Association of Regulatory Proteins in Systemic Lupus Erythematosus and Lupus Nephritis.

Completed for the Certificate in Scientific Computation
Summer 2021

Maria Basham
Scientific Computation and Data Analysis
Department of Statistics and Data Sciences
University of Texas at Austin

Layla Guyot

[Supervising Faculty's Signature]

_____

Dr. Layla Guyot
Lecturer
Department of Statistics and Data Sciences.

Abstract

Systemic Lupus Erythematosus is a rare, multi-systemic, autoimmune disease that can damage many parts of the body, including skin, joints, kidneys, heart, blood vessels, and brain(Choi et al. 2012). The exact pathogenesis of SLE is still unknown but it seems that inflammation, dysregulation of immune system, and events caused by active disease manifestation can lead to complications like, cancer, arteriosclerosis, gastrointestinal bleeding, renal failure, including renal dialysis, transplantation, and infections, which are some the most significant causes of mortality in SLE patients(Doria 2006) This study aims to find possible common pathways between systemic lupus erythematosus and lupus nephritis pathologies.

Bioinformatics is crucial to generate new knowledge from combined information. Microarray technology can provide an understanding of patterns of expression of genes in response to factors to which the organism is exposed. This sophisticated expression profiling contributes to the understanding of biological mechanisms further facilitating the identification of possible pathways and the role of the proteins involved (Food and Agriculture Organization of the United Nations 2007). This study intended to investigate the common and pathways involving SLE and LN pathogenesis from their protein interactions.

A gene expression analysis of one microarray datasets of two lymphocytes (CD4+ and CD8+) was conducted to identify differentially expressed genes. The source of the sample is of the patients' blood cells, lymphocytes, to identify shared expressions involved in patients with SLE and patients with LN.

The most significant genes were mapped to the search tool for retrieval of interacting genes (STRING) to acquire protein-protein interaction (PPI) networks. Finally, Enrichr, a gene list enrichment analysis tool, was utilized for the functional enrichment of clusters.

The enriched pathological pathways are consistent with the pathogenesis of SLE and LN. However, the LN condition makes it more susceptible to the initiation of more complicated pathways such as viral and bacterial infections.

Introduction

Systemic Lupus Erythematosus (SLE) is a systemic autoimmune disease that occurs when your body's immune system, which normally helps protect the body from infection and disease, attacks your own tissues and organs. Some of the most common symptoms of lupus are inflammation, which causes the immune system to activate and begin to recognize and attack its own body tissues causing permanent scarring and jeopardizing the function of certain organs and systems in the body.

There are several kinds of lupus; SLE is the most common and can affect many parts of the body. Some types of lupus and their causes include Discoid lupus that usually causes red rashes, subcutaneous lupus that inflicts sores after sun exposure , and neonatal lupus, which is rare and affects newborns, and even drug induced lupus

caused by certain medicines. Some of the most common lupus symptoms happen when the disease causes inflammation in organs, including symptoms like arthritis, muscle pains, fever, hair loss, stomach pain, breathing complications, swollen glands, sun sensitivity , mouth ulcers, and “butterfly rash”. (Johns Hopkins Lupus Center 2021). While lupus is thought to develop due to an interaction between genetic susceptibility and environmental triggers, the exact cause of this disease is not well known. Efforts have been made in trying to understand the SLE pathogenesis through the study of the mechanisms that underlie the signaling activation of antibody responses, however it has not been entirely understood yet.

Prediction

T cell activation plays a major role in SLE disease to fight and adapt to intruders and it can be both defensive and detrimental for the body. Lymphocytes, like regulatory T cells, are direct fighters of foreign invaders;they produce cytokines, which are biological substances that help activate other parts of the immune system, and macrophages, that clean up the invaders and the dead tissue after an immune response. Aberrant T cell expression is linked to abnormal T cell activation, thus, deregulation of T cell receptor signals plays a major role in SLE disease.

