Transcriptomic Profiling of the Hippocampus in a Mouse Model of Post-Traumatic Stress Disorder Using RNA-Seq Analysis

BIN 508 : Next Generation Sequence Analysis and Informatics

<u>Term Paper</u>

Taha Ahmad - 2546125

Submission Date: 20.06.2025

Instructor: Prof Dr YESIM AYDIN SON

Table of Contents

1. Abstract
2. Biological Background
3. Dataset Description
4. Workflow
5. Quality Metrics Summary (FastQC & MultiQC)
6. STAR Alignment Results
7. Infer Experiment Analysis
8. FeatureCounts Summary
9. DESeq2 Outputs and Statistical Results
10. Heatmap Analysis
11. PCA and Sample Distance Matrix
12. GOseq Enrichment Findings
13. KEGG Pathway Interpretation
14. Conclusion
15. References

Galaxy history : https://usegalaxy.eu/u/taha.ahmad/h/copy-of-test

Abstract:

Post traumatic stress disorder (PTSD) is a severe neuropsychiatric condition which is characterized by dysregulation of neural circuits particularly within the hippocampus, a region of the brain which is critically involved in memory consolidation and emotional regulation. To investigate the transcriptomic alterations associated with PTSD we performed RNA-Seq analysis on the hippocampal tissue obtained from a murine PTSD model and control mice. High throughput sequencing data were processed using quality control, adapter trimming and alignment against the *Mus musculus* GRCm39(mm39) reference genome. Differential gene expression analysis was conducted using DESeq2 followed by functional enrichment analysis using Gene Ontology (GO) and KEGG pathway analysis. The findings show significant transcriptional changes affecting genes involved in synaptic function, ribosomal activity and cellular signalling pathways. The results contribute to the growing understanding of molecular mechanisms underlying PTSD and may provide insight for future therapeutic targets.

Biological Background:

Post-traumatic stress disorder (PTSD) is a complex psychiatric condition arising from exposure to traumatic events. It is marked by symptoms of avoidance, hyperarousal, and persistent re-experiencing, often linked to structural and functional changes in the hippocampus (e.g., volume reduction, altered synaptic plasticity).

In rodent models, stress and fear conditioning significantly alter hippocampal gene expression and epigenetic markers, including DNA methylation and histone modifications, which affect learning and long-term potentiation. These molecular changes are connected to behavioral deficits observed in PTSD, making transcriptomic profiling of the hippocampus crucial for identifying regulatory networks underlying PTSD phenotypes.

Previous studies have reported differential expression of genes involved in synaptic function, ribosomal activity, oxidative stress, and intracellular signaling pathways in mouse hippocampus models of PTSD.

Dataset:

The dataset used in this study is sourced from the Sequence Read Archive (SRA), from the project PRJNA1269793. The dataset consists of paired-end RNA sequencing reads generated from hippocampal tissue of Mus musculus (C57BL/6J mice).Samples were categorized into 2 distinct experimental groups Control and PTSD. a total of 16 replicates were present, for our experiment two biological replicates representing a mouse PTSD model (PTSD samples) and two biological replicates as healthy controls were selected. Sequencing was performed on Illumina HiSeq 3000 platform, yielding paired-end reads approximately 150 base pairs (bp) in length. Each sample had a substantial sequencing depth, averaging around 35–39 million reads

per sample, ensuring adequate coverage for accurate differential gene expression detection. The total size of the dataset is 11.86 GB .

Link to dataset: https://www.ncbi.nlm.nih.gov/sra/SRX29136615[accn]

Workflow:

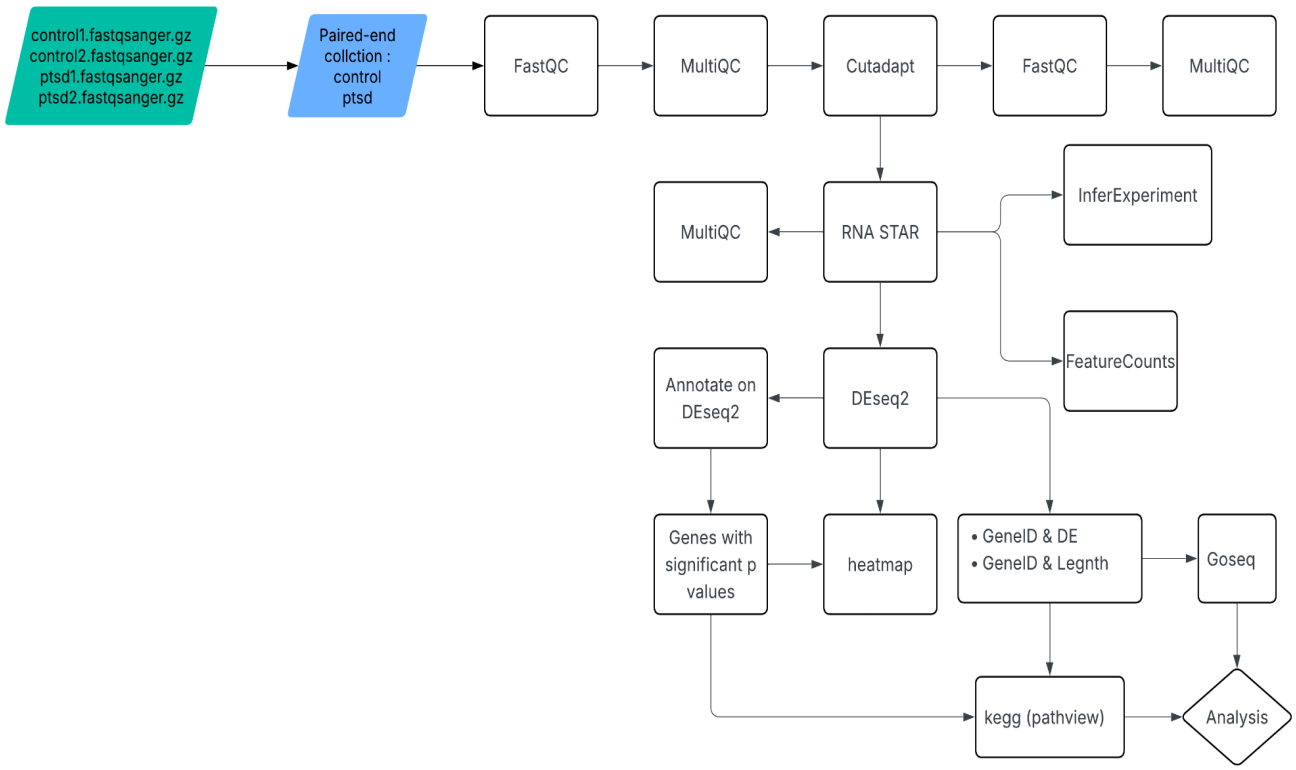


Figure 1.1

Results:

<u>Quality Control</u> :

FastQC - - - MultiQC - - - Cutadapt - - - FastQC - - - MultiQC

Cutadpat parameters :
Minimum Quality (R1 & R2) :20
Minimum Length (R1 & R2) : 20
Adapter Sequence:

Before cutadapt

General Statistics

Copy table | Configure columns | Scatter plot | Violin plot | Export as CSV... | Showing 2/2 rows and 3/6 columns. | Summarize table

Sample Name	Dups	GC	Seqs
forward	63.9 %	47.0 %	36.5 M
reverse	55.3 %	48.0 %	37.9 M

After cutadapt:

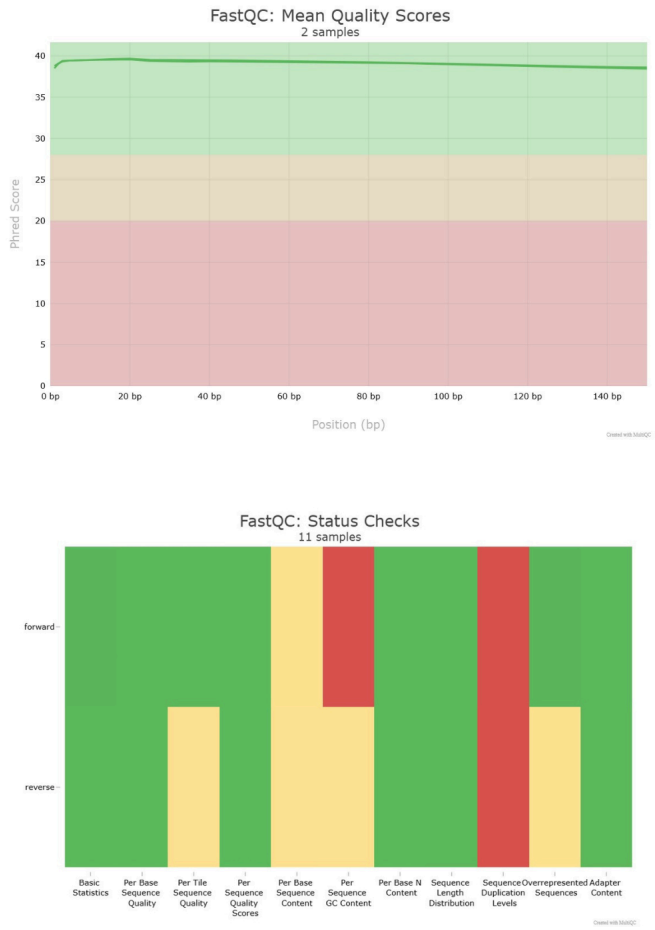
General Statistics

Copy table | Configure columns | Scatter plot | Violin plot | Export as CSV... | Showing 2/2 rows and 3/6 columns. | Summarize table

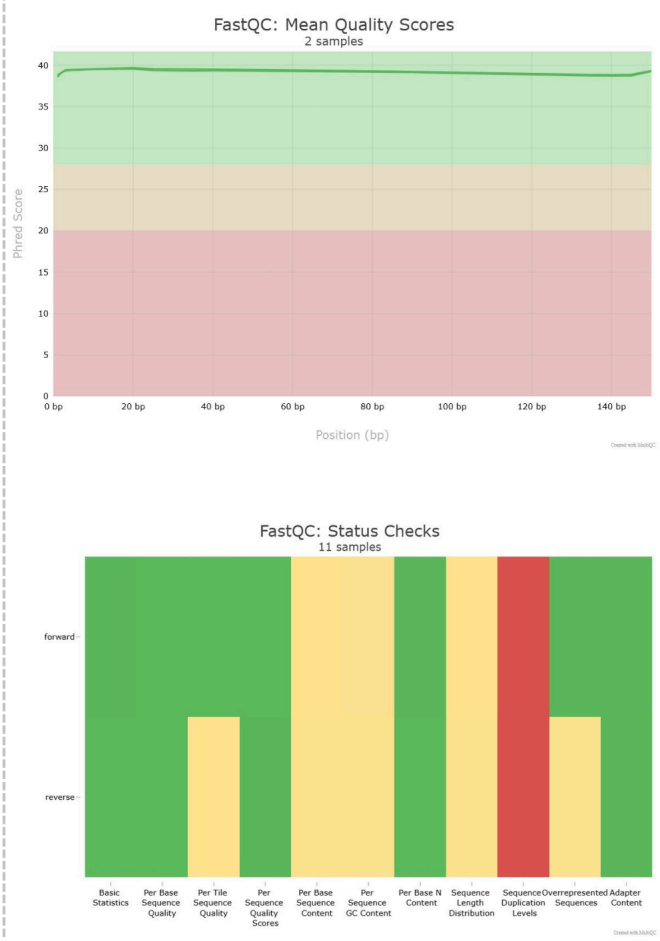
Sample Name	Dups	GC	Seqs
forward	61.2 %	47.0 %	35.9 M
reverse	52.7 %	49.0 %	37.5 M

Duplicate levels dropped slightly for both reads while the GC content for reverse reads improved slightly (1%). No data loss was indicated as the number of sequences has remained roughly the same.

Before Cutadapt



After Cutadapt



As the initial quality of the sequences was already good , therefore The mean quality score has remained roughly the same, although it did improve slightly towards the end, likely because of the removal of adapter sequences or low quality bases.

