PNAS

Main Manuscript for

RNA-seq Analysis Reveals Molecular Mechanisms of ADSC Stemness in Rat Adipose Tissue.

Mohit Poudel*[1], Zhihua Hua[2]

[1]Department of Biological Sciences, Alvarado-Serrano Lab, Ohio University, Athens, OH, 45701
[2]Department of Environment and Plant Biology, Hua Lab, Ohio University, Athens, OH, 45701

Email: mp067823@ohio.edu

PNAS strongly encourages authors to supply an ORCID identifier for each author. Do not include ORCIDs in the manuscript file; individual authors must link their ORCID account to their PNAS account at www.pnascentral.org. For proper authentication, authors must provide their ORCID at submission and are not permitted to add ORCIDs on proofs.

Author Contributions: Mohit Poudel worked for manuscript preparation and Zhihua Hua provided R scripts for data analysis.

Competing Interest Statement: No competing interests exits.

Classification: Biological Sciences: Biophysics and Computational Biology.

Keywords: RNA sequencing, Adipose-derived stromal cells (ADSCs), hypergeometric test

This PDF file includes:

Abstract

Adipose-derived stromal cells (ADSCs) hold immense potential for tissue repair and therapeutic interventions due to their unique properties. In this study, we investigated the gene expression profiles of ADSCs transplanted into the inguinal adipose tissues of nude rats to assess their stemness properties. The method involved downloading raw reads from Green Fluorescent Protein (GFP) transgenic rats, followed by quality control using FastQC, adapter trimming, alignment, and RNA-seq data analysis using various R packages. Our analysis revealed significant enrichments in Gene Ontology (GO) Biological Process terms, including DNA replication processes, immune system functions, and hemopoiesis, among genes with notable expression changes. Cellular Component ontology analysis highlighted alterations in chromosomal organization, extracellular matrix components, and DNA replication machinery. Molecular Function ontology analysis identified enrichments in DNA-related activities, transcription regulation, signaling regulation, and DNA-protein interactions. These findings contribute valuable insights into optimizing ADSC culture conditions in vitro and understanding their therapeutic potential in tissue regeneration and disease treatments.

Significance Statement

This study holds significant implications for regenerative medicine and therapeutic interventions utilizing adipose-derived stromal cells (ADSCs). By analyzing the gene expression profiles of ADSCs transplanted into rodent adipose tissues, we unveil critical insights into the molecular mechanisms governing ADSC stemness properties. Our comprehensive analysis, integrating RNA-seq data analysis methodologies with bioinformatics tools, elucidates key biological processes such as DNA replication, immune system regulation, and cellular organization. These findings not only contribute to refining ADSC culture conditions for enhanced therapeutic efficacy but also pave the way for targeted interventions in tissue engineering and disease treatments, leveraging the regenerative potential of ADSCs in clinical applications.

Main Text

Introduction

Adipose-derived stromal cells (ADSCs) are remarkable cells with the potential to repair tissues, making them increasingly important in medical treatments (1). Being abundant in adults, adipose tissues are involved in obesity and associated metabolic disorders (2). The mesenchymal cells derived from those tissues possess unique qualities such as immunomodulation, angiogenesis, and stimulation of extracellular matrix (3). Due to their versatility in cell types, and the ability to release beneficial substances, ADSCs have shown promise in treating challenging health conditions like graft-versus-host disease (4), Crohn's disease (5), heart attacks (6), lung ailments (7), and issues with bones and cartilage (8). Moreover, ADSCs are believed to exhibit enhanced efficacy in wound healing and tissue regeneration, particularly for skin, bone, and cartilage (9). Their accessibility, abundance, and minimally invasive collection make them highly favorable compared to similar cells (10). Nonetheless, the therapeutic advantages of ADSCs predominantly rely on their high cell number and high quality (11). However, ensuring their long-term vitality during laboratory cultivation presents challenges as they naturally lose some functions over time (12, 13). There is a notable knowledge gap about the

optimal lab conditions that mimic the body's environment to maintain ADSCs' potency (13), creating an urgency to explore the optimal culture conditions for ADSCs. Original cellular microenvironment in the organism is the optimal conditions for cell which can be mimicked in the culture to maintain stemness. Additionally, investigations into reintroducing ADSCs into animal models aim to bolster their resilience and therapeutic effectiveness (11, 14).