TCR responses to microbial infection and adaptive immune response, and its signals are critical for T cell development. Defects of TCR signals have been implicated in multiple human diseases. Protein receptors play an equal role in the regulatory process of activation and responses, and are indispensable for the homeostasis of the inflammatory system(Bai 2017). Considering that the body's regulatory systems, like immune cells, greatly influence immune function, this study considers the Lymphocytes,CD4+, and CD8+ to try to find a statistically significant difference in the read counts between SLE and Healthy experimental conditions.

Methods

Gene expression data set selection

The present study considers two subgroups of cell signaling proteins CD4+ and CD8+, that play a major role in this regulatory process of activation and responses, and are indispensable for the homeostasis of the inflammatory system(Bai 2017). The source of the sample is of the patients “whole blood” to identify shared expressions involved in patients with SLE and patients with LN.

Expression datasets regarding SLE, CD4+, and CD8+ were obtained from the national center for biotechnology information(NCBI) gene expression omnibus (GEO) archive(Barrett et al. 2012). The dataset with accession number GSE49454, which contains the gene expression in the form of matrix, was selected for this study. The dataset of whole blood comprised 177 samples of which 157 are SLE patients, and 20 are matched controls cases. In addition, 2 groups of SLE were created, one with SLE and Healthy control that contained 97 sample cases, and one with SLE and LN with 60 sample cases.

Dataset preprocessing

The data obtained as series matrix files and related annotations were obtained from the GEO database and preprocessed using R statistical programming language. We divided the cases into different groups based on the pathogenesis SLE and LN and healthy control types. The data was normalized using a quantile normalization technique function for quality control. This normalization makes the empirical distribution of gene expression of Illumina array to be the same(Qiu 2013).The raw expression levels were log2 transformed to moderate the variance across the mean, thereby improving the distances for visualization methods.

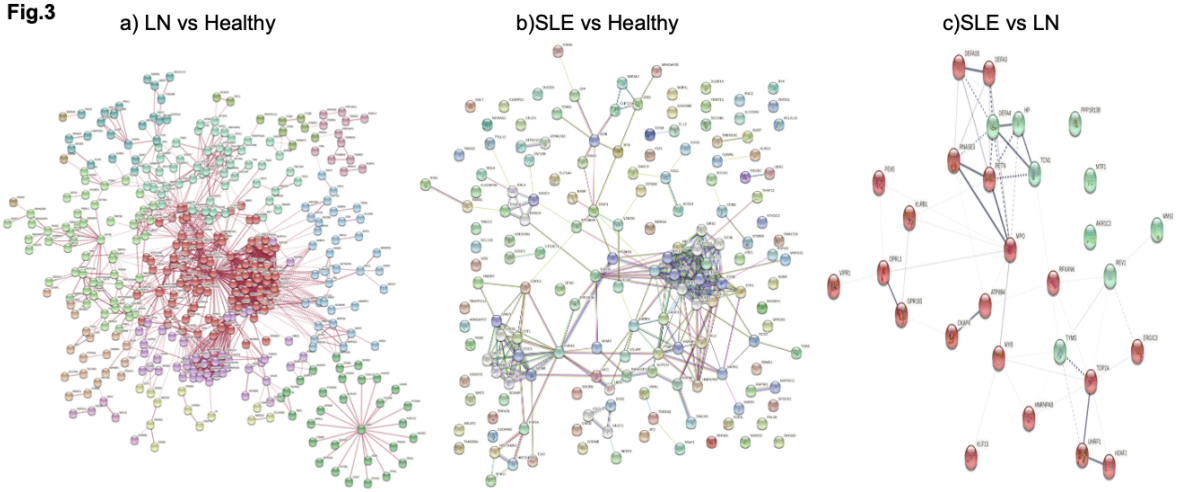
Identifying differentially expressed genes (DEGs)

