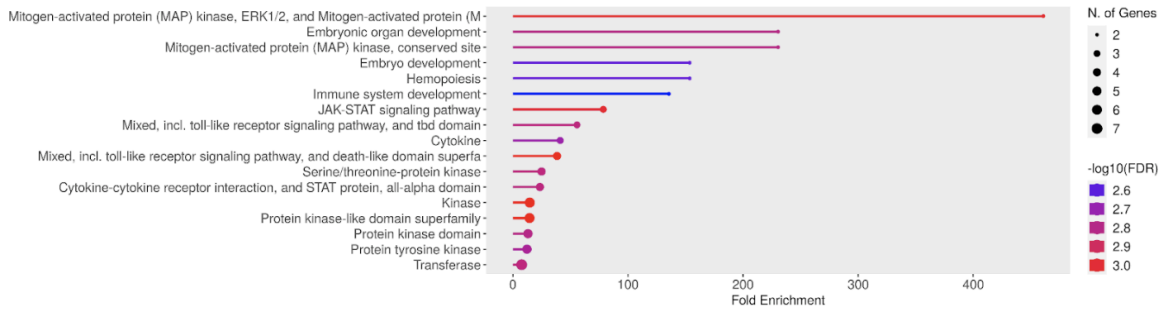


Figure 3-Chart-network enrichment of top DEGs: This fold enrichment chart shows that the top DEG are mostly associated with or enriched in the pathways, mitogen-activated protein (MAP) kinase, ERK1/2 and Mitogen-activated protein (fold enrichment about 450).

Further, these enriched pathways showed connections with each other. Specifically, the pathways mitogen-activated protein (MAP) kinase phosphatase and Mitogen-activated protein Kinase, Transferase, and Protein Kinase were significantly enriched and connected with each other by sharing 20% or more genes (Figure 4).

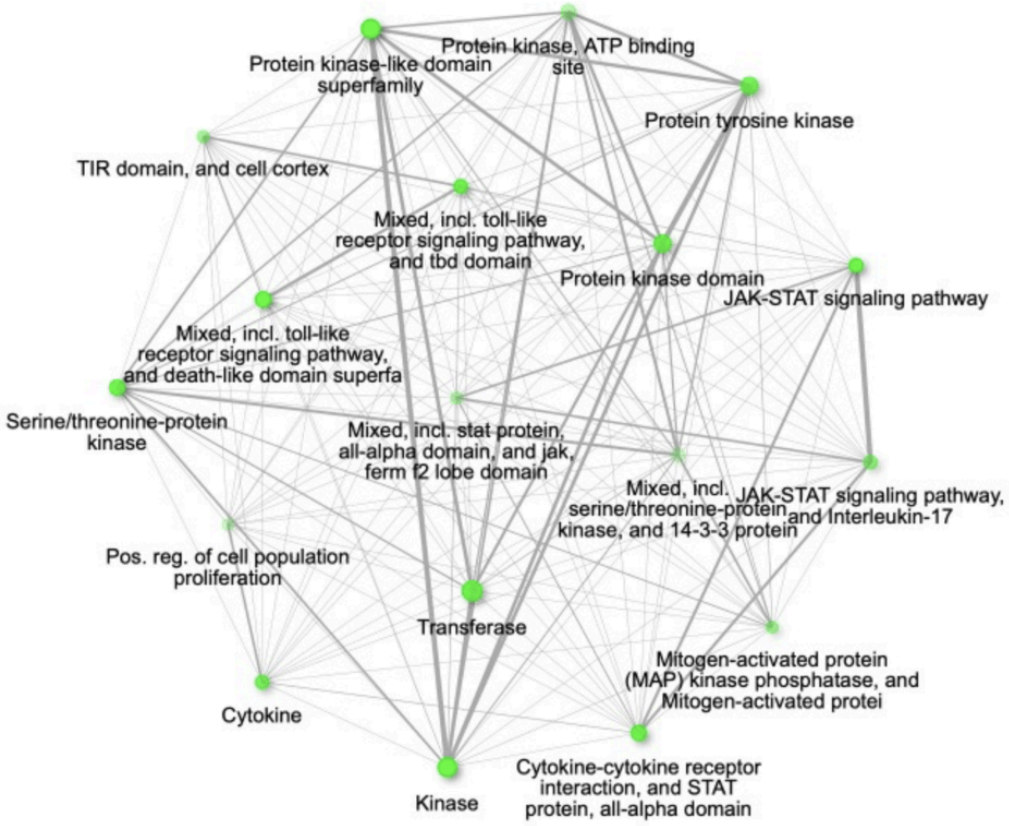


Figure 4: Relationship between enriched pathways. The significantly enriched pathways that are connected and share 20% or more genes are mitogen-activated protein (MAP) kinase phosphatase and Mitogen-activated protein, Kinase, Transferase, and Protein Kinase.Two pathways (nodes) are connected if they share 20% (default) or more genes. Darker nodes are more significantly enriched gene sets. Bigger nodes represent larger gene sets. Thicker edges represent more overlapped genes.

Although ShinyGo was not able to display or specify any genes in the results, further correlation among significant pathways (Figure 5) shown in a Hierarchical clustering tree of clustered pathways indicate that the pathways have many shared genes.

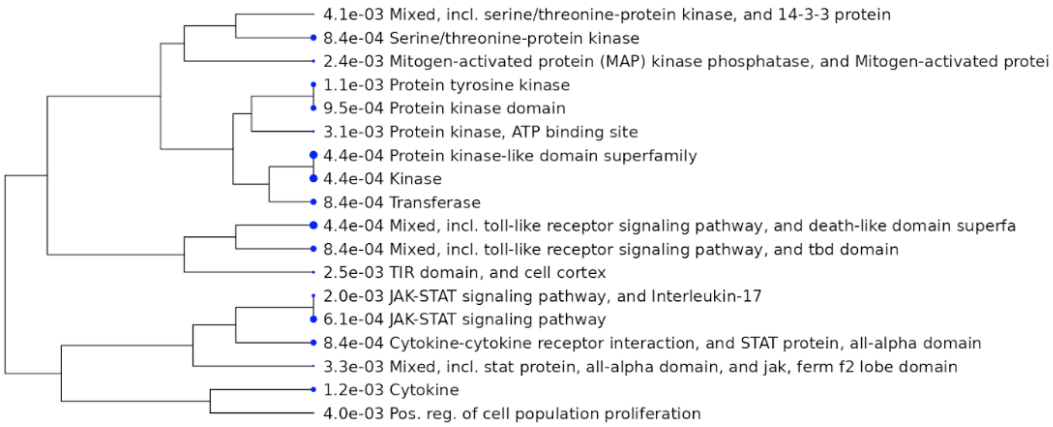


Figure 5-Hierarchical clustering tree: Summary of the correlation among significant pathways listed. Pathways with many shared genes are clustered together. Bigger dots indicate more significant P-values. Clustered together are pathways related to protein kinase, JAK-STAT signaling pathway, and interleukin, Mixed including Toll-lke receptor signaling pathway and Cytokinin-cytokinin receptor interaction, STAT protein, all alpha domain.

Overall, Table 1 summarizes the results from the enrichment pathways of the top 39 DEGs (Figure 3), the relationship between significantly enriched pathways (Figure 4) and correlation among significant pathways (Figure 5).

Table 1. Summary of significantly enriched pathways and their correlation with each other.

Pathways Significantly Enriched	**Function and Potential link to SARS-2-Cov-2**
mitogen-activated protein kinase MAPK	MAPK sends the intracellular signals in the body. Sars Cov2 or Covid19 links to MAPK because when someone is infected with Covid19 the MAPK is activated in their inflammatory system.
Mitogen-activated protein	Sends intracellular signals throughout the body. Sends the

	signals when someone has covid.
ERK1/2	Regulates stimulated cellular processes. When the MAPK is activated when someone is infected with SARS Cov2, the ERK1/2 is also triggered in the process.
Interleukin17	Interleukin17 activates and mobilizes the cell. When there is a higher Interleukin level the chance of catching Covid19 is higher.
JAK STAT signaling pathway	Binds DNA and allows the transcription of genes. This is linked to Covid 19 because it amplifies the pathological effects of Covid 19.

DISCUSSION

The purpose of the research was to find out why viruses spread or replicate so quickly. The scientific question being investigated was why the SARS-Cov-2 virus replicated so fast, and how they do so in relation to the genes that are expressed during infection. SARS-CoV-2 replicates virally through the lungs and leads to different clinical outcomes. This can include people who are asymptomatic, to people who are near death (W. Joost Wiersinga p et al 2020). Inflammation of the lungs happens during severe infection (Duan et al, 2024). And as we know, COVID-19 was one of the biggest pandemics that humans ever faced.

A previous study showed results that suggested that early-stage COVID-19 should be treated by trying to stop the virus from replicating, while later-stage therapy should focus on reducing inflammation (Duan et al, 2024). This study provided new information into why SARS-CoV-2 initially causes mild symptoms but can become potentially deadly for some patients about a week after infection. The researchers demonstrated that different stages of illness are linked to the virus behaving differently in two distinct groups of cells (Duan et al, 2024). The team discovered that when SARS-CoV-2 infects lung lining cells initially, two viral proteins operate within these cells. One protein activates the immune system, while another

paradoxically blocks this activation, leading to minimal inflammation (Duan et al, 2024). Additionally, the virus can use a different pathway to enter immune cells. This pathway limits viral replication and prevents the production of the immune signal-blocking protein. As a result, the first protein can trigger excessive inflammation associated with severe symptoms (Duan et al, 2024).

Therefore, to find out what genes are differentially expressed and linked to the rapid replication of SARS-CoV-2, we went to various bioinformatics databases and tools to collect data and conducted the bioinformatics research (Figure 1). Our goal was to go a step further and potentially identify genes that are differentially expressed and linked to the rapid replication spreading of SARS-CoV-2.

The bioinformatics database used in this study was GSE268196 where researchers originally collected preserved lung cells from rhesus macaques (monkeys) infected with SARS-CoV-2 on days 0, 3, 7, 14, and 28-35 after infection. They had 71 samples in total, divided into three groups: aIFNg (22 samples), aIL10 (24 samples), and control (24 samples) (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE268196). These scientists' results suggested that SARS-CoV-2 might spread more easily in lungs with less immune activity and that recent lung inflammation can affect the severity of COVID-19 across different people.

However in our bioinformatics study, we did not see any of our top DEGs enriched in immunity. Instead, our results showed that the top DEGs were significantly enriched in the biological pathways related to inflammation; mitogen-activated protein (MAP) kinase (MAPK) phosphatase and Mitogen-activated protein. Kinase, Transferase(no previous study that links to Sars Cov2), and Protein Kinase (Figures 3, 4 and 5). This is similar to a previous study by who found that MAPK-related biomarker levels were higher in participants who tested positive for SARS-CoV-2 (89 people) compared to those who tested negative (29 people) (Cusato et al, 2023). The researchers therefore suggested that MAPK-related biomarkers might influence COVID-19 symptoms and could help identify people with COVID-19 who are at risk for inflammation complications.

In addition, the presence of these MAPK enriched inflammation-related pathways in our results is supported by a previous study that showed that SARS-CoV-2 can infect key immune cells, triggering strong inflammatory signals (Junqueira et al, 2022). The researchers examined fresh blood samples from COVID-19 patients at Massachusetts General Hospital's emergency department. They compared these samples with those from healthy people and patients with other respiratory conditions. They also studied lung tissue from autopsies of people who died from COVID-19 (Junqueira et al, 2022). The team discovered that SARS-CoV-2 can infect two types of immune cells that act as early responders to infection: monocytes in the blood and macrophages in the lungs. When these cells become infected, they die a fiery death called pyroptosis. As they die, they release a burst of powerful inflammatory alarm signals (Junqueira et al, 2022).

Therefore, the results in our study can be used to further analyze the potential use of these inflammation MAPK-related proteins and pathways as anti-inflammatory medication for COVID-19 patients.

Limitations

Because our research used data from experiments done by other researchers, one limitation is that the enriched pathways we found in this study will need to be studied more in the laboratory or in a clinical setting before anti-inflammatory medication can be potentially developed in future studies.

Conclusion

The goal of the research was to identify genes that are expressed differently and how they can help explain why SARS-CoV-2 viruses replicate so quickly. Before the research was started, the hypothesis was that the SARS-CoV-2 virus replicates by taking over the host's gene expression machinery and upregulating genes related to replication. However, our study shows that instead, the genes that were differentially expressed in hosts infected with SARS-CoV-2, were more associated with proteins and pathways related to inflammation, including the mitogen-activated protein kinase (MAPK) pathway. When a person is infected with the SARS COV2 virus, their inflammatory system is impacted. When this happens the MAPK sends a signal to the rest of the body. Since COVID-19 often causes uncontrolled

inflammatory responses through pathways like MAPK, these proteins and pathways shown in our study could be targeted in future studies to provide customized anti-inflammatory treatments for COVID-19 patients, in addition to the regular early use of antiviral drugs.

Supplementary Results

R Script used to generate GEO2R Results

Top 39 DEGs

References

1. Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A., & Rai, A. (2016, April 1). Identification of differentially expressed genes in RNA-seq data of Arabidopsis thaliana: A compound distribution approach. Journal of computational biology : a journal of computational molecular cell biology.
2. Barrett, T., Suzek, T. O., Troup, D. B., Wilhite, S. E., Ngau, W. C., Ledoux, P., … & Edgar, R. (2005). NCBI GEO: mining millions of expression profiles—database and tools. Nucleic acids research, 33(suppl_1), D562-D566.
3. Chenchula S., Karunakaran P., Sharma S., Chavan M. Current evidence on efficacy of COVID-19 booster dose vaccination against the Omicron variant: A systematic review. J Med Virol. 2022 Jul;94(7):2969-2976.
4. Cusato J, Manca A, Palermiti A, Mula J, Costanzo M, Antonucci M, Trunfio M, Corcione S, Chiara F, De Vivo ED, Ianniello A, Ferrara M, Di Perri G, De Rosa FG, D'Avolio A, Calcagno A. COVID-19: A Possible Contribution of the MAPK Pathway. Biomedicines. 2023 May 16;11(5):1459. doi: 10.3390/biomedicines11051459. PMID: 37239131; PMCID: PMC10216575.
5. Duan, T., Xing, C., Chu, J. et al. ACE2-dependent and -independent SARS-CoV-2 entries dictate viral replication and inflammatory response during infection. Nat Cell Biol 26, 628–644 (2024). https://doi.org/10.1038/s41556-024-01388-w
6. Ge, S. X., Jung, D., & Yao, R. (2020). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics, 36(8), 2628-2629. (http://bioinformatics.sdstate.edu/go/)
7. Geall, A. J., Mandl, C. W., Ulmer, J. B. RNA: The new revolution in nucleic acid vaccines. Semin Immunol. 2013;25(2):152–159. doi: 10.1016/j.smim.2013.05.001.

8. Hu, B., Guo, H., Zhou, P. et al. Characteristics of SARS-CoV-2 and COVID-19. Nat Rev Microbiol 19, 141–154 (2021). https://doi.org/10.1038/s41579-020-00459-7
9. Junqueira, C., Crespo, Â., Ranjbar, S. et al. FcγR-mediated SARS-CoV-2 infection of monocytes activates inflammation. Nature 606, 576–584 (2022). https://doi.org/10.1038/s41586-022-04702-4
10. Kim, T. H., Choi, S. J., Lee, Y. H., Song, G. G., & Ji, J. D. (2014). Gene expression profile predicting the response to anti-TNF treatment in patients with rheumatoid arthritis; analysis of GEO datasets. Joint Bone Spine, 81(4), 325-330. (GEO2R)
11. Lundstrom, K. Replicon RNA viral vectors as vaccines. Vaccines. 2016;4:39. doi: 10.3390/vaccines4040039.
12. Philip V Kovski, Annika Kratzil, Silvio Steiner, Hanspeter Stalder, Volker Thiel. Corona Virus and Replication: Implications for SARS-CoV-2
13. Rawaa S. Al-Kayali, Mohomed F. Kashkash, Azzam H. Alhussein Alhajji, Abdullah Khouri. Activation of tuberculosis in recovered Covid-19 patients: a case report, 2023
14. Raman R., Patel K. J., Ranjan K. COVID-19: Unmasking Emerging SARS-CoV-2 Variants, Vaccines and Therapeutic Strategies. Biomolecules. 2021 Jul 06;11(7)
15. Rosa, S. S., Prazeres, D. M., Azevedo, A. M., & Marques, M. P. (2021). mRNA vaccines manufacturing: Challenges and bottlenecks. Vaccine, 39(16), 2190-2200.
16. Rosa SS, Prazeres DMF, Azevedo AM, Marques MPC. mRNA vaccines manufacturing: Challenges and bottlenecks. Vaccine. 2021 Apr 15;39(16):2190-2200. doi: 10.1016/j.vaccine.2021.03.038. Epub 2021 Mar 24. PMID: 33771389; PMCID: PMC7987532.
17. Ross, J. mRNA stability in mammalian cells. Microbiol Mol Biol Rev. 1995;59:423–450.
18. Schlake, T., Thess, A., Fotin-Mleczek, M., Kallen, K.-J. Developing mRNA-vaccine technologies. RNA Biol. 2012;9(11):1319–1330. doi: 10.4161/rna.22269
19. Schoch, C. L., Ciufo, S., Domrachev, M., Hotton, C. L., Kannan, S., Khovanskaya, R., … & Karsch-Mizrachi, I. (2020). NCBI Taxonomy: a comprehensive update on curation, resources and tools. Database, 2020, baaa062.

20. Sharma A., Ahmad Farouk I., Lal SK. COVID-19: A Review on the Novel Coronavirus Disease Evolution, Transmission, Detection, Control and Prevention. Viruses. 2021 Jan 29;13(2)
21. Vogel A. B., Lambert L., Kinnear E., Busse D., Erbar S., Reuter K. C., et al. Self-Amplifying RNA Vaccines Give Equivalent Protection against Influenza to mRNA Vaccines but at Much Lower Doses. Mol Ther. 2018;26(2):446–455. doi: 10.1016/j.ymthe.2017.11.017
22. Walensky RP, Walke HT, Fauci AS. SARS-CoV-2 Variants of Concern in the United States-Challenges and Opportunities. JAMA. 2021 Mar 16;325(11):1037-1038.
23. Wiersinga W. J., Rhodes A., Cheng A. C., Peacock S. J., Prescott H. C. Pathophysiology, Transmission, Diagnosis, and Treatment of Coronavirus Disease 2019 (COVID-19): A Review. JAMA. 2020;324(8):782–793. doi:10.1001/jama.2020.12839
24. Zhang Y, Bharathi V, Dokoshi T, de Anda J, Ursery LT, Kulkarni NN, Nakamura Y, Chen J, Luo EWC, Wang L, Xu H, Coady A, Zurich R, Lee MW, Matsui T, Lee H, Chan LC, Schepmoes AA, Lipton MS, Zhao R, Adkins JN, Clair GC, Thurlow LR, Schisler JC, Wolfgang MC, Hagan RS, Yeaman MR, Weiss TM, Chen X, Li MMH, Nizet V, Antoniak S, Mackman N, Gallo RL, Wong GCL. Viral afterlife: SARS-CoV-2 as a reservoir of immunomimetic peptides that reassemble into proinflammatory supramolecular complexes. Proc Natl Acad Sci U S A. 2024 Feb 6;121(6):e2300644120. doi: 10.1073/pnas.2300644120. Epub 2024 Feb 2. PMID: 38306481; PMCID: PMC10861912.

Rithvik Lal

Using Bioinformatics to Uncover the Gene Expression of Genes in Affected and non Affected Patients with Cystic Fibrosis

ABSTRACT

Background: Cystic fibrosis (CF) is a genetic disorder caused by mutations in the CFTR gene, leading to the production of thick, sticky mucus that primarily affects the lungs and digestive system. This results in chronic respiratory infections, difficulty breathing, digestive issues, and reduced life expectancy. While treatments like CFTR modulators have improved outcomes, challenges remain, including limited access to therapies, progressive lung damage, and the lack of treatments for rare mutations. Researchers aim to better understand CFTR gene expression to develop targeted therapies, explore gene-editing technologies, and personalize treatments, ultimately improving quality of life and reducing the burden of this disease.

Methods: Bioinformatics tools and databases were used to analyze MRSA data. Data was collected from the National Center for Biotechnology Information. Gene Expression Omnibus (GEO), which included 2 groups, affected and non affected. Differentially expressed genes (DEGs) were identified by removing data with a p-value higher than 2.20e-12. The top 50 DEGs were analyzed using ShinyGO, which incorporated KEGG.

Results: The original total number of genes was 18331, which was then analyzed down to the top 50 DEG's, with major expression across the group. The use of ShinyGO showed 2 key pathways: Cell adhesion Molecules, and Asthma which showed key genes such as MHC, NLGN4Y, MHC-2, which are associated with inflammation and immune responses.

Conclusion: The study finds genes that could be markers of cystic fibrosis in patients and indicate inflammation and immune deficiencies, These genes can be later tested in laboratories to find possible therapies for CF.

KEYWORDS: Cystic fibrosis, MRSA, NCBI, GEO, KEGG, GO, MHC, NLGN4Y, MHC-2, Cell adhesion molecules, Asthma

INTRODUCTION

Cystic Fibrosis is an autosomal recessive disorder that targets the CFTR protein causing the build up of Air Surface Liquid (ASL). These fluids, also called secretions, are usually thin and slippery to protect the body's internal tubes and ducts and make them smooth pathways, But in people with CF, a changed gene causes the secretions to become sticky and thick. This causes the blocking of pathways to locations such as the lungs, digestive tract, and other organs (1)

What can be done to curve the significant abnormalities when it comes to CFTR. Understanding the Gene in the long arm of chromosome 7. (1) The gene mutation that causes CF has been discovered but currently researchers must identify ways of affecting said gene. Since 2012, recent approaches have enabled the identification of small molecules targeting either the CFTR protein directly or its key processing steps, giving rise to novel promising therapeutic tools yet there is nothing concrete.(2)

The disease is genetic in nature and is most commonly found in the caucasian race more than any other race, and remains one of the most common life-shortening disorders in the white population, with an estimated median survival age of 33.4 years in the US. There is no cure but options may become more available due to continued study of the disorder. (2)

It occurs due to a mutation in the long arm of Chromosome 7 and focuses on cells that produce mucus and sweat in the body. This causes issues with a protein called CFTR.(1)

Patients with the same CFTR mutation can exhibit vastly different disease severities. This variability may be influenced by modifier genes, epigenetics, and environmental factors, which are not fully understood.

The goal of this research is to use bioinformatics to identify and expand the knowledge of cystic fibrosis (CF). Specifically, we used NCBI GEO, ShinyGo, KEGG, and GO bioinformatics tools.NCBI GEO2R is an interactive online tool provided by the National Center for Biotechnology Information that allows users to compare two or more groups of samples in a Gene Expression

Omnibus (GEO) dataset to identify differentially expressed genes. In addition, ShinyGo, KEGG, and GO bioinformatics tools were used to further analyze the results obtained from GEO2R. ShinyGo is an online platform for statistical and graphical visualization of biological data, helping to explore patterns, correlations, and trends in high-throughput data.

KEGG (Kyoto Encyclopedia of Genes and Genomes) is a comprehensive database that integrates genomic, chemical, and systemic functional information, enabling the understanding of high-level functions and utilities of biological systems, such as pathways and networks (14).GO (Gene Ontology) is a framework for the standardized representation of gene and gene product attributes across species and databases, providing a structured vocabulary for biological processes, cellular components, and molecular functions (15).These bioinformatics tools together allowed for a comprehensive analysis and interpretation of the datasets, contributing to the expansion of our understanding of CF-related genetic and molecular mechanisms.

Our hypothesis is that there will be a significant difference in gene expression when we compare healthy and infected samples. Advancements in CF treatment have significantly increased life expectancy, which was once only a few years post-diagnosis. Continued research can help further extend and improve the quality of life for patients.

METHODS

Collection of Datasets

GEO2R (www.ncbi.nlm.nih.gov/geo/geo2r/) is an online tool designed for comparing and analyzing gene expression across different sample groups. It shows data as visual data charts Accessed the NCBI library. I used the GEO database to search for cystic fibrosis microarray data. I identified a dataset from research on cystic fibrosis that found that CFTR deficiency in cystic fibrosis disrupts lipid metabolism, promotes chronic oxidative stress, and drives a persistent pro-inflammatory state in airway epithelial cells. I then processed this GEO dataset for my study.

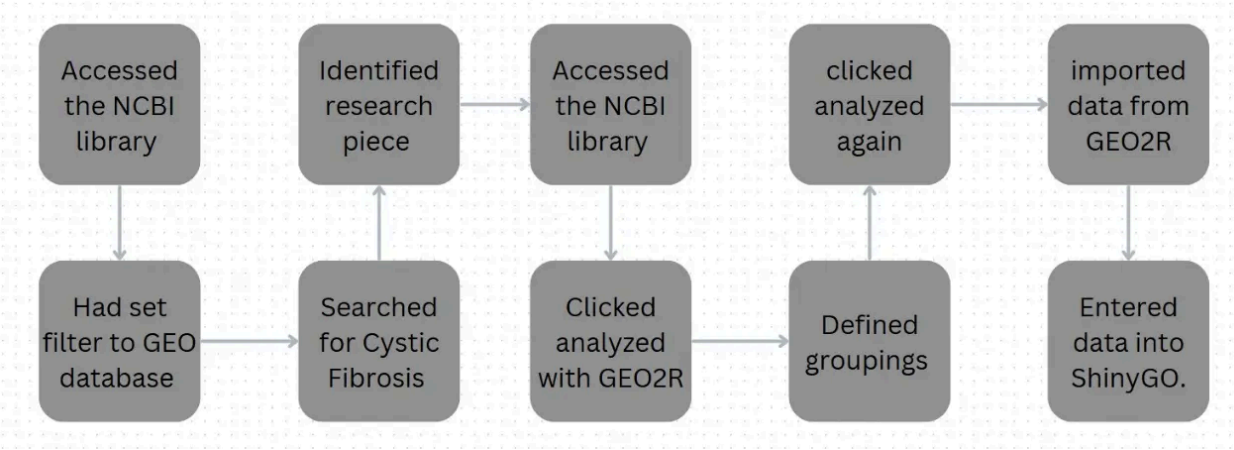


Figure 1: Methods Flow chart. This flow chart shows the methodological approach used in this research.

Data Analysis of GEO2R Results

To process the datasets, I first defined or categorized the data into two groups; affected patients and non-affected patients, and then utilized the no-code GEO2R 'Analyze' tool, which operates with pre-programmed R scripts. Analysis with this tool generated multiple graphs of gene expression results which we analyzed. Using the volcano plot, I examined gene expression differences between the various sample groups. From the Venn diagram results, I evaluated the number of expressed genes and identified overlapping genes among the different human sample groups.

Narrowing down the number of differentially expressed genes (DEGs) to top 50

Next, I applied statistical analysis to identify the top 50 most significantly expressed genes from my samples. I downloaded the top differentially expressed genes (DEGs) data into a Google Sheet and selected the top 50 DEGs by removing any data with a p-value higher than 2.20e-12.

Further Analysis of Top Differentially Expressed Genes using ShinyGO, KEGG and GO Bioinformatics tools

To analyze the top 50 identified DEGs and determine the functions and participating key pathways for key genes, I used the ShinyGo bioinformatics tool. ShinyGO is a online bioinformatics program that allows for the creation of KEGG pathway visualizations when imputing large sets of gene data. ShinyGo also generates Gene ontology or GO results to help determine potential functions of key genes from the top 50 DEGs. Altogether, ShinyGo,

the GO bioinformatics and KEGG databases (7, 14, 15) were used to find out what the top genes potentially do in patients affected and non-affected with Cystic Fibrosis.

RESULTS

Identification of Genes Expressed Differently in the CF Samples

The GEO2R bioinformatics tool was utilized to identify differentially expressed genes (DEGs). The results were visualized using a volcano plot and a Venn diagram shown in Figure 2. In the volcano plot comparing affected versus non affected versus. The genes were observed to be differentially expressed. A portion of the genes were very even, while a slight number displayed lower expression levels. In the plot, red dots represent highly expressed (upregulated) genes, blue dots indicate less expressed (downregulated) genes. The venn Diagram produced a total of 7229 genes expressed between each other out of the total of 18331 genes in the data set.

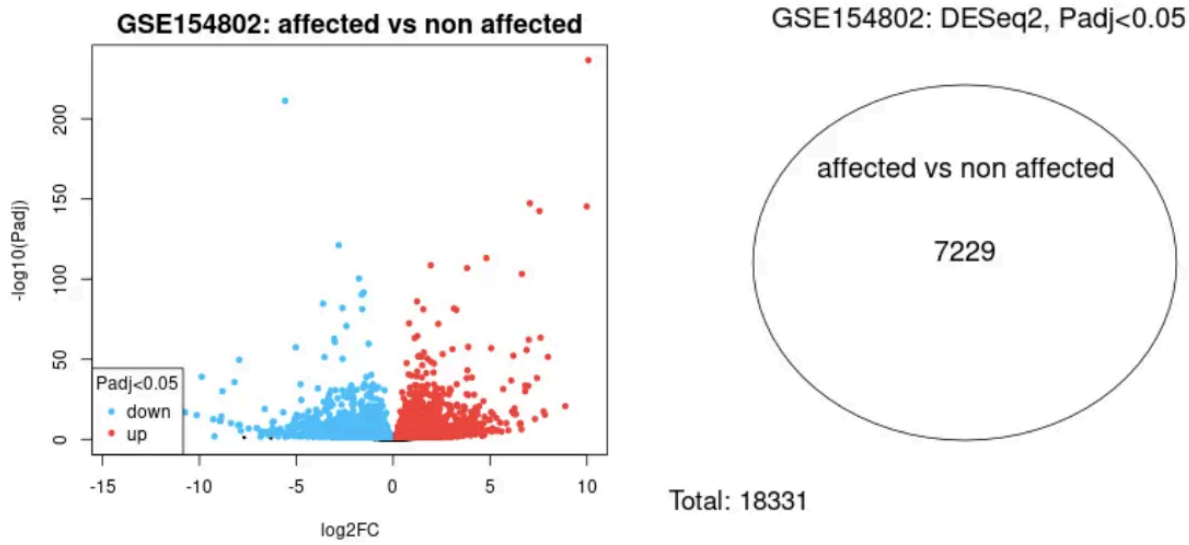


Figure 2 Volcano diagram and Venn diagram. **(A)** The volcano map of GSE154802 . **(B)** The volcano map of GSE40611. Upregulated genes are colored in red; downregulated genes are colored in blue. **(C)** The two datasets showed an overlap of 7229 genes.

Statistical Analysis of (50) Differentially Expressed Genes (DEGs)

To identify the top 50 DEG'S from 18331 total any p-value greater than 2.20e-12 was excluded. geo2r data.

Enrichment Analysis to Determine Potential Functions of the Identified Genes

Through ShinyGO I identified two pathways, Cell adhesion Molecules and asthma from the KEGG analysis. Through this I identified the key genes of MHC-2, and NLGN4Y **(Figure 3)**

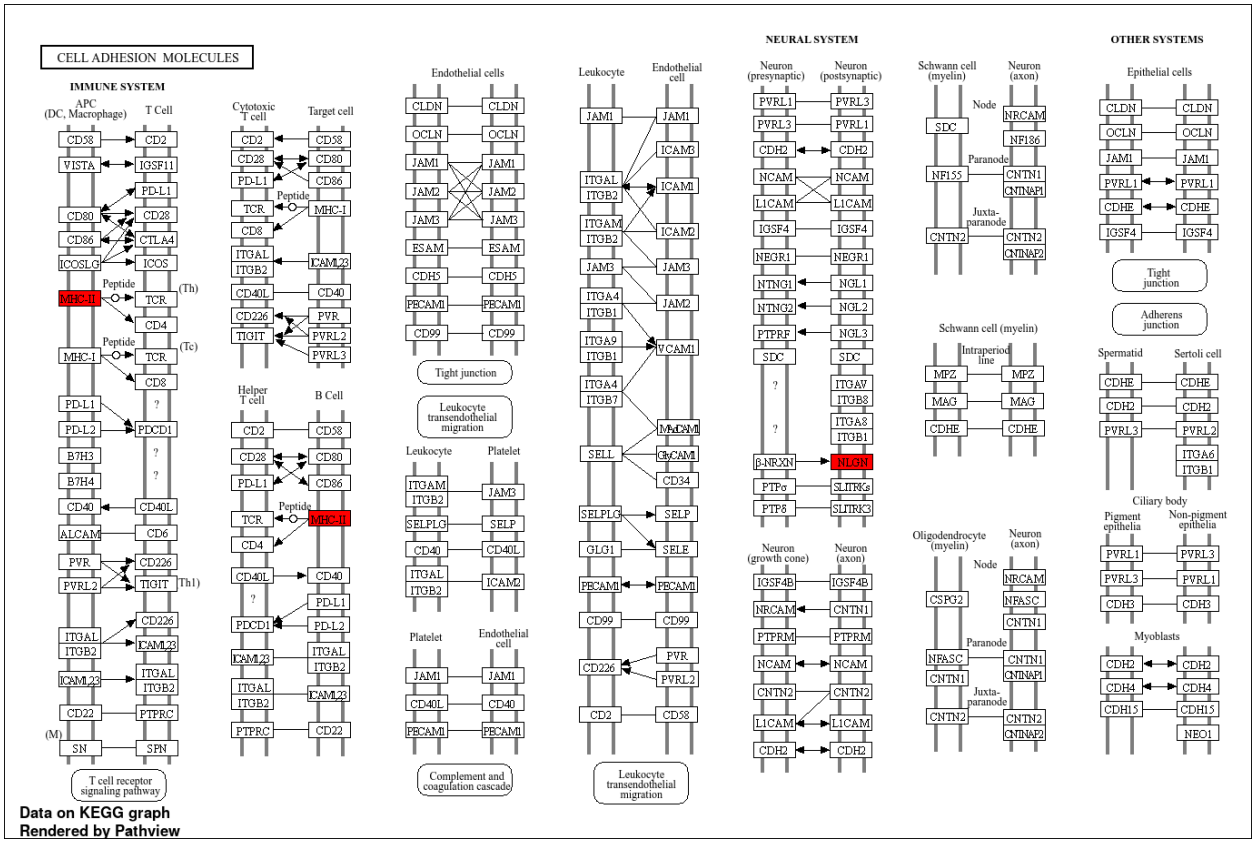


FIGURE 3: Key pathway and key genes associated with the top 50 DEGs: Cell adhesion Molecules. Key genes highlighted in Red: MHC-2, and NLGN4Y

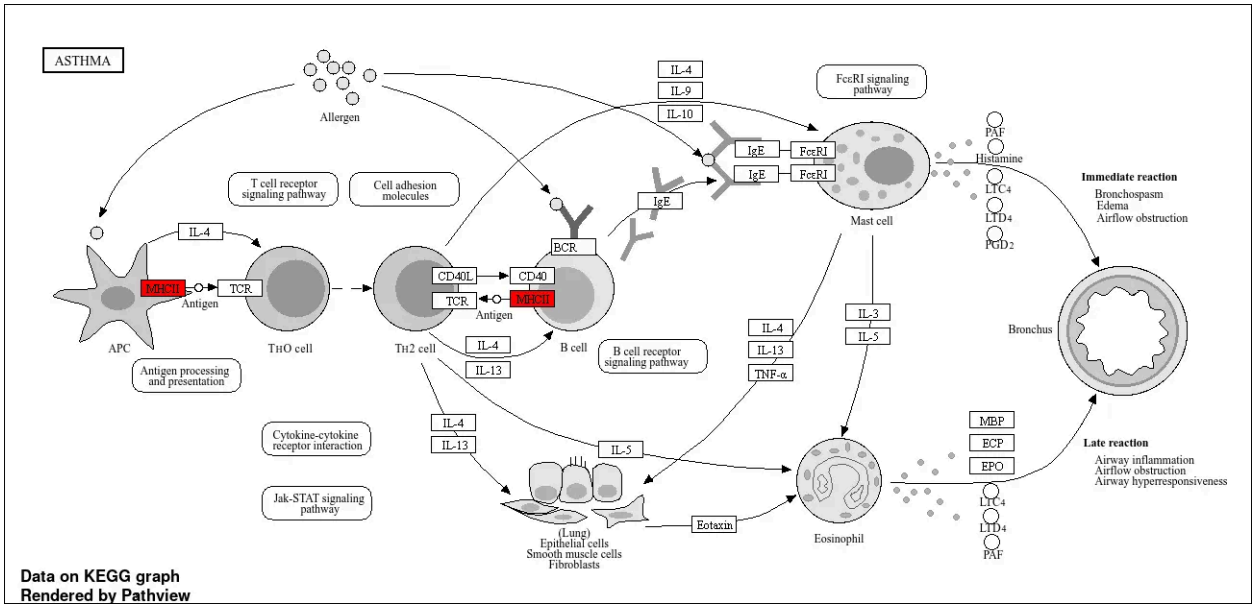


FIGURE 4: Key pathway and key genes associated with the top 50 DEGs: Asthma. Key genes highlighted in Red: MHC II

DISCUSSION

The main goal of the research paper is to identify genes expressed differently in individuals with cystic fibrosis and without. During which genes related to cell adhesion and Asthma **(Figure 3 and 4)** both exhibited severe differences compared to each other. The original hypothesis of the study was that there would be significant differences in expression of genes which was half proven correct due to the findings showing few gene differences but rather major differences **(Figure 2).**

The findings of our the study (Table 1) are backed up by other studies as several discuss CF's relation to the epithelial cells (4). First, using ShinyGo functional and enrichment analysis bioinformatics tool, our study identified cell adhesion molecules and Asthma as the key pathways associated with the top 50 DEGs identified.

Table 1: Summary of the prevalent genes and enriched pathways found in this analysis analysis

Key Genes	Key Pathways	Connection to Cystic Fibrosis

MHC-II	Cell adhesion Molecules/Asthma	MHC-II molecules are critical for presenting antigens to CD4+ T cells, facilitating the adaptive immune response. In CF, chronic lung infections (e.g., with Pseudomonas aeruginosa or Staphylococcus aureus) lead to heightened immune activation. Variations in MHC-II genes can influence immune response intensity and may determine susceptibility to recurrent infections and inflammation in CF patients.
NLGN4Y	Cell adhesion Molecules	Neuroligins are synaptic adhesion molecules involved in neural development and function. While NLGN4Y is not directly implicated in the pathology of CF, its relevance may emerge in studying CF-associated comorbidities, such as cognitive or neurodevelopmental issues linked to chronic hypoxia or inflammation. For instance, chronic illness and systemic inflammation in CF could indirectly affect neurological

		development or function, where molecules like NLGN4Y might play a modulatory role.
MHC	Asthma	MHC system or an immune-related protein, its role could involve modulating the immune response in CF

Cystic fibrosis has been documented to be linked to cell adhesion molecules (11). Adhesion molecules are cell surface proteins that are involved in binding cells to one another with an extracellular matrix (ECM) (16). In layman's terms it acts like molecular glue helping cells stick to one another. For example, ATS journal noted in a study how the adhesion molecules may be linked to an increase in inflammation in CF patients (11). The NLGN4Y gene is defined as a key regulator of cell adhesion and immune response modulation, playing a critical role in cellular interactions within the extracellular environment (12). It functions by encoding a protein that facilitates intracellular binding and immune signaling. The connection of NLGN4Y to CF is that its altered expression may contribute to heightened inflammation and chronic infections seen in CF patients, particularly in the lungs, by affecting the adhesion properties of epithelial cells and immune cell interactions (12).

The MHC-2 class molecule facilitates the adaptive immune response by presenting antigens to T-cells. (Its major connection to CF is that, when patients with CF experience chronic lung infections (e.g., with Pseudomonas aeruginosa or Staphylococcus aureus), it leads to heightened immune activation. This persistent immune response can contribute to lung damage and disease progression. Further research in laboratory settings is needed to fully elucidate this connection. Asthma was another key pathway identified in our study. Asthma is defined as a chronic inflammatory airway disease characterized by reversible airway obstruction, bronchial hyperresponsiveness, and increased mucus production (13). It connects to CF because both conditions involve airway inflammation, mucus hypersecretion, and immune dysregulation. This also requires further testing in a laboratory environment.

Conclusion and Future Direction

By studying these key pathways and genes, researchers can uncover potential target pathways for personalized interventions, such as immunomodulators or anti-inflammatory therapies, to reduce complications in CF. The identified genes can be tested in the laboratory by scientists in the laboratory or clinical trials to determine if possible cures can be identified from it.

Limitations

In our case, since we rely on bioinformatics datasets derived from microarray experiments conducted by other researchers, a key limitation is that the identified genes will require further investigation in laboratory or clinical settings before progressing toward vaccine development.

References

1. Gibson, R. L., Burns, J. L., & Ramsey, B. W. (2003). Pathophysiology and management of pulmonary infections in cystic fibrosis. American journal of respiratory and critical care medicine, 168(8), 918-951.
2. Fanen, P., Wohlhuter-Haddad, A., & Hinzpeter, A. (2014). Genetics of cystic fibrosis: CFTR mutation classifications toward genotype-based CF therapies. The international journal of biochemistry & cell biology, 52, 94-102.
3. Castellani, C., & Assael, B. M. (2017). Cystic fibrosis: a clinical view. Cellular and molecular life sciences, 74(1), 129-140.
4. Rommens, J. M., Iannuzzi, M. C., Kerem, B. S., Drumm, M. L., Melmer, G., Dean, M., … & Collins, F. S. (1989). Identification of the cystic fibrosis gene: chromosome walking and jumping. Science, 245(4922), 1059-1065.
5. Saiman, L., & Siegel, J. (2004). Infection control in cystic fibrosis. Clinical microbiology reviews, 17(1), 57-71.
6. Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, Alexandra Soboleva, NCBI GEO: archive for functional genomics data sets—update, Nucleic Acids Research, Volume 41, Issue D1, 1

January 2013, Pages D991–D995, https://doi.org/10.1093/nar/gks1193
7. Minoru Kanehisa, Susumu Goto, Shuichi Kawashima, Akihiro Nakaya, The KEGG databases at GenomeNet, Nucleic Acids Research, Volume 30, Issue 1, 1 January 2002, Pages 42–46, https://doi.org/10.1093/nar/30.1.42
8. Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, et al. SRplot: A free online platform for data visualization and graphing. PLoS ONE [Internet]. 2023 Nov 9 [cited 2025 Jan 13];18(11):e0294236–6. Available from: https://pubmed.ncbi.nlm.nih.gov/37943830/
9. Smith B, Williams J, Schulze-Kremer S. The ontology of the gene ontology. AMIA Annu Symp Proc. 2003;2003:609-13. PMID: 14728245; PMCID: PMC1480173.
10. C;, Marion CR;Izquierdo M;Hanes HC;Barrios. "Asthma in Cystic Fibrosis: Definitions and Implications of This Overlap Syndrome." Current allergy and asthma reports. Accessed January 21, 2025. https://pubmed.ncbi.nlm.nih.gov/33560464/.
11. Circulating adhesion molecules in cystic fibrosis | American Journal … Accessed January 21, 2025. https://www.atsjournals.org/doi/full/10.1164/ajrccm.157.4.9704134.
12. "NLGN4Y Neuroligin 4 Y-Linked [Homo Sapiens (Human)] – Gene – NCBI." National Center for Biotechnology Information. Accessed January 21, 2025. https://www.ncbi.nlm.nih.gov/gene/22829.
13. Jesenak M;Durdik P;Oppova D;Franova S;Diamant Z;Golebski K;Banovcin P;Vojtkova J;Novakova E; "Dysfunctional Mucociliary Clearance in Asthma and Airway Remodeling – New Insights into an Old Topic." Respiratory medicine. Accessed January 21, 2025. https://pubmed.ncbi.nlm.nih.gov/37516275/.
14. Kanehisa M;Furumichi M;Sato Y;Kawashima M;Ishiguro-Watanabe M; "Kegg for Taxonomy-Based Analysis of Pathways and Genomes." Nucleic acids research. Accessed January 26, 2025. https://pubmed.ncbi.nlm.nih.gov/36300620/.
15. Ashburner M;Ball CA;Blake JA;Botstein D;Butler H;Cherry JM;Davis AP;Dolinski K;Dwight SS;Eppig JT;Harris MA;Hill DP;Issel-Tarver L;Kasarskis A;Lewis S;Matese JC;Richardson JE;Ringwald M;Rubin GM;Sherlock G; "Gene Ontology: Tool for the Unification of Biology. the Gene Ontology Consortium." Nature genetics. Accessed January 26, 2025. https://pubmed.ncbi.nlm.nih.gov/10802651/.

16. Alberts, Bruce. “Cell-Cell Adhesion.” Molecular Biology of the Cell. 4th edition., January 1, 1970. https://www.ncbi.nlm.nih.gov/books/NBK26937/.

Lauren Lee

The Association Between Family Meals and Cardiovascular Disease

ABSTRACT

Background: Although many socioeconomic risk factors for cardiovascular disease have been extensively analyzed, factors related to family dynamics and household meal consumption have yet to be better understood. This scoping literature review was conducted to provide a comprehensive understanding of the association between family meals and the risk of cardiovascular disease.

Methods: A scoping review of Google Scholar and PubMed with keywords related to family meals and their impacts on cardiovascular disease was utilized. Some keywords were "cardiovascular disease" and "family meals". A broad inclusion criteria was used, including articles from any geographical location. Literature articles were excluded.

Results: Five articles were reviewed, from a broad review of social relationships and cardiovascular disease to a more focused study that analyzes specific family diet interventions to mitigate cardiovascular risk. Among these articles are topics of the frequency of family meals as well as having a meal companion, both of which have a positive association with cardiovascular outcomes. We also found that an increased frequency of shared mealtimes in families had a positive impact on cardiovascular health.

Discussion: This scoping review suggests that family meal patterns and positive meal consumption habits, which are influenced by social interactions with close relatives, help reduce cardiovascular disease risks. Social interactions have a strong impact on individuals' food selections and the way their meals are consumed. However, further studies would be needed to validate these findings.

INTRODUCTION

It is well known that some diets are closely related to an increase in risk of cardiovascular disease, including diets of high cholesterol, excess calories, and lack of fruits or vegetables. Food selection may initially seem to be a very personal decision, but evidence suggests that humans, as social creatures, interact with others which often leads to changes in the type of food consumed as well as our eating habits [1]. Social interaction in an environment where family members care for each other's health, along with

the selection of healthy diet and meal habits, further reinforces a favorable impact on cardiovascular risk.

Cardiovascular disease (CVD) is the leading cause of death globally, responsible for around 17.9 million deaths each year [2]. Morbidities associated with CVD are often caused by heart attacks and strokes, and even in the case of survival, overall health and lifespan are diminished [3]. In 2021, 1 in every 5 deaths in the USA were due to CVDs [3]. The World Health Organization (WHO) estimates that over 75% of premature CVD is preventable [4].

Dietary habits are well established to have a strong correlation with developing CVDs. For example, a high intake of salt, fat, and excess calories has frequently been shown to lead to CVDs [3]. Many of these factors are related to not only personal but also interpersonal dynamics, especially in the setting of a household.

Dietary patterns are closely tied to the choices people make on what to eat as a family. These choices have a significant impact on major CVD risk factors [3]. There is a paucity of literature reviews on this matter; with this scoping review, the goal is to synthesize our understanding of CVD and family meals.

METHODS

Google Scholar and PubMed were used to complete this scoping review. The key search terms used were “cardiovascular disease”, “cardiovascular health”, “family meals”, and “family support”. Articles that were focused on family meal dynamics and their impacts on cardiovascular health, regardless of age, sex, or geographical location were included and literature reviews were excluded.

RESULTS

The study “Social Relationships and Cardiovascular Disease Risk Factors: Findings from the Third National Health and Nutrition Examination Survey” found trends between risk factors for cardiovascular disease and the frequency of interactions in individual and organizational relationships [5]. The frequency of interactions was defined with a questionnaire of how many times per week the subject visited or called friends/family and attended

clubs or church [5]. The five risk behaviors studied were smoking, physical inactivity, inadequate fruit and vegetable consumption, not having a blood pressure check within the previous 12 months, and not having a cholesterol check in 12 months [5]. With a more frequent presence of individual or organizational relationships, study participants tended to eat healthy amounts of vegetables, get their blood checked more often, and tended to smoke less (or not at all) [5].

The influence of social relationships is shown to be a factor in the overall health of a family. In the study "Effects of a Family Diabetes Self-Management Education Intervention on the Patients' Participating Supporters", family members of patients with diabetes were encouraged to attend diabetes education classes along with their patients and follow the same meal recommendations as the patient [6]. The study concluded that shared meals and exposure to health education of an entire family (instead of just the patient) would therefore increase the health of the participating family members along with the patient [6]. The participants' health was indicated through BMI, physical activity, and rates of consumption of healthy or unhealthy food [6].

The inherent connection between a patient's health and the health of their residing family members is, in part, due to shared meals. The frequency of shared mealtimes in families was shown to have an impact on cardiovascular health in the study "Effects of Family Meal Frequency on Risk Factors for Cardiovascular Disease in Korean Elderly Males and Females" [7]. Data was extracted from the Fifth Korea National Health and Nutrition Examination Survey, which concerned elderly Koreans and their risk factors for cardiovascular disease such as blood pressure, BMI, and cholesterol levels [7]. As the frequency of meals they shared with family or friends increased, their blood sugar, cholesterol, and BMI reached healthier amounts; for example, as the number of shared meals increased by one standard deviation, the interpretation of the path coefficient indicated that the standard deviation of blood sugar in males decreased by 0.024, the BMI in males decreased by 0.38, and the total cholesterol in females decreased by 0.047 [7].

This pattern was seen in another Korean study, "Association between eating alone and cardiovascular diseases in elderly women: a cross-sectional study

of KNHANES 2016 data" [8]. As women ate more meals alone, they were less aware of nutrition labels [8]. Just 29.5% of menopausal women in Korea who ate more than two meals a day alone were aware of nutrition labels, compared to an awareness of 43.3% of women who ate more than two meals a day with others [8]. Women who ate alone were 2.61 times more likely to have angina than women who ate with others [8]. The presence of an "eating companion" was therefore shown to be in improving CVD outcomes in menopausal women [8]. Eating alone was associated with poor food choices, such as fatty and processed foods, as well as the consumption of excess calories [8]. Having a "companion" to eat with encouraged healthier meal choices and meal consumption behavior [8].

DISCUSSION

Study findings suggest that positive interpersonal relationships have favorable impacts on modifying assessed cardiovascular risk factors, including the choice of a healthy diet [5]. Additionally, even the mere presence of an "eating companion" was shown to work in favor of reducing cardiovascular risk [8]. Elderly citizens who shared 3 meals per day with their family had better results than those who only shared 1 or no meals a day with others [7]. Though a lack of participation hindered some testing results, a connection is likely to emerge in future trials between a patient's diet regimen and the health of other family members living with them [6].

A limitation of the study is the paucity of experimental research papers related to the association between family meals and CVD risk. Future studies are needed to further support these findings. The geographic diversity of the study also makes it difficult to standardize the findings accounting for each country's different diet cultures.

A prime example of this is "A three-arm randomized controlled trial using ecological momentary intervention, community health workers, and video feedback at family meals to improve child cardiovascular health: the Family Matters study design", which has not published results yet [9]. The purpose of the Family Matters study design is to see how much of an impact the frequency and quality of family meals would have on cardiovascular disease risks in American children [9]. The study also aims for behavioral changes by using real-time interventions: text messages would be sent out inquiring about the family meal planner's stress and what sort of meal they had planned [9]. A community health worker might be sent to help make a meal

more healthy [9]. They predicted that these interventions would improve a child's BMI percentile and diet quality, as well as that of their family members, effectively reducing the general risk of cardiovascular disease [9].

This is a prospective study in the pediatric population that studies changes in cardiovascular risk through family dynamics and modulation of food selections; however, a prospective study in the adult population will be needed to further validate the findings for the adult population.

References

1. Lichtenstein, A. H., Appel, L. J., Vadiveloo, M., Hu, F. B., Kris-Etherton, P. M., Rebholz, C. M., Sacks, F. M., Thorndike, A. N., Van Horn, L., & Wylie-Rosett, J. (2021). 2021 Dietary Guidance to Improve Cardiovascular Health: A scientific statement from the American Heart Association. Circulation, 144(23).
2. World Health Organization: WHO. (2019, June 11). Cardiovascular diseases.
3. Heart disease risk factors. (2024, May 15). Heart Disease.
4. Stewart J., Manmathan G., Wilkinson P. Primary prevention of cardiovascular disease: A review of contemporary guidance and literature. JRSM Cardiovascular Disease, 2017; 6
5. Ford, E. S., Ahluwalia, I. B., & Galuska, D. A. (2000). Social Relationships and Cardiovascular Disease Risk Factors: Findings from the Third National Health and Nutrition Examination Survey. Preventive Medicine, 30(2), 83–92
6. Felix, H. C., Narcisse, M. R., Long, C. R., & McElfish, P. A. (n.d.). Effects of a family diabetes self-management education intervention on the patients' supporters. Families Systems & Health, 38(2), 121–129.
7. Jeong, T., Jung, M., Lee, J., & Kim, T. (2015, January 1). Effects of family meal frequency on risk factors for cardiovascular disease in Korean elderly males and females. PubMed Central (PMC)
8. Choi, H., Kim, H., & Kang, S. (2021). Association between eating alone and cardiovascular diseases in elderly women: a cross-sectional study of KNHANES 2016 data. Menopause, 29(1), 82–88.
9. Berge, J. M., Trofholz, A. C., Aqeel, M., Norderud, K., Tate, A., Fertig, A. R., Loth, K., Mendenhall, T., & Neumark-Sztainer, D. (2023). A three-arm randomized controlled trial using ecological momentary intervention, community health workers, and video feedback at family

meals to improve child cardiovascular health: the Family Matters study design. BMC Public Health, 23(1).

Samreen Mehta

Using Bioinformatics Tools to Identify Possible Target Genes for Alzheimer's Disease Treatments

ABSTRACT

Background: Alzheimer's disease is a progressive, advanced brain disorder that mainly targets older adults. Alzheimer's disease causes the accumulation of amyloid plaques which significantly affects memory, thoughts, and behavior. With no cure, this study aims to find target genes for the treatment of Alzheimer's disease using bioinformatics to analyze gene enrichment and expression.

Methods: I utilized bioinformatics tools including NCBI GEO2R and SR Plot, for analysis of KEGG and GO, to study dataset GSE52022. Within the dataset, I defined the samples into groups entitled littermate and mutant. GEO2R was used to identify differentially expressed genes, these differentially expressed genes were then further analyzed using SR Plot for GO and KEGG pathway enrichment.

Results: Genes with a P-value greater than 0.5 were excluded to find DEGS, and fold change values were utilized to identify upregulated and downregulated genes. There were 30 top Differentially Expressed Genes, with 15 upregulated genes and 15 downregulated genes. The GO analysis indicated enrichment in the biological process ontology, and importation into the cell was the most enriched function. The GO analysis also revealed key genes including TREM2 and TYROBP. KEGG analysis emphasized the Tuberculosis pathway and genes including Dectin-1 (CLEC7A), CALM, and CathepsinS.

Conclusion: Genes like TREM2, TYROBP, and CLEC7A are vital in microglial activation and neuroinflammation in Alzheimer's disease. Targeting these genes may enhance the clearance of amyloid-beta plaques and reduce neurodegeneration, providing insights that could inform future research and therapeutic strategies.

INTRODUCTION

Despite significant advancements in our knowledge of Alzheimer's Disease (AD), one persistent issue is the absence of a cure or disease-modifying treatments, which is what this research aims to address (1). The scientific

question that is being investigated is determining what genes are involved in the risk of developing Alzheimer's disease.

Alzheimer's disease is a progressive, advanced brain disorder that mainly targets older adults (2). Further, AD causes the accumulation of amyloid plaques which has significant effects on memory, thoughts, and behavior. It is the leading cause of dementia, caused by the buildup of abnormal proteins that disrupt brain function. Over time, it can lead to severe cognitive decline, preventing individuals from living independently. Alzheimer's disease has impacted over 15 million people nationwide and poses a dangerous public health crisis (3), as cases of this disease expand, it will continue to put financial and emotional strain on family members and healthcare organizations. Therefore, addressing and understanding Alzheimer's is a staple in decreasing its harmful impact and improving brain health across the board.

So far, we know that Alzheimer's disease causes progressive cognitive decline, which gravely impacts the amount of care and treatment that a diagnosed individual would need.

Bioinformatics studies on Alzheimer's disease aim to better comprehend the biological mechanisms driving the conversion of normal/MCI into AD pathology, which may ultimately open the door to disease-modifying treatments or cures (4). Research has identified multiple genes that are linked to Alzheimer's risk and examined how genetic variations impact the progression of the disease. Bioinformatics tools, including NCBI and GEO2R (5), analyze data including brain imaging (6), protein interaction, and even modified mice to find potential indicators of the disease. Moreover, these bioinformatics tools aid with predicting early signs of Alzheimer's which supports the process of treatment strategies.

A major setback with Alzheimer's is that there is currently no cure for this disease (1). While there are medications that people can take to reduce the symptoms, they do not offer a permanent cure or stop the progression of Alzheimer's (1). This impacts bioinformatics studies because researchers have devoted innumerable hours to identifying genes and molecular mechanisms that can lead to a treatment or a cure for this disease.

The goal of my research is to help identify significant genes with altered expressions that can help us identify treatment for this disease.

METHODS

Data Collection and Analysis of GEO2R Data

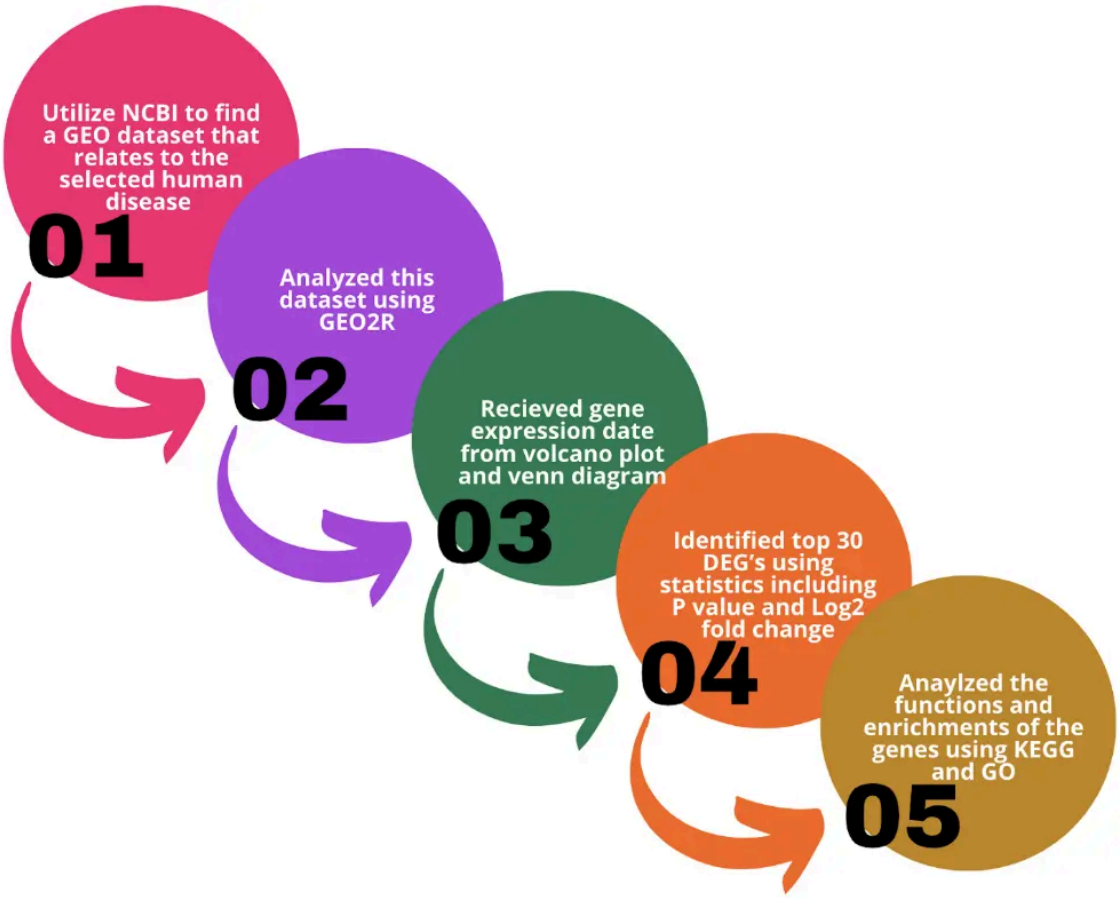


Figure 1: Research Methodology: Provides a summary of the procedures used in this study.

IDENTIFICATION OF THE TOP DIFFERENTIALLY EXPRESSED GENES

To identify the most significant differentially expressed genes (the top 30 or 40), statistical analysis was applied. This process used p value and log fold change to prioritize the most important genes based on their differential expression across samples.

Data Analysis Using SRPlot, KEGG, and GO Bioinformatics Tools

Then SRPlot (7), KEGG (8), and GO (9) bioinformatics tools and databases were utilized to analyze the functions of these top genes. SRPlot is a

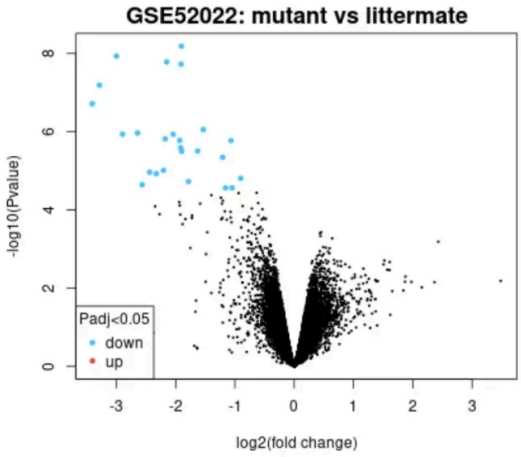
bioinformatics tool that provides a wide variety of plots and different graph types for scientists to analyze (7). The KEGG bioinformatics tool that organizes information regarding genes and their behaviors (8). The GO tool is another large resource meant for cataloging genes and their functions (9). Altogether, these tools helped uncover the potential roles of the genes in Alzheimer's Disease.

RESULTS

Identification of Differentially Expressed Genes

The GEO2R bioinformatics tool was utilized to determine differentially expressed genes (DEGs). The visualization of the analysis results was in the form of a Venn diagram and a volcano plot. Based on the results from the volcano plot, genes that were expressed differently between the two groups, mutant and littermate, can be identified.

In this plot, the blue dots represent genes that were downregulated; and less active. The black dots represent genes that did not display significant differences in expression. In this case, there are no red dots, but if there were, the red dots would be a representation of genes that are upregulated; and more active. Originally, the Venn diagram produced a sum of 45076 genes. Furthermore, 25 genes overlapped each other within the sample groups.



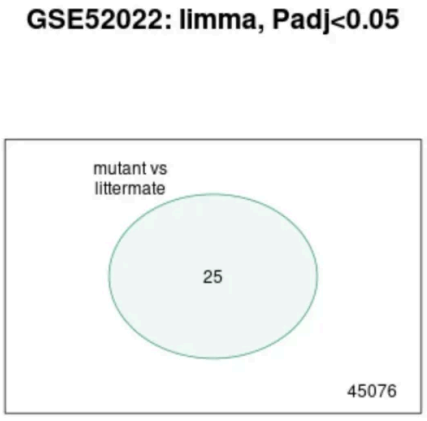


Figure 2: Differentially Expressed Genes: The volcano plot shows us how genes are expressed differently between the two groups. while the Venn diagram shows if any genes overlap or if they are different between the two groups.

Identification of (30 or 40 or 50) Statistically Significant Differentially Expressed Genes (DEGs)

To narrow down the genes, genes with a P-value greater than 0.5 were excluded, and fold change values were utilized to identify upregulated and downregulated genes. Based on the statistical methods used, P-value and fold change, there were 30 top Differentially Expressed Genes (DEGs) from the total of 45102. Out of the 30 DEGs, there were 15 upregulated genes and 15 downregulated genes. Gene Dataset- Alzheimers

Potential Functions and Enrichment of the Identified Genes and/or pathways

To determine the potential functions of the genes, the SR Plot enrichment tool was utilized. From the KEGG results, I received an overall pathway analysis and I identified a tuberculosis pathway.

The pathway analysis KEGG result indicated that Tuberculosis (TB) was the most prominent In Alzheimer's disease (AD). The bacterium that causes TB can result in inflammation in areas of the nervous system as well as prolonged immune responses, which contribute to the progression of AD. Through the Tuberculosis path, significant genes including Dectin-1, CALM, and CathepsinS were identified.

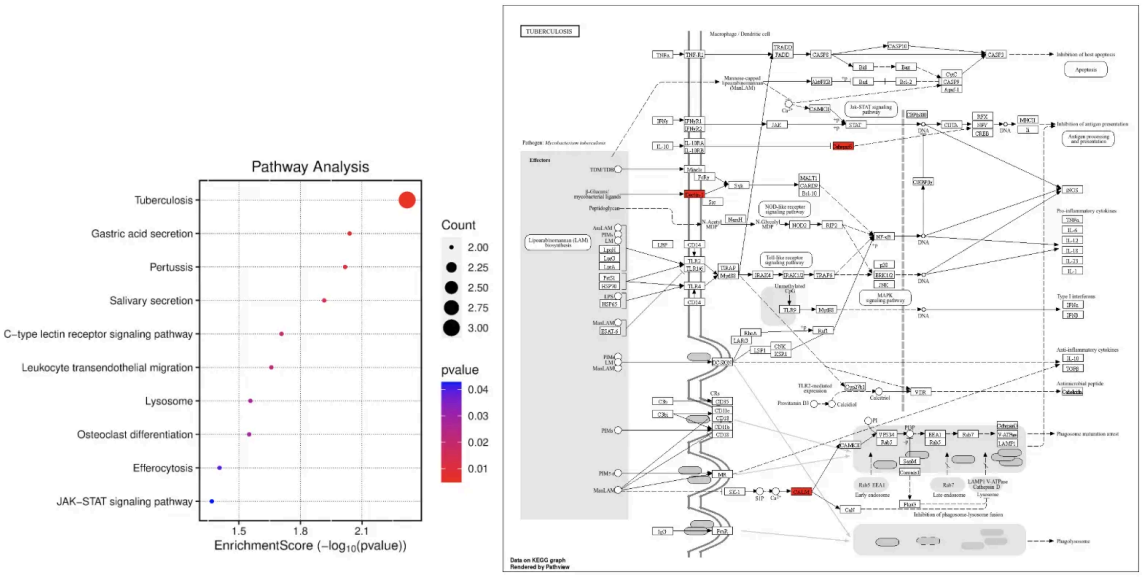


Figure 3: Functional Analysis from KEGG: The KEGG Pathway analysis results indicate the biological pathways that are significantly enriched within a set of genes. In this case, Tuberculosis is extremely prominent. In the Tuberculosis Pathway, the significant genes within the pathway are highlighted in red. These genes include Dectin-1, CALM, and CathepsinS.

From the GO results of the Three Ontologies, the bar graph shows three ontologies and how enriched they are. The blue represents Molecular Functions (MF), the green represents Cellular Components (CC), and the orange represents Biological processes (BP). In this display, BP is slightly more enriched than the other two ontologies. The most significantly enriched BP was ‘import into cell.’ The cnet plot provided is used to visualize the relationship between genes and their associated pathways.

This cnet plot displays the most significant genes in red, which represents a high-fold change. From this graph two main genes are highlighted: TREM2 and TYROBP.

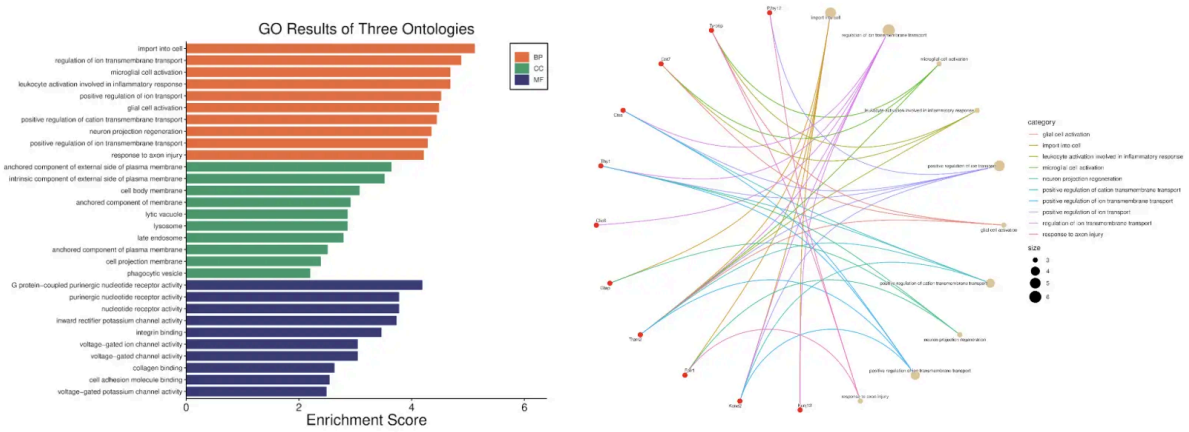


Figure 4: BP, CC, and MF are the three ontologies that are included in this graph. The graph reveals that the top 30 DEGS are most enriched in BP and MF. Overall, the enrichment scores were greater than or around 3, (middle of the x-axis). The Cnet plot is a display of key Biological Process genes and their relation to specific functions and activities. The red dots represent significant, upregulated genes.

DISCUSSION

Summary of Findings

The main goal of this research was to identify significant genes that could be used as potential targets for a cure or effective treatment. Results from the analyses conducted in GEO2R and SR plots helped to define key genes that showed high enrichment and were involved in prevalent pathways.

Interpretation of Results

The key genes that were identified in the graphs could potentially be used as targets for AD treatments. TREM2, TYROBP, and Dectin-1, also known as CLEC7A (Figures 5 and 7) play key roles in AD.

Comparison with Previous Studies

TREM2 was a significant gene that appeared in the BP cnet plot (Figure 7). This gene plays a crucial role in a variety of things, including positive regulation of ion transmembrane transport, positive regulation of cation transmembrane transport, glial cell activation, positive regulation of ion transport, and much more. Enhancing TREM2 signaling could result in a cure for AD. One distinguishing characteristic of AD is a lack of TREM2 signaling or a decreased efficiency because it impairs microglial function that is meant

to clear amyloid-beta plaques (10). Moreover, if therapies that enhanced TREM2 signaling were engineered, researchers could potentially improve microglial phagocytosis of amyloid-beta which would reverse the progression of this disease.

TYROBP was another key gene that occurred in the BP cnet plot (Figure 7). It interacts with TREM2 and plays a crucial role in glial cell activation. Targeting its role in immune response regulation could offer a potential cure for AD. TYROBP is heavily involved in enhancing the phagocytic activity of microglia which helps to clear amyloid-beta peptides and apoptotic neurons, which contribute to the progression of AD. Moreover, when there is neuroinflammation, TYROBP plays a role by restraining it by reducing the production of microglia-mediated cytokine (11). Because these actions can protect against neuroinflammation and plaque building, factors that contribute to AD, enhancing TYROBP's role could help reverse disease progression.

Dectin-1 (CLEC7A) is a receptor that is activated around amyloid plaques. Normally, microalgia like CLEC7A is helpful to clear waste and protect the brain. In this case, it is too activated, which worsens AD progression. Researchers found that CLEC7A microalgia in model mice contributed to neuroinflammation and blood vessel damage. When these microalgia were blocked, blood vessels were shielded and inflammation decreased significantly (12). This evidence suggests that blocking or controlling CLEC7A when it is too activated could help reduce brain inflammation, which would help treat AD. For instance, treatments that limit CLEC7A activity could be a potential solution to AD.

Table 1. Summary of Key Genes Identified and their Connection to Alzheimer's Disease

Gene	Function	Gene Name	Connection
TREM2	-Contributes to microagial activation-Regulates inflammation-Neuroprotection	Triggering Receptor Expressed on Myeloid Cells 2	Clears amyloid-beta plaques that play a part in AD

TYROBP	-Helps with glial cell activation-Enhances the phagocytic activity of microalgia-Exhibits inflammatory response regulation	TYRO3, AXL, and MER tyrosine kinase binding protein	Restrains neuroinflammation by reducing the production of microglia-mediated cytokine
CLEC7A	-Helps microaglia become overactive and harmful-Leads to Inflammation-Can cause Blood Vessel Damage	C-type lectin domain family 7 member A	If regulated, the microglial ability to clear amyloid beta plaques and reduce neuroinflammation will be enhanced.

Implications

Regulating TREM2 and TYROBP could enhance microglial activity to clear amyloid-beta plaques, while controlling CLEC7A may reduce harmful inflammation and amyloid beta plaques. These findings could lead to targeted drugs, gene therapies, or early interventions to slow or prevent Alzheimer's disease.

Limitations

In our case, since we use bioinformatics datasets from microarray experiments conducted by other researchers, one limitation is that the identified genes will need to be further studied in the laboratory or clinical environment before developing a cure.

Future Directions

The identified genes can be tested in the laboratory by scientists who can conduct clinical trials or other studies to determine if any treatments or a cure could be determined using the pathways we have identified.

References

1. Lane, C.A., Hardy, J. and Schott, J.M. (2018), Alzheimer's disease. Eur J Neurol, 25: 59-70. https://doi.org/10.1111/ene.13439
2. Selkoe DJ. Defining Molecular Targets to Prevent Alzheimer Disease. Arch Neurol. 2005;62(2):192–195. doi:10.1001/archneur.62.2.192

3. Blennow K, de Leon MJ, Zetterberg H. Alzheimer's disease. Lancet. 2006;368(9533):387-403. doi:10.1016/S0140-6736(06)69113-7.
4. Khanna S, Domingo-Fernández D, Iyappan A, Emon MA, Hofmann-Apitius M, Fröhlich H. Using multi-scale genetic, neuroimaging and clinical data for predicting Alzheimer's disease and reconstruction of relevant biological mechanisms. Sci Rep. 2018;8:11173. doi:10.1038/s41598-018-29433-3.
5. Tanya Barrett, Stephen E. Wilhite, Pierre Ledoux, Carlos Evangelista, Irene F. Kim, Maxim Tomashevsky, Kimberly A. Marshall, Katherine H. Phillippy, Patti M. Sherman, Michelle Holko, Andrey Yefanov, Hyeseung Lee, Naigong Zhang, Cynthia L. Robertson, Nadezhda Serova, Sean Davis, Alexandra Soboleva, NCBI GEO: archive for functional genomics data sets—update, Nucleic Acids Research, Volume 41, Issue D1, 1 January 2013, Pages D991–D995, https://doi.org/10.1093/nar/gks1193
6. Reiman EM, Jagust WJ. Brain imaging in the study of Alzheimer's disease. Neuroimage. 2012;61(2):505-516. doi:10.1016/j.neuroimage.2011.11.075.
7. Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., Wu, S., & Wang, Y. (2023). SRplot: A free online platform for data visualization and graphing. PLOS ONE, 18(11), e0294236. https://doi.org/10.1371/journal.pone.0294236
8. Kanehisa, M. The KEGG Database. 91-103. https://doi.org/10.1002/0470857897.ch8
9. Louis du Plessis, Nives Škunca, Christophe Dessimoz, The what, where, how and why of gene ontology—a primer for bioinformaticians, Briefings in Bioinformatics, Volume 12, Issue 6, November 2011, Pages 723–735, https://doi.org/10.1093/bib/bbr002
10. Ulrich JD, et al. Elucidating the role of TREM2 in Alzheimer's disease. Neuron. 2017;94(2):237-248.
11. Ma, J., Jiang, T., Tan, L. et al. TYROBP in Alzheimer's Disease. Mol Neurobiol 51, 820–826 (2015). https://doi.org/10.1007/s12035-014-8811-9
12. Shi, H., Yin, Z., Koronyo, Y. et al. Regulating microglial miR-155 transcriptional phenotype alleviates Alzheimer's-induced retinal vasculopathy by limiting Clec7a/Galectin-3+ neurodegenerative microglia. acta neuropathol commun 10, 136 (2022). https://doi.org/10.1186/s40478-022-01439-z.

Gabriel Muraszczuk

Bioinformatics – Driven Analysis of Vortioxetine's Impact on Glioblastoma: Identifying Key Differentially Expressed Genes for Treatment Strategies

ABSTRACT

Background: Glioblastoma multiforme (GBM) is the most aggressive primary brain tumor in adults, characterized by rapid growth, invasive behavior, and poor survival outcomes. Current therapies, including surgery, radiation, and chemotherapy, provide limited benefit due to treatment resistance and the blood-brain barrier. Drug repurposing offers a promising strategy to overcome these challenges. This study investigates the effects of Vortioxetine, an antidepressant capable of crossing the blood-brain barrier, on gene expression in glioblastoma LN229 cells to identify potential biomarkers and therapeutic vulnerabilities.

Methods:RNA sequencing data from the GEO dataset GSE214968 was analyzed using GEO2R to identify differentially expressed genes (DEGs) between untreated control (0-hour) and Vortioxetine-treated (24-hour) groups. A total of 9485 DEGs were identified and visualized through a volcano plot and Venn diagram. The top 50 DEGs (25 upregulated and 25 downregulated) were further analyzed using SRplot for Gene Ontology (GO) and KEGG pathway enrichment.

Results: RNA5S family genes showed significant downregulation, suggesting impaired ribosome biogenesis and protein synthesis in glioblastoma cells, potentially indicating therapeutic vulnerability. In contrast, genes such as NRARP, DLL3, COL6A3, and FREM2 were linked to pathways that support tumor progression, invasion, and resistance mechanisms. Pathways of interest included Ribosome biogenesis in eukaryotes, Ribosome, NOTCH signaling, and ECM-receptor interaction.

Conclusion: This study highlights the potential of Vortioxetine to disrupt ribosome biogenesis, presenting a novel therapeutic strategy for glioblastoma. Differentially expressed genes and enriched pathways identified in this analysis could serve as biomarkers or targets for future research to improve treatment outcomes in GBM.

Keywords: Glioblastoma Multiforme, Differentially Expressed Genes, Bioinformatics, Gene Ontology, KEGG Pathways, Ribosome Biogenesis, NOTCH Signaling, ECM-Receptor Interaction

INTRODUCTION

Glioblastoma Multiforme (GBM) is an aggressive brain tumor which originates from astrocytic glial cells. It grows rapidly, is highly invasive, and is the most common primary malignant central nervous system tumor in adults, accounting for 48.6 % of all such tumors. GBM is more prevalent amongst men and presents with symptoms such as headaches, seizures, and neurological decline. Treatment typically involves excessive surgery, radiation therapy, and chemotherapy; however outcomes remain poor, with a median survival rate of approximately 15 months (1).

The problem or challenge being addressed in this research is that GBM remains an incurable tumor. Achieving clear surgical margins is challenging due to the risks of damaging healthy brain tissues responsible for critical functions. GBM often exhibits resistance to radiation therapy, and excessive doses can lead to severe side effects, such as radiation necrosis and cognitive decline. Chemotherapy is also limited in effectiveness due to the Blood-Brain Barrier, which restricts many drugs from reaching the tumor. These challenges underscore the need for innovative treatment strategies, such as drug repurposing, to improve GBM outcomes (2).

This research investigates the potential application of Vortioxetine as a repurposed drug for the treatment of GBM. The repurposing of existing medication such as Vortioxetine, an antidepressant, is of primary interest. The question aims to understand the gene expression dynamics in response to Vortioxetine treatment and identify highly or lowly expressed genes that may provide insights into potential biomarkers for personalized targeted therapies for the patients (3).

Glioblastoma is a highly aggressive and deadly brain tumor originating from astrocytes, a type of glial cell that supports nerve cells. Classified as a grade IV glioma by the World Health Organization, GBM is the most malignant form of brain cancer and accounts for about half of all malignant brain

tumors. The annual incidence rate is approximately 3.2 per 100,000 people worldwide and only 5.5% of patients survive beyond 5 years post-diagnosis. Its high morbidity and mortality rates make GBM a critical area of research (2).

What is known so far is that GBM is highly invasive, fast growing, and composed of heterogeneous cell populations, making it resistant to many treatment modalities. The Blood-Brain Barrier further complicates treatment by limiting drug delivery. Current treatment multidisciplinary team approaches, such as surgical resection, followed by radiotherapy and chemotherapy, have shown limited success. Researchers are exploring novel treatment approaches for GBM, including immunotherapy, oncolytic viral therapy, and drug repurposing. Repurposed drugs, such as vortioxetine, are of particular interest due to established safety profiles, cost-effectiveness, and potential to contribute to the development of personalized gene targeted therapies (2).

So far, biomarker research in Glioblastoma has focused on identifying genetic and molecular targets to guide treatment and predict patient outcomes. Biomarkers such as EGFR amplification, MGMT promoter methylation, and IDH1 mutations have been extensively studied. These markers are crucial for understanding glioblastoma's heterogeneity and tailoring treatments to individual patients. However, the discovery of additional biomarkers remains essential for advancing personalized therapy. Bioinformatics tools, such as GEO2R, have proven instrumental in identifying novel biomarkers, as they allow researchers to analyze large data sets and correlate genetic expression with survival outcomes (4).

Glioblastoma's heterogeneity presents significant challenges for bioinformatics research. The tumor comprises a mix of cell types, each with unique genetic profiles, making it difficult to draw consistent conclusions about treatment responses. Additionally, the blood-brain barrier limits the ability of drug responsive genetic data. Bioinformatics studies also face challenges in data standardization and reproducibility, as data sets often come from diverse experimental conditions. Despite these hurdles, bioinformatics tools are invaluable for undercovering hidden patterns in gene expression and identifying potential therapeutic targets, which are otherwise hard to detect through traditional methods (1).

The primary goal of this research was to utilize bioinformatics tools to analyze the impact of vortioxetine on glioblastoma LN229 cells. The objective was to identify critical differentially expressed genes (DEGs) and pathways, with the ultimate aim of discovering novel therapeutic strategies and potential biomarkers for targeted therapies in glioblastoma patients (2).

The hypothesis of this research was that the treatment with antidepressant drug vortioxetine would result in significant difference in gene expression compared to untreated control samples. A specific prediction was made that some genes in the treated group would be upregulated (shoving higher gene expression) and some would be downregulated (lower expression).

This research is based on data available utilizing NCBI's GEO2R bioinformatic tools that supports data about vortioxetine leading to significant changes in the expression of specific genes in glioblastoma LN229 cells. These differentially expressed genes may play a critical role in tumor progression or suppression and serve as potential biomarkers for personalized therapies (3). NCBI GEO 2R is a Bioinformatic tool that is provided by the National Center for Biotechnology Information (NCBI) within the Gene Expression Omnibus (GEO) database. It was used to perform differential gene expression analysis by comparing two experimental conditions (control vs.treatment). It allowed us to determine top differentially expressed genes and visualise them with graphs such as Volcano Plot, Venn Diagram and others. Log2 fold change and p-values were selected for data manipulation to create a narrow-downed list, from 9485 differentially expressed genes, to the most important 50 up and down regulated genes for further analysis (6). In addition, SR plot, KEGG and GO bioinformatics was used to further analyze the results obtained from GEO2R. SR Plot is the bioinformatic tool that is used for functional analysis and is able to generate visual and graphical outputs of most important selected genes based on their symbols (7). KEGG is a comprehensive database resource that helps to identify enriched biological pathways in gene lists for mapping and functional annotation (8). GO is used to describe the biological processes, molecular functions, and cellular components in which gene products are involved. KEGG and GO work in synergy (9).

This research is important because Glioblastoma is highly resistant to conventional therapies, in part due to its genetic heterogeneity and the rapid evolution of treatment resistance mechanisms (5). This research is important because it utilizes bioinformatics to identify differentially expressed genes (DEG's) that could serve as biomarkers for a glioblastoma prognosis and treatment response. By focusing on vorteoxetine's impact on gene expression in glioblastoma cells, the study may uncover novel therapeutic targets that bypass existing resistance pathways. This knowledge has the potential to contribute to the design of personalized and more effective therapeutic approaches for glioblastoma, improving patient outcomes and survival rates.