In tissue engineering, the transplantation of cells into living organisms (in vivo cell transplantation) is a crucial step for preserving the unique characteristics of various cell groups. This process creates an environment that mimics natural stimuli, aiding in cell maintenance. For instance, studies have effectively used the omentum and subcutaneous tissues of nude rodents to culture epithelial cells, chondrocytes, and other organs/tissues (15). These models promote enhanced cell communication with their surroundings, resulting in graft proliferation, angiogenesis, and functional improvements(16) . Notably, transplanting bone marrow MSCs under the skin has shown their ability to adapt to the local environment, form aggregates that stimulate blood vessel growth, and regulate circulating cytokines (17). Building on these findings, we explored whether in vivo transplantation culture could help maintain or enhance the stemness of ADSCs (adipose-derived stem cells). Our investigation focused on analyzing the RNA sequence data after transplanting ADSCs into the inguinal adipose tissues of nude rodents to assess their stemness properties. Particularly, we conducted a comprehensive analysis of the ADSCs' gene expression profiles under control and treatment culture conditions, revealing potential molecular mechanisms at play.

To perform the analysis, nucleotide bases of RNA obtained by sequencing RNA of Green Fluorescent Protein (GFP) transgenic rat (*Rattus norvegicus)* are analyzed using bioinformatics pipelines starting with Quality Control using FastQC (18), followed by trimming of adaptors and alignment. Feature count data obtained after alignment were then fed into different R packages for downstream analyses. Specifically, Gene ontology annotation and Gene Ontology enrichment analysis were performed using statistical tools like hypergeometric testing (19). The findings contribute to optimizing ADSC culture in vitro, leading to improved properties and enhanced understanding of their therapeutic potential.

Results

Initially, we obtained the feature count data with six columns, each representing a replicate (3 control and 3 treatment) and 21,924 rows, each representing a gene. Average feature count was lesser in all treatment replicates as compared to their corresponding control replicates; however significant number of outliers are observed in each category (See Figure 1). However, the total feature counts of the replicates are of almost same number, as illustrated by the histogram in figure 2.

Strong correlation is observed between the log of all feature counts data. For instance, log of feature counts between control 1 and control 2 replicates exhibits strong positive correlation (Figure 3), which is also true for the linear relation between control 1 and treatment 2 (Figure 4). We obtained 14455 rows after fitting with the formatted matrix as required by DESeq2 and filtering for the rows with less than 4 counts. PCA plot between all the replicates (Figure 5) reveals distinction between control and replicates, as controls are towards the left of the plot while treatments are towards right edge. Higher variation

is observed between the treatments as compared to controls, however Y-axis accounts for only 15% variation, which is much lesser than the variation explained by X-axis (ie: 68%).

The heatmap in figure 6 exhibits significant differentiation in gene expression among different genes of all replicates. Most of the genes in control are upregulated, whereas most of the genes are downregulated in treatment replicates. The proportion of unregulated genes is minimal. Likewise, a MA plot (Figure 7) is obtained with log fold change in Y-axis and mean of normalized counts in X-axis. This plot illustrates higher variance of log fold change in lower mean of normalized counts. The variance of log fold change is reduced as the mean of normalized counts goes increasing.

According to the wald test (table 1), gene ENSRNOG00000060879 has a log2 fold change of -0.819, but the p-value (0.055) is greater than the typical significance threshold of 0.05 (5%). The adjusted p-value (p-adjusted) is also greater than 0.05, indicating that this result is not statistically significant after correcting for multiple testing. Similarly, for the other genes listed in table 1, none of them have p-values or adjusted p-values below 0.05, suggesting that the differences in expression between the treatment group and the control group are not statistically significant for these genes. However, with the significance level of 0.1, 8% (1151) of the total genes are found to be significantly upregulated, whereas 6.6 % (958) of the total genes are found to be downregulated. We also observed 0.15 % (21) of the outliers from the wald test. Similarly, table 2 illustrates the result of wald test with p-adjusted value < 0.1 and absolute value of log2fold change also <1. Using the same data, we plotted a scatter plot (figure 8) that shows statistical significance (log transformed p-adjusted values) in vertical axis versus magnitude of change (log2 fold change) in the horizontal axis. It demonstrates that the number of upregulated genes is higher than downregulated genes as a higher number of dots are on the right side as compared to the left side.

Lastly, the scatter plot in figure 9 illustrated the distribution of the normalized counts of top genes in both control and treatment categories revealing higher variation in the counts between two groups, ranging from 0.5 in control to around 400 in treatment replicates. This result indicates the variation in the expression of cells due to the treatment applied.

Discussion

The hypergeometric test results as demonstrated in table 3 reveal significant enrichments in specific Gene Ontology (GO) Biological Process terms among the selected genes with significant log2 fold changes and adjusted p-values. Firstly, genes associated with DNA replication processes, as indicated by terms such as "DNA replication" (GO:0006260) and "DNA-templated DNA replication" (GO:0006261), show remarkably significant enrichments, suggesting a strong upregulation of DNA replication-related genes. Additionally, the term "DNA metabolic process" (GO:0006259) also exhibits significant enrichment, further emphasizing the heightened activity in DNA-related biological processes. Moving beyond DNA processes, the results also highlight enrichments in immune system-related processes. The GO term "immune system process" (GO:0002376) shows significant enrichment, indicating an increased expression of genes involved in immune system functions. This finding aligns with the

broader context of the study, suggesting potential interactions or regulatory mechanisms between DNA processes and immune responses.