The linear models for microarray data, was used to identify the differentially expressed genes from the selected T cells, CD4+ and CD8+ lymphocytes. This function treats each gene as an independent analysis and thus improves inference by sharing information across samples(Wu et al. 2015). To identify the set of genes that are going to be used in our enricher analysis, gaining statistical power is important while maintaining some bound error. The FDR method, introduced by Benjamini-Hochberg, is defined as the expected proportion of falsely rejected hypotheses among the set of rejected hypotheses if there is at least one rejection, and zero otherwise. This function is included in the R function “p. Adjust” under the names “BH”, and “fdr”(Loon 2017). The Benjamini-Hochberg procedure guarantees to control for false discovery rate and assumes independent test statistics. To apply the BH method we considered genes with an adjusted p-value < 0.05 and the logarithm of fold change (logFC) $> \pm 0.5$.

Protein-protein interaction(PPI)

The DEG values were then uploaded to the search tool for retrieval of interacting genes/proteins (STRING) database to visualize potential significant interactions between the differentially expressed genes(Mering et al. 2018). Full network analysis was conducted applying active interaction sources, including text mining, experiments, databases, and co-expression, neighborhood, gene fusion, and co-occurrence, with a medium level of confidence, as well as species limited to homo sapiens. Results of local network clusters STRING were considered to create the visualization based on the number of significant clusters. Fig. 3 shows the protein networks involved in this study.

Fig.3



PPI network of LN vs Healthy, SLE vs Healthy and SLE vs LN. each cluster is a set of connected nodes as illustrated in a discrete color

Functional and pathway enrichment analysis

The differentially expressed genes/proteins obtained were further analyzed with Enrichr, an integrative web-based software that presents an alternative approach to rank enriched terms. Enrichr allows the evaluation of annotations with its extensive gene-set libraries(Chen et al.2013). The ontology and pathways categories contained the gene set libraries, Gene Ontology Biological Process, and 2021 KEGG 2021 Human, that were used to perform enrichment analysis. To obtain the cluster grams with the enriched terms, the significant terms and pathways were selected with the threshold of adjusted p-value < 0.05. GO biological processes 2021 shows two significant pathways, and KEGG 2021 human showed eight pathways of significant results are listed in Tables 1,2, respectively.enriched terms. Enrichr allows the evaluation of annotations with its extensive gene-set libraries(Chen et al.2013). The ontology and pathways categories contained the gene set libraries, Gene Ontology Biological Process, and 2021 KEGG 2021 Human, that were used to perform enrichment analysis. To obtain the cluster grams with the enriched terms, significant terms and pathways were selected with the threshold of adjusted p-value < 0.05. KEGG 2021 humans showed eight significant pathways, and GO biological processes 2021 showed two significant pathways. The results are listed in Tables 1,2, respectively.

Table 1 KEGG pathways in the LN vs Healthy group with eight significant values.

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	Ribosome	2.241e-10	6.566e-8	4.34	96.42
2	Coronavirus disease	2.948e-9	4.319e-7	3.35	65.86
3	Th17 cell differentiation	0.00009983	0.009750	3.19	29.39
4	Th1 and Th2 cell differentiation	0.0001873	0.01372	3.29	28.21
5	Epstein-Barr virus infection	0.0004635	0.02716	2.28	17.51
6	Human T-cell leukemia virus 1 infection	0.0006492	0.03170	2.18	16.00
7	Legionellosis	0.001205	0.04642	3.58	24.06
8	Transcriptional misregulation in cancer	0.001267	0.04642	2.19	14.58
9	T cell receptor signaling pathway	0.002111	0.06873	2.62	16.14
10	Inflammatory bowel disease	0.003325	0.08325	3.06	17.45