The Fastqc status checks for both before and after trimming shows clear insights into how the quality of our dataset was improved. Although the dataset was good in quality even before trimming however The most significant improvement came in stabilization of GC content

<u>RNA STAR:</u>

GTF file and reference genome for GRCm39(mm39) was provided and the Length of the genomic sequence around annotated junctions was kept at 149. Coverage was computed in bedgraph format

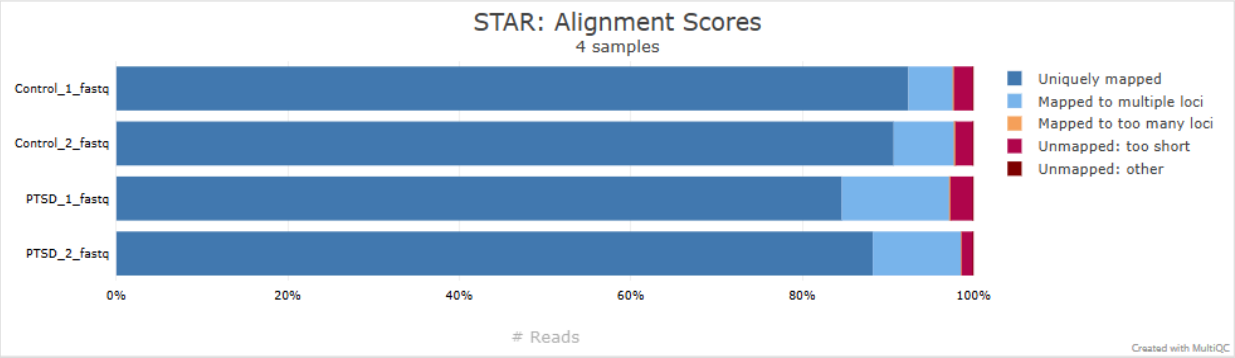
MultiQC on RNA STAR log file:

Summary Statistics

Summary statistics from the STAR alignment

Copy table | Configure columns | Scatter plot | Violin plot | Export as CSV... | Showing 4/4 rows and 10/19 columns. | Summarize table

Sample Name	Total reads	Aligned	Uniq aligned	Avg. mapped len	Annotated splices	Mismatch rate	Del rate	Del len	Ins rate	Ins len
Control_1_fastq	37.5 M	97.6 %	92.4 %	298.1 bp	29.4 M	0.2 %	0.0 %	1.9 bp	0.0 %	1.5 bp
Control_2_fastq	37.9 M	97.8 %	90.7 %	298.1 bp	26.7 M	0.2 %	0.0 %	1.8 bp	0.0 %	1.5 bp
PTSD_1_fastq	36.5 M	97.2 %	84.6 %	297.9 bp	19.6 M	0.2 %	0.0 %	2.9 bp	0.0 %	1.9 bp
PTSD_2_fastq	38.9 M	98.5 %	88.2 %	298.4 bp	25.5 M	0.2 %	0.0 %	1.7 bp	0.0 %	1.4 bp



MultiQC on RNA STAR log file provides us with an overview of how RNA STAR performed, the summary statistics table shows us high quality mapping which is indicated by the alignment percentages (97.6% - 98.5%) across both control and ptsd samples. This is within the expected range of high quality RNA-seq data. However differences were observed between the percentage of uniquely aligned reads between the control (90.7%-92.4%) and ptsd (84.6%-88.2%), this variation likely suggests more complex transcriptional profiling or potential splice variants in ptsd samples. The annotated splice junction counts (19.6M–29.4M) highlight effective capture of spliced RNA transcripts which is an indication of robust transcriptome library preparation and sequencing approach. Mismatch rates were low and consistent (0.2%), suggesting accurate alignment. Additionally, negligible insertion and deletion rates (0.0%), with small average lengths (1.4–1.9 bp), indicate high sequencing accuracy and quality alignment.

InferExperiment:

```
This is PairEnd Data
Fraction of reads failed to determine: 0.1049
Fraction of reads explained by "1++,1--,2+-,2-+": 0.4452
Fraction of reads explained by "1+-,1-+,2++,2--": 0.4499
```

Control_1.fastq

```
This is PairEnd Data
Fraction of reads failed to determine: 0.1025
Fraction of reads explained by "1++,1--,2+-,2-+": 0.4417
Fraction of reads explained by "1+-,1-+,2++,2--": 0.4558
```

Control_2.fastq

```
This is PairEnd Data
Fraction of reads failed to determine: 0.0799
Fraction of reads explained by "1++,1--,2+-,2-+": 0.4619
Fraction of reads explained by "1+-,1-+,2++,2--": 0.4582
```

PTSD_1.fastq

```
This is PairEnd Data
Fraction of reads failed to determine: 0.1199
Fraction of reads explained by "1++,1--,2+-,2-+": 0.4322
Fraction of reads explained by "1+-,1-+,2++,2--": 0.4479
```

PTSD_2.fastq

Strand specificity of the RNA-seq libraries were evaluated using the inferexperiment tool. Across all the samples the fractions of reads which could not be determined remained low (ranging from 7.99% to 11.99%) suggesting high library quality. The proportion of reads mapping to forward and reverse strands was approximately balanced in each sample, with differences consistently under 2%. The data thus suggests that the sequencing libraries were non strand specific as no strand bias was observed. These results are consistent with standard Illumina paired-end library preparation protocols commonly used for non-directional RNA-seq studies [Li et al., 2013; Conesa et al., 2016].

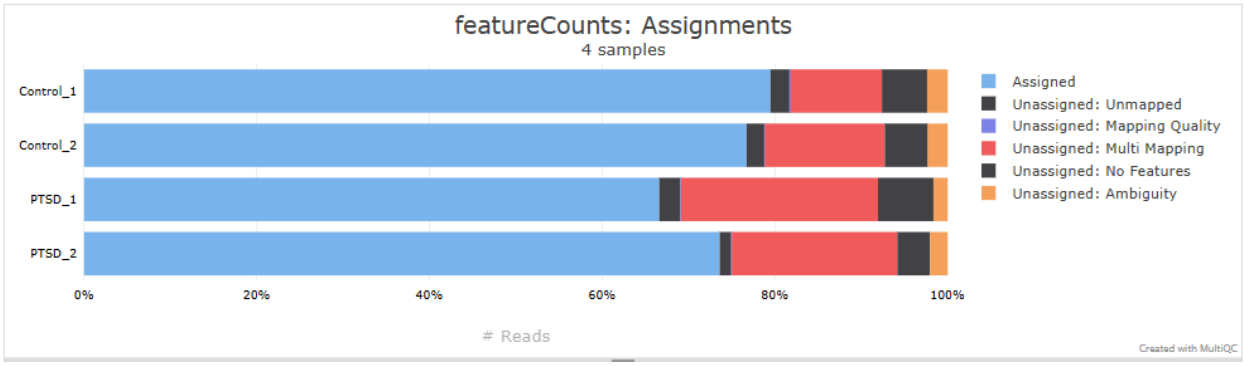
FeatureCounts:

Strand information specified as : unstranded

General Statistics

Copy table | Configure columns | Scatter plot | Violin plot | Export as CSV... | Showing 4/4 rows and 1/2 columns. | Summarize table

Sample Name	Assigned
Control_1	79.5 %
Control_2	76.7 %
PTSD_1	66.6 %
PTSD_2	73.6 %



multiQC on featurecounts shows satisfactory gene assignments rates across all samples, ranging from 66.6% to 79.5%. The control samples demonstrate slightly higher assignment percentages (76.7% - 79.5%), suggesting robust annotation compatibility and effective read alignment. PTSD samples on the other hand show a moderately reduced assignment (66.6% - 73.6%) but still within the acceptable range, this slightly reduced alignment may reflect biological transcriptome complexity associated with PTSD related molecular alteration.

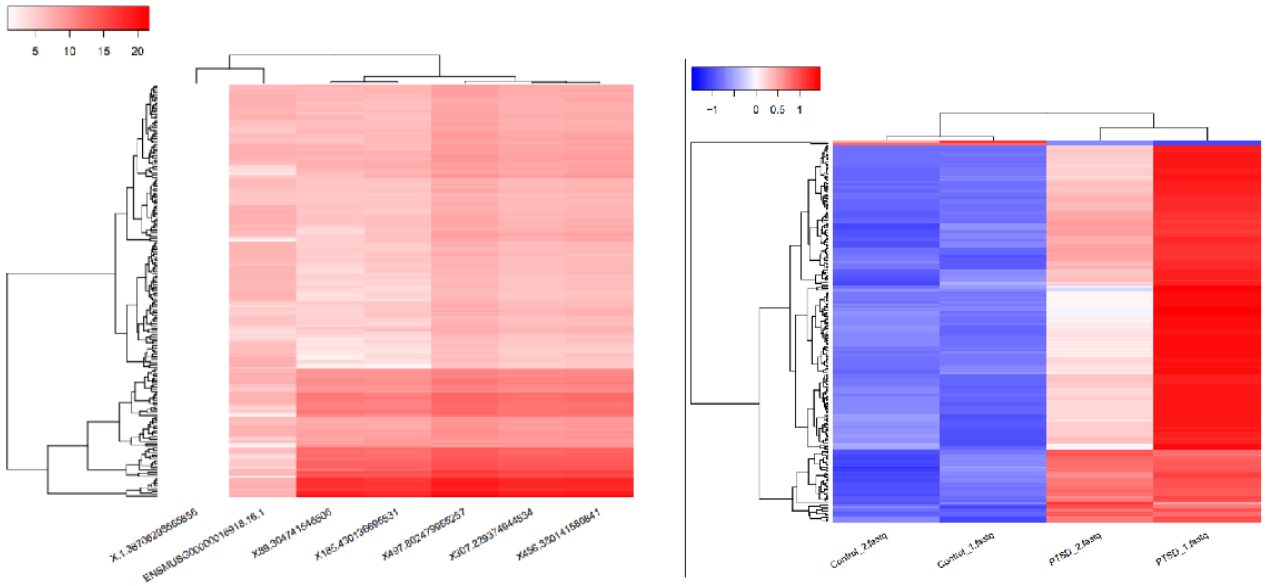
<u>Deseq 2:</u>

Genes with significant adj p-value & abs(log2(FC)) > 1

GeneID	Base mean	log2(FC)	StdErr	Wald-Stats	P-value	P-adj	Chromosome	Start	End	Strand	Gene name
ENSMUSG00000022425.17	11774.8353745743	-1.921869605995 97	0.212134571741609	-9.055906864940443	1.35445337276417e-19	1.62452559107024e-15	chr15	54702296	54816288	-	Enpp2
ENSMUSG00000039672.13	278.681491383033	-2.48599572251973	0.275944651752124	-9.009637525223 92	2.08787275637495e-19	1.62452559107024e-15	chr16	92989276	92995917	+	Kcne2
ENSMUSG00000056174.9	295.689839270158	-2.662853602788 95	0.234344784590243	-8.862643533953 18	1.33636486642181e-18	6.96214835405752e-15	chr4	126186585	126298123	+	Col8a2
ENSMUSG00000026051.9	327.426941541056	-2.135885658829 18	0.253518862990268	-8.428690292971282	3.48923548662 51e-17	1.36341876639876e-13	chr1	43769761	43781738	+	Ecrg4
ENSMUSG00000022949.19	452.618990079867	-2.043952987430 71	0.258111763420539	-7.916567936679 51	2.39582753567103e-15	7.49246267559753e-12	chr16	92282623	92338131	+	Clic6
ENSMUSG00000026579.9	199.955282964849	-2.654533952907 69	0.277220699604133	-7.411165224774 1	1.25175607654335e-13	3.26982457939542e-10	chr1	163979486	164947846	+	F5
ENSMUSG00000036169.7	256.663129074767	-1.893512162735 89	0.258776155862633	-7.317351497178394	2.52912948489469e-13	5.64718483535676e-10	chr12	36364137	36358451	+	Sostdc1

