ANALYSIS OF ENDOCRINE INTERACTIONS AND SEX DIFFERENCES VIA TISSUE PAIR GENE EXPRESSION CORRELATIONS

A Research Paper submitted to the Department of Biomedical Engineering
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Biomedical Engineering

Felipe Barraza
Jonathan Blichar

By

Emmanuel Enoch Edu Jr.

May 7th, 2021

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

ADVISOR
Shannon Barker, Department of Biomedical Engineering
Warren Anderson, Center for Public Health Genomics

Analysis of Endocrine Interactions and Sex Differences Via Tissue Pair Gene Expression Correlations

Jonathan J. Blichar[a1], Felipe Barraza[b1], Emmanuel E. Edu Jr.[c1]

[a] Email: jb2ar@virginia.edu
[b] Email: fb7yg@virginia.edu
[c] Email: eee4fj@virginia.edu
[1] University of Virginia, Department of Biomedical Engineering

Abstract

Inter-tissue communication is an essential mechanism that maintains physiological homeostasis. Communication in the endocrine system occurs when cells in an origin tissue secrete proteins that interact with target tissue receptors after traveling via the bloodstream. These interactions are referred to as endocrine signaling. Genes are responsible for tissue communication via the endocrine system, and to what degree each tissue participates is relatively unknown. Our group sought to produce a program that could be utilized to address both knowledge gaps. To analyze endocrine interactions between tissues,we created an R package as a bioinformatics approach that streamlines the process of identifying gene correlations between tissue pairs using gene expression data. The human tissue sample data used in the development of the package were acquired from the Genotype-Tissue Expression (GTEx) database. Tissue-gene datasets were then formatted into appropriate matrices, normalized, and quality controlled in Rstudio. This was a preliminary step in the development of the R package. Our software package is equipped with the ability to accept two sets of tissue-gene data and analyze gene expression correlations between tissue pairs while producing several visualizations depicting the results of the analysis. This fully functional package was then applied to identify novel sex differences in putative endocrine interactions.

Keywords: gene expression, ligand, gene correlation, endocrine signaling

Introduction

Cardiovascular diseases (CVDs) account for 25% of deaths in the United States, and are the leading cause of global mortality (CDC, 2018; Khera and Kathiresan 2017). Obesity, metabolic syndrome, and type II diabetes are associated with CVDs (Virani et al., 2020). Adipose tissues are significantly involved in obesity and metabolic dysregulation (Attie and Scherer, 2019). Abdominal adiposity is linked to cardiometabolic diseases (Emdin, 2017). These diseases were found to have a genetic heritability of around 50%, with obesity having the highest heritability of up to 70% (Khera and Kathiresan, 2017; Herrera and Lindgren, 2010; Karastergiou et al., 2012). These findings, and the high variability of type II diabetes and CVDs in the population, prompted scientists to investigate the relationship between body fat distribution, genetics, and metabolic diseases. One way heritability can affect individuals simply by their sex. Men and women have different fat distribution patterns that impact their risk for metabolic and cardiovascular diseases. Men store adipose tissue around their abdominal region, which leads to an increased risk for CVDs. In comparison, women store adipose tissue in the gluteofemoral region, and were found to have improved systemic metabolism compared to their male counterparts, despite having higher total body fat (Karastergiou et al. 2012). These findings implicate a significant difference in adipose tissue function in men and women. Disease genetic heritability has led researchers to create computational methods that are capable of modeling inter-tissue interactions to better understand the communication pathways that lead to disease states. This

growing collection of work has spawned the field of bioinformatics, a combination of the emergent field of data science, and biological studies.

Bioinformatics is a rapidly growing field and has great potential for research and clinical applications. As the field grows and information derived from the analysis of large cohort biological data becomes highly accurate, this information could form the foundation of medical informatics. Medical informatics is the application of bioinformatics approaches to the fields of clinical medicine and biomedical research (Chang, 2005). Genetics is one of the most rapidly growing fields in medicine and genetic data is essential for the growth of bioinformatics used in clinical settings (Reches et al., 2019). Medical informatics has extensive potential to expand and individualize the treatment options for patients (Chang, 2005). The implementation of bioinformatics methods into clinical settings will supplement phenotype-based diagnosis with information on patient genotype (Reches et al., 2019). This will also allow for the correlation between disease condition phenotype and genotype for more accurate diagnosis and efficacious therapeutics. Bioniformatics approaches based solely on genetics and DNA alignment are not the only applications of these methods in clinical settings. Bioinformatics can be used to analyze protein function and structure, discover signalling pathways between tissues, and search for biomarkers that indicate susceptibility to disease states like cancer (Chang, 2005).

One group of researchers developed a pipeline that utilized gene expressions to model endocrine tissue interactions in mice (Seldin et al., 2018). This was achieved by calculating gene expression correlations between two endocrine tissues. The study used a variety of inbred strains of mice, and five endocrine tissues. Their computational software was able to identify previously known interactions, validating the results of their work, while also discovering novel interactions that were previously undocumented. Due to their results being supported by previous research, the computational method presented by Seldin et al. holds promise for application in human subjects. In their work, they invite future researchers to apply their pipeline to human data to see if novel interactions in humans could be discovered.