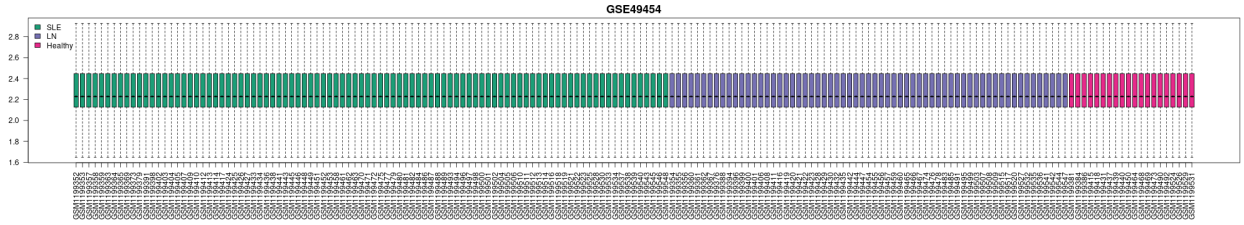
Tabe 2 GO biological pathways for SLE and LN group with two significant values.

Index	Name	P-value	Adjusted p-value	Odds Ratio	Combined score
1	bundle of His cell action potential (GO:0086043)	0.00002473	0.007170	443.67	4706.27
2	AV node cell action potential (GO:0086016)	0.00005182	0.007514	266.17	2626.53
3	positive regulation of striated muscle cell differentiation (GO:0051155)	0.0008490	0.06155	53.18	376.07
4	positive regulation of muscle cell differentiation (GO:0051149)	0.0008490	0.06155	53.18	376.07
5	regulation of muscle cell differentiation (GO:0051147)	0.001428	0.06901	40.27	263.85

Results

Data quality control

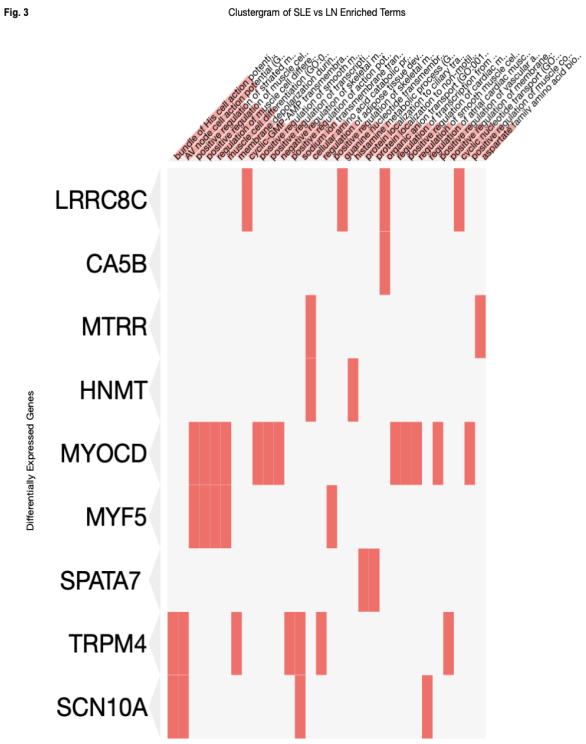
To ensure a similar distribution of the data, the limma package function, normalized between arrays, was applied. The box-set of the groups after normalization is illustarted in Fig. 1, showing a different color for each group.



Network Analysis of the DEGs

A total of 138 DEGs were identified between LN-healthy and SLE-LN,1196 DEGs between SLE-Healthy and LN-Healthy, and 28 DEGs were identified in the intersection

of the three groups. To further understand the relationship between the differentially expressed genes, clustergram analysis and visualization were applied. Each row contains one of the DEGs observations and the enriched terms variable dimensions are the columns. Enricher's implemented scoring method uses hierarchical analyses of data to find the ranks for each observation, and plot their cumulative distributions based on the membership values 0 (white square), and 1 (red square) according to their respective clusters. The interaction enrichment indicates that the proteins are at least partially biologically connected(Kuleshov et al. 2016). Fig. 2 shows the clustergram of LN vs Healthy KEGG. Fig.3 shows the clustergram of SLE vs LN GO biological process.



Enrichment analysis of clusters

The first enriched cluster is associated with ribosomal proteins which are involved in the genetic information processing pathway. The cytoplasmic translation process uses essentially ribosome proteins for assembly and protein translation. Its mis regulations are associated with the activation of the tumor suppressor p53 pathway in response to ribosomal stress, which provokes the accumulation of ribosome-free ribosomal proteins, and other functions like immune signaling, and development(Zhou et al.).