Furthermore, the term "hemopoiesis" (GO:0030097) is significantly enriched, implying an enhanced activity in blood cell formation-related genes. This observation could be of particular interest as our study involves aspects related to hematopoiesis or blood cell development. Additionally, genes involved in regulating immune system processes, represented by the term "regulation of immune system process" (GO:0002682), also show significant enrichment, highlighting the dynamic regulatory mechanisms influencing immune responses. In summary, the hypergeometric test results indicate significant enrichments in DNA replication, DNA metabolic processes, immune system functions, and hemopoiesis among the genes with notable changes in expression levels, providing valuable insights into the biological processes affected in our RNA-seq dataset.

The hypergeometric test results with the Cellular Component ontology as tabulated in table 4, reveal significant enrichments in various cellular components among the genes with significant log2 fold changes and adjusted p-values. Firstly, the term "chromosome" (GO:0005694) shows a highly significant enrichment, indicating an increased presence of genes associated with chromosomal structures among the upregulated genes. This suggests potential alterations or activities related to chromosomal organization or function. Additionally, genes related to the "extracellular region" (GO:0005576) exhibit significant enrichment, pointing towards changes in extracellular matrix components or secreted proteins among the upregulated genes. This finding is complemented by the enrichment of terms like "external encapsulating structure" (GO:0030312) and "extracellular matrix" (GO:0031012), further highlighting alterations in external cellular components and extracellular matrix-related genes.

Moreover, the "MCM complex" (GO:0042555), a key player in DNA replication, shows significant enrichment, suggesting heightened activity or regulation of DNA replication processes among the upregulated genes. This finding underscore potential alterations in DNA replication machinery or processes within the cellular context. Other enriched terms include "extracellular space" (GO:0005615), indicating changes in genes associated with extracellular processes among the upregulated genes. Collectively, these results provide valuable insights into the cellular structures and processes affected in our RNA-seq dataset, including chromosomal organization, extracellular signaling, DNA replication, and cellular organization within the extracellular space.

The hypergeometric test results focusing on Molecular Function ontology ("MF") in table 5 highlight significant enrichments in specific molecular functions among the genes with significant log2 fold changes and adjusted p-values. Notably, terms related to DNA-related activities stand out prominently. The term "DNA helicase activity" (GO:0003678) exhibits a highly significant enrichment, indicating an increased presence of genes associated with DNA helicase functions among the upregulated genes. This suggests potential heightened activity or regulation of DNA unwinding processes. Similarly, the term "DNA binding" (GO:0003677) shows significant enrichment, reflecting alterations in genes involved in DNA binding activities among the upregulated genes. This finding aligns with the broader context of DNA-related processes being influenced within the cellular environment. Moreover, terms like "ATP-dependent activity, acting on DNA" (GO:0008094) and "catalytic activity, acting on DNA" (GO:0140097) also exhibit significant enrichments, indicating potential changes in enzymatic activities related to

DNA processes among the upregulated genes. These terms encompass functions such as ATP-dependent actions and catalytic activities specifically targeting DNA molecules.

Furthermore, specific helicase activities, such as "helicase activity" (GO:0004386) and "single-stranded DNA helicase activity" (GO:0017116), show significant enrichments, suggesting increased activities in DNA unwinding processes and single-stranded DNA manipulation among the upregulated genes. Additionally, terms related to DNA-binding transcription activator activities, such as "DNA-binding transcription activator activity" (GO:0001216) and "DNA-binding transcription activator activity, RNA polymerase II-specific" (GO:0001228), exhibit significant enrichments, indicating potential changes in gene regulatory processes involving DNA-binding transcription factors. Lastly, terms like "signaling receptor regulator activity" (GO:0030545) and "single-stranded DNA binding" (GO:0003697) also show significant enrichments, pointing towards alterations in regulatory and binding activities associated with signaling pathways and single-stranded DNA interactions among the upregulated genes.

In summary, our results reveal significant enrichments in various molecular functions related to DNA processes, transcription regulation, signaling regulation, and DNA-protein interactions among the genes with notable changes in expression levels. However, it's important to note a limitation of our study, which is limited computational power. Due to this limitation, we could not assess the differential gene expression of multiple treatments simultaneously, potentially affecting the comprehensiveness of our analysis.