Basing our work off that of Seldin et al., we hope to contribute an efficacious tool to the scientific community. We plan to automate the computational method presented in Seldin et al., and compile it into an R package that expedites the correlation analyses for inter-tissue communications. We will be utilizing human expression data, so as to fulfill the proposed future directions of Seldin et al. We develop a publicly available package that we will then apply ourselves to explore the differences between male and female endocrine interactions, since this dynamic has not been extensively studied.

Results

In order to achieve our objectives for this project, we split the work into three specific aims. These would act as a roadmap for our work, with each aim serving a purpose in the consequent one. Because this restricted our work into a sequential order, all three members of our group worked on each aim together, and did not move forward until all members accomplished their respective tasks.

Specific Aim 1 - Process and format human inter-tissue communication data.

Our first aim was to acquire and perform preprocessing on human gene expression data. We utilized the Genotype-Tissue Expression (GTEx) project as our source for such data (GTEx Consortium, 2013). The GTEx project is a public database that contains a multitude of human data readily available for download, with the objective of providing the scientific community the resources to study gene expression and regulation in humans. The data was essential for ensuring that the scripts used to develop our R-package worked properly throughout aim two, but it was also critical for performing a sex based analysis of endocrine interactions in aim three. Following a vignette developed by our advisor in Anderson et al. (Anderson et al., 2020), we were able to download, annotate, normalize, and quality control the gene expression data. The data collected came from 49 unique tissues, all of them sampled from 838 consenting donors. We normalized the expression data using log base 2 for standardization between tissue samples. Quality control was performed by evaluating the expression of the *XIST* gene, a gene exclusively expressed in females, to confirm male and female donor identification.

Specific Aim 2 - Development of an R Package

An R package was developed with the primary objective of analyzing gene expressions between tissue pairs. The processing of human RNA-Seq TPM data conducted earlier laid the foundation for the development of this package. The current edition of this package titled, ‘geneExpCor’ is successfully able to accept two pairs of tissue-gene data, conduct a biweight midcorrelation analysis on both data sets, and return the p-values and correlation coefficients associated with the analysis while producing a series of scatter plots and histograms depicting

gene expression correlations. This package is publicly available on Github and details pertaining to the intricacies of the package are documented in both the package vignette and function manual.

R Package Features and capabilities

The package is equipped with 3 subfunctions, one primary function, and an additional validation function. Each of the three subfunctions performs a particular task, gene and donor matching, variance filtration, and finally ligand filtration followed by the main correlation analysis. The primary function called 'corr_analysis,' wraps together our subfunctions and has five parameters: tissue gene matrix 1, tissue gene matrix 2, variance percentage, p-value variance, and variance plotting. The first two parameters are tissue gene expression matrices that contain data on gene expression values for each of the subject donors. When the function performs the analysis, the tissue matrices are assumed to be formatted with the 'subjects' as the columns and the associated 'genes' as the rows. This is an important note since the 'corr_analysis' function assumes imputed data abides to the format listed above to produce results based on the initial parameters set by the user. Variance percentage is another parameter that allows the user to filter out genes with low variability. Filtering genes by variability allows for the specific targeting of the most variable genes between both tissue matrices resulting in more significant gene correlations between the tissues. A histogram for gene variances between adipose and liver tissue is presented in Figure 1, with the vertical line indicating the cutoff based on the user defined percentile value. Other than filtering by variability, a ligand filtration process is automatically conducted on the first inputted tissue matrix to isolate ligand-encoding genes. Utilizing the ligand encoding genes from the data is an essential factor in analyzing endocrine interactions because ligands have an active role in mediating endocrine signaling between cells (Hoopes, 2014). The next parameter of the main function is p-value variance which essentially filters for the negative log base 10 p-values above the specified threshold. Adjusting the percentage of p-values provides another layer of detail when determining the kind of visualizations associated with the analysis.

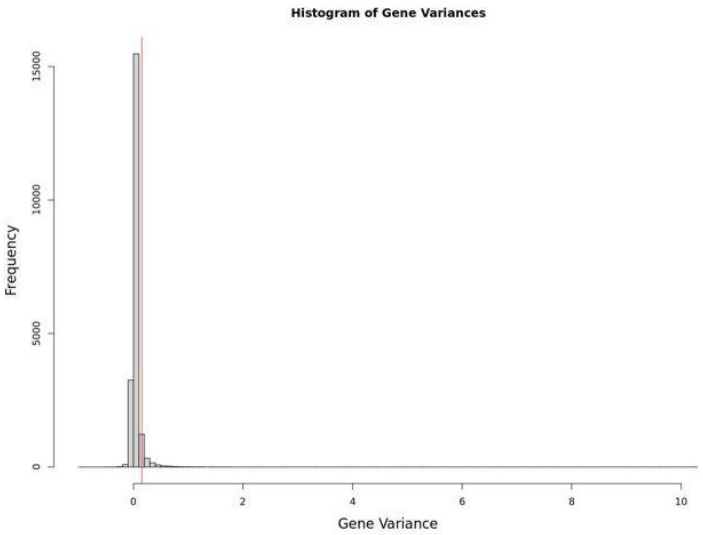

Fig. 1. Histogram of gene variances for matching genes in adipose and liver tissue. The red line is an indicator for cutoff variance based on the input parameter by the user.