The second cluster is associated with viral infectious disease, coronavirus. Some viruses like COVID-19 have evolved ways to destabilize RIG-like receptor (RLR) signaling to enhance their survival. Immunity responses to viruses rely largely on a set of antiviral factors like type-1 IFNs, which are secreted by infected cells, to activate innate immune responses leading to a cascade of signals for the recruitment of

inflammatory cells that produce cytokines(Jerala 2017). Cytokines, regulators of T cell differentiation, can kill tissue and damage the organs. Failure of the signaling regulation process can lead to overproduction of inflammatory cytokines potentially resulting in serious disorders(Koyama et al. 2007)

The third cluster is associated with the immune system and Th17 cell differentiation pathway. T helper type 17 cells are involved in mucosal immunity and autoimmune disorders. This type of cells are proinflammatory, releasing cytokines leading to the recruitment of neutrophils and macrophages to sites of inflammation, which in turn produce more cytokines and proteases that further exacerbate the immune response. Thus, Th17 cells can have beneficial and pathogenic effects. The persistent secretion of Th17 promotes chronic inflammation and may be involved with rheumatoid arthritis, multiple sclerosis, psoriasis, systemic lupus erythematosus, asthma, and inflammatory bowel disorders(Moran 2021).

The fourth cluster is associated with the immune system and Th1 and Th2 cell differentiation pathway. These cells play a unique and important role in immunity. Cytokines are the most effective regulator, but transcription factors can also play a major role in cell differentiation. CD4+ helper T cells (Th) are divided into two groups, type I helper T lymphocytes (Thl) and type II helper T cells (Th2) that synthesize different characteristic cytokines, IFN-Y, and IL-4, respectively. The overexpression of Th2 cells can lead to inappropriate immune responses, and overexpression of Th1 is associated with autoimmune diseases like rheumatoid arthritis(Berger 2000).

The fifth cluster is associated with viral infectious disease and Epstein-Barr Virus pathway. This virus has the ability to become dormant for life in B lymphocytes after infection. It is known to be common in autoimmune diseases like lupus but the mechanisms are unknown. Recent findings suggest an involvement with gene regulation, and also to work through human transcription factors, which bind DNA and affect gene expression. The EBV virus is thought to drive the activation of genes that contribute to an individual’s risk to develop an autoimmune disease(National Institutes of Health 2018).

The sixth cluster is associated with viral infectious disease and human T-cell leukemia virus 1 infection (HTLV-1). This type of infection is retroviral that uses ribonucleic acid (RNA) as their genetic material, this means that it behaves backwards from the original way, which is that DNA makes RNA, and RNA makes proteins. This virus binds to its receptor in the surface of immune cells, CD4 T cells in the case of AIDS, and once the viral DNA is inserted to the host cell genome, it uses the host cell's machinery to produce new viral components like viral RNA and viral proteins. In humans this type of retrovirus causes a form of cancer called adult T-cell leukemia (ATL) (Healthline 2018).

The seventh cluster is associated with bacterial infectious disease and the legionellosis pathway. Legionellosis is a potentially fatal infectious disease associated with Legionella species. Part of its mechanisms for continued intracellular replication

includes manipulation of host cell death and survival. The legionella bacteria are found in freshwater environments and water containing this bacteria can spread in droplets small enough for people to breathe. Legionella causes serious types of lung infection (pneumonia) and people with weak immune systems, and underlying illnesses are at increased risk of infection(CDC 2021).

The eighth enriched cluster is associated with cancer, human disease transcriptional and misregulation pathways. The genes encoding transcription factors (TFs) in tumor cells are often amplified, deletead, rearranged or suffer loss-of-function mutations that reduce or abolish protein production. Thus, these fusion proteins have aberrant transcriptional functions that alter expression of targeted genes, leading to alteration of cellular properties that contribute to the tumorigenic process(KEGG 2020).

