Validating the effects of deregulation of endoplasmic reticulum (ER) stress on unfolded protein response in deer mice.

Enoch Adelekan

Department of Mathematics and Computer Science

HNTH 391: Honors Thesis Seminar I

Dr. Shrikant Pawar

November 23, 2022

Enoch Adelekan
HNTH 391: Honors Thesis Seminar I
Dr. Shrikant Pawar
November 23, 2022

Validating the effects of deregulation of endoplasmic reticulum (ER) stress on unfolded protein response in deer mice.

Thesis Statement

1. Introduction
2. Method
3. Results
4. Conclusion
5. Discussion
6. References

Introduction

Co-expression networks help researchers to draw connections between genes of unknown functions and biological processes to infer the function of these genes and finding disease-related genes. With advances such as RNA-sequencing in transcriptomics, this information can also be collected for non-coding genes and splice variants, which are unincluded in co-expression networks built from microarray data. Data from RNA-sequencing or microarrays is commonly used in building co-expression networks. RNA sequencing data offers the benefits of higher accuracy and specificity as it allows a higher resolution for identifying tissue-specific expression and distinguishes the expression profiles of closely related paralogues - homologous genes locates at different parts of the genome that arise from gene duplications.

While co-expression networks provide relevant information in understanding the biological process of an organism, it requires a lot of resources in terms of computing power and storage. The size of the network grows at a quadratic rate (n^2) for every addition to the network. To address this, the network can be built in blocks from subsets of the entire data, and then combined at a later stage in the analysis. However, there is no clarity on how much this affects the results of the analysis when the number of blocks scale.

The aim of this project is to validate the effects of the deregulation of endoplasmic reticulum (ER) stress on unfolded protein response in deer mice. Aging can be characterized by depressed metabolism, compromised proteostasis, and increased inflammation. However, the causes of these symptoms are not well understood. While differential expression is a widely used technique to explore and understand the molecular foundations of neural aging, it often obscures

larger expression patterns. In contrast, co-expression analysis allows for identification of networks of genes and their regulators.

This project is in collaboration between researchers at Claflin University and University of South Carolina. As a contributor to this project, I am responsible for building a platform that enables investigators of varying academic backgrounds to visualize and carry out comparative analysis on transcriptomic data generated from the laboratory procedures carried out at University of South Carolina.

Method

This project, in collaboration with the University of South Carolina, is an interdisciplinary research endeavor and it is sectioned in two parts – data generation and data analysis. The data generation process was facilitated by University of South Carolina's Peromyscus Genetic Stock Center and Claflin University's Biology Department. The data analysis, on the other hand, is to be carried out exclusively by the Claflin University's Mathematics and Computer Science department.

In the data generation stage, researchers at the University of South Carolina were focused on obtaining RNA sequencing data from various outbred deer mice tissues. Male and female deer mice groups of six were the species of focus; they were studied from birth up till they reached 3 years of age. In six-month intervals since the start of the study, samples were collected from each mouse across the following tissues: brain, liver, lungs, intestine, and kidney. Over the course of the study, 360 individual specimens were collected - 6 time points * 6 mice per group * 2 groups

* 5 tissues. The research team from Claflin University's biology department was responsible for tissue isolation and RNA extraction from the generated samples.

The data analysis phase is yet to commence, however the outline for the flow of the project is as follows: clarification, design, implementation, testing, and deployment. The aim of the data analysis phase is to analyze the generated data to gain insight and make conclusive judgements based on this. More importantly, the analysis phase is also expected to give rise to an accessible platform for researchers that simplifies visualization and comparative analysis on genomic and transcriptomic data. The clarification segment of this phase is focused on identifying the details on how the platform should function i.e., the type of data that the platform should analyze, the scope of the data being analyzed, the types of analysis expected of the platform, the speed of analysis required, and many more.

This is followed by the design phase where the architecture of the platform is designed to meet the requirements specified from the clarification phase. After this, the implementation phase is the actual development of the platform with code. For this project, the primary programming languages to be used are JavaScript and Python. After implementation, the platform will be tested to ensure that all its modules are working as expected. This phase would help in discovering bugs so they can be fixed before the platform is deployed for public use. The final segment is deployment where access to the platform is now made public.

Other research has been conducted utilizing a similar procedure – co-expression analysis – in answering the main question of their projects:

a) A research project titled, Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models, was focused on developing a consensus atlas of the human brain transcriptome in Alzheimer's disease, based on 2114 postmortem samples from previous studies. In this research, further investigation was conducted to examine the overlap with 251 expression signatures from mouse brain RNA sequencing experiments, including many neurodegenerative disease models. Towards accomplishing this, first, random and fixed-effects meta-analyses of differential expression across the brain regions were performed to define the unregulated and downregulated transcripts in AD (Alzheimer's Disease) brains.

   Next, co-expression analyses were performed using five different algorithms to generate brain region-specific co-expression modules. The results showed that similar co-expression patterns were observed in each brain region irrespective of the analysis method used. This was good practice to validate the co-expression modules in the brain and eliminate cases that may have come up by chance.