Tissue Pair Expression Analysis

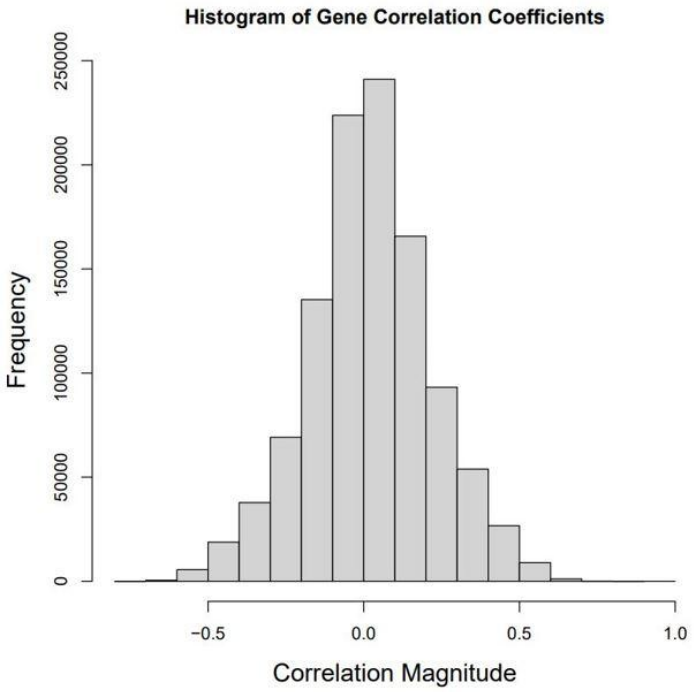

Fig. 2. Histogram of biweight midcorrelation coefficients for adipose to liver analysis.

A biweight midcorrelation method is used within the 'corr_analysis' function to appropriately analyze tissue-gene correlations. Biweight midcorrelation is a median-based method that is resilient to outliers and is therefore the chosen method for identifying gene expression correlations. When the 'corr_analysis' function is run, the results of the biweight midcorrelation method are returned as matrices of p-values and correlation coefficients, which are then utilized by the function to produce several scatter plots and histograms depicting gene-gene correlations between the two tissues. The histogram output plot for our example analysis using adipose to liver tissue analysis is presented in Figure 2.

Identifying Tissue Pair Interactions

We utilized two methodologies for identifying significant gene interactions in our analysis pipeline, one of which

was utilized by Seldin et al. (Seldin et al., 2018), and the second was proposed as a viable alternative. The proposed method is to utilize a p-value cutoff. The second requires the evaluation of Ssec scores, which measure the interaction of origin tissue genes on the entirety of a target tissue. Due to the number of genes being analyzed during the calculation of correlation coefficients, thousands of gene pairs have p-values close to zero. Rather than utilize a p-value cutoff for significance which can be difficult to gauge without utilizing multiple hypothesis correction like false discovery rate or bernoulli methods, our program will isolate p-values of magnitude lower than a percentile cutoff. Negative log base 10 is applied to each of the p-values after filtration for ligand encoders for effective visualization of the p-value distribution. Gene pairs identified based on the highest magnitude negative log base 10 p-values are visualized in a scatter plot of their expression in the origin and target tissues. An example output for analysis using adipose and liver tissue data for the p-values distribution and expression plots are provided in Figure 3 and 4. Figure 3 includes a vertical line indicating the percentile cutoff for gene pairs based on a user defined input parameter. Our package also generates linear regression lines for each isolated gene pair expression scatter.

$$Ssec = \sum(-log(origin\ tissue\ pvalues))/\#\ of\ donors \quad [1]$$

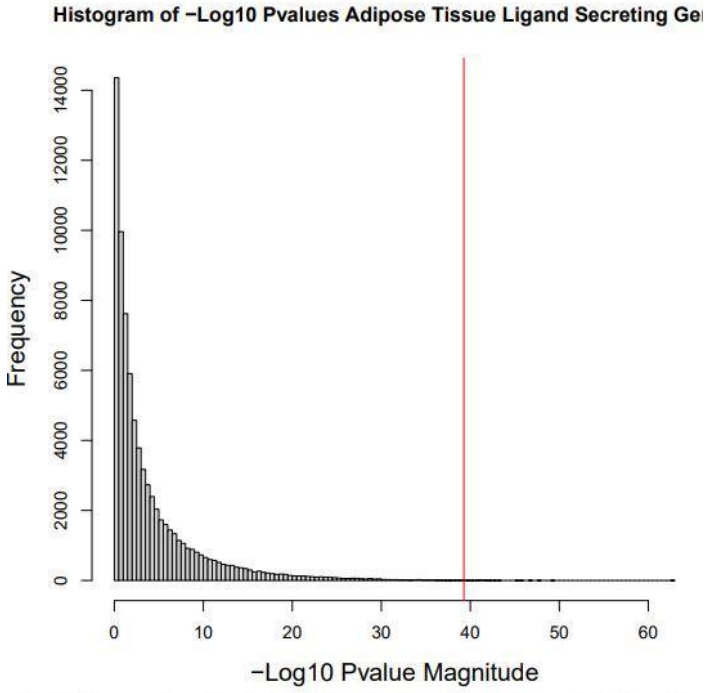

Fig. 3. Histogram of negative log base 10 p-values for adipose to liver analysis. The red vertical line is an indicator for cutoff of p-values used in downstream visualizations based on a user defined input parameter.