METHODS

Data Collection Using Bioinformatics Tools

This research was conducted using the Gene Expression Omnibus (GEO) bioinformatics tool provided by the National Center for Biotechnology Information (NCBI). NCBI GEO2R is the powerful bioinformatic data tool that is a user-friendly data analysis platform which enables researchers to perform different gene expression analysis of datasets stored in GEO (6). The dataset GSE214968, titled Effect of Vortioxetine treatment on gene expression of LN-229 glioblastoma cells over time, was used to identify differentially expressed genes (DEGs) in glioblastoma following treatment with Voltioxetine over a 24-hour period. The GSE214968 dataset includes RNA sequencing data from LN229 glioblastoma cells (a commonly used cell line in cancer research) collected at baseline and across six time points after drug treatment (3).

To prepare the data for this research, two groups with samples were defined: Control group – baseline samples collected at 0 hours with no treatment and Treatment group – samples collected 24 hours after treatment with Vortioxetine. The no-code GEO2R bioinformatics tool was then used to process these groups and generate initial results, including the identification of differentially expressed genes (DEGs). The programming language R was incorporated for data analysis and visualisation R script.

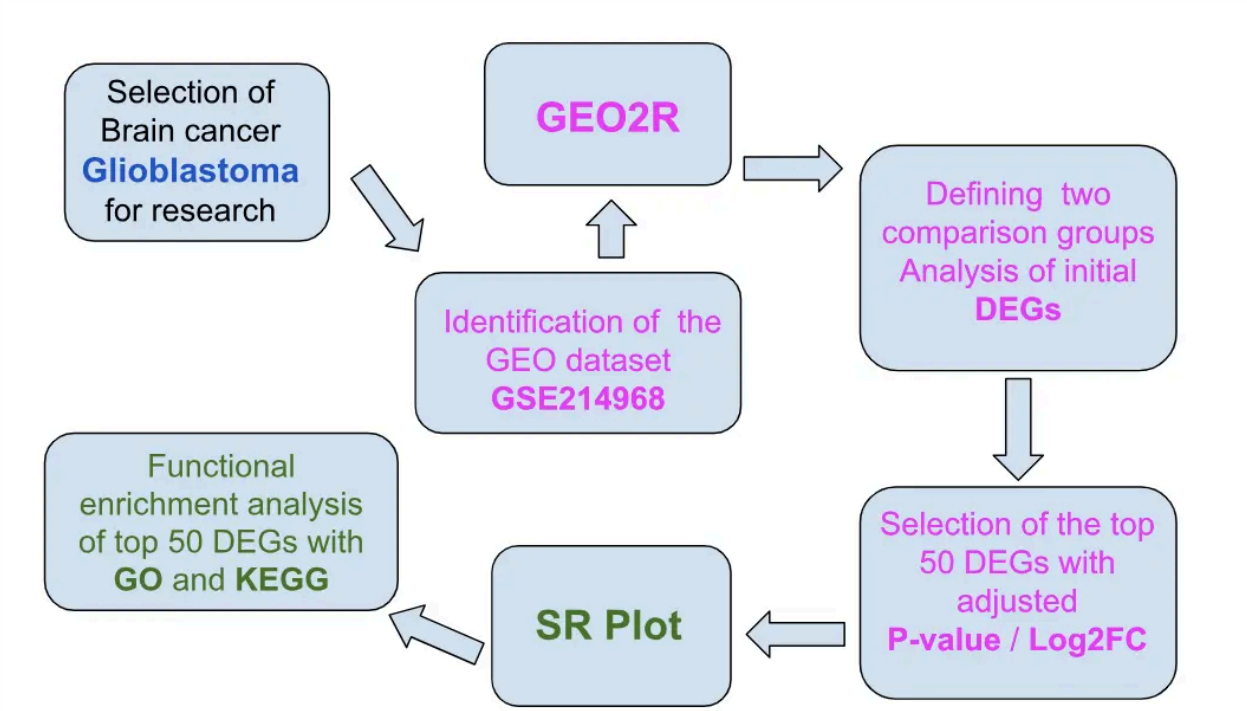


Figure 1. Research Methodology: The steps and bioinformatics tools used in this study.

Data Analysis of GEO2R Data

The data was analyzed and visualization graphs and outputs were generated. Two visual results, the Volcano Plot and Venn Diagram, were chosen. The Volcano Plot was analyzed to identify significantly differentially expressed genes, while the Venn Diagram was used to examine the number of gene expressions in the groups. Volcano plot analysis focused on identifying the highly expressed (Red) genes and low expression (Blue) genes. Venn Diagram analysis focused on visually highlighting a subset number of DEGs identified between the) 0-hour control and the 24-hour vortioxetine-treated glioblastoma samples.

Identification of the Top Differentially Expressed Genes

To identify the most significant differentially expressed genes (DEGs), further statistical analysis was performed. To support this selection process, a Venn diagram was utilized to visualize the distinctions in the data. The total list of 18326 genes generated with the GEO2R database was uploaded into Google Sheets. From this dataset a total of 9485 DEGS were identified based on the applied thresholds. A cut-off adjusted p-value (p.adj) of less than 0.05 and a

log2 fold change (log2FC) threshold of +/-3 were used to filter the results. These criteria ensured the selection of the most statistically significant genes. Using these parameters, the study focused on a narrowed list of 50 DEGs, which included: 25 Upregulated Genes (log2FC > +3) and 25 Downregulated Genes (log2FC < -3)

Further Data Analysis Using SRPlot, KEGG, and GO Bioinformatics Tools

SRPlot, KEGG, and GO bioinformatics tool databases were utilized to analyze the functions of these top 50 differentially expressed genes (DEGs). These tools allowed for functional enrichment analysis and provided insights into the biological processes, cellular components, molecular functions and biological functions of the selected DEGs, with the ultimate goal of uncovering their potential roles in glioblastoma. The gene symbols and corresponding log2 fold changes (log2 FC) were input into SR Plot online server, which generated a variety of graphical representations. These included results from Gene Ontology (GO) and KEGG pathways analysis, which summarized the functions of DEGs and mapped them to specific biological pathways relevant to glioblastoma. Three Ontologies, Pathway Enrichment Bar, and C Net Plot graphs were selected from Go and KEGG to demonstrate biological significance and the interconnected networks relevant to top 50 DEGs in glioblastoma cells. 14 genes that showed high enrichment and important functions in glioblastoma were chosen for further study. Ribosomal genes were identified as significantly downregulated in the CNET plot for Ribosome Biogenesis in Eukaryotes and Ribosome pathway.

RESULTS

Identification of Differentially Expressed Genes

The first step in identifying differentially expressed genes (DEGs) Involved the use of the GEO2R bioinformatics tool, provided by the National Center of Biotechnology Information (NCBI). This tool facilitated the comparison of two defined groups within the dataset: a control group (0-hour baseline samples) and a treatment group (24-hours post-treatment with Vortioxetine). GEO2R generated a comprehensive list of all genes and DEGs based on log2FC threshold set at 0 and adjusted p-value (p.adj) less than 0.05. From the initial analysis using GEO2R of 18326 genes, a total amount of 9485 were identified as differentially expressed between the control and treatment groups, reflecting the impact of Vortioxetine treatment on gene expression in glioblastoma LN229 cells. These DEGs were visualised using two key figures:

a volcano plot and a Venn diagram. The volcano plot (Figure 2a) provides a visual representation of the DEGs identified. Red dots represent genes that are highly expressed genes (significantly upregulated genes – top right region of the plot – positive fold change, significant P- value). Blue dots represent lowly expressed genes (significantly downregulated genes – top left region of the plot – negative fold change significant, P- value). Black dots represent genes with no difference in expression (non-significant genes -close to 0 threshold (X and Y) of fold change and p-value). The Venn Diagram illustrates a total 18326 number of genes that were identified in glioblastoma LN229 cells after 24 hours of treatment with Vortioxetine, highlighting 9485 genes as differentially expressed genes (DEGs).

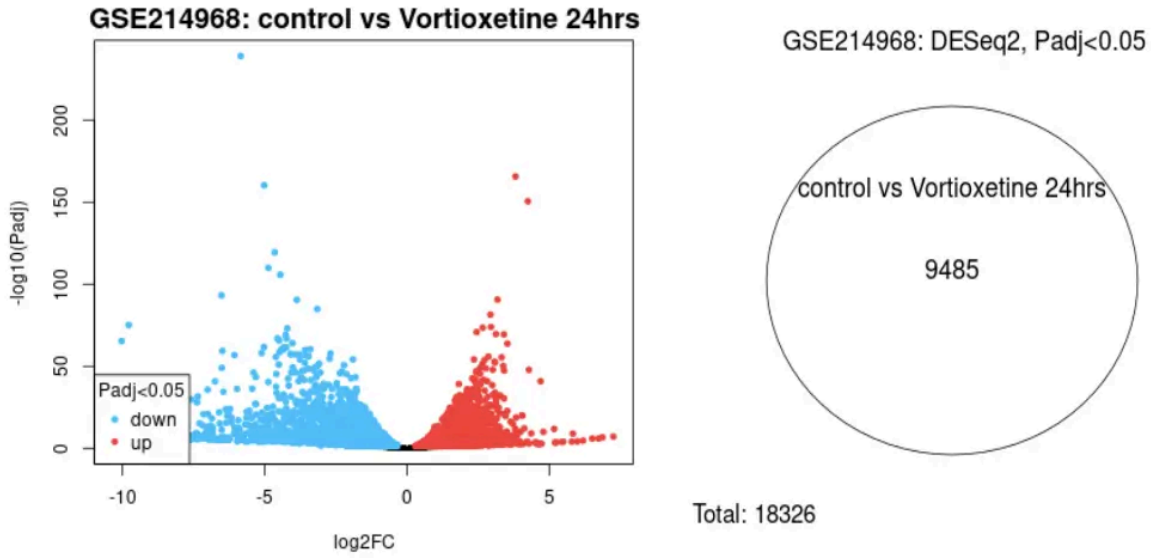


Figure 2. Differentially Expressed Genes in GEO2R: (a) Volcano Plot, (b) Venn Diagram visualize gene expression changes in glioblastoma LN229 cells after 24 hours treatment with Vortioxetine, using the control group (0-hour, untreated) as the baseline. (a) Genes with significant upregulation (higher expression) are highlighted in red, and those with significant downregulation (lower expression)are in blue. Black dots represent genes with minimal expression change. (b) illustrates a total 18326 genes number that were identified in glioblastoma LN229 cells after 24 hours of treatment with Vortioxetine, highlighting 9485 genes as differentially expressed genes (DEGs).

Identification of 50 Statistically Significant Differentially Expressed Genes (DEGs)

To refine the initial list of differentially expressed genes (DEGs), statistical thresholds were applied to identify the most significant genes from the GEO2R generated list of 9485 DEGs for future analysis. The selection criteria included: Log2 fold change (lig2FC) threshold:+/-3, Adjusted p-value (p.adj) at < 0.05. Using these criteria, a total of 50 statistically significant DRGs were identified, consisting of 25 upregulated genes (log2FC >+3) and 25 downregulated genes (logeFC <-3). This subset was selected based purely on statistical metrics and is summarized in a table, which included the gene symbol, name, log2FC and adjusted p-values. full 50 data for further analysis log2fold +-

Potential Functions and Enrichment of the Identified Genes

To explore the biological roles and pathways associations of the 50 statistically significant differentially expressed genes (DEGs), functional enrichment analysis was performed using SRPlot, including GEne Ontology (GO) and KEGG pathway analysis. These analyses provided insights into the molecular, cellular, and biological mechanisms impacted by the identified DEGs, with particular emphasis on their relevance to glioblastoma biology. Gene Ontology (GO) enrichment analysis bar graph categorized the DEGs into three ontologies (Figure 3a). Higher enrichment score indicates stronger statistical significance and expressed in -log10(p-value). Biological Process (BP) highlighted processes such as somitogenesis, Notch signaling pathway, chemokine-mediated signaling pathway with enrichment score ranging from about 3.5 highest to 2 being the lowest. Cellular Components (CC) emphasized structural terms like Ribosome with significantly high enrichment score at about 12.

Finally, Molecular Function presented the most highly enriched score. Structural constituent of ribosome was about 13 in p-value, indicating that genes in this group are significantly present. The CCNet plot (Figure 3b) was analyzed to narrow down on some DEGs and the pathways they follow. The collagen containing extracellular matrix stood out as a significant pathway and connection with COL6A3 and FREM2 genes was noted.These genes are significantly upregulated and are known to to be a critical factors in tumor invasion and aggressiveness.

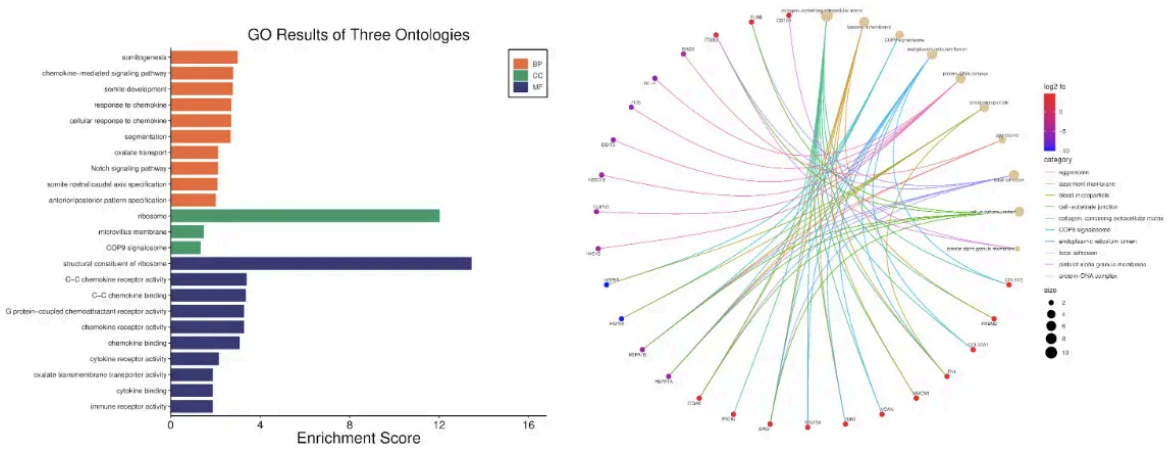


Figure 3. GO Pathway enrichment and DEGs in SR Plot: (a) Enrichment Three Ontologies Bar graph, (b) Cellular Component (CC) CNet plot provide insight into biological, cellular, and molecular roles and determine how likely it is that the observed association with specific DEGs occurred by chance. COL6A and FREM2 are significantly upregulated and show strong connection to the collagen-containing extracellular matrix pathway.

KEGG pathway analysis highlighted four key biological pathways enriched with DEGs. The Pathway Enrichment Bar graph (Figure 4a) shows Ribosome biogenesis in eukaryotes with the highest enrichment score 12.5, calculated -log10 p-value, closely followed by Ribosome pathway. In contrast, Notch signaling pathway and ECM-receptor interaction exhibit a significant drop to about 2-2.5 enrichment score level. Visualisation of CNet plot in KEGG pathway (Figure 4b) enables studying of these specific pathways and their association with individual statistically significant DEGs and highlighting their relevance to glioblastoma biology. Highly expressed COL6A3 and FREM2 in ECM-receptor interaction together with lowly expressed NRARP and DLL3 genes in Notch signaling pathway were identified from the GO analysis. New 10 DEGs in RNA5s family genes were depicted and identified as significantly downregulated in Ribosome and Ribosome biogenesis in eukaryotes pathways.

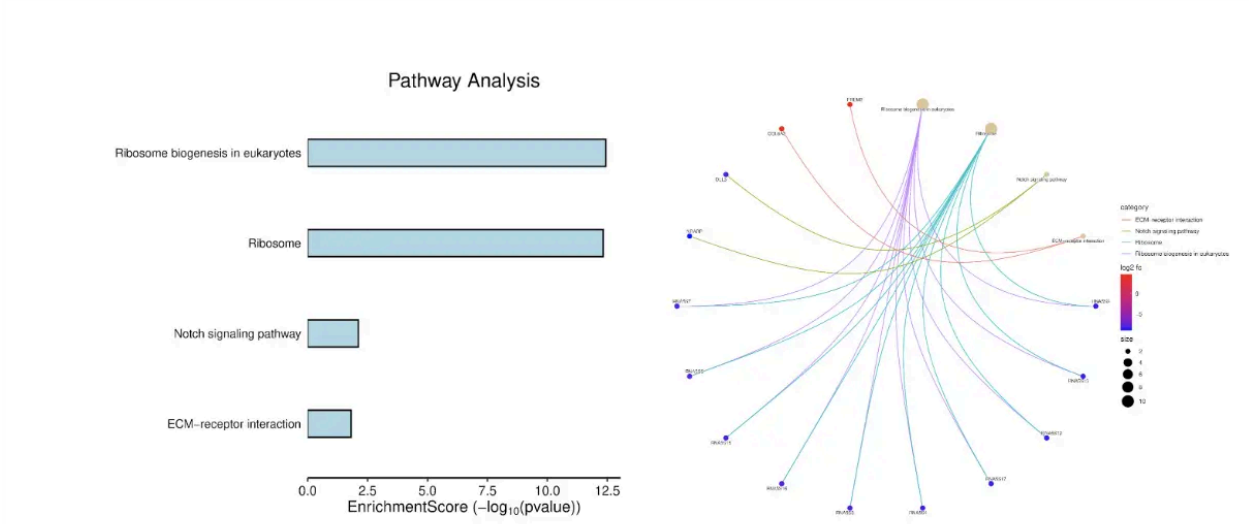


Figure 4: KEGG pathway enrichment and DEGs in SR Plot: (a) Pathway Enrichment Bar Graph, (b) CNet Plot provide comprehensive analysis of the biological pathways associated with the differentially expressed genes (DEGs). Among four significantly enriched pathways, the Ribosime and Ribosome biogenesis in eukaryotes pathways display the highest enrichment score. Additionally, 10 RNA5S family genes are prominently associated with these pathways.

The table below provides a comprehensive summary of the key differentially expressed genes (DEGs) identified in this study, their associated pathways, biological functions, and their relevance to glioblastoma. This table integrates findings of GO and KEGG enrichment analyses, highlighting both tumor-promoting and tumor-disrupting genes. By detailing the biological roles of these DEGs within their respective pathways, this table emphasizes their potential biomarkers or therapeutic targets in glioblastoma research.

Table 1. DEGs and Pathways identified in these Studies

Gene Code	Gene Name	Pathway	Functions	Glioblastoma Connection
RNA 5S (…7, 5, 15, 16, 3, 4, 17, 12, 13, 6)	Ribosomal RNA 5S	Ribosome biogenesis in eukaryotes	Essential for ribosome assembly and function; supports	Low expression disrupts ribosome function, affecting

			protein synthesis.	protein synthesis (4).
RNA 5S (…7, 5, 15, 16, 3, 4, 17, 12, 13, 6)	Ribosomal RNA 5S	Ribosome	Integral to ribosome structure and efficient protein translation.	Reduced activity may impair protein translation in tumor cells (7).
NRARP	Notch Regulated Ankyrin Repeat Protein	NOTCH signaling pathway	Regulates NOTCH signaling; involved in cell differentiation.	Altered signaling influences tumor progression and differentiation (10).
DLL3	Delta like Protein 3	NOTCH signaling pathway	Ligand for NOTCH signaling; governs cell fate decisions.	Abnormal expression impacts glioblastoma cell proliferation (10).
COL6A3	Collagen VI alpha-3	ECM-receptor interaction	Provides structural support in ECM; promotes cell adhesion.	Overexpression facilitates tumor invasion and metastasis (12).
FREM2	FRAS-related extracellular matrix protein 2	ECM-receptor interaction	Mediates cell-ECM signaling; critical for tissue integrity.	Upregulation enhances glioblastoma cell migration and adhesion (12).

DISCUSSION

Summary of Findings

This study aimed to investigate the impact of vortioxetine on gene expression in glioblastoma LN229 cells using bioinformatics tools to identify differentially expressed genes (DEGs) and their associated pathways. Using GEO2R, a total of 1836 DEGs were identified, including 25 upregulated and 25 downregulated genes. KEGG pathway analysis highlighted four key pathways: Ribosome biogenesis in eukaryotes, Ribosome, Notch signaling pathway, and ECM-receptor interaction (8,12). Downregulation of the RNA5S family genes suggested disruption of ribosome biogenesis, potentially impairing protein synthesis in glioblastoma cells. Conversely, COL6A3 and FREM2 were highly upregulated, linking them to extracellular matrix remodeling and tumor progression (5,6). Go enrichment analysis further categorized these DEGs into biological processes, cellular components, and molecular functions, reinforcing their critical roles in glioblastoma biology.

Interpretation of Results

The findings underscore the dual nature of the DEGs in glioblastoma. The RNA5S family genes were strongly associated with the ribosome biogenesis pathways, and their significant downregulation suggests that Vortioxetine may disrupt ribosomal function and protein synthesis, presenting a potential therapeutic vulnerability (4, 7). On the other hand, the upregulation of COL6A3 and FREM2 in the ECM-receptor interaction pathway supports their role in extracellular matrix remodeling, a key process in glioblastoma cell adhesion, migration, and invasion (5, 6, 15). Furthermore, the identification of NRARP and DLL3 within the Notch signaling pathway highlights the importance of this pathway in tumor progression and glioblastoma stem-like cell maintenance (9, 13). Collectively, these findings align with the hypothesis that Vortioxetine affects key molecular pathways in glioblastoma, influencing both tumor-suppressive and tumor-promoting mechanisms.

Comparison with Previous Studies

The results are consistent with previous research emphasizing the importance of ribosome biogenesis in cancer. Upregulation of ribosomal activity has been shown to be critical for tumor growth, while its disruption can impair protein synthesis and reduce tumor viability (4, 7). Similarly, the involvement of COL6A3 and FREM2 in ECM remodeling aligns with studies demonstrating the role of the extracellular matrix in glioblastoma Invasion and metastasis (5, 6,15). The findings also corroborate earlier research on the role of the NOTCH signaling pathway in glioblastoma, particularly its

regulation of stem-like cells (ability to self-renew, ability to differentiate into multiple cell types, and rapid growth and division) behavior and therapy resistance (9, 13). Unique to this study, however, is the impact of Vortioxetine on these pathways, offering new insights into its potential for repurposing as a glioblastoma therapy.

Implications

These findings have significant implications for glioblastoma research and treatment. The down regulation of RNA5S genes suggests that targeting ribosome biogenesis may be a viable therapeutic strategy, potentially limiting the tumor's ability to sustain high protein synthesis demands. Meanwhile, the up-regulated genes COL6A3 and FREM2 present potential targets for therapies aimed at reducing ECM-driven tumor invasiveness (5,6). Importantly, the ability of Vortioxetine to cross the blood-brain barrier (BBB), as highlighted in advancements in neurotechnology and drug delivery (17), makes it an attractive candidate for glioblastoma treatment. By modulating glioblastoma-related pathways and bypassing the BBB, Vortioxetine could offer a dual advantage in addressing the challenges of tumor resistance and drug delivery.

Limitations

A limitation of this study lies in its reliance on publicly available bioinformatics datasets derived from microarray experiments conducted by other researchers. While these datasets offer valuable insights, the findings may not fully capture the complexity of glioblastoma in clinical settings due to the absence of direct experimentation on human or animal samples (10,11). Additionally, the use of a single glioblastoma cell LN229 limits the generalizability of the results to other glioblastoma subtypes or patient populations. The identified DEGs and pathways will require further validation in laboratory or clinical environments to confirm their roles and therapeutic potential (13).

Further Directions

Future research should focus on experimentally confirming the identified DEGs and pathways in laboratory or clinical environments. For instance, the downregulated RNA5S family genes could be investigated for their roles in ribosome biogenesis and its disruption in glioblastoma cells. Similarly, the upregulated genes COLA3 and FREM2 could be explored for their

contributions to ECM remodeling and tumor invasion, with a view to developing anti-invasive therapies (5,6,16). Given Vortioxetine's ability to cross the BBB, further studies should assess its therapeutic efficacy in glioblastoma and its potential synergistic effects when combined with other treatments (17). Expanding research to include diverse glioblastoma cell lines and patient-derived samples would provide a more comprehensive understanding of these pathways' roles in glioblastoma biology (9, 11). Future research could expand on these findings to develop targeted therapies and biomarkers for improved glioblastoma management and patient survival (15, 16).

References

1. Kanderi T, Munakomi S, Gupta V. Glioblastoma Multiforme. [Updated 2024 May 6]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2024 Jan-. Available from: https://www.ncbi.nlm.nih.gov/books/NBK558954/
2. Grochans S, Cybulska AM, Simińska D, Korbecki J, Kojder K, Chlubek D, Baranowska-Bosiacka I. Epidemiology of Glioblastoma Multiforme–Literature Review. Cancers. 2022;14(10):2412. https://doi.org/10.3390/cancers14102412
3. Lee S, Weiss T, Bühler M, et al. High-throughput identification of repurposable neuroactive drugs with potent anti-glioblastoma activity. Nat Med. 2024;30(12):3196-3208. https://doi.org/10.1038/s41591-024-03224-y
4. El Atat O, Naser R, Abdelkhalek M, Habib RA, El Sibai M. Molecular targeted therapy: A new avenue in glioblastoma treatment. Oncol Lett. 2022;25(2):46. https://doi.org/10.3892/ol.2022.13632
5. Menyhárt O, Fekete JT, Győrffy B. Gene expression-based biomarkers designating glioblastomas resistant to multiple treatment strategies. Carcinogenesis. 2021;42(6):804-813. https://doi.org/10.1093/carcin/bgab024
6. Clough E, Barrett T. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res. 2024;52(D1):D138-D144. https://doi.org/10.1093/nar/gkad965
7. Wang D. Using Bioinformatics to Find Differentially Expressed Genes in Patients Using SGLT2i. Rising Researchers by MoonPrep. September 2024. Available from:

https://risingresearchers.com/journal/exploring-genetic-mechanisms-of-type-2-diabetes-using-bioinformatics-tools-and-databases/
8. Kanehisa M, Furumichi M, Sato Y, Matsuura Y, Ishiguro-Watanabe M. KEGG: biological systems database as a model of the real world. Nucleic Acids Res. 2025;53(D1):D672-D677. https://doi.org/10.1093/nar/gkae909
9. Rodriguez A, Monleon Getino A. BioFunctional: A Comprehensive App for Interpreting and Visualizing Functional Analysis of KEGG Pathways and Gene Ontologies. bioRxiv. 2024. https://doi.org/10.1101/2024.10.08.616405
10. Tang D, Chen M, Huang X, et al. SRplot: A free online platform for data visualization and graphing. PLoS One. 2023;18(11):e0294236.https://doi.org/10.1371/journal.pone.0294236
11. Yin X, Wu Q, Hao Z, et al. Identification of novel prognostic targets in glioblastoma using bioinformatics analysis. Biomed Eng Online. 2022;21(26):1-13. https://doi.org/10.1186/s12938-022-00995-8
12. Bautista A, Romero R. Identification of Potential Therapeutic Targets and Biomarkers for Glioblastomas Through Integrative Analysis of Gene Expression. bioRxiv. 2024. https://doi.org/10.1101/2024.03.04.583392
13. Gene Ontology Consortium. The Gene Ontology Resource: 20 Years and Still Going Strong. Nucleic Acids Res. 2019;47(D1):D330-D338. doi:10.1093/nar/gky1055. Available at: https://academic.oup.com/nar/article/47/D1/D330/5144766 .
14. Kressler D, et al. Ribosome Biogenesis in Eukaryotes: More Than One Way to Make a Ribosome. BMC Biol. 2019;17:56. doi:10.1186/s12915-019-0699-6. Available at: https://bmcbiol.biomedcentral.com/articles/10.1186/s12915-019-0699-6.
15. Sánchez-Álvarez M, et al. Ribosomal Proteins in Glioblastoma: Role in Cancer Stem Cells. Oncogenesis. 2018;3(1):12. doi:10.1038/s41389-018-0044-8. Available at: https://www.nature.com/articles/s41389-018-0044-8.
16. Müller M, et al. Inhibition of Ribosome Biogenesis in Glioblastoma. BMC Biol. 2019;17:32. dBoronat S, et al. Targeted Drug Delivery to Glioblastoma: Challenges and Advances in Crossing the BKumar S, et al. Overcoming the Blood-Brain Barrier in Glioblastoma: Advances in

Nanotechnology and Drug Delivery. Brain Sci. 2020;10(11):858. doi:10.3390/brainsci10110858. Available at: https://doi.org/10.2147/IJN.S243223

17. Kumar S, et al. Overcoming the Blood-Brain Barrier in Glioblastoma: Advances in Nanotechnology and Drug Delivery. Brain Sci. 2020;10(11):858. doi:10.3390/brainsci10110858. Available at: https://doi.org/10.2147/IJN.S243223.

Siddharth Nanda

Impact of Socioeconomic Factors on Cardiovascular Health Disparity

ABSTRACT

Background: Cardiovascular disease (CVD) is one of the largest chronic conditions globally. The risk for CVD is closely associated with socioeconomic status, as lower SES communities are impacted by CVD at a higher rate. This scoping review aims to summarize the current literature around socioeconomic disparities in CVDs.

Methods: This review utilized PubMed and Google Scholar search engines to find relevant articles on the association between SES and CVD. Specific keywords regarding SES and CVD, such as socioeconomic, economic status, cardiovascular health, and cardiovascular disease, and inclusion criteria of original research published in the last 14 years, which helped identify the studies used in this review..

Results: Data from cross sectional analysis of respondents from Italy, Bosnia and Herzegovina, and the United states illustrated a clear correlation between CVD and SES. The data from the retrospective cohort studies also found a large gap in the CVD risk factors of lower SES populations in comparison to individuals with higher income and education.

Discussion: The data indicates that there is a large disparity between higher and lower SES populations. Further efforts should focus on developing initiative to decrease the disparity in cardiovascular health seen globally and reduce the impact of CVD on underprivileged communities.

INTRODUCTION

Background

Cardiovascular disease (CVD) is currently one of the most prevalent chronic conditions affecting millions. CVD, being a non-communicable chronic disease, refers to a multitude of conditions that impact the cardiovascular system such as heart disease, heart attack, heart failure, stroke, and arrhythmia. Many of these conditions develop due to atherosclerosis (ASCVD), which is a form of heart disease that is caused by plaque build up in the arterial walls [1]. CVD impacted an estimated 775 million individuals in 2022 and accounted for around 19.8 million deaths globally [2–3]. Disproportionate CVD mortality rates have been identified in less economically advantaged regions, with less economically stable regions of

Eastern Europe having rates of cardiovascular related mortalities of 432.3 per 100,000 compared to high-income parts of the Asia Pacific with rates of 73.6 deaths per 100,000 [2,4]. Lack of resources and poor environment makes management and treatment of CVD difficult for less economically stable populations. Deficiencies in the management of one's cardiovascular health due to lifestyle and environment, such as exercise, access to quality food choices and healthcare, transportation, and financial priorities can make CVD prevention and treatment far from ideal [5]. Five major modifiable CVD risk factors, including high body-mass index (BMI)/obesity, high low-density lipoprotein cholesterol (LDLC)/hyperlipidemia, blood pressure (BP)/hypertension (HT), diabetes, and smoking, have been identified to contributed to the risk of CVD and mortality [3–6]. These modifiable risk factors make CVD a preventable disease with proper treatment and lifestyle choices; however, without proper resources, the management of these modifiable risk factors become difficult. Economic barriers to medications, treatment, and therapy limit the ability for lower SES individuals to properly manage and identify the modifiable risk factors that contribute to CVD. This consequently results in an increased risk for CVD in less economically advantaged populations, making addressing SES disparity an essential part of global cardiovascular health.

Previous studies have established a connection between SES and CVD by relating different socioeconomic demographics to modifiable CVD risk factors [4,7]. Furthermore, regional cohort studies have especially found connections between education and CVD risk and premature mortality [8–9]. While there is a known correlation between SES and CVD, there remains a lack of research that quantifies the CVD gap between higher and lower SES populations.

Objective

The purpose of this review is to analyze the extent of the effect of SES disparity on CVD and CVD risk factors between higher and lower income and SES populations. Specifically by analyzing cross-sectional research and cohort data between different SES populations, this study will establish an understanding on the disparity of cardiovascular health based on socio-economic factors.

METHODS

Search Technique

For this review, PubMed and Google Scholar search engines were utilized to identify articles looking at the correlation between socioeconomic factors and cardiovascular health. For PubMed, the advanced search feature was utilized to identify the following key terms: socioeconomic, economic status, economic stability, cardiovascular health, and cardiovascular disease. The inclusion criteria for this scoping review was studies published within 14 years, clinical trials, meta-analyses, and randomized control trials. Exclusion criteria included systematic reviews and review articles. These criteria helped find original research on the topic of SES and CVD. Cross sectional studies based on surveys and retrospective cohort studies were analyzed for the purposes of this review.

Data Organization

The data retrieved was split between the cross sectional studies and retrospective cohort studies with the data from the studies being reorganized. Some statistics were recalculated due to percentages being split between genders within an identified population for a characteristic rather than the overall population being split between genders and percentages for characteristics within each gender were found. This system for identifying the composition of characteristics in the observed populations was implemented for all studies that have not previously kept their statistics in this format for comparative purposes.

RESULTS

Cross Sectional Studies

Data from a 2004-2005 National Survey conducted in Italy was analyzed as a cross sectional study to identify the correlation between SES and management of cardiovascular risk factors [9]. The observed population in Table 1 depicts the distribution of the respondents based on gender, age, education, occupational class, and risk factors [9]. The respondents were primarily composed of males and females aged 40-49, composing 40.4% and 38.9% of their respective gender groups [9]. Educational distribution varied between genders, with males having a 40.2% rate of having graduated high school or further while females had a more bimodal distribution between primary school and high school, composing 34% and 35.2% of their group respectively [9]. Even greater disparity was shown in education as males predominantly composed the lower occupational class

but had 29.4% of their group working in the high occupational class compared to only 10.8% of women working in that class [9]. Low unemployment rates of around 3.9% for males was much lower than the 28.6% of unemployed females [9]. For the CVD risk factors identified in the population, women tended to have higher rates of hypertension, while men had higher rates of diabetes, smoking, and higher BMI [9]. Around 49.7% of males in this population showed BMI levels around being pre-obese [9]. Furthermore, females tended to monitor their CVD risk factors more often than males but the majority of both genders monitored all their levels except for only 43.2% of males monitoring their weight at least once a year [9]. Overall the demographic composition identified in this study was mostly those with an above average education level and there was an even distribution between occupational classes [9]. CVD risk factor rates for hypertension, diabetes, and obesity were relatively low in the observed population [9].

Table 1. Demographics of respondents observed in Italy (Damiani et al., 2011)

Country	Characteristic	% Identified	
Italy (2004-2005) [9]	Age	Male	Female
	40-49	40.4	38.9
	50-59	34.3	34.1
	60-69	25.4	27.1
	Education		

≤ Primary School	25.3	34
Secondary School	34.6	30.8
≥ High School	40.2	35.2
Occupational class		
Unemployed	3.9	28.6
Low	37.8	30.2
Medium	28.8	30.4
High	29.4	10.8
Risk factors		
Hypertension	17.9	20.1
Diabetes	5.4	4.5
Former Smoker	32.9	16
Current Smoker	29.7	19
Pre-obese (25≤BMI<30)	49.7	30.2
Obese (BMI≥30)	13.5	11.4

Preventive behaviors (___ monitored at least once every (#) years)		
Cholesterol (5)	78.9	82.4
Glycemia (1)	56	59.5
Blood Pressure (1)	67.3	72.1
Weight (1)	43.2	59.9

This same study also found that higher educated individuals were more likely to monitor their CVD risk factors more often than lower educated populations being more defined in male populations compared to females [9]. This trend was also found between occupation classes as higher classes tended to manage their BP and weight at a higher rate than lower occupational classes [9]. For men, those who had an education of primary school and under had a 0.34 rate of regular monitoring of weight, 0.63 for BP, and 0.75 for cholesterol, while those with a high school education and above monitored those factors at a rate of 0.51, 0.70, and 0.83 respectively, which was also reflected in the female population [9]. This illustrates that those with a higher education tended to monitor their weight, blood pressure, and cholesterol at a higher rate [9].

Table 2 put together the demographic data from a Cross Sectional study of a 2010 national survey in the middle-income nation of Bosnia and Herzegovina [10]. The distribution of the observed population was composed mostly of males and females aged 45-64 with 57.9% of males having a middle education, while 54.5% of females had a lower education [10]. For the labor force of Bosnia and Herzegovina, 67.7% of men and 41.1% of women were either employed or looking for work [10]. The wealth index of the respondents were bimodally distributed with 44.2% and 35.7% of men and women respectively being in the upper class and 37% and 43.6% of men and women in the lower class [10]. For the risk factors seen in the population, out of all participants, 36.7% had HT, 13% had Diabetes with

sugar levels over 126 mg/dL, 49.2% were either current or former smokers, 32.8% were considered pre-obse, and 30.1% were obese [10].

Table 2. Demographics of respondents observed in Bosnia and Herzegovina (Janković et al., 2015)

Country	Characteristic	% Identified*	
Bosnia and Herzegovina (2010) [10]	Age	Male	Female
	25–44	34.1	31.1
	45–64	44	41
	≥65	21.9	28
	Education		
	Low	32	54.5
	Middle	57.9	38.8
	High	10.1	6.7
	Employment Status		

Out of Labor Force	32.3	58.8
In Labor Force	67.7	41.1
Wealth Index		
Lower	37	43.6
Middle	18.8	20.7
Upper	44.2	35.7
Risk Factors	All Participants	
Hypertension	36.7	
Diabetes	13	
Former Smoker	16.8	
Current Smoker	32.4	
Pre-obese (25 ≤ BMI < 30)	32.8	
Obese (BMI ≥ 30)	30.1	

*Data may be recalculated to represent percent per gender of the whole observed population to match other studies for comparative purposes.

Two separate studies observing survey data from American respondents analyzed individual data from respondents in the past decade, and Table 3 depicts SES and CVD data from respondents from the different surveys from 2005 – 2006 [7,11]. The respondents had a generally even distribution in education and a 19.6% unemployment rate [11]. Family income to poverty ratio had 40% of individuals right above the poverty line with a ratio of 100% – 299% with 22.8% in the 300%-499% group and 19.3% above 500% [11]. CVD risk factors observed in the 2006 respondents had 55.18% and 50.97% of males and females with HT [7]. Around 17.21% and 14.48% of males and females took diabetes medications and 41.1% and 33.89% of males and females were on cholesterol medications (antihyperlipidemics) [7]. The rate of smokers seen in men and women were 16.19% and 13.26% respectively, which are significantly lower than the numbers found internationally in Italy and Bosnia and Herzegovina [7,10].

Table 3. Demographics of respondents observed in the United States (Jenkins & Ofstedal, 2014) and (He et al., 2021)

Country	Characteristic	% Identified	
United States (2005-2006) [7,11]**	Age [7]	Male	Female
	50–59	38.92	37.11
	60–69	31.86	28.76
	70–79	19.38	20.75
	80–89	9.85	13.39

Education [11]	All Participants
≤ Highschool	27.8
Highschool	24.1
Some College	28.5
≥ College Graduate	19.6
Employment Status [11]	
Unemployed	19.6
Employed, Student, or Retired	80.4
Family Income to Poverty Ratio % [11]	
≤100%	16.9
100%-299%	41
300-499%	22.8
≥500%	19.3

Risk Factors [7]	Male	Female
Hypertension	55.2	51.0
Diabetes	16.2	13.3
Current Smoker	16.2	13.3
On Cholesterol Medication	41.1	33.9

**Combined data from two different surveys may not represent the same observed populations.

Various trends identified by the 1999-2018 study on US respondents found correlations between education and income with CVD risk factors [11]. Between education levels, those with a college education and above had lower mean BMI, BP, hemoglobin A1c (HbA1c), smoking rate, and 10 year risk of ASCVD compared to those at lower education levels [11]. Those with an education below a highschool diploma had the highest rates of 10 year ASCVD risk and HbA1c [11]. Trends between different levels of family income found that those with a family income to poverty ratio of ≥500% had lower BMI, HbA1c, smoking rate, and 10 year risk of ASCVD with the inverse being true for those under the poverty line [11].

Retrospective Cohort Studies

The data from the retrospective cohort study on a Korean population illustrates the rates of CVD risk factors and mortalities between different SES levels with demographic data shown in Table 4 [12]. The age distribution of the observed population has the oldest mean population for the higher SES group, while the middle SES was the youngest on average[12]. Cardiovascular health tends to deteriorate with age and the rates of CVD risk factors tended to be the greatest in the high SES population, while also having the oldest mean age; however, rates of HT and diabetes tended to align closest to the low SES group [12]. Middle SES individuals had the

lowest rates of HT of 36.5 compared to the high SES and low SES groups with 41.7% and 40.5% respectively [12]. The middle SES group also had the lowest rate for diabetes of 10% and the lowest mean BMI of 23.4 kg/mg2, while having the highest rate of current smokers [12]. Despite lower observed CVD risk factors in low SES compared to high SES groups, in respect to CVD mortality, low SES populations had a much higher rate of cardiovascular deaths of 1.2% compared to the high and middle groups who had 0.7% and 0.8% respectively [12]. This illustrates a trend of older groups inherently having higher rates of CVD risk factors; however, despite the age gap between the high and low SES populations their cardiovascular health was similar, while the youngest middle SES group had significantly lower rates of CVD risk factors [12].

Table 4. CVD statistics from observed population in Korean (Sung et al., 2020)

Country	**Characteristic**	**Identified (% or mean if specified)**		
Korea (2020) [12]	SES:	High	Middle	Low
	Age (mean)	46.2	40.8	43.4
	CVD mortalities	0.7	0.8	1.2
	CVD Risk Factors			

	Hypertension	41.7	36.5	40.5
	Diabetes	11.7	10	11.3
	Former Smoker	6.6	4	2.4
	Current Smoker	31.7	36.6	27.8
	BMI	23.9	23.4	23.2
	Cholesterol	198.2	192.1	191.8

Another study analyzed populations from two different regions: Australia, and Asia, using education as the measurement of SES to identify the relationship between SES and CVD [8]. CVD risk factor rates tended to increase among the "primary education" or "no education" groups showing an inverse relationship between education and CVD risk factors [8]. Diabetic rates were around 4.4% for the lowest educational group and 2.1% for the highest educational group [8]. Smoking rates were also more than double between the highest and lowest SES groups with 10% and 22% respectively [8].

Table 5. CVD statistics from observed population in Australia (Woodward et al., 2015)

Continent	Characteristic	Identified (% or mean if specified)		
Australia (2015) [8]	Education:	Tertiary	Secondary	≤ Primary
	Age (mean)	56	58	61

	CVD Risk Factors			
	Systolic BP	132	135	136
	Diabetes	2.1	3.3	4.4
	Smokers	10	19	22
	BMI	25.3	26	26.9
	Cholesterol	98.6	100.4	101.4

The observed populations in Asia (Table 6) illustrated a similar distribution of ages, but the secondary education group was the youngest with an average age of 48 [8]. CVD risk factors trended the same direction as the Australian population as the lowest education group had higher rates and mean measured CVD risk factors than the highest SES group [8]. However, mean total cholesterol levels trended the opposite direction as the tertiary education group had a mean total cholesterol of 88.1 mg/dL while the lowest education group had 86.3 mg/dL [8]. There was higher disparity in the smoking rates between SES compared to the Australian population and higher rates of smoking as the lowest SES group was composed of 42% current smokers and 16% in the highest group [8].

Table 6. CVD statistics from observed populations in Asia (Woodward et al., 2015)

Continent	Characteristic	Identified (% or mean if specified)

Asia (2015) [8]	Education:	Tertiary	Secondary	≤ Primary
	Age (mean)	50	48	55
	CVD Risk Factors			
	Systolic Blood pressure	123	124	124
	Diabetes	4.3	4.6	4.7
	Current Smokers	16	30	42
	BMI	22.7	22.9	23.3
	Cholesterol	88.1	87.2	86.3

DISCUSSION

Though a correlation between SES and CVD had been previously identified, a quantified measure of the disparity in cardiovascular health caused by economic factors was generally unknown. The key finding of this review is that there is a strong correlation between SES and CVD, with a large gap in CVD risk factors between lower and higher SES populations. In Bosnia and Herzegovina, higher disparity in rates of CVD risk factors of HT, smoking, diabetes, and cholesterol were found, which aligns with the high SES inequality shown in the bimodal distribution of wealth index in the respondent population. Another factor that was utilized as a measure of SES was education and occupational status, which showed a strong correlation

with CVD risk factors. The highest educated groups were associated with increased management rate for CVD risk factors, while having lower BMI, BP/HT, rate of diabetes, and smokers. The lowest educational groups exhibited the highest rates of CVD risk factors and lower rates of management in comparison to higher educated groups, which was also found between occupational classes. This trend could be identified in most analyzed populations with the exception of South Korea, who illustrated an inverse relationship with higher CVD risk rates associated with the highest SES group, which could be explained by the age distribution of the population with the majority of higher SES individuals being much older than the lower and middle SES groups. Despite this relationship found in South Korea, the low SES population had the highest mortality rates from CVD compared to the high SES group. Another relationship found throughout this review was high socio-economic disparity found between genders as males tended to have higher occupation and educational status compared to females; however, no explicit trend was found regarding CVD risk disparity between males and females in this review. The data shows that despite factors such as age and gender, lower SES populations based on educational attainment and occupational status tended to be disproportionately impacted by higher rates of CVD risk factors and tended to monitor their cardiovascular health less compared to higher SES groups. This shows a clear deviation in health related behavior between different SES populations that manifests itself through worsened cardiovascular health and higher rates of CVD in less advantaged populations.

The implications of understanding the gap in health and treatment of different SES populations and their CVD risk can help lower SES communities' susceptibility towards CVD. With proper knowledge on the importance and extent of the disparity between higher and lower income individuals can help make steps towards taking down the economic and educational barriers to healthcare and treatment that lower SES people face. Though this review is limited in its global scope of the issue of CVD and only looks at socioeconomic factors such as education, employment, and income without considering the multitude of economic and global factors that could impact the cardiovascular health of a population. Furthermore, data utilized for the purposes of this review may not be reflective of current populations and trends, especially given the potential confounding variables that could be playing a role in different countries. Additionally, self-reported data from

surveys introduces biases into the responses that may have skewed the data. This review, however, is still essential towards understanding the detrimental effect of CVD on lower SES communities. It is advised that further research and initiative is taken towards helping lower SES communities with their cardiovascular health to eliminate SES disparities in CVD.

References

1. What is Cardiovascular Disease. (2024). American Heart Association.
2. Mensah, G. A., Habtegiorgis Abate, Y., Abbasian, M., Abd-Allah, F., Abdollahi, A., Abdollahi, M., Morad Abdulah, D., Abdullahi, A., Abebe, A. M., Abedi, A., Abedi, A., Olusola Abiodun, O., Ali, H. A., Abu-Gharbieh, E., Abu-Rmeileh, N. M. E., Aburuz, S., Abushouk, A. I., Abu-Zaid, A., Adane, T. D., … Roth, G. A. (2023). Global Burden of Cardiovascular Diseases and Risks, 1990-2022. Journal of the American College of Cardiology, 82(25), 2350–2473. https://doi.org/10.1016/j.jacc.2023.11.007
3. Christina Magnussen, Francisco M. Ojeda, Darryl P. Leong, & et al. (2023). Global Effect of Modifiable Risk Factors on Cardiovascular Disease and Mortality. New England Journal of Medicine, 389(14), 1273–1285. https://doi.org/10.1056/NEJMoa2206916
4. Schultz, W. M., Kelli, H. M., Lisko, J. C., Varghese, T., Shen, J., Sandesara, P., Quyyumi, A. A., Taylor, H. A., Gulati, M., Harold, J. G., Mieres, J. H., Ferdinand, K. C., Mensah, G. A., & Sperling, L. S. (2018). Socioeconomic status and cardiovascular outcomes: Challenges and interventions. Circulation, 137(20), 2166–2178. https://doi.org/10.1161/CIRCULATIONAHA.117.029652
5. Sells, M. L., Blum, E., Perry, G. S., Eke, P., & Presley-Cantrell, L. (2023). Excess Burden of Poverty and Hypertension, by Race and Ethnicity, on the Prevalence of Cardiovascular Disease. Preventing Chronic Disease, 20. https://doi.org/10.5888/PCD20.230065
6. Arafa, A., Lee, H. H., Eshak, E. S., Shirai, K., Liu, K., Li, J., Anni, N. S., Shim, S. Y., Kim, H. C., & Iso, H. (2021). Modifiable Risk Factors for Cardiovascular Disease in Korea and Japan. In Korean Circulation Journal (Vol. 51, Issue 8, pp. 643–655). Korean Society of Cardiology. https://doi.org/10.4070/kcj.2021.0121
7. Jenkins, K. R., & Ofstedal, M. B. (2014). The Association Between Socioeconomic Status and Cardiovascular Risk Factors Among

Middle-Aged and Older Men and Women. Women and Health, 54(1), 15–34. https://doi.org/10.1080/03630242.2013.858098

8. Woodward, M., Peters, S. A. E., Batty, G. D., Ueshima, H., Woo, J., Giles, G. G., Barzi, F., Ho, S. C., Huxley, R. R., Arima, H., Fang, X., Dobson, A., Lam, T. H., & Vathesatogkit, P. (2015). Socioeconomic status in relation to cardiovascular disease and cause specific mortality: A comparison of Asian and Australasian populations in a pooled analysis. BMJ Open, 5(3). https://doi.org/10.1136/bmjopen-2014-006408
9. Damiani, G., Federico, B., Bianchi, C. B. N. A., Ronconi, A., Basso, D., Fiorenza, S., & Sassi, F. (2011). Socio-economic status and prevention of cardiovascular disease in Italy: Evidence from a national health survey. In European Journal of Public Health (Vol. 21, Issue 5, pp. 591–596). https://doi.org/10.1093/eurpub/ckq075
10. Janković, J., Erić, M., Stojisavljević, D., Marinković, J., & Janković, S. (2015). Socio-economic differences in cardiovascular health: Findings from a cross-sectional study in a middle-income country. PLoS ONE, 10(10). https://doi.org/10.1371/journal.pone.0141731
11. He, J., Zhu, Z., Bundy, J. D., Dorans, K. S., Chen, J., & Hamm, L. L. (2021). Trends in Cardiovascular Risk Factors in US Adults by Race and Ethnicity and Socioeconomic Status, 1999-2018. JAMA – Journal of the American Medical Association, 326(13), 1286–1298. https://doi.org/10.1001/jama.2021.15187
12. Sung, J., Song, Y. M., & Hong, K. P. (2020). Relationship between the shift of socioeconomic status and cardiovascular mortality. European Journal of Preventive Cardiology, 27(7), 749–757. https://doi.org/10.1177/2047487319856125.

Thien Nguyen

The Impact of Socioeconomic Factors on Cardiovascular Health

ABSTRACT

Background: Socioeconomic status plays a significant role in the cardiovascular health of an individual. The purpose of this paper is to synthesize and affirm the specific socioeconomic factors associated with changes in cardiovascular health.

Methods: This scoping review paper synthesized results from 5 research papers that were conducted across over 20 countries, including the United States, Canada, and Sweden. Pubmed and Google Scholars were the search engines of choice with key terms being socioeconomics and cardiovascular diseases, Education and Cardiovascular Health, and/or Occupation and Cardiovascular Diseases.

Discussion: Given the critical role education plays in cardiovascular health, further initiative should focus on investing in more equitable education to improve the overall cardiovascular health of the general population.

INTRODUCTION

Socioeconomic status is a measurement of the economic and social environment of one's life. There is an inverse relationship between such socioeconomic status and cardiovascular diseases [1]. For example, both educational level and occupation are associated with cardiovascular health. [2]. Countries of lower macro-socioeconomic status have a much larger risk of a stroke incident when compared to countries of higher macro-socioeconomic status [3]. In part, the association between educational level and cardiovascular health may be due to diet, with college graduates more likely to eat healthier items, such as whole fruit, vegetables, and whole grains, while people of lower educational status consume more saturated fat and sodium [4]. The objective of this paper is to synthesize the recent findings that describe socioeconomic status's impact on cardiovascular health.

METHODS

Search Strategy

The search was conducted through PubMed, and the search terms were socioeconomics and cardiovascular diseases, Education and Cardiovascular Health, and/or Occupation and Cardiovascular Diseases

Inclusion and Exclusion Criteria

The inclusion criteria was that the sample size was at least 1000 participants or a survey was nationally done, the paper was in English and that the sample size represented a general population (no research focusing on only one demographic). The exclusion criteria was if the sample size was too small, only focused on a certain group rather than a general population, papers not in English, or if the sample population only included one county or one city.

RESULTS

Study 1: Statistical Analysis from National Self-Reported Surveys in the United States

Across Black, Mexican, and White populations, hypertension was less common amongst individuals with a high school education or above [5]. In addition, individuals with less than a high school education were more likely to engage in smoking and less likely to be physically active with Mexican women having one exception (Mexican women with a high school education or higher were more likely to smoke) [5]. However, obesity rates didn't follow quite the same trend; although white individuals with a high school education or above were less likely to be obese, blacks and Mexicans with a high school education or above were more likely to be obese [5].

Study 2: Cohort study in 20 countries

Education level has a strong inverse correlation with major cardiovascular disease rates in high and middle income countries. However, this relationship is weaker in low income countries; between non-secondary educated individuals, secondary educated, and higher level educated individuals, secondary educated had the highest rates of major cardiovascular disease in poor countries [6]. There is a weak association between wealth and cardiovascular episodes [6]. In addition, education and healthy lifestyle choices of cessation from smoking, healthy eating, and physical activity have a positive correlation, with the exception of lower income countries for physical activity: non-secondary and secondary educated individuals and more physical activity than post secondary educated individuals [6].

Study 3: Statistical Analysis from US wide Cross-Sectional surveys

For risk factors of cardiovascular diseases, men tended to engage in riskier lifestyle choices than women. There is a gradient between education level and smoking for all sexes, and a gradient for women between education level and HDL cholesterol for women [7]. The correlation between income and cardiovascular disease and occupation and cardiovascular disease were weaker than the correlation between education and cardiovascular disease. However, one notable trend was that managers and executives displayed lower average blood pressure, and were the least likely to smoke [7].

Study 4: Case Control Study in Japan

Manual workers were the most likely to face the stroke subtypes of cerebral infarction, intracerebral hemorrhage, and subarachnoid hemorrhage amongst the 81 standard Japanese categories of occupations [8]. Amongst professional careers, the risk of the three stroke subtypes listed above were less prominent, with the exception of healthcare professionals, such as doctors, dentists, and veterinarians, and journalists/writers,who were at greater risk of cerebral infarction and subarachnoid hemorrhage [8]. Another exception was musicians, stage designers, and natural resource engineers, who were at greater risk of subarachnoid hemorrhage [8]. In addition, women were generally at a greater risk of cerebral hemorrhage and acute myocardial infarction for almost every occupation category [8].

Study 5: Meta-analysis in China

There is a strong correlation between educational level and cardiovascular diseases, where those with lower educational levels had a higher risk for all-cause mortality, stroke-specific mortality, recurrent stroke, and cardiovascular episodes, even with adjustment for lifestyle choices, such as smoking and drinking, age, and sex [9]. This association was independent of income [9]. This was attributed to often poor life management for cardiovascular health often found in lower educated communities [9]. There is an increasing gradient of cardiovascular episodes from college educated to middle school educated to primary school educated to illiterate, with those who are illiterate displaying the highest rate of cardiovascular complications [9].

DISCUSSION

While income and wealth are often thought to be the socioeconomic factor with the greatest impact on cardiovascular health, this scoping review has falsified this misconception. Education seemingly has a much larger impact on cardiovascular health, with occupation also playing a small role. However, it should be noted that education level is often tied with wealth despite wealth being independent of cardiovascular health. For example, a person of higher income is more likely to acquire higher level education, but should they get poor education, their wealth is not a determinant in their cardiovascular health. Occupation is a relatively weak determinant compared to education, but the fact that labor-intensive blue collar workers are more likely to experience strokes while managers and executives have healthier blood pressures cannot be ignored; a physically demanding career takes a toll on the body. The data demonstrates the importance of addressing inequitable education, particularly in the United States, where educational funding is based on local taxing. Policy makers should focus on offering American children equal educational opportunities starting from primary school, in turn ensuring improved cardiovascular health for the nation. For blue collar workers, employers should make an effort to at least ease the labor intensiveness of their worker, whether through some automation or better equipment.

One limitation of this review is that 1 of the 5 sources is over 20 years old. However, this only emphasizes that this trend has been going unaddressed for a while, with newer studies yielding results that further strengthen the credibility of the older one.

References

1. Schultz, W. M., Kelli, H. M., Lisko, J. C., Varghese, T., Shen, J., Sandesara, P., … & Sperling, L. S. (2018). Socioeconomic status and cardiovascular outcomes: challenges and interventions. Circulation, 137(20), 2166-2178.
2. Hahad, O., Gilan, D. A., Chalabi, J., Al-Kindi, S., Schuster, A. K., Wicke, F., Büttner, M., Tüscher, O., Lackner, K. J., Galle, P. R., Konstantinides, S., Daiber, A., Wild, P. S., & Münzel, T. (2024). Cumulative social disadvantage and cardiovascular disease burden and mortality. European journal of preventive cardiology, 31(1), 40–48.
3. Sommer, I., Griebler, U., Mahlknecht, P., Thaler, K., Bouskill, K., Gartlehner, G., & Mendis, S. (2015). Socioeconomic inequalities in

non-communicable diseases and their risk factors: an overview of systematic reviews. BMC public health, 15, 914
4. Psaltopoulou, T., Hatzis, G., Papageorgiou, N., Androulakis, E., Briasoulis, A., & Tousoulis, D. (2017). Socioeconomic status and risk factors for cardiovascular disease: Impact of dietary mediators. Hellenic journal of cardiology : HJC = Hellenike kardiologike epitheorese, 58(1), 32–42.
5. Mensah, G. A., Mokdad, A. H., Ford, E. S., Greenlund, K. J., & Croft, J. B. (2005). State of disparities in cardiovascular health in the United States. Circulation, 111(10), 1233-1241.
6. Rosengren, A., Smyth, A., Rangarajan, S., Ramasundarahettige, C., Bangdiwala, S. I., AlHabib, K. F., … & Yusuf, S. (2019). Socioeconomic status and risk of cardiovascular disease in 20 low-income, middle-income, and high-income countries: the Prospective Urban Rural Epidemiologic (PURE) study. The Lancet Global Health, 7(6), e748-e760.
7. Winkleby, M. A., Jatulis, D. E., Frank, E., & Fortmann, S. P. (1992). Socioeconomic status and health: how education, income, and occupation contribute to risk factors for cardiovascular disease. American journal of public health, 82(6), 816–820. https://doi.org/10.2105/ajph.82.6.816
8. Fukai, K., Furuya, Y., Nakazawa, S., Kojimahara, N., Hoshi, K., Toyota, A., & Tatemichi, M. (2021). A case control study of occupation and cardiovascular disease risk in Japanese men and women. Scientific Reports, 11(1), 23983.
9. Che, B., Shen, S., Zhu, Z., Wang, A., Xu, T., Peng, Y., Li, Q., Ju, Z., Geng, D., Chen, J., He, J., Zhang, Y., & Zhong, C. (2020). Education Level and Long-term Mortality, Recurrent Stroke, and Cardiovascular Events in Patients With Ischemic Stroke. Journal of the American Heart Association, 9(16), e016671. https://doi.org/10.1161/JAHA.120.01667.

Somya Nog

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

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.

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 $p\text{-value} < 0.05$ 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).

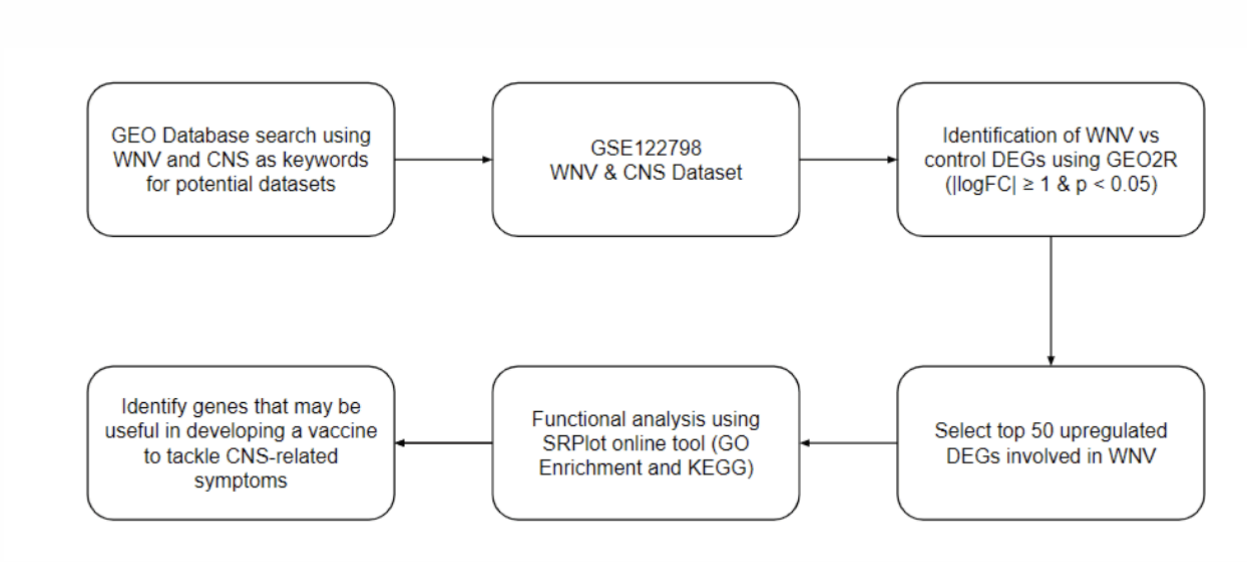


Figure 1: Summary of the methodology undertaken in this study. Overview of steps and bioinformatics tools and databases used in this study

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).

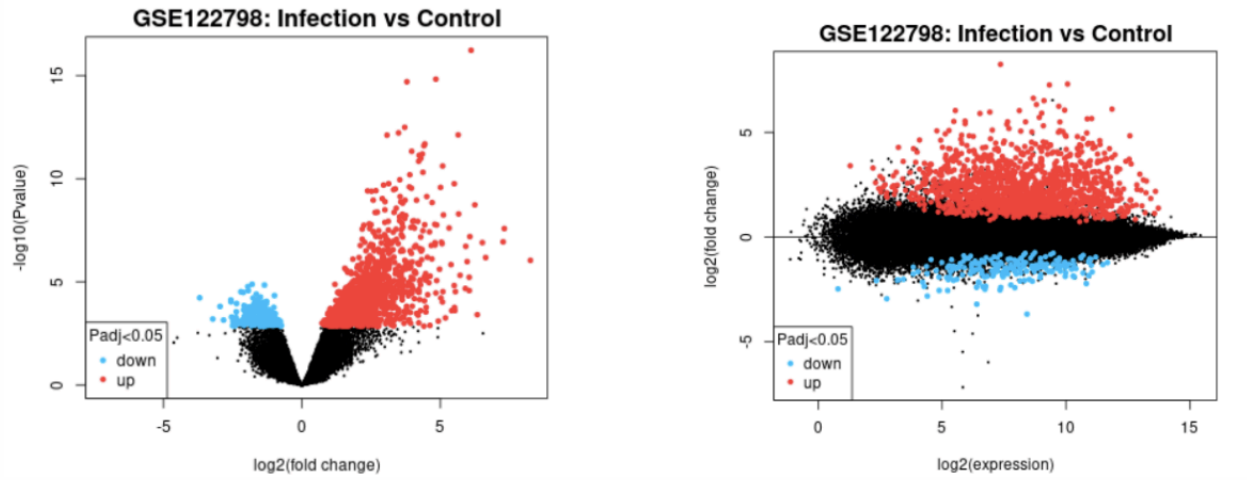


Figure 2a: Volcano plot illustrating the scope of differential expression in samples taken from patients with WNV infection vs healthy patients.

Figure 2b: Mean difference plot illustrating the scope of differential expression in samples taken from patients with WNV infection vs healthy patients.

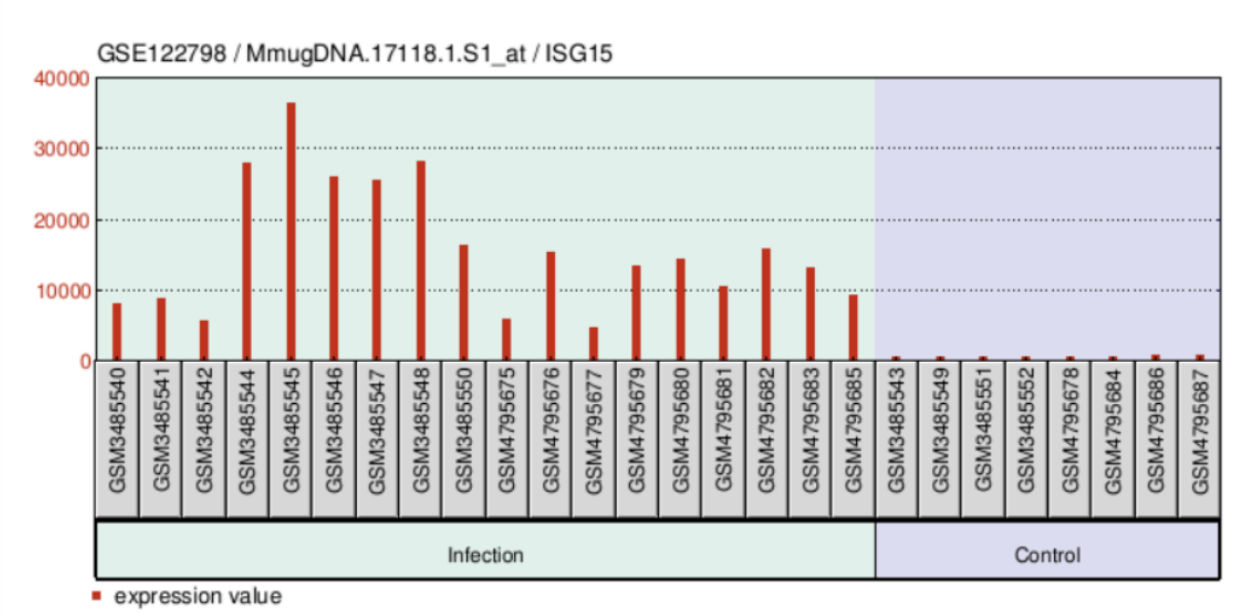


Figure 3: A bar plot-style depiction of the different levels at which the top DEG, ISG15, was expressed amongst the infection (n = 18) and control (n = 8) samples.

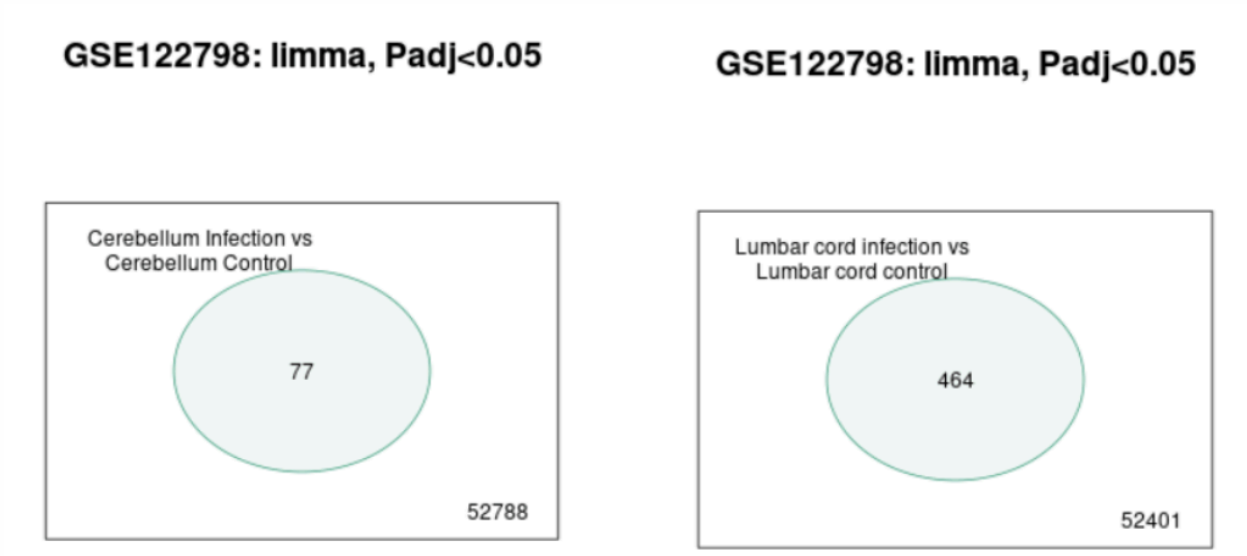


Figure 4a: A venn diagram illustrating the quantities of genes expressed at similar levels between the infection (n = 9) and control (n = 4) groups collected from the cerebellum.

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).

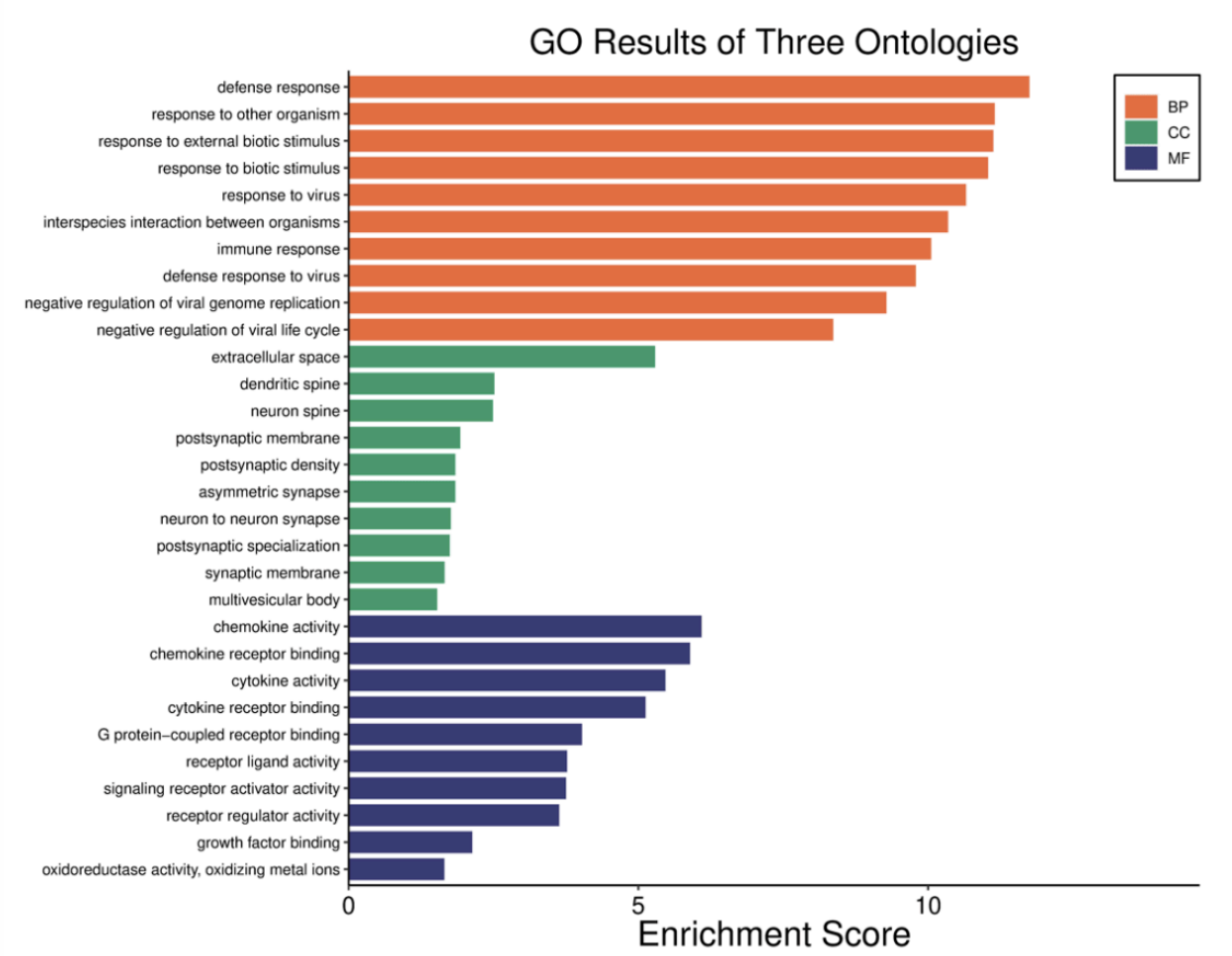


Figure 5a: The GO pathway enrichment results analyzing the level of representation of various biological processes (BP), cellular components (CC), and molecular functions (MF) within the top DEGs.

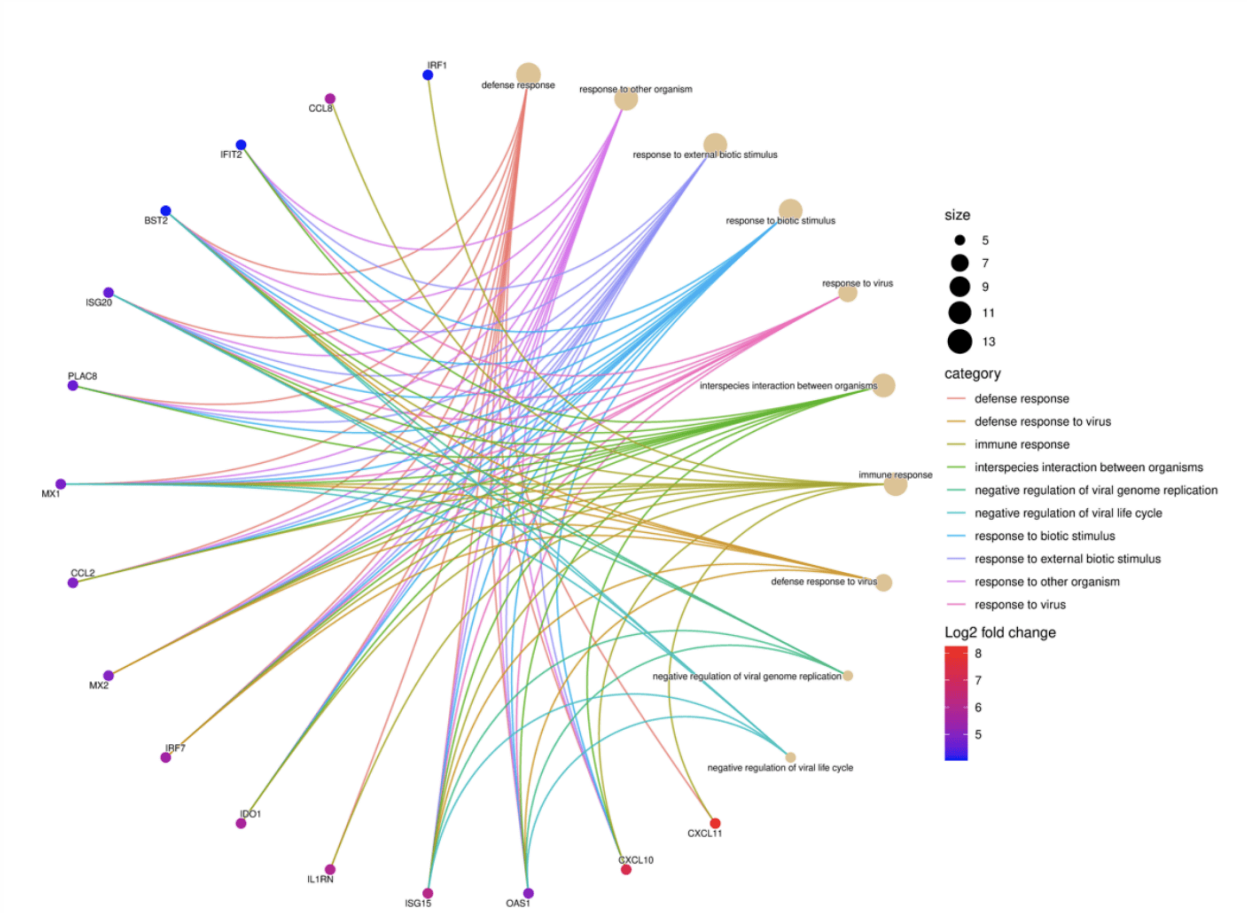


Figure 5b: Another graphic depicting the multiple roles the top 16 DEGs play in the infection process. The dots, or nodes, represent genes / biological functions, with each colored connection representing a biological function and the color of the dot representing the level at which each gene is upregulated, where red is the most significant.

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

1. 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.
2. 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.
3. Treatment and prevention of West Nile virus disease. (2024). Centers for Disease Control and Prevention.

4. 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.
5. 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.
6. 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.
7. 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.
8. Hay-McCullough, E., & Morrison, J. (2021). Contributions of Ubiquitin and Ubiquitination to Flaviviral Antagonism of Type I IFN. Viruses, 13(5), 763.
9. 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.
10. 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.
11. Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. Nucleic acids research, 28(1), 27-30.
12. 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.
13. 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.

14. 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.
15. 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.
16. Petersen, L. R., Brault, A. C., & Nasci, R. S. (2013). West Nile virus: review of the literature. Jama, 310(3), 308-315.
17. 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.
18. Perng, Y. C., & Lenschow, D. J. (2018). ISG15 in antiviral immunity and beyond. Nature Reviews Microbiology, 16(7), 423-439.
19. 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.
20. Rosa, S. S., Prazeres, D. M., Azevedo, A. M., & Marques, M. P. (2021). mRNA vaccines manufacturing: Challenges and bottlenecks. Vaccine, 39(16), 2190-2200.
21. 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.
22. 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.
23. Ulbert, S. (2019). West Nile virus vaccines–current situation and future directions. Human Vaccines & Immunotherapeutics, 15(10), 2337-2342.
24. 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.

25. 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.

Esha Nuttaki

Using Bioinformatics to Identify Potential Target Genes for Multiple Sclerosis Treatments

ABSTRACT

Background: Multiple Sclerosis (MS) is a progressive autoimmune disease affecting the central nervous system (CNS), primarily in young Caucasian adults. It damages the myelin sheath around neurons, leading to symptoms such as numbness, pain, muscle weakness, and cognitive issues. With no cure and limited treatment options, this study aims to identify target genes for MS treatment using bioinformatics tools to analyze gene expression and enrichment.

Methods: Bioinformatics tools NCBI GEO2R and SRplot for GO and KEGG analysis were used to study dataset GSE239698. Groups included RELA (RelA-deficient), REL (Rel-deficient), and WT (wild-type, normal). GEO2R was used to identify differentially expressed genes (DEGs), which were further analyzed with SRPlot for GO and KEGG pathway enrichment.

Results: The analysis identified 27 DEGs, with 24 upregulated and 3 downregulated. GO analysis revealed enrichment in terms including CD83, RIPK2, and the IKK pathway. KEGG analysis highlighted pathways such as NF-kB, cIAP1, and cIAP2. The NF-kB pathway was a common finding in both analyses, indicating its significant role in MS.

Discussion: Key genes like cIAP1/2, CD83, and RIPK2 play crucial roles in inflammation and immune regulation in MS. Inhibiting these genes could offer therapeutic benefits, though careful monitoring is needed to avoid side effects such as immune dysfunction or excessive cell death. This study provides valuable insights into gene-pathway relationships in MS, paving the way for future research and potential treatments targeting these pathways.

KEYWORDS: Multiple Sclerosis (MS), GEO2R, SR Plot, GO, KEGG, NCBI, cIAP1/2, CD83, RIPK2

INTRODUCTION

Multiple Sclerosis (MS) is an autoimmune disease that affects the brain and spinal cord, or the central nervous system (CNS) (1). During MS, the immune system mistakenly attacks the myelin sheath of neurons in the brain, spinal cord, and optic nerves. The immune system sends immune

cells, such as Tconv cells, to the CNS, where they attack the myelin sheath. This destruction of the myelin sheath causes inflammation around it, ultimately damaging it. The damage to the myelin causes nerve signals to slow down and become less efficient. As the disease progresses, the damaged myelin turns into sclerosis (scar tissue), which makes it extremely difficult for nerve signals to pass through the CNS (2). Multiple Sclerosis (MS) has no cure and limited treatment options. In this research paper, the relationship between MS and gene expression with different genes was studied to find a possible treatment or advancement in treatment for MS patients.

The scientific question being investigated in this study is which gene is most expressed or relevant in the context of MS. The research conducted in this experiment will help identify which genes are highly enriched and expressed in MS. Additionally, it will help identify possible genes that could be targeted for MS treatments.

MS is known to not have a specific cause, but rather many factors that may increase the likelihood of developing it. MS is most commonly diagnosed in young adults and is more prevalent in women, who are more likely than men to develop the disease (1). It is also more common in individuals of Caucasian descent. Other factors, such as smoking, obesity, and exposure to toxins, can increase the risk of MS (3). MS can affect the physical, mental, and cognitive health of those affected (1).

Bioinformatics continues to help researchers learn more about MS. Aspects that have already been studied using bioinformatics include genetic risk factors, Tconv involvement in inflammation, environmental and molecular factors that influence MS, and components of the NF-kB pathway. However, the exact cause of MS is still unknown, and there is no cure for this disease. Bioinformaticians still face challenges in diagnosing the disease in its early stages.

The objective of this research is to use bioinformatics tools and databases to further investigate and identify genes that can be targeted for treatments of Multiple Sclerosis. The bioinformatics tools used in this study were the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO2R), Scientific and Research Plot Tool (SRplot), Kyoto Encyclopedia of Genes and Genomes (KEGG), and Gene Ontology (GO). Specifically, NCBI GEO2R is a database where genomic data is submitted by researchers conducting experiments for others to analyze the results (4). Additionally, SRplot, KEGG, and GO bioinformatics were used to further analyze the results obtained from GEO2R. SRplot is a web server used to generate graphs and visualizations for many bioinformatics areas, including genomics (5). KEGG is a bioinformatics database that assigns functional meanings to genes and genomes and deals with biological pathways at molecular and higher levels (6). GO is an online bioinformatics database regarding gene and genome functions (7).

My hypothesis is that the gene RelA will be the most highly expressed gene in MS patients and the least expressed in healthy patients. This research is important for advancing treatment options for MS patients. By highlighting the potential of targeting specific genes as a possible treatment, this study lays the groundwork for future research that could lead to the development of new therapies for MS.

METHODS

Collection of Data Using NCBI GEO Bioinformatics Tool

To start the research process, the dataset GSE239698 was selected from the National Center for Biotechnology Information (NCBI) GEO sets (4). The chosen dataset, titled “NF-κB transcription factors RelA and c-Rel selectively control CD4+ T-cell function in multiple sclerosis and cancer [bulk_human]”, was originally created to explore how Tconv cells react without the RelA and c-Rel genes.

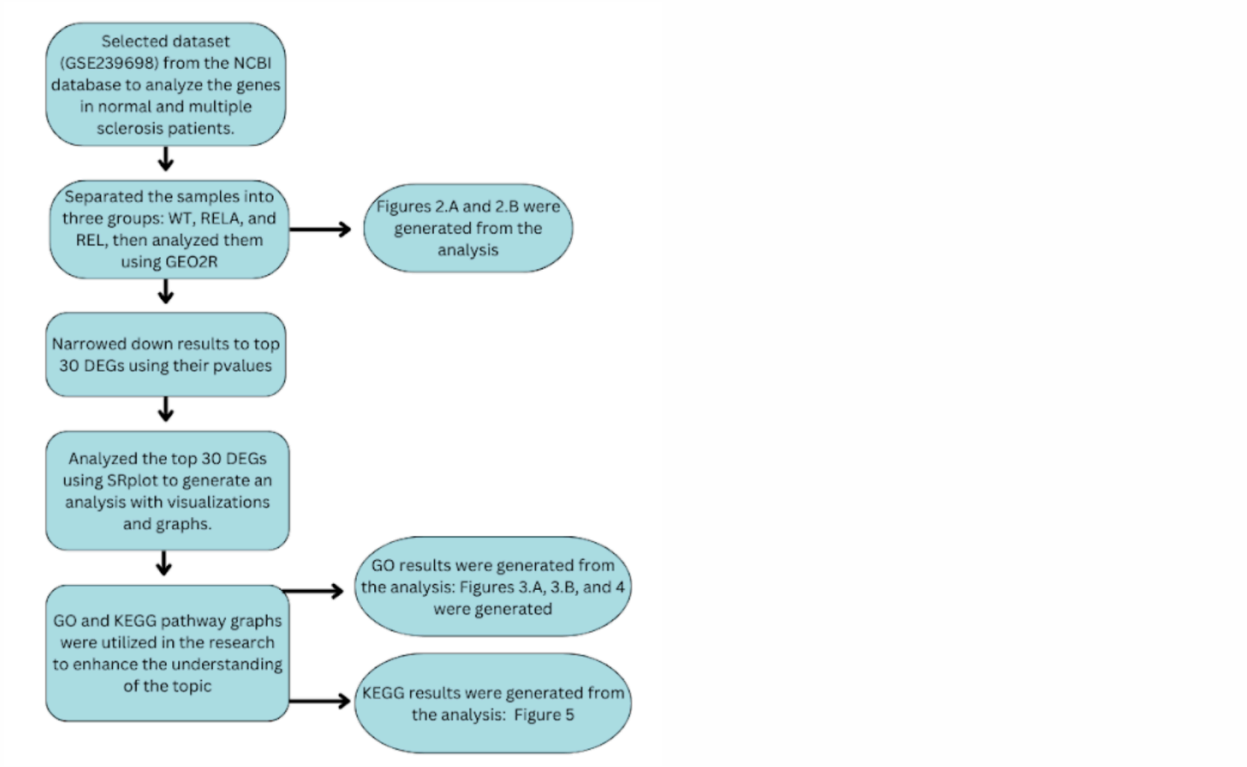


Figure 1. Research Methodology: Flow chart of the summary of methods and bioinformatics tools used in this research.