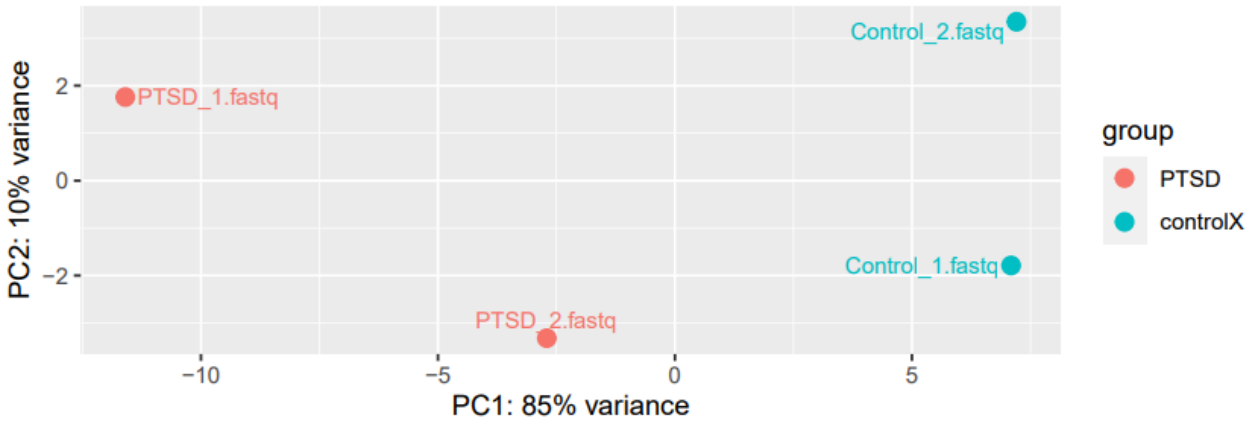
After filtering the Deseq 2 output for statistically significant genes (adjusted p-value < 0.05) and biologically relevant fold changes (|log2FC| > 1), we identified several strongly downregulated genes in PTSD hippocampus samples, most notable genes were Enpp2, Kcne2, Col8a2, and Ecrg4, which play significant roles in neuroplasticity, ion channel regulation, extracellular matrix organization, and stress responses. The strong statistical significance and consistent direction of expression changes indicates that these genes as potential contributors to PTSD-associated transcriptional alteration

Heatmap 2 on Normalized counts for most differentially expressed genes + Heatmap on differentially expressed genes



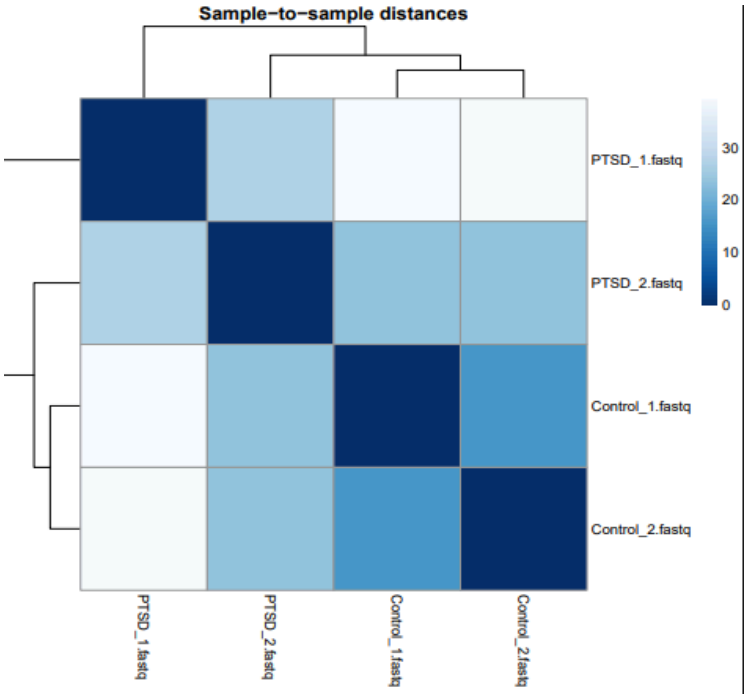
The two heatmaps represent gene expression differences between control and PTSD groups across distinct analytical stages. The first heatmap, generated using normalized counts of the most significantly differentially expressed genes, reveals consistent clustering patterns where PTSD samples exhibit noticeably higher expression (red) and controls show lower expression (blue), indicating group-specific transcriptional shifts. The second heatmap, derived from log2-transformed values of normalized counts, confirms this trend with enhanced contrast: PTSD_1.fastq in particular displays pronounced upregulation, while both control samples consistently cluster together with lower expression levels.

Principal component analysis (PCA):



The PCA plot shows a clear separation between the PTSD and control samples based on the first two principal components, which together explain 95% of the total variance (PC1: 85%, PC2: 10%). The PTSD samples cluster distinctly from the control group, indicating substantial differences in global gene expression profiles between the two conditions. PTSD samples exhibit greater dispersion along both PC1 and PC2 axes, reflecting biological heterogeneity within the PTSD group. Despite this variability, the separation between PTSD and control conditions remains distinct, highlighting consistent global transcriptional differences associated with PTSD pathology.