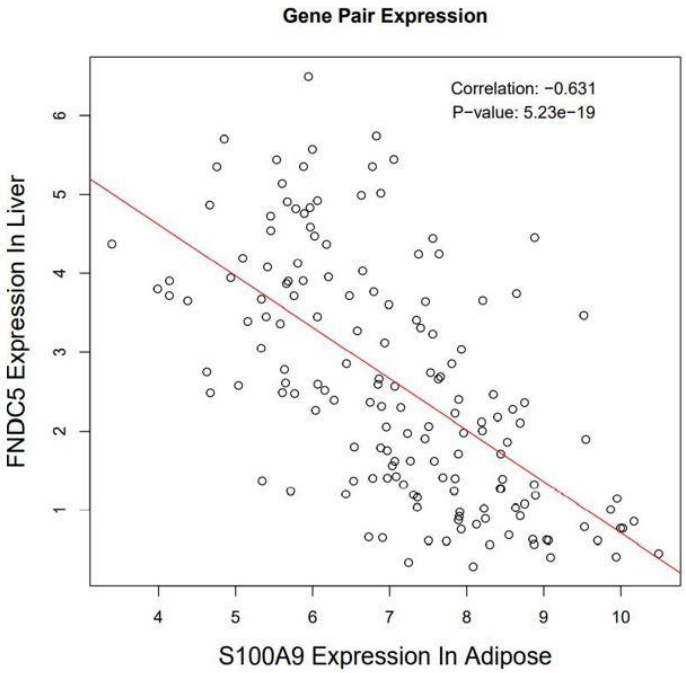

Fig. 4. Scatter plot of gene expression for lowest magnitude p-value gene pair from adipose to liver analysis. The red line is a linear regression of the two gene's expression

Ssec scores gauge the interaction of one origin ligand encoder gene across all genes within the target tissue. An equation for how these scores are calculated is presented in Equation 1. The scores are divided by the number of donors present in the target tissue matrix to standardize and allow for comparison between tissue pair analyses. Table 1 provides gene symbol, Ssecs, and p-values of adipose tissue ligand encoder interactions with liver tissue for the top 10 highest magnitude scores. Measuring the significance of these scores is not apparent however. Because the distribution of Ssec scores is unknown, we opted to produce a null distribution to quantify the significance of observed values in our analysis. To do this, we permute the donor data of the target liver tissue matrix used in the correlation analysis 1000 times and calculate the random Ssec score each time. Then, we calculate the p-value of the observed Ssec from our unaltered analysis on the null distribution. An example

Gene Symbol	Ssec Score	P-value
S100A9	79.80128	.014
S100A8	72.69494	.027
CXCL12	68.56354	.042
NAMPT	66.52788	.05
ANGPTL4	63.51723	.059
TNFSF10	63.25181	.061
VEGFA	62.42633	.068
CX3CL1	56.76088	.12
IL34	55.9022	.128
AZGP1	52.71073	.158

Table 1. Ten highest magnitude Ssec scores for adipose tissue ligand encoders with corresponding p-values

plot for the validation of our Ssec scores is shown in Figure 5.

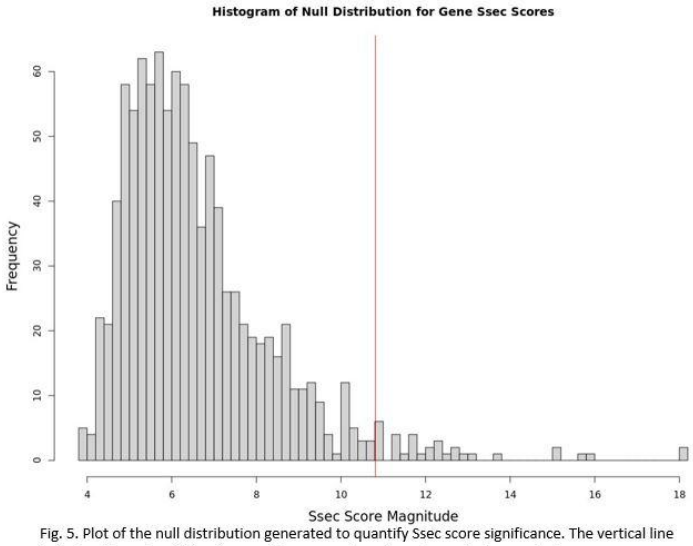

Fig. 5. Plot of the null distribution generated to quantify Ssec score significance. The vertical line indicates the value of the observed Ssec score from the adipose to liver analysis.