GEO2R Analysis

For the GEO2R analysis of this dataset, the 10 samples were separated into 3 groups: RELA, REL, and WT. The RelA group consisted of 3 samples, the REL group consisted of 3 samples, and the WT group consisted of 4 samples. Once the groups were defined, GEO2R analysis was performed to identify the differentially expressed genes (DEGs) and generate a variety of graphs depicting gene expression. The programming language used for the GEO2R analysis can be found in the R script (R script). From the analysis results, the Volcano Plot, Venn Diagram, and DEGs were utilized in this study. The DEGs generated from the GEO2R analysis all had a p-value less than 0.05.

Narrowing down differentially expressed genes using statistics

To identify the most significant differentially expressed genes, the entire list of DEGs from GEO2R was downloaded and transferred to a Google Sheet. Out of the 17,331 DEGs obtained from the GEO2R analysis, 30 were chosen

for deeper analysis (Top 30 DEGs). These 30 genes were selected based on having the lowest p-values among the 17,331 genes.

SR Plot Analysis to determine the potential functions of the genes

To further analyze the 27 DEGs, the SR Plot online web server was used to assign functions to the genes. The gene symbols and log2 fold changes (log2FC) were input into GO and pathway enrichment analysis in SR Plot for functional analysis. Upon completion of the analysis, a variety of graphs and visualizations were generated. These visualizations were grouped into two key categories: GO and Pathway. The GO and KEGG Pathway graphs were explored to find specific genes that were highly enriched. The genes that showed high enrichment and important functions in MS were chosen for further study.

RESULTS

Identification of Differentially Expressed Genes

The first bioinformatics tool used to identify the DEGs of the chosen dataset was GEO2R in NCBI. The DEGs were presented in the form of a table as a result of the GEO2R analysis.

In the volcano plot generated from the GEO2R analysis, the red dots represent upregulated genes, meaning they are highly expressed, and the blue dots represent downregulated genes, meaning they are expressed less than usual. The black dots represent genes that are not statistically significant.

From the Venn diagram produced from the GEO2R analysis, a total of 17,331 genes were identified. The Venn diagram revealed that the WT vs RelA comparison had 23 overlapping DEGs, the RelA vs REL comparison had 8 overlapping DEGs, and the overlap between WT vs RelA and RelA vs REL had 21 overlapping DEGs.

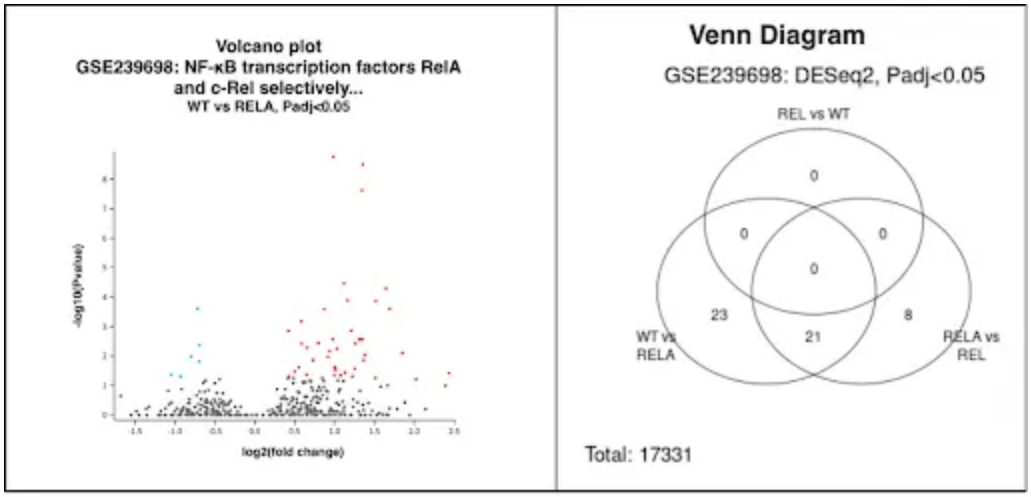


Figure 2. Identification of differentially expressed genes from collected samples: A Volcano plot shows gene expression sorted by -log10 (Pvalue and log2(fold change) between the RelA and WT groups. Each dot is a gene that is differentially expressed among the different groups. The upregulated genes are shown as red dots, while the downregulated genes are shown as blue dots. The top DEG from this graph was the NFKB2, followed closely by STARD10 and ANKRD33B. **B.** The Venn diagram shows the relationship between the RELA, REL, and WT groups, revealing that many DEGs overlap between the groups.

Identification of 30 Statistically Significant Differentially Expressed Genes (DEGs)

The statistical method used to narrow down the DEGs from the table was the P-value. Genes with the lowest P-values were selected because a low P-value indicates statistical significance. A total of 27 top DEGs were identified based on their P-values. Of the 27 genes identified, 3 were downregulated and 24 were upregulated. Top 27 DEGs

Potential Functions and Enrichment of the Identified Genes

The bioinformatics enrichment tool used to determine the potential functions of the genes was SR Plot. The most significant pathway identified from the KEGG results was the NF-kB (Nuclear Factor kappa B) signaling pathway. This pathway plays a crucial role in MS. In MS, the NF-kB pathway becomes overactive in immune and brain cells, leading to inflammation and driving the immune system to attack the body's own tissues (8). Using the pathway graph in Figure 3, genes involved in the NF-kB pathway were identified.

The genes that repeatedly appeared and were considered the most significant from the NF-kB pathway graph were cIAP1 and cIAP2 (Cellular Inhibitor of Apoptosis Protein 1 and 2). These genes play a crucial role in activating the NF-kB pathway. When cIAP1 and cIAP2 become overactivated, they can directly lead to the overactivation of the NF-kB pathway (9).

The pathway that repeatedly appeared in the GO results was the IκB kinase (IKK) pathway. IKK is a protein complex that is key for regulating NF-kB activation (10).

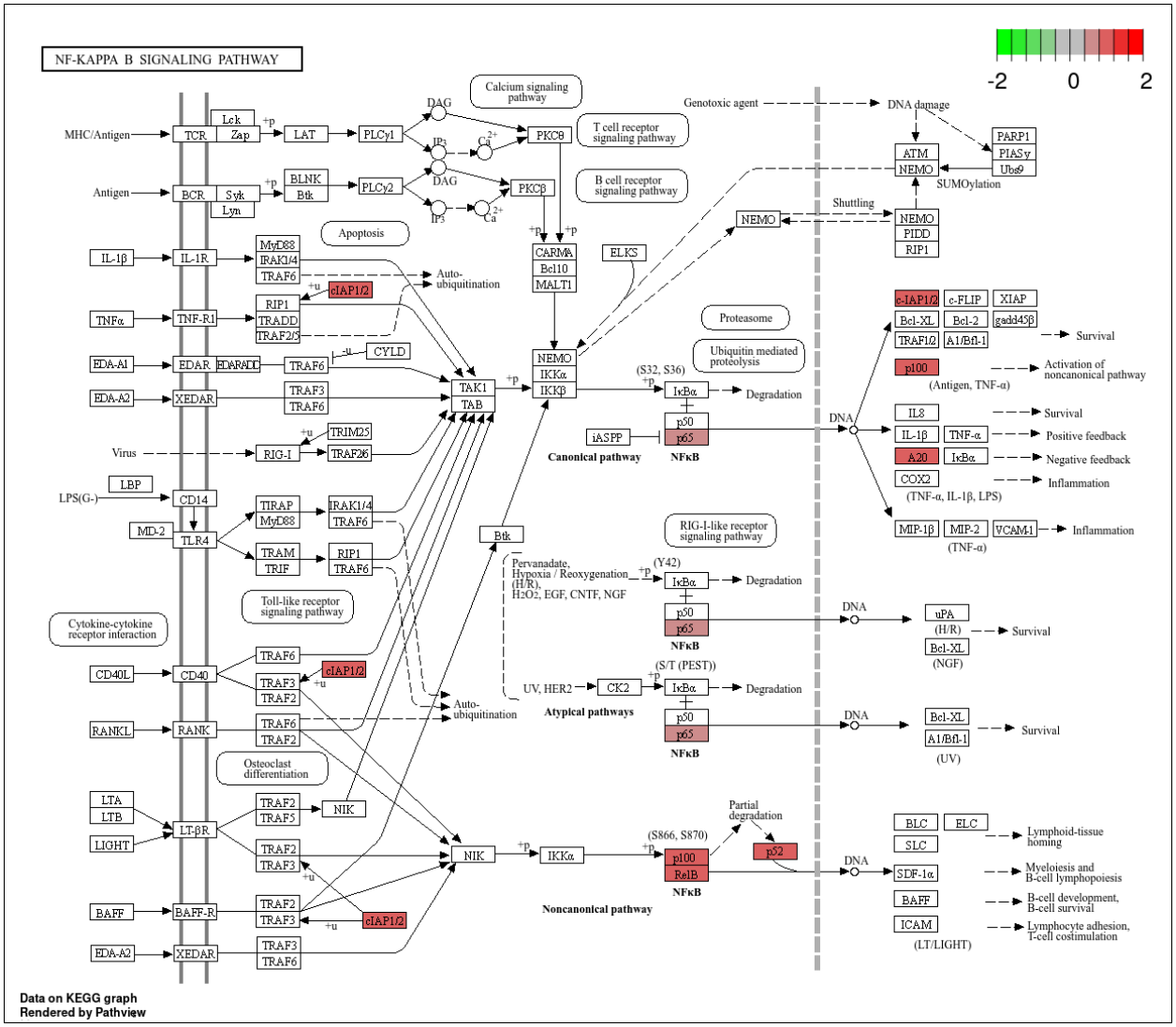


Figure 3. KEGG Results NF-KAPPA B signaling pathway: In this graph, green represents low significance, while red represents high significance in the pathway. The most significant genes in this pathway were identified as cIAP1 and cIAP2 (cIAP1/2).

The GO results highlighted two main genes: Cluster of Differentiation 83 (CD83) and Receptor-Interacting Protein Kinase 2 (RIPK2). The CD83 gene is important for dendritic cell activation and regulation. However, at elevated levels, CD83 is associated with the worsening of autoimmune diseases (11). The RIPK2 gene encodes a protein that helps activate dendritic cells. When RIPK2 is activated, it can lead to inflammation within the cell (12).

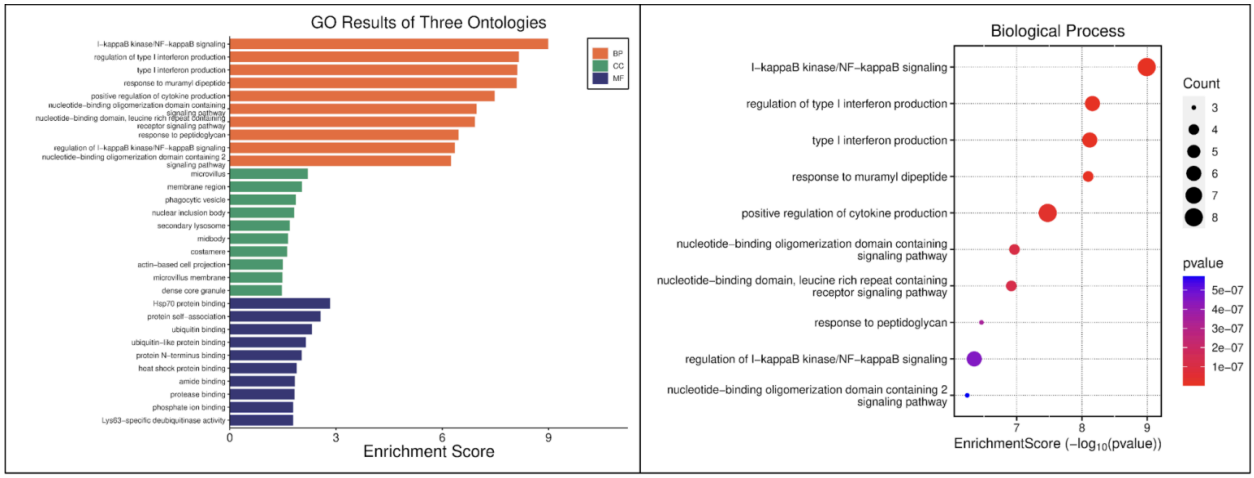


Figure 4. GO Analysis: A. The Go Results of Three Ontologies bar graph displays three ontologies and their enrichment within the genes analyzed. Orange represents Biological Processes (BP), green represents Cellular Components (CC), and blue represents Molecular Functions (MF). The BP ontology is significantly more enriched than CC and MF. The most enriched result in BP was the I-kappaB kinase (IKK) pathway. B. The Biological Processes dot plot shows the enrichment score, p-value, and count for each biological process. The top biological process identified is the I-kappaB kinase (IKK) pathway.

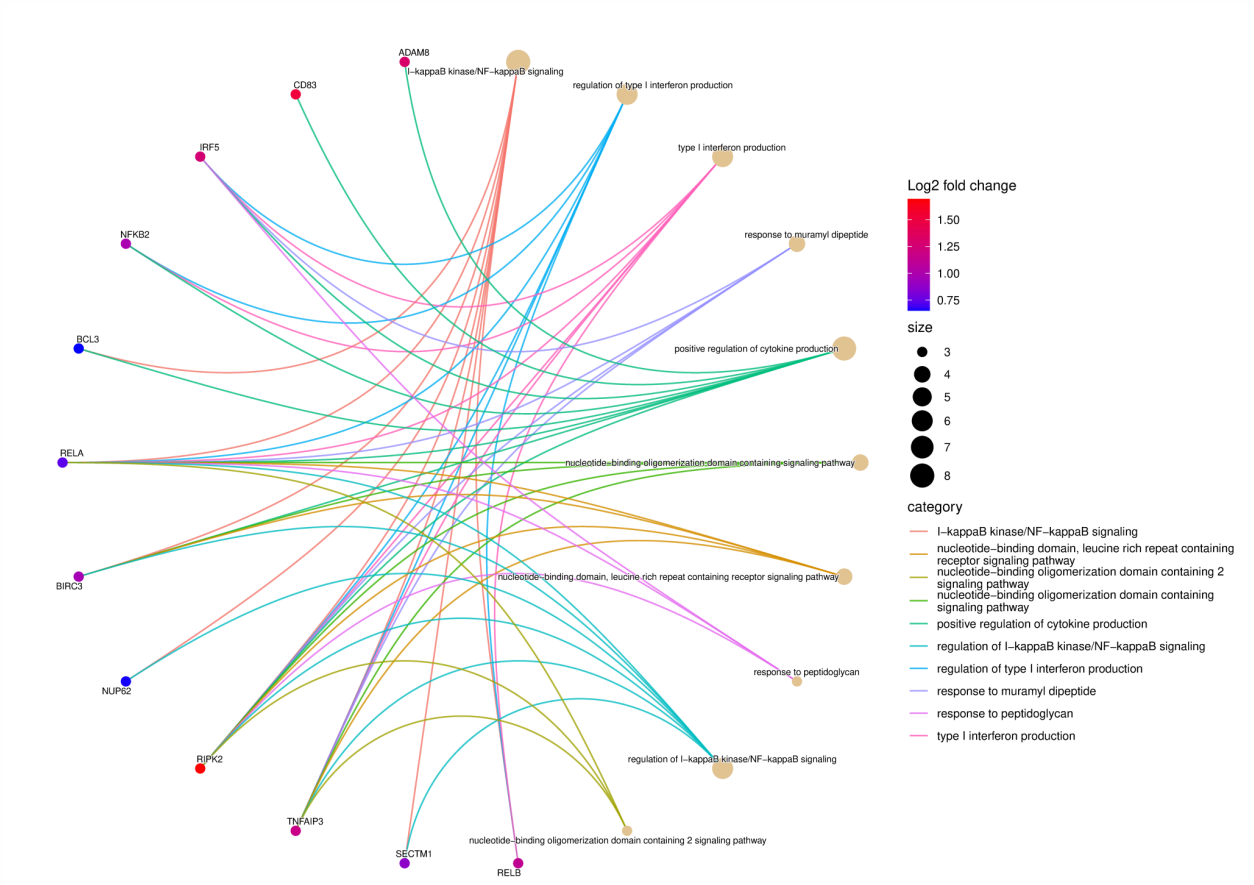


Figure 5 BP cnetplot: This cnetplot displays significant genes and the categories they are connected to. The most significant genes are shown in bright red, which represents a high Log2 fold change (Log2fc). The top genes in this plot are CD83 and RIPK2.

DISCUSSION

The main goal of this research was to identify key genes that could serve as potential targets for treatment using various bioinformatics tools. The results from the GEO2R and SRplot analyses revealed several genes and pathways that were consistently involved in significant pathways and showed high enrichment.

The significant genes identified in the graphs and tables could potentially serve as targets for MS treatments. The cIAP1/2, CD83, and RIPK2 genes (Figures 3 and 5) play key roles in the inflammation associated with MS (9, 10, 11, 12, 13, 14).

cIAP1 and cIAP2 (cIAP1/2) are two genes that appeared repeatedly in the NF-kB pathway graph (Figure 3). These genes play a crucial role in activating the NF-kB pathway and regulating immune cells that attack the CNS. Previous studies have shown that when cIAP1/2 genes become dysfunctional, they send uncontrolled signals to immune cells, leading to tissue damage and inflammation observed in MS (10). By precisely targeting these genes, their inhibition in MS patients could prevent the unchecked immune responses. However, since cIAP1/2 genes are important for preventing apoptosis, their inhibition could result in excessive cell death, which may lead to severe side effects. If cIAP1/2 inhibitors become a viable treatment for MS, close patient monitoring would be essential to prevent potentially dangerous side effects.

CD83 plays a key role in the maturation of dendritic cells, which act as messengers between different parts of the immune system and help activate immune cells. While healthy dendritic cells are crucial for the body's defense, abnormal dendritic cells contribute to inflammation and tissue damage seen in MS (13). Further research may reveal that inhibiting CD83 could prevent abnormal dendritic cells from maturing, potentially offering a treatment for MS. However, such a treatment would require extensive laboratory and animal studies before testing in humans. If not carefully controlled, inhibiting CD83 could halt the maturation of all dendritic cells, severely compromising the immune system and increasing vulnerability to infections (11).

RIPK2 is a crucial gene that activates the NF-kB pathway and helps regulate dendritic cells. It also plays a role in regulating cytokines, which are proteins released by cells to facilitate communication between them. Elevated levels of pro-inflammatory cytokines contribute to the inflammation seen in MS (14). In the future, RIPK2 inhibitors may be used as a treatment, but further research is needed to fully understand their effects. By targeting RIPK2 in MS patients, it may be possible to prevent inflammation driven by high cytokine levels and dysfunctional dendritic cells. With careful research, testing, and clinical trials, RIPK2 inhibitors could become a promising treatment for MS.

Table 1. Summary of Identified Genes and Pathways

Gene	Function	Gene Name	Connection
cIAP1/2	• Regulates NF-kB signaling • Prevents apoptosis(cell death) • Involved in cytokine production	Cellular Inhibitor of Apoptosis Protein 1 and 2	Keeps immune cells, which attack the CNS, alive.
CD83	• Protein that matures dendritic cells • Regulates T-cell response • Regulates immune activation	Cluster of Differentiation 83	Elevated levels of CD83 is linked to MS inflammation.
RIPK2	• Activates NF-kB pathway • Regulates dendritic cell activation and inflammatory cytokine production	Receptor-Interacting Protein Kinase 2	Promotes neuroinflammation and activated immune cells which attack the CNS.

Overall, the genes identified in this study could be tested in the laboratory or through clinical trials to determine whether the benefits of targeting these genes outweigh the potential risks.

One limitation of this research was the small sample size of the dataset used. Additionally, the study relied solely on secondary research, and no

access to a laboratory was available for further investigation of the genes. The identified genes will need to be studied in greater depth in a laboratory setting before any potential treatments for MS patients can be fully developed.

References

1. Yang, M., Sun, L., Han, J., Zheng, C., Liang, H., Zhu, J., & Jin, T. (2019). Biological characteristics of transcription factor RelB in different immune cell types: implications for the treatment of multiple sclerosis. Molecular Brain, 12(1). https://doi.org/10.1186/s13041-019-0532-6
2.
3. Guilhem Lalle, Raphaëlle Lautraite, Khaled Bouherrou, Plaschka, M., Pignata, A., Voisin, A., Twardowski, J., Marlène Perrin-Niquet, Stéphan, P., Durget, S., Tonon, L., Ardin, M., Degletagne, C., Alain Viari, Dutron, L. B., Davoust, N., Postler, T. S., Zhao, J., Christophe Caux, & Caramel, J. (2024). NF-κB subunits RelA and c-Rel selectively control CD4+ T cell function in multiple sclerosis and cancer. The Journal of Experimental Medicine, 221(6). https://doi.org/10.1084/jem.20231348
4.
5. Adamczyk- Sowa M, Gębka- Kępińska B, Kępiński M. MULTIPLE SCLEROSIS – RISK FACTORS. Wiadomości Lekarskie. 2020;73(12):2677–82.
6. Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Maxim Tomashevsky, et al. NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Research [Internet]. 2012 Nov 26 [cited 2025 Jan 13];41(D1):D991–5. Available from: https://pubmed.ncbi.nlm.nih.gov/23193258/
7. Tang D, Chen M, Huang X, Zhang G, Zeng L, Zhang G, et al. SRplot: A free online platform for data visualization and graphing. PLoS ONE [Internet]. 2023 Nov 9 [cited 2025 Jan 13];18(11):e0294236–6. Available from: https://pubmed.ncbi.nlm.nih.gov/37943830/
8. Minoru Kanehisa, Miho Furumichi, Tanabe M, Sato Y, Morishima K. KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research [Internet]. 2016 Oct 26 [cited 2025 Jan 13];45(D1):D353–61. Available from: https://pubmed.ncbi.nlm.nih.gov/27899662/

9. The Gene Ontology Resource: 20 years and still GOing strong. Nucleic Acids Research [Internet]. 2018 Nov 5 [cited 2025 Jan 13];47(D1):D330–8. Available from: https://pubmed.ncbi.nlm.nih.gov/30395331/
10. Lin W, Yue Y, Stone S. Role of nuclear factor κB in multiple sclerosis and experimental autoimmune encephalomyelitis. Neural Regeneration Research [Internet]. 2018 Jan 1 [cited 2025 Jan 17];13(9):1507–7. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC6126134/
11. Mahoney DJ, Cheung HH, Mrad RL, Plenchette S, Simard C, Enwere E, et al. Both cIAP1 and cIAP2 regulate TNFα-mediated NF-κB activation. Proceedings of the National Academy of Sciences [Internet]. 2008 Aug 19 [cited 2025 Jan 18];105(33):11778–83. Available from: https://www.pnas.org/doi/abs/10.1073/pnas.0711122105
12. Solt LA, May MJ. The IκB kinase complex: master regulator of NF-κB signaling. Immunologic Research [Internet]. 2008 Jul 14 [cited 2025 Jan 18];42(1-3):3–18. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC2965074/
13. Sajad Karampoor, Hamid Zahednasab, Masoud Etemadifar, Hossein Keyvani. The levels of soluble forms of CD21 and CD83 in multiple sclerosis. Journal of Neuroimmunology [Internet]. 2018 Apr 13 [cited 2025 Jan 18];320:11–4. Available from: https://www.sciencedirect.com/science/article/abs/pii/S016557281830064X
14. Shaw PJ, Barr MJ, Lukens JR, McGargill MA, Chi H, Mak TW, et al. Signaling via the RIP2 Adaptor Protein in Central Nervous System-Infiltrating Dendritic Cells Promotes Inflammation and Autoimmunity. Immunity [Internet]. 2011 Jan 1 [cited 2025 Jan 18];34(1):75–84. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC3057380/
15. AH Nuyts, Lee W, R Bashir-Dar, ZN Berneman, Cools N. Dendritic cells in multiple sclerosis: key players in the immunopathogenesis, key players for new cellular immunotherapies? Multiple Sclerosis Journal [Internet]. 2013 Jan 31 [cited 2025 Jan 19];19(8):995–1002. Available from: https://pubmed.ncbi.nlm.nih.gov/23369893/
16. Zhang JM, An J. Cytokines, Inflammation, and Pain. International Anesthesiology Clinics [Internet]. 2007 Jan 1 [cited 2025 Jan

19];45(2):27–37. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC2785020/
17. Yan, J., Winterford, C. M., Catts, V. S., Pat, B. K., Pender, M. P., McCombe, P. A., & Greer, J. M. (2018). Increased constitutive activation of NF-κB p65 (RelA) in peripheral blood cells of patients with progressive multiple sclerosis. Journal of Neuroimmunology, 320, 111–116. https://doi.org/10.1016/j.jneuroim.2018.04.002
18. Manuel, A. M., Dai, Y., Freeman, L. A., Jia, P., & Zhao, Z. (2020). Dense module searching for gene networks associated with multiple sclerosis. BMC Medical Genomics, 13(S5). https://doi.org/10.1186/s12920-020-0674-5
19. Riaz B, Islam S, Ryu H, Sohn S. CD83 Regulates the Immune Responses in Inflammatory Disorders. International Journal of Molecular Sciences [Internet]. 2023 Feb 1 [cited 2025 Jan 19];24(3):2831. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC9917562/.

Ankan Pal

The Effect of Education Inequality Upon Cardiovascular Disease Risk

ABSTRACT

Background: Cardiovascular disease is the leading cause of mortality, accounting for 15.6 million deaths in 2010 and 17.8 million deaths in 2017. CVD is often affected by non-medical factors known as social determinants of health.

Methods: Google Scholar and PubMed were used to find four original studies. Our exclusion criteria consisted of other literature reviews and any studies that compared groups with each other such as men to women.

Results: Among the studies reviewed, higher education is associated with a statistically significant decrease in CVD risk. One study showed that 1-ll of education leads to a 9% decrease in their participant's risk of developing CVD. Another study showed a 20%-40% decrease in CVD risk from the least to most educated men and women.

Discussion: Overall, education attainment and inequality were inversely associated with CVD risk. This indicates that countries and large organizations should use their influence and power to encourage, introduce, and improve education in developing countries.

INTRODUCTION

Cardiovascular disease (CVD) has increased the mortality rates in the United States since 2010[1]. Most commonly, nonmedical factors such as social determinants of health greatly impact CVD risk in all ages [2]. In 2010 alone, CVD caused an estimated 15.6 million deaths globally [3], and this number increased to 17.8 million in 2017 [1]. Additionally, CVD brings a catastrophic burden on health care and productivity. In 2010, CVD was the leading source of an estimated US $863 billion in healthcare and productivity costs; these numbers are only bound to worsen with a projected estimate of $30 trillion by 2030 [4].

Some of the major risk factors of CVD are cardiometabolic, behavioral, environmental, and social risk factors [5]. Among them, education leads to more informed decisions among individuals, more secure employment, and higher income to afford adequate healthcare and insurance [6]. Nevertheless, it is difficult for lower-income families to afford higher-level

education, which becomes detrimental to their health since occupation is one of the strongest predictors of any disease [7]. Additionally, many adolescents are unaware of the risk they put on themselves with their education since those years are pivotal in their development [8]. Globally, 17% of children (258 million) are either not in school or don't have access to any education [10]. These statistics are worse in developing countries, as 31% in sub-Saharan Africa and 21% in Central Asia are not in school compared to 3% in Europe and America [10].

This study aims to summarize the literature around education's impact on an individual's risk for developing CVD or experiencing a cardiovascular event in their life.

METHODS

This scoping review used Google Scholar and PubMed to find relevant studies using the key search terms: education, school, educational attainment, cardiovascular disease, and cardiovascular health. Among the studies with titles that fit the key terms, we excluded other literature reviews and any studies that compared groups with each other such as men to women because this study involves analyzing the effect of different levels of education and not race or gender.

RESULTS

In a 26-year study, 13,948 participants (56% female and 27% African) were followed up with their physicians and recorded 4512 CVD events and 2401 non-CVD-related deaths [11]. Amongst men, grade school level had a 59% risk of CVD; 52.5% for high school education (without graduation); 50.9% for high school (with graduation); 47.2%for vocational school, 46.4% for college (with or without graduation); and 42.2% for graduate/professional school [11].

In a second study, the Tromsø Study, 12,000 participants enrolled in a self-reported questionnaire [6]. It was documented that for every 1-level increase in education, the participant's risk of CVD or experiencing a cardiac-related event decreased by 9% [6]. However, after adjusting for covariates, the association was weaker [6].

A third study showed that the least educated men had a 21% higher risk of developing CVD than the most educated and a 41% higher risk for the least educated women [12].

All three studies above are consistent in the inverse correlation between education risk and CVD risk, however, one study displayed no correlation between CVD risk and education. In a randomized clinical trial, a program tested three groups–long-term intervention (LTI), short-term intervention (STI), and a control) for whom all the participants started with a baseline CVH score (10.3 for LTI, 10.3 for STI, and 10.5 for the control) and were analyzed at the end of the short-term intervention (STI), and the end of the long-term intervention (LTI) [8]. By the first analysis, the LTI group had a difference of 0.44 points in the CVH score change between the controls, and the STI group had a difference of 0.18. But, at the end of the 4 years, the differences between CVH score changes of the STI and the LTI against the control were 0.13 and 0.12 points [8]. The program did not provide any benefit nor was it disadvantageous to any participants, and there were no adverse events reported [8].

DISCUSSION

Cardiovascular disease has been the leading cause of mortality and morbidity across the globe. CVD has also been mediated by multiple factors, some of which are nonmedical factors like social determinants of health. This study analyzed the effect of education and how it affects individuals' risk of developing CVD.

From the reviewed studies, education was found to have a significant impact on CVD risk. Differences in education led to statistically distinct CVD risk among patients. In some cases, a single-level increase in education was associated with a 9% decrease in the risk of CVD in the participant's life [6] while the difference was strikingly even greater in a second study : the least educated men had a 21% higher risk of CVD and women had a 41% higher risk than the most educated men and women, respectively [12]. It's likely that education may have an impact on CVD risk, as higher education leads to more knowledge around risk factors and behaviors associated with CVD, which then allows for healthier informed decisions regarding lifestyle choices. Despite the associations seen in 3 studies, there was one study that displayed no significance on how different levels of education impacted CVD

risk or experiencing any cardiac-related issue in their lifetime [8]. This may have been due to a younger age in which the effects were indistinguishable or absent. Since most of the participants were teens, the effects may have not been as prominent compared to an elderly person.

One of the limitations of the searched studies is self-reported data risked recall bias. Additionally, the era in history in which people were born may have also affected educational attainment, as the oldest participants were in grade school during the Depression, and the youngest were born during WWII, a time of racially segregated schools [11]. Minorities may have had a decreased education from segregation and the war may have reduced education from enlistments so fewer young adults had a chance to extend their education and instead were fighting for their country.

Education is important, not only for a stable income but for the maintenance of health. Large organizations and other developed countries should look into foreign relations of developing countries to improve or even introduce education. Additionally, it's imperative to improve the quality of education in low-income neighborhoods even in developed nations, including to implement programs to help retain students in school.

References

1. Mensah G., Roth G., Fuster V. The Global Burden of Cardiovascular Diseases and Risk Factors: 2020 and Beyond. Journal of the American College of Cardiology. 2019; Volume 74: Pages 2529-2532
2. Powell-Wiley, T. M., Baumer, Y., Baah, F. O., Baez, A. S., Farmer, N., Mahlobo, C. T., Pita, M. A., Potharaju, K. A., Tamura, K., & Wallen, G. R. (2022). Social Determinants of Cardiovascular Disease. Circulation Research, 130(5), 782–799. https://doi.org/10.1161/CIRCRESAHA.121.319811
3. Mensah G., Roth G., Moran A., Narula J. The Global Burden of Cardiovascular Diseases, 1990–2010. Global Heart. 2014; Volume 9: Pages183-184
4. Bloom, D. E., Cafiero, E., Jané-Llopis, E., Abrahams-Gessel, S., Bloom, L. R., Fathima, S., … & Weiss, J. (2012). The Global Economic Burden of Noncommunicable Diseases (No. 8712). Program on the Global Demography of Aging.

5. Roth, G. A., Mensah, G. A., Johnson, C. O., Addolorato, G., Ammirati, E., Baddour, L. M., Barengo, N. C., Beaton, A. Z., Benjamin, E. J., Benziger, C. P., Bonny, A., Brauer, M., Brodmann, M., Cahill, T. J., Carapetis, J., Catapano, A. L., Chugh, S. S., Cooper, L. T., Coresh, J., Criqui, M., … GBD-NHLBI-JACC Global Burden of Cardiovascular Diseases Writing Group (2020). Global Burden of Cardiovascular Diseases and Risk Factors, 1990-2019: Update From the GBD 2019 Study. Journal of the American College of Cardiology, 76(25), 2982–3021. https://doi.org/10.1016/j.jacc.2020.11.010
6. Cathro, C. J., Brenn, T., & Chen, S. L. F. (2023). Education Level and Self-Reported Cardiovascular Disease in Norway-The Tromsø Study, 1994-2016. International journal of environmental research and public health, 20(11), 5958. https://doi.org/10.3390/ijerph20115958
7. Kustaa Piha, Mikko Laaksonen, Pekka Martikainen, Ossi Rahkonen, Eero Lahelma, Interrelationships between education, occupational class, income and sickness absence, European Journal of Public Health, Volume 20, Issue 3, June 2010, Pages 276–280, https://doi.org/10.1093/eurpub/ckp162
8. Santos-Beneit, G., Fernández-Alvira, J. M., Tresserra-Rimbau, A., Bodega, P., de Cos-Gandoy, A., de Miguel, M., Ramírez-Garza, S. L., Laveriano-Santos, E. P., Arancibia-Riveros, C., Carral, V., Orrit, X., Rodríguez, C., Carvajal, I., Haro, D., Peyra, C., Martínez-Gómez, J., Álvarez-Benavides, A., Estruch, R., Lamuela-Raventós, R. M., Fernández-Jiménez, R., … Fuster, V. (2023). School-Based Cardiovascular Health Promotion in Adolescents: A Cluster Randomized Clinical Trial. JAMA cardiology, 8(9), 816–824. https://doi.org/10.1001/jamacardio.2023.2231
9. Tamburini, R. (n.d.). World poverty statistics 2024. Social Income. https://socialincome.org/en/int/world-poverty-statistics-2024
10. Schmelkes, S. Recognizing and overcoming inequity in education. United Nations. 2022.
11. Kubota, Y., Heiss, G., MacLehose, R. F., Roetker, N. S., & Folsom, A. R. (2017). Association of Educational Attainment With Lifetime Risk of Cardiovascular Disease: The Atherosclerosis Risk in Communities Study. JAMA internal medicine, 177(8), 1165–1172. https://doi.org/10.1001/jamainternmed.2017.1877
12. Petrelli, Alessio et al. Education inequalities in cardiovascular and coronary heart disease in Italy and the role of behavioral and biological

risk factors. Nutrition, Metabolism and Cardiovascular Diseases, 2022, Volume 32, Issue 4, 918 – 928.
https://www.nmcd-journal.com/article/S0939-4753(21)00538-X/fulltext

Eshika Paruchuri

Utilizing Bioinformatics to Investigate Differences in Regulation of Gene Activity Between Whipple's Disease Affected and Asymptomatic Populations

ABSTRACT

Background: Whipple's disease is a rare and complex chronic infection caused by Tropheryma whipplei, primarily affecting the digestive system and potentially spreading to other organs. The infection begins in the small intestine, leading to symptoms like diarrhea, stomach pain, and weight loss. As it progresses, it can affect other organs, causing neurological issues such as memory loss and confusion. The disease primarily affects middle-aged males in North America and Europe, but the way it enters the human body and its origin remain unclear. Misdiagnosis is common due to its wide range of symptoms. This disease's wide-ranging symptoms and rarity make it difficult to diagnose and treat effectively. What makes this disease especially mysterious is that it is extremely rare, with only a few hundred cases around the world, and due to its wide array of effects across different organs in the body, its molecular mechanisms and patterns are poorly understood. This wide range of symptoms often leads to misdiagnosis, as it can mimic more common diseases like rheumatoid arthritis, celiac disease, or neurological disorders. The main objective of this research is to determine the effect of the levels of expression of certain genes in affected patients. This will be accomplished with the use of various bioinformatics tools to analyze and understand the significance of expression levels of certain genes found in the affected samples.

Methods: The bioinformatics databases that were used to conduct this research included the website for the National Center for Biotechnology Information (NCBI), NCBI's gene expression omnibus (GEO2r) tool and ShinyGO. GEO2R was used for analyzing pre existing primary research based on a study comparing two types of samples involved with Whipple's disease, unaffected and affected. The GEO Dataset that was used was #GSE20209. The primary research focused on using whole genome microarray analysis to pinpoint gene expression changes in interleukin-16-knock-out bone-marrow-derived macrophages after infection with T. whipplei. Following a 6-hour bacterial exposure, researchers discovered a specific gene expression pattern that set infected samples apart from uninfected ones. To

validate these results, they employed real-time PCR to measure the expression levels of several key genes within the same RNA samples, confirming the predicted macrophage response. The samples from the dataset were defined into two groups: affected and control (unaffected). There were three samples in each group, totaling 6 samples. After performing an analysis of the sample, the top 50 most differentially expressed genes (DEGs) were selected for further analysis via ShinyGO. ShinyGO produced a collection of various charts by assorting these top 50 DEGs into various molecular pathways, and associated molecules.
Results: A total of 53518 DEGs were identified upon analysis. As shown in Figure 1, there seemed to be more upregulated genes than downregulated ones. The top 8 GO terms upon enrichment using ShinyGO include ECM-receptor interaction, TNF signaling pathway, Rheumatoid arthritis, IL-17 signaling pathway, AGE-RAGE signaling pathway in diabetic complications, Amoebiasis, Protein digestion and absorption, and C-type lectin receptor signaling pathway.
Conclusion: Overall, the study revealed that the genes coding for the ECM-receptor interaction and TNF signaling pathway were the most significant in the enriched sample. Collagen was particularly repetitive, suggesting its role as a potential biomarker for future diagnosis or clinical trials. These findings can be used to understand how Whipple's disease moves across the body for infection as well as proper prevention.

INTRODUCTION

Whipple's disease is a rare chronic infectious disease caused by the gram-positive bacterium T. whipplei (Antunes & Singhal, 23). The bacterial infection begins in the digestive system, where vital organs reside, such as the small intestines. The bacteria penetrates the layers of the small intestines, causing lots of inflammation and poor digestion of the food passing through it leading to symptoms such as diarrhea, stomach pain and cramping, and weight loss (Boumaza et al., 2022). As the bacteria makes its way into other regions of the body, such as the nervous system, it can cause neurological symptoms including dementia, memory loss, confusion, difficulty finding words, lack of control of eye movements, and even personality changes (Song et al., 2023). Prior research has found that people who are usually male, middle-aged, and located in North America and Europe are more prone to this infection (Melas et al., 2021). Further studies have established the fact that the bacteria can reside in locations such as

farming areas, where the water can be contaminated (Johnson, 2021). However, the bacteria's true origin and its exact way of entry into the body are unknown.

Existing research has proven the many symptoms of this disease as well as the best way to keep the effects of the bacteria under control: antibiotic treatment (Bai, 2004). Inspection often involves a biopsy of the affected tissue, typically from the small intestine, showing characteristic findings under a microscope, such as the presence of periodic acid-Schiff (PAS) positive macrophages (Bakkali et al., 2008). Polymerase chain reaction (PCR) testing for T. whipplei DNA can also confirm the diagnosis (Dolmans 2017).

This study investigates how gene expression levels differ between individuals with Whipple's disease and unaffected controls. Using bioinformatics tools, this research analyzes gene expression data from infected and non-infected samples to identify significant differentially expressed genes. The goal is to uncover key molecular pathways involved in the disease and enhance understanding of its progression and diagnosis. This aim is to analyze gene expression differences between Whipple's disease patients and unaffected controls. The hypothesis is that there will be notable differences in gene expression patterns. This research is crucial for improving the understanding of Whipple's disease and its molecular mechanisms. Identifying key gene expression changes could enhance diagnostic accuracy and lead to potential biomarkers for early detection. The findings may also inform new treatment strategies, making the research valuable for better managing this rare and complex disease.

METHODS

NCBI and Primary Research

This research utilized bioinformatics databases from the National Center for Biotechnology Information (NCBI), the GEO2R tool (NCBI, "GEO2R"). The dataset being used was focused on a primary study that analyzed gene expression changes in response to T. whipplei infection using affected macrophages from bone marrow. The accession number for this dataset is #GSE20209. The primary research involved a whole genome microarray analysis of interleukin-16-knock-out bone-marrow-derived macrophages, examining how gene expression was altered after a 6-hour exposure to the

bacteria (GEO2R, 2010). Distinct gene expression patterns differentiated infected samples from uninfected ones and validated these findings using real-time PCR on the same RNA samples.

GEO2R Analysis

The dataset included two groups: affected (infected) and control (unaffected), with three samples in each group, making a total of six samples. The two groups were defined such that 3 samples were put into 'affected', and three into 'control'. Analysis was performed via the "Analyze with GEO2R" option. The GEO2R bioinformatics tools use R programming that is pre-set with AI and machine learning algorithms. The linked R script gives more information about how the program works: R Script used to generate GEO2R results. Thereafter, the top 50 most significantly differentially expressed genes (DEGs) were identified and chosen for further examination using ShinyGO.

ShinyGO Bioinformatics Analysis

The top 50 DEGs were uploaded to a Google sheet in order, and their gene symbols were taken. These 50 gene symbols were then uploaded into the empty box on the ShinyGO bioinformatics website. After entering the symbols, analysis was performed using the "submit" option which yielded various charts and diagrams on the right, ranging from charts sorted by log fold, number of genes, pathways, and more. After comprehending these charts, the top two most enriched pathways, the ECM-receptor interaction and TNF signaling pathway were acknowledged as the most significant in the affected sample involving Whipple's disease.

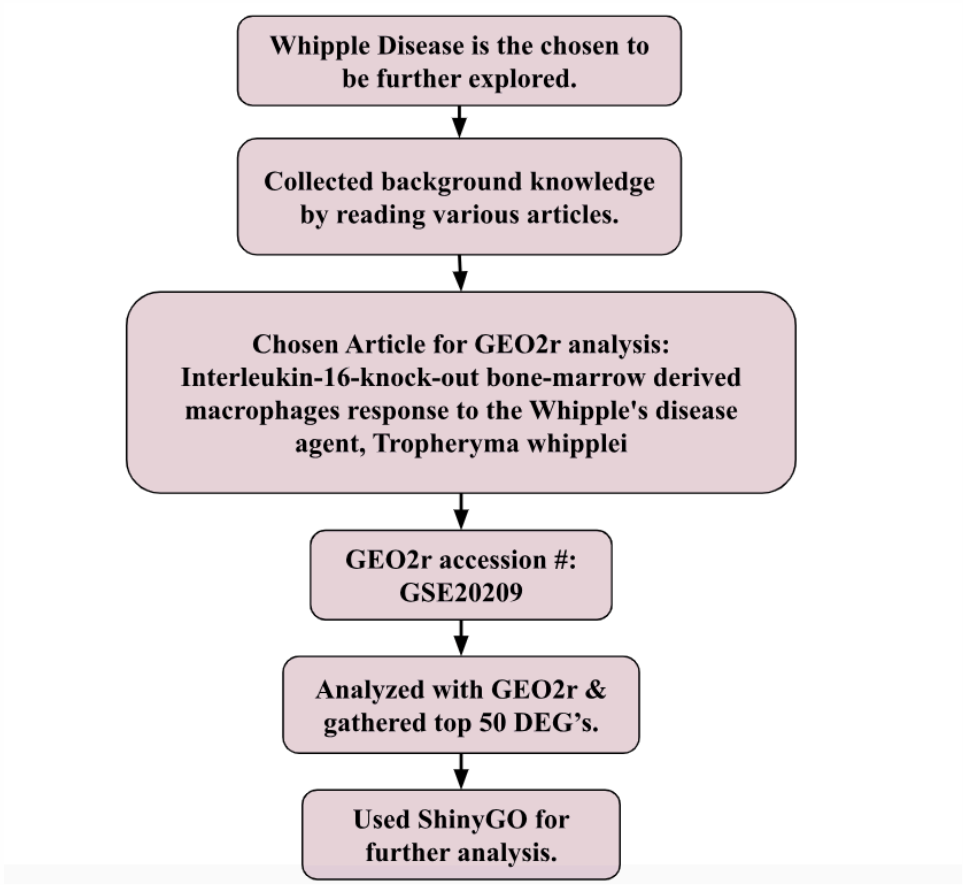


Figure 1. Showing experimental design.

Steps to obtain the results

To obtain the results, multiple bioinformatics tools and databases were utilized, including the National Center for Biotechnology Information (NCBI), GEO2R, and ShinyGO. The process began by identifying a relevant GEO dataset focused on Whipple's disease, specifically Dataset #GSE20209. This dataset included two groups: affected (infected) and control (unaffected), with three samples in each group, totaling six samples. The primary research involved using a whole genome microarray to investigate gene expression changes in bone-marrow-derived macrophages after a 6-hour exposure to T. whipplei bacteria.

The analysis started by accessing GEO2R, a tool within NCBI, which enables users to compare gene expression in different sample groups. By selecting the "Analyze with GEO2R" option, the expression levels of genes in affected versus control samples were compared. GEO2R uses R programming

pre-configured with AI and machine learning algorithms to carry out this differential gene expression analysis. This comparison identified a total of 53,518 differentially expressed genes (DEGs). Among these, the top 50 most significantly different DEGs were selected for further pathway analysis.

These top 50 DEGs were then organized in a Google Sheet, where their gene symbols were listed and copied. The next step was to input these gene symbols into the ShinyGO tool, a web-based application used for functional enrichment analysis and visualization. By entering the gene symbols into the designated field and selecting the "submit" option, ShinyGO generated various charts and diagrams. These visual outputs included pathway analyses, gene ontology terms, and molecular interactions based on the inputted DEGs.

From the generated results, the pathways were assessed, with a particular focus on the most enriched pathways relevant to Whipple's disease. Among the outputs, the ECM-receptor interaction and TNF signaling pathways emerged as the most significant in distinguishing affected samples from controls.

RESULTS

GEO2R Results: Identification of Differentially Expressed Genes

In this research, 50 DEGs were identified based on fold change and p values. There were 25 upregulated and 25 downregulated genes included in the GEO2R Top 50 DEGs Google Sheet. Using ShinyGO, these top DEGs were analyzed to produce charts containing the matching processes, pathways, and associated molecules. Some of the top pathways include ECM-receptor interaction, TNF signaling pathway, Rheumatoid arthritis, and the IL-17 signaling pathway. The top two were taken into account due to the number of DEGs associated with them.

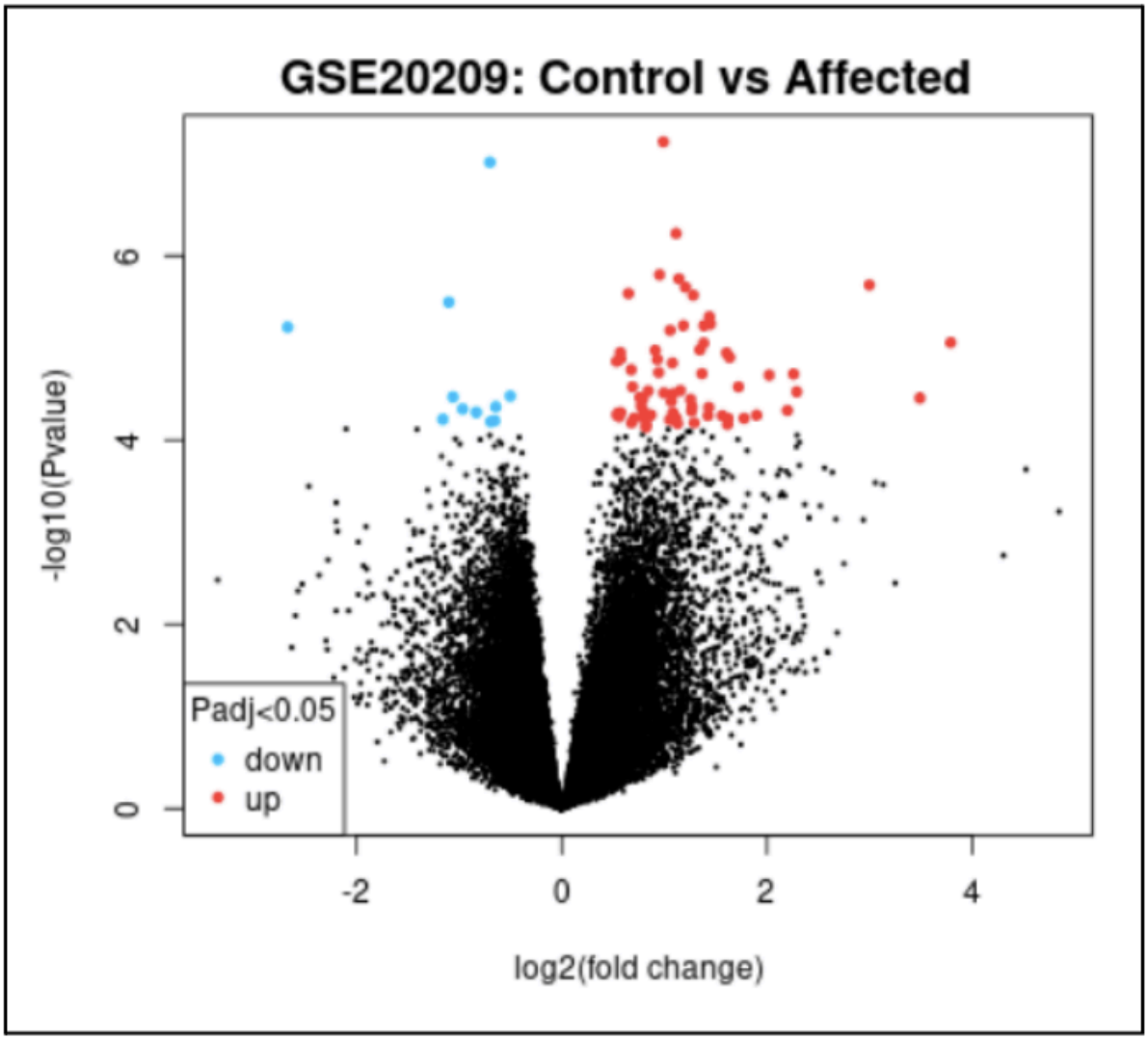


Figure 2. Volcano Plot. This volcano plot shows 52518 differentially expressed genes (DEGs) represented with each dot, both red and blue. Red dots represent upregulated genes, and blue dots represent downregulated genes. More genes were turned “on.”

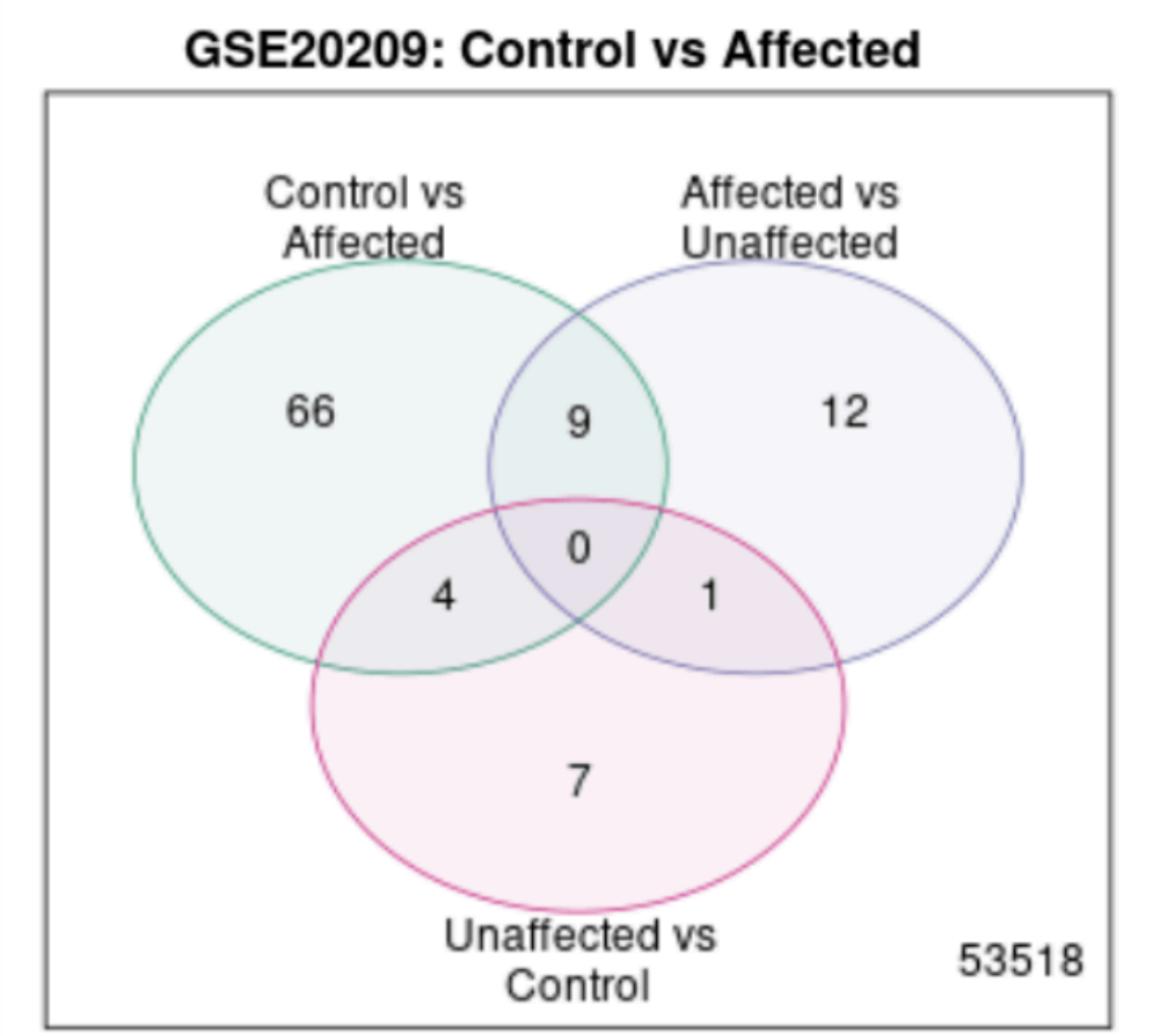


Figure 3. Venn Diagram. Within a total of 53518 genes, there were no common genes expressed in both groups, when taking "unaffected" into account, however, there is some regional overlapping between any two given groups.

Identification of Significantly Expressed Genes and Key Pathways in Whipple's Disease Using ShinyGO Analysis

To further analyze the identified differentially expressed genes, the ShinyGO bioinformatics tool was used and it produced a collection of various charts by assorting these top 50 DEGs into various molecular pathways, and associated molecules.

Enrichment FDR	nGenes	Pathway Genes	Fold Enrichment	Pathways (click for details)
8.6E-03	3	88	30	ECM-receptor interaction
8.8E-03	3	112	23.6	TNF signaling pathway
1.4E-02	4	354	9.9	PI3K-Akt signaling pathway
2.3E-02	3	200	13.2	Focal adhesion
2.3E-02	3	214	12.3	Lipid and atherosclerosis
3.2E-02	2	104	16.9	C-type lectin receptor signaling pathway
3.2E-02	2	93	18.9	IL-17 signaling pathway
3.2E-02	2	106	16.6	Parathyroid hormone synthesis secretion and action
3.2E-02	2	100	17.6	AGE-RAGE signaling pathway in diabetic complications
3.2E-02	2	103	17.1	Protein digestion and absorption
3.2E-02	2	102	17.3	Amoebiasis
3.2E-02	3	331	8	Human papillomavirus infection
3.2E-02	2	92	19.1	Rheumatoid arthritis

Figure 4. Enrichment Chart. In this chart, fold enrichment was taken into consideration among all factors for measurement of how significant the pathways are. Fold Enrichment measures the magnitude of enrichment. Higher values indicate stronger enrichment and are an important metric of effect size. As shown, the ECM-receptor interaction and TNF signaling pathway have the highest enrichment scores with 30 and 23.6, respectively indicating their significance.

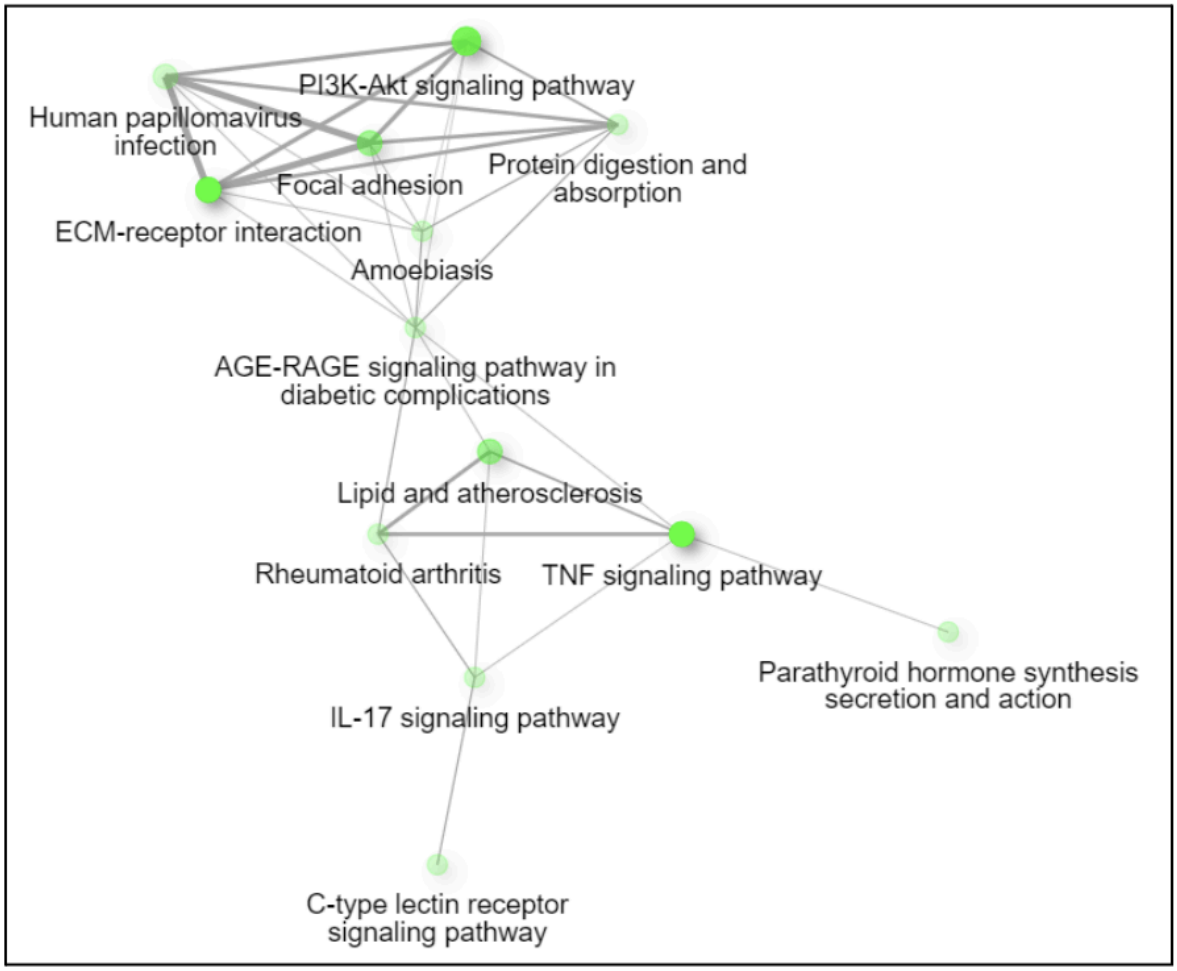


Figure 5. Network. This interactive plot shows the relationship between enriched pathways. Two pathways (nodes) are connected if they share 20% or more genes. Brighter nodes indicate more significantly enriched gene sets and larger nodes represent bigger gene sets.

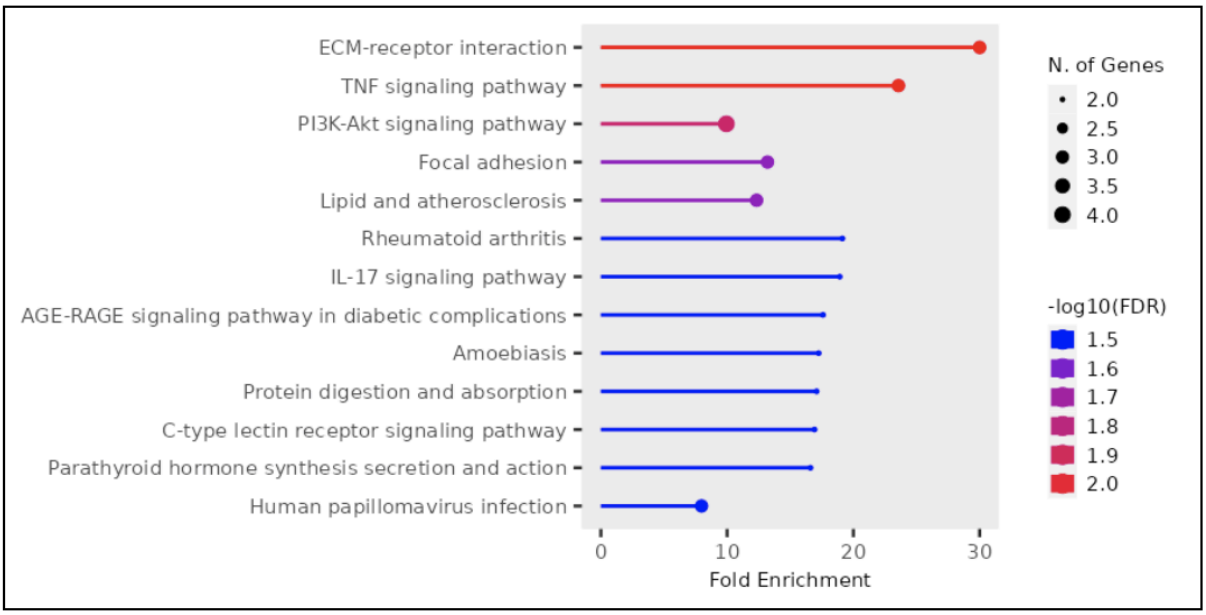


Figure 6. Lollipop chart. The thicker dots represent more enriched pathways with more genes involved from the selected 50 DEGs. The most enriched signaling pathways were the ECM-receptor interaction, TNF signaling pathway, and the PI3K-Akt, but only the top two were taken into account in this study due to their FDR value and color, represented as a scale of blue to red, with 2.0 as the brightest red shade. Only the top two pathways had an FDR value of 2.0.

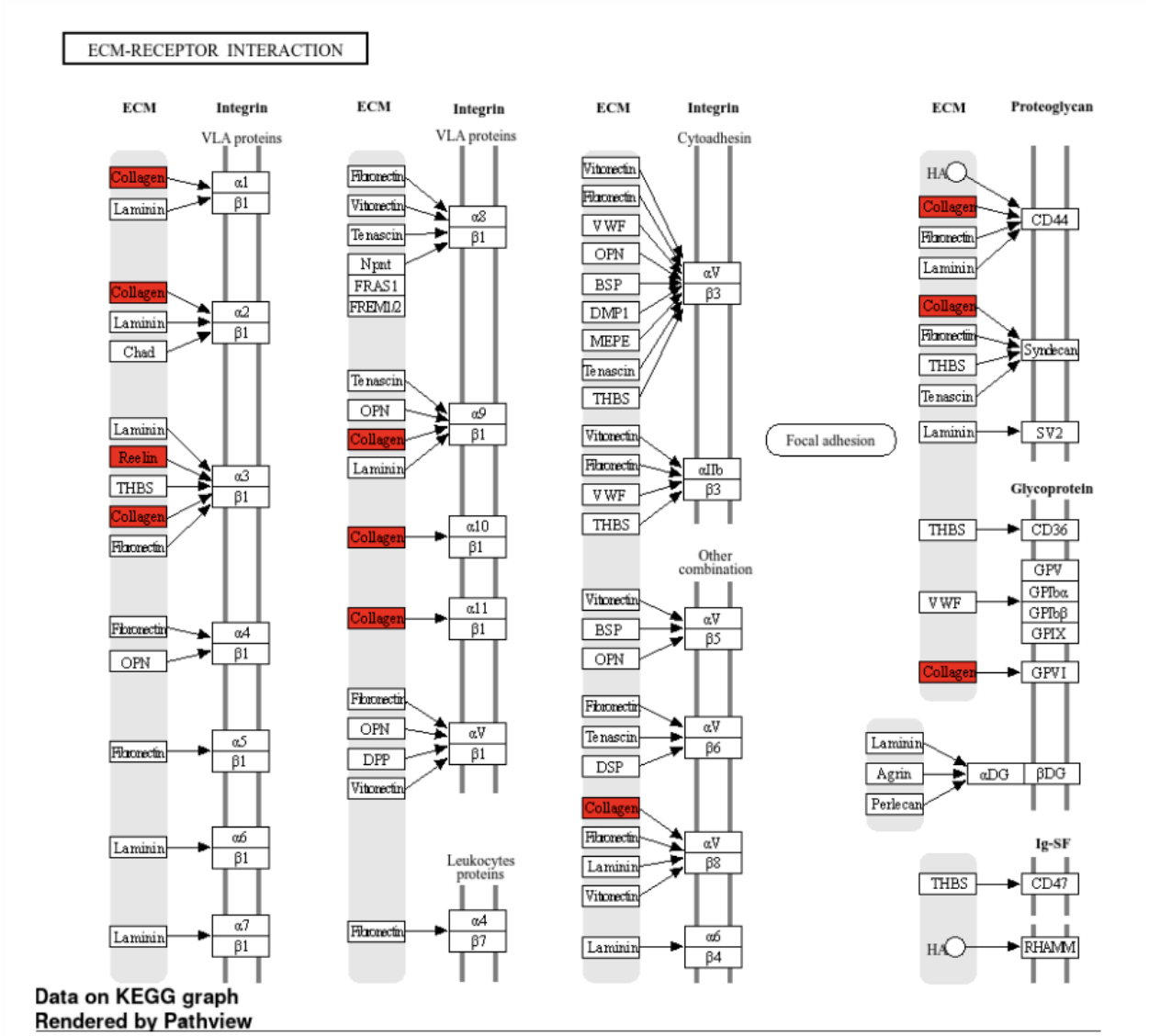


Figure 7. ECM-receptor interaction KEGG Pathway. Integrin proteins and proteoglycan were involved in the ECM-Receptor interaction. Collagen plays a large role in this pathway due to its repetition as shown in red, along with one other gene, Reelin.

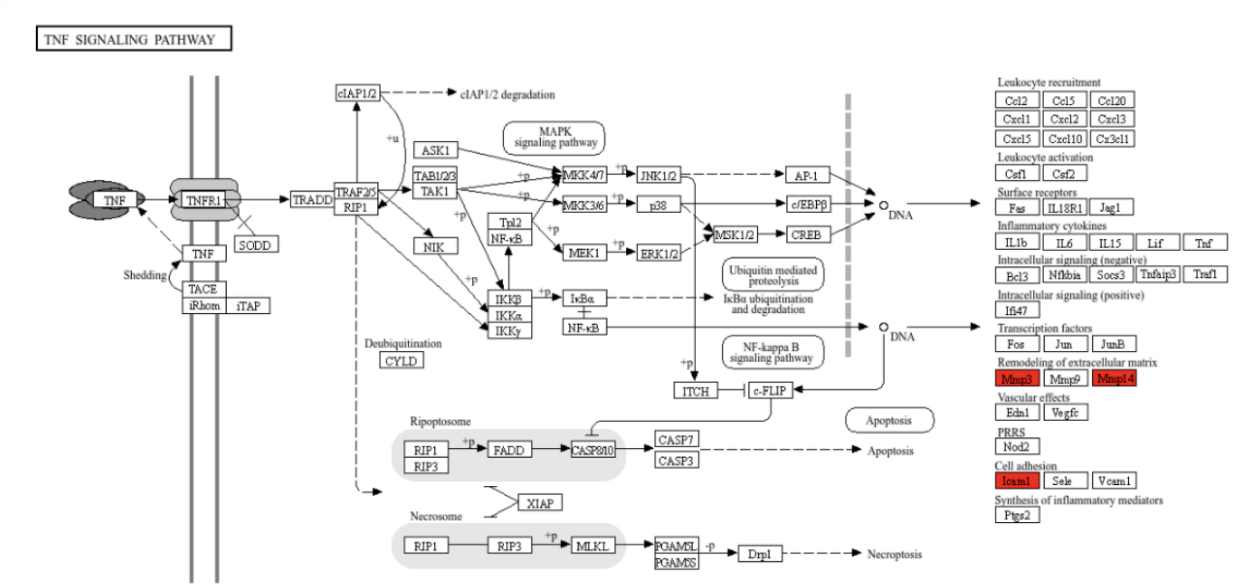


Figure 8. TNF Signaling Pathway. The genes Mmp3, Mmp14, and 1cam1 are involved in the remodeling of the extracellular matrix and cell adhesion in the TNF signaling pathway, respectively.

Table 1. Summary of Findings

Significant Genes from Enrichment analysis (based on repetition)	**KEGG Pathways**	**Common Results**
COL2A1	ECM-receptor interaction & COL2A1 and Reelin	COL2A1
	TNF signaling pathway & Mmp3, Mmp14, and 1cam1	

Supplementary Materials

R Script: R Script used to generate GEO2R results
Top 50 DEGs: GEO2R Top 50 DEGs Google Sheet

The primary goal of this research was to investigate gene expression differences between individuals affected by Whipple's disease and unaffected controls. Using bioinformatics tools such as GEO2R and ShinyGO, the study analyzed gene expression data from macrophages exposed to T. whipplei. The analysis identified the top 50 differentially expressed genes (DEGs) as shown in the Google sheet, with significant involvement in pathways (figure 5) such as ECM-receptor interaction (figure 6) and TNF signaling (figure 7). These findings suggest that specific genes, particularly those related to immune response and inflammation, play a crucial role in the disease's progression and could serve as potential biomarkers for diagnosis and treatment.

The findings of this study align with previous research on Whipple's disease, particularly in terms of the role of the extracellular matrix (ECM) and the TNF signaling pathway (Wang et al., 2020). Previous studies have also indicated that Whipple's disease involves significant disruptions in ECM interactions and chronic inflammation, often driven by TNF signaling (Jarosz-Griffiths et al., 2019). For example, the upregulation of ECM-related genes and inflammatory pathways has been documented in other research focusing on the immunological and pathological aspects of Whipple's disease (Friebel et al., 2022).

However, this study adds a nuanced perspective by identifying specific gene expression patterns and pathways involved, particularly the role of collagen (figure 6) and the ECM-receptor interaction (figures 5 and 6) in disease progression. This detailed gene expression analysis and pathway enrichment extend previous knowledge by providing more specific insights into the molecular mechanisms underlying Whipple's disease.

ECM-receptor interaction

The extracellular matrix (ECM) and its interaction with cellular receptors play a significant role in the pathogenesis of Whipple's disease. The ECM is a complex network of proteins, including collagen, that provides structural support to tissues and facilitates cell signaling (Büttner, 2018). In Whipple's disease, the interaction between the ECM and its receptors is disrupted by the presence of the bacterium Tropheryma whipplei, leading to chronic inflammation and tissue damage. This disruption contributes to symptoms such as diarrhea, malabsorption, and systemic manifestations as the disease

progresses. Specifically, the upregulation of ECM-related genes in affected individuals suggests that changes in ECM composition may be involved in the immune response and tissue remodeling processes seen in Whipple's disease. Such alterations can exacerbate the disease's impact on the digestive system and other organs, leading to the diverse and severe symptoms observed in patients (Soles, 2023).

TNF signaling pathway

The TNF (Tumor Necrosis Factor) signaling pathway is closely linked to the inflammatory processes observed in Whipple's disease. TNF is a cytokine that plays a critical role in the body's immune response, particularly in regulating inflammation (Holbrook et al., 2019). In Whipple's disease, the bacterium Tropheryma whipplei triggers an abnormal immune response, leading to the activation of the TNF signaling pathway. This pathway contributes to the chronic inflammation seen in the disease, which damages the intestinal lining and other tissues. The persistent activation of TNF signaling can exacerbate symptoms such as abdominal pain, diarrhea, and systemic inflammation, as well as potentially contribute to the neurological symptoms associated with the disease (Parameswaran & Patial, 2010). The dysregulation of TNF and its associated pathways underscores the complexity of Whipple's disease and highlights the importance of targeting this pathway in potential therapeutic strategies.

COL2A1

Collagen, particularly the type II alpha 1 chain gene (COL2A1), plays a significant role in the pathology of Whipple's disease. Collagen is a primary component of the extracellular matrix (ECM) and is crucial for maintaining the structural integrity of tissues. In Whipple's disease, the expression of collagen-related genes like COL2A1 can be altered due to the chronic inflammation caused by Tropheryma whipplei. This can lead to excessive deposition of collagen and other ECM components, contributing to fibrosis and tissue remodeling (Di Stefano et al., 1998). Such changes are associated with the characteristic symptoms of Whipple's disease, including malabsorption and gastrointestinal distress, as the normal architecture of the intestinal lining is disrupted. Furthermore, the involvement of collagen in the ECM-receptor interaction and TNF signaling pathways suggests that it may play a role in the disease's progression by exacerbating inflammation and tissue damage.

In summary, the identification of collagen and ECM-receptor interactions as key factors in Whipple's disease opens up potential avenues for developing diagnostic biomarkers and therapeutic targets. For instance, collagen-related genes like COL2A1 could be further explored as biomarkers for early detection or progression of the disease.

A limitation of this research is that the study's analysis was based on only six samples (three affected and three controls), which may not be representative of the broader population or account for variability in the disease. Further, because we used bioinformatic datasets derived from microarray experiments conducted by other researchers, a limitation of this study is that the identified genes will need further investigation in a laboratory or clinical setting.

There are several avenues for future research. First, this research can be continued in a lab or clinical setting to validate the role of collagen (e.g., COL2A1) and other identified biomarkers in Whipple's disease. This involves testing these biomarkers in larger and more diverse patient cohorts to confirm their utility. Second, more research can be done to develop and refine diagnostic assays based on validated biomarkers. This could include creating blood tests or tissue assays to detect early signs of Whipple's disease or assess disease progression.

As for potential vaccinations in the future, identifying and validating specific antigens of T. whipplei that could be targeted by a vaccine. Research into the bacterium's surface proteins or secreted factors might reveal potential vaccine targets.

Conclusion

This study found that the ECM-receptor interaction and TNF signaling pathway are connected to Whipple's disease through altered gene expression, particularly involving high levels of collagen and other extracellular matrix components in patients suffering from Whipple's disease.

DISCUSSION

Healthcare Access and Utilization linked to Insurance Status

The significant association between insurance status and healthcare utilization aligns with existing literature highlighting the role of insurance in facilitating access to healthcare services. Insured individuals had higher odds of engaging in timely cholesterol screening and discussions about exercise with healthcare professionals [3]. However, the non-significant association between insurance status and discussions about weight or eating habits, when controlling for a usual source of care, suggests that factors such as accessibility and age play a role [1].

Usual Source of Care

The robust relationship between having a regular source of care and various healthcare utilization metrics underscores the importance of continuity in healthcare services. Individuals with a usual source of care had increased odds of physician visits, timely screenings for blood pressure and cholesterol, and discussions about lifestyle factors [4]. This finding aligns with the notion that consistent healthcare relationships contribute to more proactive preventive care.

Moderation by CVD Clinical Risk/Disease

The moderation analysis did not reveal significant differences in the relationships between insurance status, a usual source of care, and CVD prevention factors based on the presence or absence of CVD clinical risk/disease [4]. This suggests that the benefits of insurance and a regular source of care extend across different risk profiles, emphasizing their universal relevance in promoting cardiovascular health.

Implications for Public Health Policy

The studies' results underscore the importance of addressing healthcare access barriers to enhance cardiovascular health. Policies aimed at increasing insurance coverage and promoting the establishment of a regular source of care may contribute to better preventive care. Efforts to facilitate discussions about lifestyle factors within healthcare settings could further enhance the impact of these policies.

Limitations and Future Directions

While this paper provides valuable insights, some limitations should be acknowledged. Studies included with a cross-sectional design limits the establishment of causality. Additionally, any self-reported data, especially

regarding lifestyle behaviors, may be subject to bias. Future research could explore longitudinal data and incorporate objective measures to strengthen the evidence base.

Conclusion

In conclusion, this study contributes to the ongoing discourse on the interplay between healthcare access and cardiovascular health. The findings emphasize the need for holistic approaches in healthcare policies to address disparities and enhance preventive care, ultimately contributing to improved cardiovascular outcomes within diverse populations.

References

1. Antunes, C., & Singhal, M. (2023, July 4). Whipple Disease. StatPearls – NCBI Bookshelf.
2. Bai, J. C., Mazure, R. M., Vazquez, H., Niveloni, S. I., Smecuol, E., Pedreira, S., & Mauriño, E. (2004). Whipple's disease. Clinical Gastroenterology and Hepatology, 2(10), 849–860. https://doi.org/10.1016/S1542-3565(04)00387-8
3. Bakkali, N., Fenollar, F., Biswas, S., Rolain, J., & Raoult, D. (2008). Acquired resistance to trimethoprim-sulfamethoxazole during Whipple disease and expression of the causative target gene. The Journal of Infectious Diseases, 198(1), 101–108. https://doi.org/10.1086/588706
4. Boumaza, A., Azzouz, E. B., Arrindell, J., Lepidi, H., Mezouar, S., & Desnues, B. (2022). Whipple's disease and Tropheryma whipplei infections: From bench to bedside. The Lancet Infectious Diseases, 22(10), e280–e291. https://doi.org/10.1016/s1473-3099(22)00128-1
5. Büttner, P., Ueberham, L., Shoemaker, M. B., Roden, D. M., Dinov, B., Hindricks, G., Bollmann, A., & Husser, D. (2018). Identification of central regulators of calcium signaling and ECM–Receptor interaction genetically associated with the progression and recurrence of atrial fibrillation. Frontiers in Genetics, 9, 162. https://doi.org/10.3389/fgene.2018.00162
6. Di Stefano, M., Jorizzo, R. A., Brusco, G., Cecchetti, L., Sciarra, G., Loperfido, S., Brandi, G., Gasbarrini, G., & Corazza, G. R. (1998). Bone mass and metabolism in Whipple's disease: The role of hypogonadism. Scandinavian Journal of Gastroenterology, 33(11), 1180–1185. https://doi.org/10.1080/00365529850172548

7. Dolmans, R. a. V., Boel, C. H. E., Lacle, M. M., & Kusters, J. G. (2017). Clinical manifestations, treatment, and diagnosis of Tropheryma whipplei infections. Clinical Microbiology Reviews, 30(2), 529–555. https://doi.org/10.1128/cmr.00033-16
8. Friebel, J., Schinnerling, K., Geelhaar-Karsch, A., Allers, K., Schneider, T., & Moos, V. (2022). Intestinal barrier dysfunction mediates Whipple's disease immune reconstitution inflammatory syndrome (IRIS). Immunity Inflammation and Disease, 10(5), 622. https://doi.org/10.1002/iid3.622
9. GEO Accession viewer. (2010). https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE20209
10. GEO2R. (n.d.). NCBI – U.S. National Library of Medicine. https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html
11. Holbrook, J., Lara-Reyna, S., Jarosz-Griffiths, H., & McDermott, M. F. (2019). Tumour necrosis factor signalling in health and disease. F1000Research, 8, 111. https://doi.org/10.12688/f1000research.17023.1
12. Jarosz-Griffiths, H. H., Holbrook, J., Lara-Reyna, S., & McDermott, M. F. (2019). TNF receptor signalling in autoinflammatory diseases. International Immunology, 31(10), 639–648. https://doi.org/10.1093/intimm/dxz024
13. Johnson, S. (2021, July 12). Whipple's Disease. Healthline. https://www.healthline.com/health/whipples-disease
14. Melas, N., Amin, R., Gyllemark, P., Younes, A. H., & Almer, S. (2021). Whipple's disease: The great masquerader—a high level of suspicion is the key to diagnosis. BMC Gastroenterology, 21(1), 664. https://doi.org/10.1186/s12876-021-01664-1
15. Parameswaran, N., & Patial, S. (2010). Tumor necrosis factor-α signaling in macrophages. Critical Reviews in Eukaryotic Gene Expression, 20(2), 87–103. https://doi.org/10.1615/critreveukargeneexpr.v20.i2.10
16. ShinyGO 0.80. (n.d.). http://bioinformatics.sdstate.edu/go/
17. Soles, A., Selimovic, A., Sbrocco, K., Ghannoum, F., Hamel, K., Moncada, E. L., Gilliat, S., & Cvetanovic, M. (2023). Extracellular matrix regulation in physiology and in brain disease. International Journal of Molecular Sciences, 24(8), 7049. https://doi.org/10.3390/ijms24087049

18. Song, X., Duan, R., Duan, L., & Wei, L. (2023). Current knowledge of the immune reconstitution inflammatory syndrome in Whipple disease: A review. Frontiers in Immunology, 14, 1265414. https://doi.org/10.3389/fimmu.2023.1265414
19. Wang, Y., Zhang, H., Chen, Q., Jiao, F., Shi, C., Pei, M., Lv, J., Zhang, H., Wang, L., & Gong, Z. (2020). TNF-α/HMGB1 inflammation signalling pathway regulates pyroptosis during liver failure and acute kidney injury. Cell Proliferation, 53(6), 12829. https://doi.org/10.1111/cpr.12829.

Karthik Peddu

Using Bioinformatics to Study the Change in Gene Expression During Eye Infections Caused by Bacteria

ABSTRACT

Eye complications have always been a problem in the healthcare field. The biggest of these complications has been eye infections. However as bacteria starts to build up antibiotic resistance, old treatments will one day stop working. Therefore, along with the continued use of antibiotic treatment, new solutions can be created to target the genes that are being affected by eye infections so that the genes are being expressed normally. However, there is a lack of research on what genes should be targeted. Due to this research gap, this research study was focused on identifying genes that are regulated differently between a cornea with bacterial eye infection and normal cornea. This was done using multiple bioinformatics tools and databases such as GEO, NCBI, SR Plot, GO and KEGG to visualize, identify and analyze these genes that are being differentially expressed. Results show that the genes with the label Rac, CXCL5, MMP5, and IL1β are all differentially expressed between the normal and infected cornea. Additionally all these genes are associated with the immune system and immune response. Therefore, these 4 genes are potential biomarkers and that can help contribute to the creation of solutions and treatment that will be effective against eye infections.

KEY WORDS: Bacterial eye infections, Gene expression, Bioinformatics, GEO, NCBI, SR Plot, GO, KEGG

INTRODUCTION

Currently there are many eye complications that many people experience due to bacteria (Watson et al, 2018). These include eye infections such as Conjunctivitis, Keratitis, Endophthalmitis (Holland, 2017). No matter what the severity or pain associated with each eye infection, one thing that they have in common is that they can occur from bacterial infection (Holland, 2017).