For future directions, authors should aim for a broader approach, such as evaluating multiple treatment levels simultaneously to gain a more comprehensive understanding of gene expression changes under different conditions. Additionally, incorporating other bioinformatics analyses, such as KEGG pathway enrichment analysis, would provide a

Materials and Methods

This study uses RNA sequence data utilized in a different study aimed at quantifying gene expression in different treatment levels (20). The original author (20) has used GFP-transgenic rats (female, 6-8 weeks, weighting 130-150 g) obtained from the Experimental Animal Center of Daping Hospital of Army Medical University (Chongqing, China). As control, 3 replicates of ADSCs were directly collected for further investigation, whereas 3 replicates were resuspended in PBS at concentration of 1*10^7 cells/ml, followed by injection of 100 ml of each cell suspension into bilateral inguinal adipose tissue of nude rats. The adipose tissues were separated, minced and digested with 0.075 % collagen II solution for 20 min at 37℃ (20). As reproducing all RNA extraction and resequencing was beyond the scope of this study, we bypassed the wet lab procedures by downloading the sequenced files submitted by authors as BioProject: PRJNA1088406 (SRA: SRP495408).

Although the study (20) have used 9 samples in total, we selected 6 samples (3 control and 3 treatment) due to computational limitations. The SRA accession IDs of the samples studied in this sample are: SRR28352479, SRR28352480, SRR28352481, SRR28352485, SRR28352486, SRR28352487, out of which first three represent control and other three represent treatment. Raw fastq data of all 6 samples were obtained using fastq-dump tool of SRA-toolkit version 3.1.0 (21). The quality of

raw fastq files were assessed using FASTQC tool , and impurities such as adapters and short-unwanted reads were trimmed using trimmomatic (22). To align the trimmed files, genome index dataset of rat (*Rattus norvegicus)* was created in STAR (23); using genome and annotation file of rat downloaded from Ensemble database (24). Followed by the alignment, expression count of each gene was performed by installing subprogram featureCounts (25) inside conda environment. The text file obtained from feature count was preprocessed using Microsoft Excel and saved as a csv file. Each row of the feature count file represents a single gene, whereas each column represents the counts of expressed gene normalized as FPKM (Fragments per Kilobase per million mapped fragments).

By utilizing the dplyr package in RStudio (26), the csv file was restructured in the format suitable for downstream analysis. This step involved procedures, such as renaming the columns as control1, control2, control3, treatment1, treatment2 and treatment3. The count data was summarized and plotted for visualization. As data from different cells give biased results if compared directly, they were normalized by calculating size factors via DESeq2 package of R (27). To obtain PCA plot, custom R functions dist2 and hclust2 were created. The dist2 function takes a dataframe as input, calculates correlation coefficient of the transposed matrix, subtracts each correlation value from 1, and then converts the resulting matrix into a distance matrix using Pearson correlation as the distance metric. Similarly, hclust2 function was created to perform hierarchical clustering using complete linkage method.

Groups with similar values in expression were created by ordering the rows based on their standard deviation in descending order. The cutoff values were calculated as median ± twice the standard deviation. All the outliers (values beyond cutoff) were replaced by the cut-off values and saved into a new matrix. The matrix was visualized with rainbow color palette resulting in the heatmap of the matrix obtained. The function ‘results()’ of R library DESeq2 was used on the dataset to obtain statistical information such as log2fold change, p-values and adjusted p-values. This step yielded a concise summary of the differential expression analysis results, aiding in identification and prioritization of the genes that show significant changes in expression between experimental conditions. Additionally, MA plot was generated using plotMA function of DESeq2 package, with the result obtained by setting p-value = 0.01. Furthermore, a volcano plot was generated from the same dataset, using log2Fold change as X-axis and -log10 of adjusted p-values as Y-axis. To obtain the top genes, the data were sorted in descending order of the absolute values of the log2fold change of each row. Normalized count of all top genes in each replicate (3 control and 3 treatment) were then plotted as scatter plot to visualize the difference in count of the most expressed genes among the replicates.

For further analysis, two R-packages: AnnotationDbi (28) and org.Rn.eg.db – organismal database for rat(29) were installed from R - bioconductor. We then

applied the mapIds function of org.Rn.eg.db package to map Ensembl Identifiers to Entrez Gene Identifiers and stored the mapping results into a new column 'ENTREZ' in the original dataframe. This step allowed us to annotate gene identifier from one type (Ensembl) to another type (Entrez Gene). Similarly, Ensemble Ids in the dataframe were mapped to the gene symbols of organism database.

Additionally, we installed three more commonly used R libraries –GO.db, GOstats (30) and gage (31) for gene ontology annotation and enrichment analyses (32, 33). For these analyses, we subset the dataframe by selecting only those rows (genes) which have adjusted p-values less than 0.05. Out of the selected genes, any gene that has log2fold change value more than 1 was considered as significantly upregulated, whereas those having log2fold change value less than -1 were considered significantly downregulated. Furthermore, hypergeometric tests were performed separately for upregulated and downregulated genes using hyperGtest function from GOstats package. The result of GO enrichment analysis was stored in a variable and summarized for interpretation. Moreover, Gene ontology category was obtained, using ontology function for the parameters for GO enrichment analysis. We set the code to focus on the Cellular component (CC) aspect of Gene Ontology. The enrichment analysis result obtained was stored in the variable upCC and summarized for interpretation. Similarly, Gene Ontology (GO) enrichment analysis was performed focusing on Molecular Function (MF) and summary of the results was stored along with the statistical information, which are presented and discussed above.