There were not significant KEGG clusters found in the SLE vs Healthy, and SLE vs LN groups. However, GO biological processes found two significant clusters in the SLE vs LN group. Cluster one is associated with the cardiac conduction system, the electrical pathway. The bundle of His is a specialized muscle bundle connecting the atrial and ventricular chamber of the hearth. The electrical impulses of this system result in a synchronized contraction of the ventricles, thus, malfunction of AV conduction system can lead to cardiac arrhythmia(ACHA 2008).

The second cluster is associated with the atrioventricular node cell action potential pathway. The AV node is a gatekeeper between the atria (upper chamber), and the ventricles (lower chamber), and its main role is to conduct the impulse from the atria to the ventricles. The AV node serves as a backup pacemaker when there are problems with the sinus node, which sets the pace for the heart. Failure of AV node function can lead to a condition called hearth block(Cyril et al. 2021).

Discussion

Overall, this study shows the strong impact that transcriptional and regulatory systems have in the pathogenesis of SLE and LN. TCR recognizes self or foreign antigens and sends signals that lead to the generation of second messengers and subsequent activation of signaling cascades, which control many aspects of T cell biology(Gorentla and Zhong 2012).

The LN vs Healthy group presented the most DEGs and an analysis of its functional clusters presented the complexity of regulatory proteins and their impact in the body. The strongest associations among the studied samples include misregulation of genetic information processing, viral infection diseases like coronavirus, Epstein-Barr, and human T-cell leukemia virus, Th 17, Th1, and Th2 differentiation, bacterial infections diseases like legionellosis, including cancer transcriptional and misregulation pathways. These results indicate that a high prevalence of viral and bacterial infections is most common among lupus nephritis patients. This can be due to drug therapies like immunosuppressants compared to the SLE group whose most common drug treatment involves steroids.

Even though there was no clustering in the SLE vs LN, and SLE vs Healthy groups, enrichment analysis suggests that these proteins have more interaction among themselves than what would be expected from a random set, indicating that the proteins are at least partially biologically connected.

The SLE vs LN group showed no clustering, however, three were two GO biological processes pathways associated with cell signaling involved in cardiac conduction, as well as cell to cell signaling and regulation of atrioventricular cardiac muscle cell action. These findings suggest that people with SLE and LN present more cardiac manifestations like arrhythmia.

Limitations for the study can arise from the sample's profile. The study included blood samples of SLE and LN patients that were on common drug therapies like hydroxychloroquine, azathioprine, prednisone, methotrexate, and others.

Areas for future exploration may include the role of phosphorylation, dephosphorylation in cell signaling, mRNa breakdown, catabolism, and degradation. Moreover, blood sampling from SLE patients should be conducted in different stages of the disease to perform more accurate expression comparisons and design effective systemic or targeted therapies for SLE in the future.

Acknowledgements

I wish to show my appreciation to Dr. Layla Guyot for the support, assistance, and suggestions in the realization of this study. I thank Dr. Enamul Huq for the useful discussion. I thank my family for their patience and support. Finally, I want to thank my late mother who inspired me to do this study. I will be ever grateful for her encouragement, and am sorry that she has not lived to see me graduate.