Bacteria are prokaryotic microorganisms that are able to survive at many places around the world. Some bacteria are good and do well for the earth

and help humans, however there are also bacteria that can cause harm. Due to this bacteria has a large impact on public health (Doron, 2008). Bacteria are transmitted to humans through air, water, food, or living vectors. The principal modes of transmission of bacterial infection are contact, airborne, droplet, vectors, and vehicular. Doron, 2008). Due to its ability to be transferred easily in humans bacteria are able to get into the eye leading to eye infections. Additionally one of the places that are most vulnerable to these bacterial infections is Africa. The main bacterial eye infection found in Africa is Trachoma. Trachoma is a bacterial infection that affects the eyes and is caused by the bacterium Chlamydia trachomatis (Ferede et al., 2021). More specifically Africa is the most affected continent with 27.8 million people with Trachoma. Nearly 50% of the global burden of active trachoma is highly distributed in three countries: Ethiopia, Malawi, and Nigeria which are all a part of Africa (Ferede et al., 2021).

In addition to eye infections in Africa, bacterial eye infections rates are increasing all over the world. Therefore, there is a great need to investigate how to slow down these bacterial infections that have risen. One way to investigate solutions to these bacterial eye infections is studying gene regulation during a bacterial eye infection. Studying gene regulation during bacterial eye infection has been used before and has been effective at identifying gene regulated differences between an infected and non infected eye. (Diehn et al., 2005). However, a challenge of studying gene expression in the eye is the technical obstacles associated with extracting sufficient quantities of high quality RNA from the tissues. Due to this, not every eye infection's effect on gene regulation has been studied (Diehn et al., 2005).

Therefore, to fill this gap, the goal of this research was to understand how gene regulation changes between each eye infection. Also, the goal of this research was to identify and uncover how differences in gene expression can be used to create more solutions to combat eye infection.

The question this research study asks is, how does gene regulation change in an organism after having an eye infection due to bacteria? More specifically, are genes upregulated (highly expressed) or downregulated due to eye infections? I hypothesized that there will be differences in gene expression in each eye infection and that gene expression does not always

go down or up during an eye infection. Results from this study can be used by scientists to create more solutions to combat eye infections.

METHODS

Gene Expression Microarray Data Collection

To investigate the question of gene expression during an eye infection caused by bacteria, this research utilized bioinformatics tools and databases. One of these databases is Gene Expression Omnibus (GEO). GEO is a database in NCBI that contains high-throughput molecular data, primarily focused on gene expression generated by DNA microarray technology (Barrett & Edgar, 2006). NCBI is an online resource that facilitates the use of databases and software and performs research on advanced methods of computer-based information processing for analyzing the structure and function of biologically important molecules. It is a popular database and used frequently by many researchers and scientists (Wojtkiewicz et al., 2016). However, as both NCBI and GEO focus on biological data, GEO is a part of NCBI and is focused more on experiments between two groups (Clough & Barret, 2016) by using pre-program AI and ML algorithms in the R programming language (Supplementary Data: R script for Dataset GSE58291). GEO stores a billion individual gene expression measurements, collected from over 100 organisms, submitted by over 1500 laboratories, addressing a wide range of biological phenomena (Barrett & Edgar, 2006).

In order to utilize GEO, keywords were used in order to find databases related to gene expression during eye infections caused by bacteria. These keywords include: Eye, Bacteria, Retina, Keratitis, conjunctivitis. From the use of the keywords, GEO was able to analyze and find microarray datasets that highlight bacteria's effect on gene regulation during eye infections. From here the GSE58291 was discovered. Pathways involved in the pathophysiology of this disease were studied in the experiment by comparing the human transcriptome profile of tissue from culture-proven bacterial and fungal keratitis (n=7 and n=8 respectively) with normal non-infected cadaveric corneal tissue (C, n=12) using Illumina HT12 v4 microarrays (Chidambaram et al., 2017).

Analyzing Difference in Gene Expression using GEO2R Bioinformatics Tool

GEO2R allows users to identify and visualize differential gene expression based on the R- based GEO data analysis. It is a part of GEO and highlights genes that show significant differences in expression between two groups. It uses R programming language algorithms and includes a number of statistical categories, such as P-values, t-statistics, fold change, and gene annotations in order to highlight differences in gene expression (Sofi et al. 2022). The R programming script used to generate the data in this research can be found here: R script for Dataset GSE58291

Based on the GEO2R analysis of the GSE58291 dataset , the genes with the most significant gene expression difference were displayed as differentially expressed genes (DEGs). DEGs are genes that show significant changes in their activity levels under different experimental conditions (Anjum et al., 2016). These conditions can include different tissues, treatments, diseases, stages of development, or environmental factors (Anjum et al., 2016). A gene expression was considered significant if the P value was less than 0.05. From here the top 20 upregulated expressed genes and 20 down regulated expressed genes were chosen (Supplementary Data: Top 20 Downregulated and Upregulated Genes) for further analysis.

Investigating Potential Functions and Biological Pathways of the identified Differentially Expressed Genes Using SR Plot Bioinformatics Tool

The chosen 20 upregulated expressed genes and 20 down regulated expressed genes were further analyzed using a bioinformatics tool called SRplot (http://www.bioinformatics.com.cn/srplot). SRplot is used to visualize multiple types of biological and molecular components that occur in many different types of organisms (Tang et al., 2023). SRplot is able to do this by carrying out Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. GO works by visualizing many processes by creating graphs and other visualizations. These processes include biological process (BP), cellular component (CC), and molecular function (MF) (Zhang et al., 2014). BP highlights many cellular functions that are necessary to normal behavior. CC highlights and analyzes the various components of the cell and how they play a part in a location of a gene. Lastly MF focuses on activity at the molecular level such as structural molecular activity (Zhang et al., 2014)

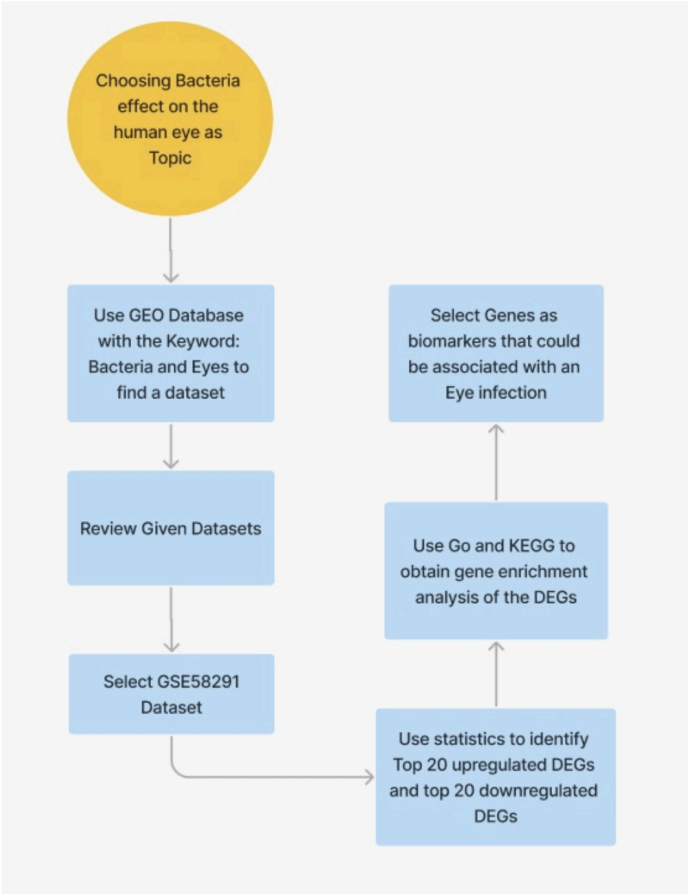


Figure 1: Method Workflow. The use of Bioinformatic tools and databases namely, GEO2r, SRplot, GO, and KEGG were used. A total of 40 differentially expressed genes (DEGs) were analyzed.

RESULTS

Identification of Differentially Expressed Genes

Differentially expressed genes (DEGs) are genes that are different expressed between two experimental conditions (Anjum et al, 2016). In this study, the DEGs that were analyzed were 20 upregulated genes and 20 down regulated genes. The DEGs were obtained from infected cornea. In order for a DEG to be identified, there needs to be a statistical significant difference or change in read counts or expression levels between the two experimental conditions. (Anjum et al., 2016).These DEGs and their expression levels can be studied under different environments, conditions, treatments, and stages for further analysis and interpretation(Anjum et al., 2016). Additionally these DEGs can be up-regulated (exhibiting increased expression) or down-regulated

(exhibiting decreased expression) between two experimental groups. Identifying DEGs is important in gene expression analysis. It provides valuable insights into the molecular mechanisms that eventually lead to many biological processes and conditions (Anjum et al., 2016).

In this study, the DEG's were obtained by a study done by Tien et al (2019) by using GEO2R (A part of NCBI (https://www.ncbi.nlm.nih.gov/geo/)) to analyze the dataset of the study: GSE58291 dataset. The study contains two experimental groups and a control group. One experimental group had the cornea with a bacteria infection, while the other experimental group had the cornea with a fungal infection. The control group had a healthy, non-infected cornea. Due to the fact that the study focuses on the difference in gene expression due to presence of bacteria or not, only the control and the experimental group that had a bacterial infection were selected for the study. The two groups were compared and the DEG's were identified (Figure 2).

Figure 2a is a volcano plot that shows genes that were upregulated and downregulated when the experimental group that contained a bacterial infection was compared to the control group. The upregulated genes are the red dots on the plot, the downregulated genes are the blue dots on the plot, and the genes that show no statistical significant evidence in change of gene regulation are the black dots. In order for a gene to have statistical significant evidence of change in gene regulation the p-value of the change in gene regulation between the 2 groups has to be less than 0.05. Figure 2b shows all total genes that were analyzed by GEO2R. The number inside the green oval highlights the genes that show significant difference in expression between the control group and the experimental group that contains a bacterial infection. Figure 3 shows a UMAP plot. This plot shows how each group differentiates from each other in regards of approximation and projection.

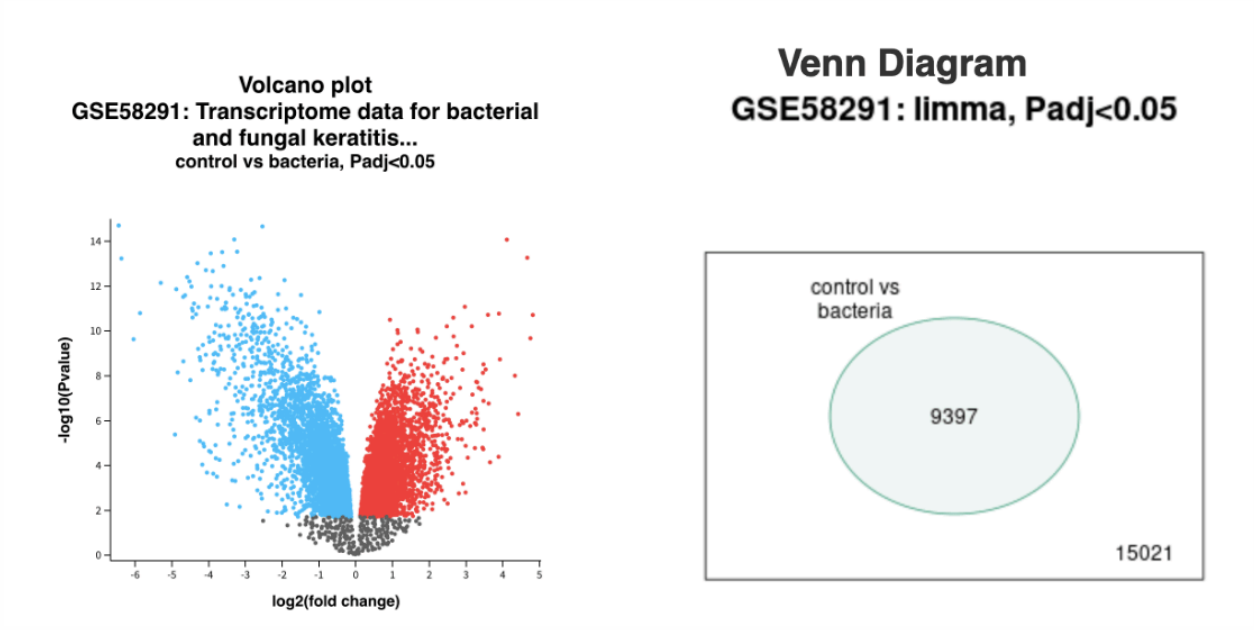


Figure 2: Difference in Gene expression between a control and bacteria experimental group. (a) Volcano plot of the genes between control group and bacteria experimental: Red dots represent genes that were upregulated, while blue dots represent genes that were downregulated. Black represents no significant difference in gene expression between control and bacterial infected groups. (b) Venn Diagram showing the total number of genes identified in this study (15021) when comparing between the control group versus bacteria experimental group (9397).

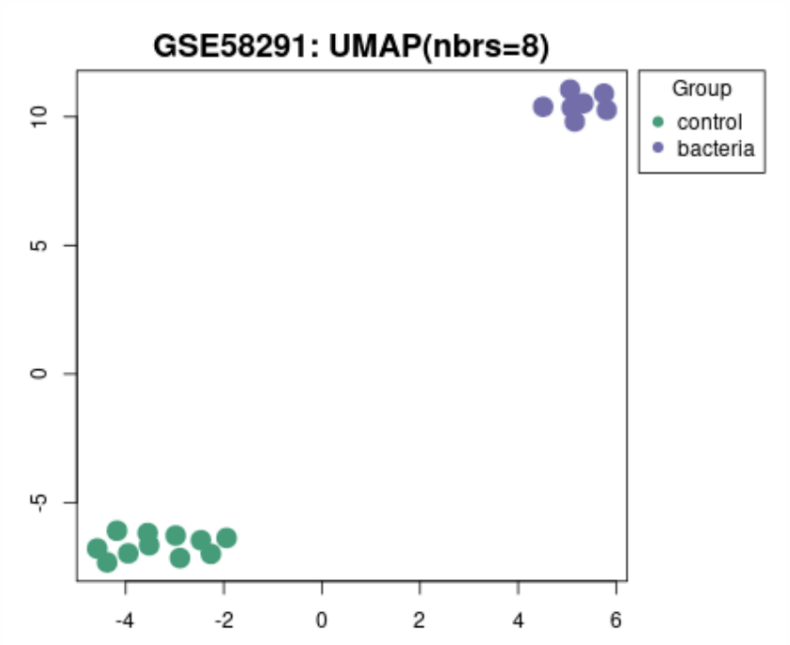


Figure 3: Association between different samples. How samples in the control group and samples of the bacterial experimental group align with other groups using the UMAP plot.

Using SRplot Bioinformatics Tool to Analyze DEGs

Using SRplot (http://www.bioinformatics.com.cn/srplot) the 20 upregulated and downregulated genes that had been selected were analyzed. This is due to the KEGG pathway enrichment analysis and GO function. Figure 4 and 5 were obtained through KEGG. Figure 4 highlights a Chemokine Signaling Pathway that was obtained from KEGG. Chemokine is responsible for signaling the migration of cells, specifically white blood cells which help with an immune response (Hughes & Nibbs, 2018).

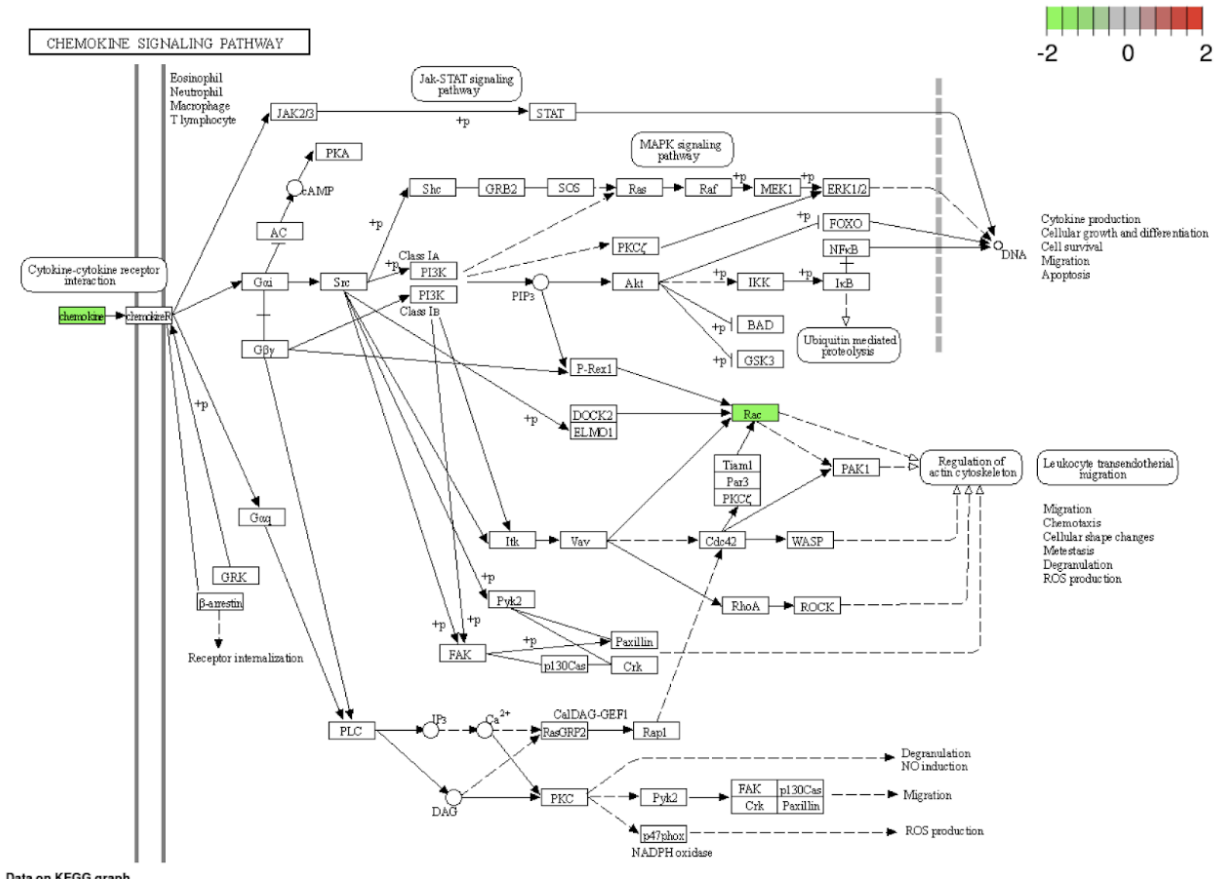


Figure 4: Chemokine Signaling Pathway. Green boxes represent areas and genes that are down regulated.

Results in figure 5 highlight an IL-17 Pathway. This Pathway is a key component in the immune system, inflammation, and T cell activation (Zenobia & Hajishengallis,2016). Figure 6 was obtained through the GO Analysis. The shows three ontologies and the respective enrichment score. The orange bars represent Biological Processes (BP) such as neutrophil chemotaxis and neutrophil migration. The green Bar represents cellular components (CP) such as tertiary granule and tertiary granule membrane. The dark blue bars represent molecular functions (MF) such as G Protein-coupled receptor binding and receptor ligand activity.

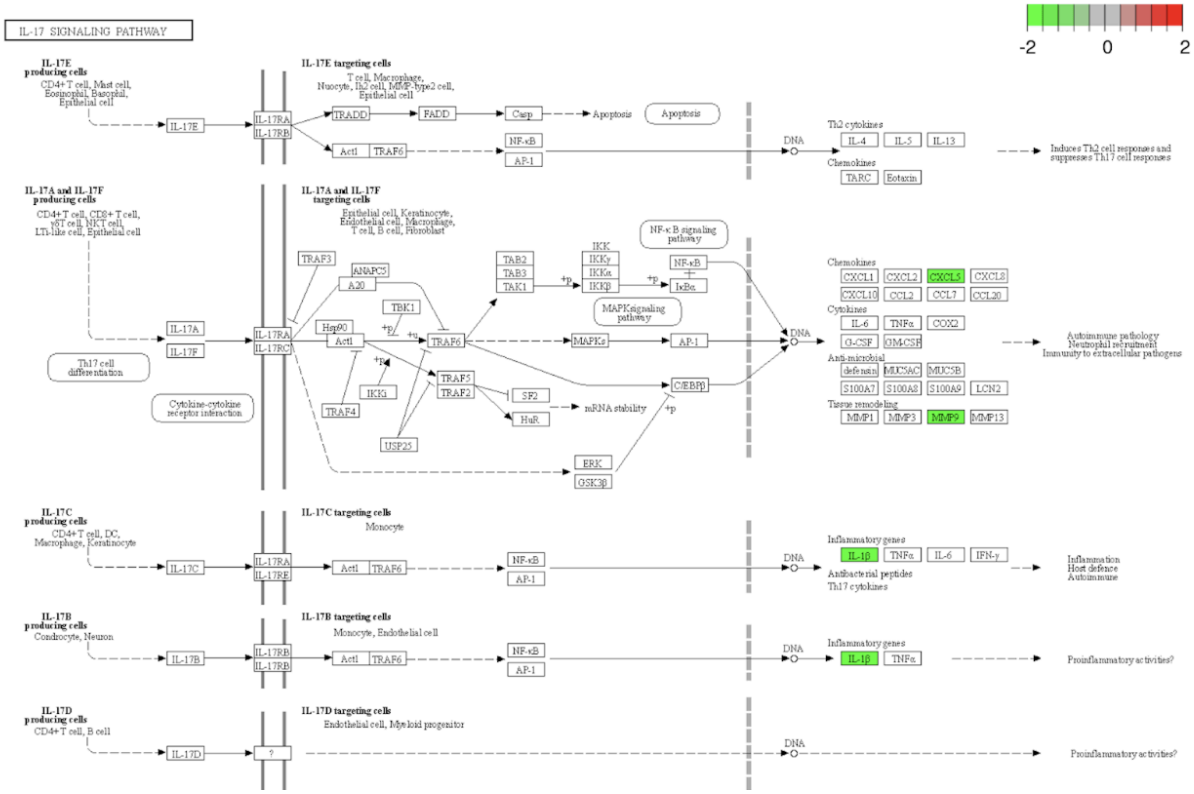


Figure 5: an IL-17 Pathway. Green boxes represent areas and genes that are down regulated.

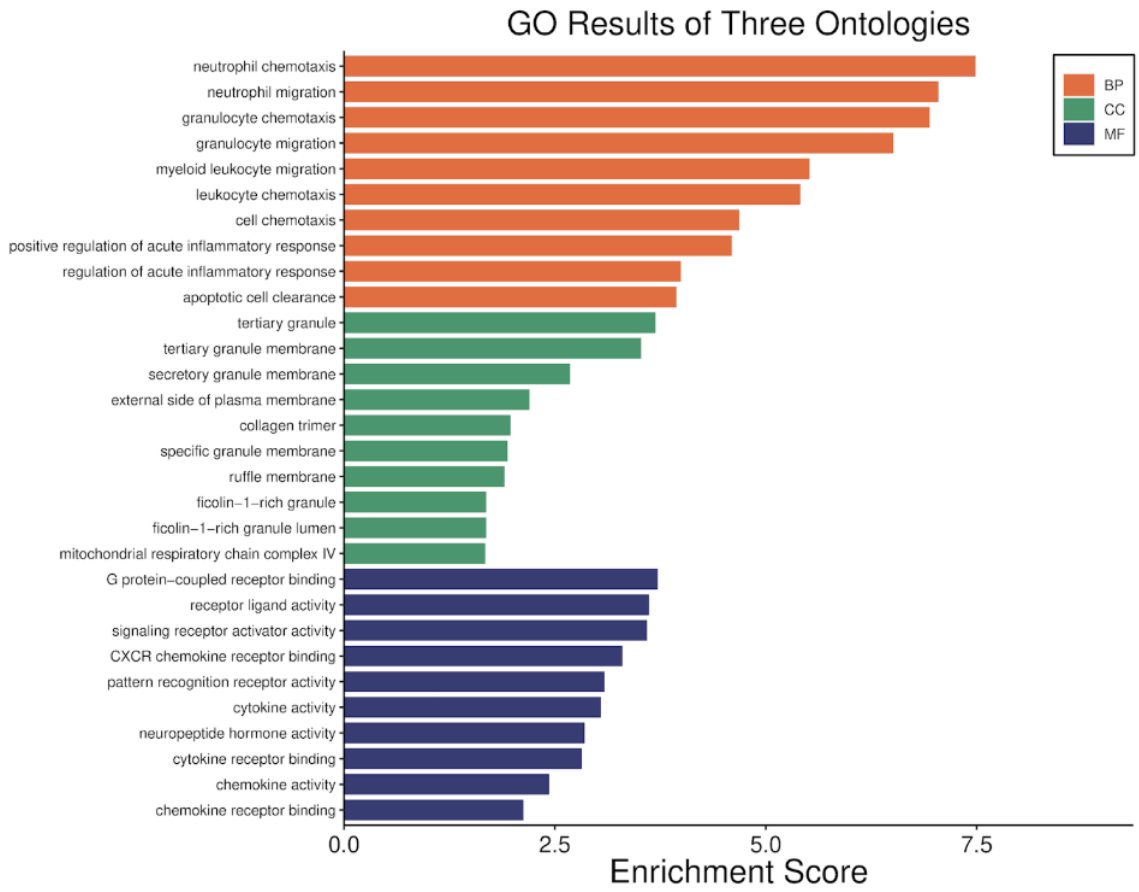


Figure 6: GO results of three Ontologies. Orange indicates biological processes (BP), Green indicates cellular components (CC), and blue indicates molecular functions (mf). As seen in the enrichment schore, the top DEGs are most enriched in the biological processes (BP) gene ontology term.

Table 1: Summary of Identified Downregulated Genes Associated with Immune System

DOWN REGULATED GENES ASSOCIATED WITH THE IMMUNE SYSTEM
Rac
CXCL5
MMP9
IL-1B

DISCUSSION

Eye Infections are an ongoing problem in primary care. Most of these infections are caused by bacteria and are usually solved with antibiotic treatment (Prescr, 2018). However as more and more antibiotic treatment is used so does the chance of antibiotic resistance (Salam et al., 2023). Therefore the goal of this research study is to identify genes that are affected during an eye infection so that new solutions can target and influence the genes to act normally, leading to decrease in eye infections in people.

I Identified differentially expressed genes (DEGS) from GEO2R, which analyzed a NCBI study. Figure 2A highlights that there are many genes that are upregulated as well as downregulated. Figure 2B also highlights the difference in gene regulation by showing that 9397 genes were differentially expressed between the control group and the bacteria experimental group. Additionally GEO2R showed the UMAP plot. The UMAP plot highlights the gene similarity between groups. More specifically It highlights how the two groups in the study are different from each other but the samples in each group are similar (FIgure 3) Statistics was then used to choose the top 20 unregulated genes and top 20 downregulated genes from all the DEGs. Using the SRplot online tool, the 20 upregulated genes and 20 down regulated genes were examined through the GO and KEGG analysis. The KEGG analysis of the 40 combined DEGs showed two signaling pathways. The first signaling pathway is the Chemokine signaling pathway. The Chemokine signaling pathway is a part of the immune system and plays a part in the transportation of white blood cells (Hughes & Nibbs. 2018) KEGG highlights that the gene with the label Rac is downregulated when the cornea has a bacterial infection. (Figure 5) This downregulation of the gene indicates that the whole pathway is not occurring at the rate intended leading to decreased immune response. The second signaling pathway is the IL1β Signaling pathway. The IL1β gene is also associated with immune response and is usually transcribed to recognize pathogen-associated molecular patterns (Yin et al. 2024) KEGG highlights that the gene labeled CXCL5, MMP5, and IL1β are all differentially expressed between the normal and infected cornea. (Figure 6) Due to differences in gene regulation there is also a difference in this pathway’s ability to be involved in immune response

and recognizing pathogens. Additionally GO analyzed the 40 combined DEGs. Based on GOs analysis it created a plot that highlights the association between the genes and either biological processes, cellular components, or molecular functions. Based on the graph it highlights how the 40 DEGS have the most association with biological processes due to the higher enrichment scores of the orange boxes.(Figure 7). Therefore these difference in genes cause biological processes to be altered during an eye infection.

Due to the fact that this study uses bioinformatics tools and analyzes past research studies there are some limitations. Firstly, the dataset studied in GEO2R only analyzed two types of bacteria during an eye infection. Therefore it is uncertain how genes are regulated with different types of bacteria. Therefore a future study could be analyzing different bacterial microorganisms in the eye as well as in a laboratory or clinical study instead of using a bioinformatics database.

Conclusion

Based on this research study 4 genes were indicated that potentially have association to an eye infection. These genes are Rac, CXCL5, MMP5, and IL1β, which were all downregulated. Before the research study we hypothesized there will be differences in gene expression in each eye infection and that gene expression does not always go down or up during an eye infection. I failed to reject my hypothesis based on figure 2a due to the fact that it shows that some genes were downregulated while others were upregulated. However, the genes that were associated with the pathways analyzed in KEGG were all genes that were downregulated. The findings of this study can help other healthcare professionals and other researchers to further study the relationship between gene regulation and bacterial eye infections. From here new treatment can be potentially developed in order to combat this ongoing problem of bacterial infections especially in the human eye.

SUPPLEMENTARY DATA

Supplementary Data 1: Using R Script to generate DEG results using the bioinformatics tool GEO2R
R script for Dataset GSE58291

Supplementary Data 2: Google Sheet with DEGS with a P value < 0.05 and their logFC values
Top 20 Downregulated Genes
Top 20 Upregulated Genes
Additional GO and KEGG illustrations

References

1. Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A., & Rai, A. (2016). Identification of Differentially Expressed Genes in RNA-seq Data of Arabidopsis thaliana: A Compound Distribution Approach. Journal of computational biology : a journal of computational molecular cell biology, 23(4), 239–247. https://doi.org/10.1089/cmb.2015.0205
2. Barett, T., & Edgar, R. (2006, August 29). [19] gene expression omnibus: Microarray data storage, submission, retrieval, and analysis. Gene Expression Omnib us: Microarray Data Storage, Submission, Retrieval, and Analysis.
3. Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus Database. Methods in molecular biology (Clifton, N.J.), 1418, 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5
4. Chidambaram, J. D., Kannambath, S., Srikanthi, P., Shah, M., Lalitha, P., Elakkiya, S., Bauer, J., Prajna, N. V., Holland, M. J., & Burton, M. J. (2017). Persistence of Innate Immune Pathways in Late Stage Human Bacterial and Fungal Keratitis: Results from a Comparative Transcriptome Analysis. Frontiers in cellular and infection microbiology, 7, 193. https://doi.org/10.3389/fcimb.2017.00193
5. Diehn, J. J., Diehn, M., Marmor, M. F., & Brown, P. O. (2005). Differential gene expression in anatomical compartments of the human eye. Genome biology, 6(9), R74. https://doi.org/10.1186/gb-2005-6-9-r74
6. Doron, S., & Gorbach, S. L. (2008). Bacterial Infections: Overview. International Encyclopedia of Public Health, 273–282. https://doi.org/10.1016/B978-012373960-5.00596-7
7. Ferede, A. T., Dadi, A. F., Tariku, A., & Adane, A. A. (2017). Prevalence and determinants of active trachoma among preschool-aged children in Dembia District, Northwest Ethiopia. Infectious diseases of poverty, 6(1), 128. https://doi.org/10.1186/s40249-017-0345-8
8. Hughes, C. E., & Nibbs, R. J. B. (2018). A guide to chemokines and their receptors. The FEBS journal, 285(16), 2944–2971. https://doi.org/10.1111/febs.14466

9. Jewell, T. (2023, February 13). Infected eye: 8 common causes. Healthline. https://www.healthline.com/health/infected-eye
10. Salam, M. A., Al-Amin, M. Y., Salam, M. T., Pawar, J. S., Akhter, N., Rabaan, A. A., & Alqumber, M. A. A. (2023). Antimicrobial Resistance: A Growing Serious Threat for Global Public Health. Healthcare (Basel, Switzerland), 11(13), 1946. https://doi.org/10.3390/healthcare11131946
11. Sofi, M. Y., Shafi, A., & Masoodi, K. Z. (2022). Gene expression omnibus. Chapter 9 – Plant genomic data and resources at NCBI. https://www.sciencedirect.com/science/article/abs/pii/B9780323911283000019
12. Tian, R., Zou, H., Wang, L., Liu, L., Song, M., & Zhang, H. (2020). Analysis of differentially expressed genes in bacterial and fungal keratitis. Indian journal of ophthalmology, 68(1), 39–46. https://doi.org/10.4103/ijo.IJO_65_19
13. Watson, S., Cabrera-Aguas, M., & Khoo, P. (2018). Common eye infections. Australian prescriber, 41(3), 67–72. https://doi.org/10.18773/austprescr.2018.016
14. Wojtkiewicz, M., Wiederin, J., & Ciborowski, P. (2006). National Center for Biotechnology Information. Gene Expression Omnibus: Microarray Data Storage, Submission, Retrieval, and Analysis.
15. Wojtkiewicz, M., Wiederin, J., & Ciborowski, P. (2016, April 1). Proteomic database search and analytical quantification for mass spectrometry. Proteomic Profiling and Analytical Chemistry (Second Edition). https://www.sciencedirect.com/science/article/abs/pii/B9780444636881000112
16. Yin, W., Godbout, J. P., & Sheridan, J. F. (2024, August 25). Interleukin-1 beta in psychosocial stress. Chapter 7 – Interleukin-1 beta in psychosocial stress. https://www.sciencedirect.com/science/article/abs/pii/B9780128175583000007
17. Zenobia, C., & Hajishengallis, G. (2015). Basic biology and role of interleukin-17 in immunity and inflammation. Periodontology 2000, 69(1), 142–159. https://doi.org/10.1111/prd.12083
18. Zhang, J., Xing, Z., Ma, M., Wang, N., Cai, Y. D., Chen, L., & Xu, X. (2014). Gene ontology and KEGG enrichment analyses of genes

related to age-related macular degeneration. BioMed research international, 2014, 450386. https://doi.org/10.1155/2014/450386

Nikhil Pesati

Using Bioinformatics to Locate Genes Affected by Staphylococcus aureus and their Effects on Sepsis

ABSTRACT

As the largest and most vulnerable organ system, the skin is teeming with bacteria that are usually commensal. However, there are some that are opportunistic pathogens like Staphylococcus aureus which find a way to invade the skin and go into the bloodstream. The harms brought by S. aureus include conditions like sepsis, which is an inflammatory response that is unregulated, thus damaging organs inside the body and leading to death. While there have been experiments trying to identify how to neutralize the effect of S. aureus, there is still a lack of clarity regarding the genes that are being dysregulated by S. aureus in the human blood that cause sepsis. Therefore, to fill this gap, this research aimed to find genes that can be used in gene therapy to stop people infected by S. aureus from dying of sepsis. This research used bioinformatics tools called GEO2R to find genes that are active in sepsis. We first identified the top 30 genes that are expressed differently between infected patients and healthy individuals. Then we used another bioinformatics tool, SR Plot, to analyze these genes via gene ontology (GO) and KEGG bioinformatics databases. This analysis found two key genes, UTY and KDM5D, that affect sepsis by influencing histone demethylase activity which is a factor that is closely related with sepsis inflammatory function and behavior. With more testing and research, these genes might help develop gene therapy to prevent sepsis deaths caused by S. aureus.

INTRODUCTION

This research is addressing the potential for a decrease in deaths by sepsis for human beings. Sepsis is a syndrome that causes organ failure due to an incorrect response to an infection (Raymund B Dantes, Lauren Epstein, 2018). In the United States alone, sepsis causes about 265 thousand deaths a year. (Raymund B Dantes, Lauren Epstein, 2018) Staphylococcus aureus is one of the main causes of sepsis for humans (McAdow et. al, 2011). S. aureus normally resides on the human skin and is known to be an opportunistic pathogen as it causes many different types of infections (Parlet

et. al, 2019). The bacteria causes a lot of infections as it frequently invades the skin of humans and goes into our bloodstream (Thomer et. al, 2016).

When S. aureus enters the bloodstream, it triggers an inflammatory response (Kwiecinski and Horswill, 2020). This inflammatory response can lead to sepsis as the excessive spread of inflammation throughout the body causes damage to organs in the body (Kwiecinski and Horswill, 2020). Annually, there are more than 35 thousand causes of sepsis related to the invasion of S. aureus (Thomer et. al, 2016). Although it is known that S. aureus leads to sepsis, scientists do not know what genes are related to the pathogenesis of sepsis and thus cannot create drugs that cure or inhibit sepsis (Li. et al, 2021). Therefore, this current research was conducted to fill this gap.

Specifically, the goal of this research study was to find out if there are key genes that S. aureus dysregulates in the human blood and if so what the key genes are. By doing this, scientists can gather a list of genes that should be targeted in future drugs to prevent sepsis such as gene therapy. Gene therapy is a medical technique that stops a disease by correcting the genetic issue causing it (Scheler and Krebsbach, 2009). To do this, scientists either add a new gene to the cells to fight the disease or replace the faulty gene with a healthy copy (Gonçalves and Paiva, 2017). Therefore, genes identified in this research study can be potentially used for gene therapy.

In order to conduct this research, different bioinformatics tools and databases were used. Specifically, I used gene expression data from public sources that is commonly used by research scientists today (Rung and Brazma, 2013). Analyzing gene expression using advanced methods like microarrays is important because they have promising uses in early diagnosis, finding potential cures or predicting outcomes for many diseases (Kulasingam and Diamandis, 2008). Therefore, in my research, I identified and used a Genome Expression Omnibus (GEO) dataset GSE237960 that used a microarray method tested the effects of S. aureus and E. coli on human blood as a way to figure out which genes are being dysregulated. These scientists were also trying to neutralize sepsis as these two bacteria are major causes of sepsis.

For my study I therefore used several bioinformatics tools namely, GEO2R (http://www.ncbi.nlm.nih.gov/geo/geo2r), SR Plot (http://www.bioinformatics.com.cn/srplot), to see if there are genes that could be the real cause of the infections by identifying which genes are expressed differentially between infected patients and healthy individuals and finally determine their potential functions and biological pathways using gene ontology (GO) and KEGG bioinformatics databases.

METHODS

Collecting Gene Expression Microarray Data from GEO2R Bioinformatics Tool

In order to investigate and identify the genes which through dysregulation by S. aureus play a role in a human being infected by sepsis, this research used a variety of bioinformatics tools. The overall process by which the research was conducted is shown in Figure 1 for clarity. In order to find a dataset, I sifted through the Gene Expression Omnibus (GEO) bioinformatics tool that is part of the NCBI website (Edgar et al., 2002). The National Center for Biotechnology Information (NCBI) is a website that contains a plethora of genomic data in order to allow researchers to access and conduct experiments with (Barrett et. al, 2009). NCBI's main benefit is the fact that it contains analysis tools which allow researchers to not only obtain but also analyze the data that they obtained (Edgar et. al, 2002). Inside the GEO, there is one such tool called GEO2R which was used for the research. GEO2R uses algorithms to highlight the differentially-expressed genes (DEGs) between two groups that a user defined from the samples in the dataset (Clough and Barrett, 2016). DEGs are genes that have a difference in expression levels based on the state of the sample (Anjum et al, 2016).

In order to find a dataset that contained genes affected by S. aureus, I conducted a keyword search around S. aureus and sepsis. From this search, I got a dataset, GSE237960, that contains samples of unaffected human whole-blood (n = 4) as well as samples of human blood that was infected with heat-killed S. aureus (n = 3). This dataset, which is the GSE237960 dataset, was conducting an experiment that focused on the different signaling effects of S. aureus and E. coli in healthy and unaffected whole blood and the blood that is infected with heat-killed S. aureus. I excluded E. coli when using the GEO2R bioinformatic tool, as it is unrelated to the research that is being conducted.

Identifying DEGs that are statistically significant

Upon use of GEO2R, the difference between the S. aureus infected human blood and the unaffected human blood from the dataset GSE237960 was generated using pre-programmed AI and ML algorithms using the R programming language (Supplementary Data: GEO2R Analysis R-script). The results are shown as differentially expressed genes (DEGs) in the volcano plot of Figure 2 with some genes upregulated and some genes downregulated. For this study, we decided to consider differentially expressed genes (DEG) that had a p-value < 0.05 and $|\log FC| \geq 1$ as statistically significant (Supplementary Results: Top 30 DEGs).

Using SR Plot Bioinformatics tool to determine functions and biological pathways of selected DEGs

In order to analyze the results that were obtained for key genes related to sepsis, the top 30 DEGs (Supplementary Results: Top 30 DEGs) were picked and imported into the SR Plot bioinformatics tool (https://www.bioinformatics.com.cn/srplot). SR Plot is a bioinformatics tool that contains a lot of algorithms that allow users to visualize and analyze inputted biological data through graphs and other outputs (Tang et. al, 2023). The two analyses that I performed with SR Plot are called Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analysis. GO contains ontologies, a set of vocabularies, that are used to describe biological process (BP), cellular component (CC), and molecular function (MF) (Harris et. al, 2004). KEGG is a database used for the analysis of genomic sequences and to find biological pathways that genes might be involved in (Kanehisa and Goto, 2000).

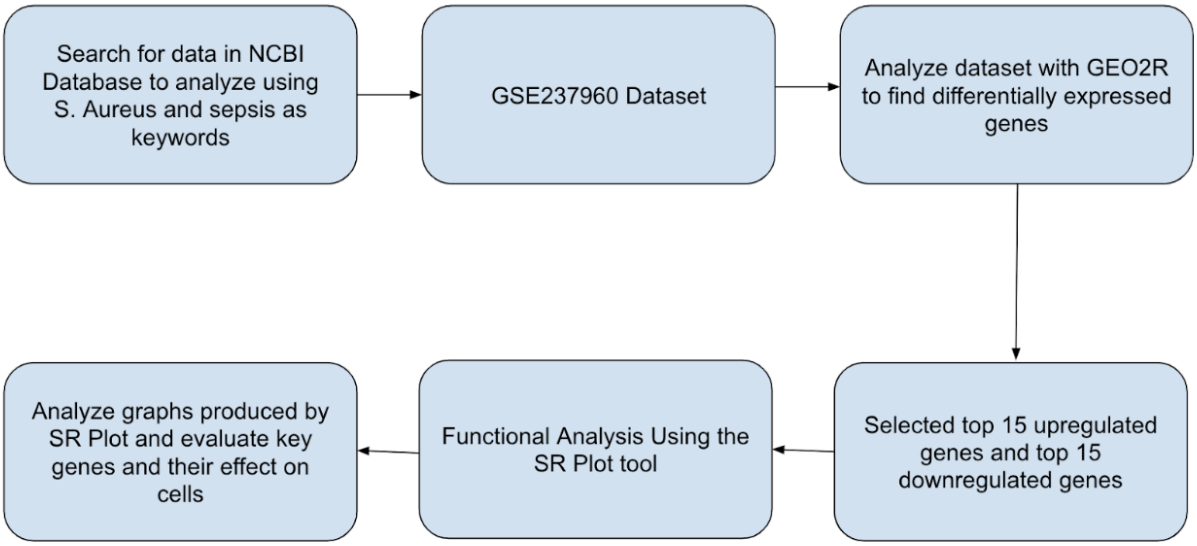


Figure 1: Methodology Flowchart: Overview of the methodology used in this bioinformatics study

RESULTS

Identification of top DEGs

In this study, differentially-expressed genes (DEGs) were analyzed to figure out which genes play a role in promoting and creating sepsis. DEGs are genes that have a statistically significant change in expression level when they are put in different experimental conditions (Chen et. al, 2021). The aforementioned different experimental conditions that DEGs change in are, in this research, the healthy and unaffected whole blood and the blood that is infected with heat-killed S. aureus. The DEGs that are analyzed are either upregulated, which means that they are expressed at a higher level in one experimental condition than then the other, and downregulated meaning that it is expressed at a lower level (Xie et. al, 2020).

Using GEO2R, the top DEGs in the GSE237960 dataset between the control and the S. aureus infected cells were shown clearly. In the volcano plot below (Figure 2), the red dots signify upregulated DEGs, the blue represent downregulated genes, and the black are not statistically significant enough ($p \geq 0.05$) to be DEGs (Figure 2). After getting the list of DEGs downloaded, the top 15 upregulated DEGs and the top 15 downregulated DEGs were picked after narrowing the list down by deleting the DEGs that were not statistically significant ($p \geq 0.05$) (Supplementary Results: Top 30 DEGs).

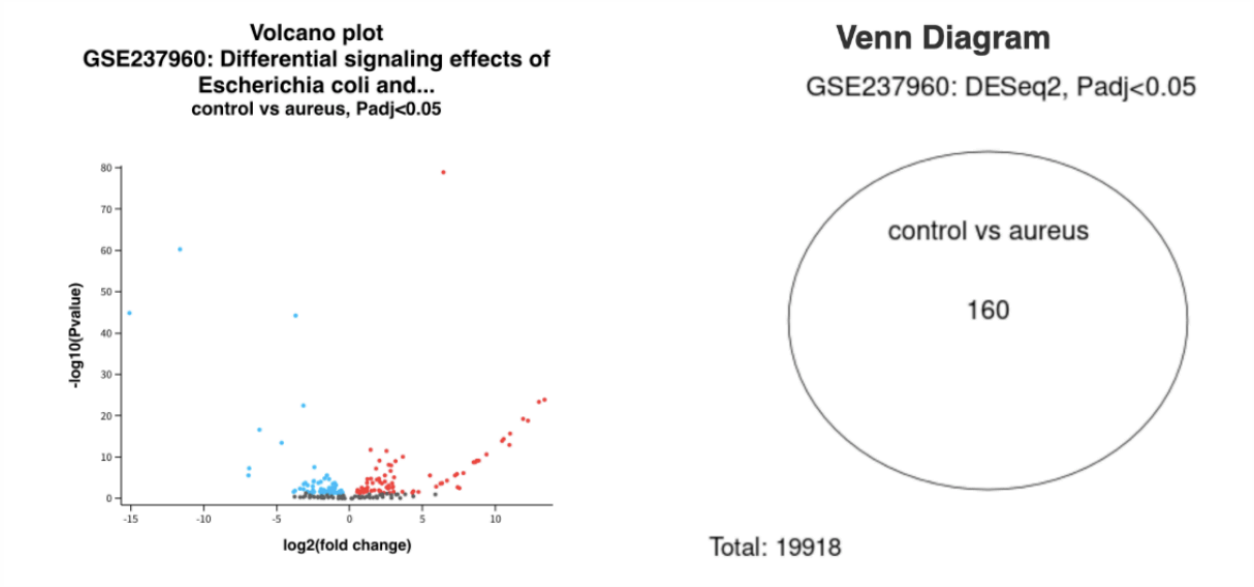


Figure 2a. Volcano plot showing the magnitude of differential expression in the healthy samples vs the samples infected with S. aureus. The blue dots represent genes that are up-regulated.while the blue dots represent genes that are down-regulated.

Figure 2b. A venn diagram portraying the amount of genes expressed at similar levels between the control (n = 4) and the infected (n = 3) groups. The 160 represents genes that re expressed differently between the control and infected. On the other hand, the 19918 shows all the genes identified in this study.

Functional analysis of the top DEGs

After identifying the top 30 DEGs, the genes were imported into SR Plot bioinformatics tool in order to understand how the dysregulation of these genes affected the cell and its function as well as evaluate whether this change in cell function could be related to sepsis. Using the SR Plot bioinformatics tool, GO and KEGG pathway enrichment analysis were carried out on the DEGs. The results showed that histone lysine demethylation, histone demethylation, and protein demethylation were the major biological processes that were affected by the DEGs (Figure 3). It also shows that the histone methyltransferase complex is the most enriched cellular component and the histone demethylase activity is the most enriched molecular function (Figure 3).

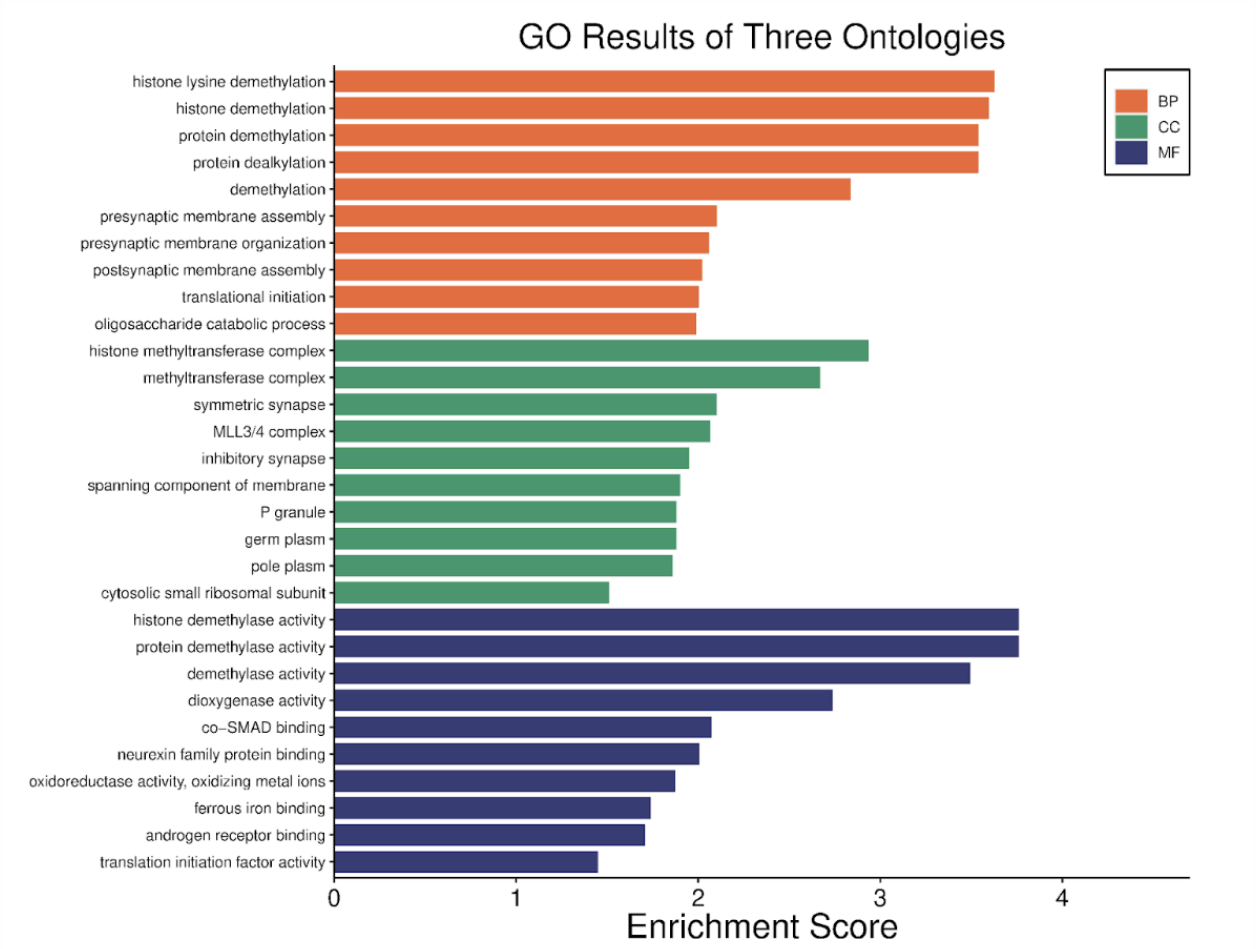


Figure 3. The GO pathway enrichment results. These results show the main biological processes (BP), cellular components (CC), and molecular functions (MF) represented within the top DEGs.

Further, the SR Plot revealed several genes that were involved in the major KEGG pathways (Figure 7) as well as the major genes that played a part in the biological processes (BP) (Figure 4), cellular components (Figure 5), and molecular functions (Figure 6). For the BP, the significant genes included UTY, KDM5D, NLGN4Y, RPS4Y1, and EIF1AY and they were associated with the biological processes of demethylation, protein dealkylation, and translational initiation. For the CC, the significant genes included UTY, KDM5D, and DDX3Y and they affected the histone methyltransferase complex and MLL3/4 complex. The gene ontology terms for the MF that came from the genes UTY, KDM5D, USP9Y, NLGN4Y, and EIF1AY were related to histone demethylase activity and protein demethylase activity.

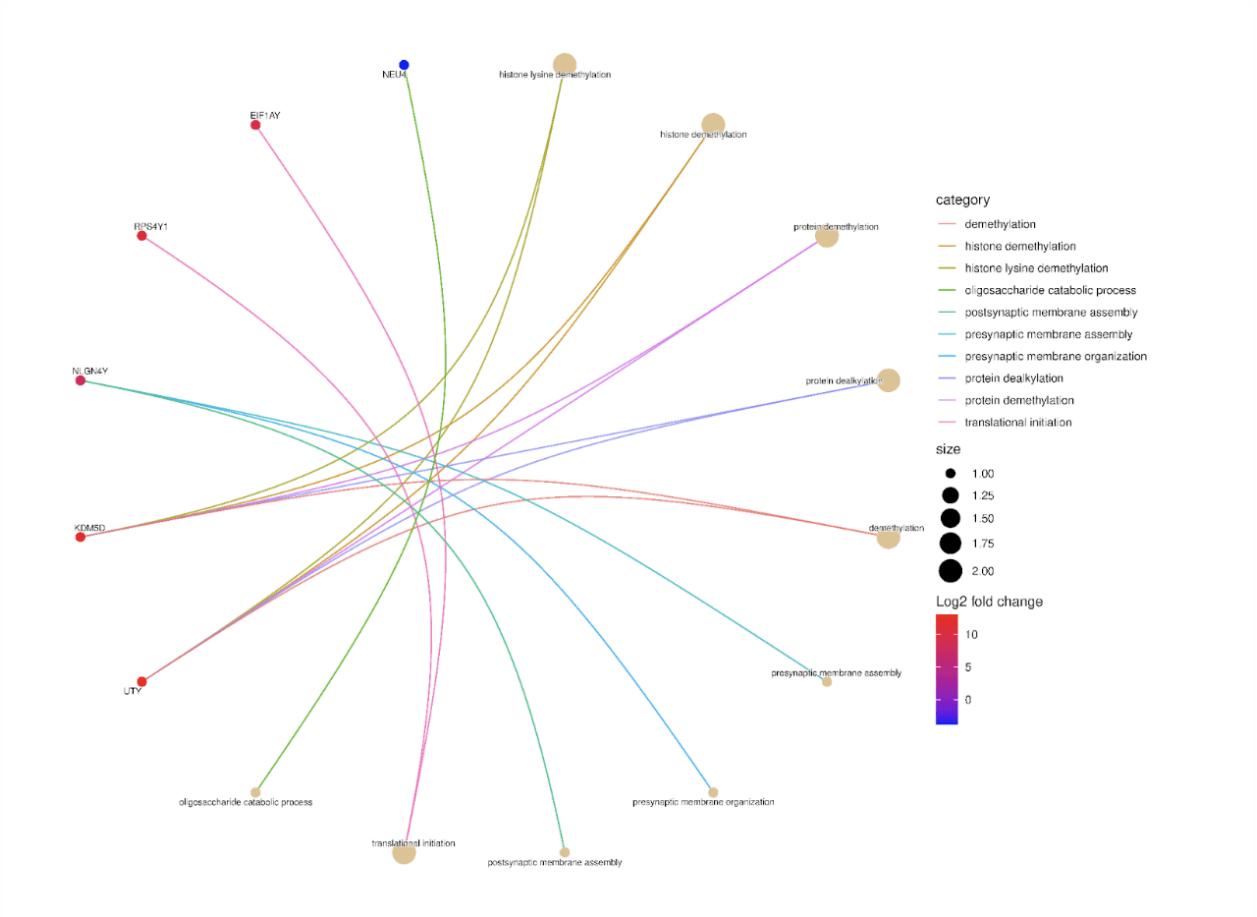


Figure 4. Effect the top 30 DEGs have on the biological process. The dots are genes that are red and blue and the strings lead to the biological processes that the genes are related to. The color of the gene dot shows the level at which each gene is dysregulated with red being the most significant.

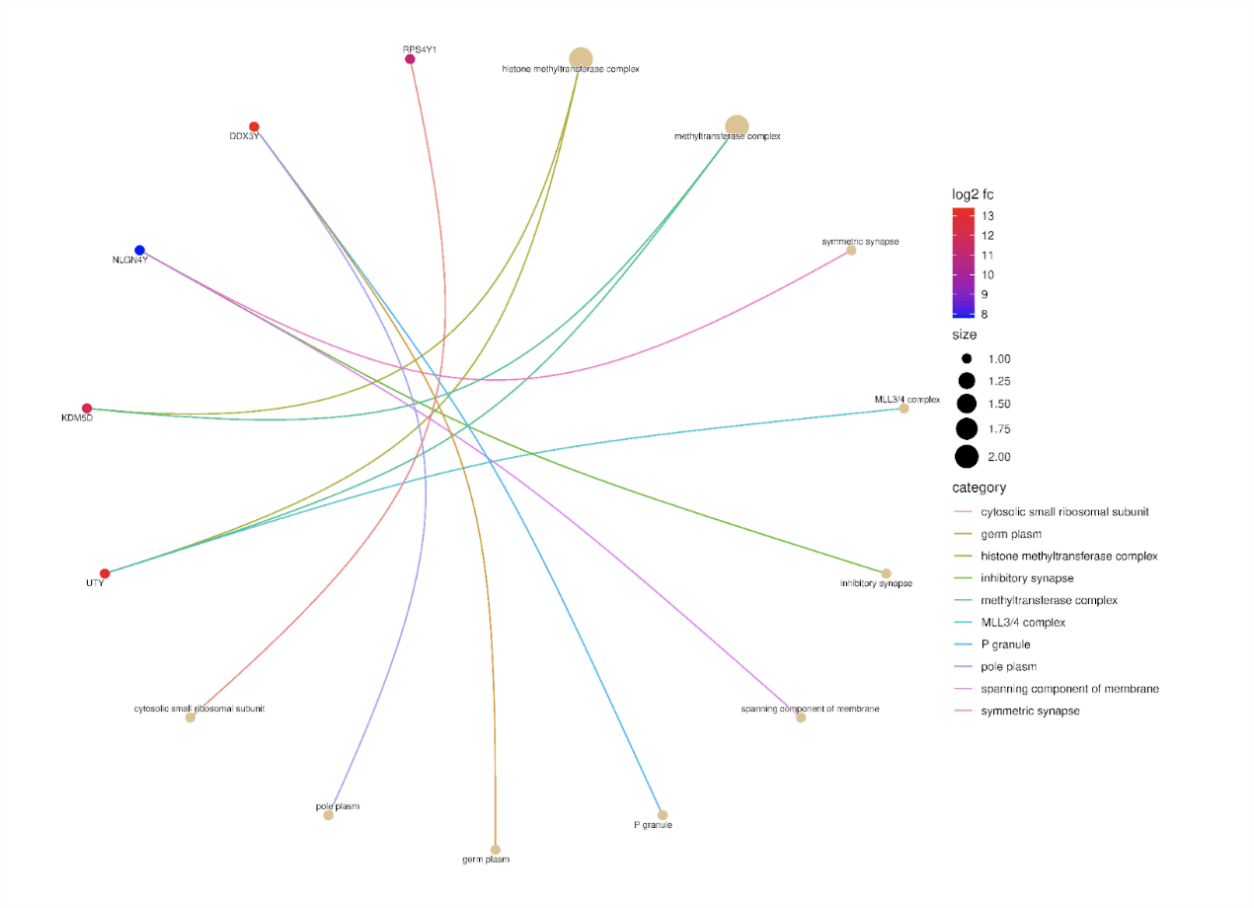


Figure 5. Effect the top 30 DEGs have on the cellular components. The dots are genes that are red and blue and the strings lead to the part of the cell that the genes are related to. The color of the gene dot shows the level at which each gene is dysregulated with red being the most significant.

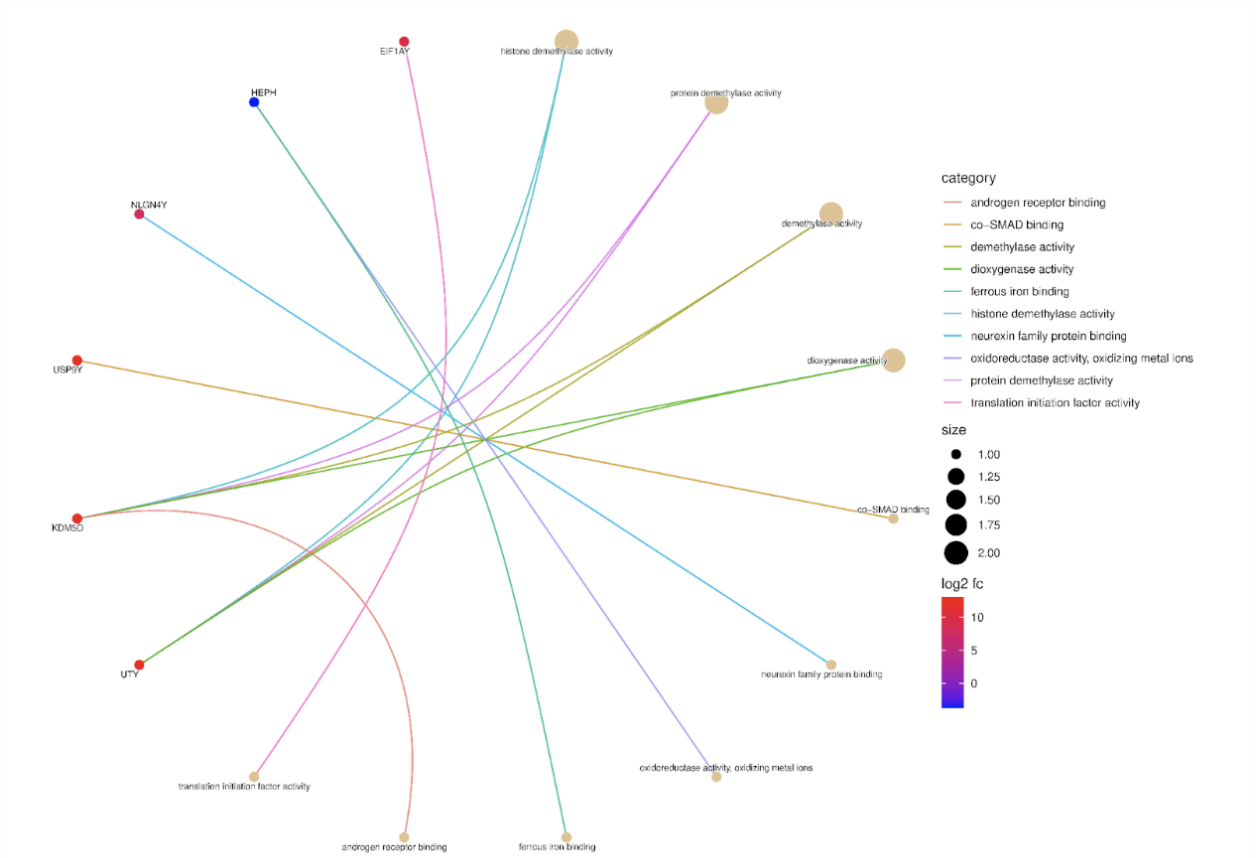


Figure 6. Effect the top 30 DEGs have on the molecular functions of the human blood. The dots are genes that are red and blue and the strings lead to the molecular functions that the genes are related to. The color of the gene dot shows the level at which each gene is dysregulated with red being the most significant.

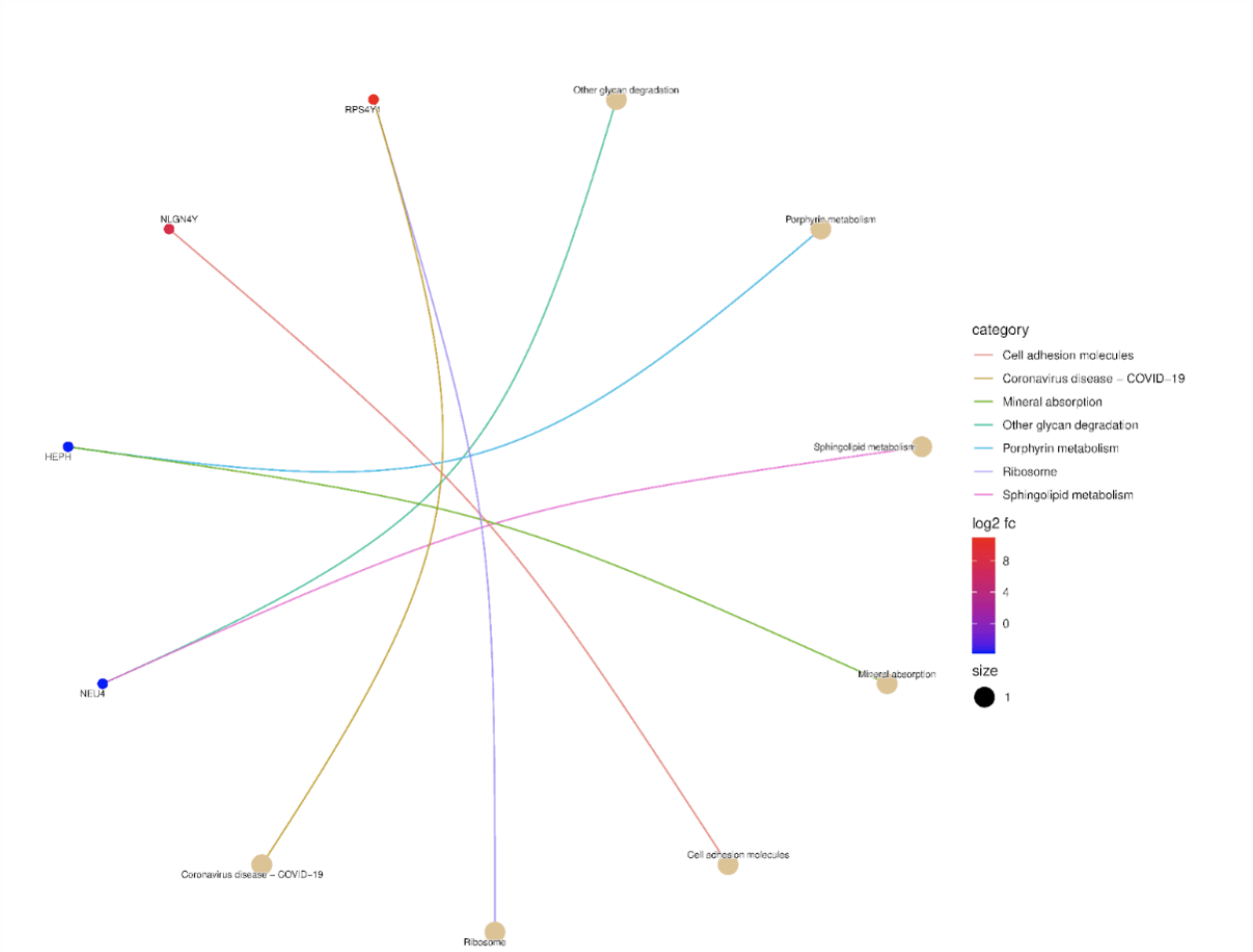


Figure 7. This graph shows the effect the top 30 DEGs have on biological pathways. The dots are genes that are red and blue and the strings lead to the biological pathways that the genes are related to. The color of the gene dot shows the level at which each gene is dysregulated with red being the most significant.

DISCUSSION

The goal of this research was to understand how S. aureus affects the genes in human blood cells when it invades into the bloodstream and determine if it was possible to create a drug that stopped S. aureus from causing sepsis. In order to figure this out, several bioinformatic tools were used to identify genes that are differentially expressed by the S. aureus signaling effects in patients versus healthy individuals and thus allow scientists interested in stopping sepsis to test these genes and see if the problem lies with the genes that were dysregulated. Dysregulated genes are differentially expressed genes that reflect statistically significant changes in RNA level or

DNA expression in patients compared to healthy controls. In the case that they are, scientists can target the gene and neutralize the effects.

In order to fix these genes that were dysregulated by S. aureus, scientists can use gene therapy. Gene therapy is a medical approach that prevents a disease by fixing the genetic problem that causes it (Scheler and Krebsbach, 2009) To do gene therapy, scientists usually introduce a new gene into the cells that fights the disease or create a copy of the gene that is not altered by the disease to replace the one that is faulty (Gonçalves and Paiva, 2017).

When analyzing the top DEGs, it became clear that many genes were coding for histone demethylation at methyltransferase proteins. Methyltransferase proteins are enzymes that add one or more methyl groups to their substrates which include DNA, RNA, and protein (Boriack-Sjodin and Swinger, 2016). This change in the amount of certain methyltransferase proteins correlates to a promotion of sepsis as they play a crucial role in inflammation in the human body (Wu et. al, 2023). These methyltransferase proteins can change how genes involved in inflammation work thus potentially making the inflammation more or less harmful (Wu et. al, 2023). The methylation, the process of adding methyl groups to substrates, by these methyltransferase proteins can lead to faultiness in key mechanisms that regulate inflammation (Wu et. al, 2023) An example of this is the methylation of H3K4, which is a specific site where methylation can occur on the histone protein H3 (Wu et. al, 2023) The methylation of H3K4 supports the production of inflammatory cytokines which makes the inflammation larger (Wu et. al, 2023). This methylation can not only activate inflammatory genes but also downregulate methyltransferases that regulate the inflammation (Wu et. al, 2023). This methylation ultimately leads to prolonged inflammation and poor recovery, which are both symptoms experienced by people with sepsis (Wu et. al, 2023). This is also the case for demethylation as it can also change immune-related pathways when it has abnormal expression levels (Qu et. al, 2023). Overall, not only can methylation promote sepsis, but also demethylation which occurs in the samples infected by S. aureus.

From the results in Figure 4, Figure 5, and Figure 6, it is clear that there are two main genes that have been upregulated due to the infection by S. aureus. These two genes are UTY and KDM5D and they promote histone

demethylation (Table 1). These two genes are the ones that were highlighted in this research because in Figure 4, Figure 5, and Figure 6, they are both red, meaning that there was a major change in expression level for the two genes. The upregulation of genes coding for histone methylation as well as the fact that these two genes are red in the graphs prove that these two genes promote sepsis due to the existence of S. aureus. Overall, S. aureus boosts histone demethylase activity through the top DEGs. Scientists can use the top DEGs, UTY and KDM5D (Table 1), in gene therapy by altering it to control histone demethylation, thus reducing inflammation and stopping sepsis from causing more deaths.

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

GENE ID	FUNCTION
UTY	Activates histone demethylation, which has either a regulatory or detrimental effect to the inflammatory response that is taken against S. aureus invasion into the bloodstream
KDM5D	Also activates histone demethylation, which has either a regulatory or detrimental effect to the inflammatory response that is taken against S. aureus invasion into the bloodstream

Due to the fact that the bioinformatics dataset that was used in this study was from an experiment performed by other researchers, the key DEGs identified in my study (Table 1) need to be tested by other scientists in the laboratory or clinical setting to validate their potential use in gene therapy for sepsis. Additionally, in the future, I would like to examine more DEGs in the dataset I used because I was only able to identify 2 genes that had an effect on sepsis and there could be more hidden in the dataset. I also think it would be interesting if experiments could be done focusing on gene expression levels that occur during sepsis in humans as opposed to gene expression in the bacteria itself.

Conclusion

Overall, this research study's purpose was to identify genes that can be potentially used in gene therapy to prevent patients who have been infected by S. aureus from dying by sepsis. In order to do this, the GEO2R bioinformatics tool was used to get differentially-expressed genes (DEGs) that potentially had an effect on sepsis. These genes were narrowed down further into 30 major DEGs and imported into SR Plot where GO and KEGG pathway enrichment analysis was done on the genes. These tools identified two major genes, UTY and KDM5D that have an impact on sepsis through their promotion of histone demethylase activity. Through further testing and research on these genes, gene therapy could be potentially formulated that ultimately stops deaths caused by sepsis due to S. aureus.

Supplementary Results

GEO2R Analysis R-script

Identified DEGs from GEO2R

References

1. Barrett, T., Troup, D. B., Wilhite, S. E., Ledoux, P., Rudnev, D., Evangelista, C., Kim, I. F., Soboleva, A., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Muertter, R. N., & Edgar, R. (2009). NCBI GEO: archive for high-throughput functional genomic data. Nucleic acids research, 37(Database issue), D885–D890. https://doi.org/10.1093/nar/gkn764
2. Boriack-Sjodin, P. A., & Swinger, K. K. (2016). Protein Methyltransferases: A Distinct, Diverse, and Dynamic Family of Enzymes. Biochemistry, 55(11), 1557–1569. https://doi.org/10.1021/acs.biochem.5b01129
3. Chen, B., Gao, L., & Shang, X. (2021). A two-way rectification method for identifying differentially expressed genes by maximizing the co-function relationship. BMC genomics, 22(Suppl 1), 471. https://doi.org/10.1186/s12864-021-07772-2
4. Clough, E., & Barrett, T. (2016). The Gene Expression Omnibus Database. Methods in molecular biology (Clifton, N.J.), 1418, 93–110. https://doi.org/10.1007/978-1-4939-3578-9_5
5. Dantes, R. B., & Epstein, L. (2018). Combatting Sepsis: A Public Health Perspective. Clinical infectious diseases : an official publication of the

Infectious Diseases Society of America, 67(8), 1300–1302. https://doi.org/10.1093/cid/ciy342
6. Edgar, R., Domrachev, M., & Lash, A. E. (2002). Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. Nucleic acids research, 30(1), 207–210. https://doi.org/10.1093/nar/30.1.207
7. Gonçalves, G. A. R., & Paiva, R. M. A. (2017). Gene therapy: advances, challenges and perspectives. Einstein (Sao Paulo, Brazil), 15(3), 369–375. https://doi.org/10.1590/S1679-45082017RB4024
8. Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., Ringwald, M., … Gene Ontology Consortium (2004). The Gene Ontology (GO) database and informatics resource. Nucleic acids research, 32(Database issue), D258–D261. https://doi.org/10.1093/nar/gkh036
9. Kulasingam, V., and Diamandis, E. P. (2008). Strategies for discovering novel cancer biomarkers through utilization of emerging technologies. Nat. Clin. Pract. Oncol. 5, 588–599. doi: 10.1038/ncponc1187
10. McAdow, M., Kim, H. K., Dedent, A. C., Hendrickx, A. P., Schneewind, O., & Missiakas, D. M. (2011). Preventing Staphylococcus aureus sepsis through the inhibition of its agglutination in blood. PLoS pathogens, 7(10), e1002307. https://doi.org/10.1371/journal.ppat.1002307
11. Parlet, C. P., Brown, M. M., & Horswill, A. R. (2019). Commensal Staphylococci Influence Staphylococcus aureus Skin Colonization and Disease. Trends in microbiology, 27(6), 497–507. https://doi.org/10.1016/j.tim.2019.01.008
12. Rung, J., and Brazma, A. (2013). Reuse of public genome-wide gene expression data. Nat. Rev. Genet. 14, 89–99. doi: 10.1038/nrg3394
13. Scheller, E. L., & Krebsbach, P. H. (2009). Gene therapy: design and prospects for craniofacial regeneration. Journal of dental research, 88(7), 585–596. https://doi.org/10.1177/0022034509337480
14. Tang, D., Chen, M., Huang, X., Zhang, G., Zeng, L., Zhang, G., Wu, S., & Wang, Y. (2023). SRplot: A free online platform for data visualization and graphing. PloS one, 18(11), e0294236. https://doi.org/10.1371/journal.pone.0294236

15. Thomer, L., Schneewind, O., & Missiakas, D. (2016). Pathogenesis of Staphylococcus aureus Bloodstream Infections. Annual review of pathology, 11, 343–364. https://doi.org/10.1146/annurev-pathol-012615-044351
16. Wu, D., Shi, Y., Zhang, H., & Miao, C. (2023). Epigenetic mechanisms of Immune remodeling in sepsis: targeting histone modification. Cell death & disease, 14(2), 112. https://doi.org/10.1038/s41419-023-05656-9
17. Xie, J., Xu, Y., Chen, H., Chi, M., He, J., Li, M., Liu, H., Xia, J., Guan, Q., Guo, Z., & Yan, H. (2020). Identification of population-level differentially expressed genes in one-phenotype data. Bioinformatics (Oxford, England), 36(15), 4283–4290. https://doi.org/10.1093/bioinformatics/btaa523

Sahaana Rangaraj

Using Bioinformatics to Investigate Differences in Regulation of Leukocyte Gene Activity Between Schizophrenic and Neurotypical Populations

ABSTRACT

Background: Schizophrenia is a psycho-somatic disease that occurs to both men and women at different ages. It does not have a cure yet, but there are certain antipsychotics and other neurological medications that control symptoms. This disease occurs due to a combination of both environmental and genetic factors.This disease causes individuals to have hallucinations, paranoia, disorganized speech or thoughts, and sometimes act irrationally. Due to these symptoms, patients with this condition tend to be isolated. There is no certain cause for this disease. It is a composite of genetic factors and environmental factors, but the specifics of each have not been discovered yet.

Methods: Therefore, the goal of this research study is to determine the effect of the expression of significant genes in patients with schizophrenia and their potential functions and biological pathway to potentially identify genetic biomarkers for potential medications targeting specific gene regulation.

Results: To conduct this study, several bioinformatics tools and databases were used including NCBI, GEO2R, SRplot, GO enrichment analysis and KEGG enrichment analysis. Specifically, the GEO Dataset used to find the DEGs (differentially expressed genes was the accession number GSE263180. To categorize the samples within GEO2R, 2 groups were defined. One was the control population, which didn’t have schizophrenia, and the other population, which did have schizophrenia. Finally, SRplot was used to conduct GO and KEGG pathway analyses to determine the potential functions and biological pathways these genes are associated with.

Discussion: The bioinformatic tools that generated these results utilized the R programming language, which was pre-programmed by artificial intelligence and machine learning algorithms. The script used to formulate these results is linked here R Script for GEO2R Sahaana . Top 30 DEGs were identified based on fold change and p value. 15 of those genes were upregulated and 15 were downregulated. Down and up regulated DEGs. The top GO terms that were enriched in our results included mitochondrial

electron transport from NADH to ubiquinone, NADH Dehydrogenase Complex Assembly, mitochondrial respiratory chain complex 1 assembly, and mitochondrial ATP synthesis coupled electron transport. The gene ND6 was also enriched and significant. The top KEGG pathways that were enriched in our results included RNA5-8SN3, ND6, oxidative phosphorylation, and retrograde endocannabinoid signaling. The gene functions mitochondrial electron transport from NADH to ubiquinone and NADH Dehydrogenase Complex Assembly, alongside the genes ND6 and RNA5-8SN3 are the main functions and biological pathways in the studied DEGs.

INTRODUCTION

Schizophrenia is a mental and neurological disorder that can be genetically inherited, but can also be caused by environmental factors (Trifu et al., 2020). Men and women are marginally inequally at risk for this disease (SAMHSA, 2023). Schizophrenia causes hallucinations, paranoia, and disorganized speech or thoughts (Patel et al., 2014). Schizophrenic people are not generally violent, but rather take on the role of a victim more often. People with this disorder struggle to communicate to others, and justify their hallucinations. They can be very isolated and afraid, since their perspective of the world is not shared by anyone else (SAMHSA, 2023). The role that genes play in schizophrenia is not heavily researched, schizophrenia itself is not a well-researched subject (Trifu et al., 2020). The goal of this study was to use statistical analysis of a GEO2R dataset (Accession number: GSE263180) to find the genetically significant differences in the presence and expression of genes in blood leukocytes of schizophrenic participants and unaffected study participants.

Leukocytes in particular were important for this study, as compared to red blood cells (RBCs) or neurons, due to their ability to do many things other cells cannot do, like entering the cerebrospinal fluid (CSF) at its formation site, or enter capillaries located in the surface of the brain (Wang et al., 2024). It was important for us to discover the relevance of specific leukocyte genes to schizophrenia.

Bioinformatic tools such as Genome-Wide Association Studies (GWAS) and pathway analysis using GO and KEGG are used in bioinformatics studies in order to further deepen the genetic understanding behind schizophrenia (Wahbeh et al., 2024).

The genes known to play a part in schizophrenia are regulated by other genes, which have not been heavily investigated into as the direct genes that take part in affected participants exhibiting certain symptoms (Wahbeh et al., 2024). For schizophrenic mothers, studies about the effects on the fetus are between far and few, especially exploring chronic health conditions the fetus and children may experience past the neonatal stage (Vigod et al., 2022).

The goal of this study is to observe the significance and differences in down and upregulated genes between people with schizophrenia and the control population. This study's hypothesis, in terms of gene expression, is that DEGs like NADH dehydrogenase activity will present differently in people with schizophrenia compared to the control population

This study was performed to recognize the relevance of these top 30 DEGs in schizophrenic patients to urge pharmaceutical therapeutic efforts to pursue medications targeting specific gene regulation. Schizophrenia is also a very unexplored field compared to other diseases, such as Alzheimer's Disease and Parkinson's, so the addition of more research into this field is very important to come closer to better medication and therapies for affected populations (Winship IR et al., 2019).

METHODS

NCBI and Primary Research

The different bioinformatics tools and databases used in this study are summarized in Figure 1. This study used NCBI to search for a schizophrenia dataset. Once the search was set to "GEO Datasets", a search was conducted on the best dataset revolving around schizophrenia. The dataset used in this study is called "Identification of Differential Expression Genes of Blood Leukocytes for Schizophrenia," The primary research that produced this dataset consisted of peripheral blood leukocyte samples of 9 schizophrenic patients and 20 healthy patients. Then, those samples would be taken for RNA sequencing analysis. The bioinformatic analyses used in this primary research included differentially expressed gene analysis, pathway enrichment analysis, and weighted gene co-expression network analysis.

GEO2R Analysis

The data from this primary research was then separated into two groups called “C-group,” the control group, and “T-group,” the schizophrenic group. Then, the data was visualized using the “Analyze by GEO2R” setting on the NCBI database.

SRplot Analysis

The individual results were then uploaded into a Google Sheet, where using the fold change and p value columns, the top 15 most upregulated and most downregulated genes were determined. Using the 30 DEGs, all of their symbols were uploaded into SRplot, where they were analyzed and produced both results of KEGG and GO analyses.

GO and KEGG analyses

The function of the DEGs in this dataset are revealed through the GO analysis, which shows genes with functions such as mitochondrial electron transport from NADH to ubiquinone and NADH dehydrogenase complex assembly. Some biological pathways that these DEGs are involved in are retrograde endocannabinoid signaling and oxidative phosphorylation.

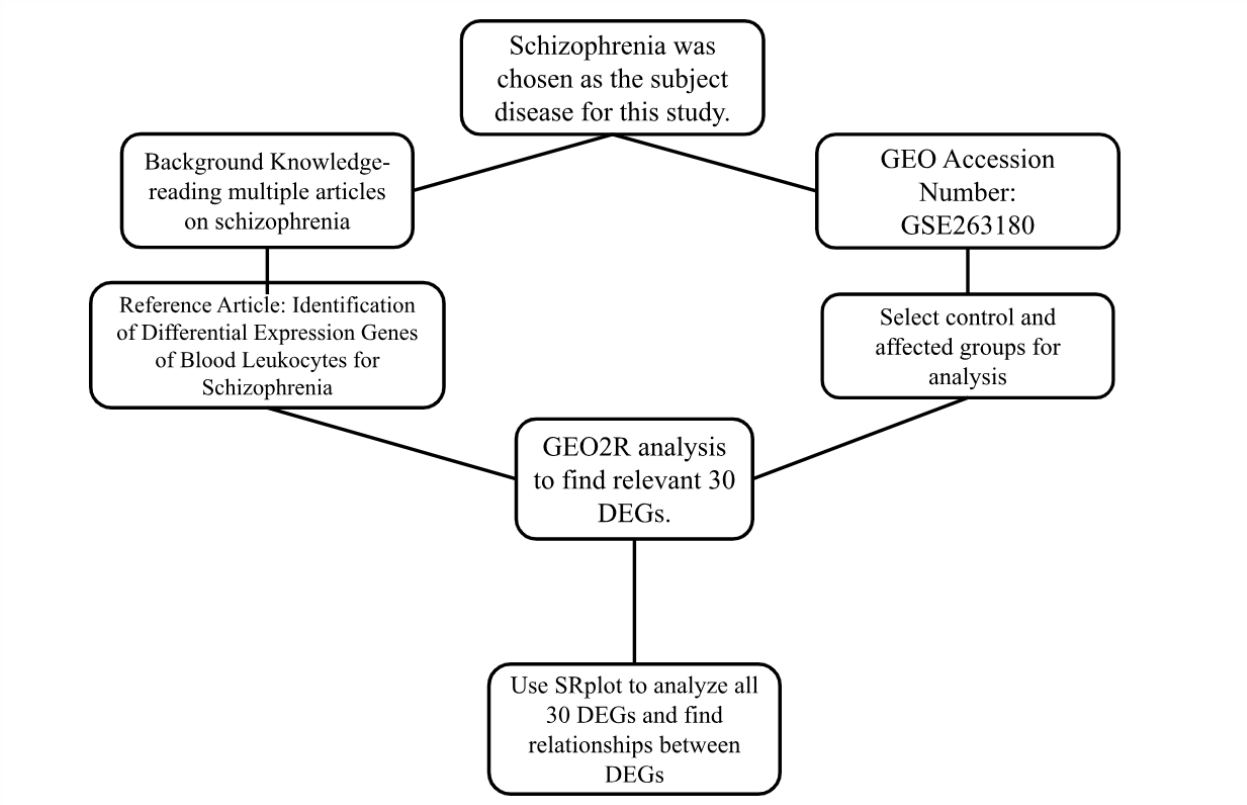


Figure 1. Summary of the methods and bioinformatics tools and databases used in this study

RESULTS

GEO2R Results

In this study, 30 DEGs were identified based on fold change and p values. There were 15 upregulated and 15 downregulated genes included in GEO2R analysis data top DEGs. Using SRplot, visualizations of this study's results were able to be produced, through which- under the assumption enrichment scores above 3 are considered enriched- mitochondrial transport of NADH to ubiquinone, NADH dehydrogenase complex assembly, mitochondrial respiratory chain complex 1 assembly, NADH dehydrogenase ubiquinone activity, and NADH dehydrogenase quinone activity were considered enriched DEGs.