Specific Aim 3 - Analyzing novel differences between male and female organ interactions

Prior research has demonstrated that some genes are differentially expressed between males and females (Zhang et al., 2009; Ellegren and Parsch 2007). Whether differential gene expression applies to ligand and receptor encoding genes, responsible for the maintenance of the endocrine system, in males and females is not well documented. We hypothesize that differential endocrine interactions between males and females can be identified through the application of our R-package. In our third aim, we applied our R-package to adipose and liver expression data split by the sex of the donor. These two tissues are known to have secreted factors and receptors that influence tissue function, making this pair an optimal choice for our analysis. Discovery of differential gene interactions between males and females based on correlation coefficients and Ssec scores can illuminate novel differences in tissue communication.

After splitting the adipose and liver tissue matrices by male and female donors, we utilized these pairings as inputs into our corr_analysis function. This process produced two sets of gene-to-gene correlation coefficients, p-values, and Ssec scores. We adapted our two methodologies for identifying tissue pair gene interactions to identify divergence between sexes. The first is taking the difference of biweight midcorrelation coefficients, and analyzing the magnitude of the difference matrix, similar to the p-value analysis mentioned previously. The second is a difference of Ssec scores. The first method allows us to single out particular gene pairs that show considerable deviation in correlation between males and females. The second method allows us to analyze differences in how each origin adipose tissue gene impacts all genes present in the liver tissue.

Matrices resulting from the use of our package on separated male and female data do not inherently match due to filtration by gene variance. Before a difference of male and female correlation coefficients can be taken, the matrices must be realigned to include genes present in both. After this formatting, female coefficients are subtracted from the male's to produce our difference matrix. Lastly, before we identify gene pairs based on the magnitude of coefficient differences, we perform one final

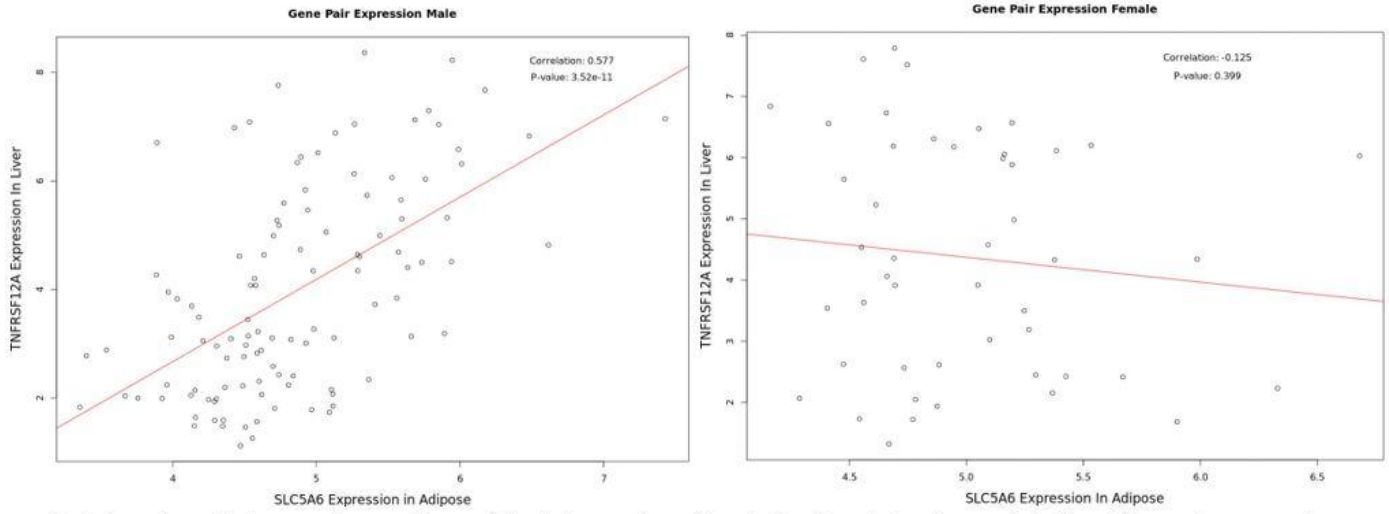

Fig. 6. Comparison plots for gene pair expression correlation between males and females for adipose to liver tissue analysis. The red lines are linear regressions of the expression data. The gene pair visualized was that with the highest magnitude correlation coefficient difference.

filtration of target tissue receptor encoding genes in the liver. This allows us to identify potentially specific genes responsible for endocrine pathways between the adipose and liver tissue. Solute carrier family 5 member 6 (SLC5A6) and Tumor necrosis factor receptor superfamily member 12A (TNFRSF12A) were the pairing with highest magnitude correlation difference of 0.702. Expression scatters for these two genes are presented below in Figure 6. There is a significant positive correlation of the genes in male donors, and a negative correlation in females. Although the p-value for these genes in females is not significant, the identification of similar differential correlations can have vast future implications.