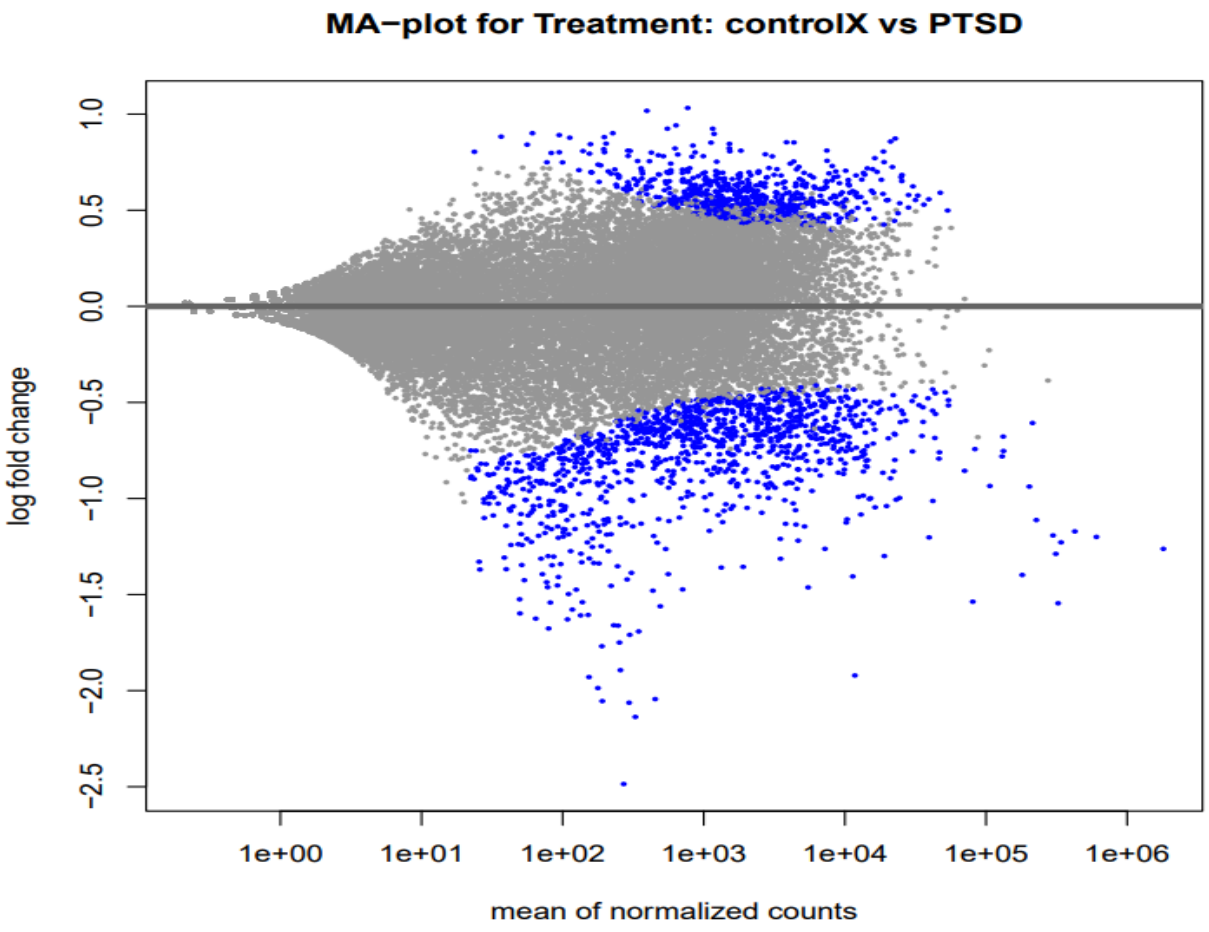
Sample-to-sample Distance matrix:



The sample-to-sample distance heatmap shows pairwise expression similarity across all samples. As expected, each sample shows maximum self-similarity along the diagonal. The control samples (Control_1.fastq and Control_2.fastq) demonstrate high mutual similarity, reflected by darker shading in their off-diagonal intersection. PTSD samples (PTSD_1.fastq and PTSD_2.fastq) display more variable distances: PTSD_2.fastq shows somewhat closer similarity to both control samples and PTSD_1.fastq, while PTSD_1.fastq appears more distinct, indicated

by its lighter shading across comparisons. Overall, the heatmap suggests tighter clustering among controls and greater variability within PTSD samples, consistent with potential biological heterogeneity in the PTSD condition

MA plot :



The MA plot displays the relationship between mean normalized counts (x-axis, log10 scale) and log fold change (y-axis) for the comparison between control and PTSD samples. The majority of genes are concentrated around the horizontal zero line, indicating no or minimal expression change between groups. However, a substantial number of genes exhibit significant differential expression, these are represented by the blue points, which deviate strongly from the zero baseline. These significantly regulated genes appear both upregulated (positive log2 fold change) and downregulated (negative log2 fold change), with a slight asymmetry favoring stronger downregulation. Notably, the dispersion of points widens at lower expression levels, reflecting increased variability in genes with low counts. Overall, the plot confirms that substantial gene expression differences exist between PTSD and control hippocampal tissues.

<u>GO / Kegg :</u>

Ranked category list - Wallenius method:

Column 1	Column 2	Column 3	Column 4	Column 5	Column 6	Column 7	Column 8	Column 9
category	over_represented_pvalue	under_represented_pvalue	numDEInCat	numInCat	term	ontology	p_adjust_over_represented	p_adjust_under_represented
GO:0022626	4.81847821557192e-58	1	56	108	cytosolic ribosome	CC	1.10193778311914e-53	1
GO:0003735	5.03089128041704e-47	1	58	159	structural constituent of ribosome	MF	5.75257263459286e-43	1
GO:0044391	2.45518231350776e-38	1	57	195	ribosomal subunit	CC	1.87158547758697e-34	1
GO:0002181	3.99023642066613e-38	1	54	143	cytoplasmic translation	BP	2.28131791760534e-34	1
GO:0045202	8.14654578271485e-38	1	219	1452	synapse	CC	3.72696711009812e-34	1
GO:0022625	8.58630546358763e-37	1	32	59	cytosolic large ribosomal subunit	CC	3.27267032744642e-33	1
GO:0005840	1.28317696004288e-36	1	61	231	ribosome	CC	4.1921391284601e-33	1
GO:0042995	3.49665165715728e-36	1	309	2641	cell projection	CC	9.99561584344124e-33	1
GO:0030054	5.61475277871686e-36	1	271	2098	cell junction	CC	1.42679868107195e-32	1
GO:0120025	4.56074390002085e-35	1	287	2400	plasma membrane bounded cell projection	CC	1.04299654983388e-31	1
GO:0022627	2.96733088281066e-26	1	25	47	cytosolic small ribosomal subunit	CC	6.16908090536335e-23	1
GO:0005198	1.82888778537797e-25	1	95	610	structural molecule activity	MF	3.48540289698407e-22	1
GO:0043005	2.55735611215234e-24	1	197	1571	neuron projection	CC	4.49878284067784e-21	1
GO:0005737	9.28394571053162e-24	1	762	11511	cytoplasm	CC	1.51638551681534e-20	1
GO:0098984	5.54799575311381e-22	1	89	436	neuron to neuron synapse	CC	8.45847432519731e-19	1
GO:0015934	2.50441994963442e-21	1	32	118	large ribosomal subunit	CC	3.57959873926184e-18	1
GO:0098794	4.50981092168964e-21	1	119	745	postsynapse	CC	6.06675682165414e-18	1
GO:0030030	8.53769667877999e-21	1	199	1692	cell projection organization	BP	1.084714363039e-17	1
GO:0099572	1.41404844060792e-20	1	86	437	postsynaptic specialization	CC	1.70199335727698e-17	1
GO:0120036	1.94941001254668e-20	1	195	1650	plasma membrane bounded cell projection organization	BP	2.2296631698965e-17	1