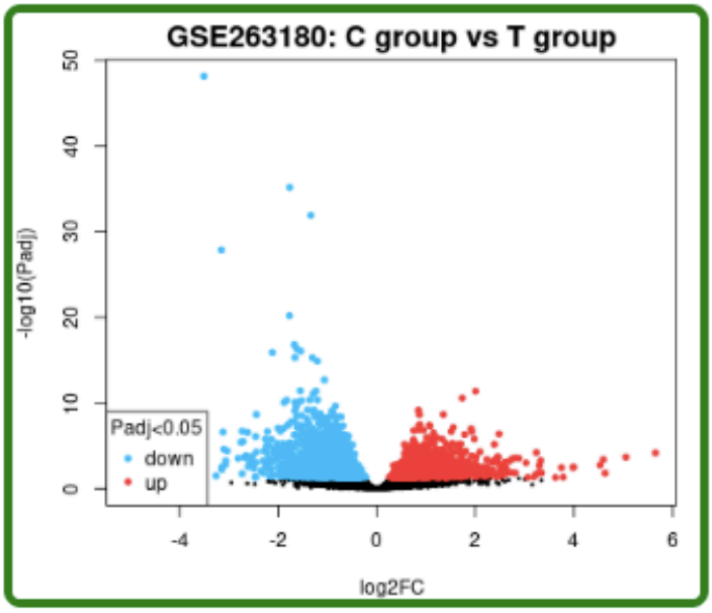


Figure 2. Volcano Plot. This volcano plot shows 3717 differentially expressed genes (DEGs) as each dot, both red and blue. The C-group is controlled (unaffected by schizophrenia), and the T-group represents schizophrenic people. Red dots represent upregulated genes, and blue dots represent downregulated genes.

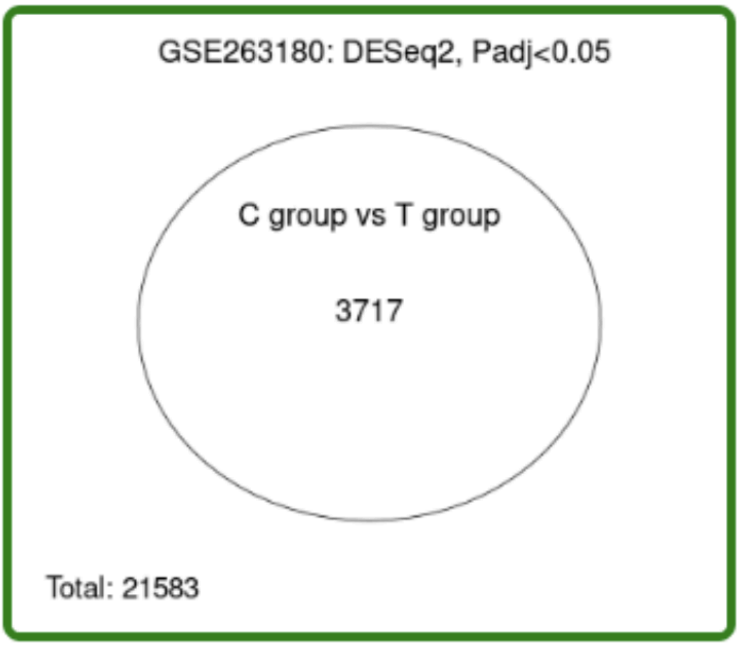


Figure 3. Venn Diagram. 3717 genes are differentially expressed (all the dots in volcano plot) out of a total 21583 total genes in the dataset.

GO Analysis Results

Using the top 30 DEGs from GEO2R analysis results, a GO analysis was performed through the server SRplot in order to obtain relevant gene functions and pathways that affect the genetic aspect of schizophrenia. The results were split into ontologies, and ranked on an enrichment score. The cellular component (CC) and molecular function (MF) ontologies ranked highest on the enrichment score scale, and hence, the gene functions within the CC and MF ontologies were focused on in more depth throughout the rest of this study. Gene functions were also separated by significance using a color-based scale, further allowing the scope of this investigation to become more focused.

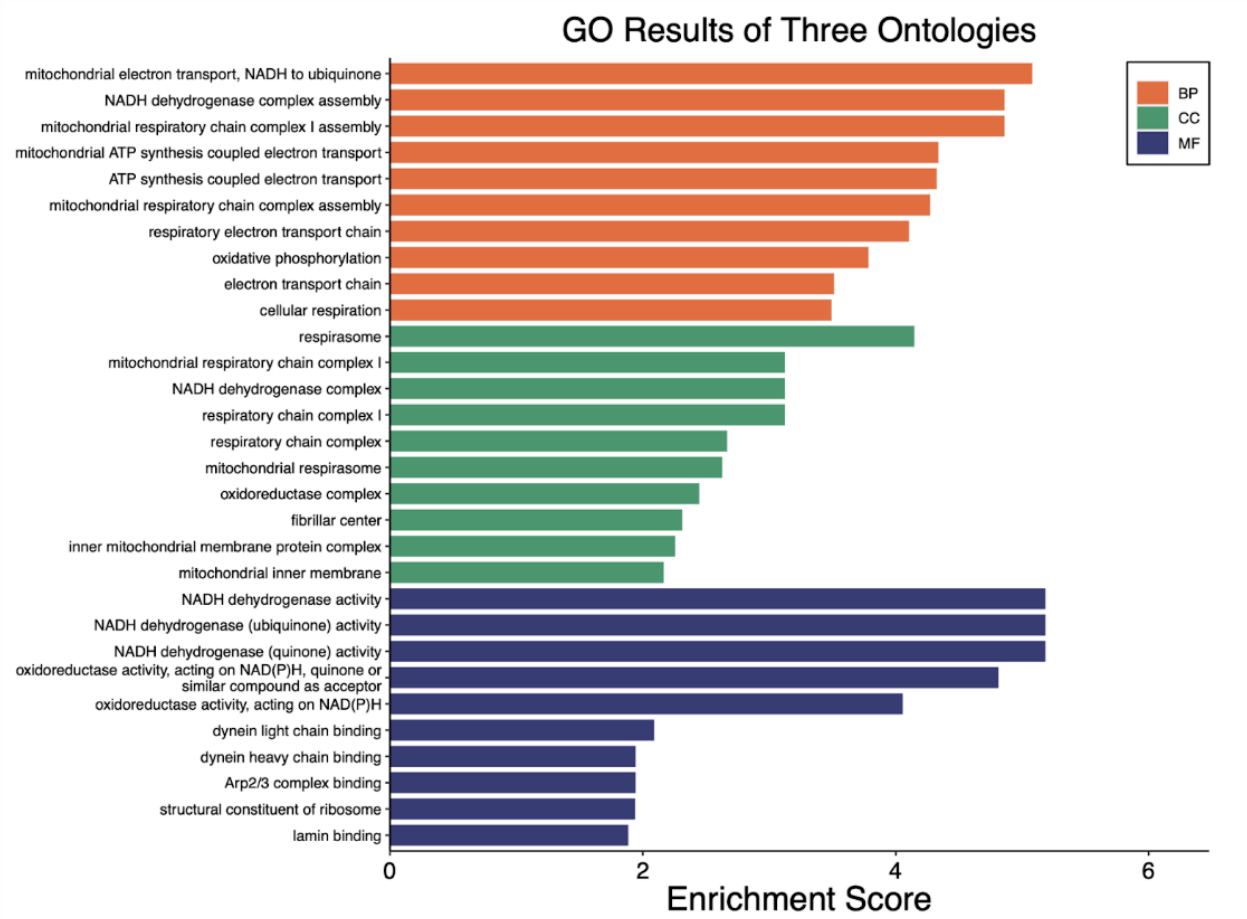


Figure 4. GO results of three ontologies. This figure displays the GO results of all three ontologies- biological processes (BP), cellular component (CC), and molecular function (MF). The most enriched ontology based on the GEO2R dataset is BP, followed by MF, and lastly with CC. Enrichment scores ranged from 0 to 5.

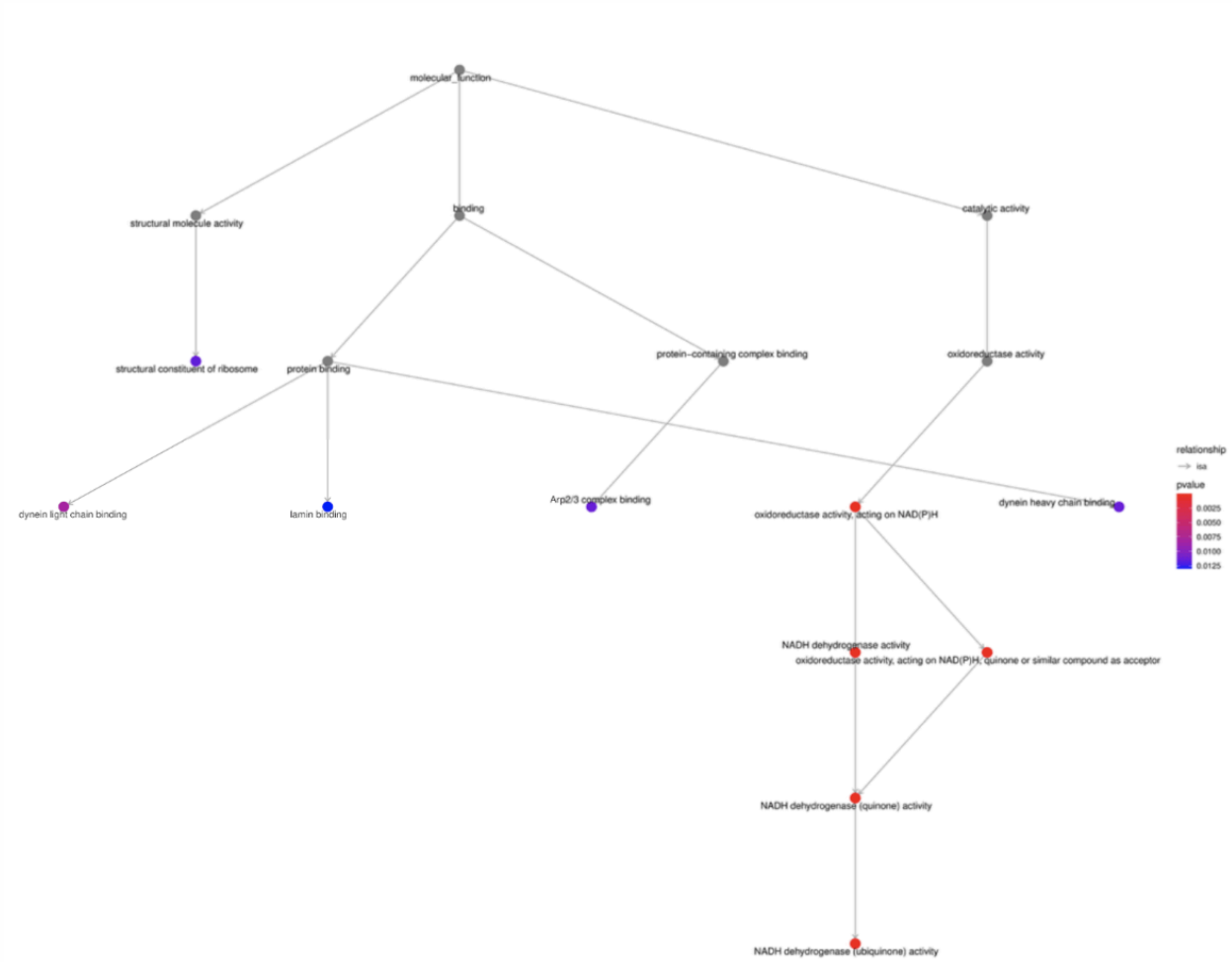


Figure 5. MF GO Plot. This plot displays the connections between BP, CC, and MF results, signified by dots connected by lines. The bottom right corner contains connections between significant results, as they are red. The more red the dot is, the more significant the result is.

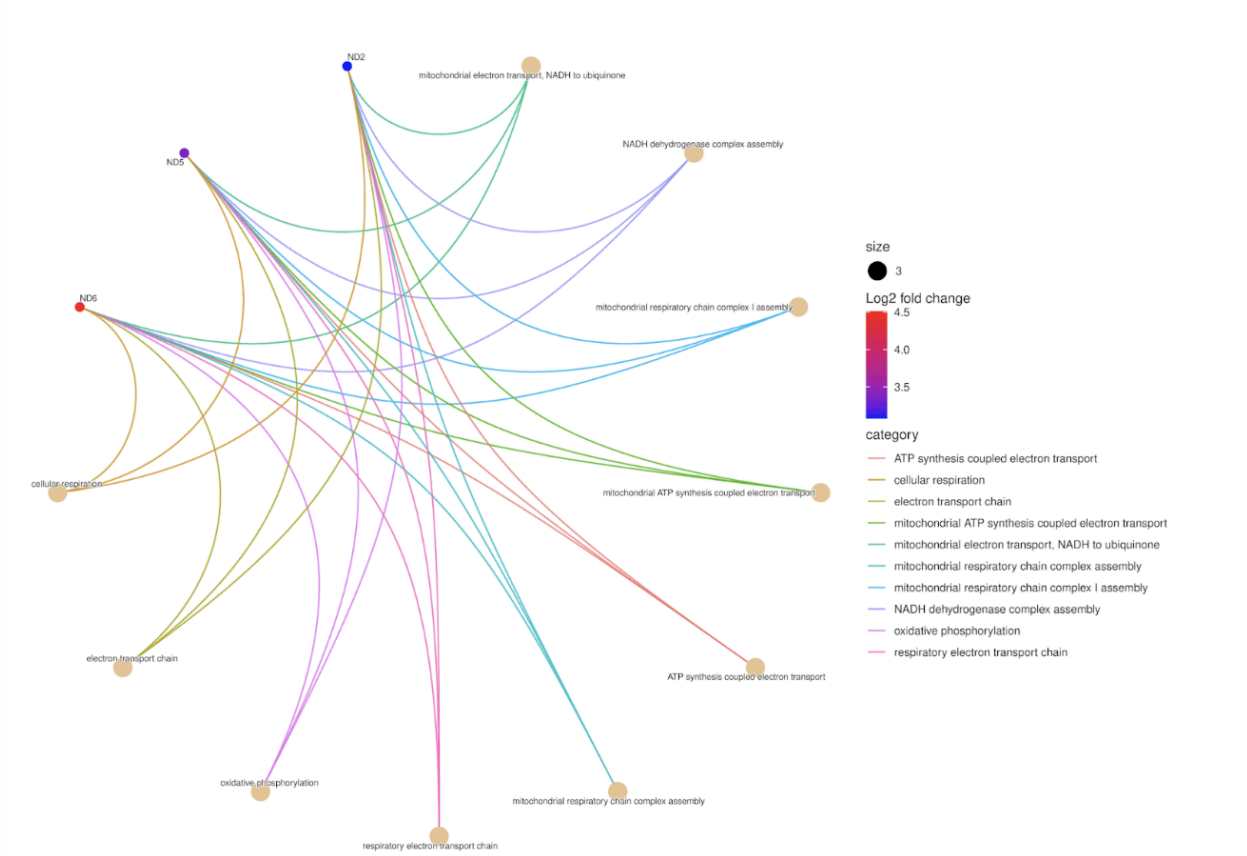


Figure 6. BP CNET plot. This plot displays the types of connections between gene functions and BP results. The most significant gene is ND6 due to its increased Log2 fold change value, and it has connections drawn to every BP result.

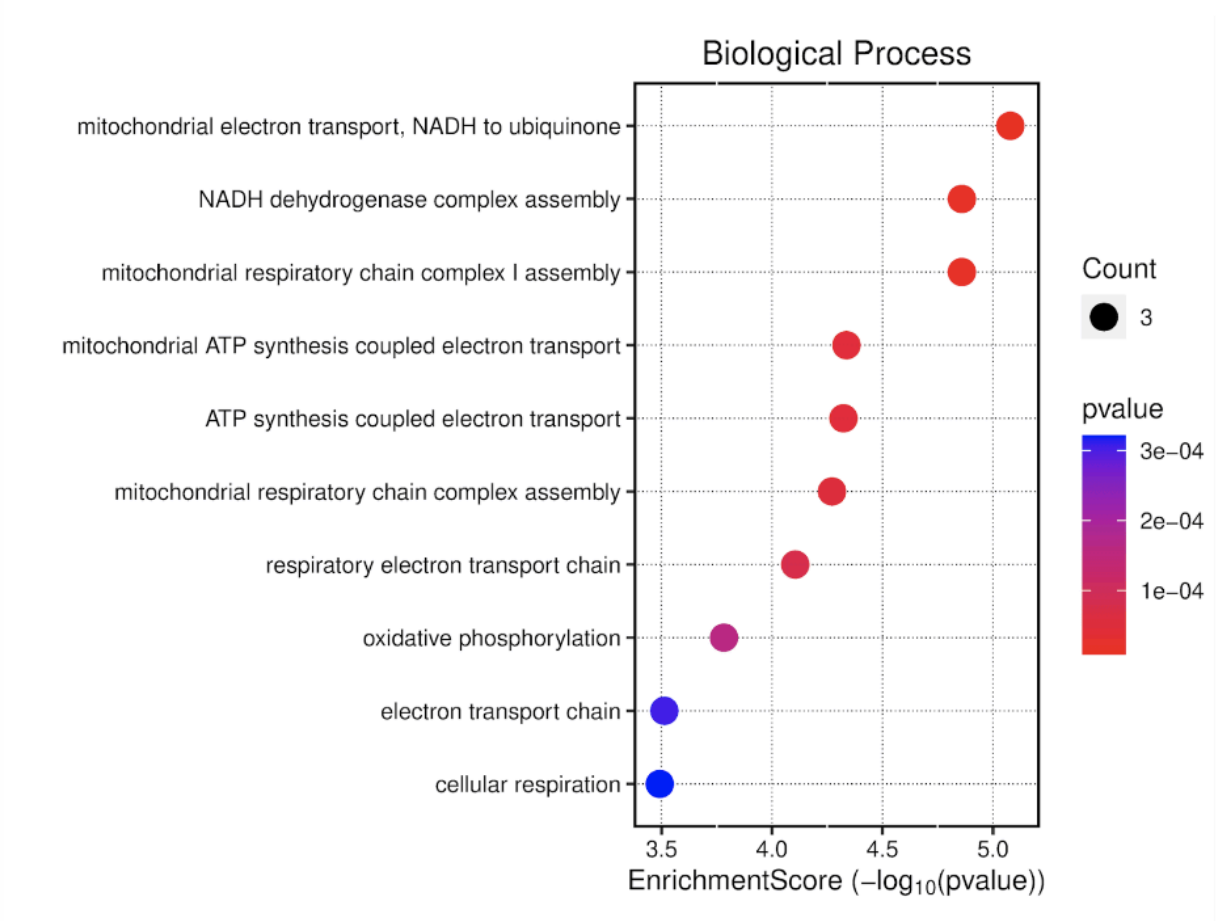


Figure 7. BP Enrichment Score Dot Plot. This plot shows the most relevant BP gene functions using a combination of the enrichment score and p-value. The top three functions are highest on the enrichment score scale, and are the most red, which makes them the most enriched and significant of the BP results.

KEGG Pathway Results

Similar to GO analysis, KEGG pathway analysis was also run in the SRplot server. It revealed the most significant and relevant gene pathways that affect schizophrenia. Associated diseases and disorders were also portrayed in these results, connecting to relevant DEGs, therefore allowing for inter-disease connections to be explored. Connecting both GO and KEGG analyses, certain parts of these results explored how gene functions and pathways in both analyses are connected in a multi-faceted network.

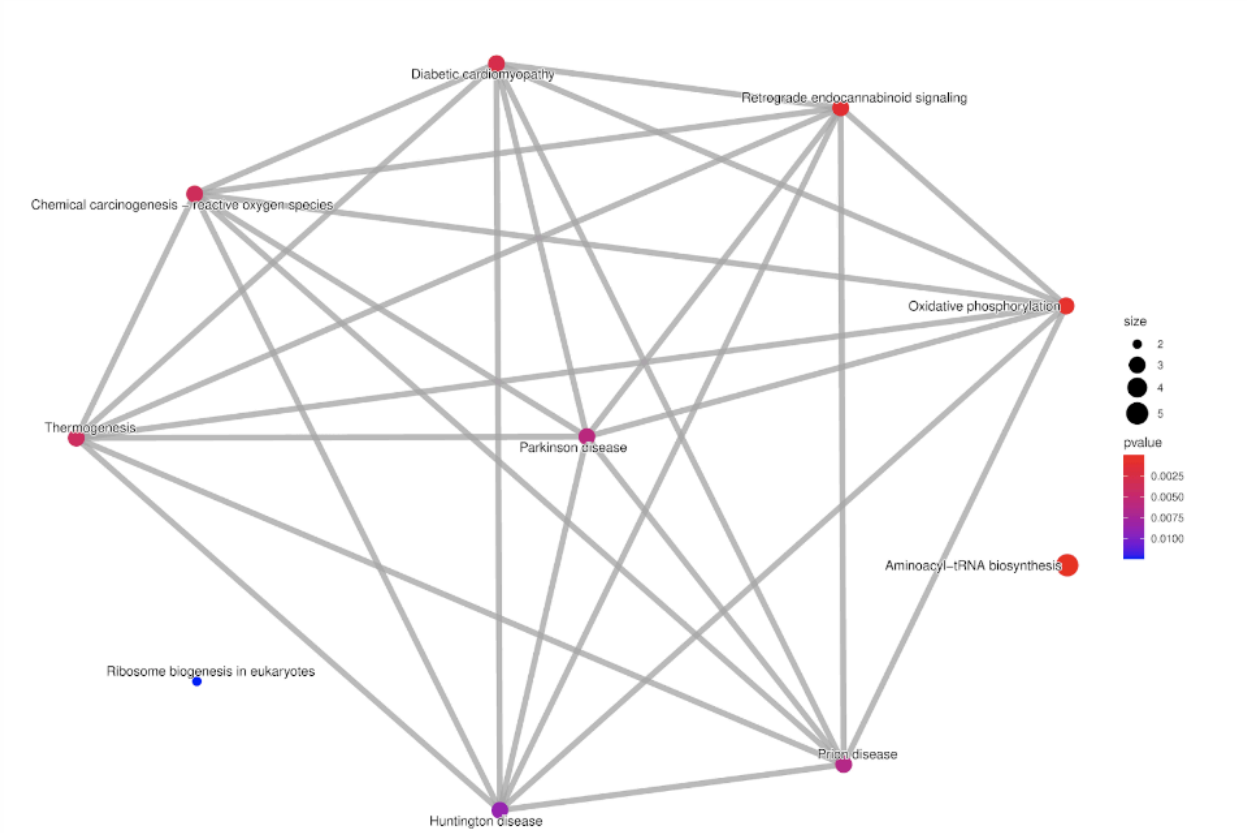


Figure 8. Pathway EMAP Plot. This plot displays the connections between the shown diseases associated with schizophrenia, separated by p-value to display the significance. The size of the dots relate to gene count enriched in a given pathway. The color of the dots reveal the significance of the pathway enrichment.

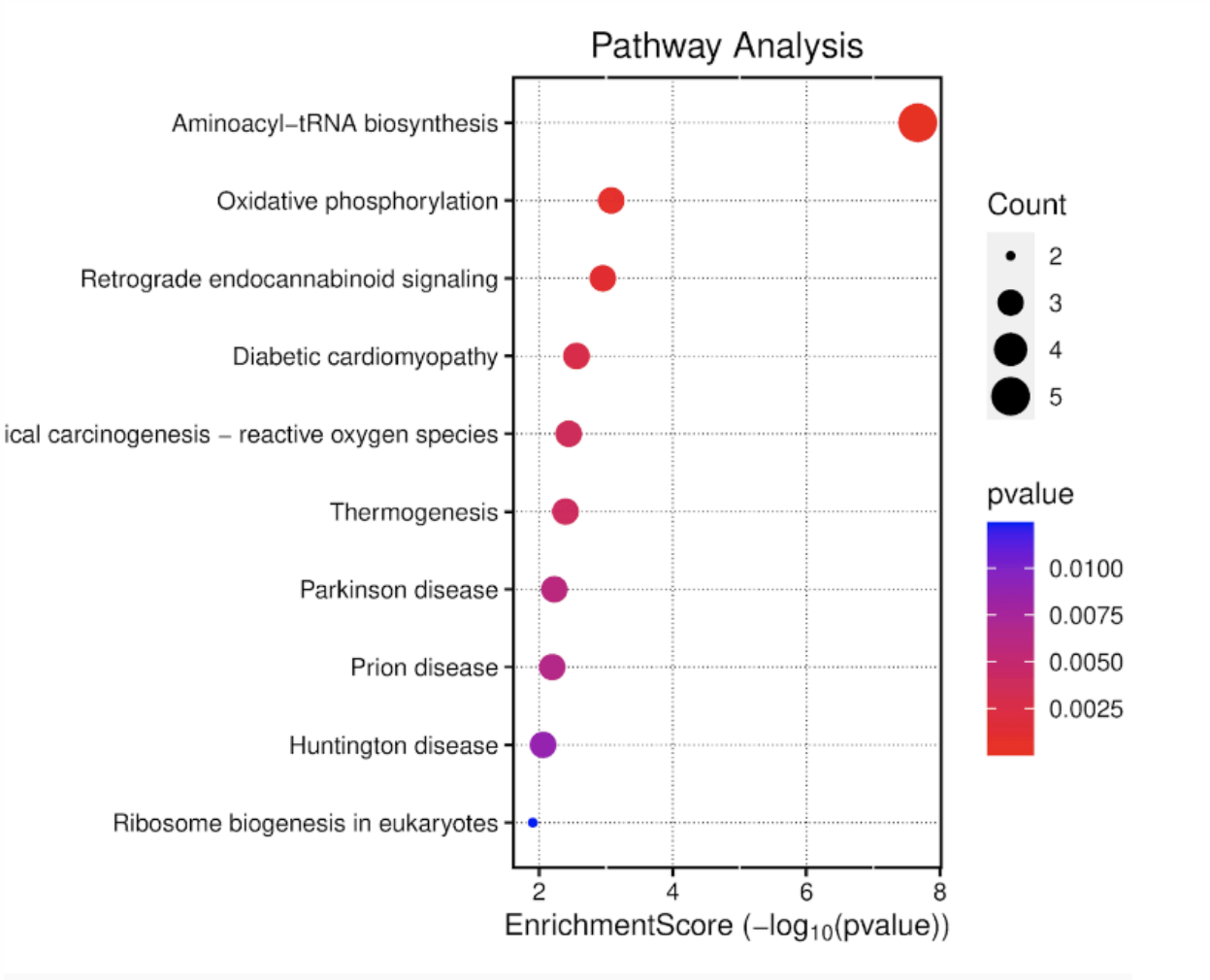


Figure 9. Pathway Enrichment Score Dot Plot. This plot displays the enrichment scores of different diseases related to schizophrenia in the KEGG pathway analysis. The top three pathway results have the greatest significance, though as seen through Figure 8, Aminoacyl-tRNA biosynthesis is disconnected from all other diseases.

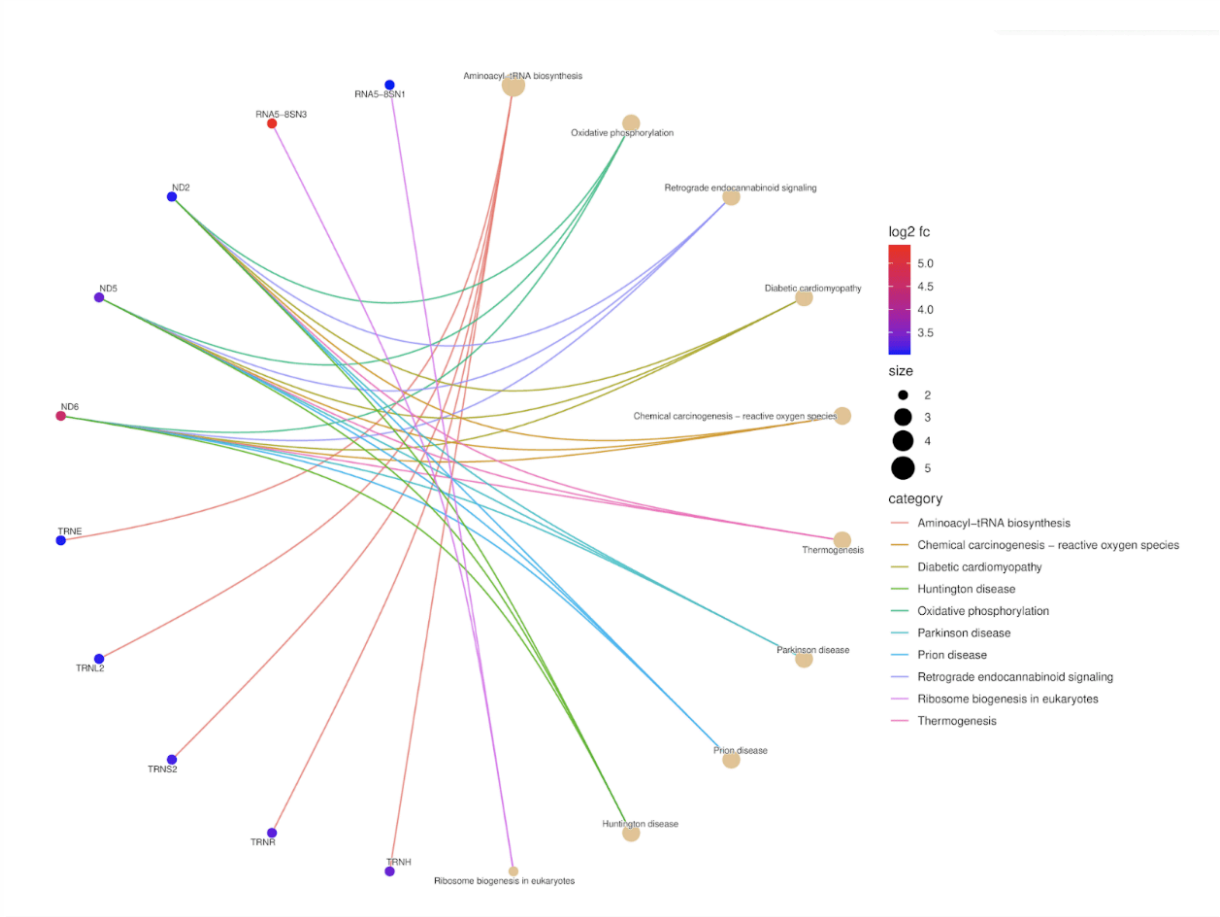


Figure 10. Pathway CNET Plot. This plot shows how gene functions and pathways in GO and KEGG analyses are connected in a network. This plot displays 10 genes that are connected to KEGG pathway analysis results. It shows how gene functions and pathways within GO terms or KEGG pathways are connected through a network. The connections are displayed through the multi-colored curved lines. Aminoacyl tRNA biosynthesis and Parkinson's Disease have the largest dots, showing their increased significance compared to the other diseases. The color of the dots display their significance. The higher the log2 fc value, or the more red the dot is, the more significant it is.

Table 1. Summary of Results

GO Results	KEGG Results	Common Results
ND6	Oxidative	ND6

	phosphorylation	
	Retrograde endocannabinoid signaling	
	Aminoacyl t-RNA biosynthesis (disconnected from all other gene functions)	
	ND6	
	RNA5-8SN3	

Supplementary Materials
R Script: R Script for GEO2R Sahaana
Top DEGs: GEO2R analysis data of top DEGs

DISCUSSION

The goal of this study, as mentioned prior, was to observe the differences in significant DEGs between people with and without schizophrenia. Out of the many gene functions presented in the results of this study, mitochondrial electron transport from NADH to ubiquinone and NADH dehydrogenase complex assembly through the gene ND6 presented as most significant, due to their low associated p-value, high log2fc values, and high enrichment scores. DEGs mainly participated in pathways such as RNA5-8SN3, ND6, oxidative phosphorylation, retrograde endocannabinoid signaling, and aminoacyl t-RNA biosynthesis pathways, as they were revealed to be the most significant due to their low p-values and high log2fc values.

ND6 Gene

The ND6 gene is located in the inner mitochondrial membrane of a cell, and is known to play a role in diseases such as Leber hereditary optic neuropathy, spinal muscular atrophy with lower extremity predominante 2B, and Leigh disease. The gene itself is responsible for enabling NADH dehydrogenase (ubiquinone) activity, mitochondrial respiratory chain complex I assembly, and mitochondrial electron transport (NCBI). Hence, these gene functions were seen as significant in the results of our GO analysis. In the past, researchers have determined that schizophrenia is

linked with mitochondrial dysfunction, specifically describing ND6 as one of many genes that function differently in patients with schizophrenia compared to an unaffected population (Verge et al., 2011). But, no study known to us has explored the relationship between leukocytic ND6 and schizophrenia.

RNA5-8SN3 Pathway

Based on the pathway CNET plot results, the main significant pathway, RNA5-8SN3, is only associated with ribosome biogenesis in eukaryotes. In the past, studies have shown that the gene group that RNA5-8SN3 is in, 5.8S ribosomal RNA, is associated with ribosome biogenesis through nucleolar protein 53 (Nop53). Nop53 is a factor for ribosome biogenesis to occur, and is also required for maturation of 5.8S to occur (Micol-Ponce et al., 2020). No prior research, to our knowledge, has associated 5.8S rRNA with schizophrenia, but rRNA gene count has been shown to be increased in people with schizophrenia compared to a healthy control sample (Chestkov et al., 2018).

Oxidative Phosphorylation

RNA5-8SN3 has not been linked to oxidative phosphorylation in prior studies known to us. Although, oxidative phosphorylation has been linked to schizophrenia through the oxidative phosphorylation system (OXPHOS) deficiencies (Bergman et al., 2016). Scientists have revealed that OXPHOS deficiencies are commonly seen in patients with disorders that may cause psychosis symptoms, like schizophrenia. The inverse relationship has also been proven- patients with OXPHOS deficiencies can exhibit symptoms of psychosis (Bergman et al., 2016).

Retrograde Endocannabinoid Signaling

RNA5-8SN3 has also not been linked to retrograde endocannabinoid signaling in prior studies known to us. Endocannibinoids (ECB), however, are linked to schizophrenia (Desfossés et al., 2010). To discuss what this signaling gene function is, it is essential to understand the meaning of "endocannabinoid". This word refers to neurotransmitters in the body that structurally represent compounds in cannabis (Grinspoon, 2021). ECBs have been proven to increase symptoms of schizophrenia due to their active role in the parts of the human brain that are disturbed by schizophrenia disorder (Desfossés et al., 2010). Endocannabinoid signaling in general (not retrograde), is associated with schizophrenia due to its association with

regulating dopamine (Covey et al., 2017). This association is relevant due to the relationship between the symptoms of schizophrenia and dopamine. An excess or underproduction of dopamine has increased effects of different schizophrenia symptoms (Davis et al., 1991).

The significance of this study's results is presented by adding more information to a small collection of schizophrenia genetic research. Also, the knowledge of RNA5-8SN3 pathway's relationship with schizophrenia can support further research into the disorder and possibly produce positive results supporting treatment plans for schizophrenia. In relevance to this study in particular, due to our use of bioinformatic datasets as a result of microarray experiments facilitated by other scientists, a limitation this study faces is our identified genes having to be further studied in a lab or clinical environment first before beginning the process of developing a vaccine or any other therapeutic medication. The 30 DEGs identified in this study can be further investigated and tested by researchers in a laboratory or clinical trial to determine the significance of the relationship between the RNA5-8SN3 and ND6 gene functions/pathways and schizophrenia in order to develop more accurate medications for schizophrenics that have less debilitating side effects.

Conclusion

The goal of this study was to observe the differences in the presence and expression of down and upregulated genes in between people with schizophrenia and a healthy population using a variety of bioinformatic tools like GEO2R and SRplot. This study's hypothesis predicted that DEGs, like NADH dehydrogenase activity, will present differently in people with schizophrenia compared to the healthy population. This hypothesis was proven to be true through the results of this study. This study found that the ND6 gene and RNA5-8SN3 pathway are connected to schizophrenia through mitochondrial dysfunction and increased rRNA gene count in cells of patients suffering from schizophrenia. Using this study's findings, further research can be facilitated on the significance of schizophrenia in relation to the RNA5-8SN3 and ND6 gene pathway's relationship in order to develop medications that more accurately target symptoms of schizophrenia to permit affected populations to have the best quality of life possible.

References

1. Adamu, M. J., et al. (2023). Unraveling the pathophysiology of schizophrenia: Insights from structural magnetic resonance imaging studies. Frontiers in Psychiatry. https://www.frontiersin.org/articles/10.3389/fpsyt.2023.1188603/full
2. Bergman, O., & Ben-Shachar, D. (2016). Mitochondrial oxidative phosphorylation system (OXPHOS) deficits in schizophrenia: Possible interactions with cellular processes. Canadian Journal of Psychiatry, 61(8), 457-469. https://doi.org/10.1177/0706743716648290
3. Chestkov, I. V., et al. (2018). Abundance of ribosomal RNA gene copies in the genomes of schizophrenia patients. Schizophrenia Research. https://www.sciencedirect.com/science/article/abs/pii/S092099641830001X
4. Covey, D. P., et al. (2017). Endocannabinoid modulation of dopamine neurotransmission. Neuropharmacology, 124, 52-61.
5. Davis, K. L., et al. (1991). Dopamine in schizophrenia: A review and reconceptualization. The American Journal of Psychiatry, 148(11), 1474-1486.
6. Desfossés, J., et al. (2010). Endocannabinoids and schizophrenia. Pharmaceuticals, 3(10), 3101-3126. https://doi.org/10.3390/ph3103101
7. Frota, I. J., et al. (2023). Decrease in cognitive performance and increase of the neutrophil-to-lymphocyte and platelet-to-lymphocyte ratios with higher doses of antipsychotics in women with schizophrenia: A cross-sectional study. BMC Psychiatry, 23(1), 558. https://doi.org/10.1186/s12888-023-05050-x
8. GEO2R. (n.d.). NCBI. https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html
9. Grinspoon, P. (2021). The endocannabinoid system: Essential and mysterious. Harvard Health. https://www.health.harvard.edu/blog/the-endocannabinoid-system-essential-and-mysterious-202108112569
10. Havdahl, A., et al. (2022). Associations between pregnancy-related predisposing factors for offspring neurodevelopmental conditions and parental genetic liability to attention-deficit/hyperactivity disorder, autism, and schizophrenia: The Norwegian Mother, Father and Child Cohort Study (MoBa). JAMA Psychiatry, 79(8), 799-810. https://doi.org/10.1001/jamapsychiatry.2022.1728

11. Micol-Ponce, R., et al. (2020). SMALL ORGAN4 is a ribosome biogenesis factor involved in 5.8S ribosomal RNA maturation. Plant Physiology, 184(4), 2022-2039. https://doi.org/10.1104/pp.19.01540
12. MT-ND6 mitochondrially encoded NADH dehydrogenase 6 [Homo sapiens (human)] – Gene – NCBI. (n.d.). National Center for Biotechnology Information. https://www.ncbi.nlm.nih.gov/gene/4541
13. Patel, K. R., et al. (2014). Schizophrenia: Overview and treatment options. P & T: A Peer-Reviewed Journal for Formulary Management, 39(9), 638-645.
14. Schizophrenia. (n.d.). SAMHSA. https://www.samhsa.gov/mental-health/schizophrenia
15. Seeman, M. V. (2018). Women who suffer from schizophrenia: Critical issues. World Journal of Psychiatry, 8(5), 125-136. https://doi.org/10.5498/wjp.v8.i5.125
16. Simoila, L., et al. (2020). Schizophrenia and pregnancy: A national register-based follow-up study among Finnish women born between 1965 and 1980. Archives of Women's Mental Health, 23(1), 91-100. https://doi.org/10.1007/s00737-019-0948-0
17. Tang, D., et al. (2023). SRplot: A free online platform for data visualization and graphing. PLOS ONE, 18(11), e0294236. https://doi.org/10.1371/journal.pone.0294236
18. Trifu, S. C., et al. (2020). Genetics of schizophrenia (review). Experimental and Therapeutic Medicine, 20(4), 3462-3468. https://doi.org/10.3892/etm.2020.8973
19. Verge, B., et al. (2011). Mitochondrial DNA (mtDNA) and schizophrenia. European Psychiatry, 26(1), 45-56.
20. Vigod, S. N., et al. (2022). Maternal schizophrenia and the risk of a childhood chronic condition. Schizophrenia Bulletin, 48(6), 1252-1262. https://doi.org/10.1093/schbul/sbac091
21. Wahbeh, M. H., et al. (2024). A functional schizophrenia-associated genetic variant near the TSNARE1 and ADGRB1 genes. HGG Advances, 5(3), 100303. https://doi.org/10.1016/j.xhgg.2024.100303
22. Wang, F., et al. (2024). Identification of differentially expressed genes of blood leukocytes for schizophrenia. Frontiers in Genetics. https://www.frontiersin.org/articles/10.3389/fgene.2024.1398240/full
23. What is schizophrenia? (n.d.). Psychiatry.org. https://psychiatry.org/patients-families/schizophrenia/what-is-schizophrenia

24. Winship, I. R., et al. (2019). An overview of animal models related to schizophrenia. Canadian Journal of Psychiatry, 64(1), 5-17. https://doi.org/10.1177/0706743718773728

Maya Sammanna

Investigating Genetic Associations with Parkinson's Disease Using Bioinformatics

ABSTRACT

Introduction: Parkinson's Disease (PD) is a progressive neurodegenerative disorder characterized by motor (tremors, stiffness, and bradykinesia) and non-motor symptoms (pain, cognitive impairment, and sleep disturbances). However, the exact causes of Parkinson's disease are not fully understood, but there is growing evidence indicating a significant genetic contribution. Therefore, to fill this gap, this study aimed to identify differentially expressed genes (DEG's) in PD patients compared to healthy controls, shedding light on genetic factors and pathways involved in PD pathogenesis.
Methods: Various bioinformatics databases and tools were used to conduct this study. First RNA sequencing data from the Gene Expression Omnibus (GEO) database (GSE168496) was analyzed. This study used post-mortem brain tissue samples from 8 PD patients and 8 healthy controls. The GEO2R bioinformatics tool was used to analyze and identify differentially expressed genes (DEGs) between PD patients and healthy samples. Based on fold change and p-value criteria, the top 30 DEGs were subjected to Gene Ontology (GO) and KEGG enrichment analyses using SR Plot to determine potential functions and biological pathways of the top DEGs.
Results: The GEO2R and SR Plot analyses identified several DEGs between PD patients and healthy samples, including SLC6A3 and PITX3, which were upregulated in PD patients. GO enrichment analysis revealed several enriched biological pathways and cellular components associated with the top DEG's, such as locomotory behavior, dopaminergic neuron differentiation, and neurotransmitter transport. KEGG pathway analysis highlighted active pathways, including dopamine metabolism and synaptic transmission. The enriched pathways underscore the involvement of dopaminergic systems in PD.
Discussion: The findings emphasize the relevance of dopaminergic pathways in PD, with SLC6A3 and PITX3 identified as key genes. These genes play crucial roles in dopamine reuptake and neuron differentiation, respectively, aligning with PD pathology. The results suggest potential biomarkers and therapeutic targets, contributing to the understanding of PD's genetic basis.

This study identified significant DEGs and enriched pathways in PD, highlighting genes like SLC6A3 and PITX3. These insights advance the understanding of PD's genetic foundations and suggest targets for therapeutic intervention.

KEYWORDS: Parkinson's disease, Bioinformatics, NCBI, GEO, dopaminergic pathways, SR Plot, Gene Ontology, KEGG

INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative disorder that affects movement (Khan et al, 2018). It causes symptoms such as tremors, stiffness, and bradykinesia (slowness of movement). The exact causes of Parkinson's disease are not fully understood, but there is growing evidence indicating a significant genetic contribution (Clarke, 2007). The research aims to address specific genetic factors associated with Parkinson's disease, which could lead to improved diagnostic methods and treatment approaches.

This study aims to determine if there is a difference in the expression of specific genes in people with Parkinson's disease compared to those who are healthy. The research specifically seeks to identify which genes exhibit changed expression patterns in Parkinson's disease patients, with the goal of enhancing our understanding of the genetic foundations of the illness.

Parkinson's disease is the second most common neurodegenerative disorder after Alzheimer's disease (Alexander, 2004). It primarily affects individuals over the age of 60, although early-onset Parkinson's can occur. The disease is characterized by the degeneration of dopamine-producing neurons in the substantia nigra, a region of the brain crucial for movement control (Klockgether, 2004). This loss of dopamine leads to the hallmark motor symptoms of Parkinson's disease. Non-motor symptoms, such as cognitive decline, mood disorders, and sleep disturbances, also significantly impact patients' quality of life (Tomlinson et al, 2013).

Parkinson's disease affects millions of people worldwide, with a substantial economic, health and social burden (Kowal et al, 2013). The exact cause of neuron degeneration in Parkinson's is not fully understood, but a combination of genetic, environmental, and lifestyle factors is believed to

contribute. Several genes, such as SNCA, LRRK2, and PARK2, have been implicated in familial forms of Parkinson's, but these account for around 10% small percentage of cases (Clarke, 2007). The majority of cases are sporadic (a type of Parkinson's), with a more complex interplay of genetic predisposition and environmental triggers.

Bioinformatics has become an essential tool in Parkinson's disease research, enabling the analysis of large-scale genomic and transcriptomic data. For example, previous bioinformatics studies have identified various genetic variants and expression profiles associated with Parkinson's disease (Diao et al., 2012). These studies have provided insights into potential molecular pathways involved in the disease, such as mitochondrial dysfunction, oxidative stress, and protein aggregation (Feldman, J et al, 2023).

However, despite advancements, much remains unknown about the genetic basis of Parkinson's disease (Yao et al., 2021). Specifically, the complete spectrum of genes involved and their precise roles in the disease's progression are not fully clear (Gasser, 2009). Additionally, the interaction between genetic factors and environmental influences in sporadic Parkinson's cases requires further investigation (Pang et al., 2019).

Therefore, the goal of this research is to analyze gene expression profiles in Parkinson's disease patients compared to healthy controls to identify differentially expressed genes. The hypothesis is that there will be significant differences in the expression levels of certain genes between the two groups. These differences in gene expression may reveal new molecular targets for therapeutic intervention and improve understanding of the disease mechanism.

This research is crucial for identifying genetic factors and understanding their role in Parkinson's disease can lead to earlier diagnosis, personalized treatment approaches, and the development of new therapeutic targets. Given the growing prevalence of Parkinson's disease, particularly in aging populations (Marras et al., 2018), this research has significant implications for public health. Therefore, advancing the knowledge of Parkinson's disease through bioinformatics ultimately benefits patients and healthcare systems worldwide.

METHODS

Study Design

This study aimed to explore differentially expressed genes (DEGs) in Parkinson's disease (PD) patients compared to healthy controls. A differentially expressed gene (DEG) is a gene that shows statistically significant differences in expression levels (how "active" a gene is) between two or more groups, such as diseased versus healthy patients. This difference can indicate the gene's involvement in specific biological processes or conditions (Anjum et al., 2016). The goal was to unravel their potential implications through Gene Ontology (GO) (Thomas et al., 2021) and KEGG (Kanehisa, 2000) enrichment analyses, providing a comprehensive understanding of the genetic underpinnings of Parkinson's disease and the molecular mechanisms possibly involved. A summary of the methodology used in this study is summarized in Figure 1.

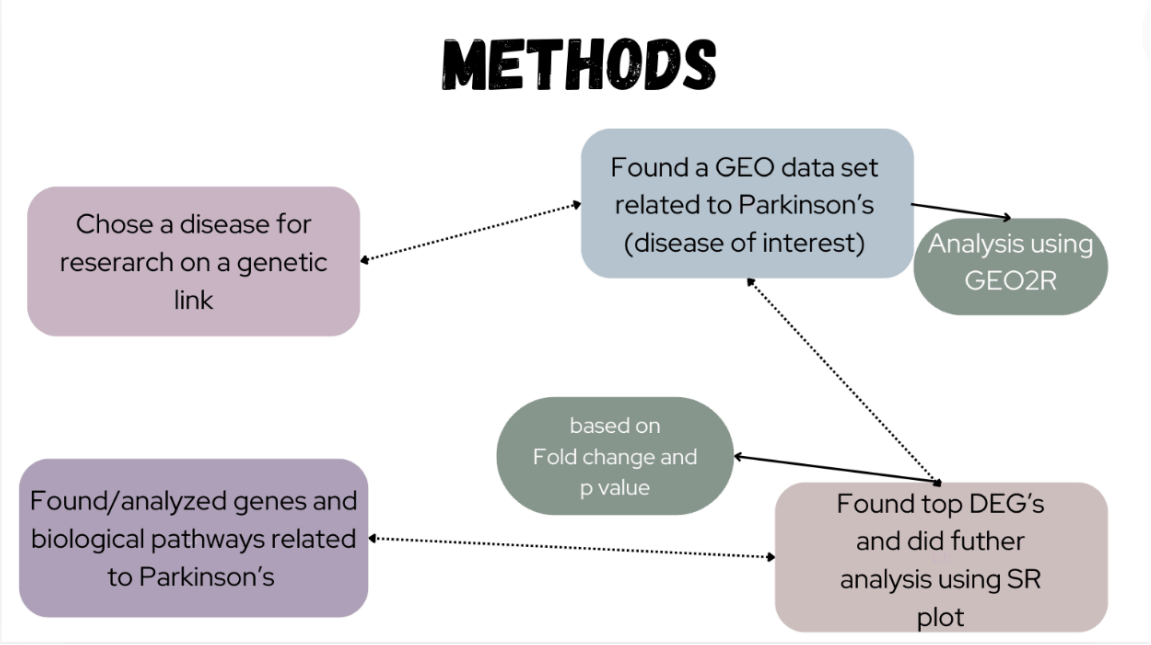


Figure 1. Methods flowchart: The flowchart above displays the methods used in this study.

Data Collection

Data for this investigation were sourced from the Gene Expression Omnibus (GEO) database (Geo, n.d.) under accession number GSE168496. The Gene Expression Omnibus (GEO) is a public database where researchers store and share genetic data from various studies (Barrett et al., 2012). The National Center for Biotechnology Information (NCBI) is a central resource for

biomedical genomic information. It provides access to many databases and tools to help analyze biological data (Kalbfleisch et al., 2010). GEO2R is an online tool provided by GEO that allows users to compare gene expression levels between sample groups and simplifies the process of identifying DEG's in data sets in the GEO database (Geo, n.d.-a). The dataset included brain tissue samples from both PD patients and age-matched healthy controls, consisting of 16 participants: 8 PD patients and 8 healthy controls. These samples were obtained post-mortem. GEO2R was employed for data preprocessing and normalization to ensure consistency and comparability across samples. This process involved categorizing samples into 'Control' and 'Parkinson's' groups based on disease status. The GEO2R analysis used the following AI and ML-programmed R Script to generate the DEG data.

Identifying Differentially Expressed Genes (DEGs)

To identify DEGs, GEO2R utilized statistical methods and algorithms capable of analyzing RNA sequencing data (Solano et al., 2024). DEGs or differentially expressed genes are genes that show statistically significant differences in expression levels between different conditions or groups (Zhang et al., 2004). Following DEG identification, a filtering approach was implemented based on two criteria: a high fold change value and a low p-value. A high fold change value indicates that a gene is significantly upregulated/downregulated in healthy controls compared to PD patients, suggesting potential involvement in disease pathology. The second criterion, a low p-value, was applied to mitigate the risk of false positives in experiments where numerous genes are tested simultaneously.

Identifying Top DEGs

From the pool of identified DEGs, the top 30 were selected for further analysis based on their significance, with a cutoff from the p-value 4.36E-03 and a cutoff of the log fold change value 1.398. The list of all and the top thirty DEG's chosen can be found in the Supplementary Data Top 30 DEGs.

Functional Enrichment Analysis: GO and KEGG

These top DEGs were subjected to GO and KEGG bioinformatics enrichment analysis tool SR Plot (SRplot – Science and Research Online Plot, n.d.), which categorized them into biological processes, molecular functions, and cellular components (SRplot- Science and Research Online Plot, n.d.). This bioinformatics analysis is a way to provide some insights into the specific

functional roles of these DEGs in the context of Parkinson's disease. It may reveal their involvement in key biological processes such as neuroinflammation, mitochondrial function, and neurotransmitter regulation.

Gene Ontology (GO) is a framework that categorizes genes based on their functions, biological processes, and cellular locations. This helps gain an understanding of gene roles and interactions (Gene Ontology Overview, 2024). Kyoto Encyclopedia of Genes and Genomes (KEGG) is a database that maps genes and their products to biological pathways, illustrating how they interact in various cellular processes and disease mechanisms (Kanehisa, 2000).

Additionally, KEGG (KEGG Database, n.d.) pathway analysis using SR Plot identified biological pathways potentially disrupted in PD. This analysis highlighted pathways such as dopamine metabolism, oxidative stress response, and synaptic transmission pathways, shedding light on critical mechanisms underlying PD pathogenesis and potential therapeutic targets. Enrichment analysis involved calculating scores and p-values to assess the significance of associations between DEGs and specific KEGG pathways.

Data Presentation and Interpretation

Visual tools like bar plots, dot plots, and cnet plots were used to effectively present and interpret the data. This helped to identify potential gene functions and key biological themes and insights. The analysis laid the groundwork for further research into targeted therapies and interventions aimed at reducing the impact of the disease on affected individuals.

RESULTS

GEO2R Analysis

Analysis of this dataset using GEO2R from the volcano plot and Venn diagram reveals that some genes were differentially expressed between control (healthy) and PD patients **(Figures 2 and 3).**

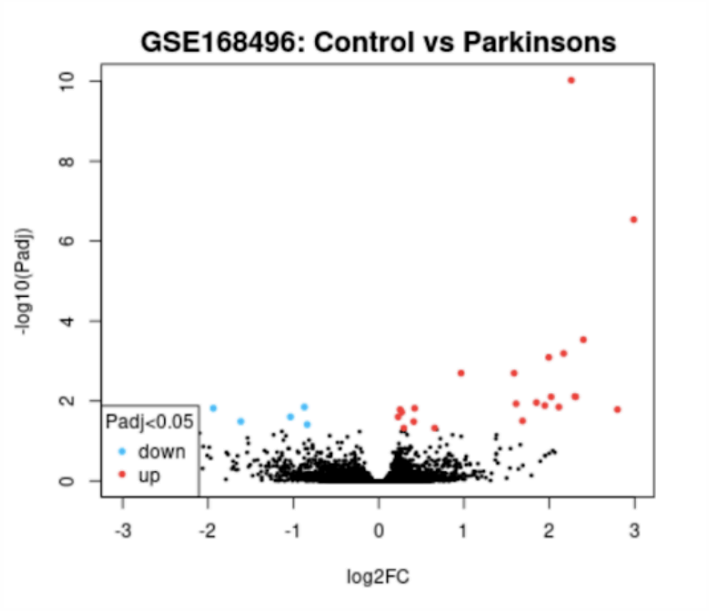


Figure 2. GEO2R Volcano Plot. The resulting volcano plot of the data after GEO2R analysis is pictured above. The blue dots represent genes that are downregulated in Parkinson's patients, and the red dots represent genes that are upregulated in Parkinson's patients. The black dots are genes that have the same amount of expression between Parkinson's patients and healthy controls.

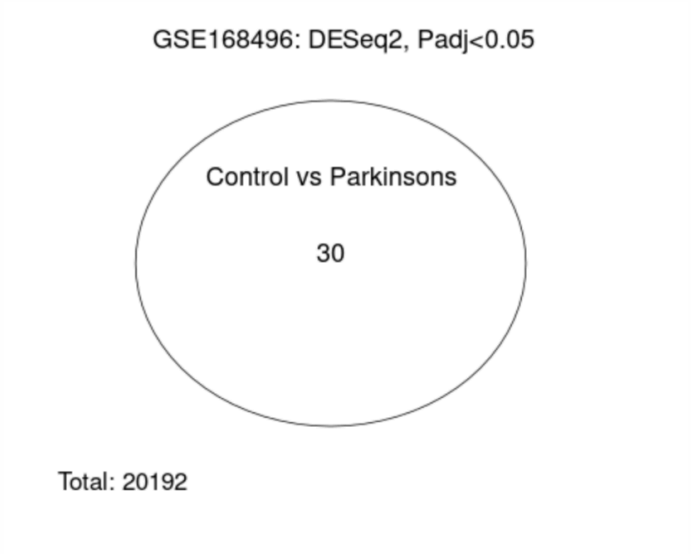


Figure 3. GEO2R Venn Diagram. The following Venn diagram was taken as the result of the GEO2R analysis. The number thirty inside the circle represents the number of genes that were differentially expressed between the two groups, out of a total of 20192 genes.

Differentially Expressed Gene Selection

Based on the fold change and p-value with a cutoff of 4.36E-03 for the p-value and a cutoff of 1.398 for the log fold change value, the top thirty DEG's were selected for further analysis. These genes can be found in the **supplementary results here:** Top 30 DEGs From these selected genes, all thirty were upregulated.

SR Plot GO Enrichment Analysis

To determine the potential functions of the top 30 DEGs and their associated biological pathways, results from the GO enrichment pathway analysis in SR plot reveal several enriched pathways **(Figure 4).** The top 30 differentially expressed genes with an enrichment score cutoff of 6 were considered significant. Notably, biological processes (BP) such as locomotory behavior, dopaminergic neuron differentiation, neurotransmitter transport, amine transport, and dopamine metabolic processes show high enrichment scores. Similarly, the cellular component (CC) pathway of dopaminergic synapse also stands out. However, molecular function (MF) pathways exhibit low activity. The dot plot of biological processes **(Figure 5)** further confirms the significant activity of pathways like locomotory behavior, dopaminergic neuron differentiation, and neurotransmitter transport.



Figure 4. GO enrichment results from SR plot. The orange color represents BP (biological processes), the green represents CC (cellular component), and MF (molecular function). The bar graph displays that mostly biological processes and some cellular components are involved.

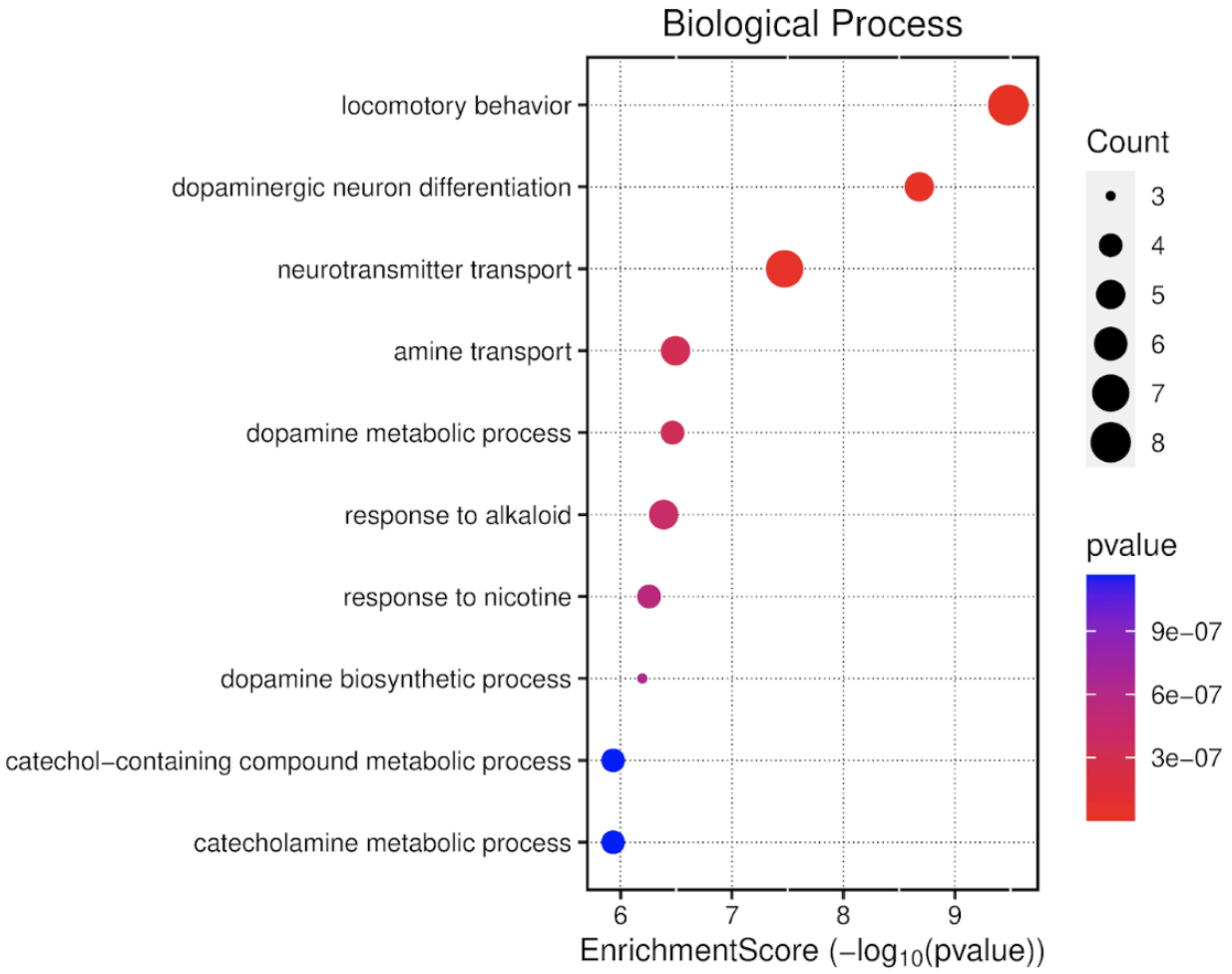


Figure 5. Biological Processes Dotplot. This shows enrichment scores for various biological process (BP) pathways. Dot size (count) reflects the number of genes in each pathway. Redder dots indicate higher statistical significance (lower p-value). A higher enrichment score suggests these pathways are more likely to be relevant to your study genes.

In the cellular component pathways, dopaminergic synapse, synaptic vesicle, and exocytic vesicle are notably enriched (Figure 6). The emapplot of biological processes (Figure 7) emphasizes the high enrichment of locomotory behavior and neurotransmitter transport pathways, while the emapplot of the cellular component (Figure 8) highlights the activity of pathways such as dopaminergic synapse and synaptic vesicle.

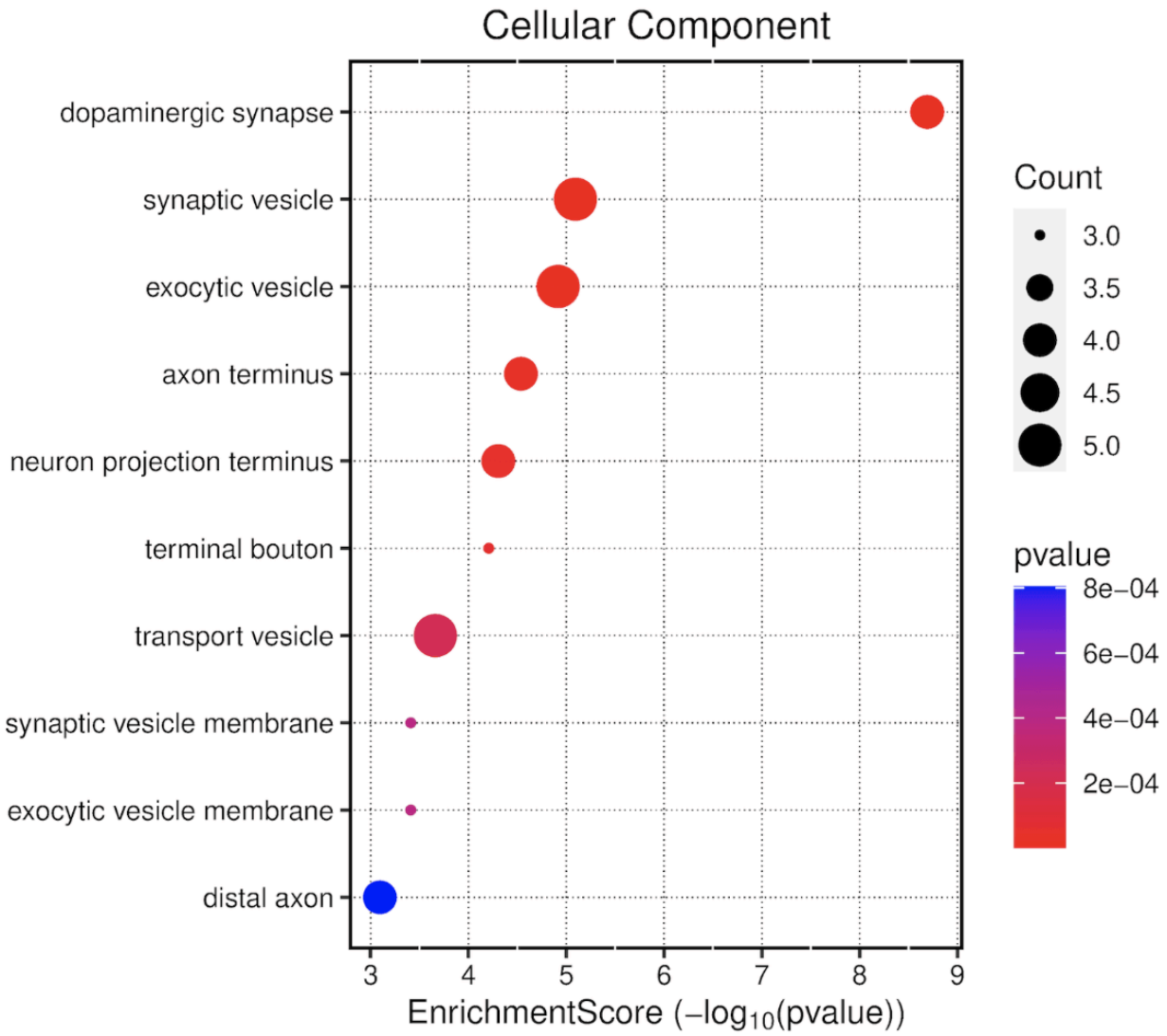


Figure 6. Cellular Component Dotplot. This dot plot shows enrichment scores for different GO terms related to cellular components (CC). The size of the dot (count) depicts the number of genes associated with that GO term. A more red-colored dot indicates a lower p-value (higher statistical significance), meaning the GO term/function is more relevant than by chance.

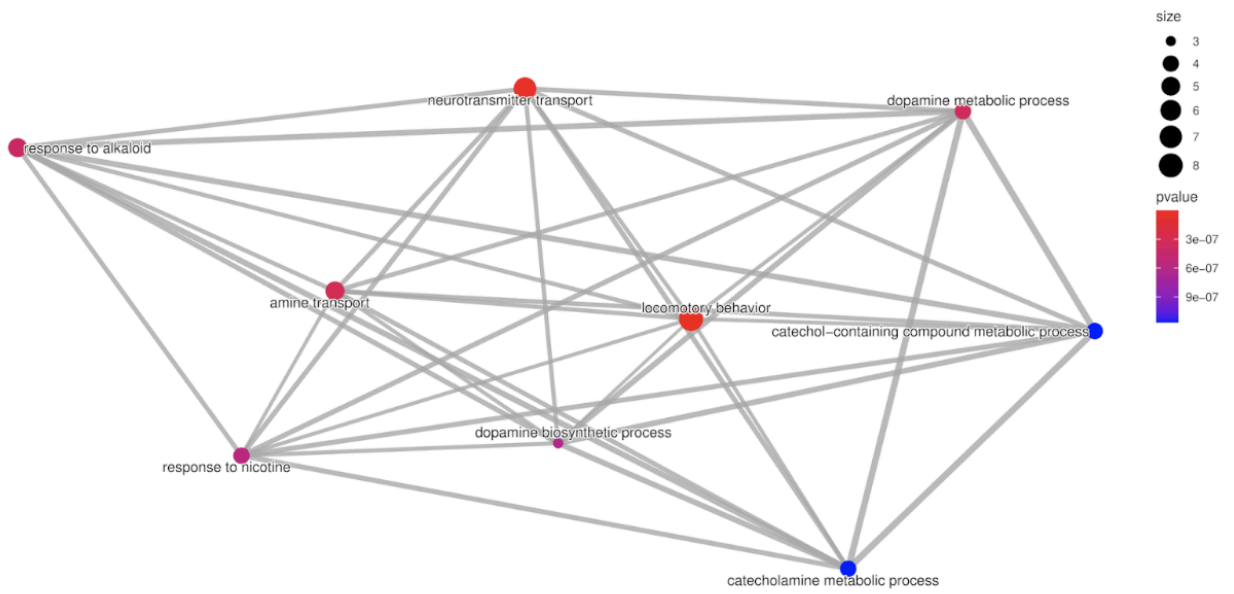


Figure 7. Emapplot of Biological Processes. This plot visualizes the results of enrichment analysis for pathways of biological processes. The larger circles represent more enriched pathways and the redder color represents more statistical significance.

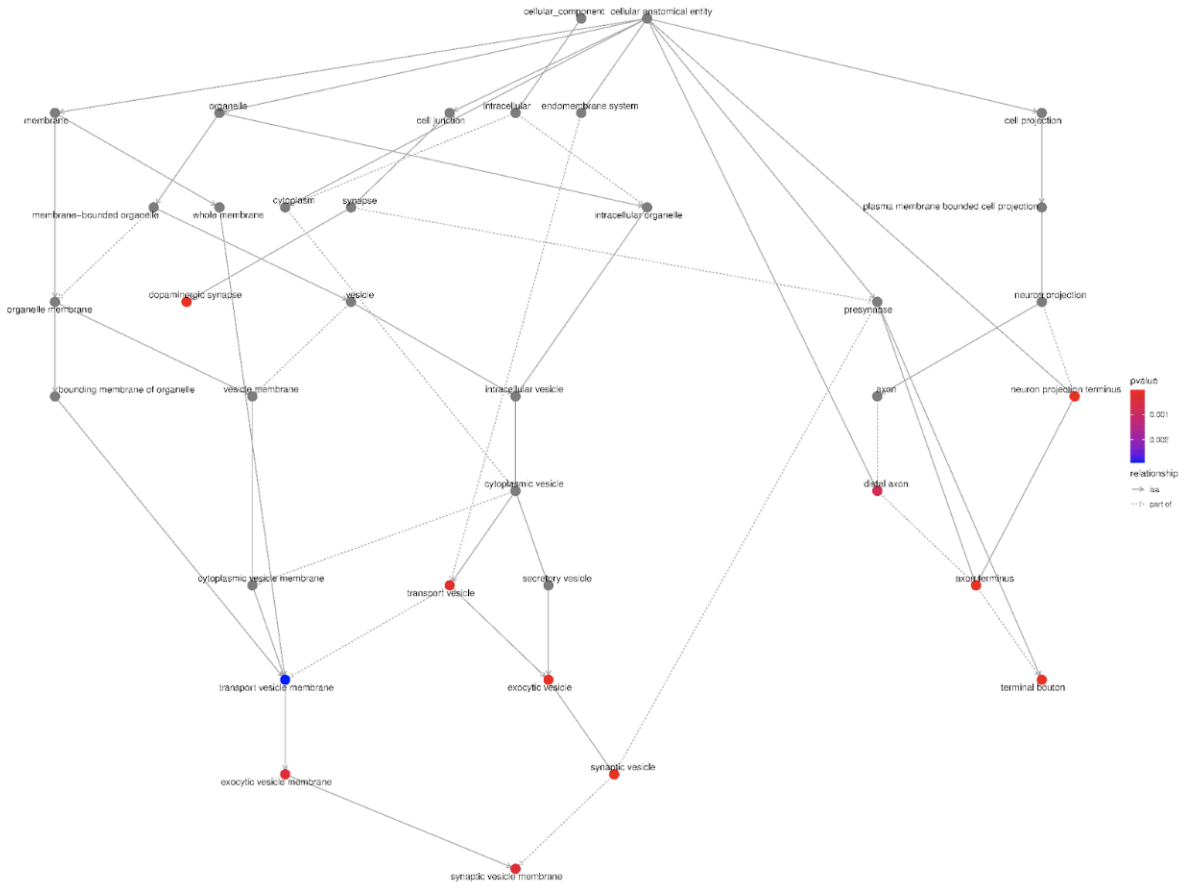


Figure 8. Emapplot for Cellular Component. The plot above displays the results of enrichment analysis for pathways of the cellular component. The larger circles represent more enriched pathways and the redder color represents more statistical relevance.

From the biological processes cnet plot (Figure 8), several genes are linked to enriched pathways. For instance, SLC6A3 is connected to pathways such as locomotory behavior, response to alkaloid, response to nicotine, and neurotransmitter transport, while PITX3 is also associated with these pathways. In the cellular component cnet plot (Figure 9), SLC6A3 is involved in the dopaminergic synapse pathway.

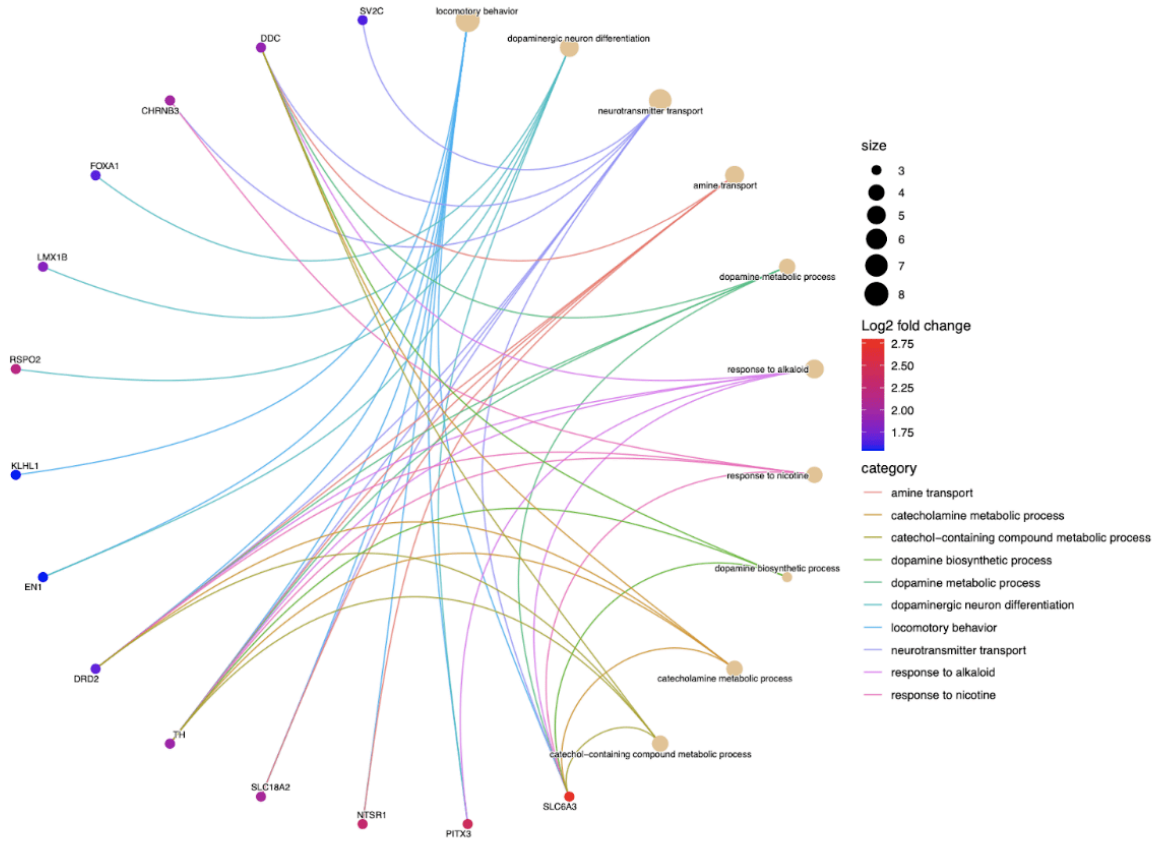


Figure 9. Cnet Plot for Biological Processes

The BP cnet plot displays the genes associated with different pathways within biological processes and connected by different colored lines. A higher log fold change value represents higher statistical importance, and the size represents the count.

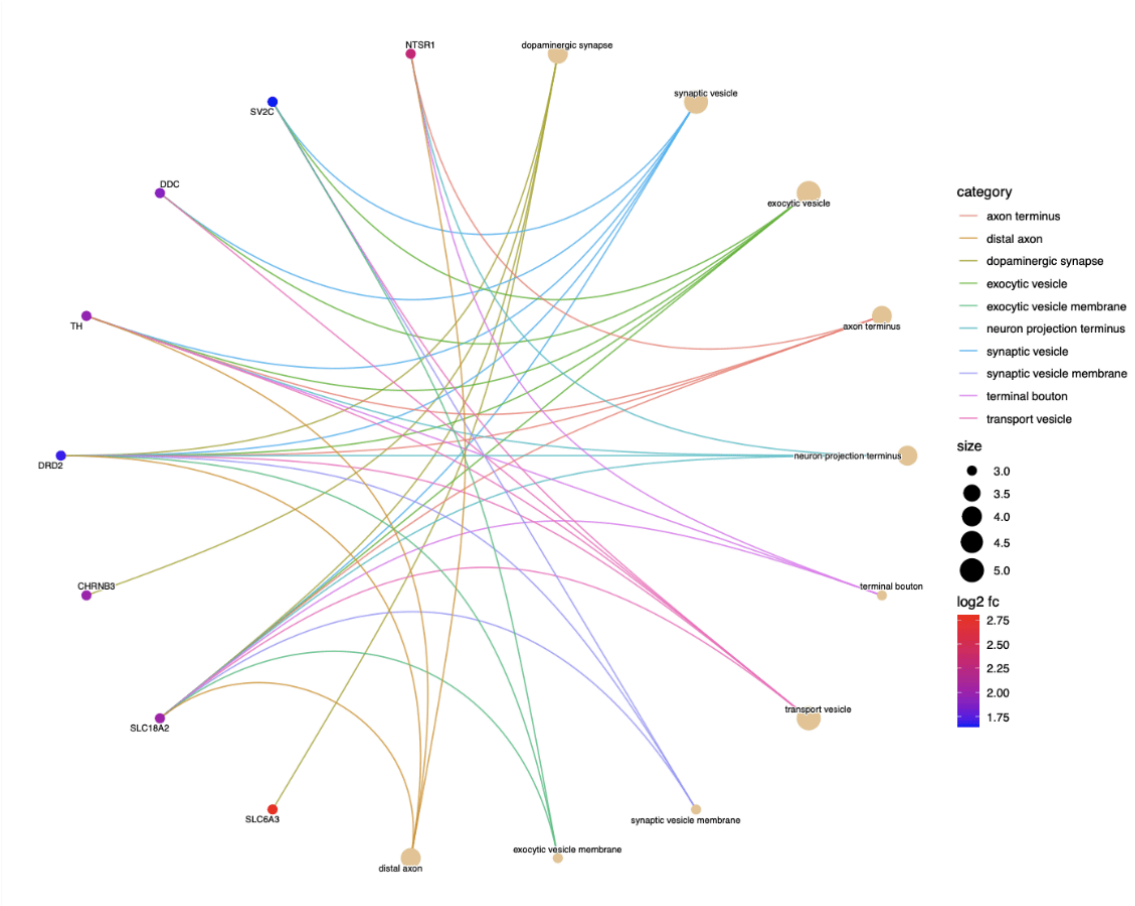
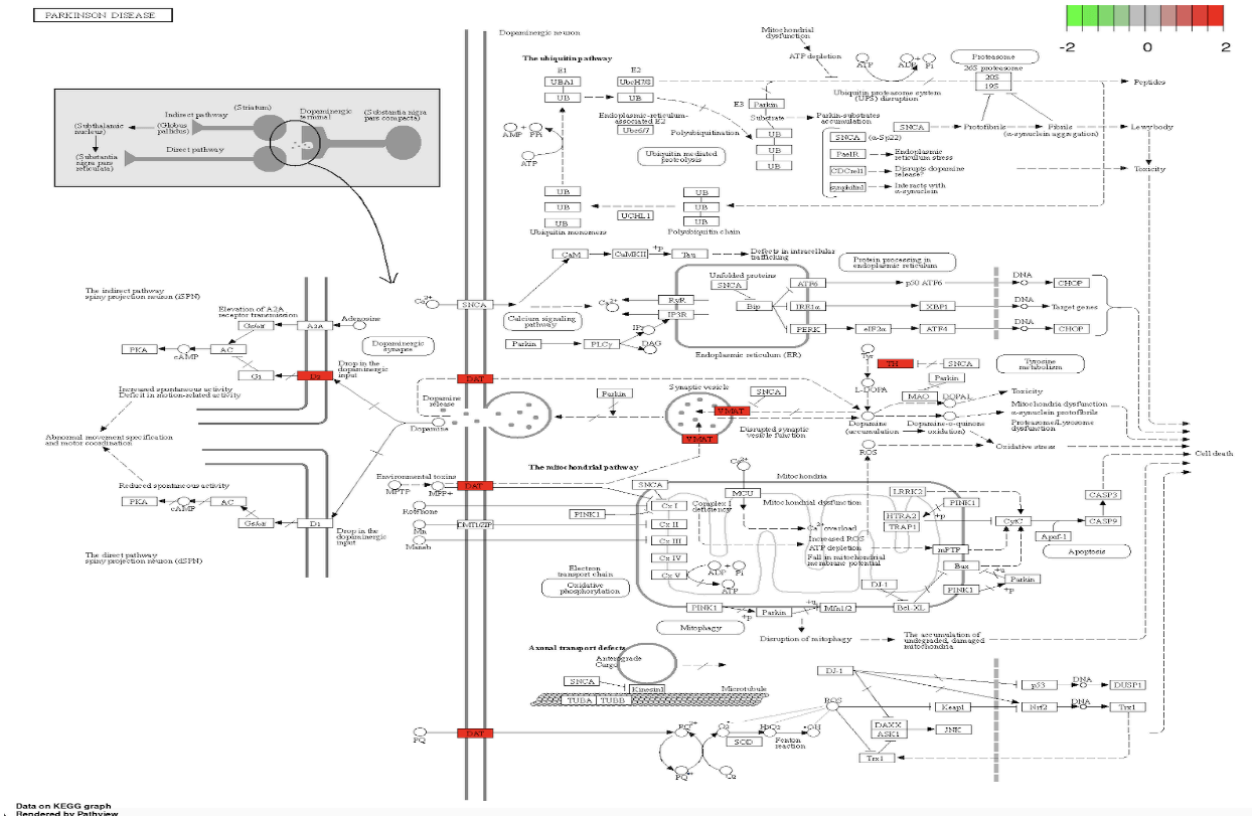


Figure 10. Cnet plot for Cellular Component. The CC cnet plot displays the genes associated with different pathways within the cellular component, showing an association between the gene and pathway by different colored lines. A higher log fold change value represents higher statistical importance, and the size represents the count/number of genes associated.

In the path view of Parkinson's disease (Figure 10), genes such as DAT, D2, VMAT, and TH are active at various stages of the disease. Some key biological pathways and functions are locomotory behavior, dopaminergic synapse, and dopaminergic neuron differentiation. Genes such as SLC6A3 and PITX3 were prevalent in this analysis.