The second method utilized in our analysis pipeline is a difference in Ssec scores. Similar to our methodology above, genes retained after variance filtration differ after function utilization. Therefore, matching of ligand encoding genes between male and female results is the first step. Then, after solving for the male and female difference, we can identify important ligand encoders in adipose that have large impacts on liver tissue gene function differences between sexes. To compare the magnitudes, Figure S1 displays the Ssec differences histogram for matched genes. In Table 2, we display the top ten adipose ligand encoders with differential impacts on the liver. Genes with positive Ssec differences have larger effects on the liver in males than females, while genes with negative Ssec differences demonstrate the opposite effect. Adrenomedullin (ADM) and Vascular endothelial growth factor A (VEGFA) have the largest absolute value differences, and an analysis on their function along with the gene pair identified based on correlation differences will be analyzed in the discussion.

Discussion

Utilizing our bioinformatics program, we identified multiple genes that could play a critical role in differential tissue communication in males and females. ADM is a vasodilator protein, and prior research has identified adipose tissue as a major source of the protein ADM in the body. It was also identified that the expression of ADM increased significantly in an obese mice model (Nambu, 2005). VEGFA plays a key role in angiogenesis and tissue remodeling, but research has also identified it as potentially influencing metabolism and insulin resistance (Elias, 2013). Hypothesizing why we see differences between males and females for these genes is difficult to speculate without more concrete understanding of their function. SLC5A6 is a multivitamin transport protein encoder and TNFRSF12A is an apoptosis modulator and signal protein encoder (NCBI, 2021). Research linking these two genes is limited, even less so for identifying potential interaction differences based on sex.

Gene Symbol in Adipose	Ssec Difference
ADM	-29.285
ANGPTL1	-28.441
ANGPTL4	-27.989
AREG	-25.817
C1QB	-23.885
THBS1	20.422
TNFSF10	26.946
TNFSF14	28.895
VCAN	29.160
VEGFA	34.966

Table 2. Ssec differences for top ten highest absolute value magnitudes of adipose to liver sex-based analysis.

While we have demonstrated the capabilities of our package to perform gene expression analysis and particular sexed based analysis, there are some limitations to our work that should be taken into consideration. Despite the package being developed in tandem with the use of GTEx sourced expression data, the package was produced to accommodate the use of human expression data sourced from other databases. This has yet to be tested; so we assume that non-GTEx sourced information will work as expected. A majority of donors in the GTEx data were male. Some tissue expression matrices had relatively low female donors, such as the liver with only 62 donors with even fewer after tissue pair matching. Increasing the number of donors in general, but especially for females, could increase the accuracy of our correlation analysis. Our package is written exclusively in R, and functions best when in an Rstudio environment. This demands knowledge of the programming language from the user, and could pose as a barrier for entry to those unfamiliar with R. It is entirely possible to translate our software to other programming languages, like Python, but we cannot confirm it would provide a novel purpose in those settings. This package also uses gene expression levels, instead of protein levels in its analysis. It cannot be assumed that differential expression of genes will necessitate similar differential translation of proteins. As such, performing a correlation analysis on protein abundance could complement our findings.

Future directions in the enhancement of our R-package coincide with some of our limitations. Continued improvement of the package will include providing variations that function using other programming languages. It is also important to experimentally validate correlation results from our package analysis. It may be pertinent to confirm whether the differential correlations in gene expression lead to differential translation of proteins. Just as our package was applied to do a group analysis based on sex, the program could be applied to other group based analysis. This could include analysis in groups separated by disease state, age, and geographical location. Genes that aren't retained after filtration, both by variance

and which genes are ligand encoders, may also be pertinent information to analyze in the differences between male and female tissue interactions.

Bioinformatics and data science methodologies are rapidly being implemented in the field of biomedical engineering as tools for better understanding large cohorts of biological information. Study of gene expression in humans is just one of many applications that data science is well suited to address. Our R-package presents the scientific community with an efficient tool for such analysis. With it, we demonstrated a specific application of the program through the analysis of differential endocrine interactions between sexes.

Materials and Methods

Human Gene Expression Data

Gene expression data was downloaded from the GTEx project website using the command line, and stored on the University of Virginia's High-Performance Computing system, Rivanna. The data was then annotated using subject sample identifiers, normalized using log base two, and split into separate tissues using the R programming language in an RStudio environment.

Matrix Processing

Sample donors were matched between tissue pairs to ensure samples contained gene expressions for the relevant tissues. ENSEMBL gene identifiers were renamed into the corresponding HUGO symbols to match genes between matrix pairs (Anderson et al., 2020), and then later to filter for ligands in the origin tissue. The dataset containing the list of ligand secreting genes was provided by the R package GSEAplot (Reinaltt, 2021). Genes were then filtered for high variance, in order to better serve later correlation methods. For our sex based analysis, gene information was rematched after male and female data was processed due to differential retention of certain genes during variance filtration. Finally, for the analysis of gene-gene pairs utilizing p-values, target tissue genes were filtered based on known receptor encoders to better capture endocrine interactions between sexes.

Gene Expression Correlations

Inter-tissue correlation coefficients and corresponding p-values were calculated using the R package WGCNA (Langfelder, 2012). We utilized the biweight midcorrelation method, which is median based, making our package analysis less susceptible to outliers. Gene pairs of significantly low p-value magnitude were isolated and expression data from the initial tissue matrix was called to visualize correlation.