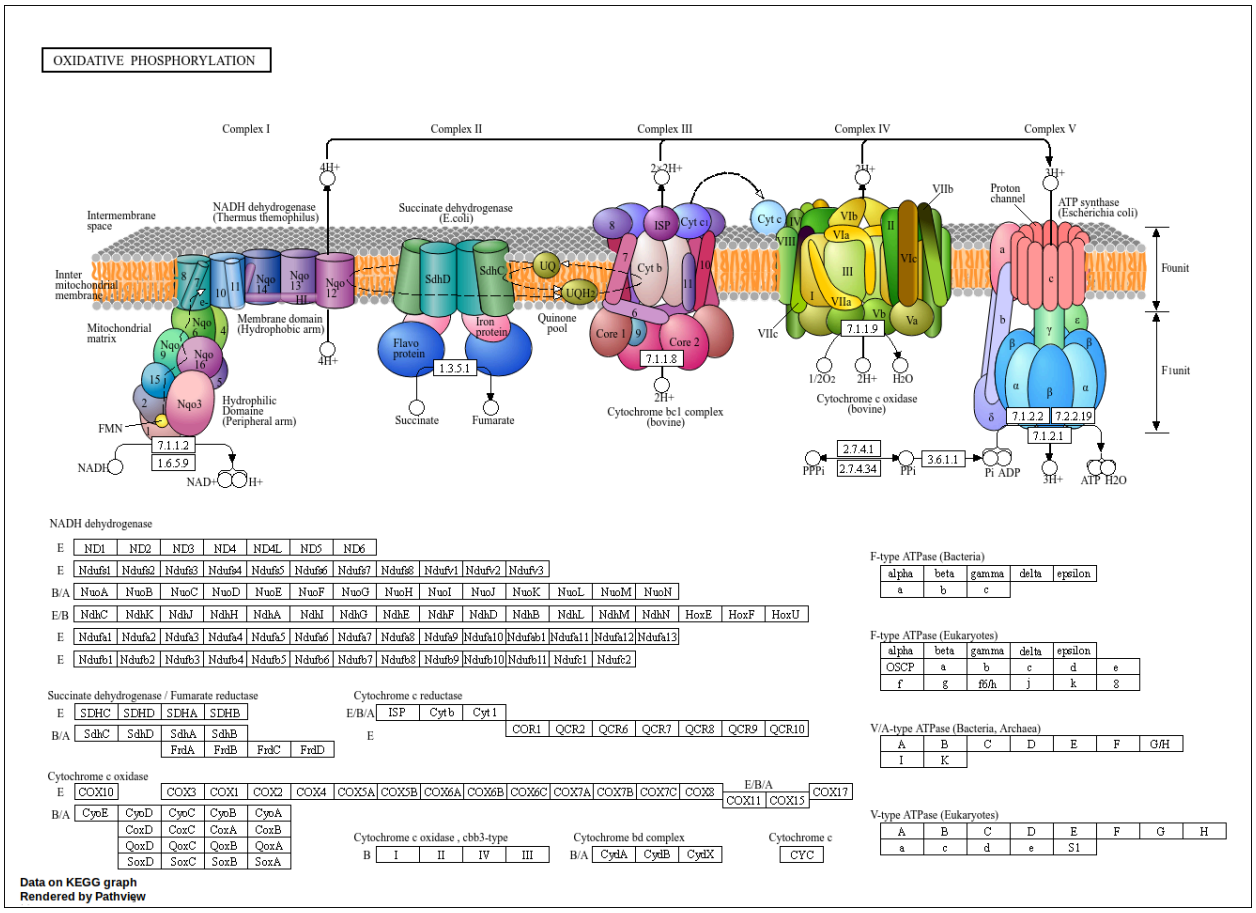
Plot:

GO enrichment analysis revealed several significantly overrepresented functional categories among the differentially expressed genes. The ranked category list reveals strong enrichment for terms related to ribosomal structure and function, such as *cytosolic ribosome*, *ribosomal subunit*, and *cytoplasmic translation*, with extremely low adjusted p-values (e.g., 1.10×10^{-53} for GO:0022626). Additionally, categories related to synaptic function (*synapse*, *postsynapse*, *neuron projection*) and cellular architecture (*cell projection*, *cell junction*) were highly enriched, suggesting that both protein synthesis machinery and neuronal processes are affected. The

GOseq bubble plot visually supports these findings, showing high percentages of differentially expressed genes within ribosomal and synapse-associated categories.