Figure 11. Diagram of Parkinson's Pathway. The pathview above shows the pathway of Parkinson's disease and at which stage different genes are involved. The highlighted colors represent the level of enrichment associated with these genes.



<u>Supplementary Materials</u>

Supplementary Table 1: Top 30 DEGs.

Supplementary Figure 1: R Script used to generate DEGs

DISCUSSION

This study investigated the genetic factors associated with Parkinson's disease by using different bioinformatics tools. First, we compared gene expression profiles between affected individuals and healthy controls in the GEO Dataset ID GSE168496. The GEO2R analysis displayed that there were several upregulated and downregulated genes between the two groups (Figure 2). Key findings include significant differential expression in genes involved in biological and molecular pathways such as locomotory behavior, dopaminergic synapse function, and dopaminergic neuron differentiation. Notably, genes like SLC6A3, which is essential for dopamine reuptake (Reith et al. 2021), and PITX3, which plays a role in dopaminergic neuron development (Wang et al. 2023), were prevalent in this analysis. These results highlight the central role of dopaminergic pathways in Parkinson's disease and identify potential biomarkers and therapeutic targets for further investigation.

The Gene Ontology (GO) and KEGG pathway analyses reveal that the differentially expressed genes (DEGs) are significantly involved in key biological processes and pathways relevant to Parkinson's disease. The GO enrichment analysis highlighted several critical pathways, including locomotory behavior, dopaminergic neuron differentiation, and several others. These findings suggest that the upregulated DEGs are heavily involved in the dopaminergic systems and processes disrupted in Parkinson's disease. The enrichment in the dopaminergic synapse displays the importance of this cellular component in the disease. Similarly, KEGG pathway analysis identified several enriched pathways associated with dopaminergic signaling and neurodegenerative processes, aligning with the known pathology of Parkinson's disease. The presence of key genes like SLC6A3 and PITX3 in these pathways reinforces their relevance for understanding and targeting Parkinson's disease mechanisms.

This study adds to the understanding of the genetic foundations of Parkinson's disease (PD) by examining the roles of both SLC6A3 and PITX3 genes. The findings suggest that the SLC6A3 gene is associated with cognitive performance and brain activation in PD patients. This aligns with the research of Habak et al. (2014), who found that PD patients with a certain version of the SLC6A3 gene (the 9-repeat allele) showed less brain activity in the prefrontal cortex and premotor cortex compared to patients with a different version (the 10-repeat allele). Additionally, another study indicated a specific SNP (a variation in a single nucleotide) in SLC6A3 is associated with increased PD risk in Caucasians (Jiménez-Jiménez et al. 2013). They found that an SNP in the promoter region of the SLC6A3 gene is linked to a higher risk of PD in Caucasians. Other studies observed that variations in the PITX3 gene are linked to PD. Bäckström et al. (2017) found that PITX3 variants are associated with early-onset PD in Caucasians. Furthermore, Zhai et al. (2014) reported that the PITX3 C allele increases the risk of developing Parkinson's disease dementia (PDD) and visuospatial dysfunction in PD patients, as well as a higher incidence of visual hallucinations. These findings suggest that PITX3 significantly contributes to the neurobiology and genetic landscape of PD. These studies underscore the importance of considering genetic and demographic diversity when investigating PD, as the impact of genetic variants such as those in SLC6A3 and PITX3 can be highly specific to different populations. Future research

should continue to explore these variations to better understand how genetic factors interact with demographic characteristics in Parkinson's disease.

The findings of this study have significant implications for the understanding and treatment of Parkinson's disease (PD). Identifying key genetic factors, such as variations in the SLC6A3 and PITX3 genes that influence PD risk and progression can help grasp a better understanding of the underlying mechanisms of the disease. This knowledge could be critical in developing targeted therapies aimed at these specific genetic variations. For example, treatments that modulate the activity of the SLC6A3 gene could potentially enhance cognitive function and brain activity in PD patients. Additionally, a better understanding of the role of PITX3 in early-onset PD and dementia could lead to the development of preventive strategies or therapies that specifically target the pathways affected by this gene. Ultimately, these findings could lead to personalized medicine approaches, where treatments are tailored based on an individual's genetic profile, thereby improving the efficacy and outcomes of PD treatments. Moreover, these insights might contribute to the development of diagnostic tools that can identify individuals at higher risk of PD, allowing for earlier intervention and management of the disease.

Table 1. Summary

Category	**Pathway/function**	**Genes involved**
Biological Processes	Locomotory behavior	SLC6A3, PITX3
Biological Processes	Dopaminergic neuron differentiation	PITX3
Cellular Component	Dopaminergic synapse	SLC6A3
Cellular Component	Synaptic Vesicle	SLC6A3

A key limitation of the study is that it relies on RNA sequencing data from experiments conducted by other researchers. This is a concern because, while these findings point to potentially important genes and pathways, they are based on pre-existing data rather than direct experiments conducted in our own lab. Therefore, additional laboratory or clinical research is needed to validate these findings. Without conclusive validation, there is uncertainty

about the accuracy and applicability of our results for developing new treatments or vaccines.

Future research should focus on several key areas. First, the identified genes and pathways should be further investigated in the laboratory through experimental studies to confirm their roles in Parkinson's disease. Additionally, clinical trials should be conducted to assess how targeting these genes might influence disease progression or symptom management in patients. By translating our bioinformatics findings into experimental and clinical settings, researchers can determine the practical implications of these genes and pathways for developing targeted therapies or preventive strategies for Parkinson's disease.

Conclusion

This study aimed to identify differences in the expression of genes of individuals with Parkinson's Disease (PD) and healthy control patients. The hypothesis was that certain genes would exhibit differences in expression between these two groups. The analysis, using GEO2R and SRPlot, confirmed the hypothesis, and displayed several differentially expressed genes associated with biological pathways and cellular components involved in PD. Genes such as SLC6A3 and PITX3 were linked to pathways such as locomotory behavior, dopaminergic synapse, and dopaminergic neuron differentiation, highlighting their potential roles in the disease mechanism. These findings contribute to our understanding of the genetic factors underlying Parkinson's disease and suggest new molecular targets for therapeutic intervention. Future research should focus on validating these DEGs in larger, diverse populations and exploring the interactions between genetic and environmental factors in PD. Further studies could also investigate the therapeutic potential of targeting these genes and pathways to develop more effective treatments for Parkinson's disease.

References

1. About GEO2R – GEO. NCBI, https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html. Accessed 11 July 2024.
2. Alexander, G. E. "Biology of Parkinson's Disease: Pathogenesis and Pathophysiology of a Multisystem Neurodegenerative Disorder."

Dialogues in Clinical Neuroscience, 2004, pp. 259–280. https://doi.org/10.31887/dcns.2004.6.3/galexander.
3. Anjum, A., Jaggi, S., Varghese, E., Lall, S., Bhowmik, A., & Rai, A. (2016). Identification of Differentially Expressed Genes in RNA-seq Data of Arabidopsis thaliana: A Compound Distribution Approach. Journal of Computational Biology, 23(4), 239–247. https://doi.org/10.1089/cmb.2015.0205
4. Bäckström, D., Domellöf, M. E., Granåsen, G., Linder, J., Mayans, S., Elgh, E., Mo, S. J., & Forsgren, L. (2017). PITX3 genotype and risk of dementia in Parkinson's disease: A population-based study. Journal of the Neurological Sciences, 381, 278–284. https://doi.org/10.1016/j.jns.2017.08.3259
5. Barrett, T. "NCBI GEO: Mining Millions of Expression Profiles–Database and Tools." Nucleic Acids Research, 2004, pp. D562–D566. https://doi.org/10.1093/nar/gki022.
6. Barrett, T., Wilhite, S. E., Ledoux, P., Evangelista, C., Kim, I. F., Tomashevsky, M., Marshall, K. A., Phillippy, K. H., Sherman, P. M., Holko, M., Yefanov, A., Lee, H., Zhang, N., Robertson, C. L., Serova, N., Davis, S., & Soboleva, A. (2012). NCBI GEO: archive for functional genomics data sets—update. Nucleic Acids Research, https://doi.org/10.1093/nar/gks1193
7. Clarke, C. E. "Parkinson's Disease." BMJ, 2007, pp. 441–45. https://doi.org/10.1136/bmj.39289.437454.ad.
8. Corti, O., et al. "Inflammation and Regulatory T Cell Genes Are Differentially Expressed in Peripheral Blood Mononuclear Cells of Parkinson's Disease Patients." Scientific Reports, 2021. https://doi.org/10.1038/s41598-021-81961-7.
9. Diao, H., Li, X., Hu, S., & Liu, Y. (2012). Gene Expression Profiling Combined with Bioinformatics Analysis Identify Biomarkers for Parkinson Disease. PloS One, 7(12), e52319. https://doi.org/10.1371/journal.pone.0052319
10. Diekelmann, S., et al. "What Genetics Tells Us About the Causes and Mechanisms of Parkinson's Disease." Physiological Reviews, 2010, pp. 1075-1112. https://doi.org/10.1152/physrev.00022.2010.
11. Fanning, S., et al. "Parkinson's Disease: Proteinopathy or Lipidopathy." npj Parkinson's Disease, 2019. https://doi.org/10.1038/s41531-019-0103-7

12. Feldman, J., et al. Cognitive Neuroscience: LRRK2 Parkinson Disease. 2nd ed., W.W. Norton, 2023. https://www.ncbi.nlm.nih.gov/books/NBK1208/
13. Gasser, T. (2009). Molecular pathogenesis of Parkinson disease: insights from genetic studies. Expert Reviews in Molecular Medicine. https://doi.org/10.1017/s1462399409001148
14. Gene Ontology Consortium. "Gene Ontology Consortium: Going Forward." Nucleic Acids Research, 2015. https://doi.org/10.1093/nar/gku1179.
15. Gene Ontology overview. (2024, July 24). Gene Ontology Resource. https://geneontology.org/docs/ontology-documentation/
16. Geo. (n.d.-a). About GEO2R – GEO – NCBI. https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html
17. GEO. (n.d.-b). Home – GEO – NCBI. https://www.ncbi.nlm.nih.gov/geo/
18. Habak, C., et al. (2014). Dopamine transporter SLC6A3 genotype affects cortico-striatal activity of set-shifts in Parkinson's disease. Brain. doi: 10.1093/brain/awu251
19. Jiménez-Jiménez, F. J., et al. (2013). PITX3 and Risk for Parkinson's Disease: A Systematic Review and Meta-Analysis. European Neurology, 71(1–2), 49–56. doi: 10.1159/000353981
20. Kalbfleisch, T., et al. (2010). Genomics, bioinformatics, and computational biology. In Elsevier eBooks. doi: 10.1016/b978-0-08-046884-6.00236-0
21. Kanehisa, M. (2000). KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Research. doi: 10.1093/nar/28.1.27
22. Kanehisa, M., et al. "KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs." Nucleic Acids Research, 2017. https://doi.org/10.1093/nar/gkw1092.
23. KEGG Database. (n.d.). https://www.genome.jp/kegg/kegg1.html
24. Khan, A. U., et al. "Awareness and Current Knowledge of Parkinson's Disease: A Neurodegenerative Disorder." International Journal of Neuroscience, 2018. https://doi.org/10.1080/00207454.2018.1486837.
25. Klockgether, T. "Parkinson's Disease: Clinical Aspects." Cell and Tissue Research, 2004, pp. 115–120. https://doi.org/10.1007/s00441-004-0975-6.

26. Kowal, S. L., et al. “The Current and Projected Economic Burden of Parkinson’s Disease in the United States.” Movement Disorders, 2013, pp. 311–318. https://doi.org/10.1002/mds.25292.
27. Marras, C., et al. “Prevalence of Parkinson’s Disease across North America.” npj Parkinson’s Disease, 2018. https://doi.org/10.1038/s41531-018-0058-0
28. Nussbaum, R. L., & Ellis, C. E. “Alzheimer’s Disease and Parkinson’s Disease.” New England Journal of Medicine, 2003, pp. 1356–64. https://doi.org/10.1056/nejm2003ra020003.
29. Opara, J. A., et al. “Quality of Life in Parkinson`s Disease.” Journal of Medicine and Life, 2012, pp. 375–381. www.ncbi.nlm.nih.gov/pmc/articles/PMC3539848/.
30. Pang, S. Y., et al. “The interplay of aging, genetics and environmental factors in the pathogenesis of Parkinson’s disease.” Translational Neurodegeneration, 2019. https://doi.org/10.1186/s40035-019-0165-9
31. Rajalahti, T., et al. “Discriminating Variable Test and Selectivity Ratio Plot: Quantitative Tools for Interpretation and Variable (Biomarker) Selection in Complex Spectral or Chromatographic Profiles.” Analytical Chemistry, 2009, pp. 2581–2590. https://doi.org/10.1021/ac802514y.
32. Reith, M. E. A., et al. “The dopamine transporter gene SLC6A3: multidisease risks.” Molecular Psychiatry, 2021. https://doi.org/10.1038/s41380-021-01341-5
33. Solano, L. E., et al. “PRRGO: a tool for visualizing and mapping globally expressed genes in public gene expression omnibus RNA-Sequencing studies to PageRank-scored gene ontology terms.” bioRxiv (Cold Spring Harbor Laboratory), 2024. https://doi.org/10.1101/2024.01.21.576540
34. SRplot – Science and Research online plot. (n.d.). https://www.bioinformatics.com.cn/srplot
35. Snyder, H. R. “Understanding Neurodegenerative Diseases.” Nature Reviews Neuroscience, 2022, pp. 124–136. https://doi.org/10.1038/s41583-022-00500-1.
36. Takeda, A., et al. “Bioinformatics tools and Parkinson’s Disease.” Bioinformatics, 2020. https://doi.org/10.1093/bioinformatics/btz892.
37. Van Lieshout, K., et al. “Genetic Etiology of Parkinson’s Disease and Parkinsonism: The Emerging Role of VPS35.” Annual Review of

Neuroscience, 2018. https://doi.org/10.1146/annurev-neuro-080317-061930.
38. Vos, T., et al. “Global, Regional, and National Burden of 369 Diseases and Injuries for 204 Countries and Territories, 1990–2019: A Systematic Analysis for the Global Burden of Disease Study 2019.” The Lancet, 2020. https://doi.org/10.1016/s0140-6736(20)30925-9.
39. Wang, Y., et al. “Integrating Bioinformatics Analysis and Experimental Validation to Identify Biomarkers of Parkinson’s Disease.” Experimental and Therapeutic Medicine, 2020. https://doi.org/10.3892/etm.2020.8775.
40. Zhao, J., et al. “The Role of LRRK2 in Parkinson’s Disease Pathogenesis.” Molecular Neurodegeneration, 2020. https://doi.org/10.1186/s13024-020-00394-w.
41. Zhou, Q. “Investigating Pathological Pathways in Parkinson’s Disease through Single-Cell Transcriptomics.” Neuroscience Letters, 2022. https://doi.org/10.1016/j.neulet.2022.136923.
42. Zhu, L., et al. “The Link between Neuroinflammation and Parkinson’s Disease: What the Latest Research Shows.” Journal of Neurochemistry, 2024. https://doi.org/10.1111/jnc.15898.

Dev Sathyan

Social Isolation During the COVID-19 Pandemic and its Impact on Cardiovascular Disease

ABSTRACT

Background: Social isolation (SI) has traditionally been thought of as a physiological problem, but research has shown that it can increase the risk of cardiovascular diseases (CVDs) by 30%. People affected by SI were most likely physically inactive and resorting to unhealthy choices.

Methods: This study is a systematic literature review that reviewed studies from Pubmed. Included studies looked at COVID-19, social isolation, loneliness, cardiovascular diseases, and physical inactivity. The search excluded systematic reviews, literature reviews, and studies published after 2019.

Results: During the pandemic, people experienced high SI because of the fear of contracting COVID-19. In 2020, among young men in Poland, 20.48% said they had mild SI, 36.87% said they had moderate SI, and 42.63% said they had severe SI. With increased levels of SI, physical inactivity and participation in unhealthy habits such as smoking increased. The weight gain and unhealthy habits led to a higher risk of developing CVDs.

Discussion: The high SI during the pandemic led to an increased risk of developing CVD. Also during the pandemic, people started to have increased anxiety from the risk of contracting COVID-19.

INTRODUCTION

While social isolation (SI) is primarily thought of as a psychological issue, recent research shows that SI can increase the risk of cardiovascular disease (CVD) by 30% [1]. Those who experience more SI are more likely to experience CVD as they may resort to unhealthy lifestyle choices, including smoking and being physically inactive.

If people are physically inactive for prolonged periods, they are more likely to gain weight and face obesity[2]. Obesity can cause indirect CVDs such as hypertension and myocardial infarction [2]. Differences in physical activity may contribute to the increased risk of ill-health and poor well-being associated with SI [3]. The increased physical inactivity from SI causes

obesity, in turn, increasing the risk for CVD. Obesity is known to increase the risk of developing CVD significantly [2], and it also accelerates the rate of heart failure [2].

The study aims to summarize the current literature on the association between SI and CVDs. By understanding the impact of SI on CVDs, preventative measures, and early interventions can be considered to mitigate the risk of CVDs. This study hypothesizes that increased social isolation will lead to a higher risk of CVD, mitigated by factors such as decreased physical activity and increased unhealthy food and drink consumption.

METHODS

This study used PubMed to search for key terms, including social isolation, COVID-19, loneliness, cardiovascular diseases, and physical inactivity. As part of this search, all literature reviews, systematic reviews, and studies before 2019 were excluded. Studies prior to 2019 were excluded as COVID-19 was a predominant factor that led to the rise in social isolation.

RESULTS

SI is the leading cause of physical inactivity

Studies show that SI directly relates to physical inactivity [4]. Loneliness was associated with an increased risk of physical inactivity [4]. In Europe, a study said that highly isolated people are more likely to become physically inactive [5]. However, in other European countries, high SI resulted in an inadequate diet without fruit and vegetables [5]. Furthermore, a study showed that greater SI in the adult population is related to reduced everyday physical activity and greater sedentary time [3].

SI during the COVID-19 pandemic

The high SI from the pandemic caused people to develop mental health conditions [6]. In 2020, among young men in Poland, 20.48% said they had mild SI, 36.87% said they had moderate SI, and 42.63% said they had severe SI [7]. During the pandemic, people started becoming lonely, and they took substances which led to overdoses and deaths [6]. After the restrictions were eased mental health recovered relatively fast [6]. However,

some still need mental help to recover fully [8]. In Sao Paulo, the SI from March to May 2020 increased from 43.07% to 50.71% [8].

CVDs during Covid pandemic

A study in Sao Paulo shows that the COVID-19 pandemic caused a decrease in hospitalizations and deaths from CVDs [8]. This may have been due to the fear of contracting COVID-19 from the hospital [8]. A study in Brazil said that hospital admissions decreased but the in-hospital deaths from CVDs increased [9]. This may have been due to inadequate CVD management planning [10]. Another study in the U.S. says that contracting COVID-19 increases the risk of developing CVD [10]. People infected who survived the first 30 days had an increased risk of inflammatory heart disease [10]. Additionally, another study recorded a substantial monthly drop in CVD hospitalizations by 42.60% from March to May 2020, during which social isolation increased by 7.64% (43.07% to 50.71%) [8].

DISCUSSION

This paper investigates the connection between SI, during and after the COVID-19 pandemic, and CVDs. A paper said SI during the pandemic caused people's blood pressure to rise [11]. Also in the same paper, it noted that lonely people had a 30% higher risk of getting CVDs [11]. Loneliness causes people to participate in unhealthy activities such as being physically inactive[3]. Because the people were physically inactive they started gaining weight, which in turn increased the individual's risk of developing CVD. During the pandemic, people started having increased anxiety [7]. Some people were isolated from their families and began to develop the first stages of depression. Also, people who developed CVD didn't go to the hospital to seek medical care to decrease the chances of getting COVID-19, which only worsened their CVD conditions and were most likely fatal at times[8]. This study has some limitations such as using data from other countries. Another limitation is different countries have COVID-19 pandemic restrictions and

This study shows how SI during the pandemic caused an increased risk of CVDs. The findings from this study may be used by future researchers to investigate the relationship between SI and CVD outcomes on a global scale.

References

1. Hakulinen, C., Pulkki-Råback, L., Virtanen, M., Jokela, M., Kivimäki, M., & Elovainio, M. (2019). Coronary artery disease. BMJ Journals, 104(18). https://doi.org/10.1136/heartjnl-2017-312663
2. Ashraf, M. J., & Baweja, P. (2013). Obesity: The "Huge" Problem in Cardiovascular Diseases. Missouri Medicine, 110(6), 499–504. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6179812/
3. Vasan, S., Lambert, E., Eikelis, N., & Lim, M. H. (2022). Impact of loneliness on health-related factors in Australia during the COVID-19 pandemic: A retrospective study. Health & Social Care in the Community. https://doi.org/10.1111/hsc.13948
4. Hawkley, L. C., Thisted, R. A., & Cacioppo, J. T. (2009). Loneliness predicts reduced physical activity: Cross-sectional & longitudinal analyses. Health Psychology, 28(3), 354–363. https://doi.org/10.1037/a0014400
5. Delerue Matos, A., Barbosa, F., Cunha, C., Voss, G., & Correia, F. (2021). Social isolation, physical inactivity and inadequate diet among European middle-aged and older adults. BMC Public Health, 21(1). https://doi.org/10.1186/s12889-021-10956-w
6. Jeffers, A., Meehan, A. A., Barker, J., Asher, A., Montgomery, M. P., Bautista, G., Ray, C. M., Laws, R. L., Fields, V. L., Radhakrishnan, L., Cha, S., Christensen, A., Dupervil, B., Verlenden, J. V., Cassell, C. H., Boyer, A., DiPietro, B., Cary, M., Yang, M., & Mosites, E. (2022). Impact of Social Isolation during the COVID-19 Pandemic on Mental Health, Substance Use, and Homelessness: Qualitative Interviews with Behavioral Health Providers. International Journal of Environmental Research and Public Health, 19(19), 12120. https://doi.org/10.3390/ijerph191912120
7. Justyna Mojsa-Kaja, Klaudia Szklarczyk-Smolana, Niedzielska-Andres, E., Kurpińska, A., Suraj-Prażmowska, J., & Walczak, M. (2023b). COVID-19-related social isolation and symptoms of depression and anxiety in young men in Poland: Does insomnia mediate the relationship? PLOS ONE, 18(5), e0285797–e0285797. https://doi.org/10.1371/journal.pone.0285797
8. Lima, L. R. de, Ponte, P. F. A. da, Dias, L. N., Silvestre, M. H. L., Suen, P. J. C., & Mansur, A. de P. (2022a). Social Isolation, Hospitalization, and Deaths from Cardiovascular Diseases during the COVID-19 Epidemic in São Paulo Metropolitan Area in 2020. International Journal

of Environmental Research and Public Health, 19(17), 11002. https://doi.org/10.3390/ijerph191711002

9. Normando, P. G., Araujo-Filho, J. de A., Fonseca, G. de A., Rodrigues, R. E. F., Oliveira, V. A., Hajjar, L. A., Almeida, A. L. C., Bocchi, E. A., Salemi, V. M. C., & Melo, M. (2021). Redução na Hospitalização e Aumento na Mortalidade por Doenças Cardiovasculares durante a Pandemia da COVID-19 no Brasil. Arquivos Brasileiros de Cardiologia. https://doi.org/10.36660/abc.20200821
10. Xie, Y., Xu, E., Bowe, B., & Al-Aly, Z. (2022). Long-term cardiovascular outcomes of COVID-19. Nature Medicine, 28(28), 1–8. https://doi.org/10.1038/s41591-022-01689-3
11. Bu, F., Zaninotto, P., & Fancourt, D. (2020). Longitudinal associations between loneliness, social isolation, and cardiovascular events. Heart, 106(18), 1394–1399. https://doi.org/10.1136/heartjnl-2020-316614

Sana Sidhu

How Do Social Isolation and Loneliness Affect Cardiac Health?

ABSTRACT

Background: Social isolation and loneliness have become increasingly prevalent in the United States, presenting as significant cardiovascular risks including coronary heart disease and strokes. These conditions are associated with increased mortality rates and share similarities with established risk factors such as smoking and alcoholism. This study aims to explore how social isolation and loneliness impact our cardiovascular system, in turn addressing gaps in our current knowledge so strategies can be created to reduce the risk of heart disease nationwide.

Methods: A scoping review using PubMed and Google scholar was conducted with the following key terms: social isolation, cardiovascular health, depression, loneliness, heart risks, and heart failure. Articles included looked at the impact of social isolation and loneliness on both genders specifically related to heart health.Literature reviews, single-gender studies, and articles focusing on overall health effects unrelated to cardiac health were excluded

Results: Social isolation and loneliness significantly increase the chances of heart failure and cardiovascular diseases, with socially isolated individuals having a 17% higher risk of heart failure and a 42% higher risk of developing heart conditions. Loneliness affects a large portion of the population, ranging from 5% to 30%, and is associated with a 29% higher risk of heart disease and a 32% higher risk of stroke. Social isolation and loneliness also lead to significant changes in weight and higher blood pressure levels.

Discussion: Social isolation significantly impacts the body's systems, affecting heart rate and blood pressure. It increases the risks of cardiovascular diseases like heart failure and strokes, and doubles mortality chances. While both genders are affected, females tend to experience more severe effects. Addressing social isolation and loneliness can decrease the risk of heart failure, regardless of a patient's genetic background. Promoting programs to mitigate the impact of social isolation and loneliness could increase social support and improve heart health outcomes.

INTRODUCTION

Social isolation has been a prevalent issue in America in the past few years [1]. The uptick in social isolation and loneliness is affecting our health. Loneliness has been seen as an important risk factor of developing coronary heart disease and stroke [2]. Those with weak social connections are at a 29% increase of developing a heart disease and at a 32% increase of a stroke [2,3]. As such, social isolation and loneliness are important to pay attention to in cardiovascular health.

There is a clear correlation between social isolation, loneliness, and increased mortality rates [4]. These social determinants are comparable with well-established risk factors of death such as drinking alcohol and smoking [4]. Feeling lonely and being socially isolated can increase the risk of coronary heart disease (CHD) even for those with no other risk factors [2]. Hypertension, known to lead to numerous heart related issues, is increased by loneliness and social isolation as well [5,6]. Additionally, loneliness has a direct impact on health behaviors, such as an impact on sleep quality [7]. By understanding factors leading to loneliness and social isolation as well as investigating the impact of this isolation on cardiovascular health, proactive initiatives can be implemented to lower the percentages of heart disease across America.

Social isolation and loneliness are known factors that have a clear impact on cardiac health [2-4], however reasons behind why people experience loneliness and social isolation and their impact on cardiac health are not well-known. The purpose of this study is to address this gap by examining evidence on social isolation and how it leads to hypertension and increases mortality rates.

METHODS

Search Strategy

PubMed and Google scholar were used to search the academic literature. The following key terms were searched: “social isolation”, “cardiovascular health”, “depression”, “loneliness”, “heart risks”, and “heart failure”.

Inclusion and Exclusion Criteria

The inclusion criteria focused on articles that included isolation, loneliness, male and female genders, and how social isolation/loneliness affects the heart. Articles were excluded if they were literature reviews, only about one gender, or looked at the impact on overall health as opposed to heart health specifically.

RESULTS

There is a direct relationship between social isolation/loneliness and heart failure (HF) [8]. Those who were socially isolated had a 17% risk of HF, and those who felt lonely also had a 19% risk of HF [8]. Those who are lonely and socially isolated are also 42% more likely to develop cardiovascular diseases (CVD) [9]. About 5-30% of our population are lonely, risking their heart health [10]. Social isolation and loneliness are also associated with a 29% increase in the risk of heart disease and a 32% risk of stroke [11]. To put these statistics into context, loneliness and social isolation are linked to twice the risk of death from CVD. [9].

Social isolation and loneliness also affect weight, as males tend to lose weight (when tested, 23% less in 4 weeks) and females tend to gain weight (when tested, 90% more in 4 weeks) [11]. Systolic blood pressure increases by 12% in males and 22% in females due to social isolation and loneliness [11]. Those who have reported being socially isolated or lonely are also at a 66% risk of incident CVD [9].

Table 1: Corresponding percentage increase in risk of type of heart disease among those who are lonely or socially isolated

HEART DISEASE	**INCREASED RISK**
Heart Failure (socially isolated)	17%
Heart Failure (loneliness)	19%
Cardiovascular diseases	42%
Heart Disease	29%
Stroke	32%
Systolic blood pressure (males)	12%

Systolic blood pressure (females)	22%

DISCUSSION

Social isolation impacts the autonomic nervous system which controls functions like heart rate and blood pressure [11]. It increases the risk of CVD, including HF and strokes, and doubles the chance of mortality [4]. Social isolation affects both genders similarly, but females tend to experience more severe effects [11]. This could be due to the different hormones females have, and how they feel emotions more strongly compared to males. Females gain a lot of weight, while males lose some weight [11]. Social isolation raises blood pressure in both genders, with females often showing a more significant increase, highlighting a greater impact on women [11].

There is a pattern where the more socially isolated or lonely an individual is, the more likely they are to experience HF [8]. Trying to decrease social isolation or loneliness can decrease the chances of HF, even if you have a genetic risk for that condition [2]. Policy-makers and hospital administration interested in improving heart health outcomes should consider funding programs that increase social support networks. Healthcare providers should consider talking to their patients about feelings of social isolation and loneliness as part of a routine checkup. Helping those who are isolated or lonely may help reduce the risks of HF, CVD, and mortality.

Social desirability bias may exist, resulting in an underreporting of true feelings of social isolation and loneliness among the participants. There could have been other confounding factors as well, such as socioeconomic status or lifestyle factors, that could have contributed to the risk of CVD.

Based on the findings of this scoping review, it’s important to be aware of the effects of social isolation and loneliness on cardiovascular health, and to prioritize initiatives both in our community and the medical system that offer help to those who are feeling socially isolated or lonely.

References

1. Holt-Lunstad J, Smith TB, Layton JB. Social relationships and mortality risk: A meta-analytic review. Annual Review of Sociology. 2010;35:87-109.
2. Ding D, Del Pozo Cruz B, Green MA, Bauman AE. Physical activity and sedentary behavior among children and adolescents: A latent class analysis approach. European Journal of Preventive Cardiology. 2020;25:1387-1398.
3. Santini ZI, Jose PE, Cornwell EY, Koyanagi A, Nielsen L, Hinrichsen C, Meilstrup C, Madsen KR, Koushede V. Social disconnectedness, perceived isolation, and symptoms of depression and anxiety among older Americans (NSHAP): a longitudinal mediation analysis. The Lancet Public Health. 2020;5(1)
4. Steptoe A, Shankar A, Demakakos P, Wardle J. Social isolation, loneliness, and all-cause mortality in older men and women. PLOS Medicine. 2013;10(8)
5. Katz JN, Chang LC, Sangha O, Fossel AH, Bates DW. Can comorbidity be measured by questionnaire rather than medical record review? Journal of Clinical Epidemiology. 1996;49(1):11-18.
6. Ginsburg GS, Willard HF. Genomic and personalized medicine: Foundations and applications. The New England Journal of Medicine. 2009;360:1811-1818.
7. Hawkley LC, Cacioppo JT. Loneliness and health: Potential mechanisms. Psychosomatic Medicine. 2002;64(3):407-417.
8. National Institute on Aging. (2023). Loneliness and Social Isolation Linked to Serious Health Conditions. National Institutes of Health, 6, 1-8.
9. National Institute on Aging. (2022). The Impact of Social Isolation on Cardiovascular Health: A Meta-Analysis. National Institutes of Health, 9, 1-12.
10. Kachan, D. (2021). The Impact of Social Isolation on Cardiovascular Health: A Meta-Analysis. Journal of Health Psychology, 23, 1-15.
11. Chen, Y. (2021). The Impact of Social Isolation on Mental Health during the COVID-19 Pandemic. Frontiers in Psychology, 12, Article 609157.

Kavya Singh

Dissociative Identity Disorder and Broken Memories

ABSTRACT

Background: Dissociative identity disorder (DID) significantly impacts memory processing and brain function, leading to memory gaps, amnesia, and separate neural activation patterns between alters.

Objectives: This study explores the cognitive and neurological impacts of DID, specifically on how different alters display different brain activity and how these differences alter and prevent memory encoding, recall, and integration between alters.

Methods: By using PubMed and Google Scholar, a total of 5 studies were reviewed for this paper.

Results: The studies used in this paper reveal that traumatic alters often display heightened activity in areas that deal with sensory and motor processing, while normal alters use different neural patterns altogether. The impaired memory transfer between alters results in memory gaps and inconsistencies, which have been found to be caused by structural and functional brain differences.

Conclusion: This research highlights the impact that DID has on memory and neural processes, highlighting the need for continued research and new, more effective treatments to better the lives of DID patients.

INTRODUCTION

Dissociative identity disorder (DID) patients regularly experience memory gaps and fragmentation, along with cognitive impairments, making their daily lives harder than they already are [1].

DID patients often have two distinct personality types: normal and emotional or traumatic [2]. The normal alter commonly functions during “day-to-day” life, while the traumatic alter is associated with memories, emotions, and experiences of past trauma [2]. Despite having two personality types, it is common for DID patients to have more than two alters [2].

Historically, dissociative identity disorder has been overlooked, misunderstood, and not thought of as a mental health disorder [1, 2]. However, due to recent technological advancements, such as MRIs, fMRIs,

and other brain imaging techniques, evidence regarding brain differences in neural pathways between alters, and DID vs regular patients has come to light [5]. Evidence shows that a major brain difference between DID and normal patients is the size of their hippocampus -the region of the brain associated with memory function [4]. It has been found that the hippocampus of a DID patient's brain is significantly smaller than that of a normal patient's brain, suggesting a major reason for a loss of memories as well as memory fragmentation and gaps [4].

This study aims to synthesize the findings of multiple research papers to explore how dissociative identity disorder impairs memory processes and brain function. Specifically, it looks at the different neural patterns behind alters as well as the impact of DID on the encoding and recall processes of memories and memory fragmentation. By using a combination of fMRI studies, clinical memory tests, and patient reports, this paper looks into the neurological and psychological basis of DID. These findings aim to improve the understanding the general public has, reduce stigma, and aid researchers and scientists in improving the lives of those who suffer from dissociative identity disorder.

METHODS

This study was a systematic review of studies that explored how DID impacts memory. Both Google Scholar and Pubmed were used to find information for this paper. Keywords were DID, memory loss, amnesia, and memory fragmentation. English articles that focus on memory fragmentation in relation to DID were referenced in this paper while non-English articles and review articles were excluded.

RESULTS

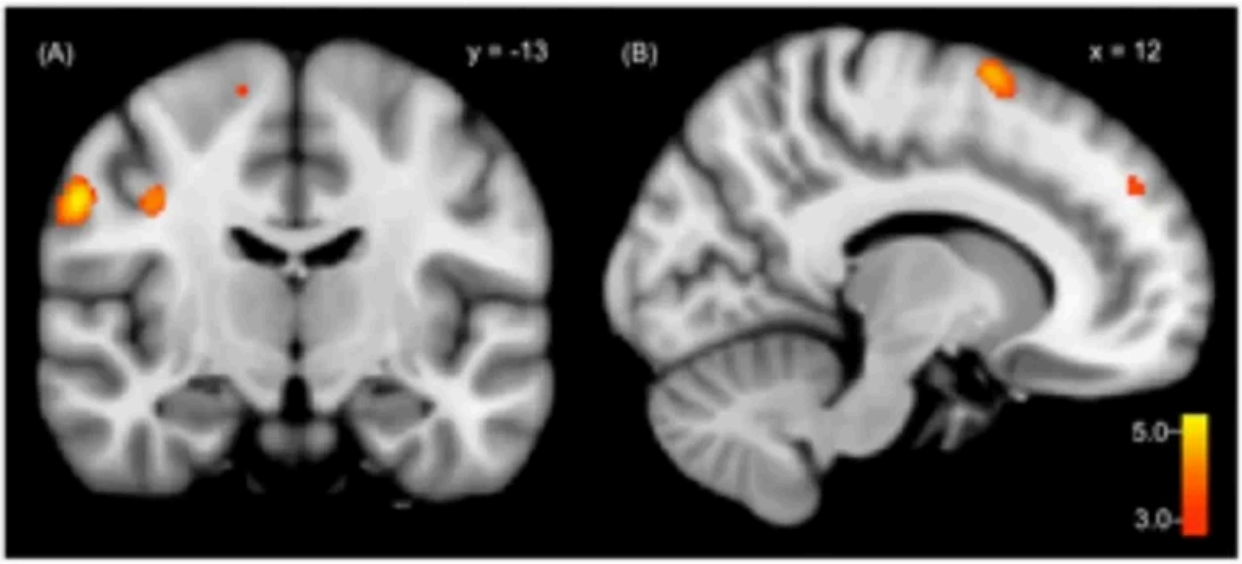
Table 1 summarizes the key findings, main idea, and importance of each study on the impact of DID on memories. These studies were used as they focused on brain pattern activation between different alters, memory transfer, as well as fMRI scans of DID patients, all of which are necessary to look at when trying to understand how DID impacts memory processing.

Table 1. Summary of studies on memory and brain function in patients with dissociative identity disorder (DID)

Study	Focus	Key Findings	Relevance to Research Question
Reinders et al. (2003) [3]	Patterns of brain activation in alters	Brain activity between traumatic and normal alters greatly differs, suggesting emotional responses to traumatic memories also differ between types of alters.	Helps to explain how DID alters impact and often leads a different memory processing, affecting storage and recall
Peters et al. (1998)[4]	Memory transfer between alters	DID patients displayed significant memory impairment when trying to recall memories and when switching between alters.	Displays that memory does not transfer properly between alters, adding to memory gaps
Schlumpf et al. (2013) [5]	Resting-state brain activity in patients with DID	Identified patterns of brain activation in normal and emotional alters in response to external stimuli	Displays how different DID alters utilize different parts of the brain, potentially impacting memory recall and processing
Vissia et al. (2022)[2]	Lack of neural activity during memory-related tasks	DID patients experience less activation in the parietal lobe when completing tasks related to	Explains how there is a lack of brain activation in areas that aid in memory encoding in patients with DID, leading to an overall lack of memories.

		attention and memory	
Van der Hart et al. (2005)[1]	Loss of memory in DID patients	DID patients often experience memory gaps or amnesia in between alters	Supports the argument that memory loss is a key DID symptom

Figure 1: fMRI of DID brain activity



fMRI scans display a significant increase in regional cerebral blood flow in the emotional or traumatic alters' brains compared to the normal alters brains. The emotional alter displays increased activation of areas that deal with processing sensory information and motor activity while the normal one appears stable [5].

The studies used in this review paper highlight the major impact that dissociative identity disorder has on memory and brain function, particularly on how DID patients recall and process memories. Table 1 displays how the research supports the theory that memory loss, gaps, and fragmentation are all incredibly common in patients with DID. The first study explained that brain activity significantly differs between traumatic (emotional) and normal alter, with traumatic alters displaying lower amounts of regional cerebral blood flow when exposed to traumatic memories, events, and ideas, indicating a changed pattern of memory processing [3]. This supports the

hypothesis that the areas of the brain that are involved in memory recall are altered when switching between alters, causing memory gaps [3].

Similarly, Peters et al. (1998) found that patients with DID struggled with transferring memories between alters [4]. These findings highlight the difficulty in memory combination and integration across alters and further explain the amnesia-like symptoms that patients with DID often report. This aligns with the findings of Van der Hart et al. (2005), who saw that memory fragmentation is a key symptom of DID, with patients being unable to recall significant life events, usually during dissociation [1].

When specifically looking at brain activity, Schlumpf et al. (2013) provided evidence that different alters utilize different regions of the brain [5]. The traumatic or emotional alter had a more active prefrontal cortex and sensory-motor areas of the brain, suggesting that the traumatic alters may rely more on these regions, possibly because of heightened levels of anxiety due to past trauma [5]. On the other hand, the normal alter showed greater activation in the thalamus, which is used for sensory and motor processing, indicating that the different brain areas may impact memory recall in the presence of an alter [5]. Further research similarly displayed that parts of the brain that are used to encode, process, and create memories did not activate in DID patients, overall leading to a lack of memory formation and, therefore, memory gaps and amnesia [2].

All the evidence from these studies goes to show that DID disrupts memories, processing, and recall within DID patients [4]. Not only are there differences in the activation of the brain, but the transfer of information between alters is not done properly, leading to memory fragmentation as well as gaps [4]. These findings further support the idea that DID is a mental health disorder that impacts both neural and memory processes.

DISCUSSION

This review synthesizes findings from multiple studies that look into the cognitive and neural functions behind Dissociative Identity Disorder (DID), specifically looking into memories and neural differences between alters. Together, these studies highlight the biological and physiological basis of DID and provide information regarding the massive impact it has on memory, brain activity, and transition between alters.

Neural Differences Between Alters

Studies have displayed distinct patterns of regional cerebral blood flow (rCBF) in traumatic and normal alters when responding to emotion-provoking stimuli [3]. Specifically, the traumatic alters exhibited reduced rCBF levels when exposed to a traumatic story compared to the normal alter [3]. The findings state that specialization within the different regions of the brain allows for different alters to process emotional memories in different ways [3]. The reduced rCBF in trauma-related alters suggests the presence of coping mechanisms put in place to prevent emotional distress [3]. It also supports the hypothesis that parts of the brain related to emotions and emotional regulation, such as the prefrontal cortex and limbic system, are activated differently depending on the alter's role and context [3].

Similarly, in Schlumpf et al.'s study, researchers backed up these findings as they revealed different neural activation patterns of alters as they reacted to external stimuli [5]. The emotional alter had increased activity in the prefrontal cortex, somatosensory cortex, and motor areas, while the normal alter had elevated perfusion in the thalamus [5]. The difference in brain activation suggests that alters are not just made by the subconscious mind but also have a distinct neurological backing for each one [5]. The increased activity in the temporal lobe of DID patients also indicates long-lasting anxiety that is often associated with their trauma [5].

Memory Gaps and Impairment

Peters et al. explored the struggles that many DID patients face when it comes to memory transfer between alters [4]. The study explained that while normal or control patients could easily complete memory tests, DID patients had trouble with the same tasks, specifically when the tested information was encoded by one alter and then asked of another [4]. This suggests that DID patients do not properly transfer memories between alters, supporting the notion that memory encoding and retrieval in DID patients is state-dependent and potentially affected by the smaller-than-average hippocampus and other brain structures [4].

These findings align with the idea that DID patients failed to activate the parietal lobe (the part of the brain that is needed for attention and working

memory) during memory tasks [2]. Instead, DID brains relied on newly created, different neural pathways, leading to poor performance on memory-related tests and, therefore, proof of memory gaps [2]. This discovery further supports the idea that normal brains and DID brains have a number of differences, and the specific differences in DID brains have led to memory fragments [2].

Additionally, it has been found that memory fragmentation is a common symptom of DID, with participants reporting memory gaps when it comes to recalling traumatic events and life experiences [1]. These memory gaps are likely due to dysfunction in the hippocampus, as a lack of hippocampus size has been known to be a cause of an inability to form proper memories [1]. These findings, specifically the fact that over 90% of participants in this study reported amnesia, display the significance and the commonality of amnesia in DID patients [1].

Integration of Findings

All of these studies convey the idea that the dissociations experienced in DID are seen through both neural activity and psychological symptoms. The activation of different regions of the brain by different alters highlights the connection between trauma, multiple identities, and memory [4]. The studies suggest that the brain creates coping mechanisms that have led to the formation of alters, which have new specialized neural pathways that support the different alters at the cost of properly integrated memory.

Relevance

These findings are significant in both the world of medicine and neuroscience. By noting the neurological background of DID, these studies challenge the notion that DID is only a psychological disorder by proving that it has a number of neurological impacts. This knowledge can help advance treatment methods and overall increase the public understanding of DID. For this reason, understanding the neural systems that back DID and its symptoms can aid in improving the lifestyle of DID patients.

Future Pathways

Despite the progress made by these studies, several gaps remain in the research. Due to the types of studies and the rarity of DID, the sample sizes for these experiments were small, resulting in generalizations, and

conclusions to be made for all DID patients. Future research should involve a large volume of participants, allowing for diversity and therefore generalization. Long-term studies such as longitudinal studies are necessary to better understand how the neural and cognitive patterns explored above change, mature and evolve over time. These studies should specifically look at the impacts that therapy has on the symptoms of DID patients as well as the inner workings of their minds.

Conclusion

These studies all highlight the impact DID has on memory and neural function, with evidence supporting the notion that neural activity is specific to a patient's state and alter, therefore leading to impaired memory integration between alters. These findings display the need for continuous research and studies to further our understanding of DID, allowing us to come up with better, more targeted therapies for patients. By using both psychology and medicine, these studies will be able to improve the standard of living and overall lives of those living with DID and other trauma-related disorders.

References

1. Van der Hart PhD, HildeBolt MA & Bessel A.van der KolkMD (2005) Memory Fragmentation in Dissociative Identity Disorder, Journal of Trauma & Dissociation, 6:1, 55-70, DOI: 10.1300/J229v06n01_04.
2. Vissia EM, Lawrence AJ, Chalavi S, et al. Dissociative identity state-dependent working memory in dissociative identity disorder: a controlled functional magnetic resonance imaging study. BJPsych Open. 2022;8(3):e82. doi:10.1192/bjo.2022.22
3. Reinders AA, Nijenhuis ER, Paans AM, Korf J, Willemsen AT, den Boer JA. One brain, two selves. Neuroimage. 2003;20(4):2119-2125. doi:10.1016/j.neuroimage.2003.08.021
4. Peters ML, Uyterlinde SA, Consemulder J, van der Hart O. Apparent amnesia on experimental memory tests in dissociative identity disorder: an exploratory study. Conscious Cogn. 1998;7(1):27-41. doi:10.1006/ccog.1997.0323.
5. Schlumpf YR, Nijenhuis ER, Chalavi S, Weder EV, Zimmermann E, Luechinger R, La Marca R, Reinders AS, Jäncke L. Dissociative

part-dependent biopsychosocial reactions to backward masked angry and neutral faces: An fMRI study of dissociative identity disorder. NeuroImage: Clinical. 2013 Jan 1;3:54-64

Surya Srinivasan

Exploring Genetic Mechanisms of Type 2 Diabetes Using Bioinformatics Tools and Databases

ABSTRACT

Background: Type 2 diabetes occurs when the body can not properly use insulin. This causes high blood glucose levels, which can further lead to serious side effects like kidney failure, neuropathy, and cardiovascular disease. Although there are multiple genes found to be linked to Type 2 diabetes, it is still unknown how exactly these genes contribute to the condition. Thus, to create more treatments, it is crucial to comprehend the precise genetic changes and pathways associated with Type 2 diabetes. The main objectives of this research are finding differentially expressed genes (DEGs) linked to Type 2 diabetes and understanding the pathways these genes affect.

Methods: To conduct this research, various bioinformatics tools and databases were used. First a dataset for a study on type 2 diabetes was found using the NCBI GEO2R bioinformatics Database. GSE271700 was the dataset found to match the disease. In the dataset, the samples were divided into two groups: diabetic and non-diabetic. GEO2R analysis was conducted to identify the genes that were expressed differently between diabetic and non-diabetic samples. The resulting differentially expressed genes were then refined using statistics by using the highest and lowest 15 LogFc values, to get the upregulated and down regulated genes. Finally for the functional annotation and pathway analysis of the DEGs, we used the ShinyGo bioinformatics tool.

Results: The study of the type 2 diabetes dataset identified 512 DEGs, with 273 upregulated and 239 downregulated genes, which got narrowed down to 30 total genes. From this dataset genes such as EPHA4, COL11A2, KITLG, EPHA2, and FLT1, were the top expressed genes found using KEGG. Furthermore, Gene Ontology (GO) enrichment analysis using ShinyGo showed enriched biological processes such as regulation of cell proliferation, response to insulin stimulus, and glucose homeostasis. Meanwhile, KEGG pathway analysis identified key pathways including the hematopoietic cell lineage, phospholipase D signaling, and insulin signaling pathways. All of these findings provide insights into the complex mechanisms underlying Type 2 diabetes.

Conclusion: To summarize, the results identified were DEGs between healthy and unhealthy samples. The top DEGs were highly enriched or associated with biological pathways. The Melanogenesis pathway with SCF gene significantly upregulated (red). The RAS signaling pathway with OF gene significantly upregulated (red). Finally the RAPI Signaling pathway with OF gene significantly upregulated (red). These specific genes and pathways involved can be studied further in the lab to help better understand the complexity of the disease and potential cures.

KEYWORDS: Type 2 diabetes, Bioinformatics, NCBI GEO2R, ShinyGo, Gene Ontology, KEGG

INTRODUCTION

Type 2 diabetes is a disease where the body can not use insulin properly (Helmholtz Association of German Research Centres, 2024). This is a big health problem that affects many people and can cause other serious issues like heart disease and nerve damage (Helmholtz Association of German Research Centres, 2024). My research aims to find out how changes in gene activity related to type 2 diabetes, so we can find new ways to treat diabetes. The question this research study is asking is, what changes in gene activity occur in people with type 2 diabetes, and how can these changes help us find new treatments?

Since the body can no longer produce enough insulin due to resistance, Type 2 diabetes causes additional health issues and affects a large number of people (Grandl et al., 2024). Although certain genes have been connected to type 2 diabetes (Dayeh et al., 2014; Kang et al., 2024), the exact way in which these genes cause the illness is still unknown (Gottlieb & Dyer, 2020).

So far what is known is that Type 2 diabetes is caused by insulin resistance and problems with insulin production (Xue et al., 2023). It affects the body's ability to manage blood sugar and can lead to serious health issues (Dye & Maunakea, 2024).

Studies have shown that certain pathways related to insulin and glucose metabolism are disrupted in type 2 diabetes (Lloyd et al., 2023; Kang et al., 2024). But, the specific genetic factors that lead to diabetes are unknown (Gottlieb & Dyer, 2020).

The main goal of this study is to find different genes that when expressed are correlated with Type 2 Diabetes. The genes associated with diabetes, found from this study, could help manage type 2 diabetes.

METHODS

Data Collection

To conduct this research, various bioinformatics tools were employed to analyze and interpret the data effectively **(Figure 1)**. Initially, a suitable dataset for the study on type 2 diabetes was identified using the NCBI GEO2R bioinformatics database (About GEO2R – GEO, 2024). The dataset, identified as GSE271700, contained comprehensive data relevant to the disease. In this dataset, the samples were categorized into two groups: diabetic and non-diabetic, allowing for a comparative analysis.

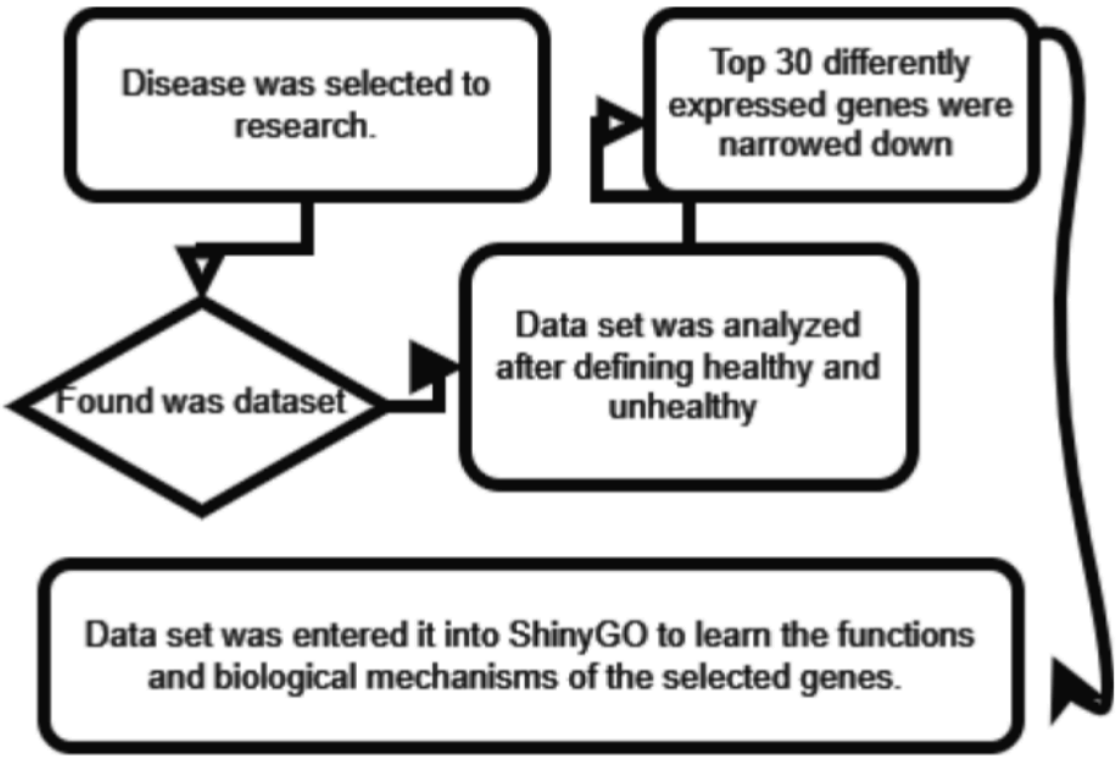


Figure 1. Flow chart indicated all the steps required to get data used for this research.

Identification of Genes Expressed Differently between Diabetic and Non-Diabetic Sample

The first step involved using the GEO2R bioinformatics analysis tool to determine which genes were differentially expressed (DEGs) between the diabetic and non-diabetic samples. This bioinformatics tool facilitated the identification of genes with significant differences in expression levels. DEGs areDifferentially expressed genes are genes that show statistically significant differences in expression levels between two or more conditions, such as between diabetic and non-diabetic samples (NCBI, 2015; Bioinformatics, 2019).

Following the identification of differentially expressed genes (DEGs), statistical methods were applied to refine the results into top 50 DEGs. Specifically, the highest and lowest 15 LogFc (log fold change) values were selected to pinpoint the most significantly upregulated and downregulated genes. This step was crucial for focusing on the genes with the most pronounced changes in expression, which are often the most biologically relevant.

Determining which functions and biological pathways the top DEGs are mostly associated with

Finally, to determine potential key biological functions and biological pathways enriched in the top DEGs, ShinyGo bioinformatics tool was used for functional annotation and pathway analysis of the DEGs. ShinyGo provides an interactive platform for understanding the biological significance of gene sets by linking them to known biological pathways and functions (ShinyGo, 2024). This tool helped elucidate the potential roles of the DEGs in the pathogenesis of type 2 diabetes, thereby offering potential insights into the underlying molecular mechanisms of the disease.

RESULTS

Identification of Genes Expressed Differently between Diabetic and Non-diabetic sample Using GEO2R Analysis

GEO2R bioinformatics tool, an interphase between NCBI and GEO was used to identify differentially expressed genes (DEGs.). GEO2R uses R programming language to generate DEG results (R-Script of this research). The GEO2R analysis revealed several genes that are expressed differently between the healthy (non-diabetic) and non-healthy (diabetic) samples as

shown in Figure 2. The volcano plot (Figure 2) reveals significant differences in gene expression, with the x-axis representing log2 or fold change and the y-axis displaying -log10 or P Value. Some genes were significantly upregulated (red dots) and others downregulated (blue dots) with Padj < 0.05. These results show differences in gene expression between the healthy and unhealthy samples.

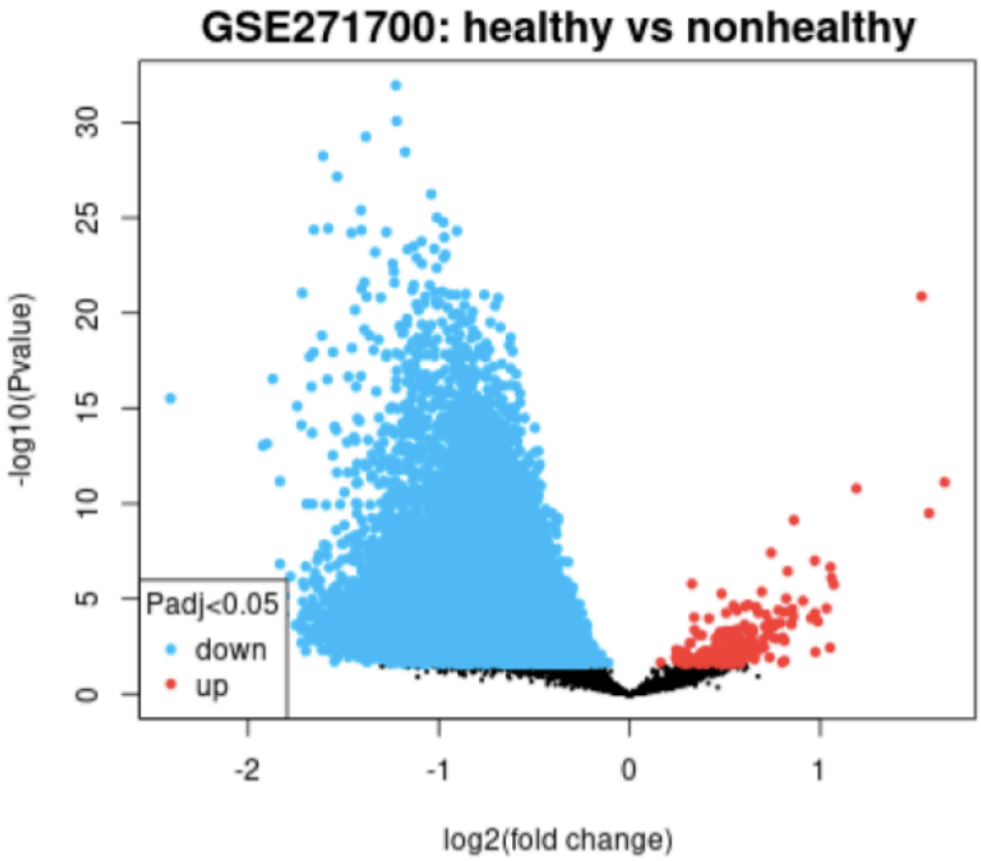


Figure 2. Genes that are downregulated (Padj < 0.05) are represented by blue dots, those that are significantly upregulated (Padj < 0.05) are shown as red dots, and genes that are not significantly differentially expressed are indicated by black dots.



Figure 3. This Venn diagram shows 11,221 differentially expressed genes between healthy and unhealthy samples out of a total of 8,937 genes tested.

Next, the Venn diagram result **(Figure 3)** reveals that out of the total number of genes analyzed, 11,221 genes were significantly differentially expressed between healthy and unhealthy **(Figure 3).** Further, these genes were narrowed down to top 30 DEGs **(Supplementary Results: Top 30 DEGs)**

Functional and Enrichment Analysis Using ShinyGo Bioinformatics Tool

In this study the gene pathways associated with diabetes were found using differential expression gene (DEG) analysis. The Pathway of Melanogenesis **(Figure 4)** illustrates the intricate signaling mechanisms involved, highlighting key components such as α-MSH, MC1R, and MITF. This pathway is crucial not only for understanding melanogenesis but also for identifying potential overlaps with diabetic gene expression patterns.

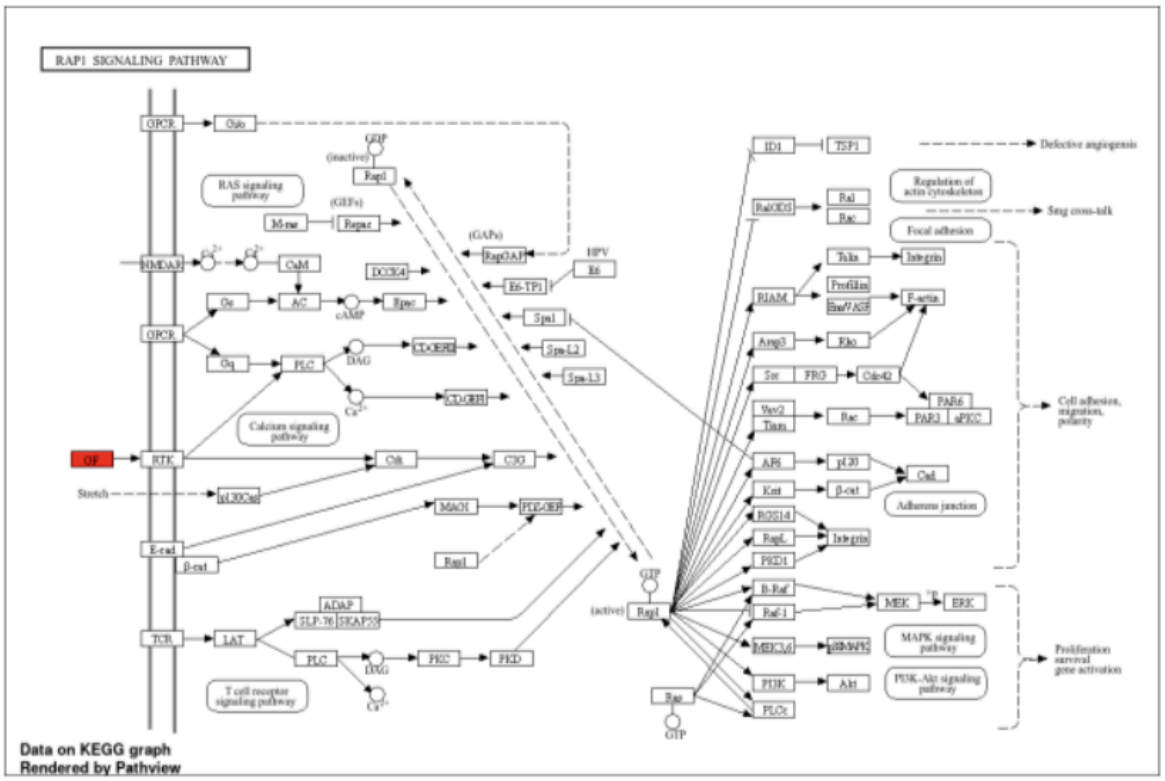


Figure 4. DEGs enriched in the RAP Signaling pathway with GF gene significantly upregulated (red).

Additionally, a network diagram of Gene Pathway Interactions (Figure 5) showcases the complex interconnections between various pathways, including the MAPK signaling pathway, Ras signaling pathway, and Phospholipase D signaling pathway. These interactions underscore the multifaceted nature of gene regulation in diabetes.

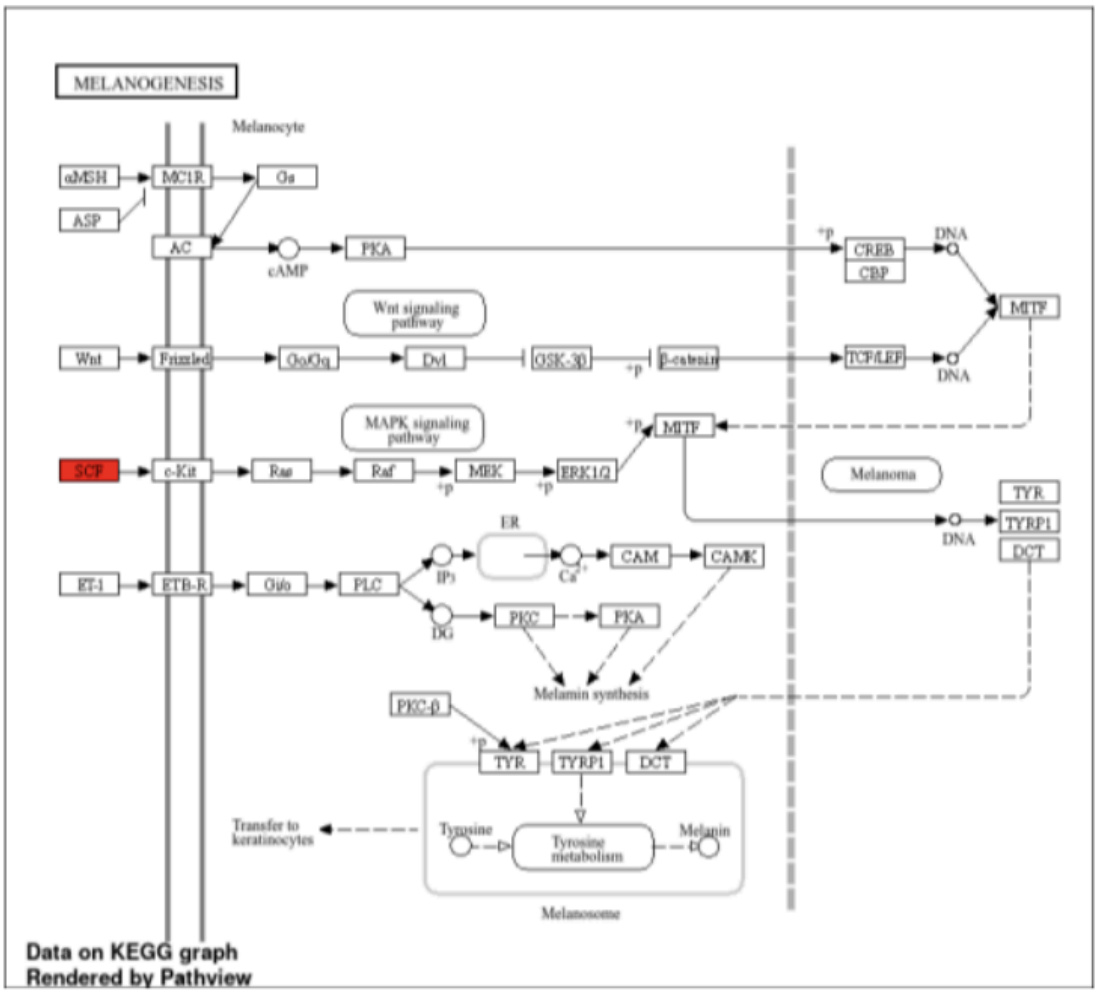


Figure 5. DEGs enriched in the Melanogenesis pathway with SCF gene significantly upregulated (red).

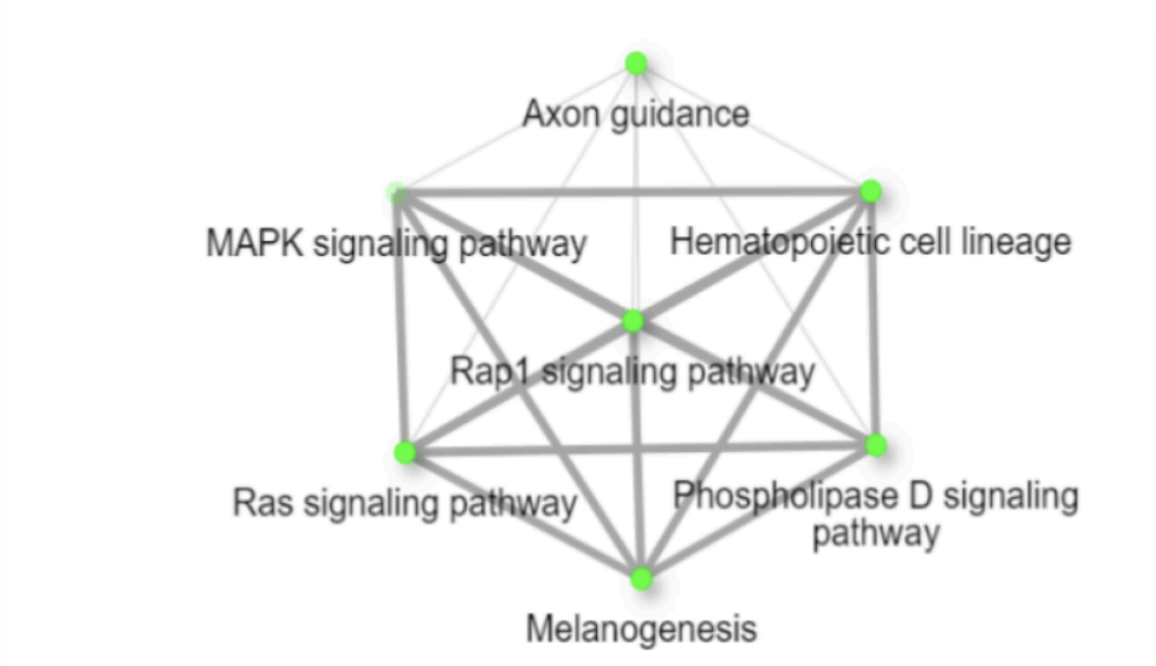


Figure 6: This KEGG Pathway Network Diagram shows the intricate relationships and cross-talk between these pathways.

The Pathway Enrichment Analysis **(Figure 6)** provides assessment of different pathways in our dataset. Hematopoietic cell lineage, Melanogenesis, and Phospholipase D signaling pathway exhibit high fold enrichment, indicating their potential roles in the disease mechanism. The MAPK signaling pathway, though showing a lower fold enrichment, is still significant,

suggesting its involvement in diabetes **(Figure 7).** These representations collectively show the complexity of gene interactions. These genes identified could be potentially studied in a more in depth study.

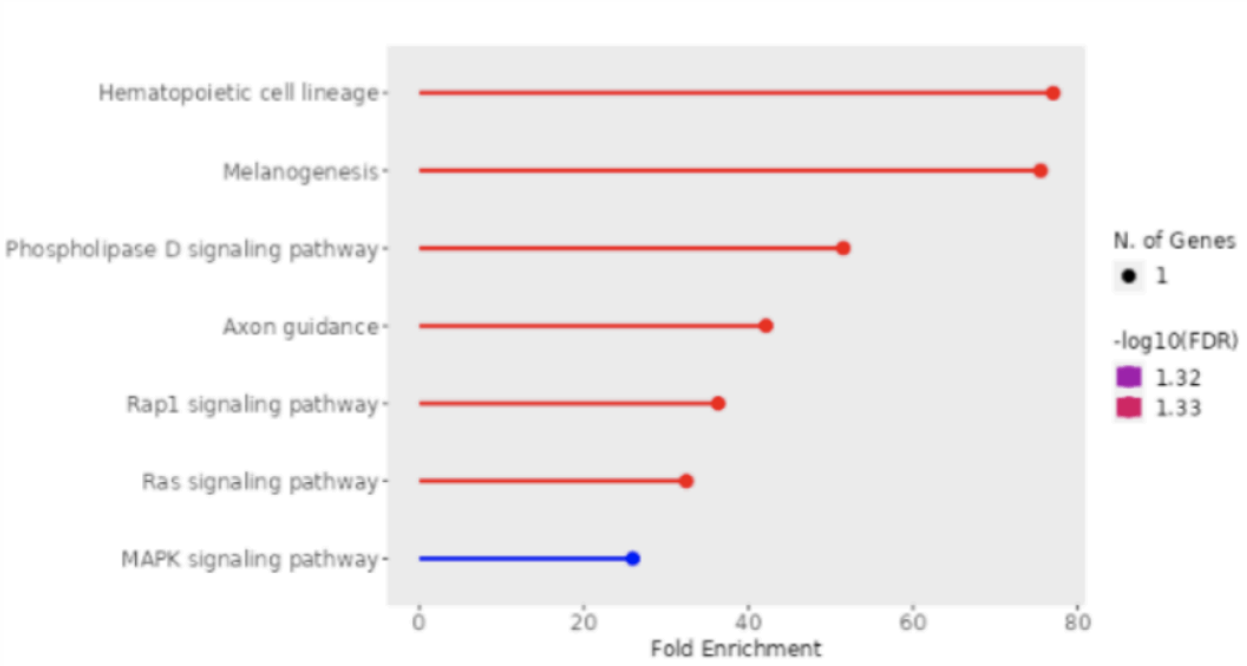


Figure 7. The chart shows the fold enrichment of various gene pathways. The color-coded bars represent the significance levels based on the -log10(FDR) values, with red indicating higher significance.

Supplementary Materials

R Script: R-Script of this research

Top DEGs: Top 30 DEGs

DISCUSSION

This study's main goal was to use bioinformatics to first identify differentially expressed genes (DEG) in diabetic samples and to investigate the potential gene functions and pathways linked to diabetes. Our research revealed that indeed some genes are expressed differently between diabetic and nondiabetic samples **(Figure 2)**.

Further, functional and enrichment analysis using Shiny Go showed enrichment of pathways namely the Insulin Secretion Pathway, Glucagon Signaling Pathway, and AGE-RAGE Signaling Pathway in Diabetic Complications as significantly upregulated in diabetic samples. The SFC gene

is part of the sarcoglycan complex, crucial for muscle integrity, and its dysfunction can affect muscle metabolism, influencing glucose metabolism and insulin sensitivity (Reference needed). Growth factors like insulin-like GF play essential roles in cell growth, differentiation, and survival, with dysregulation contributing to insulin resistance and diabetes (Kahn, 1994; McDonald & Yateman, 1999). RTKs, which are cell surface receptors for various growth factors, cytokines, and hormones, are vital for maintaining normal metabolic processes, and their dysregulation, especially in the insulin receptor, is a key mechanism in the development of type 2 diabetes (Schlessinger, 2000; White, 1998). The c-Kit gene encodes a receptor tyrosine kinase essential for cell survival and proliferation, and mutations or dysregulation in c-Kit signaling pathways can lead to metabolic imbalances associated with diabetes (Lennartsson & Rönnstrand, 2012). This indicates possible roles in the cause and development of diabetes. Significant pathways like hematopoietic cell lineage, melanogenesis, and phospholipase D are involved with signaling in diabetes **(Figure 6).**

The study's findings shed important light on the molecular processes for Type 2 diabetes. The variations in the expression of specific genes show their critical role in the pathogenesis of the illness. For example, the MAPK signaling pathway's contribution to a number of cellular functions, including growth, differentiation, and apoptosis, highlights its significance in diabetes. The discovery of these pathways and the genes linked to them may open the door to potential cures for this disease.

These results found from the study could be used to develop a potential cure for diabetes by isolating which genes to target for a cure. With the knowledge of what genes correlate to diabetes any potential cure could target the same genes to eliminate the disease. However, a limitation in the study was the small sample size, as well as the lack of diversity in sample sources, sample types. This leads to an overall limitation of bioinformatics, not directly collecting samples from human subjects or animals. This requires the experiment conducted in this research to be replicated in the lab before used for development for a cure. The identified genes were tested in the laboratory by scientists in the laboratory or clinical trials to determine if there are any potential cell inhibitors that could limit the upregulation of the gene, as well as suppress it, taking away its correlation to diabetes.

Conclusion

This study successfully found differentially expressed genes (DEGs) in diabetic samples, highlighting key genes and pathways involved in Type 2 diabetes. The top DEGs, such as SFC, RKTs, GF, and c-Kit, were linked to significant biological processes relating to diabetes. Pathway analyses revealed critical pathways like the hematopoietic cell lineage, phospholipase D signaling, and insulin signaling, which are essential in understanding the complex mechanisms of the disease.

The results show that these DEGs and associated pathways play a crucial role in the development of Type 2 diabetes. These findings provide a great foundation for further research, particularly in exploring therapeutic targets and developing new treatments for diabetes.

Future studies should include larger and more diverse sample sizes to validate these findings and identify additional genes and pathways involved in Type 2 diabetes. Laboratory experiments and clinical trials are crucial to further explore the identified genes and pathways, aiming to develop effective strategies for prevention and treatment. All in all, this study was successful in finding the upregulated genes responsible for diabetes.

References

1. Barrett, T. (2004). NCBI GEO: Mining Millions of Expression Profiles–Database and Tools. Nucleic Acids Research, 33(Database issue), D562–D566. https://doi.org/10.1093/nar/gki022
2. Dayeh, T., Volkov, P., Salo, S., Hall, E., Nilsson, E., Olsson, A. H., Kirkpatrick, C. L., Wollheim, C. B., Eliasson, L., Rönn, T., Bacos, K., Ling, C. (2014). Genome-wide DNA methylation profiling of human pancreatic islets from type 2 diabetic and non-diabetic donors. Gene Expression Omnibus, GSE53166. Available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE21232
3. Dye, C., & Maunakea, A. (2024). A type 2 diabetes DNA methylation signature in monocytes exhibits ethnic specificity preferentially for Native Hawaiians. Gene Expression Omnibus, GSE270223. Available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE270223
4. European Journal of Medical Research. (2023). Identification and validation of core genes for type 2 diabetes mellitus by integrated analysis of single-cell and bulk RNA-sequencing. Retrieved from https://eurjmedres.biomedcentral.com/articles/10.1186/s40001-023-00617-5

5. Gene Ontology Consortium. (2015). Gene Ontology Consortium: going forward. Nucleic Acids Research, 43(Database issue), D1049-56. https://doi.org/10.1093/nar/gku1179
6. Gene Ontology (GO) Tools. Retrieved from http://bioinformatics.sdstate.edu/go/
7. Grandl, G., Müller, T., Lickert, H., et al. (2024). Global, neuronal or β cell-specific deletion of inceptor improves glucose homeostasis in male mice with diet-induced obesity. Nature Metabolism. https://doi.org/10.1038/s42255-024-00991-3
8. Gottlieb, M., & Dyer, S. (2020). Information and disinformation: Social media in the COVID-19 crisis. Academic Emergency Medicine, 27(9), 844-846. https://doi.org/10.1111/acem.14076
9. Helmholtz Association of German Research Centres. (2024, February 28). Insulin-inhibitory receptor research offers hope for type 2 diabetes therapy. Medical Xpress. Retrieved July 22, 2024, from https://medicalxpress.com/news/2024-02-insulin-inhibitory-receptor-diabetes-therapy.html
10. Human Genomics. (2022). Gene expression analysis reveals diabetes-related gene signatures. Retrieved from https://humgenomics.biomedcentral.com/articles/10.1186/s40246-022-00387-4
11. IBAB Gene Identifier. Retrieved from http://resource.ibab.ac.in/GIDCON/geneid/home.html
12. Kahn, C. R. (1994). Insulin action, diabetes genes, and the cause of type II diabetes. Diabetes, 43(8), 1066-1084. https://doi.org/10.2337/diab.43.8.1066
13. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research, 45(D1), D353-D361. https://doi.org/10.1093/nar/gkw1092
14. Kang, J., Lee, H., Joo, J. Y., Song, J. M., et al. (2024). Comparison of genetic and epigenetic profiles of periodontitis according to the presence of type 2 diabetes. MedCom, 5(7), e620. https://doi.org/10.1002/mco2.620
15. Lloyd, M. W., Srivastava, A., Baker, C. N., Lek, S. H., Gerdes Gyuricza, I., Emerson, J., Barter, M. E., Tjong, H., Munger, H., Maurya, R., Schott, W., Adams, A., Gaca, M., Ngan, C. Y., Wei, C., Braun, M., Attie, A. D., Keller, M. P., Stitzel, M. L., Churchill, G. A. (2023). Identification of cross-species preserved cis-regulatory elements containing type 2 diabetes GWAS variants (pcHiC). Gene Expression Omnibus, GSE214107. Available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE214107
16. Maayan Lab. (2023). Enrichment analysis of genes upregulated in type 2 diabetes mellitus. Datasets Tools. Retrieved from https://maayanlab.cloud

17. McDonald, A., & Yateman, M. E. (1999). Growth factors in diabetes. In Joslin’s Diabetes Mellitus (pp. 373-384).
18. Rajalahti, T., Arneberg, R., Berven, F. S., Myhr, K. M., Ulvik, R. J., & Kvalheim, O. M. (2009). Discriminating Variable Test and Selectivity Ratio Plot: Quantitative Tools for Interpretation and Variable (Biomarker) Selection in Complex Spectral or Chromatographic Profiles. Analytical Chemistry, 81(7), 2581–2590. https://doi.org/10.1021/ac802514y
19. Schlessinger, J. (2000). Cell signaling by receptor tyrosine kinases. Cell, 103(2), 211-225. https://doi.org/10.1016/S0092-8674(00)00114-8
20. White, M. F. (1998). The IRS-signaling system: A network of docking proteins that mediate insulin action. Molecular and Cellular Biochemistry, 182(1-2), 3-11. https://doi.org/10.1023/A:1006828325900
21. Xue, D., Narisu, N., Taylor, D. L., Zhang, M., Grenko, C., Yan, T., Sinha, N., Erdos, M. R., Swift, A. J., Bonnycastle, L. L., Chen, S., Collins, F. S. (2023). Functional interrogation of twenty Type 2 Diabetes-associated genes using isogenic hESC-derived β-like cells [WGS]. Gene Expression Omnibus, GSE239329. Available at: https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE239329

Ibrahim Syed

Leveraging Bioinformatics to Uncover Gene Expression Differences in Amyotrophic Lateral Sclerosis (ALS): An Analysis of Key Molecular Pathways and Genes

ABSTRACT

Background:Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disease characterized by the degeneration of motor neurons, leading to muscle weakness and paralysis. Despite advances in research, the exact genetic mechanisms underlying ALS remain unclear, posing challenges for developing effective treatments.

Methods: This study aims to identify key genes and molecular pathways associated with ALS using bioinformatics tools. By analyzing gene expression data, we seek to uncover potential biomarkers and therapeutic targets that could aid in understanding and treating the disease. We utilized various bioinformatics databases and tools, including NCBI, GEO2R, GO, and KEGG, to analyze differentially expressed genes (DEGs) in ALS patients compared to healthy controls.

Results: Our analysis focused on identifying significant genes and pathways involved in the disease. The analysis revealed several genes, including FURIN, LGMN, DRD4, CD14, PRKCD, ITGB2, CYBA, and ACTN1, as significantly altered in ALS.

Discussion: These genes are involved in critical processes such as oxidative stress, immune response, and cellular structure maintenance. Pathways related to superoxide anion generation, pattern recognition receptor signaling, and actin filament organization were highlighted.

KEYWORDS: Amyotrophic lateral sclerosis (ALS), Bioinformatics, NCBI, GEO2R, SR Plot, Gene Ontology, KEGG

INTRODUCTION

Amyotrophic Lateral Sclerosis (ALS) is a neurodegenerative disease that affects nerve cells in the brain and spinal cord. ALS leads to loss of muscle control which can lead to many issues in one’s lifetime. ALS is often referred to as Lou Gehrig’s disease after a baseball player was diagnosed with it (A;,

B. D. (n.d.). However, the exact cause of the disease is still unknown. Some cases of this disease are inherited. ALS begins with common symptoms of muscle twitching and weakness in an arm or leg and also trouble swallowing or slurred speech. ALS affects control of muscle needed to move, speak, eat and breathe which leads to minimal body function (A;, B. D. (n.d.). ALS is a relatively rare disease with about 1/100,0000 cases. Mutation of superoxide dismutase 1 has been recognized as a potential cause for ALS. ALS usually starts in one specific area but eventually spreads to other parts of the body. The failure of respiratory muscles generally limits survival to 2-5 years after the disease begins. In about 50% of cases, people experience additional issues like changes in behavior, problems with executive functions, and language difficulties. For 10%-15% of patients, these issues are severe enough to be diagnosed as frontotemporal dementia (FTD) (Masrori and Van Damme). About 10% of ALS cases appear to run in families, suggesting an autosomal dominant inheritance pattern. The other 90% of cases don't have any family history of the disease and are classified as sporadic ALS (Masrori and Van Damme).

Recently, two medications, riluzole and edaravone, have been approved for treatment, offering a slight increase in survival time (Rokade et al.). However, many innovative experimental drugs are being developed. The use of bioinformatics will allow to find genes that are upregulated and downregulated in genetic pathways allowing to find DEG's to predict the disease.

Background and Significance

Amyotrophic Lateral Sclerosis (ALS) is a progressive neurodegenerative disease that affects nerve cells in the brain and the spinal cord. This disease is sometimes referred to as Lou Gehrig's disease, after the famous baseball player who was diagnosed with it. ALS leads to the loss of motor neurons, which are responsible for controlling voluntary muscle movements (Brotman et al.). The exact cause of ALS is unknown, but it is believed to be a combination of genetic and environmental factors. About 10% of cases are familial, meaning they are inherited through mutations in specific genes such as SOD1, C9orf72, TARDBP, and FUS (Masrori and Van Damme). The remaining 90% of cases are sporadic, meaning that there is no familial history linked to the diagnoses.

Despite significant research, the cause of ALS remains unknown. Genetic mutations account for about 10% of cases (familial ALS), the majority are sporadic, with no clear genetic link. Due to this research is needed to identify a way to predict ALS through DEGs and potential biomarkers.

Scientific Question

In this research study, the gene expression profiles of spinal and oculomotor tissue samples from control and individuals with sporadic ALS are compared. The primary question presented by this research is "Which specific genes and genetic pathways show upregulation or downregulation in ALS, and how can these discoveries enhance genetic testing and preventative strategies for those at risk of developing ALS?" Identifying these differences in genes will allow us to pinpoint certain areas and contribute to finding a cure for the disease. This question is crucial to pave the way for understanding the mechanisms that take place before ALS is diagnosed

Research Goals and Hypothesis

The primary goal of this research is to identify the genes and genetic pathways that are differently expressed in ALS patients compared to healthy individuals. This will help future research in developing genetic tests and preventative measures, potentially leading to earlier interventions and better management of ALS. To achieve this, various bioinformatics tools and databases, such as the National Center for Biotechnology Information (NCBI), GEO2R, and SGplot will be used.