Quantification and Statistical Analyses

The distribution of Ssec scores is unknown, and so to analyze the significance of the scores, our group adopted a method of creating null distributions for ligand encoding genes for the origin tissues. This required the permutation of donor column data 1000 times, and applying an ecdf function in Rstudio to the Ssec scores calculated from the randomly permuted data. Then the p-value for the observed value was calculated on the null distribution to quantify its significance. Similar analysis can be done on differences of Ssec scores for validating significance, but is computationally heavy and will be optimized and implemented in future versions of the package.

Github

As referenced in the text, the R script used to perform the pipeline, along with sample datasets, a vignette walkthrough, and a function manual are all available at https://github.com/Eziedu/Gene_Expression_Cors_JEF

End Matter

Author Contributions and Notes

The authors have no competing interests to declare.

Acknowledgments

We thank Warren Anderson, Ph.D., for their continued support and guidance throughout our research process. We would also like to thank Mete Civelek, Ph.D., and the other members of the Civelek lab for suggestions and feedback on our work.

References

Anderson, Warren D., Joon Yuhl Soh, Sarah E. Innis, Alexis Dimanche, Lijiang Ma, Carl D. Langefeld, Mary E. Comeau, et al. "Sex Differences in Human Adipose Tissue Gene Expression and Genetic Regulation Involve Adipogenesis." *Genome Research* 30, no. 10 (October 2020): 1379–92. https://doi.org/10.1101/gr.264614.120.

Attie, Alan D., and Philipp E. Scherer. "Adipocyte Metabolism and Obesity: Fig. 1." *Journal of Lipid Research* 50, no. Supplement (April 2009): S395–99. https://doi.org/10.1194/jlr.R800057-JLR200

Chang, Phei. "Clinical Bioinformatics." *Chang Gung Medical Journal* 28 (May 1, 2005): 201–11.

Centers For Disease Control and Prevention(CDC). "Underlying Cause of Death, (2018). https://wonder.cdc.gov/controller/datarequest/D76.

Elias Ivet, Sylvie Franckhauser, Fatima Bosch " New Insights into Adipose Tissue VEGF-A Actions in the Control of Obesity and Insulin Resistance." *Adipocyte* 2, no. 2 (November 12, 2012). https://www.tandfonline.com/doi/full/10.4161/adip.22880.

Ellegren, Hans, and John Parsch. "The Evolution of Sex-Biased Genes and Sex-Biased Gene Expression." *Nature Reviews Genetics* 8, no. 9 (September 2007): 689–98. https://doi.org/10.1038/nrg2167.

Emdin, Connor A., Amit V. Khera, Pradeep Natarajan, Derek Klarin, Seyedeh M. Zekavat, Allan J. Hsiao, and Sekar Kathiresan. "Genetic Association of Waist-to-Hip Ratio With Cardiometabolic Traits, Type 2 Diabetes, and Coronary Heart Disease." *JAMA* 317, no. 6 (February 14, 2017): 626. https://doi.org/10.1001/jama.2016.21042.

Herrera, Blanca M., and Cecilia M. Lindgren. "The Genetics of Obesity." *Current Diabetes Reports* 10, no. 6 (December 2010): 498–505. https://doi.org/10.1007/s11892-010-0153-z.

Hoopes, Laura. "Gene Expression and Regulation | Learn Science at Scitable." *Nature Education* (2014). https://www.nature.com/scitable/topic/gene-expression-and-regulation-15/.

Karastergiou, Kalypso, Steven R. Smith, Andrew S. Greenberg, and Susan K. Fried. "Sex Differences in Human Adipose Tissues – the Biology of Pear Shape." *Biology of Sex Differences* 3, no. 1 (May 31, 2012): 13. https://doi.org/10.1186/2042-6410-3-13.

Khera, Amit V., and Sekar Kathiresan. "Genetics of Coronary Artery Disease: Discovery, Biology and Clinical Translation." *Nature Reviews Genetics* 18, no. 6 (June 2017): 331–44. https://doi.org/10.1038/nrg.2016.160.

Langfelder, Peter, and Steve Horvath. "WGCNA: An R Package for Weighted Correlation Network Analysis." *BMC Bioinformatics* 9, no. 1 (December 29, 2008): 559. https://doi.org/10.1186/1471-2105-9-559.

Langfelder, Peter, and Steve Horvath. "Fast R Functions for Robust Correlations and Hierarchical Clustering." *Journal of Statistical Software* 46, no. 11 (March 2012). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3465711/.

Nambu, Takuo, Hiroshi Arai, Yasato Komatsu, Akihiro Yasoda, Kenji Moriyama, Naotetsu Kanamoto, Hiroshi Itoh, and Kazuwa Nakao. "Expression of the Adrenomedullin Gene in Adipose Tissue." *Regulatory Peptides* 132, no. 1 (December 15, 2005): 17–22. https://doi.org/10.1016/j.regpep.2005.07.006.

Reches, A., K. Weiss, L. Bazak, H. Baris Feldman, and I. Maya. "From Phenotyping to Genotyping - Bioinformatics for the Busy Clinician." *European Journal of Medical Genetics* 62, no. 8 (August 1, 2019): 103689. https://doi.org/10.1016/j.ejmg.2019.103689.