   Finally, graph clustering was applied to define brain region-specific consensus modules based in patterns of shared module gene membership, which resulted in 30 co-expression modules across seven brain regions. Overlaps between mouse expression signature and these 30 modules were evaluated by comparisons.

From the research, it was discovered that in the human brain, 30 co-expression modules were discovered from seven regions as the major sources of Alzheimer's disease transcriptional disturbances. Secondly, human-mouse overlaps highlighted responses to amyloid versus tau pathology and revealed age and sex-dependent expression signatures for disease progression.

b) A research project titled, Resilience, plasticity, and robustness in gene expression during aging in the brain of outbred deer mice, was focused on elucidating the effects of aging on coordination of the whole transcriptome. It was speculated that this could achieved by comparing the correlation of each gene in the transcriptome with every other gene.

   To execute this, the correlation of each gene in the transcriptome with every other gene was calculated and it was discovered that in about 25 % of the genes, coordination was inversed during aging. This suggested that aging was associated with retention of expression profiles for some genes and more abrupt changes in others, while more subtle but widespread changes appear protective i.e., correcting the effects of aging-associated deregulation on gene expression.

c) A research project titled, Co-expression analysis identifies neuro-inflammation as a driver of sensory neuron aging in Aplysia californica, was focused on further exploring previous work on the neural aging in the marine model *Aplysia californica*.

   Utilizing sequencing data from previous research and the WGCNA R package, sensory neuron type specific co-expression networks were constructed using variance stabilized counts from Buccal S Cluster (BSC) and Pleural Ventral Caudal (PVC) sensory neurons.

Next, bi-weighted mid-correlation in WGCNA was used to construct adjacency matrices for each sensory neuron type independently. Using the adjacency matrices, topological overlap matrices (TOM) were constructed for each neuron type, and then combines to make a consensus TOM by taking the minimum of the two modules.

Eigengenes from the initial module were hierarchically clustered to determine their similarities. The modules were merged such that those with a branch height of 0.25 or less were regarded as the same. Following this, the eigengenes for the merged modules were then recalculated, and the consensus eigengenes were correlated to animal chronological age.

In comparing the immune response of *Aplysia californica* to *Crassostrea gigas*, their proteomes were BLASTed against each other and only the top hit was selected. The resultant protein orthologs were then mapped to their respective gene and transcript identifiers.

At the end of the project, hierarchical clustering of the merged modules resulted in 12 consensus co-expression modules.

d) A research project titled, Transcriptome Network Analysis Reveals Aging-Related Mitochondrial and Proteasomal Dysfunction and Immune Activation in Human Thyroid, was aimed at investigating aging-related transcriptomic alterations in the human thyroid gland and characterized the related molecular functions.

Towards accomplishing this, publicly available RNA sequencing data of 322 thyroid tissue samples from the Genotype-Tissue Expression project were analyzed. In addition, the group created 64 data sets from RNA sequencing of normal thyroid tissue samples for validation. To evaluate the associations between aging and transcriptomic changes, a weighted gene co-expression network analysis and pathway enrichment analysis was carried out. The thyroid differentiation score was then used for further analysis to define the correlation between thyroid differentiation and aging.

At the end of the experiment, it was found that the most significant aging-related transcriptomic change in thyroid was the downregulation of genes related to mitochondrial and proteasomal functions. In addition, genes that are associated with immune processed were significantly upregulated with age, and they all overlapped with the upregulated genes in the thyroid glands affected by lymphocytic thyroiditis. Finally, these aging-related changes were not significantly different according to sex, but in terms of the thyroid differentiation, females were more susceptible to aging-related changes.

Results

No results have been generated from the current stage of my research project.

Conclusion

This project is aimed at developing an online public tool that aids researchers to visualize and carry out comparisons on genomic and transcriptomic data to gain insights right from their personal computers, making this process more accessible at it usually takes days on average and requires using a designated mainframe computer.

Discussion

As earlier stated, co-expression networks need to be constructed from the generated data in this project. There are four guiding steps to constructing a co-expression network – correlation, building, clustering, and analysis. The correlation step is where relationships between genes are defined based on measures of correlation. Some common measures of correlation are the Pearson correlation, Spearman Rank correlation, least absolute regression, and the Bayesian approach. In contrast to the Pearson and Spearman correlation measures, the least absolute regression and Bayesian approach are better at identifying causal links between genes. The next step is building the network with the correlation values calculated from the previous step. In this step, a network where each node represents a gene, and each edge represents the presence and strength of the co-expression relationship. After this, clustering of co-expressed genes in the network is carried to form modules. There is a variety of clustering methods, and each of them have an influence on the outcome of the network analysis. A few clustering algorithms include block-wise clustering and hierarchical clustering. Finally, in the analysis step, functional enrichment analysis can be carried out to interpret each module.