Acknowledgments

We extend our heartfelt gratitude to Michael Cooney and Liam Speakman for their invaluable assistance and unwavering support in writing R-scripts required throughout the preparation of this manuscript.

References

1. J. M. Gimble, *et al.*, Adipose-derived stromal/stem cells. *Organogenesis* **9**, 3–10 (2013).

2. L. Casteilla, V. Planat-Benard, P. Laharrague, B. Cousin, Adipose-derived stromal cells: Their identity and uses in clinical trials, an update. *World J. Stem Cells* **3**, 25–33 (2011).

3. R. H. Søndergaard, *et al.*, Adipose-derived stromal cells increase the formation of collagens through paracrine and juxtacrine mechanisms in a fibroblast co-culture model utilizing macromolecular crowding. *Stem Cell Res. Ther.* **13**, 250 (2022).

4. K. Zhao, *et al.*, Mesenchymal stromal cells plus basiliximab, calcineurin inhibitor as treatment of steroid-resistant acute graft-versus-host disease: a multicenter, randomized, phase 3, open-label trial. *J. Hematol. Oncol.J Hematol Oncol* **15**, 22 (2022).

5. C. Zhou, *et al.*, Autologous adipose-derived stem cells for the treatment of Crohn's fistula-in-ano: an open-label, controlled trial. *Stem Cell Res. Ther.* **11**, 124 (2020).

6. K. Bruun, *et al.*, Therapeutic applications of adipose-derived stem cells in cardiovascular disease. *Am. J. Stem Cells* **7**, 94–103 (2018).

7. J. Gao, *et al.*, Adipose-derived stem cells therapy effectively attenuates PM2.5-induced lung injury. *Stem Cell Res. Ther.* **12**, 355 (2021).

8. Single-Cell Profiles and Clinically Useful Properties of Human Mesenchymal Stem Cells of Adipose and Bone Marrow Origin - PubMed. Available at: https://pubmed.ncbi.nlm.nih.gov/31100005/ [Accessed 27 April 2024].

9. L. Mazini, L. Rochette, M. Amine, G. Malka, Regenerative Capacity of Adipose Derived Stem Cells (ADSCs), Comparison with Mesenchymal Stem Cells (MSCs). *Int. J. Mol. Sci.* **20**, 2523 (2019).

10. Y. Zhu, *et al.*, Adipose-derived stem cell: A better stem cell than BMSC. *Cell Res.* **18**, S165–S165 (2008).

11. C. Gong, *et al.*, Infusion of two-dose mesenchymal stem cells is more effective than a single dose in a dilated cardiomyopathy rat model by upregulating indoleamine 2,3-dioxygenase expression. *Stem Cell Res. Ther.* **13**, 409 (2022).

12. S. Wang, *et al.*, Effects of long-term culture on the biological characteristics and RNA profiles of human bone-marrow-derived mesenchymal stem cells. *Mol. Ther. Nucleic Acids* **26**, 557–574 (2021).

13. Y.-D. Wu, *et al.*, Effects of storage culture media, temperature and duration on human adipose-derived stem cell viability for clinical use. *Mol. Med. Rep.* **19**, 2189–2201 (2019).

14. J. Su, *et al.*, Transplantation of adipose-derived stem cells combined with collagen scaffolds restores ovarian function in a rat model of premature ovarian insufficiency. *Hum. Reprod.* **31**, 1075–1086 (2016).

15. Use of omentum as an in vivo cell culture system in tissue engineering - PubMed. Available at: https://pubmed.ncbi.nlm.nih.gov/15497386/ [Accessed 27 April 2024].

16. V. Di Nicola, Omentum a powerful biological source in regenerative surgery. *Regen. Ther.* **11**, 182–191 (2019).

17. M. A. Ghoneim, A. F. Refaie, B. L. Elbassiouny, M. M. Gabr, M. M. Zakaria, From Mesenchymal Stromal/Stem Cells to Insulin-Producing Cells: Progress and Challenges. *Stem Cell Rev. Rep.* **16**, 1156–1172 (2020).

18. Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. Available at: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ [Accessed 28 April 2024].

19. S. Falcon, R. Gentleman, “Hypergeometric Testing Used for Gene Set Enrichment Analysis” in *Bioconductor Case Studies*, F. Hahne, W. Huber, R. Gentleman, S. Falcon, Eds. (Springer, 2008), pp. 207–220.