Hypothesis: There will be differences in gene expression between oculomotor tissue sporadic and control samples in ALS and the diseased samples will have higher levels of gene expression.

METHODS

Preparation of GEO Dataset:

The dataset, GSE833, was generated from a study that compared post-mortem samples of control, healthy spinal tissue with samples from both familial and sporadic ALS-affected individuals. Gene expression was measured through microarray analysis. The GSE833 dataset analyzed the total RNA extracted from postmortem gray matter of lumbar spinal cord from 11 individuals, including five with sporadic-ALS (two samples were

confirmed to have mutations associated with familial ALS and were excluded from this study) and four normal controls. Microarray analysis, including quantification and normalization of gene expression levels, was conducted using Affymetrix GeneChip 3.1 software (Dangond et al.).

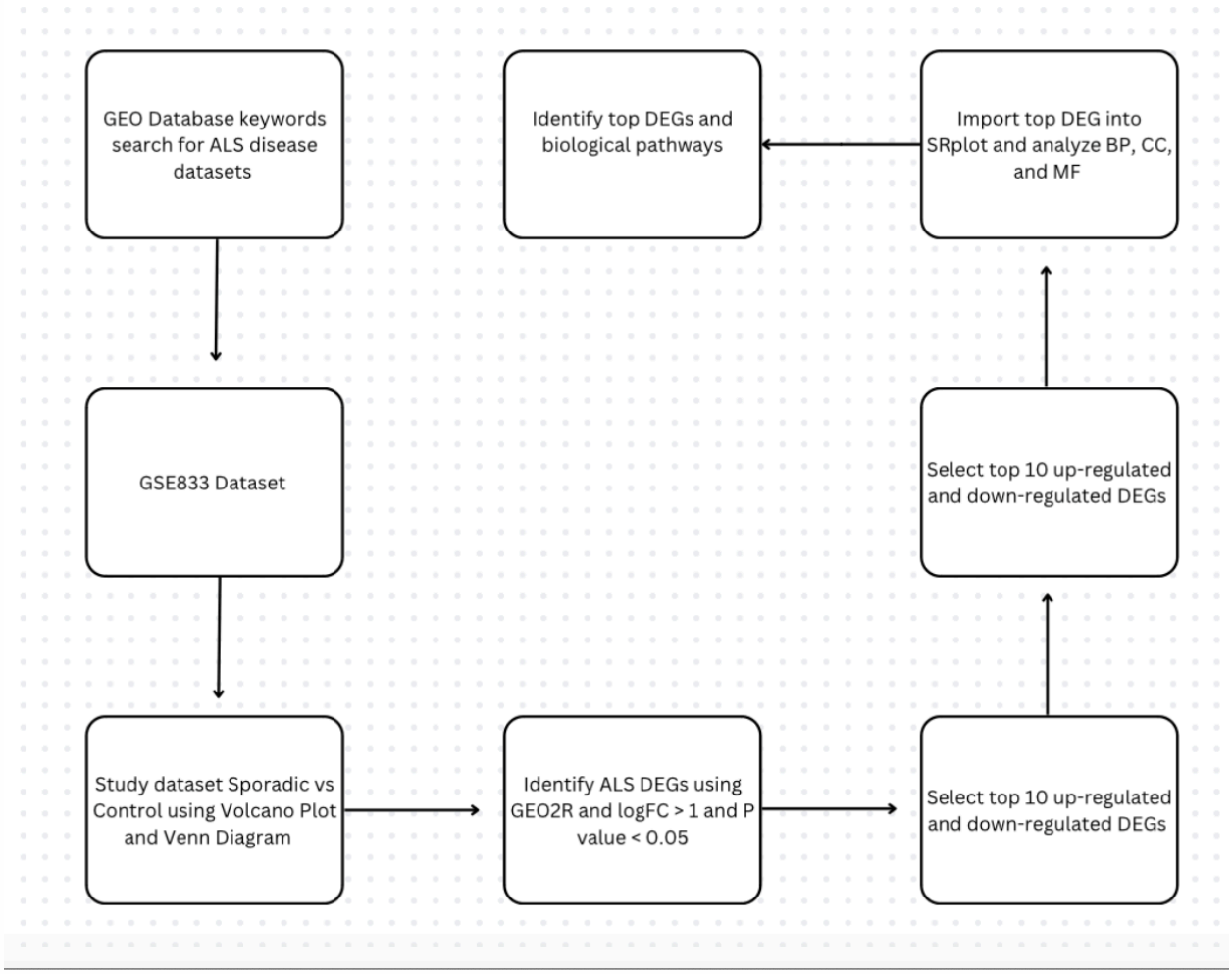


Figure 1. This flowchart shows the methods used in this study to investigate gene expression differences in Amyotrophic lateral sclerosis (ALS), and datasets GSE833 was used. First, the datasets were found using key search words in the GEOdatabase. Then DEGs were identified using GEO2R, and statistics of P-value < 0.05 and |log2FC| > 1.0. Then, the top 25 up and down regulated genres were identified and imported into SR Plot to study BP, CC, AND MF.

Collection of Data

Firstly, data was collected by researching articles on ALS and analyzing it with GEO2R. The study chosen was a data set GSE833 which was the comparison of sporadic ALS and healthy control samples. Samples were

collected from post-mortem spinal cord gray matter. The data was divided into two groups control and sporadic meaning ALS was diagnosed with no familial history. The two groups were divided and the graphs were made.

RESULTS

Identifying differentially expressed genes in ALS control vs sporadic samples
Using dataset GSE833 the top 25 upregulated and downregulated genes were identified using P-value < 0.05 and |log2FC| > 1.0. The genes were imported into SR Plot for further analysis. The R Script is also linked.

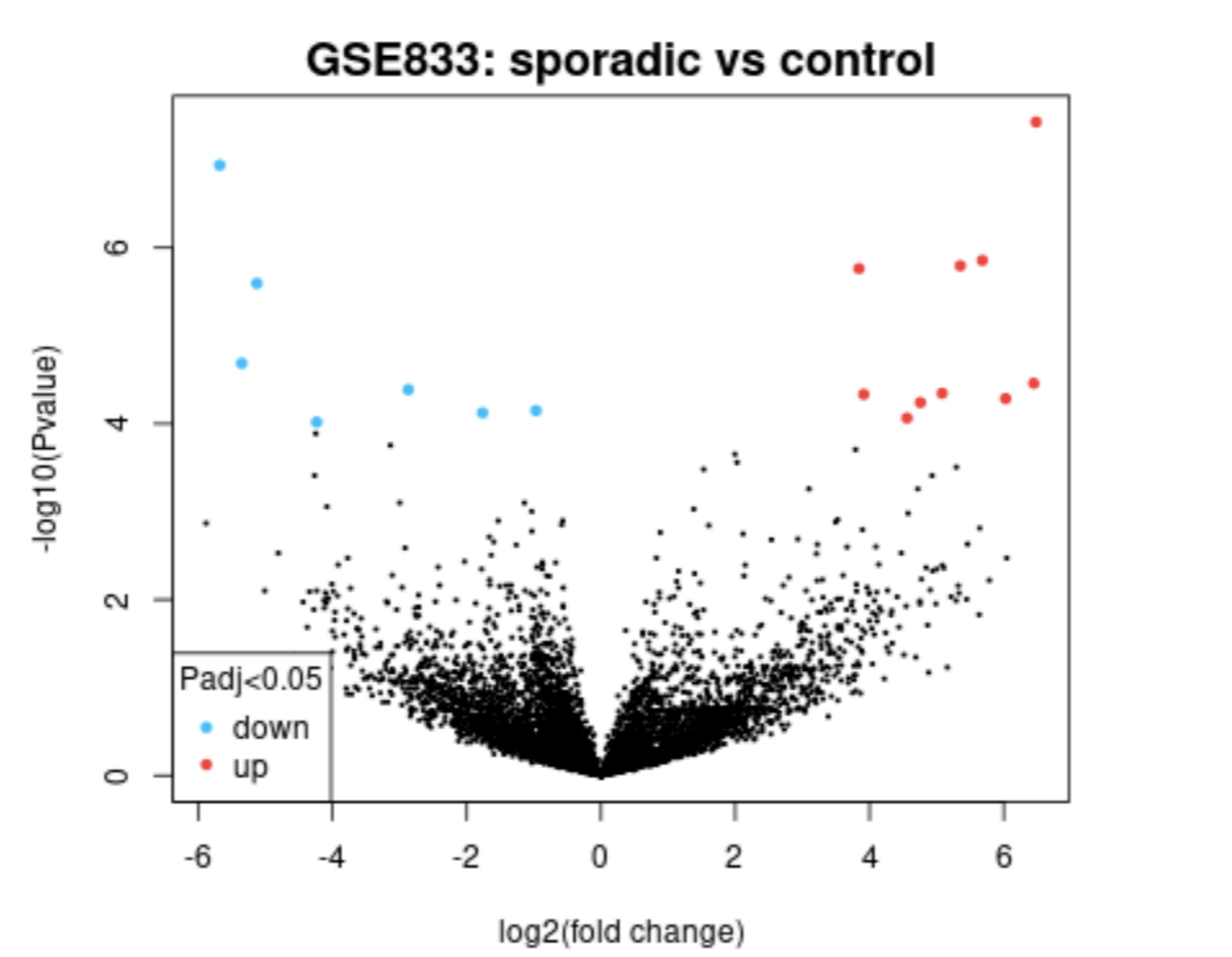


Figure 2. Volcano Plot The image is a volcano plot from a study comparing sporadic cases to a control group, labeled "GSE833: sporadic vs control." The x-axis represents the log2(fold change), while the y-axis represents the -log10(P value). Significant data points are highlighted in blue (downregulated) and red (upregulated) with an adjusted P value less than

0.05. The central black cluster represents non-significant data points. The plot visually distinguishes genes with significant expression changes between the two groups.

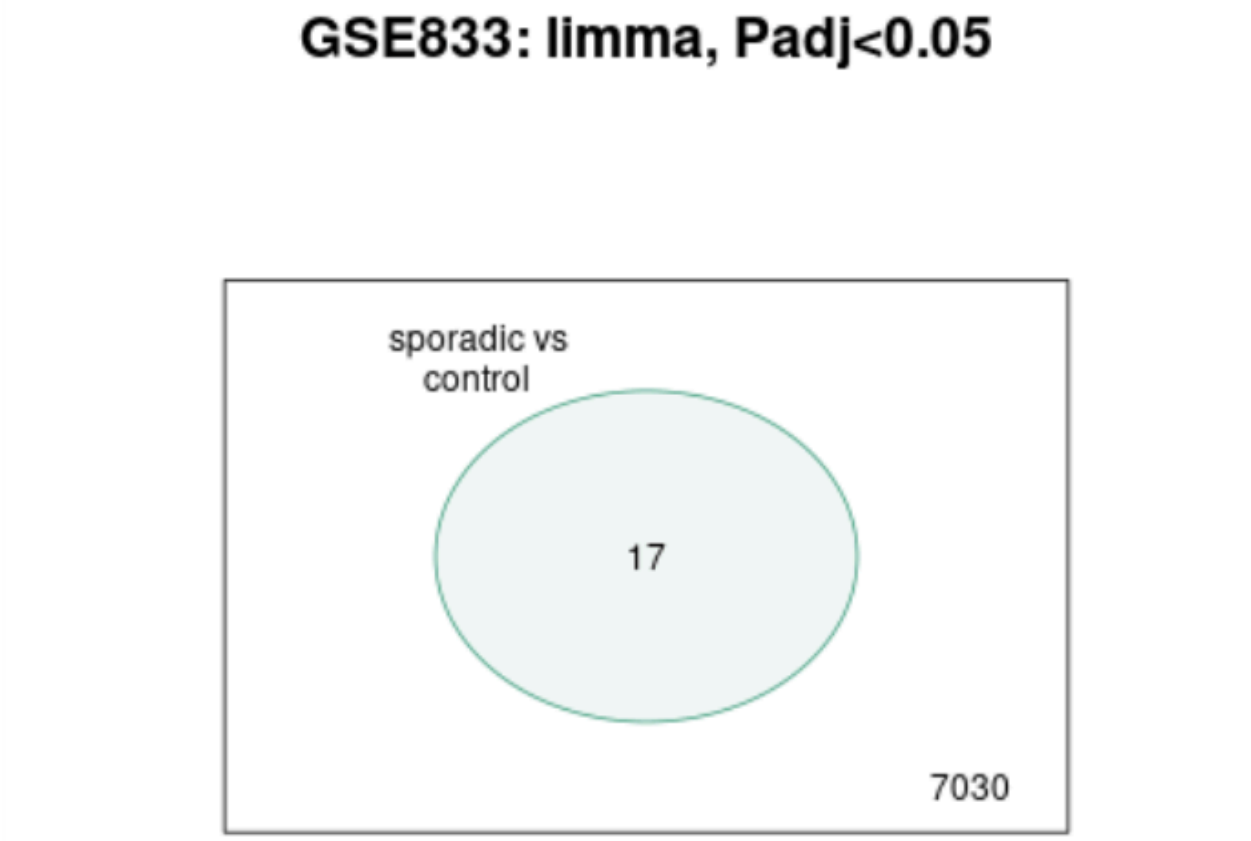


Figure 3. Venn Diagram image is a Venn diagram from a study comparing gene expression in sporadic ALS cases to a control group, labeled "GSE833: limma, Padj<0.05." The diagram shows two sets of genes: those in sporadic ALS cases and those in controls. The overlap in the center, containing the number 17, represents the genes that are significantly differentially expressed between the two groups. The number 7030 outside the overlap indicates the total number of genes analyzed in the study. This diagram visually highlights the subset of genes that exhibit significant expression changes in sporadic ALS compared to controls, with an adjusted P value of less than 0.05.

SR Plot Analysis

After identifying top 30 DEG, they were imported into SR plot to create graphs to identify biological pathways and cellular components.

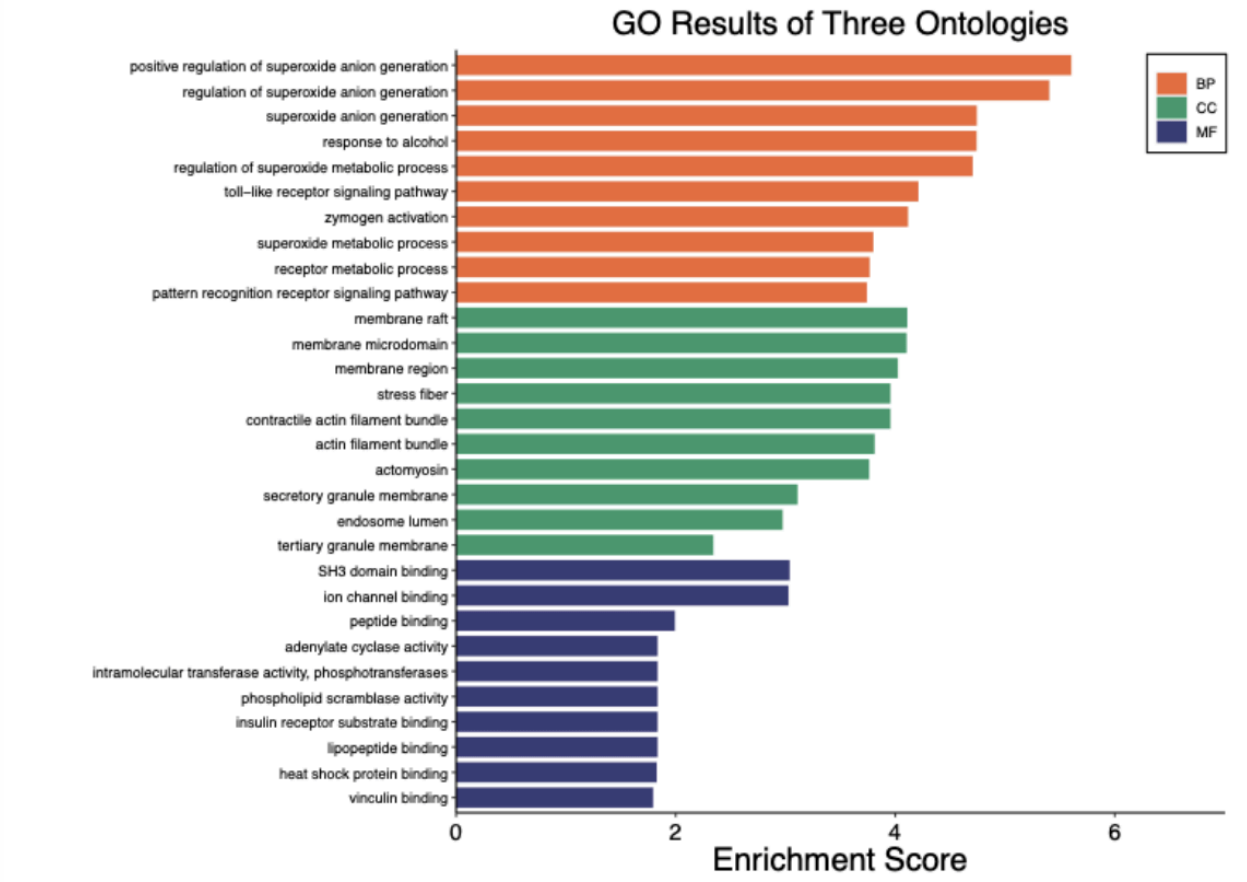


Figure 4. GO results of three ontologies. This bar plot shows the enrichment scores for Gene Ontology (GO) terms across three categories: Biological Process (BP), Cellular Component (CC), and Molecular Function (MF). The most enriched terms include positive regulation of superoxide anion generation, regulation of superoxide metabolic process, and superoxide anion generation in the Biological Process category. The Cellular Component category highlights terms like membrane raft, membrane microdomain, and membrane region. In the Molecular Function category, terms such as SH3 domain binding, ion channel binding, and peptide binding are significantly enriched. Due to these results only biological processes and cellular components will be studied as they are the most enriched.

After analyzing this data it can be seen that the biological pathway and cellular components are highly enriched. The GO analysis for these areas will be below.

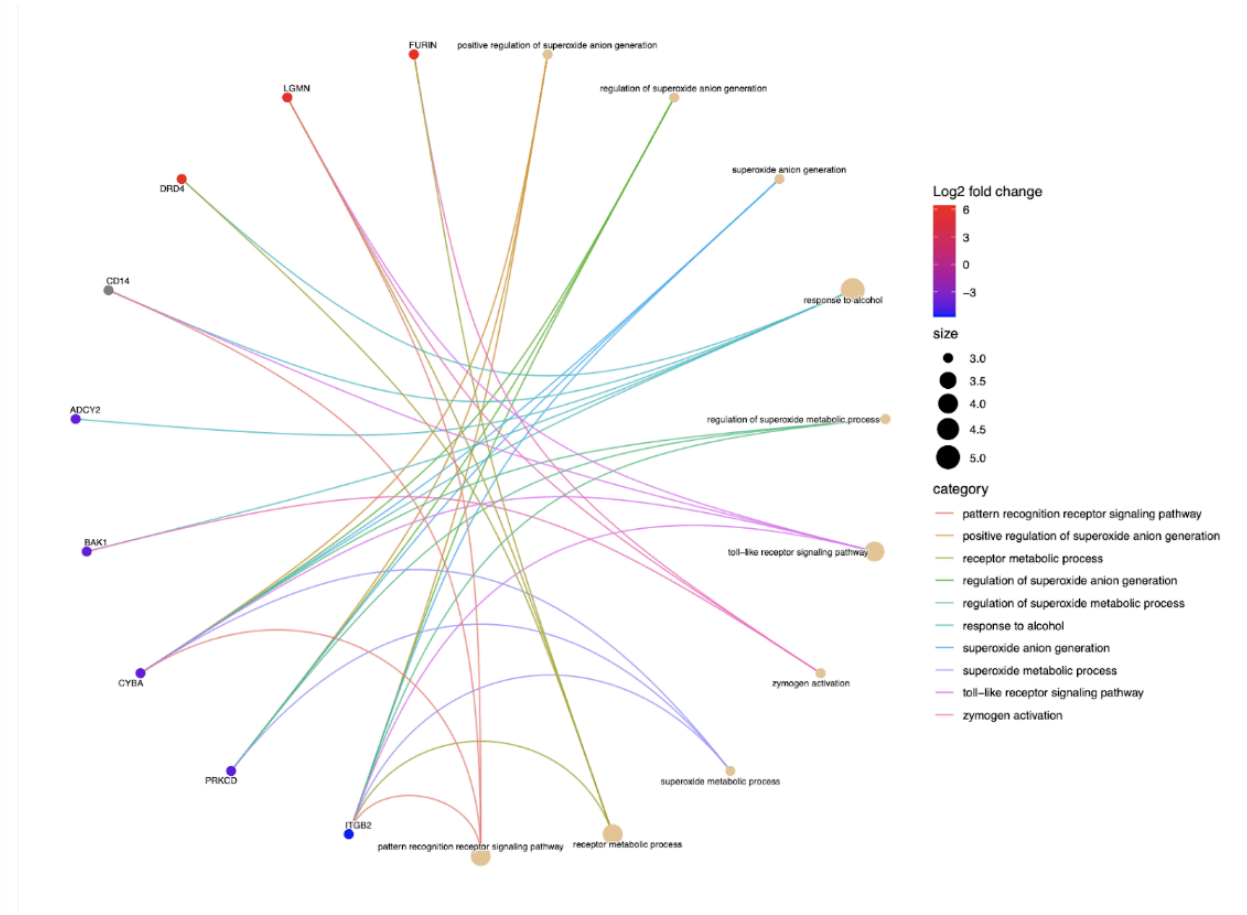


Figure 5. This cnet plot shows the connections between specific genes and their associated biological processes in sporadic ALS patients. The genes with higher log2 fold changes are highlighted in red, indicating their significant upregulation. Genes such as FURIN, LGMN and DRD4 are significantly enriched.

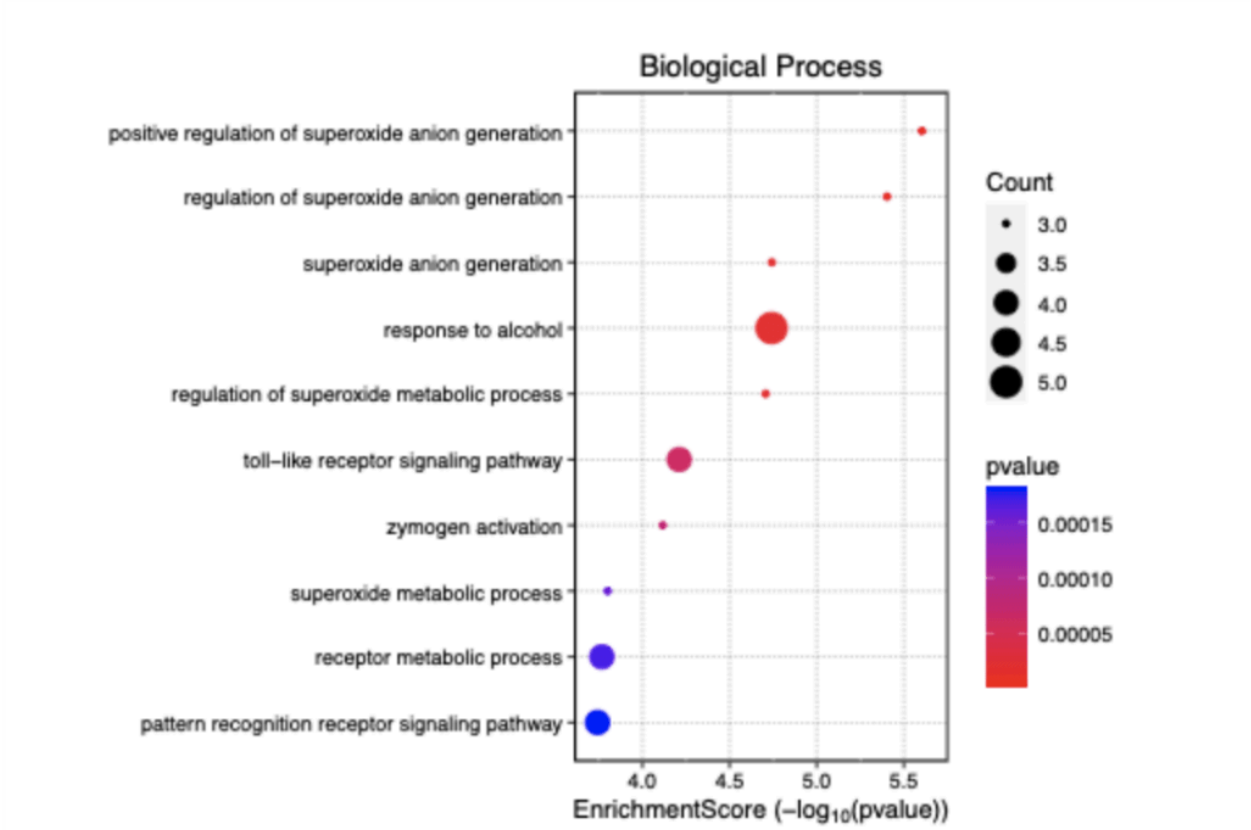


Figure 6: This dotplot visualizes the enrichment scores of various biological processes in sporadic ALS patients. The x-axis represents the enrichment score (-log10(p-value)), indicating the significance of each process. The size of the dots reflects the number of genes (count) involved in each process, while the color gradient represents the p-value, with red indicating lower p-values (higher significance).

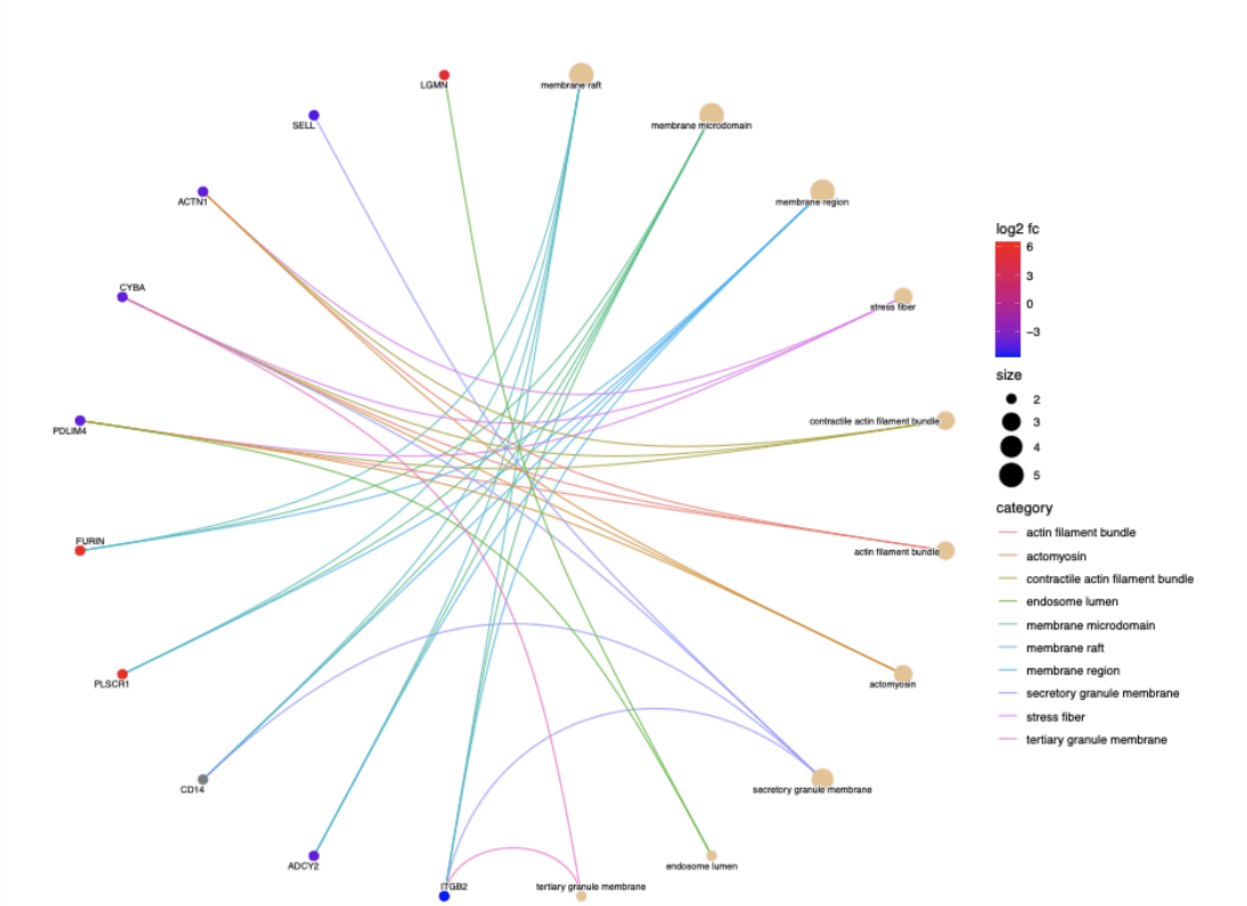


Figure 7: This cnet plot shows the relationships between specific genes and their associated cellular components in sporadic ALS patients. Genes with high log2 fold changes are highlighted in red, portraying significant upregulation. Key genes such as FURIN, LGMN, and CYBA are associated with components like membrane rafts, membrane microdomains, and stress fibers. The size of the nodes represents the significance of the enrichment scores, with larger nodes indicating higher significance. This plot highlights the involvement of various cellular structures in the pathogenesis of ALS.

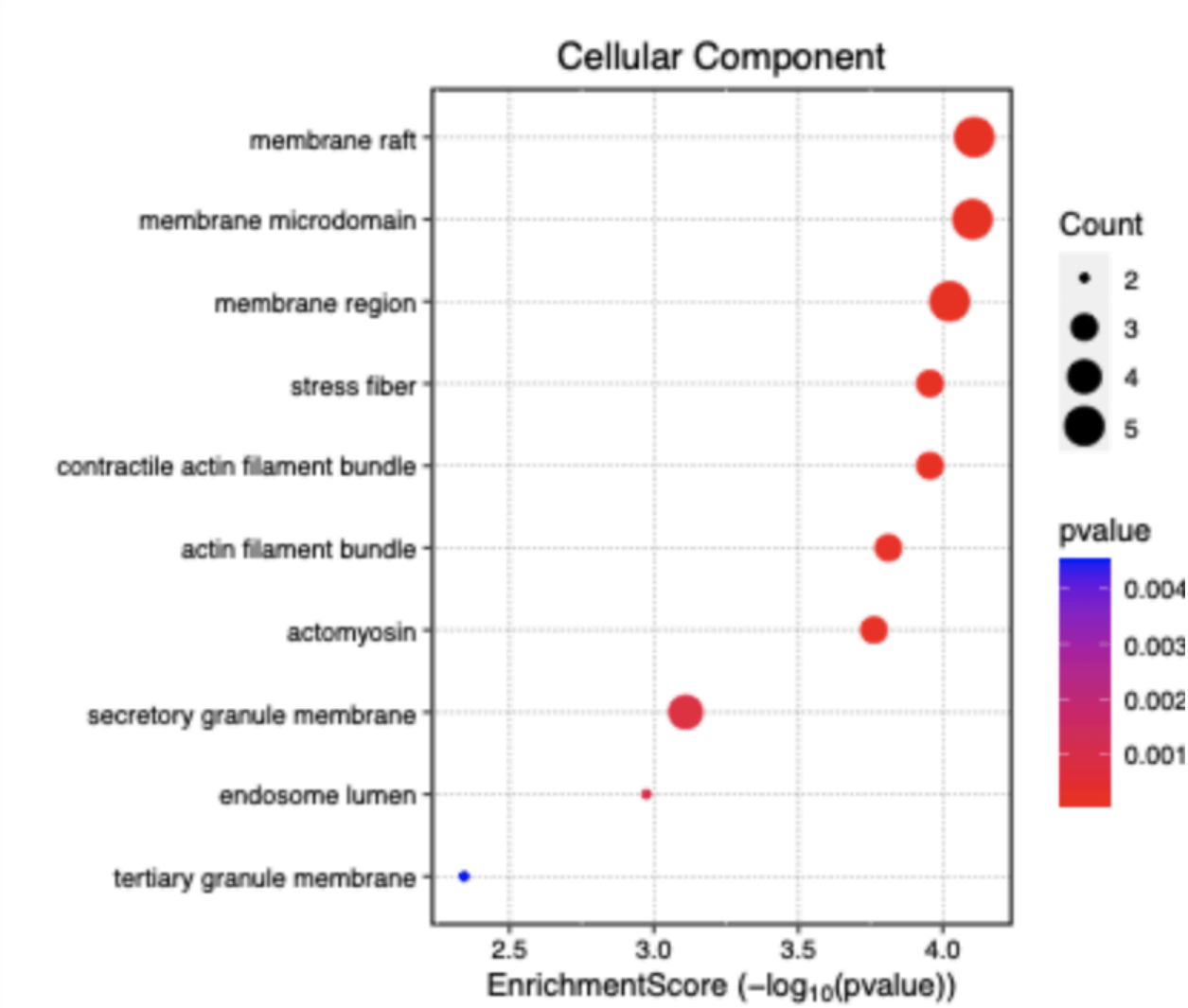


Figure 8: This dotplot shows the enrichment scores of different cellular components in sporadic ALS patients. The x-axis represents the enrichment score (-log10(p-value)), showing the significance of each component. The size of the dots reflects the number of genes (count) involved in each component, while the color gradient represents the p-value, with red indicating lower p-values (higher significance). The three most enriched cellular components are membrane raft, membrane microdomain and membrane region.

Table 1. This table shows the top DEGs from the BP and CC analysis. The name of the gene is shown on the left of the table. In the middle, the function/pathway of the gene is shown and the genes' connection to ALS is discussed.

Gene	Function/Pathway	Connection to ALS
FURIN	Positive Regulation of Superoxide Anion Generation	FURIN: Furin is a proprotein convertase involved in the processing of various precursor proteins. Its dysregulation has been implicated in several neurological conditions, including ALS. Furin influences neuroinflammation and neurodegeneration through its role in cytokine production and the activation of pro-inflammatory genes in macrophages(Zhang et al.).
LGMN	Superoxide Anion Generation	LGMN (Legumain): Legumain is an asparaginyl endopeptidase implicated in lysosomal processing and degradation. Recent studies suggest that its upregulation contributes to the pathological protein aggregation seen in ALS, potentially through mechanisms involving oxidative stress and inflammation (Thomas Sumner).
DRD4	Regulation of Superoxide Anion	DRD4 (Dopamine Receptor D4): Although

	Generation	primarily studied in the context of psychiatric disorders, alterations in dopamine signaling have been observed in ALS patients, linking DRD4 to the modulation of oxidative stress pathways that exacerbate neurodegeneration (Science Daily).
CD14	Pattern Recognition Receptor Signaling Pathway	CD14: CD14 is a pattern recognition receptor that plays a crucial role in the innate immune response. It has been linked to ALS through its involvement in microglial activation and the neuroinflammatory response, which are key components of ALS pathology (Cordova et al.).
PRKCD	Toll-like Receptor Signaling Pathway	PRKCD (Protein Kinase C Delta): This gene is part of the Toll-like receptor signaling pathway, which is crucial for the innate immune response. PRKCD is involved in neuroinflammation and has been shown to affect neuronal survival in ALS models (Zhang et al.).
ITGB2	Pattern Recognition	ITGB2 (Integrin

	Receptor Signaling Pathway	Subunit Beta 2): ITGB2 is involved in leukocyte adhesion and migration, processes that are dysregulated in ALS. The gene's role in modulating neuroinflammation through microglial activation highlights its importance in ALS pathology (Zhang et al.).
CYBA	Membrane Raft	CYBA (Cytochrome b-245 Alpha Chain): CYBA is a component of the NADPH oxidase complex, involved in reactive oxygen species (ROS) production. Increased ROS production is a hallmark of ALS, linking CYBA to oxidative stress and neuronal damage observed in the disease ("https://www.ncbi.nlm.nih.gov/gene?Db=gene&Cmd=DetailsSearch&Term=1536").
ACTN1	Actin Filament Bundle	ACTN1 (Actinin Alpha 1): ACTN1 plays a role in actin filament organization. Mutations and dysregulation in cytoskeletal proteins, including ACTN1, have been associated with ALS, affecting neuronal structure and function("ACTN1 Actinin

		Alpha 1 [Homo Sapiens (Human)] – Gene – NCBI”).

DISCUSSION

Amyotrophic lateral sclerosis (ALS) is increasing everyday and there is no clear method for recovery. Predicting the disease early on and developing a plan for prevention will halt the spread of the disease. This research study aimed to find highly enriched genes in ALS vs control patients to identify potential biomarkers. Utilizing GEO2R, NCBI and SRPlot several highly enriched genes were found as seen in Table 1 and a potential link to ALS based on previous research. Eight highly expressed genes were found in Figure 5 and 7. The analysis revealed key genes including FURIN, LGMN, DRD4, CD14, PRKCD, ITGB2, CYBA, and ACTN1 (Table 1) Specifically, CD14, another highly expressed gene in our study, expresses a surface antigen primarily found on monocytes and macrophages. It plays an important role in the innate immune response by interacting with bacterial lipopolysaccharides and viruses. A study by Zhang et al. (2020) indicated that CD14 is involved in the inflammatory response in ALS, contributing to the disease’s progression by promoting neuroinflammation (Lu et al, 2020). Also, CD14 has also been considered as a potential cure for reducing severe inflammation in SARS-CoV-2 infections (NIH, 2024).

These genes are involved in critical processes such as oxidative stress response, immune response, and cellular structure maintenance (Figure 6 & 8) . The results of this study show that these identified genes play crucial roles in ALS pathology. For example, FURIN is involved in oxidative stress responses, while CD14 and PRKDC are significant in immune response and inflammation (Table 1). These findings support the hypothesis that ALS is driven by multiple genetic interactions that impact multiple cellular processes.

Findings

The findings from this study align with previous research showcasing the role of oxidative stress and neuroinflammation in ALS progression. For example, FURIN has been linked to neurodegenerative diseases through involvement in processing neurotrophic factors and proteases, which are essential for neuronal survival (NIH, 2024). Similarly, CD14, another highly expressed

gene encodes a surface antigen primarily found on monocytes and macrophages. It works with other proteins to drive the innate immune response against bacterial lipopolysaccharides and viruses. It's also being considered as a target for treating SARS-CoV-2 infections to help reduce severe inflammation (NIH, 2024). The findings are similar to previous research that show the roles of oxidative stress and neuroinflammation in ALS disease. For example, FURIN has been connected to neurodegenerative diseases through involvement in processing neurotrophic factors and proteases essential for neuronal survival. A study shows that FURIN is crucial for the proteolytic activation of neurotrophic factors, and its dysregulation is associated with ALS and other neurodegenerative disorders (Anestopoulos et al, 2017).

A study published in Scientific Reports shows that FURIN is involved in ALS pathology, specifically in the endoplasmic reticulum stress response, which is a main feature of ALS (Yamada et al, 2018).

Implications

One implication of these genes is that it provides for clinical applications such as developing new therapy options for ALS. By understanding the molecular pathways behind ALS new treatments can be developed which target these genes and their associated pathways. This could involve gene-based therapies or small molecule inhibitors designed to combat oxidative stress or control immune responses in ALS patients.

Limitations

Some limitations of the study are that, first, the sample size is too small. Additionally, since the data from previous experiments was used, therefore, more validation from laboratory or clinical settings is required. Also, more datasets and different groups are required to validate the expression of these genes and their relevance to ALS.

Future directions

Further research should focus on validating these genes in laboratory experiments and clinical trials. Through this the functional role of these genes in ALS can be validated. This can lead to the development of gene therapies. The identified genes can be tested in the laboratory by scientists

in the laboratory or clinical trials to determine if a potential biomarker can be identified.

Conclusion

In conclusion, this study utilized bioinformatics tools and databases to explore the genetic links between specific genes and Amyotrophic Lateral Sclerosis (ALS). Our analysis focused on several key genes—FURIN, LGMN, DRD4, CD14, PRKCD, ITGB2, CYBA, and ACTN1—and their associated biological processes and pathways, such as neuroinflammation, oxidative stress, and cytoskeletal integrity.

The discussion and results show that these genes are important in the prognosis of ALS. Genes such as FURIN and LGMN are involved in oxidative stress regulation, while CD14 and PRKCD participate in neuroinflammatory pathways. These results as well as knowing what each gene means can help to understand ALS effectively and potentially develop treatments. The hypothesis of this experiment was that there will be differentially expressed genes when comparing the control and sporadic group. This was true as several DEGs were identified throughout the experiment.

Although these genes were identified. Future research should focus on experimental validation of these genetic associations and explore the molecular interactions in greater detail. Through future research more thoroughly developed molecular pathways can be formed ultimately improving patient outcomes and quality of life.

References

1. A, B. D. “Update on Genetics of Amyotrophic Lateral Sclerosis.” Current Opinion in Neurology, n.d., https://pubmed.ncbi.nlm.nih.gov/35942673/.
2. A, K. R. M. S. “Amyotrophic Lateral Sclerosis (ALS) Prediction Model Derived from Plasma and CSF Biomarkers.” PLOS ONE, n.d., https://pubmed.ncbi.nlm.nih.gov/33606761/.
3. BJ, C. R. A. “Novel Genes Associated with Amyotrophic Lateral Sclerosis: Diagnostic and Clinical Implications.” The Lancet Neurology, n.d., https://pubmed.ncbi.nlm.nih.gov/29154141/.

4. Brotman, Ryan G., et al. “Amyotrophic Lateral Sclerosis.” StatPearls Publishing, 22 Aug. 2022, www.ncbi.nlm.nih.gov/books/NBK556151/.
5. “CD14 CD14 Molecule [Homo Sapiens (Human)] – Gene – NCBI.” NCBI, www.ncbi.nlm.nih.gov/gene/929. Accessed 12 Apr. 2024.
6. Clough, Emily, and Tanya Barrett. “The Gene Expression Omnibus Database.” SpringerLink, Springer New York, 2016, link.springer.com/protocol/10.1007/978-1-4939-3578-9_5.
7. Cordova, Zuzet Martinez, et al. “Myeloid Cell Expressed Proprotein Convertase FURIN Attenuates Inflammation.” Oncotarget, vol. 7, no. 34, 5 Aug. 2016, pp. 54392–54404,www.ncbi.nlm.nih.gov/pmc/articles/PMC5342350/, doi:10.18632/oncotarget.11106. Accessed 2 Dec. 2022.
8. “CYBA Cytochrome B-245 Alpha Chain [Homo Sapiens (Human)] – Gene – NCBI.” NIH.gov, 2024, www.ncbi.nlm.nih.gov/gene/1535. Accessed 6 Aug. 2024.
9. Dameron, O., Bettembourg, C., and N. L. Meur. “Measuring the Evolution of Ontology Complexity: The Gene Ontology Case Study.” PLOS ONE, n.d., https://journals.plos.org/plosone/article?id=10.1371%2Fjournal.pone.0075993.
10. “FURIN Furin, Paired Basic Amino Acid Cleaving Enzyme [Homo Sapiens (Human)] – Gene – NCBI.” NCBI, www.ncbi.nlm.nih.gov/gene/5045.
11. “Gene Ontology Overview.” Gene Ontology Resource, 11 July 2024
12. “Genetic Link between Both Types of ALS Discovered.” ScienceDaily, 2024, www.sciencedaily.com/releases/2010/05/100505173014.htm. Accessed 6 Aug. 2024.
13. “How ALS Progresses on Genetic and Cellular Level Revealed by High-Res Spinal Cord Study.” Simons Foundation, 4 Apr. 2019, www.simonsfoundation.org/2019/04/04/als-progression-genetic-cellular/.
14. Anestopoulos, Ioannis, et al. “A Novel Role of Silibinin as a Putative Epigenetic Modulator in Human Prostate Carcinoma.” Molecules, vol. 22, no. 1, 31 Dec. 2016, pp. 62–62, www.ncbi.nlm.nih.gov/pmc/articles/PMC6155798/, doi:10.3390/molecules22010062. Accessed 17 Aug. 2024.
15. “Kegg Pathway Map (HELP).” KEGG, n.d.,

16. Masrori, P., and P. Van Damme. “Amyotrophic Lateral Sclerosis: A Clinical Review.” European Journal of Neurology, vol. 27, no. 10, 7 July 2020, pp. 1918–1929, www.ncbi.nlm.nih.gov/pmc/articles/PMC7540334/, doi:10.1111/ene.14393.
17. Rokade, Aditi V., et al. “Riluzole and Edavarone: The Hope against Amyotrophic Lateral Sclerosis.” Cureus, 7 Oct. 2022, doi:10.7759/cureus.30035.
18. Sabatelli, M., et al. “New ALS-Related Genes Expand the Spectrum Paradigm of Amyotrophic Lateral Sclerosis.” Brain Pathology, n.d., https://pubmed.ncbi.nlm.nih.gov/26780671/.
19. Tang, Doudou, et al. “SRplot: A Free Online Platform for Data Visualization and Graphing.” PLOS ONE, U.S. National Library of Medicine, 9 Nov. 2023,
20. Vidovic, M., Müschen, L. H., Brakemeier, S., Machetanz, G., & Naumann, M. “Current State and Future Directions in the Diagnosis of Amyotrophic Lateral Sclerosis.” Cells, n.d., https://pubmed.ncbi.nlm.nih.gov/36899872/.
21. Yamada, Mariko, et al. “Furin Inhibitor Protects against Neuronal Cell Death Induced by Activated NMDA Receptors.” Scientific Reports, vol. 8, no. 1, 26 Mar. 2018, doi:10.1038/s41598-018-23567-0. Accessed 4 Apr. 2022.
22. Zhang, Yi, et al. “The Emerging Role of Furin in Neurodegenerative and Neuropsychiatric Diseases.” Translational Neurodegeneration, vol. 11, no. 1, 23 Aug. 2022, doi:10.1186/s40035-022-00313-1.

Katelyn Szeto

Using Bioinformatics in Order to Analyze the Top Differently Expressed Genomes with PKP2 Knockout from iPSC-CMs

ABSTRACT

Background:Arrhythmogenic right ventricular cardiomyopathy (ARVC) is a genetic heart disease that can lead to a higher risk of sudden death. Some symptoms are: irregular heartbeat, chest pain, and shortness of breath. The most common cause of ARVC is mutations, or changes in DNA and genes related to desmosomal proteins. These are structures in heart cells that help them stick together and function properly. Currently, there are no treatments available for this disease, and this research aims to help in the search for treatment. Many details about ARVC remain unclear which could possibly aid in the development of future treatments, such as new forms of gene therapy.

Methods: To conduct this research, five bioinformatics tools were used: NCBI, KEGG, GO (Gene Ontology), GEO2R, and DAVID (Database for Annotation, Visualization, and Integrated Discovery). DAVID was used to relate the genes in ARVC to other human diseases or potentially fatal causes of ARVC. Moreover, GEO2R was used to collect and analyze data from the data set GSE253223, which was a research study that tested two potential ARVC treatments with mouse models.The two treatments that categorized two groups referenced in the data set were siNEG and siPKP2. Finally, analysis by GEO2R helped identify differentially expressed genes (DEGs) in various samples sorted according to ARVC treatments, and DAVID was used to analyze GO and KEGG enrichment pathways to determine functions and biological pathways that the genes may be related to.

Results: Results from the GEO2R analysis identified over 18,000 genes, of which 252 were differentially expressed genes (DEGs) of the two treatment groups. These results were utilized by R programming, which was made by machine learning and artificial intelligence algorithms. It was unclear which genes were upregulated and downregulated, which suggests the need for further research. The DEGs were then narrowed down to the top 30 using the p-value in statistics. From the 252 DEGs, KEGG pathway analysis of the top 30 DEGs analyzed using DAVID showed that these DEGs were mostly associated with biological pathways: chemical carcinogenesis – reactive

oxygen species, calcium signaling pathways, and TNF signaling pathways. Furthermore, the associated or enriched gene functions from the GO terms found phosphorylation to be the most enriched function, while ATP binding and calcium ion binding play indirect but important roles in ARVC. Three important genes of GSE253223 are DYNC1LI2, DDX10, and PDCD6, which were taken by the analysis of this dataset.
Conclusion: Results from this study can contribute to the broader field of personalized medicine or gene therapy where treatments and preventive strategies are tailored to an individual's genetic makeup. The KEGG pathways, chemical carcinogenesis – reactive oxygen species, calcium signaling pathways, and TNF signaling pathways show how the genes included can relate to the cause of ARVC. The GO terms, phosphorylation, ATP binding, and calcium ion binding, can also contribute to the cause of ARVC.
KEYWORDS: Arrhythmogenic right ventricular cardiomyopathy (ARVC), Bioinformatics, NCBI GEO, DAVID, GO, KEGG

INTRODUCTION

Arrhythmogenic right ventricular cardiomyopathy, also known as ARVC, is a genetic heart disease that can cause irregular heartbeats and an increased risk of sudden death (Wu, Iris, et al. 2024). Currently, there is not much known about ARVC as it is a newer disease (Joshi-Mukherjee, Rosy, et al.2008). However, there is an understanding of its side effects, how it is genetic, and how it is mostly caused by mutations in the PKP2 gene, also known as plakophilin-2 (Stanford Health Care, 2017; Wu, Iris, et al. 2024). Around 40% of ARVC cases are related to changes or mutations in PKP2 (Wu, Iris, et al. 2024). PKP2 is responsible for creating a protein called plakophilin 2. Plakophilin 2 makes structures that are called desmosomes, which are formations in heart cells that help them stick together and function properly. When there is a lack of PKP2, ARVC will occur. Although there are many details about ARVC that are unclear, new bioinformatics research has added to our understanding by discovering a large number of implicated genes and pathways.

There are no known approved treatments that are able to address or heal the genetic causes of this disease (Chantal J.M., et al, 2019). Bioinformatics research can help find a potential cure or gene therapy for this disease. What is known so far is that ARVC is primarily caused by changes in DNA or

genes called genetic mutations (Wu, Iris, et al. 2024; Chantal J.M., et al, 2019). These mutations often affect genes that code for proteins in the desmosomes. When these proteins are defective due to genetic mutations, heart muscle cells can weaken and deteriorate, particularly in the right ventricles, leading to irregular heartbeats and an increased risk of sudden death (Chantal J.M., et al, 2019; Wu, Iris et al, 2024).

Scientists hope that personalized medicine or gene therapy can someday be used to help address PKP2 gene mutations. To treat ARVC, scientists hope to use AAV9, also known as adeno associated virus 9, to deliver the normal non-mutated PKP2 gene that could possibly restore the PKP2 expression to help heal the heart (Wu, Iris, et al. 2024).

Therefore, the goal of this study is to identify additional genes, gene functions and biological pathways that can be used to create gene therapy solutions. The scientific question being investigated is: what genes are expressed differently when PKP2 is removed from heart cells, and what are the associated functions and biological pathways of these genes? The hypothesis is that there will be differences in gene expression between controls and patient samples since PKP2 is essential to making PKP2 protein to avoid ARVC.

This research study is important as it contributes to potential identifications of additional genes and biological pathways that can potentially be used to find a cure for ARVC involving gene therapy.

METHODS

Choosing a Human Disease

To conduct this research, several bioinformatics tools were used, namely, NCBI, GEO2R, DAVID, KEGG, and GO (Gene Ontology) (Figure 1). ARVC was chosen as the human disease to study because of the important need to understand the genetic causes of ARVC, develop treatments, and advance the public health knowledge of ARVC.

Finding a GEO Dataset on NCBI

The dataset that was chosen is under the accession number GSE253223, and it focuses on the effects of changing PKP2 levels on induced pluripotent stem cell derived cardiomyocytes (iPSC-CMs, also known as heart cells) over time. This dataset consists of 20 samples of homo sapiens, 4 of which have a treatment named siNeg, and the last 16 have a treatment called siPKP2 (plakophilin-2). All samples have the same cell type.

Finding the Top DEG's With the P Value

To identify genes that are expressed differently in various samples, the samples were sorted into two groups. Then, the grouped data was analyzed using GEO2R, a bioinformatics tool, which resulted in over 50,000 identified differentially expressed genes (DEGs). Only the top 30 were statistically selected using the p-value. The DEGs were chosen with p-values < 0.05.

Analyzing the top DEG's with KEGG and GO charts

To further analyze the selected top 30 DEGs to determine their functions and associated biological pathways, DAVID, a functional annotation tool, was used (david). This categorized the chosen genes into KEGG pathways and GO terms. KEGG pathway analysis identified important biological pathways, whereas GO enrichment analysis provided information about the biological processes associated with the DEGs. Finally, DAVID put the top 30 genes into KEGG (KEGG) and GO (2024-01-17).

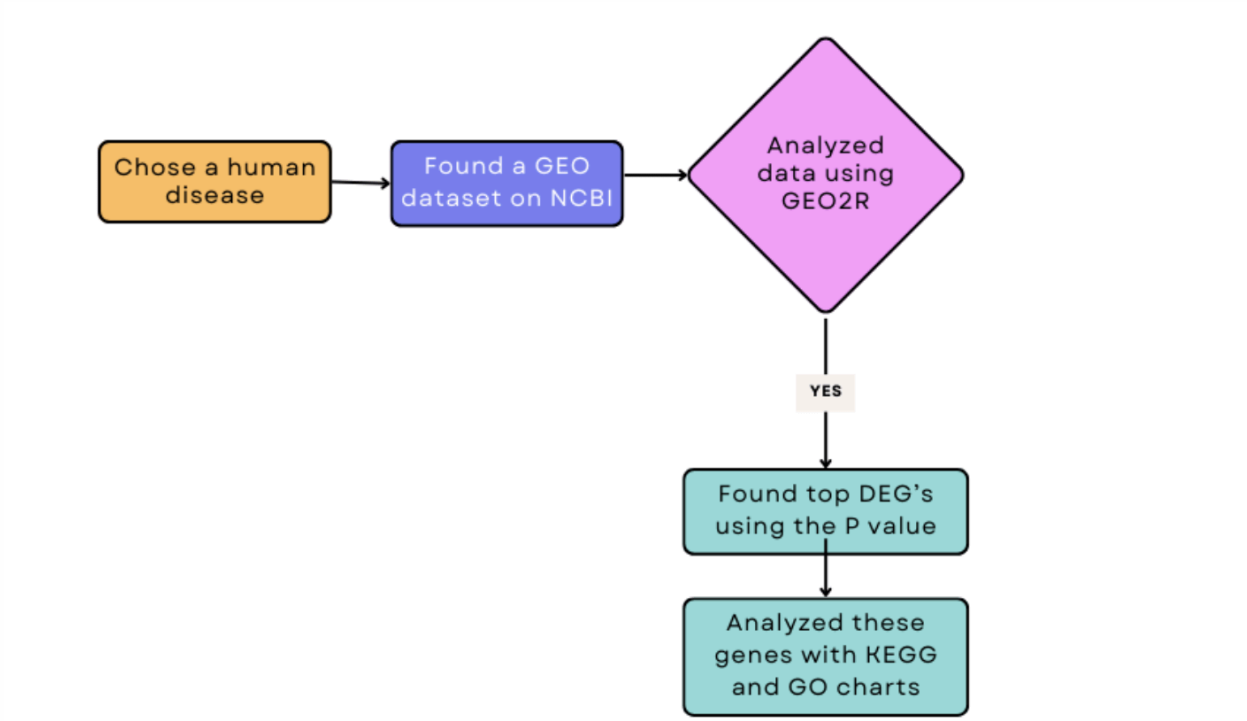


Figure 1: Methods Flowchart: Showing a summary of the methods used in this study to investigate PKP2 knockout in a mouse model.

RESULTS

Identification of Differentially Expressed Genes from GEO2R Analysis and Statistics

The DEGs that were obtained from GEO2R show how ARVC varies with each subject. Using the dataset, the top 30 DEGs were identified using the p-value (ARVC Top DEGs). All of the DEGs were upregulated from the data. GEO2R used R programming to analyze the results, which is pre-programmed with AI and ML (machine learning) algorithms. The R script detailing this can be found in this link.

Identification of Biological Pathways from DAVID Analysis

The biological pathways from the DAVID analysis show how some genes can be related to ARVC or other human diseases. Some of the pathways that were picked were the calcium signaling pathway, chemical carcinogenesis – reactive oxygen species, and the TNF signaling pathway. These three

pathways were chosen for this research because they are the most related to ARVC.

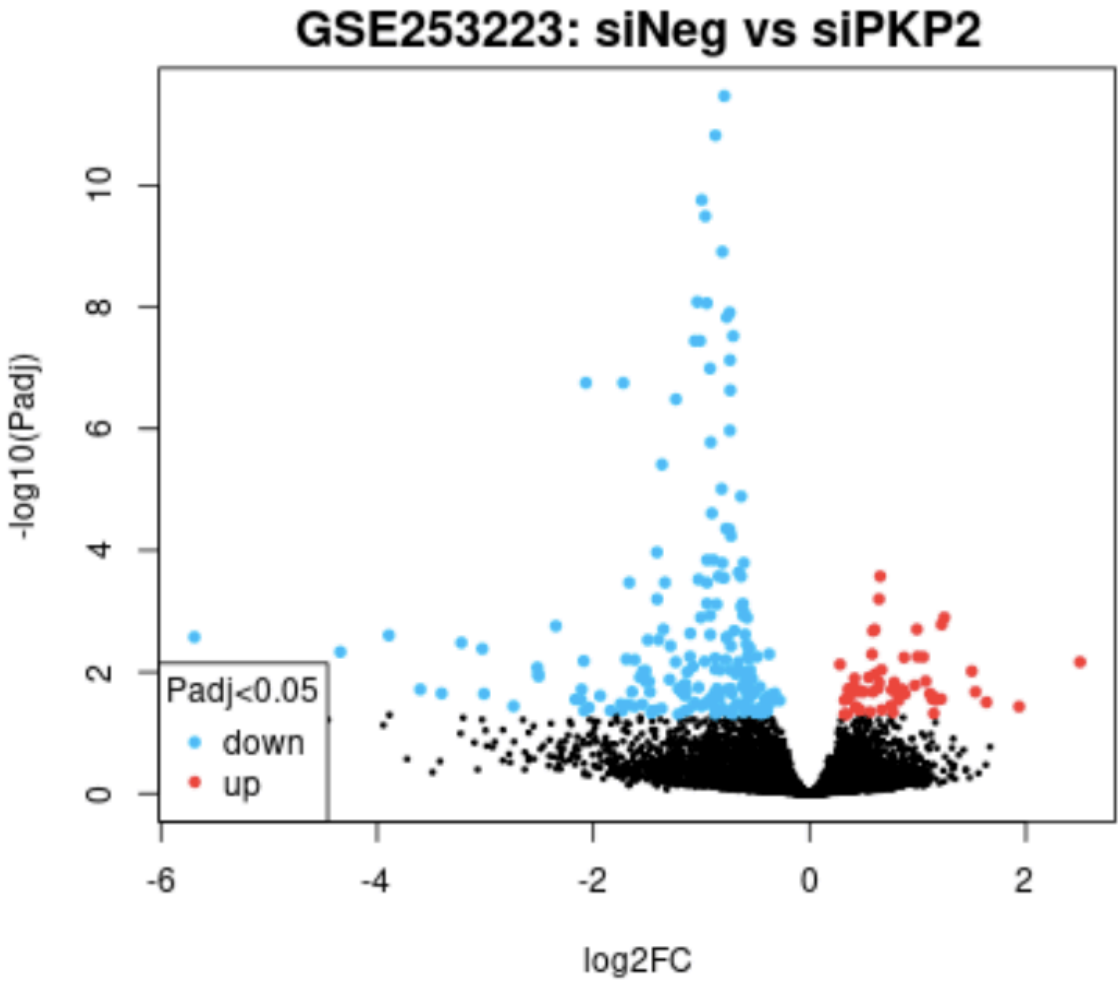


Figure 2: The volcano chart showing the upregulated and downregulated DEGs; the blue dots representing downregulation, and red dots representing upregulation. Black is neutral.

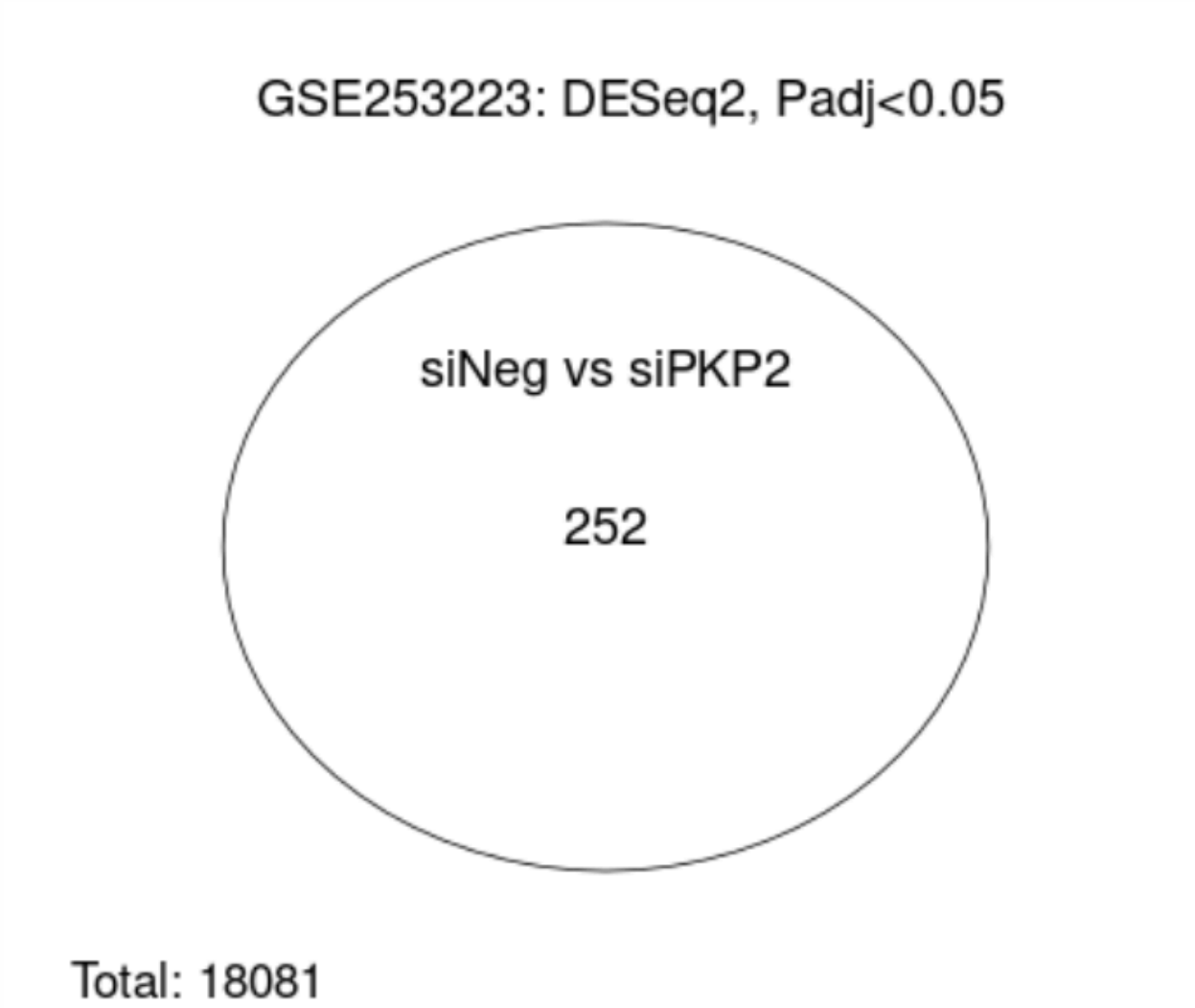


Figure 3: The venn diagram shows the similarities between the two treatments, siNEG and siPKP2

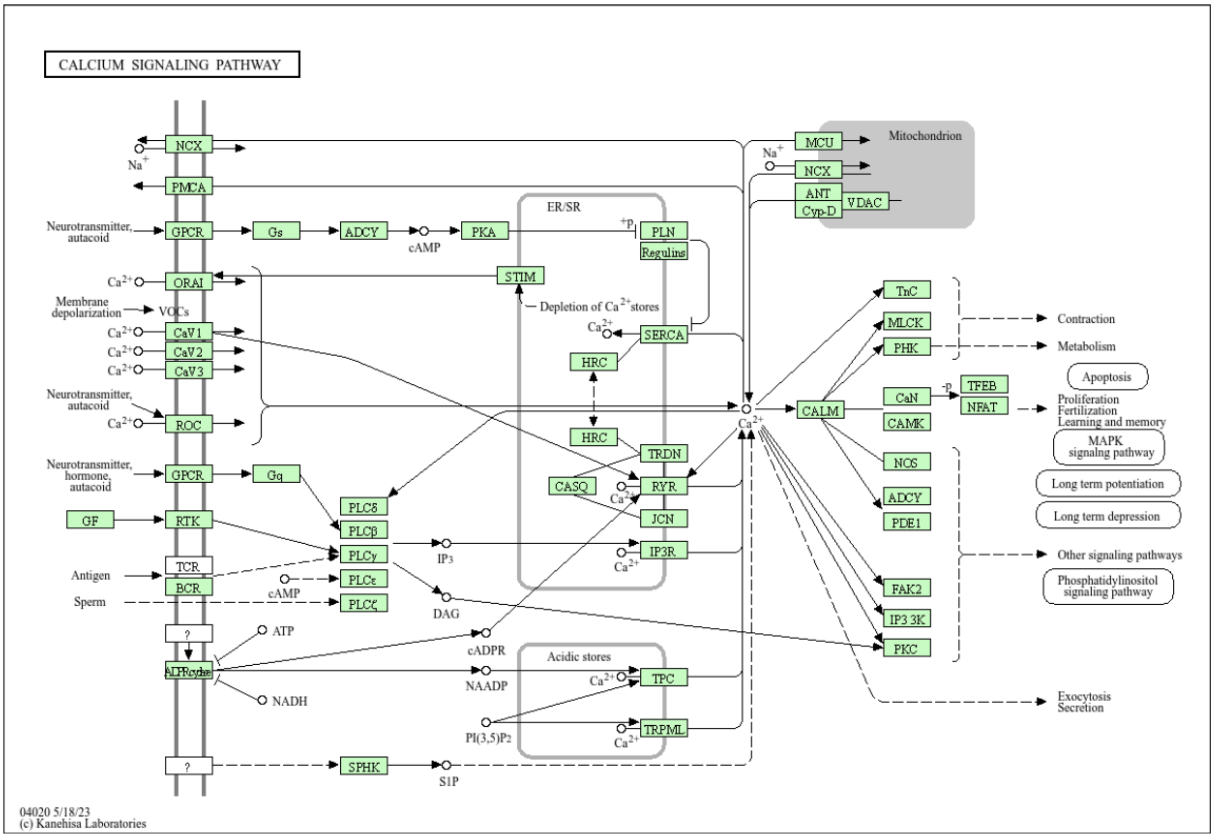


Figure 4: Shows calcium signaling In heart cells, necessary in order to maintain heart rhythm.

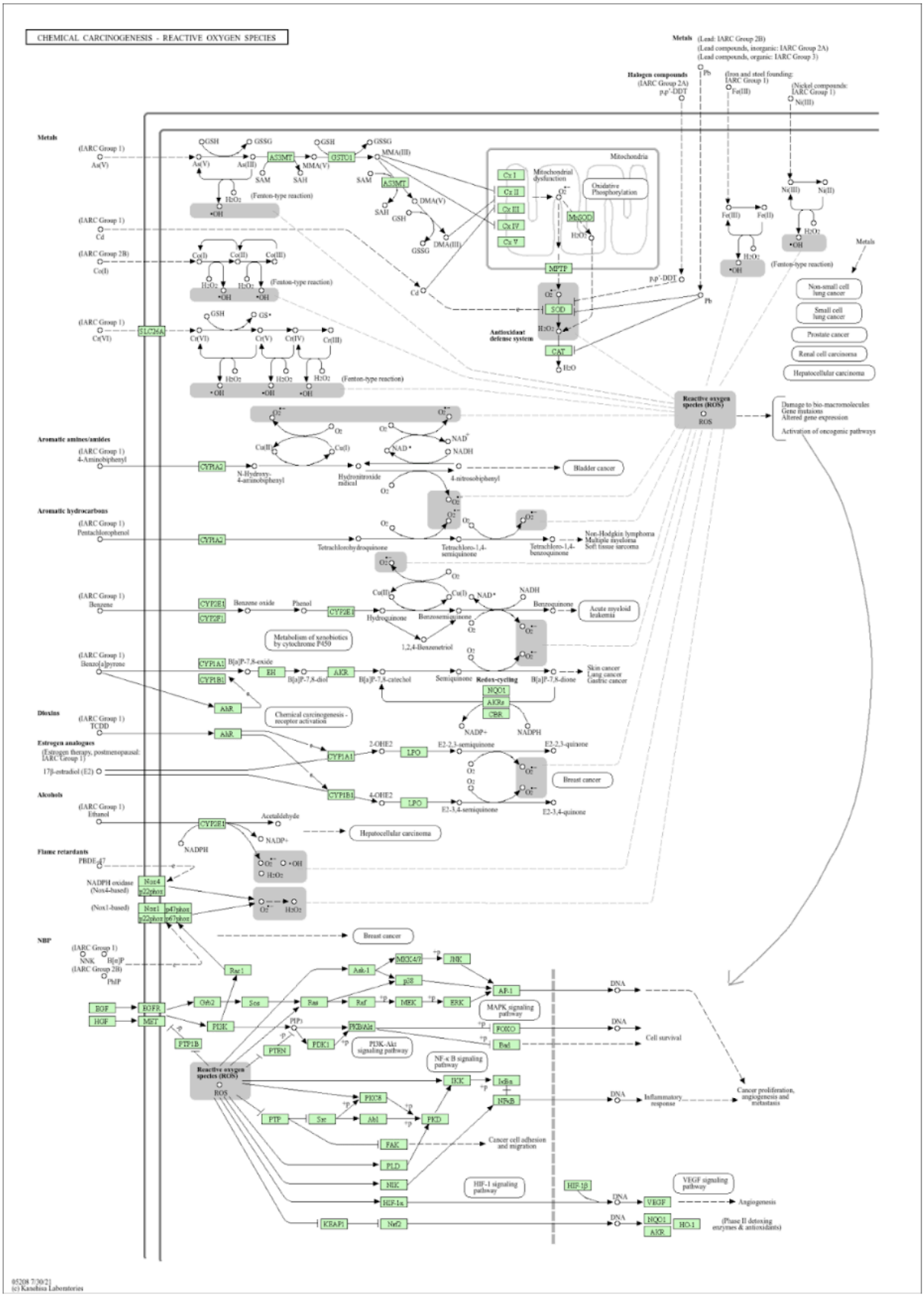

CHEMICAL CARCINOGENESIS - REACTIVE OXYGEN SPECIES
Metals
(Lead: IARC Group 2B)
(Lead compounds, inorganic: IARC Group 2A)
(Lead compounds, organic: IARC Group 3)
Halogen compounds
(IARC Group 2A)
p,p'-DDT
Pb
(Iron and steel founding: IARC Group 1)
Fe(III)
(Nickel compounds: IARC Group 1)
Ni(III)
Metals
(IARC Group 1)
As(V)
GSH
GSSG
As(III)
AS3MT
MMA(V)
GSTO1
MMA(III)
SAM
SAH
DMA(V)
DMA(III)
(Fenton-type reaction)
H2O2
•OH
Mitochondria
Cx I
Cx II
Cx III
Cx IV
Cx V
Mitochondrial dysfunction
Oxidative Phosphorylation
O2•-
Mn-SOD
MPTP
(IARC Group 1)
Cd
(IARC Group 2B)
Co(I)
Co(II)
Co(III)
(IARC Group 1)
Cr(VI)
SLC26A
GS•
Cr(V)
Cr(IV)
Cr(III)
Antioxidant defense system
SOD
CAT
H2O
Non-small cell lung cancer
Small cell lung cancer
Prostate cancer
Renal cell carcinoma
Hepatocellular carcinoma
Reactive oxygen species (ROS)
ROS
Damage to bio-macromolecules
Gene mutations
Altered gene expression
Activation of oncogenic pathways
Aromatic amines/amides
(IARC Group 1)
4-Aminobiphenyl
CYP1A2
N-Hydroxy-4-aminobiphenyl
Hydronitroxide radical
4-nitrosobiphenyl
Bladder cancer
Aromatic hydrocarbons
(IARC Group 1)
Pentachlorophenol
Tetrachlorohydroquinone
Tetrachloro-1,4-semiquinone
Tetrachloro-1,4-benzoquinone
Non-Hodgkin lymphoma
Multiple myeloma
Soft tissue sarcoma
(IARC Group 1)
Benzene
CYP2E1
CYP2F1
Benzene oxide
Phenol
Hydroquinone
Benzosemiquinone
Benzoquinone
Acute myeloid leukemia
1,2,4-Benzenetriol
Metabolism of xenobiotics by cytochrome P450
(IARC Group 1)
Benzo[a]pyrene
CYP1A1
CYP1B1
EH
AHR
B[a]P-7,8-oxide
B[a]P-7,8-diol
B[a]P-7,8-catechol
Semiquinone
Redox-cycling
B[a]P-7,8-dione
Skin cancer
Lung cancer
Gastric cancer
NQO1
AKRs
CBR
NADP+
NADPH
Dioxins
(IARC Group 1)
TCDD
AhR
Chemical carcinogenesis - receptor activation
Estrogen analogues
(Estrogen therapy, postmenopausal: IARC Group 1)
17β-estradiol (E2)
2-OHE2
LPO
E2-2,3-semiquinone
E2-2,3-quinone
4-OHE2
E2-3,4-semiquinone
E2-3,4-quinone
Breast cancer
Alcohols
(IARC Group 1)
Ethanol
CYP2E1
Acetaldehyde
Hepatocellular carcinoma
Flame retardants
PBDE-47
NADPH oxidase (Nox4-based)
(Nox1-based)
Nox4
p22phox
Nox1
p47phox
p67phox
NBP
(IARC Group 1)
NNK
B[a]P
(IARC Group 2B)
PhIP
Breast cancer
Rac1
ROR
HGF
EGFR
MET
Grb2
Sos
PI3K
Ras
Raf
MEK
ERK
MAPK signaling pathway
AP-1
DNA
Ank-1
MLK4/7
JNK
p38
PIP3
PDK1
PTEN
PI3K-Akt signaling pathway
PKB/Akt
FOXO
Bad
Cell survival
NF-κ B signaling pathway
IKK
IκBα
NFκB
Inflammatory response
Cancer proliferation, angiogenesis and metastasis
PTP
Src
Abl
PKCδ
FAK
Cancer cell adhesion and migration
PLD
NIK
HIF-1α
HIF-1 signaling pathway
HIF-1β
VEGF signaling pathway
VEGF
Angiogenesis
KRAP1
Nrf2
HO-1
(Phase II detoxing enzymes & antioxidants)
05208 7/30/21
(c) Kanehisa Laboratories

Figure 5: Chemical carcinogenesis involves oxidative stress leading to cancer. Even though PKP2 is not directly related, oxidative stress can impact cardiac cells and adhesion properties.

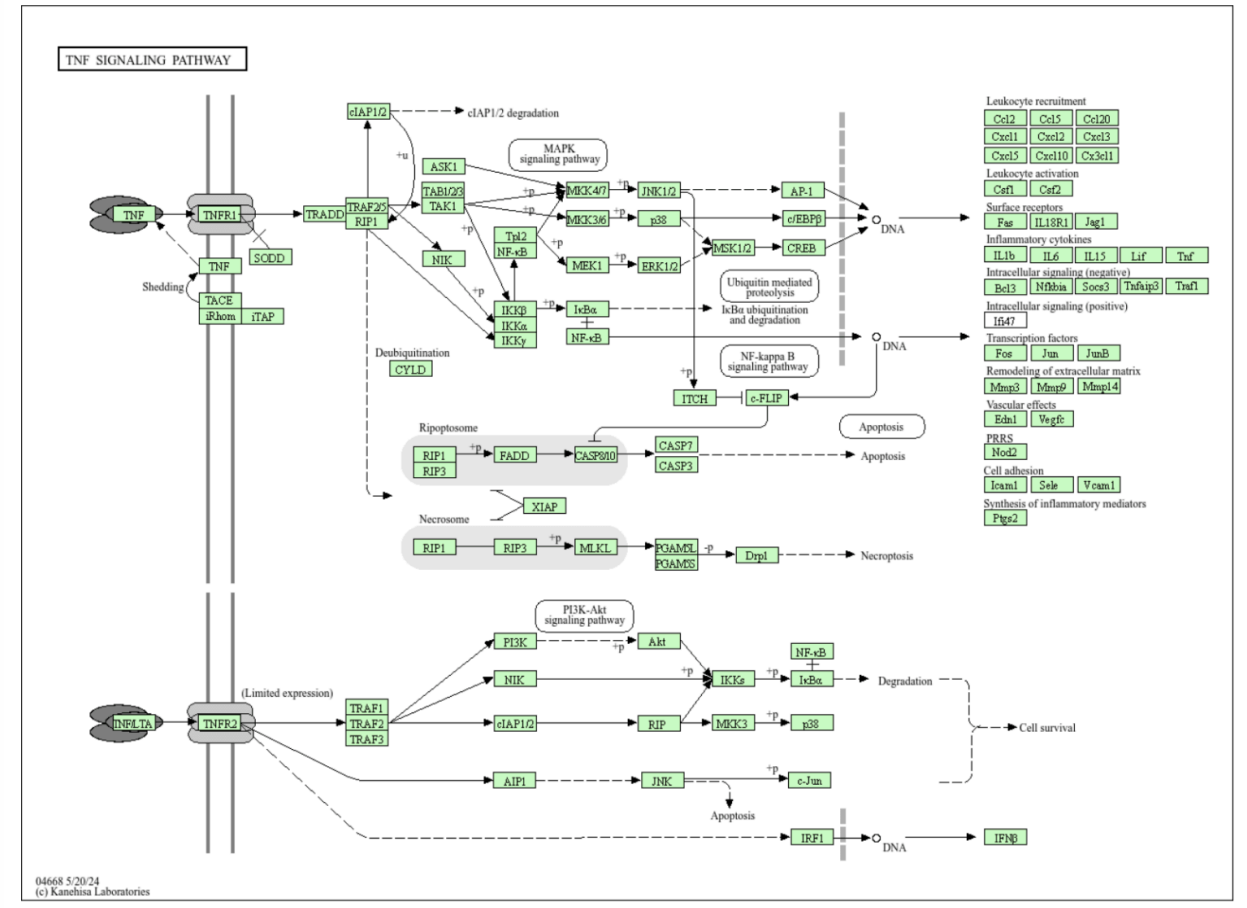


Figure 6: TNF signaling is involved with inflammation, cell survival, and immune responses. A certain cytokine in this pathway may promote inflammation in cardiac tissue, as well as fibrosis. This can lead to the replacement of normal heart muscle that contributes to heart failure.

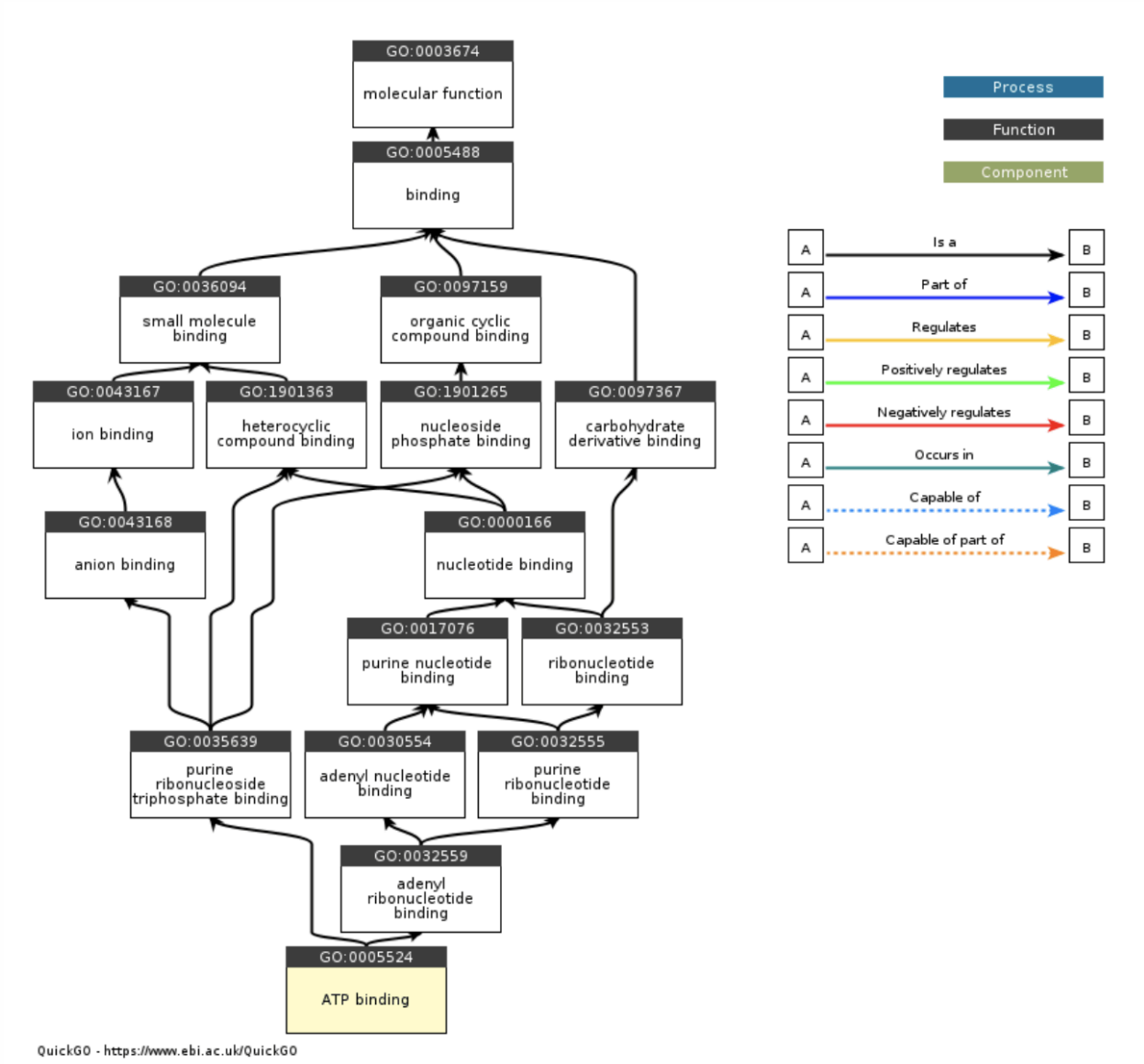


Figure 7: The role that ATP binding plays in desmosomes shows that ATP-related processes may play a role in the development of ARVC. It is possible that it can help develop ARVC through cellular pathways.