When created, co-expression networks are either signed/unsigned networks or weighted/unweighted networks. Signed networks use the absolute values of correlation coefficients. As a result, genes with negatively correlated genes are identified as co-expressed, leading to the clustering of these genes into the same modules, disrupting the network. Signed networks address this by scaling the correlation values between 0 and 1 (<0.5 is negative correlation, >0.5 is positive correlation). Whereas, in weighted networks genes are connected to one another, and these connections have weights between 0 and 1 that indicate the strength of co-

expression between these genes. In un-weighted networks, genes are either connected or not (0 or 1). They can be created from weighted networks by setting a threshold for the correlation weight.

Generally, genes with similar expression patterns in a co-expression network are co-expressed, but this isn't always the case as there are nuances that could arise from alternative splicing – the generation of multiple mRNA transcripts from a single gene. In gene-co-expression networks, expression of transcripts from the same gene is usually aggregated leading to biased co-expression signals. In co-splicing networks, this is resolved by considering the exon-expression-level distributions when calculating gene co-expression correlation. This means that for two genes to be considered as co-expressed, their splice variants must show coordinated expression. Otherwise, they are not even when the overall expression levels are correlated.

In summary, to prevent raising false positives in co-expression networks. Exon-expression-level distribution should be factored into determining the correlation between two genes.

An alternative to exon-expression-level distribution is determining the expression of different isoforms originating from the same gene. If there is an exon X, such that it produces two isoforms A and B. If the reading of isoform A maps exclusively to exon X and isoform B's reading maps to exon X and other exons, then isoform A is considered as not expressed. It is worth noting that this approach was validated by simulations, but no experimental validation has been carried out.

A challenge that could be faced in this project is inefficiency in building the co-expression network due to the quantity of data to be used in the network. To make the tool usable by all researchers, the runtime must be decreased to ensure that the data can be processed on personal computers with relatively low processing power. A solution that could work would be building the network in blocks, however not much is known about using this approach to build large co-expression networks.

References (APA)

- Gene co-expression analysis for functional classification and gene-disease predictions**;** van Dam S, Võsa U, van der Graaf A, Franke L, de Magalhães JP. Gene co-expression analysis for functional classification and gene-disease predictions. Brief Bioinform. 2018 Jul 20;19(4):575-592. doi: 10.1093/bib/bbw139. PMID: 28077403; PMCID: PMC6054162.

- Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models; Wan YW, Al-Ouran R, Mangleburg CG, Perumal TM, Lee TV, Allison K, Swarup V, Funk CC, Gaiteri C, Allen M, Wang M, Neuner SM, Kaczorowski CC, Philip VM, Howell GR, Martini-Stoica H, Zheng H, Mei H, Zhong X, Kim JW, Dawson VL, Dawson TM, Pao PC, Tsai LH, Haure-Mirande JV, Ehrlich ME, Chakrabarty P, Levites Y, Wang X, Dammer EB, Srivastava G, Mukherjee S, Sieberts SK, Omberg L, Dang KD, Eddy JA, Snyder P, Chae Y, Amberkar S, Wei W, Hide W, Preuss C, Ergun A, Ebert PJ, Airey DC, Mostafavi S, Yu L, Klein HU; Accelerating Medicines Partnership-Alzheimer’s Disease Consortium, Carter GW, Collier DA, Golde TE, Levey AI, Bennett DA, Estrada K, Townsend TM, Zhang B, Schadt E, De Jager PL, Price ND, Ertekin-Taner N, Liu Z, Shulman JM, Mangravite LM, Logsdon BA. Meta-Analysis of the Alzheimer's Disease Human Brain Transcriptome and Functional Dissection in Mouse Models. Cell Rep. 2020 Jul 14;32(2):107908. doi: 10.1016/j.celrep.2020.107908. PMID: 32668255; PMCID: PMC7428328.

- Co-expression analysis identifies neuro-inflammation as a driver of sensory neuron aging in Aplysia californica; Kron NS, Fieber LA. Co-expression analysis identifies neuro-inflammation as a driver of sensory neuron aging in Aplysia californica. PLoS One. 2021 Jun 11;16(6):e0252647. doi: 10.1371/journal.pone.0252647. PMID: 34116561; PMCID: PMC8195618.

- Transcriptome Network Analysis Reveals Aging-Related Mitochondrial and Proteasomal Dysfunction and Immune Activation in Human Thyroid; Cho BA, Yoo SK, Song YS, Kim SJ, Lee KE, Shong M, Park YJ, Seo JS. Transcriptome Network Analysis Reveals Aging-Related Mitochondrial and Proteasomal Dysfunction and Immune Activation in Human Thyroid. Thyroid. 2018 May;28(5):656-666. doi: 10.1089/thy.2017.0359. Epub 2018 Apr 23. PMID: 29652618; PMCID: PMC5952342.

- Resilience, plasticity and robustness in gene expression during aging in the brain of outbred deer mice; Soltanmohammadi E, Zhang Y, Chatzistamou I, Kiaris H. Resilience, plasticity and robustness in gene expression during aging in the brain of outbred deer mice. BMC Genomics. 2021 Apr 21;22(1):291. doi: 10.1186/s12864-021-07613-2. PMID: 33882817; PMCID: PMC8061204.