20. C. Wang, *et al.*, Effect of in vivo culture conditions on the proliferation and differentiation of rat adipose-derived stromal cells. *Mech. Ageing Dev.* 111935 (2024). https://doi.org/10.1016/j.mad.2024.111935.

21. ncbi/sra-tools. (2024). Deposited 28 April 2024.

22. A. M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* **30**, 2114–2120 (2014).

23. A. Dobin, *et al.*, STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15–21 (2013).

24. F. J. Martin, *et al.*, Ensembl 2023. *Nucleic Acids Res.* **51**, D933–D941 (2023).

25. Y. Liao, G. K. Smyth, W. Shi, featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics* **30**, 923–930 (2014).

26. H. Wicham, R. Fracois, L. Henry, K. Muller, D. Vaughan, dplyr: A Grammar of Data Manipulation. (2023).

27. M. I. Love, W. Huber, S. Anders, Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).

28. H. Pagès, M. Carlson, S. Falcon, N. Li, AnnotationDbi: Manipulation of SQLite-based annotations in Bioconductor. *R Package Version* **1** (2021).

29. S. Canzler, J. Hackermüller, multiGSEA: a GSEA-based pathway enrichment analysis for multi-omics data. *BMC Bioinformatics* **21**, 561 (2020).

30. S. Falcon, R. Gentleman, Using GOstats to test gene lists for GO term association. *Bioinformatics* **23**, 257–258 (2007).

31. W. Luo, M. S. Friedman, K. Shedden, K. D. Hankenson, P. J. Woolf, GAGE: generally applicable gene set enrichment for pathway analysis. *BMC Bioinformatics* **10**, 161 (2009).

32. M. Ashburner, *et al.*, Gene Ontology: tool for the unification of biology. *Nat. Genet.* **25**, 25–29 (2000).

33. The Gene Ontology Consortium, *et al.*, The Gene Ontology knowledgebase in 2023. *Genetics* **224**, iyad031 (2023).

Figures and Tables

Figure 1.

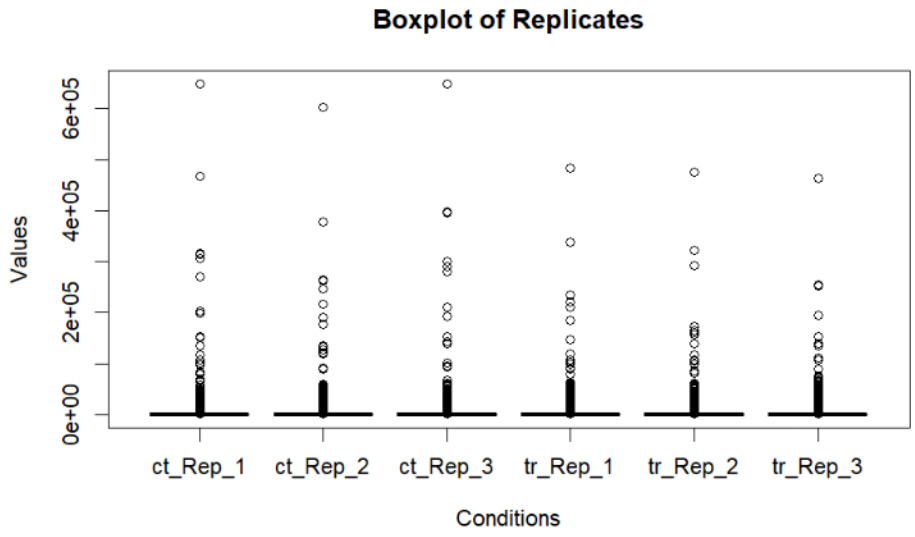
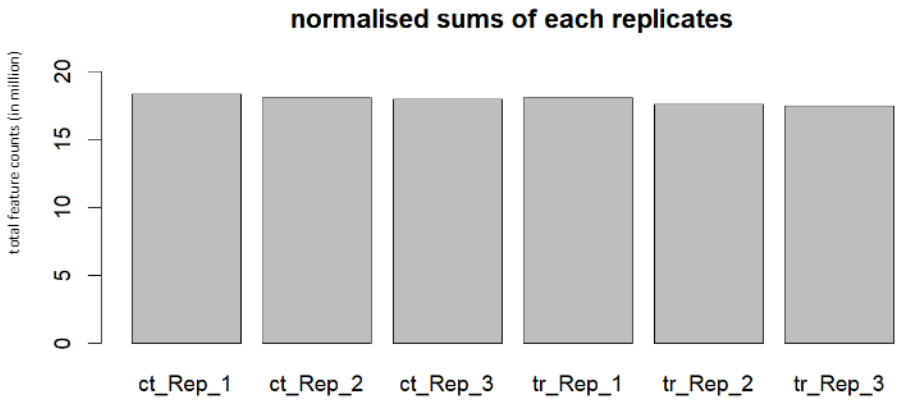


Figure 1: Distribution of feature count values of each replicate. Each dot outside the cluster represents an outlier.