Reinaltt, Kelsie, Sarah Innis, Mete Civelek, Warren Anderson. *Kelsiereinaltt/GSEAplot*. R, (2021). https://github.com/kelsiereinaltt/GSEAplot.

Seldin, Marcus M., Simon Koplev, Prashant Rajbhandari, Laurent Vergnes, Gregory M. Rosenberg, Yonghong Meng, Calvin Pan, et al. "A Strategy for Discovery of Endocrine Interactions with Application to Whole-Body Metabolism." *Cell Metabolism* 27, no. 5 (May 2018): 1138-1155.e6. https://doi.org/10.1016/j.cmet.2018.03.015.

Virani, Salim S., Alonso Alvaro, Benjamin Emelia J., Bittencourt Marcio S., Callaway Clifton W., Carson April P., Chamberlain Alanna M., et al. "Heart Disease and Stroke Statistics—2020 Update: A Report From the American Heart Association." *Circulation* 141, no. 9 (March 3, 2020): e139–596. https://doi.org/10.1161/CIR.0000000000000757.

Wilcox, Rand R. *Introduction to Robust Estimation and Hypothesis Testing*. Academic Press, 2011.

Zhang, Wei, R. Stephanie Huang, Shiwei Duan, and M. Eileen Dolan. "Gene Set Enrichment Analyses Revealed Differences in Gene Expression Patterns between Males and Females." *In Silico Biology* 9, no. 3 (2009): 55–63.

Supplementary Materials and Figures

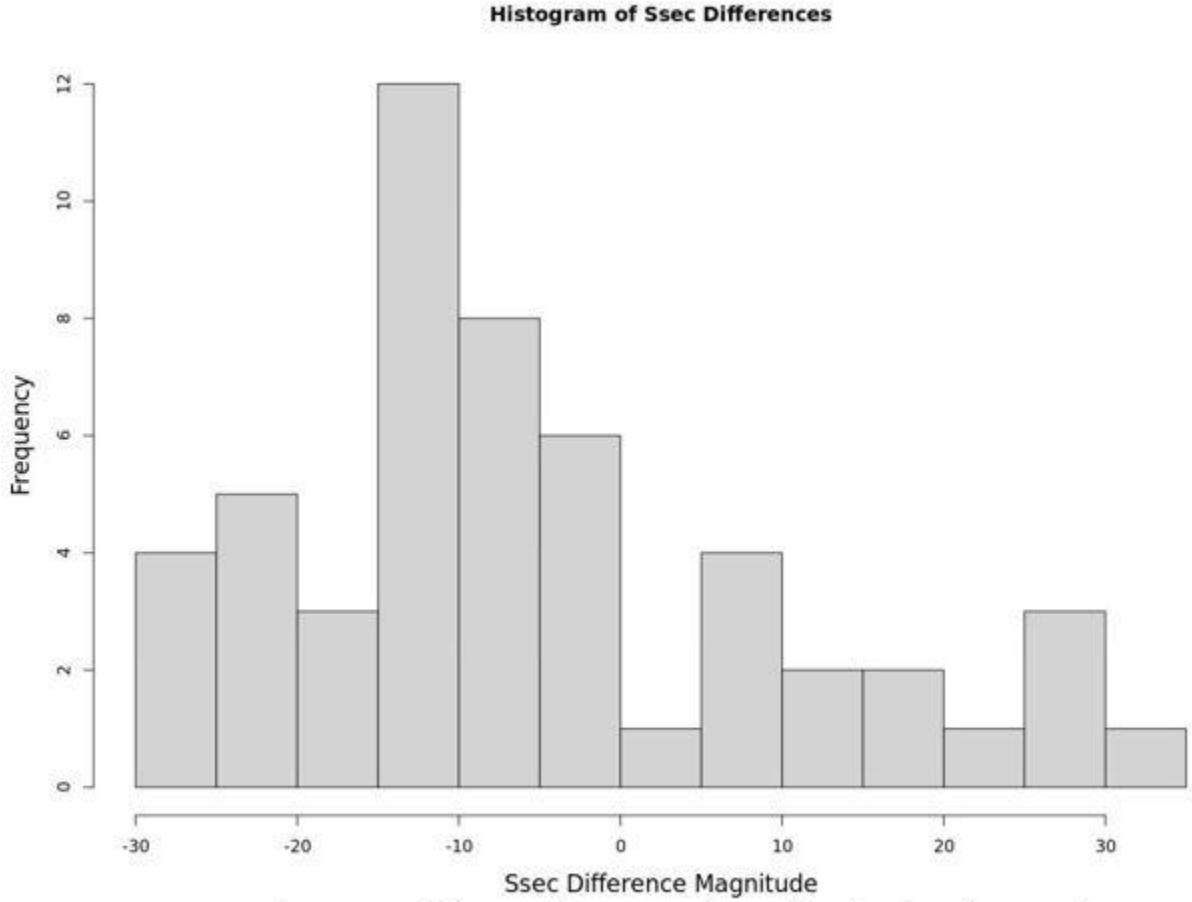

Fig. S1. Histogram of Ssec score differences between males and females for adipose to liver analysis.