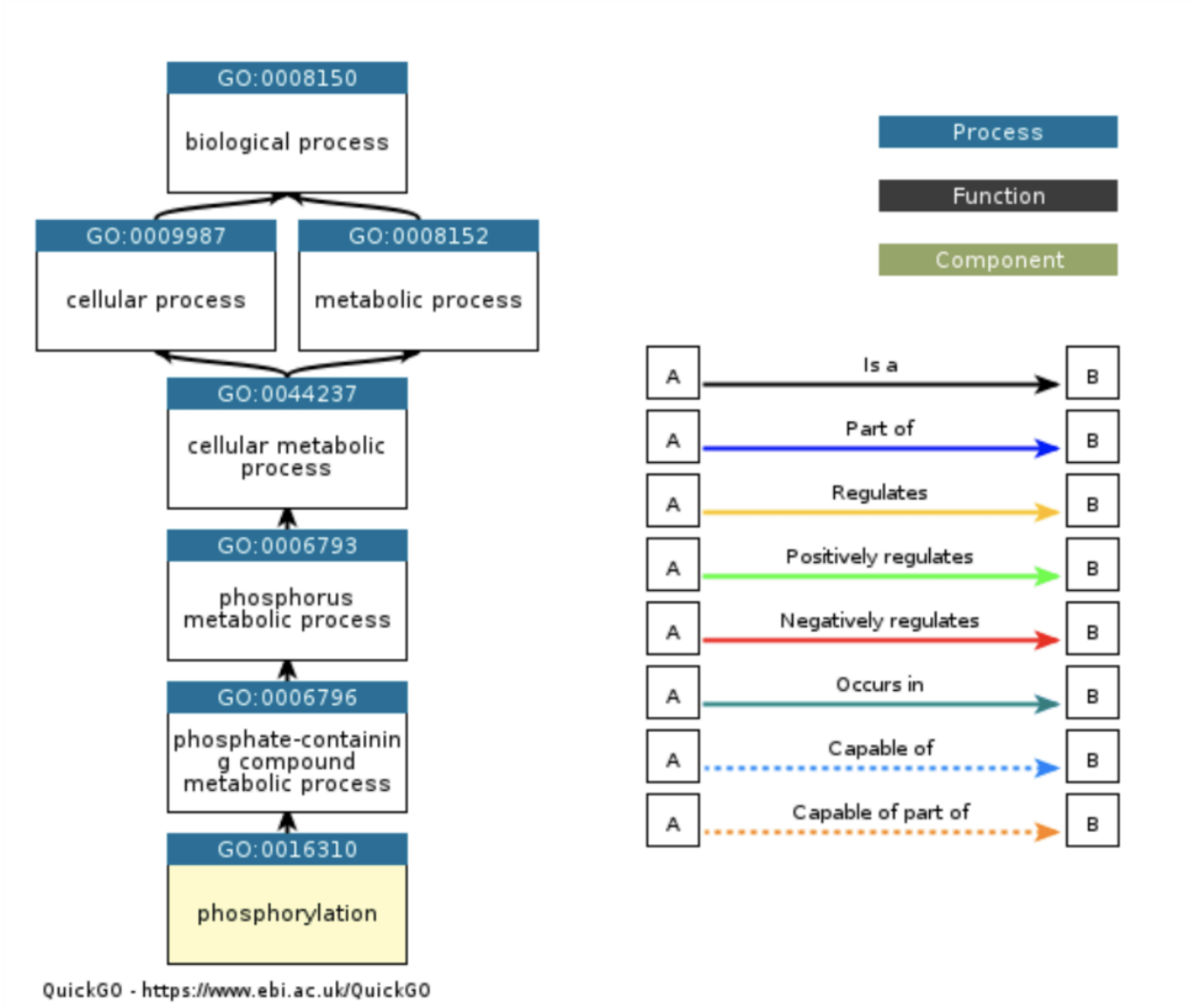


Figure 8: PKP2 can undergo phosphorylation, which may affect the role of signaling and cell adhesion. It can also influence other proteins and the structural integrity of cardiac tissue.

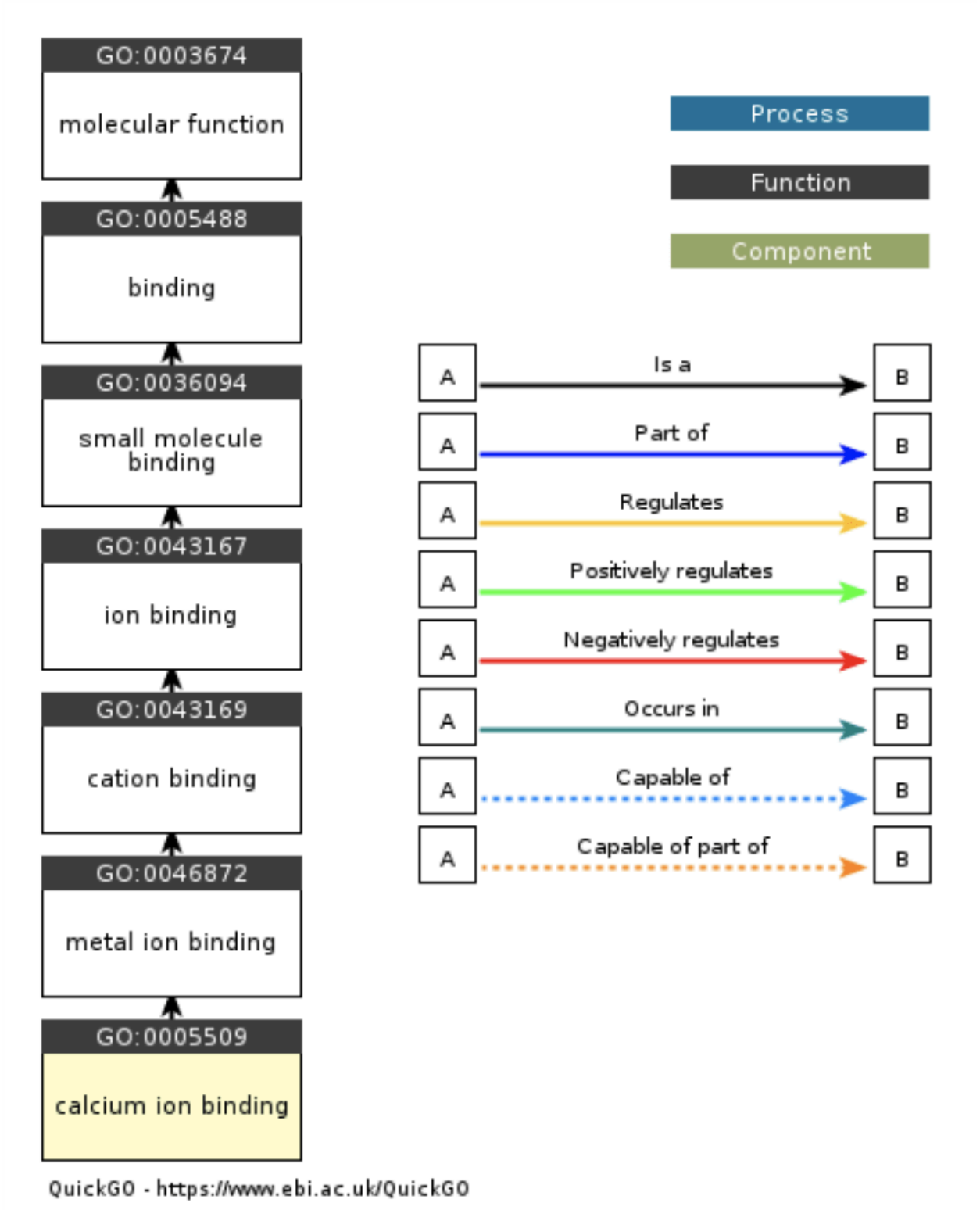


Figure 9: PKP2 does not bind calcium ions directly, but it helps other proteins inside the desmosomes.

DISCUSSION

The main goal of this research was to identify additional genes, pathways, and functions that could contribute to developing a cure for ARVC. This study revealed how it is unclear which genes are up and down regulated (figure 2), and how siNEG and siPKP2 are extremely similar to each other (figure 3). Because these results were vague, it did not support the hypothesis. In order for these results to support the hypothesis, there would need to be more and clearer results. For the DAVID results, while PKP2 is not directly related to calcium signaling pathway (figure 4), chemical carcinogenesis (figure 5), or TNF signaling (figure 6), these pathways can affect its adhesion properties (Novelli, Valeria, et al. 2018). Even though calcium ion binding (figure 9) is indirectly related to PKP2's function, it can influence the desmosomal function (Chantal J.M., et al 2019). Additionally, PKP2 can undergo phosphorylation (figure 8), which can affect its role in cell adhesion and signaling. Although ATP binding is not directly related to PKP2, it is part of the desmosomes, suggesting that ATP-related processes may depict a role in the development of ARVC (figure 7). It is possible that some cellular processes in PKP2 involve ATP binding proteins. These results align with the hypothesis by providing a better understanding of how gene expression can differ when PKP2 is absent.

Comparing these results to previous studies, the previous studies were clearer because this research was unable to use a clinical setting to find more in depth results. Other studies, like one's using mouse models, have succumbed to a more in depth understanding of PKP2's importance (Wu, Iris, et al. 2024). These results from this study can help develop a vaccine or cure for ARVC, especially given the limited data that is currently available in this topic. These results could further assist scientists or other researchers to better understand ARVC.

Limitations

Some limitations are, of course, not having as much information as preferred, and having trouble finding samples to analyze. Since there is reliance on bioinformatics datasets from microarray experiments made by other researchers, one drawback is that the genes that were identified will require additional research in a laboratory or clinical setting before a cure for ARVC can be developed. For future research, the identified genes should be

tested in a laboratory to determine if another gene could counteract or prevent the development of ARVC.

Table 1: Summary Table of key gene functions and biological pathways identified in this study

GO results	KEGG results	ARVC connection
ATP binding	Chemical carcinogenesis – reactive oxygen species	ATP – Mitochondria is the powerhouse of the cell. In ARVC, there may be a dysfunction, leading to less ATP being produced. This results in damage to cardiac cells. ROS = very oxygen reactive molecules. Excessive ROS leads to damaged DNA and proteins. This can also lead to cancer.
Phosphorylation	Calcium signaling pathway	Abnormal phosphorylation can lead to the development of ARVC. Mutations of desmosomal proteins can lead to arrhythmias, which is when calcium is disrupted
Calcium ion binding	TNF signaling pathway	In TNF-α, more apoptosis of cardiomyocytes can contribute to ARVC. Disrupted calcium homeostasis can lead to reduced cardiac

		contraction.

Conclusion

The goal of this study was to identify gene functions and pathways that can contribute to the development of a cure for ARVC. Regarding the research question, the genes picked were involved in the causes of ARVC and human diseases. The values from GEO2R were analyzed using DAVID, which resulted in GO terms and KEGG pathways. Key GO terms that were acquired included phosphorylation, calcium ion binding, and ATP binding. The KEGG pathways selected were calcium signaling pathway, chemical carcinogenesis – reactive oxygen species, and the TNF signaling pathway. Moreover, regarding the scientific question, how many genes were expressed differently when PKP2 was removed from the heart cells remains unclear. However, the biological pathways identified are closely related to the desmosomes, which helps develop ARVC. The hypothesis was not supported, accentuating that many factors that contribute to ARVC remain understood, likely due to the relatively recent discovery of this disease. Some suggestions for further research are to address other causes of ARVC and then conduct additional studies to develop a treatment for ARVC, since this study's findings only focused on one aspect of ARVC. Although there is no official cure yet, this study's findings contribute many insights which may eventually lead to the development of a cure.

References

1. "About GEO2R." (2024). NCBI GEO. https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html
2. Cerrone, M., Lin, X., Zhang, M., Agullo-Pascual, E., Pfenniger, A., Chkourko-Guskov, K., …, & Delmar, M. (2017). Plakophilin-2 is required for transcription of genes that control calcium cycling and cardiac rhythm. Nature Communications, 8(1), 1-12. https://www.nature.com/articles/s41467-017-00127-0
3. Chantal, J. M., Morais, P., Moreira, S., Leite-Moreira, A. F., & Branco, L. M. (2019). Plakophilin-2 haploinsufficiency causes calcium handling deficits and modulates the cardiac response towards stress. International Journal of Molecular Sciences, 20(17), 4076. https://www.mdpi.com/1422-0067/20/17/4076
4. Dalal, D., Nasir, K., Bomma, C., Tandri, H., Dong, J., & Tichnell, C. (2006). Clinical features of arrhythmogenic right ventricular

dysplasia/cardiomyopathy associated with mutations in plakophilin-2. Circulation, 113(13), 1641-1649. https://pubmed.ncbi.nlm.nih.gov/16549640/
5. Giordano, F. J. (2005). Oxygen, oxidative stress, hypoxia, and heart failure. Journal of Clinical Investigation, 115(3), 500-508.
6. He, Y., Rappuoli, R., De Groot, A. S., & Chen, R. T. (2010). Emerging vaccine informatics. Journal of Biomedicine and Biotechnology, 2010, 1-10.
7. Joshi-Mukherjee, R., Coombs, W., Izumo, S., & Delmar, M. (2008). Characterization of the molecular phenotype of two arrhythmogenic right ventricular cardiomyopathy (ARVC)-related plakophilin-2 (PKP2) mutations. Heart Rhythm, 5(12), 1715-1723. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2636742/
8. Kohan, R., & Pastore, A. (2020). Reactive oxygen species in cancer: A paradox between pro- and anti-tumour activities. Cancer Chemotherapy and Pharmacology, 85(4), 1-10. https://link.springer.com/article/10.1007/s00280-020-04103-2
9. Lindsey, S. H., Carbone, J. M., & Lindsey, P. W. (2005). Tumor necrosis factor signaling and the induction of heart failure: What can we learn from mouse models? Current Heart Failure Reports, 2(4), 195-201.
10. Maas, W. J., Knook, D. L., & Wisse, E. (2000). Assessment of some critical factors in the freezing technique for the cryopreservation of precision-cut rat liver slices. Cryobiology, 40(3), 261-272. https://pubmed.ncbi.nlm.nih.gov/10860624/
11. Novelli, V., Mazzanti, A., Pagan, E., Sibona, M., Robyns, T., & Calore, M. (2018). Pleiotropic phenotypes associated with PKP2 variants. Frontiers in Cardiovascular Medicine, 5, 184. https://www.frontiersin.org/articles/10.3389/fcvm.2018.00184/full
12. Pérez-Hernández, M., Tapia, M., Mechulan, A., & Álvarez, D. (2021). Transcriptomic coupling of PKP2 with inflammatory and immune pathways endogenous to adult cardiac myocytes. Frontiers in Physiology, 11, 623190. https://www.frontiersin.org/articles/10.3389/fphys.2020.623190/full
13. Saguner, A. M., Brunckhorst, C., & Duru, F. (2014). Arrhythmogenic ventricular cardiomyopathy: A paradigm shift from right to biventricular disease. World Journal of Cardiology, 6(4), 154-174.
14. Stanford Health Care. (2017). Symptoms of arrhythmogenic right ventricular dysplasia/cardiomyopathy.

https://stanfordhealthcare.org/medical-conditions/blood-heart-circulation/arrhythmogenic-right-ventricular-dysplasia-cardiomyopathy/symptoms.html
15. Wu, I., Loyer, X., Simard, F., Bortnick, A., & Seidman, C. (2024). AAV9
16. improves heart function and survival in a PKP2-deficient mouse model of arrhythmogenic right ventricular cardiomyopathy. Nature Communications, 9, 450. https://www.nature.com/articles/s43856-024-00450-w
17. Wu, I., Mereles, D., Meder, B., & Giannitsis, E. (2024). GEO accession viewer: GSE253223. NCBI GEO. https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE253223
18. Zorio, E., Medina, P., Rueda, J., Millán, J. M., de la Calle, G., & Gilabert-Estellés, J. (2009). Insights into the molecular basis of familial cardiomyopathies: From bench to bedside. Applied Microbiology and Biotechnology, 85(1), 1-18.

Ria Tandra

The Role Of The Dopamine Transporter Gene In Dopamine Dysregulation In Attention-Deficit/Hyperactivity Disorder

ABSTRACT

Background: Attention-deficit/hyperactivity disorder (ADHD) is a widespread neurodevelopmental disorder characterized by symptoms of inattention, hyperactivity, and impulsivity. The dopamine transporter gene (DAT1) plays a significant role in dopamine dysregulation. DAT1 encodes the dopamine transporter protein, which regulates the reuptake of dopamine in the brain.

Objectives: This article explores the role of the DAT1 gene and its genetic variations in ADHD pathophysiology. Additionally, this review explores patient responses to treatments like methylphenidate (MPH), and its effects on cognitive functioning.

Methods: This review article used sources from search engines such as PubMed and the National Institute of Health. Keywords searched included ADHD, DAT1, and dopamine. This article excludes sources not written in English.

Results: Results show that specific DAT1 variations are linked to symptoms of ADHD such as inattention, impulsivity, and medical treatments. Furthermore, DAT1 variations work with environmental factors and patterns in DNA methylation to influence emotional and behavioral traits in individuals with ADHD.

Conclusion: This information supports the role of the DAT1 gene in ADHD and pushes for more personalized medical approaches for treating the disorder and detecting ADHD early. Future studies should focus on interactions between the DAT1 gene and its environment and pharmacological advancements to improve ADHD treatments.

INTRODUCTION

Attention-deficit/hyperactivity disorder (ADHD) is a widespread neurodevelopmental disorder that affects millions of children and adults worldwide[1]. It is characterized by symptoms of inattention, hyperactivity, and impulsivity [1]. While the exact causes of ADHD are not fully

understood, research has linked the disorder to neurobiological factors like dopamine dysregulation [2]. Dopamine is a neurotransmitter that plays an important role in regulating executive functions like attention and motivation, which are impaired in individuals with ADHD [3]. Many genetic factors influence dopamine dysregulation in ADHD, but the dopamine transporter gene (DAT1) was found to be a key factor due to its involvement in the structure of ADHD in the brain [4].

The DAT1 gene codes for the dopamine transporter protein, which is responsible for the reuptake of dopamine from the synaptic cleft back into the presynaptic neuron [5]. This process is crucial because of how important it is to maintain the balance of dopamine levels in the brain, especially in areas that are mainly associated with cognitive function, such as the prefrontal cortex [6]. Variations or polymorphisms in the DAT1 gene, especially the 40-base pair variable number tandem repeat (VNTR) polymorphism, have been shown to impact the effectiveness of the reuptake of dopamine [7]. These genetic variations can either lead to an overactive or underactive dopamine system, which can result in defects in cognitive and behavioral functioning that ultimately could be considered symptoms of ADHD [8].

This review explores the role of the DAT1 gene in the dysregulation of dopamine associated with ADHD, with a specific focus on genetic variations, their influence on patient response to pharmaceutical treatments, and their impact on cognitive performance in patients. Research has shown that individuals with certain polymorphisms of the DAT1 gene could exhibit a variety of responses to medications such as methylphenidate (MPH), which works by inhibiting the reuptake of dopamine to increase the amount of dopamine that is available in the brain [9]. Additionally, the relationship between the DAT1 variations and cognitive functions such as spatial working memory (SWM), attention, and impulsivity has been thoroughly researched, which ultimately supports the existing information about the mechanisms by which these genetic factors contribute to ADHD symptoms [10].

This review will examine all previous research on the relationship between DAT1 gene variations, medication response, and cognitive functioning to provide a detailed understanding of how all of these factors contribute to ADHD symptomatology. Furthermore, it will utilize these findings to suggest

potential personalized medical approaches to ADHD treatment and highlight the need for further research to reveal the complexities behind the interactions between genes and their environment that are associated with ADHD.

METHODS

To obtain the reviewed articles, PubMed and Google Scholar were searched using keywords including ADHD, DAT1, dopamine, dopamine dysregulation, and polymorphism. Articles published within the last 25 years were limited to those selected based on their relevance to the research question. This review article excludes articles not written in English.

RESULTS

Albrecht et. al showed that children with ADHD exhibited lower Cue-P3 amplitudes, meaning that they exhibited decreased attention or ability to process cues [11]. This can occur when patients are unmotivated or tired, both of which are factors that can affect an individual with ADHD [11]. These children also exhibited lower CNV (contingent negative variation) amplitudes, signifying that they had reduced activity in the brain when it came to anticipating a cue [11]. Lastly, they exhibited lower Nogo-P3 amplitudes, meaning that in terms of "Go/NoGo" signals, there are problems when responding [11]. It was also found that children with the 7-repeat (7R) allele showed overall reduced cognitive functioning, coupled with lower Cue-P3 and CNV amplitudes; however, their Nogo-P3 amplitude was not impacted by the allele [11]. On the other hand, children with the DAT1 10-6 haplotype showed normal cognitive functioning and their Cue-P3 and CNV amplitudes were normal, but the Nogo-P3 amplitudes increased [11]. There was found to be no relationship between genotype and ADHD symptomatology [11].

Shang and Gau looked at how variations in the DAT1 gene might be related to ADHD and the ability to perform tasks from memory [12]. Three regions of the gene, called haplotype blocks, were identified, but none of them showed a strong relationship to ADHD overall [12]. Two specific genetic variations of the DAT1 gene, rs27048 and rs429699, showed some connection, but it wasn't strong enough to be considered significant for the experiment [12]. When these two variations were combined into a haplotype block, however, there was a strong connection to the inattentive subtype of

ADHD [12]. Additionally, a haplotype in one gene region was also tied to making more errors in cognitive functioning, and this stayed significant after further considerable testing [12]. Lastly, similar results were found when only looking at male participants [12].

Jeong et. al found no association between the 25-item Wender Utah Rating Scale (WURS-25) score and any of the DAT1 genetic markers [13]. On this scale, a score of 46 or higher suggests that an individual has a stronger chance of being diagnosed with ADHD in their childhood [14]. With a score lower than 46, it is unlikely that childhood ADHD will be present [14]. It was found that the trait of mood instability was significantly linked to the rs2937639 variation in male participants [13]. This finding was also supported by analyzing haplotype blocks with nearby variations on the gene, which showed strong associations with the mood instability trait as well [13].

Hasler et. al found significant differences in gene expression when specifically focusing on the DAT1 40-bp VNTR, DRD4 48-bp VNTR, and DRD2-141C Ins/del polymorphisms [15]. The DRD2-141C Ins/del polymorphism refers to a specific variation of the dopamine receptor gene D2, which plays a significant role in dopamine signaling in the brain [15]. This polymorphism showed no significant differences when comparing individuals with ADHD and the control [15]. The DAT1 9.5-repeat (9.5R) allele was not seen as much in individuals with ADHD and it was also linked to lower scores for impulsivity and anger [15]. However, after multiple tests, these findings were not replicated enough to be significant [15]. The DRD4 6-repeat (6R) allele was more common in individuals with ADHD because specific genotypes like the 4-repeat (4R) or 6R allele were associated with higher impulsivity scores and lower anger scores [15]. Additionally, individuals who suffered from physical abuse either in their childhood or adulthood with the 6R allele showed higher impulsivity scores compared to those without the 6R allele [15].

Loo et. al showed that the DAT1 allele influenced both cognitive functioning and electroencephalogram (EEG) responses to medication in children with ADHD [16]. Children with two copies of the 10-repeat allele (10R) showed decreased ability to complete a task compared to those with one or more copies of the 9-repeat (9R) allele [16]. Additionally, children with the 10R allele showed changes in EEG activity after administering medication. They

had increased β power in the central and parietal regions of the brain, which indicates increased levels of focus and brain activity [16]. They also had decreased θ power in the frontal region of the brain, meaning they were less relaxed or drowsy [16]. Lastly, they also showed lower θ/β ratios, which measure attention and alertness [16]. In contrast, children who possessed the 9R allele showed the opposite pattern, which suggests that based on their DAT1 allele type, children can exhibit different responses to medication [16].

Bellgrove et. al explored the possibility that the DAT1 genotype potentially influences left-sided or right-sided inattention in children with ADHD [17]. Children with the high-risk DAT1 genotype, meaning they carry two copies of the 10R allele, exhibited a rightward spatial bias and left-sided inattention, while those with the low-risk genotype showed a leftward spatial bias and right-sided inattention [17]. Additionally, children who carried the high-risk DAT1 genotype significantly improved their ADHD symptoms after being treated with methylphenidate (MPH), specifically with hyperactivity and impulsivity [17]. There was no effect on attentional cue processing [17]. These findings suggest that the DAT1 genotype–both high-risk and low-risk–plays a role in both attentional cue processing and response to MPH in individuals with ADHD [17].

Marzilli et. al looked into how children's ADHD symptoms could potentially be associated with DNA methylation at various CpG (cytosine-phosphate-guanine) sites, specifically M1, M2, M6, and M3 [18]. DNA methylation at the M1 site was strongly connected to ADHD symptoms, while methylation at the M2 and M6 were weakly correlated [18]. Both maternal and paternal environmental factors, including childhood stress, also impacted children's DNA methylation levels, but they were additionally regulated by the child's DAT1 genotype [18]. Maternal factors, in particular, showed stronger connections with DNA methylation in children with the 10/10 genotype, while paternal factors showed a stronger connection in children with the 9/x genotype [18]. These findings suggest that both genetic and environmental factors play a role in the DNA methylation patterns associated with ADHD and emotional/behavioral dysregulation in

children [18].

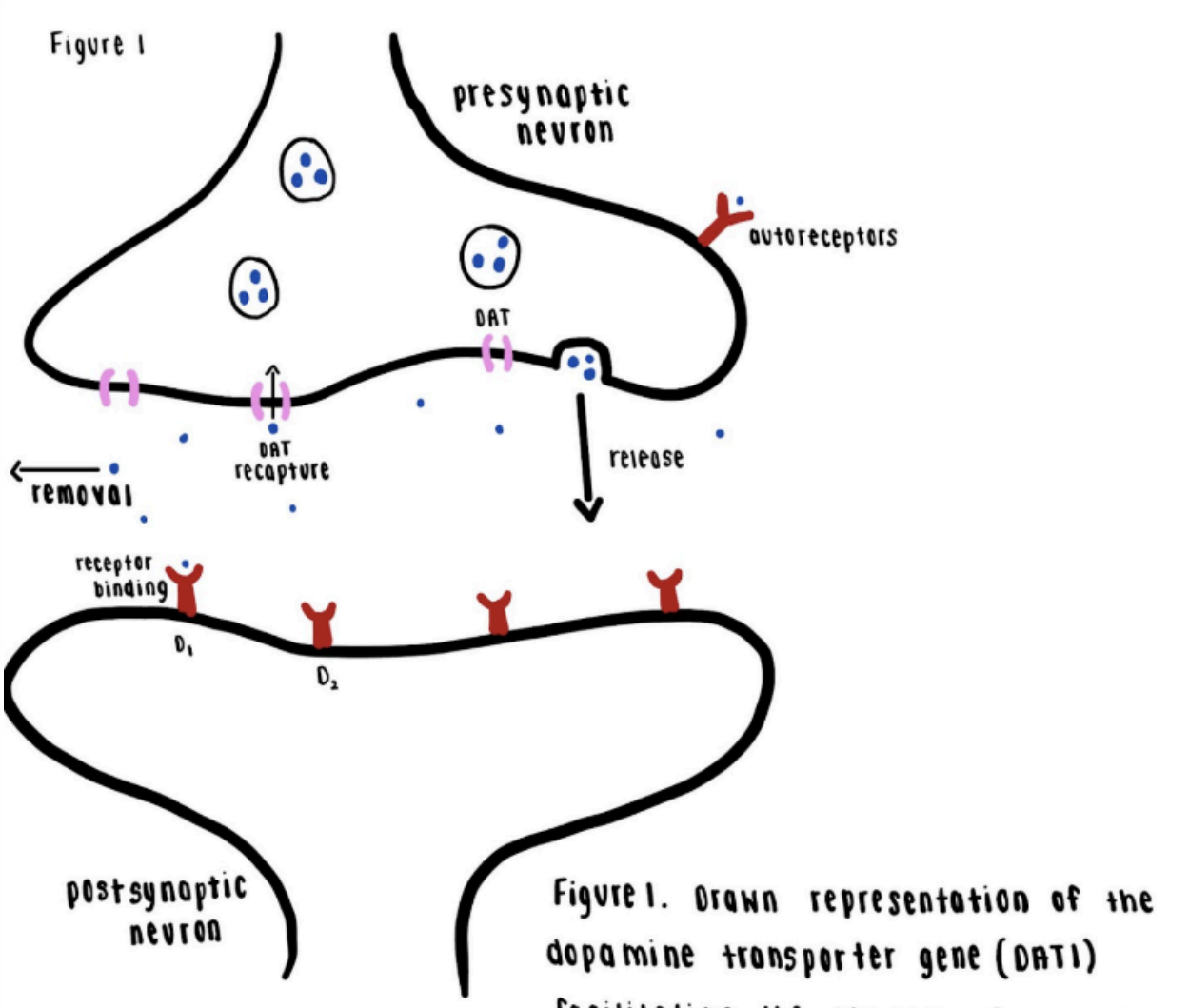


Figure 1. Drawn representation of the dopamine transporter gene (DAT1) facilitating the reuptake of dopamine from the synaptic cleft back into the presynaptic neuron.

DISCUSSION

This review article provides essential information when it comes to the complex and nuanced relationship between the DAT1 gene polymorphism and the behavioral and cognitive functioning of individuals with ADHD. These findings align with the increasing amount of research that suggests that genetic factors play a crucial role in the development of ADHD symptoms in children and adults [18]. Additionally, the data in this article continues to support the idea that these genetic factors influence a variety of emotions

and behaviors, particularly impulsivity, anger, hyperactivity, and inattention, all of which are central traits associated with ADHD [15].

The DAT1 Gene And Its Role In Emotional And Behavioral Functioning

Hasler et al. focus specifically on the role of the DAT1 polymorphism in ADHD symptoms, specifically impulsivity and inattention [15]. They found that variations in the DAT1 gene influenced the expression of impulsivity and anger, which were notably different between ADHD patients and controls [15]. This potentially implies that the DAT1 gene may impact not only the most common traits found in ADHD but also the emotional and behavioral aspects of the disorder [15].

Previous research suggests that the DAT1 gene does influence the regulation of dopamine, a neurotransmitter that is critical in the role of attention and emotional regulation [5]. Specific alleles of the DAT1 gene, such as the 9.5R allele, were associated with lower impulsivity and anger scores, supporting the idea that certain genetic polymorphisms could support some of the emotional dysregulation associated with ADHD [15]. This information helps to understand why it is important to view ADHD as a multifaceted disorder influenced by both genetic factors and emotional/behavioral factors.

The Neuropsychological Effects Of The DAT1 Polymorphism

Bellgrove et al. looked into additional information that could support the hypothesis that the DAT1 gene influences specific cognitive functions, particularly inattention and impulsivity [17]. It was found that individuals with certain DAT1 genotypes, specifically those with the 10R allele, exhibited higher trait scores of inattention [17]. These results support the idea that the DAT1 gene potentially influences cognitive processes such as attention, specifically in terms of spatial working memory (SWM) [17]. In this study, left-sided inattention was linked to the 10R allele, which further reinforces the idea that neuropsychological and cognitive performance is connected to specific DAT1 genetic polymorphisms [17].

There was also information supporting the possibility of a relationship between the DAT1 polymorphism and patient response to methylphenidate (MPH), which is a common ADHD medication [17]. The improved response to MPH in individuals with left-sided inattention, particularly those who possessed the 10R allele, suggests that genetic variations can influence not

only the development of ADHD but also the effectiveness of various treatments for the disorder [17]. This result emphasizes the importance of accounting for genetic factors when considering treatment plans for individuals with ADHD.

Implications For Preventive Programs

Marzilli et. al explores the idea that emotional and behavioral functioning in children with ADHD is not only determined by observable symptoms but can also be determined by underlying genetic factors such as DNA methylation [18]. This is important to consider when designing preventive programs for ADHD. With these programs, the rate of early interventions will increase and ADHD treatment can be approached in a more preventive way. To implement these programs, however, it's crucial to be able to understand the role of the DAT1 gene in regulating the traits and symptoms associated with ADHD.

The relationship between various DAT1 genotypes and patient response to MPH suggests that preventive programs could potentially benefit from using genetic information to predict the outcome of the treatment [17]. With the help of these programs, healthcare professionals can identify children at risk of developing ADHD or those who have a decreased response to certain ADHD treatments, allowing for early intervention or the development of alternative treatment approaches that might work better.

Relevance And Future Directions

Jeong et. al explores the emotional effects of ADHD which is relevant because of its potential to contribute to a deeper and more nuanced understanding of ADHD and its underlying genetic factors [13]. By focusing on the DAT1 polymorphism and its impact on emotional and behavioral traits found in children and adults with ADHD, this research highlights the complex nature of ADHD and acknowledges that a multifaceted approach to treatment and prevention is necessary in the future [13]. Typically, ADHD is often associated with a variety of other conditions, such as mood disorders [19]. To incorporate the genetic foundation of the emotional dysregulation that is associated with ADHD, there could be numerous benefits when it comes to treating the other conditions as well.

Future research on this topic should try to replicate these findings in larger numbers to confirm the presented data and expand on the DAT1

polymorphism and emotional/behavioral traits. Longitudinal studies could also be conducted to help clarify the role of the various DAT1 genetic variations in the development of ADHD symptoms. Furthermore, future research should also look into the relationship between genetic factors and environmental influences, such as childhood stress or family dynamics, specifically maternal and paternal influences.

Lastly, more research on the pharmacogenetic aspects of ADHD treatment is necessary. Bellgrove et al. emphasize that genetic factors like the DAT1 polymorphism can significantly influence patient response to MPH [17]. If these relationships are better understood, more effective and personalized treatments for individuals with ADHD can be developed.

Conclusion

This review covers the role of the DAT1 gene in ADHD pathophysiology, specifically its role in dopamine dysregulation, emotional and behavioral traits, and patient response to medical treatments. Variations in the DAT1 gene, like the 10-repeat allele, are linked to symptoms of ADHD such as inattention and impulsivity. These results support the importance of more personalized treatments for ADHD to increase the rates of early diagnosis.

Additionally, this information emphasizes the relationship between genes and their environment, and how it all affects ADHD symptomatology. Future research on this topic should explore those gene-environment relationships and look more into potential treatments for ADHD. Potentially using patients' genetic profiles could lead to more personalized approaches to treatment.

All in all, utilizing genetic information about ADHD can only improve all aspects of ADHD treatment and diagnosis. By continuing to research more on the role of the DAT1 gene in ADHD, every facet of this nuanced disorder can be better understood for future medical research.

References

1. Magnus W, Anilkumar AC, Shaban K. Attention Deficit Hyperactivity Disorder. [Updated 2023 Aug 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-.

2. Volkow ND, Wang G, Kollins SH, et al. Evaluating Dopamine Reward Pathway in ADHD: Clinical Implications. JAMA. 2009;302(10):1084–1091. doi:10.1001/jama.2009.1308
3. Sonne J, Goyal A, Lopez-Ojeda W. Dopamine. [Updated 2023 Jul 3]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-.
4. Olasore HS, Osuntoki AA, Magbagbeola OA, Awesu AB, Olashore AA. Association of Dopamine Transporter Gene (DAT1) 40 bp 3′ UTR VNTR Polymorphism (rs28363170) and Cannabis Use Disorder. Subst Abuse. 2023 Mar 31;17:11782218231163696. doi: 10.1177/11782218231163696. PMID: 37020726; PMCID: PMC10068503.
5. Levy F. What do Dopamine Transporter and Catechol-O-Methyltransferase Tell us About Attention Deficit–Hyperactivity Disorder? Pharmacogenomic Implications. Australian & New Zealand Journal of Psychiatry. 2007;41(1):10-16. doi:10.1080/00048670601050432
6. Olasore HS, Osuntoki AA, Magbagbeola OA, Awesu A-RB, Olashore AA. Association of Dopamine Transporter Gene (DAT1) 40 bp 3′ UTR VNTR Polymorphism (rs28363170) and Cannabis Use Disorder. Substance Abuse: Research and Treatment. 2023;17. doi:10.1177/11782218231163696
7. Šerý, O., Paclt, I., Drtílková, I. et al. A 40-bp VNTR polymorphism in the 3′-untranslated region of DAT1/SLC6A3 is associated with ADHD but not with alcoholism. Behav Brain Funct 11, 21 (2015). https://doi.org/10.1186/s12993-015-0066-8
8. Verghese C, Patel P, Abdijadid S. Methylphenidate. [Updated 2024 Oct 29]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2025 Jan-.
9. Thapar A, Stergiakouli E. An Overview on the Genetics of ADHD. Xin Li Xue Bao. 2008 Aug;40(10):1088-1098. doi: 10.3724/SP.J.1041.2008.01088. PMID: 20396407; PMCID: PMC2854824.
10. Aguiar A, Eubig PA, Schantz SL. Attention deficit/hyperactivity disorder: a focused overview for children's environmental health researchers. Environ Health Perspect. 2010 Dec;118(12):1646-53. doi: 10.1289/ehp.1002326. Epub 2010 Sep 8. PMID: 20829148; PMCID: PMC3002183.

11. Albrecht, B., Brandeis, D., von Sandersleben, H.U., Valko, L., Heinrich, H., Xu, X., Drechsler, R., Heise, A., Kuntsi, J., Müller, U.C., Asherson, P., Steinhausen, H.-C., Rothenberger, A. and Banaschewski, T. (2014), Genetics of preparation and response control in ADHD: the role of DRD4 and DAT1. J Child Psychol Psychiatr, 55: 914-923. https://doi.org/10.1111/jcpp.12212
12. Shang CY, Gau SS. Association between the DAT1 gene and spatial working memory in attention deficit hyperactivity disorder. International Journal of Neuropsychopharmacology. 2014 Jan 1;17(1):9-21.
13. Jeong SH, Choi KS, Lee KY, Kim EJ, Kim YS, Joo EJ. Association between the dopamine transporter gene (DAT1) and attention deficit hyperactivity disorder-related traits in healthy adults. Psychiatric genetics. 2015 Jun 1;25(3):119-26.
14. Kouros I, Hörberg N, Ekselius L, Ramklint M. Wender Utah Rating Scale-25 (WURS-25): psychometric properties and diagnostic accuracy of the Swedish translation. Ups J Med Sci. 2018 Dec;123(4):230-236. doi: 10.1080/03009734.2018.1515797. Epub 2018 Oct 29. PMID: 30373435; PMCID: PMC6327570.
15. Hasler, R., Salzmann, A., Bolzan, T. et al. DAT1 and DRD4 genes involved in key dimensions of adult ADHD. Neurol Sci 36, 861–869 (2015). https://doi.org/10.1007/s10072-014-2051-7
16. Loo SK, Specter E, Smolen A, Hopfer C, Teale PD, Reite ML. Functional Effects of the DAT1 Polymorphism on EEG Measures in ADHD. Journal of the American Academy of Child & Adolescent Psychiatry. 2003 Aug;42(8):986–93.
17. Bellgrove, M., Hawi, Z., Kirley, A. et al. Association between Dopamine Transporter (DAT1) Genotype, Left-Sided Inattention, and an Enhanced Response to Methylphenidate in Attention-Deficit Hyperactivity Disorder. Neuropsychopharmacol 30, 2290–2297 (2005).
18. Marzilli, E., Cerniglia, L., Tambelli, R. et al. Children's ADHD and Dysregulation Problems, DAT1 Genotype and Methylation, and their Interplay with Family Environment. Child Youth Care Forum 52, 371–399 (2023).
19. Goodman, D. W., & Thase, M. E. (2009). Recognizing ADHD in Adults with Comorbid Mood Disorders: Implications for Identification and Management. Postgraduate Medicine, 121(5), 31–41. https://doi.org/10.3810/pgm.2009.09.2049

Diana Wang

Using Bioinformatics to Find Differentially Expressed Genes In Patients Using SGLT2i

ABSTRACT

Background: Heart disease or cardiovascular disease.is one of the most prevalent diseases in the United States. Almost half of all Americans have heart disease. A lot of lifestyle choices can affect someone's risks for getting cardiovascular disease including family genetics. Recently, several genes that are linked to heart disease have been identified including new genes that increase the risk of heart disease.

Purpose: The problem our research is addressing is the lack of a cure for heart disease. Recent studies show that SGLT2i inhibitors, a medicine used to treat patients with diabetes, reduced people's risk for getting heart disease. Given the role of genes in inherited heart disease, the goal of this research is to identify additional genes, gene functions and biological pathways that can be potentially used to create new treatment solutions for heart disease such as gene therapy.

Methods: In our study, we started off by finding a dataset in GEO2R, a website that allows you to analyze data from various datasets. After separating the data into two groups; SGLT2i in use and SGLT2i not in use, we could further analyze the dataset. From there, we inserted the list of genes into a sheet and found the top 30 DEGs by sorting the list of genes into the genes with the highest log2FC value and lowest log2FC value.

The bioinformatics tool we chose to use to conduct analysis of our DEGs was SRplot. Then we inserted the gene symbols and log2FC values that corresponded to the symbols into SRplot.

After that, we decided to return back to the GEO2R dataset and included more patients into the two groups. By doing that, we could see which genes were important and compare the results. We found that there are three genes that are of importance in this study. Those three genes are also part of the top 30 DEGs..

Results: We chose to identify 30 DEGs. In total, 79 of the 17,925 genes were differentially expressed. There were 24 genes that were downregulated and 51 genes that were upregulated.

We found that histone lysine demethylation, histone demethylation, protein demethylation, protein dealkylation, histone demethylase activity, protein demethylase activity and demethylase activity have high enrichment scores. We went back a second time to the GEO2R dataset and this time, included more patients into the two groups; stgl2i in use and stgl3i not in use so the results could be more accurate and it narrowed down the DEGs from 79 to 3. From there we could see that only 3 genes are differentially expressed. SCD, TSIX, and XIST which shows their significance. Then we went back to the google sheet, where we found that the SCD, TSIX, and XIST genes were all part of the top 30 DEGs. The three genes were all downregulated.
Discussion: This study shows that differentially expressed genes were identified between samples who received SGLT2i treatment and those patients who did not. The genes of significance were SCD, XIST and TSIX. The SCD gene, also known as Stearoyl-CoA Desaturase, helps produce fats that benefit the function of cell structures. XIST and TSIX genes work together to regulate XCI, an X chromosome in women. My study provides a new direction to the development of a medicine or treatment because researchers can look into the genes of significance and SGLT2i's impact on patients with Cardiovascular disease and try to create a treatment.

INTRODUCTION

Right now, heart disease is the leading cause of death (12). Smoking, drinking alcohol, and diets high in sugar are risk factors for heart disease. Poor diets can lead to heart disease.(9) Heart disease is a disease that affects the heart and its major blood vessels. Coronary artery disease is one of the most prevalent and common types of heart disease (7) . This happens when there is a lot of buildup in central blood vessels. This buildup prevents oxygen in the bloodstream from reaching the heart. When the blood vessel becomes nearly occupied with the buildup, the person is at risk for a heart attack or stroke. (20) Blockage in a vessel prevents oxygenated blood from reaching the heart.(1) . Although we still do not know the exact cause of heart disease (20) we can identify some risk factors which increase the risk of having heart disease (20) . Poor diets, lack of exercise, family genetics, and alcohol can affect someone's risk for heart disease. How heart disease affects women has not been studied in depth yet. (22) Research has shown that hormones can affect a woman's risk for heart disease.(14) Since heart disease is different in women, than it is in men, it can lead to misdiagnosis and plenty of other issues. (14)

Our study focuses on family genetics as the cause of heart disease In regard to family. Recently, an international research team, including the CARDioGRAM and the Coronary Artery Disease Genetics Consortia, confirmed ten genetic markers already linked to heart disease (11). In addition, they discovered 13 new ones that increase the risk of heart disease (11). Genetics play a complex role in heart disease. (3) Research shows that the risk of having heart disease increases in people with more than one sibling with cardiovascular disease. (3)

Therefore, understanding how these and additional genes function is important in helping figure out how the disease develops and could potentially lead to finding new medications or treatments including gene therapy.

SGLT2i is a term used to describe a class of different medications used to treat type 2 diabetes (18). Also known as Sodium Glucose Co-Transporter 2, these treatments are used to lower blood sugar levels, and recent studies show that SGLT2 inhibitors, which treat patients with type 2 diabetes, can actually help with cardiovascular disease (6) Common SGLT2 inhibitors include canagliflozin and dapagliflozin (18).

Gene therapy is the modification of a gene to form a cure. (2) Because of the role that genes play in family heart disease and the potential for gene therapy as a treatment for heart disease, the goal of this research is to find and compare genes in the control group to genes in the group using the medicine to see the similarities and differences. We hypothesize that there will be differences in expression of genes between the control group and patients.

Our research is important because as of now, there is still no cure for heart disease.(20) Yet heart disease continues to be the #1 cause of death.(12) By looking at how Stgl2i, a group of treatments designed to reduce glucose levels in the bloodstream for patients with diabetes, affects someone's risk of heart disease, we can learn a lot about heart disease prevention. (6) Identifying the differences in gene expression between patients who don't take the treatment compared to those who take the medicine can teach us a lot about ways to reduce the risk of having cardiovascular disease.

Results from this research with gene expression may help guide other researchers and scientists on the search to find a cure for heart disease or a treatment to reduce the risk of having heart disease.

METHODS

This study used different bioinformatics tools and databases, namely, GEO2R (8) and SRPlot (1) as indicated in **Figure 1.** The dataset we used was GSE26364.

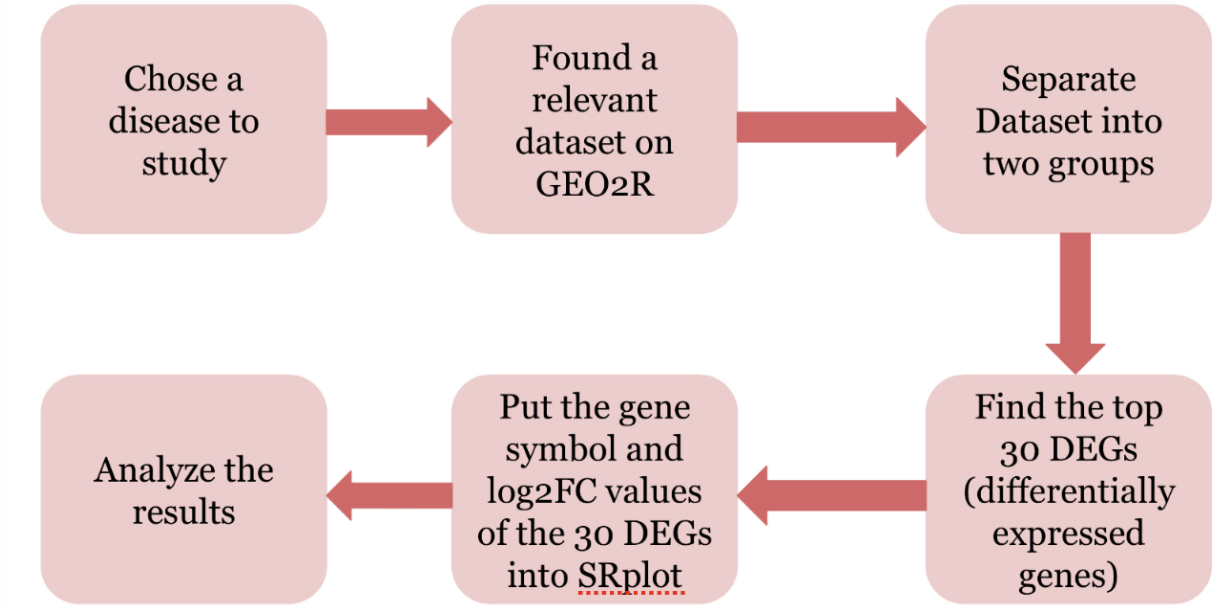


Figure 1. Flowchart showing methods and bioinformatics used in this study.

GEO and GEO2R Bioinformatics Analysis

We started off by using GEO, Gene Expression Omnibus (8), a bioinformatics tool that allows you to view other research experiments. We found an experiment similar to the topic we were studying, cardiovascular disease. From there we were able to go to GEO2R(8), and view the different patients. Then we proceeded to group the information into two groups; patients that did not use SGLT2i and patients that did use SGLT2i. After sorting the data into the two groups, we were able to analyze that data and view the different graphs. In this study, we tested to see if patients using SGLT2i had different expressed genes than patients not using SGLT2i. We tested 30 DEGs in total. 15 of the DEGs consisted of the highest log2FC value, while the rest were the lowest log2FC values. By going back to GEO2R (8) a

second time to reanalyze the control and experimental group, we were able to narrow down the DEGs to find the genes that were the most significant in this study.

SRPlot Analysis of top DEGs

After finding my 30 DEGs, we were able to go to SRPlot (1), a website that allows you to enter the gene symbols of all the DEGs, and creates charts and graphs based off the symbols.

GO Analysis

Gene Ontology (10) is a bioinformatics website to assist in describing the function of genes and the gene functions. This tool allows us to describe genes in detail. The three main categories of the bioinformatics tool are; BP (biological processes), CC (cellular component) and MF (molecular function).

KEGG Analysis

KEGG (15), a group of databases that deal with genomes, biological pathways, diseases, and chemical substances. This tool is used for bioinformatics researchers to understand the biological functions of cells.

Top 30 DEGs – Top 30 DEGs

RESULTS

Analyzing Differential Gene Expression using GEO2R

The GEO2R (8) tool that generated our results used R programming through AL and ML algorithms. This R script explains the process and is lined here – R Script

We first went to GEO (8) and found a dataset that was relevant to this topic. We chose GSE26364. From there we were able to view the dataset and separate it into two groups; SGLT2i in use and SGLT2i not in use. We decided to use my top 30 DEGs to further analyze. From there we placed the gene symbols of the top 30 DEGs into SRPlot (1).

Results from GEO2R

In **Figure 2,** we identified 79 DEGs total out of 17,925 genes. A majority of the genes were upregulated.

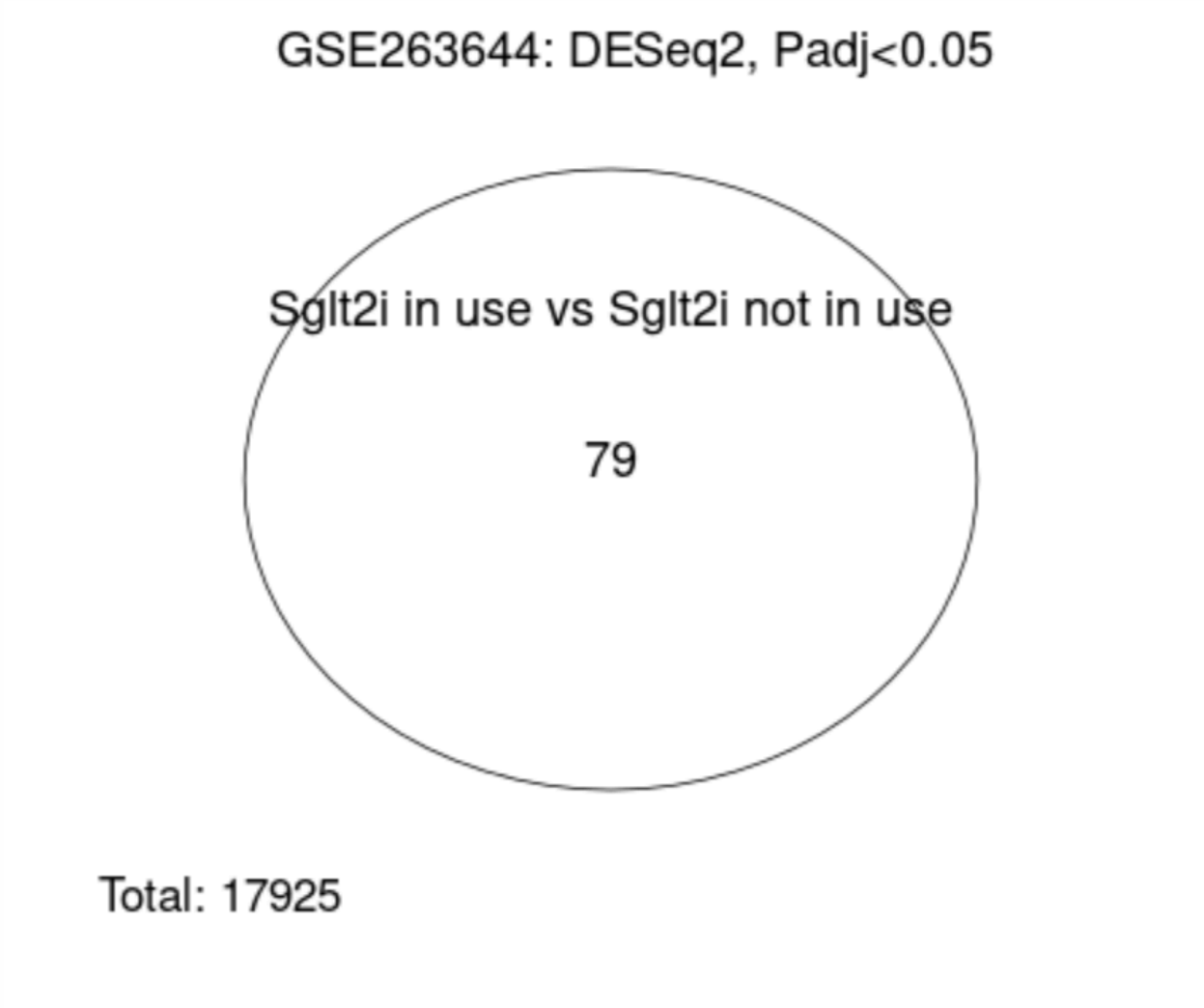


Figure 2A. Venn Diagram In this figure, 79 of the 17,925 genes were differentially expressed.

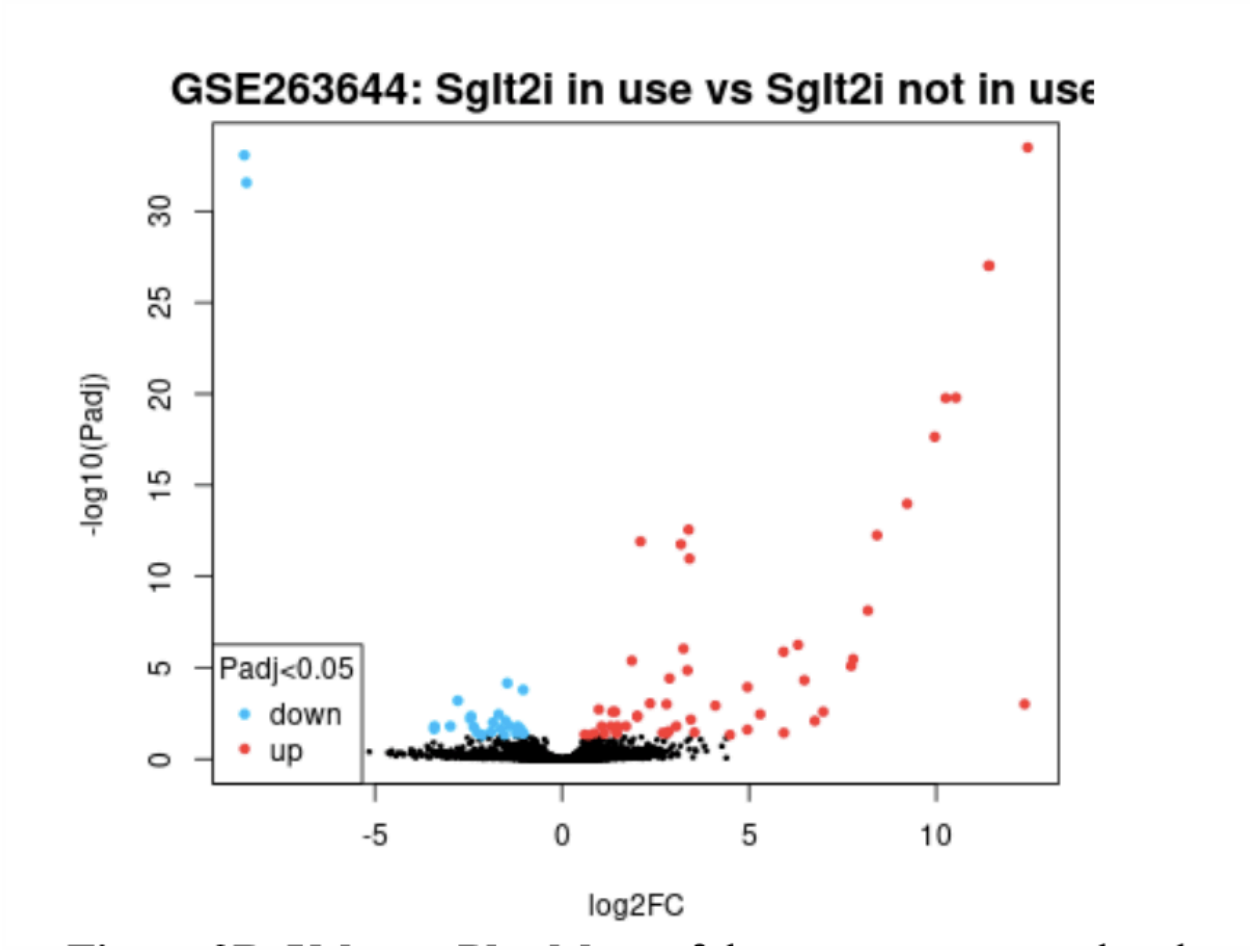


Figure 2B. Volcano Plot More of the genes are upregulated as shown on the graph. The blue dots represent downregulated genes and the black dots represent genes expressed the same in both groups

Results From GEO2R (second time)

Using GEO2R (8), the control and experimental groups were analyzed a second time and results showed, only 3 genes were found to be differentially expressed in **Figure 3.** The SCD, TSIX, and XIST genes were differentially expressed which shows their significance in this study. 3 of the 18121 genes were differentially expressed in **Figure 4**. Going back to the google sheet, we found that the SCD, TSIX, and XIST genes were all part of the top 30 DEGs.

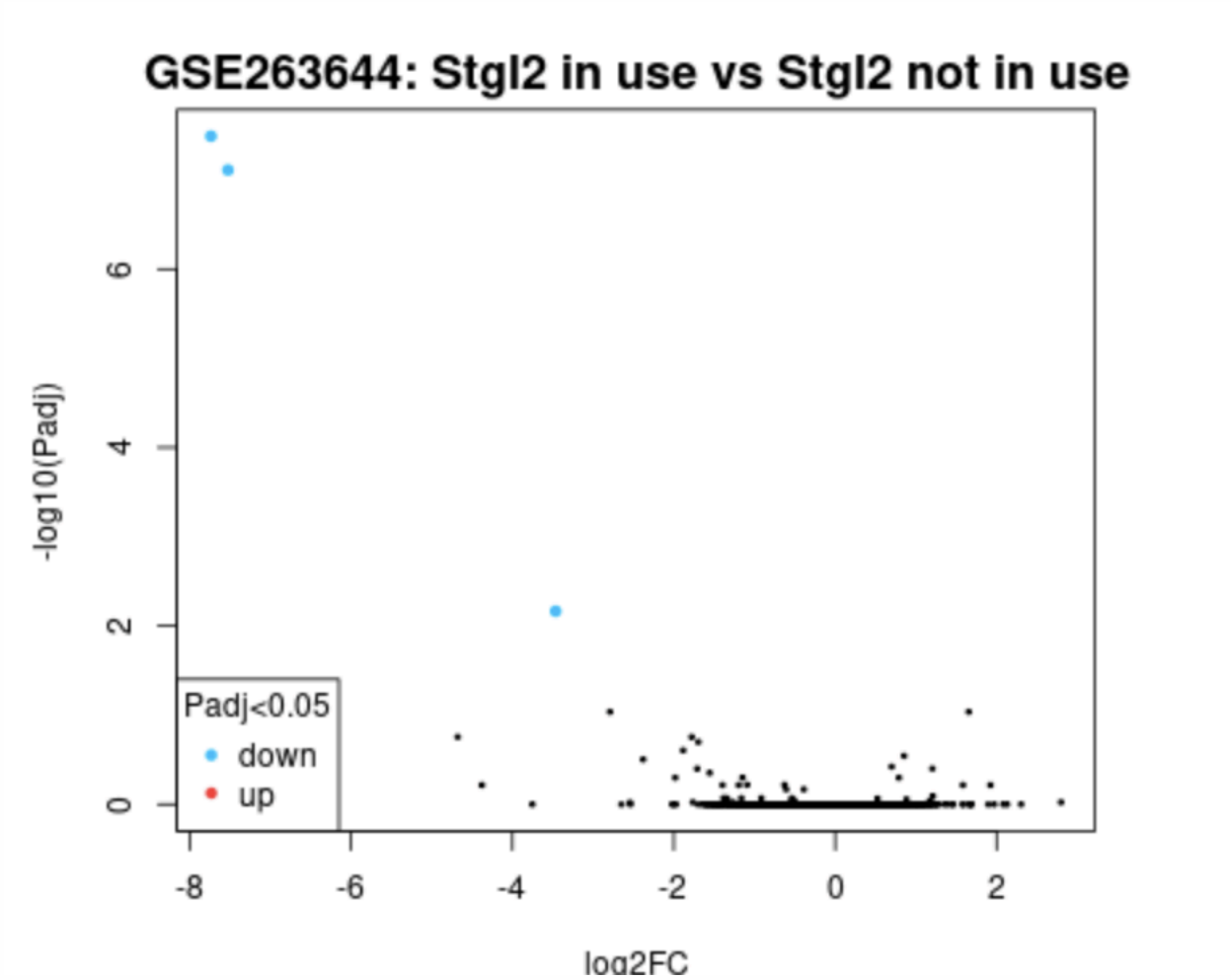


Figure 3. Volcano Plot Only 3 genes are downregulated.

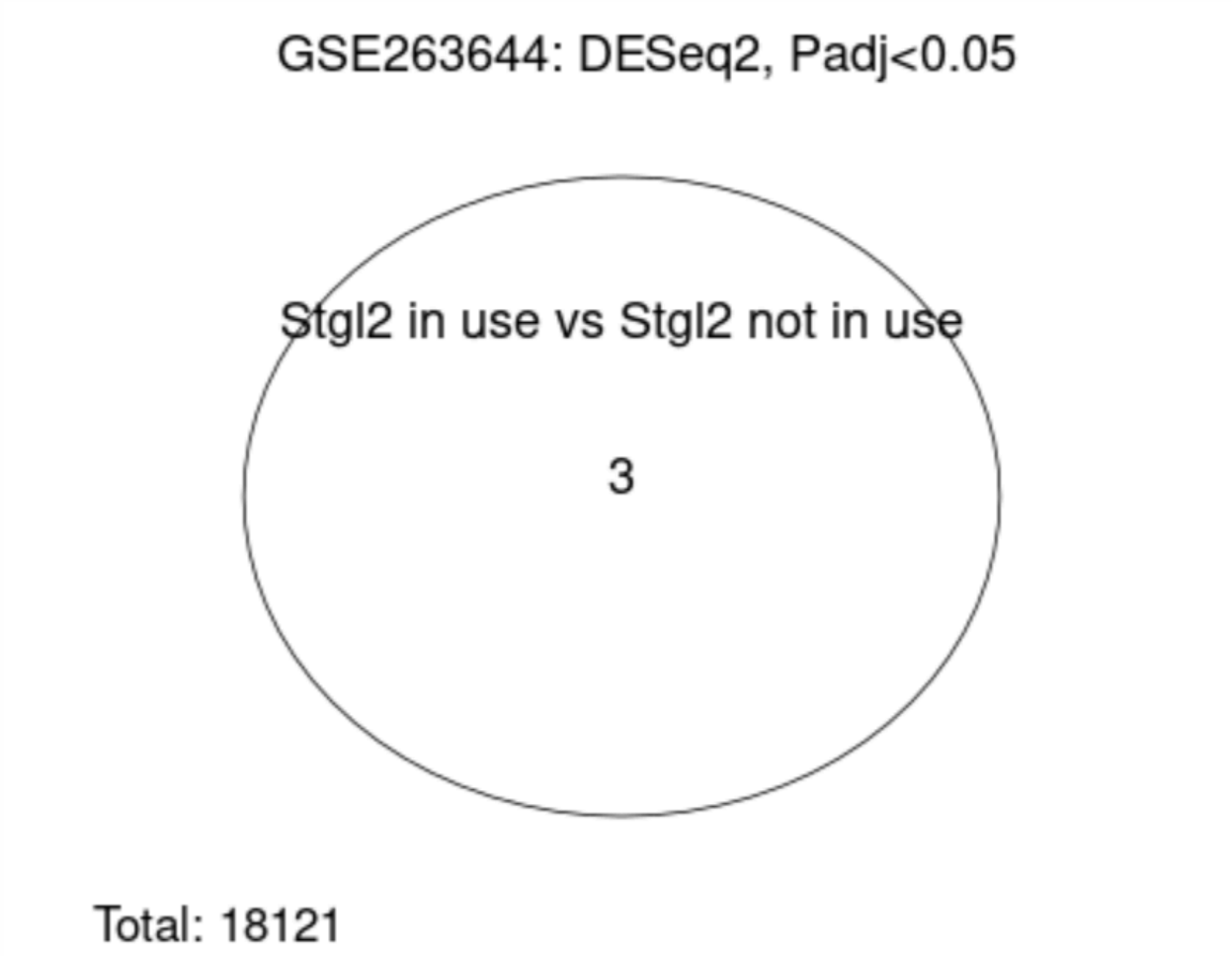


Figure 4. Venn Diagram 3 of the 18,121 genes are differentially expressed.

Results from SRplot

The bioinformatics tool we chose to use to conduct analysis of our DEGs was SRplot (1). We inserted gene symbols and log2FC values of the top 30 DEGs into SRplot. Histone lysine demethylation, histone demethylation protein, demethylation histone, demethylase activity, and protein demethylase activity all had higher enrichment scores based on **Figure 5.** Overall, the enrichment scores were greater than or around 2. The CC section had the lowest enrichment scores compared to the BP and MF in **Figure 5**. In total, we found three significant genes. SCD, XIST and TSIX **(Figure 6).**

Results From SrPlot

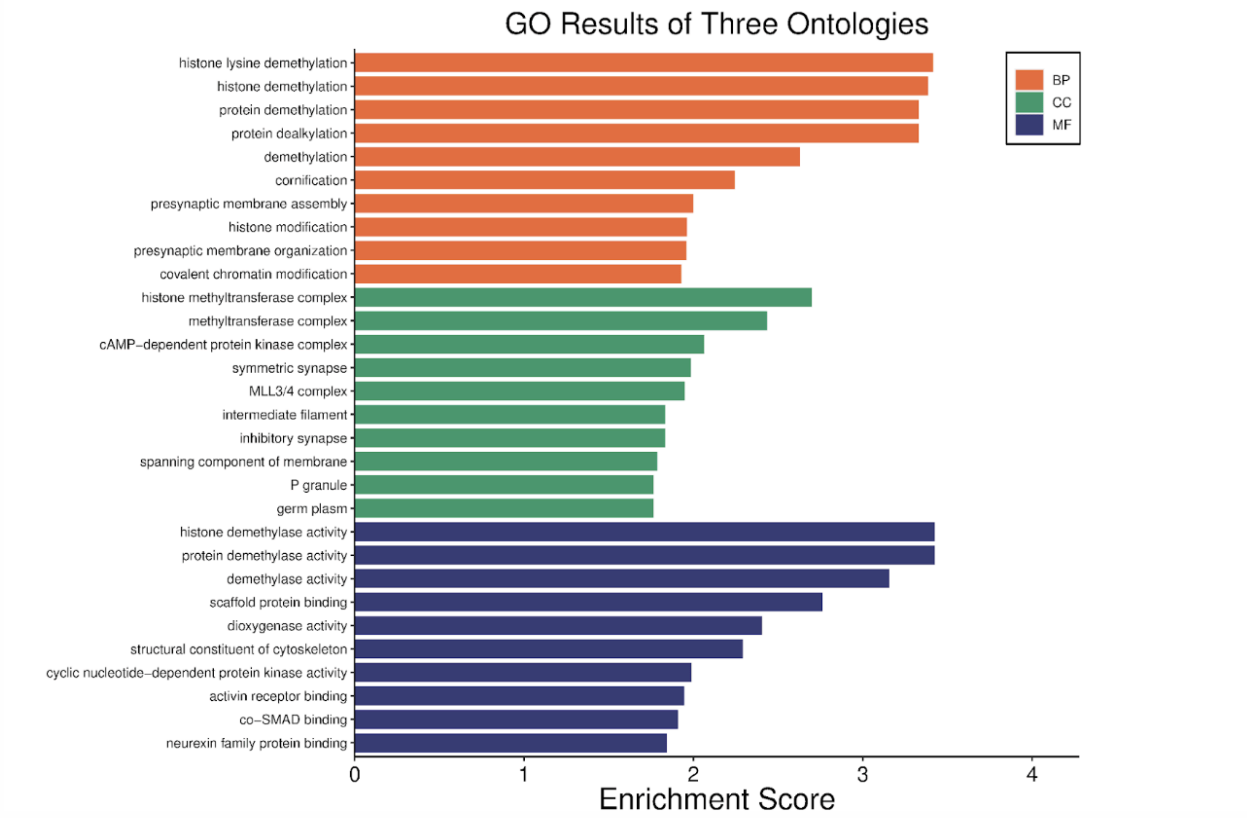


Figure 5. GO Results for BP, CC, MF Histone lysine demethylation, histone demethylation, protein demethylation, protein dealkylation, histone demethylase activity, protein demethylase activity and demethylase activity all have the higher enrichment score.

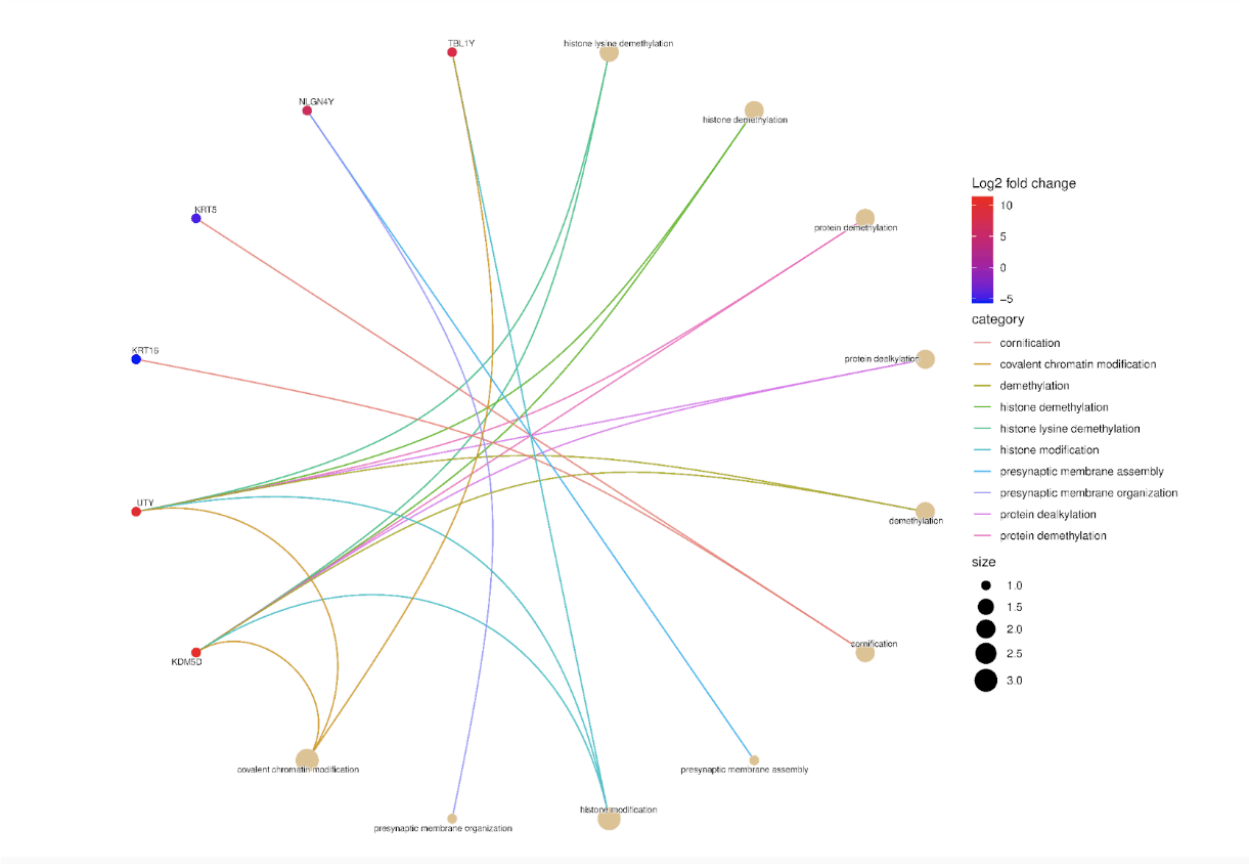


Figure 6. The size of the circles represents how significant they are. The lines connect the dots to show the relationship between genes and their biological function.

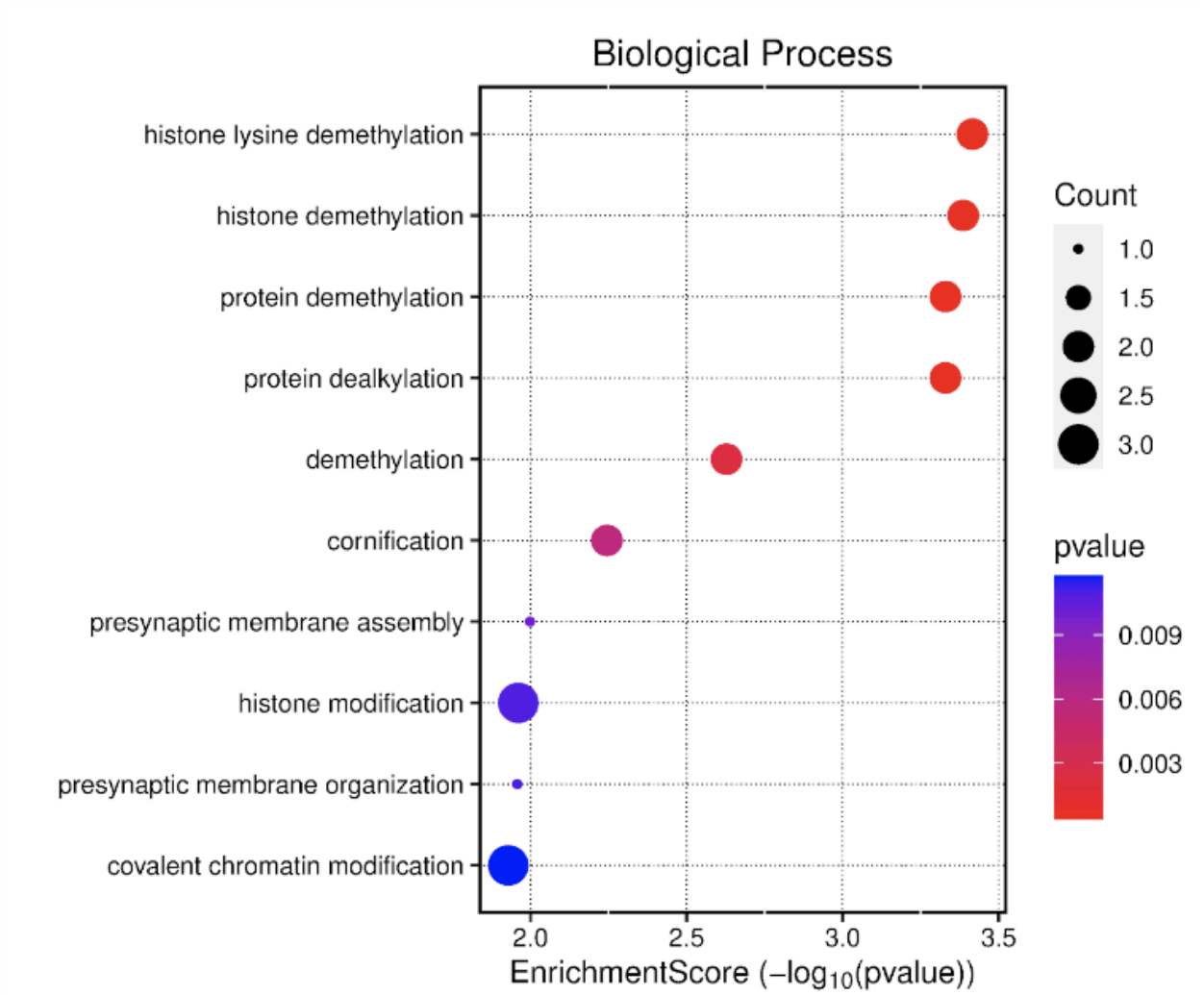


Figure 7. Dot Plot The greater the red of the circles means that the process has a lower p-value. The greater the size of the dots, the greater the significance.

Table 1. Summary of significant genes found in this study and their connection to heart disease

Gene	Function	Name	Connections to Heart Disease
SCD	-Maintains homeostasis. -Keeps balance	stearoyl-CoA desaturase	If SCD activity is reduced, the risk for cardiovascular

	of the fats in a cell.		disease increases and cardiovascular health is greatly affected.
XIST	-Work with TSIX to regulate one X chromosome	X inactive specific transcript	Disruptions with the expression of XIST have been correlated/linked to cardiovascular disease.
TSIX	-Work with XIST to regulate one X chromosome	TSIX X Inactive Specific Transcript	Problems with the regulation of the X chromosome could lead to issues with gene expression in cardiovascular processes.

Supplementary Materials
R Script: R Script
Top DEGs: Top 30 DEGs

DISCUSSION

In this research, our goal was to identify genes, gene functions and biological pathways that can be potentially used to create new treatment solutions for heart disease such as gene therapy. We hypothesized that there would be differences in gene expression between patients using SGLT2i and patients not using SGLT2i.

This study is important because it could be a potential step towards finding a solution to the disease. We learned that although patients using the SGLT2i treatment could have differentially expressed genes **(Figures 2, 3, 4, 5)** , the treatment didn't have a direct impact on the DEGs. We found DEGs between samples who received SGLT2i treatment and those who did not. We chose to look at 30 DEGs. In total, we found three significant genes. SCD, XIST and TSIX **(Figure 6).** Histone lysine demethylation, histone

demethylation, protein demethylation, and protein dealkylation were pathways that had high enrichment scores **(Figure 5).** I found that all the genes of significance have a connection to heart disease.

Our findings in **Figure 5. Figure 7** and **Table 1** were similar to other studies in this field. In one study, STGL2i proved to positively benefit cardiovascular levels by preventing DNA demethylation in promoter regions. (19) In **Figure 5**, we can see that histone demethylation and protein demethylation had higher enrichment scores. In **Figure 7**, we can see that histone lysine demethylation, histone demethylation, protein demethylation are of greater significance and all have low p-values. In our study, we found a number of DEGs between samples who received Stgl2i treatment and others that did not **(Figure 2 and Figure 4**). The results show that while SGLT2 inhibitors have a positive effect on cardiovascular risk, the treatment was not directly linked to the difference in gene expression.

In another study, SGLT2 inhibitors had positive effects on the functions of the heart like decreased oxidative stress, improved energy metabolism, decreased epicardial fat mass, etc. (15) All these benefits eventually led to ‘decreased severity of heart failure. (15) While the diabetes treatment had various benefits on the body and heart, the treatment has no direct links to differences in gene expression.

In another research article, gene therapy could be a step closer to learning about a treatment for cardiovascular disease. (2) Since we identified DEGs between the two groups, other researchers could look into gene therapy as a treatment option for cardiovascular disease. Looking into how SGLT2 inhibitors lowers cardiovascular risk could also help with the search for finding a potential cure in the future.

In this case, the study was limited because all the information gathered was from a dataset on GEO2R (8). This dataset comes from an experiment done by other scientists. Further research will need to be completed on these identified genes in a laboratory before a potential treatment or cure may be found in the future.

By looking at all our results, we can conclude that there will be differences in gene expression between patients using SGLT2i and patients not using

SGLT2i by viewing **Figure 2 and 5.** In **Figure 2**, we can see a few genes that are skewed to the right of the diagram. This skew indicated a difference in gene expression between the two groups. In **Figure 5**, a couple biological processes had higher enrichment scores. Those biological processes were histone lysine demethylation, histone demethylation, protein demethylation, and protein dealkylation, meaning they were significantly represented in the study. All these processes were of significance to our study. By identifying these important processes between the two groups, it indicated that there must have been differences in gene expression between patients using SGLT2i and patients not using SGLT2i.

Conclusion

This study provides new potential directions for the development of a potential cure because this study targets for the research of heart disease, which currently is one of the most prevalent diseases in the United States. By finding genes of significance, other researchers can further look into these genes and search for a correlation between genes and cardiovascular disease.

In this study, DEGs were identified between samples who received SGLT2i treatment and those who did not. Genes like SCD, XIST, and TSIX were significant in our study and XIST was found to have a direct link to cardiovascular disease. The top DEGs were enriched in biological pathways and molecular functions. We were able to identify genes with significant relationships with biological processes. The genes and biological processes can be studied in the lab for potential cure in the future.

References

1. Bioinformatics (n.d.). Retrieved August 6, 2024, from https://www.bioinformatics.com.cn/en/
2. Cannatà, A., Ali, H., Sinagra, G., & Giacca, M. (2020a, May 7). Gene therapy for the Heart Lessons Learned and future perspectives | circulation research. AHA ASA Journals. https://www.ahajournals.org/doi/10.1161/CIRCRESAHA.120.315855

3. Czepluch, F. S., Wollnik, B., & Hasenfuß, G. (2018, June). Genetic determinants of heart failure: Facts and numbers. ESC heart failure. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5933969/
4. Dahlöf, Björn. "Cardiovascular disease risk factors: epidemiology and risk assessment." The American journal of cardiology 105.1 (2010): 3A-9A. www.sciencedirect.com/science/article/abs/pii/S0002914909024825.
5. D'Agostino Sr, Ralph B., et al. "Cardiovascular disease risk assessment: insights from Framingham." Global heart 8.1 (2013): 11-23. www.sciencedirect.com/science/article/pii/S2211816013000057.
6. FayeRiley. "Type 2 Diabetes Medications Can Also Protect Heart Health." Diabetes UK, Diabetes UK, 16 Feb. 2022, www.diabetes.org.uk/about-us/news-and-views/type-2-diabetes-medications-can-also-protect-heart-health.
7. Franczyk B, Frąk W, Lisińska W, Młynarska E, Rysz J, Wojtasińska A. "Pathophysiology of Cardiovascular Diseases: New Insights into Molecular Mechanisms of Atherosclerosis, Arterial Hypertension, and Coronary Artery Disease." Biomedicines. 2022 Aug 10;10(8):1938. doi: 10.3390/biomedicines10081938. PMID: 36009488; PMCID: PMC9405799. www.ncbi.nlm.nih.gov/pmc/articles/PMC9405799.
8. "GEO2R: Analyze GEO Datasets." National Center for Biotechnology Information, National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html. Accessed 1 Aug. 2024.
9. Gatorworksdev. "How Does Diet Affect Your Heart Disease Risk?" Cardiovascular Institute of the South, 18 June 2024, www.cardio.com/blog/is-your-diet-increasing-your-heart-disease-risk.
10. Gene Ontology Consortium. (2000). Gene Ontology: Tool for the unification of biology. Nature Genetics, https://geneontology.org/
11. Hajar R. "Genetics in Cardiovascular Disease." Heart Views. 2020 Jan-Mar;21(1):55-56. doi: 10.4103/HEARTVIEWS.HEARTVIEWS_140_19. Epub 2020 Jan 23. PMID: 32082505; PMCID: PMC7006335.
12. "Heart Disease Facts." Centers for Disease Control and Prevention, Centers for Disease Control and Prevention, 15 May 2024, www.cdc.gov/heart-disease/data-research/facts-stats/index.html.
13. Heart Foundation. (n.d.). Family history and heart disease. Retrieved August 6, 2024, from

https://www.heartfoundation.org.au/your-heart/family-history-and-heart-disease#
14. Jhalani, Nisha. “Heart Disease in Women Is Not like Heart Disease in Men.” ColumbiaDoctors, 28 July 2023, www.columbiadoctors.org/news/heart-disease-women-not-heart-disease-men.
15. Kyoto Encyclopedia of Genes and Genomes (1995) KEGG Pathway Database www.genome.jp/kegg.
16. Lopaschuk, G. D., & Verma, S. (2020, June 22). Mechanisms of cardiovascular benefits of sodium glucose co-transporter 2 (SGLT2) inhibitors: A state-of-the-art review. JACC. Basic to translational science. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7315190/
17. National Center for Biotechnology Information (NCBI). Programs & Activities. National Library of Medicine, 2023. www.ncbi.nlm.nih.gov/home/about/programs/.
18. National Kidney Foundation. (n.d.). SGLT2 inhibitors. Retrieved August 1, 2024, from https://www.kidney.org/atoz/content/sglt2-inhibitors
19. NCBI GEO. (n.d.). GEO2R: Analyze GEO Datasets. National Center for Biotechnology Information. www.ncbi.nlm.nih.gov/geo/info/geo2r.html.
20. NHS. (n.d.). Cardiovascular disease. National Health Service. Retrieved August 1, 2024, from https://www.nhs.uk/conditions/cardiovascular-disease/
21. Scisciola, L. (n.d.). Targeting high glucose-induced epigenetic modifications at cardiac level: The role of SGLT2 and SGLT2 inhibitors. Cardiovascular diabetology. https://pubmed.ncbi.nlm.nih.gov/36732760/
22. “Women and Heart Disease.” National Heart Lung and Blood Institute, U.S. Department of Health and Human Services, www.nhlbi.nih.gov/health/coronary-heart-disease/women. Accessed 25 July 2024.
23. Verhaar, M. C., E. Stroes, and T. J. Rabelink. “Folates and cardiovascular disease.” Arteriosclerosis, thrombosis, and vascular biology 22.1 (2002): 6-13. www.ahajournals.org/doi/full/10.1161/hq0102.102190.