386 **Figure 2.**



397 Figure 2: frequency of total feature count for each replicate scaled in million.

398

Figure 3.

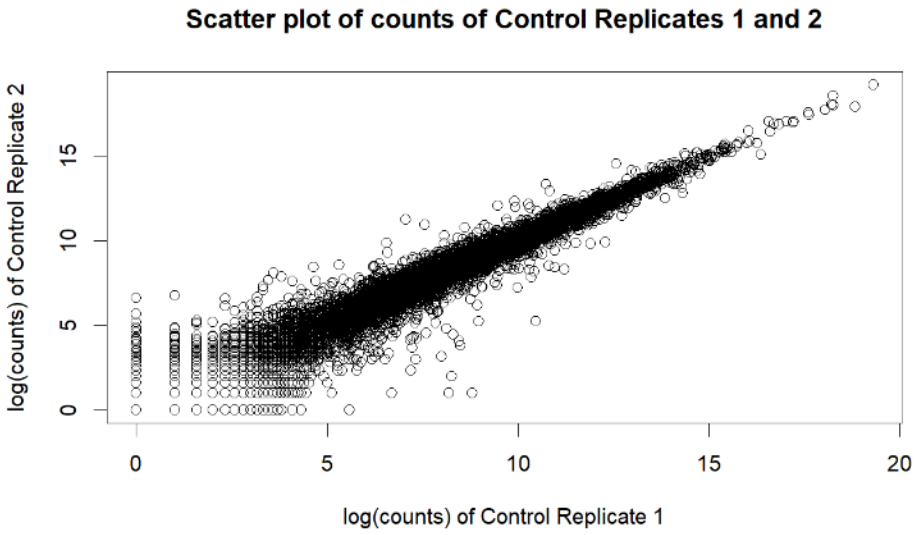


Figure 3: Scatter plot illustrating feature count values of control 1 and control 2.

Figure 4.

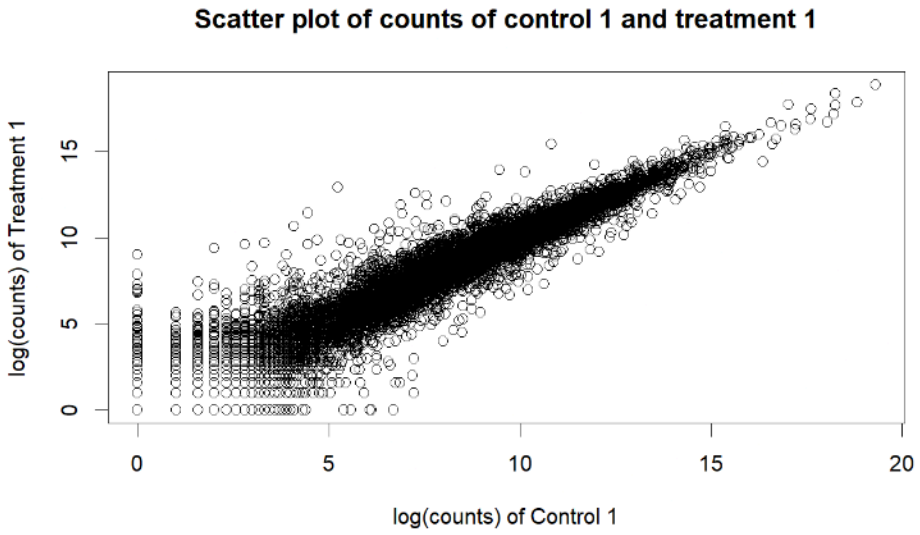


Figure 4: Scatter plot illustrating relationship between feature counts of control 1 and its corresponding treatment replicate (1).

Figure 5.

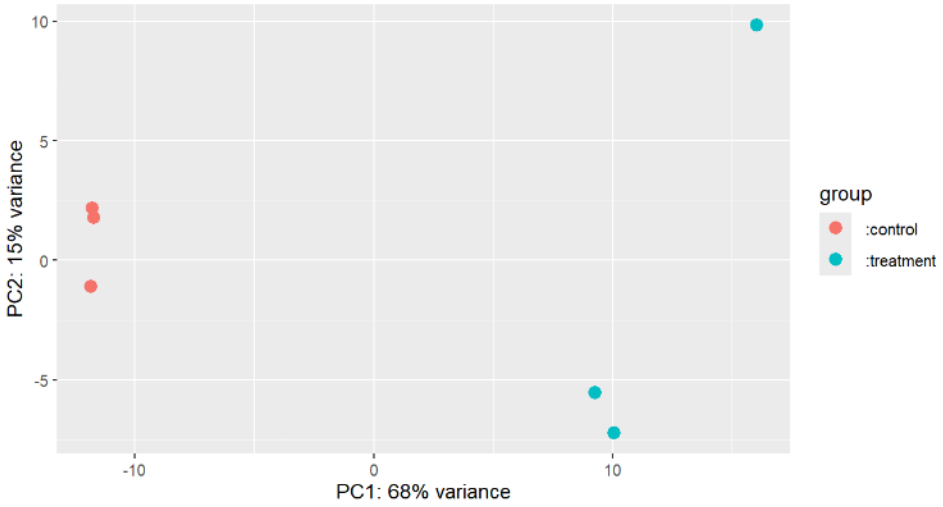


Figure 5: PCA plot illustrating variance among controlled and treatment replicate conditions.