Abbreviations

SLE: Systemic Lupus Erythematosus

LN: Lupus Nephritis

DEGs: Differentially Expressed Genes

TCR: T cell receptor

CD4+: Cluster of Differentiation 4 or co-receptor of TCR

CD8+: Cluster of Differentiation 8 or co-receptor of TCR

GEO: Gene Expression Omnibus

GO: Gene Ontology

FDR: False discovery rate

BH: Benjamini-Hochberg

logFC: Logarithm of fold change

STRING: Search Tool for the Retrieval of Interacting Genes/Proteins

PPI: Protein Protein Interaction

KEGG: Kyoto Encyclopedia of Genes and Genomes

ER: Endoplasmic Reticulum

Th17: T helper cell type 17

Th1: T helper cell type 1

Th2: T helper cell type 2

RLR: RIG-like receptor

IFN-Y: Interferon gamma

IL-4: Interleukin 4

EVB: Epstein-Barr Virus

HTLV-1: Human T-cell leukemia virus 1 infection

ATL: Adult T-cell leukemia

TFs: transcription factors

AV: Atrioventricular

R: The R project for statistical computing

References

1. Jinyoung Choi, Sang Taek Kim, Joe Craft. The pathogenesis of systemic lupus erythematosus—an update. Science Direc. December 2012;24(6):651–657. http://dx.doi.org/10.1016/j.coi.2012.10.004. doi:10.1016/j.coi.2012.10.004
2. Andrea Doria, Luca laccarino, Anna Ghirardello, Fabiola Atzen, Antonio Piccoli, Silvano Todesco. Long Term Prognosis and Causes of Death in Systemic Lupus. The American Journal of Medicine. 2006;119(8):700–706. doi:10.1016/j.amjmed.2005.11.034
3. Johns Hopkins Lupus Center. How Lupus Affects the Body. hopkins lupus.org. 2021 [accessed 2021 Jun 6]. https://www.hopkinslupus.org/lupus-info/lupus-affects-body/
4. Food and Agriculture Organization of the United Nations. State of the World's Animal Animal Genetic Resources for Food and Agriculture. Rome, Italy: Viale delle Terme di Caracalla; 2007.
5. Aiping Bai YG. Acid sphingomyelinase mediates human CD4+ T-cell signaling: potential roles in T-cell responses and diseases. Nature. 2017 Jul 27. https://www.nature.com/articles/cddis2017360. doi:10.1038/cddis.2017.360
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. 2012;41(D1):991–995.
https://academic.oup.com/nar/article/41/D1/D991/1067995. doi:10.1093/nar/gks1193
7. Xing Qiu HW&. RH. The impact of quantile and rank normalization procedures on the testing power of gene differential expression analysis. BMC Informatics. 2013;(124).
https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-124.
doi:10.1186/1471-2105-14-124
8. Di Wu, Yifang Hu, Charity W. Law, Wei Shi, Gordon K. Smyth, Matthew E Ritchie, Belinda Phipson. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids . 2015;43(7). https://academic.oup.com/nar/article/43/7/e47/2414268.
doi:10.1093/nar/gkv007
9. Von Mering DSALGDLAJSWJH-CMSNTDJHMPBLJJC. STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. Nucleic Acids. 2018;47(D1):6007–6613.
https://academic.oup.com/nar/article/47/D1/D607/5198476?login=true. doi:10.1093/nar/gky1131
10. Edward Y Chen, Christopher M Tan, Yan Kou, Qiaonan Duan, Zichen Wang, Gabriela Vaz Meirelles, Neil R Clark & Avi Ma'ayan. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013 Apr 15.
https://bmcbioinformatics.biomedcentral.com/articles/10.1186/1471-2105-14-128.
doi:10.1186/1471-2105-14-128
11. Maxim V. Kuleshov, Matthew R. Jones, Andrew D. Rouillard, Nicolas F. Fernandez, Qiaonan Duan, Zichen Wang, Simon Koplev, Sherry L. Jenkins, Kathleen M. Jagodnik, Alexander Lachmann, Michael G. McDermott, Caroline D. Monteiro, Gregory W. Gundersen, Avi Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic . 2016;44(W1):90–97.
https://academic.oup.com/nar/article/44/W1/W90/2499357?login=true. doi:10.1093/nar/gkw377
12. Xiang Zhou, Wen-Juan Liao, Jun-Ming Liao, Peng Liao, Hua Lu. Ribosomal proteins: functions beyond the ribosome. PubMed. 2015;7(2):92–104.
https://pubmed.ncbi.nlm.nih.gov/25735597/.
doi:Zhou,X.,Liao,W.J.,Liao,J.M.,Liao,P.,&Lu,H.(2015).Ribosomalproteins:functionsbeyondtheribosome.Journalofmolecularcellbiology,7(2),92–104.https://doi.org/10.1093/jmcb/mjv014
13. Kopitar-Jerala N. The Role of Interferons in Inflammation and Inflammasome Activation. Frontiers in Immunology. 2017 Jul 25.
https://www.frontiersin.org/articles/10.3389/fimmu.2017.00873/full.
doi:10.3389/fimmu.2017.00873
14. Shohei Koyama, Ken J. Ishii, Himanshu Kumar, Takeshi Tanimoto, Cevayir Coban, Satoshi Uematsu, Taro Kawai and Shizuo Akira. Differential Role of TLR- and RLR-Signaling in the Immune Responses to Influenza A Virus Infection and Vaccination. The Journal of Immunology. 2007;179(7):4711–4720. https://www.jimmunol.org/content/179/7/4711/tab-article-info.
doi:10.4049/jimmunol.179.7.4711