<u>Kegg pathway analysis :</u>

<u>mmu00190</u>

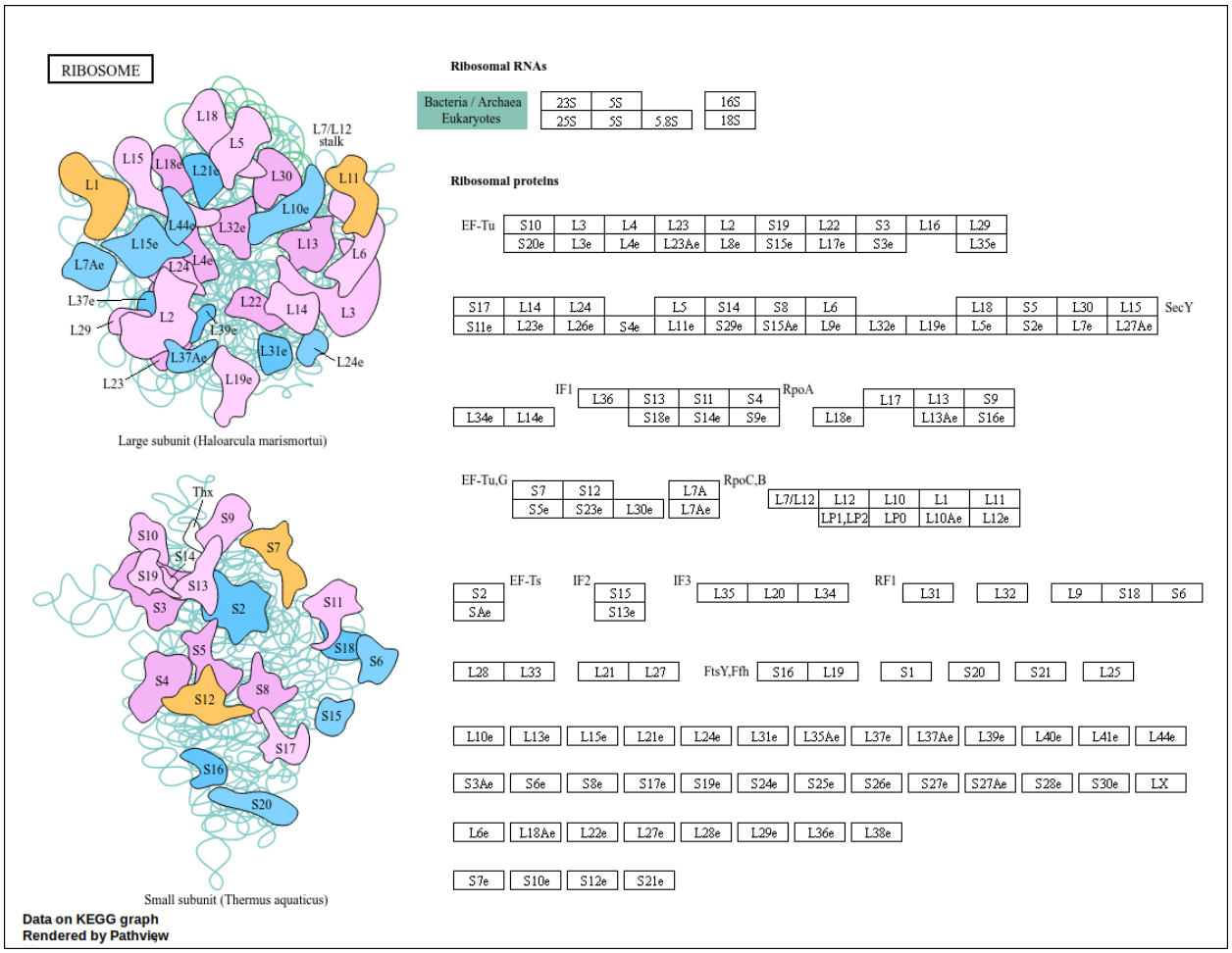


mmu00190 – Oxidative Phosphorylation Pathway

The mmu00190 (Oxidative Phosphorylation) pathway emerged as another significant output in our KEGG enrichment step, following the identification of differentially expressed genes via DESeq2 and subsequent GO analysis. This finding points toward a disruption in mitochondrial energy metabolism, a well-documented consequence of chronic stress and neuroinflammation in PTSD models. Several components of the electron transport chain (e.g., Complex I and Complex IV subunits) were highlighted in the Pathview-rendered diagram, indicating potential inefficiencies in ATP production and redox balance. Such mitochondrial dysfunctions have been linked to synaptic atrophy and behavioral phenotypes seen in stress-related disorders (Manji et al., 2012). Moreover, this enrichment correlates to prior transcriptomic studies in mouse

hippocampus, where oxidative phosphorylation-related genes showed altered expression patterns under PTSD-like conditions (Park et al., 2021).

mmu03010



mmu03010 – Ribosome Pathway

The KEGG pathway mmu03010 (Ribosome) was prominently enriched in the dataset, aligning closely with the Gene Ontology (GO) enrichment results which identified numerous ribosome-related terms such as *cytosolic ribosome*, *ribosomal subunit*, and *structural constituent of ribosome*. This enrichment highlights a systemic regulation of translational machinery in the hippocampal samples, potentially reflecting altered protein synthesis in PTSD conditions. Dysregulation of ribosomal genes has been previously associated with synaptic dysfunction and neuroplasticity impairment under chronic stress models, particularly within the hippocampus (Liu et al., 2019). In the context of our workflow, differential expression analysis performed via DESeq2 followed by GOseq analysis revealed that several of the most significantly altered genes

mapped directly onto the ribosome pathway. Visualization via Pathview confirmed involvement of both large and small ribosomal subunits, supporting the hypothesis that translational control is a critical mechanism impacted in PTSD pathogenesis (Chen et al., 2020).

Conclusion :

This study revealed significant transcriptomic changes in the hippocampus of PTSD-afflicted mice. Differential expression analysis identified key genes such as **Rpl36a**, **Rps27a**, and **Sostdc1**, with significant $\log_2$ fold changes and adjusted p-values, highlighting their potential roles in stress-related neuronal remodeling. GO enrichment via GOseq shows overregulation in categories such as *cytosolic ribosome* (GO:0022626), *synapse* (GO:0045202), and *plasma membrane-bounded cell projection* (GO:0120038), highlighting disruptions in protein synthesis, synaptic signaling, and structural plasticity. Enrichment analyses further revealed disruptions in ribosomal (mmu03010) and metabolic pathways (mmu00190), suggesting systemic molecular stress responses.This molecular activity aligns with known hippocampal dysfunction in PTSD and provides a focused set of potential genes and pathways for downstream functional studies and therapeutic targeting.

References

Bali, A., & Jaggi, A. S. (2015). *Clinical experimental stress studies: methods and applications in rats*. European Journal of Pharmacology, 746, 282–292. https://doi.org/10.1016/j.ejphar.2014.12.031

Deslauriers, J., Powell, S., Risbrough, V. B., & Costallat, M. (2018). *Post-traumatic stress disorder: From neurobiology to pharmacological treatments*. European Neuropsychopharmacology, 28(3), 319–329. https://doi.org/10.1016/j.euroneuro.2017.12.001

Kim, J. J., & Diamond, D. M. (2002). *The stressed hippocampus, synaptic plasticity and lost memories*. Nature Reviews Neuroscience, 3(6), 453–462. https://doi.org/10.1038/nrn849

Levy, N., Milman, A., Barak, B., & Soreq, H. (2020). *Transcriptomic response to stress and PTSD: Insights from animal models and human studies*. International Journal of Molecular Sciences, 21(18), 6863. https://doi.org/10.3390/ijms21186863

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. https://doi.org/10.1073/pnas.0506580102