15. Ellen Margaret Moran , Charles Institute of Dermatology, University College Dublin, Ireland. Th17 Cells. British Society for Immunology. 2021.
https://www.rndsystems.com/research-area/th17-cells
16. Berger A. Th1 and Th2 responses: what are they? The BMJ. 2000 Aug 12:321–424.
https://www.bmj.com/content/321/7258/424.1.full. doi:10.1136/bmj.321.7258.424
17. National Institutes of Health. Epstein-Barr virus and autoimmune diseases. National Institutes of Health. 2018 Apr 24.
https://www.nih.gov/news-events/nih-research-matters/epstein-barr-virus-autoimmune-diseases
18. Healthline. What Is a Retrovirus? Healthline. 2019.
https://www.healthline.com/health-news/covid-19-updates#5/17/21-2:26-p.m.-PDT-U.K.-Prime-Minister-Boris-Johnson-urges-heavy-dose-of-caution-as-country-reopens
19. CDC. Legionella (Legionnaires' Disease and Pontiac Fever). U.S. Department of Health & Human Services; 2021. https://www.cdc.gov/legionella/about/index.html
20. KEGG. Transcriptional misregulation in cancer - Homo sapiens (human). 2020.
https://www.genome.jp/kegg-bin/show_pathway?hsa05202
21. Rhodes LA, editor. The Heart and its Electrical System. Adult Congenital Heart Association; 2008.
https://www.uclahealth.org/heart/workfiles/Adult_Congenital/heartanditselectricalsystem.pdf
22. Patra, Cyril, Zhang Xinping, Mark F Brady. Physiology, Bundle of His. 2021.
https://www.ncbi.nlm.nih.gov/books/NBK531498/
23. Balachandra K Gorentla 1 X-PZ. T cell Receptor Signal Transduction in T lymphocytes. National Library of Medicine. 2012 Oct 27. https://pubmed.ncbi.nlm.nih.gov/23946894/. doi:10.4172/2155-9899.S12-005.
24. Michael Goldberg, Janice Fischer, Leroy Hood, Leland Hartwell. Genetics:From Genes to Genomes. New York, NY: McGraw Hill; 2021.
25. van Loon W. The Power of the Benjamini-Hochberg Procedure. The Netherlands: Leiden Mathematical Institute; 2017. https://www.math.leidenuniv.nl/scripties/MastervanLoon.pdf26. Elham Karimizadeh, Ali Sharifi-Zarchi, Hassan Nikaein, Seyedehsaba Salehi, Bahar Salamatian, Naser Elmi, Farhad Gharibdoost and Mahdi Mahmoudi. Analysis of gene expression profiles and protein-protein interaction networks in multiple tissues of systemic sclerosis. BMC Medical genomics. 2019;12(199).
https://bmcmedgenomics.biomedcentral.com/track/pdf/10.1186/s12920-019-0632-2.pdf. doi:10.1186/s12920-019-0632-2

Computer code

The code for the differential expression analysis can be found here.