Figure 6.

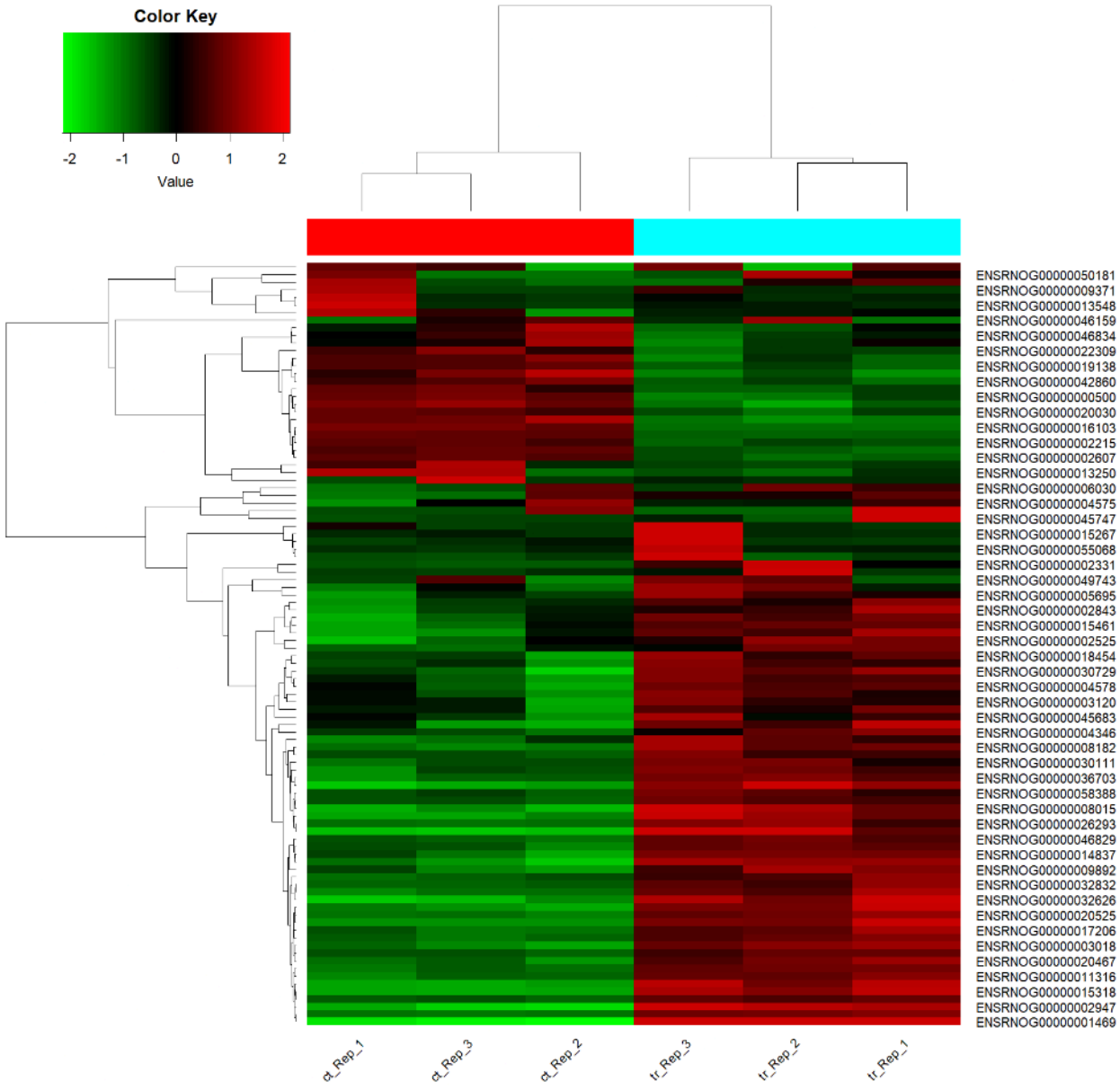


Figure 6: Heatmap representing changes in gene expression level. Each column represents a replicate, whereas each row represents a unique gene. Red color represents upregulated genes and green represents downregulated genes, whereas black represents unchanged expression.

Figure 7.

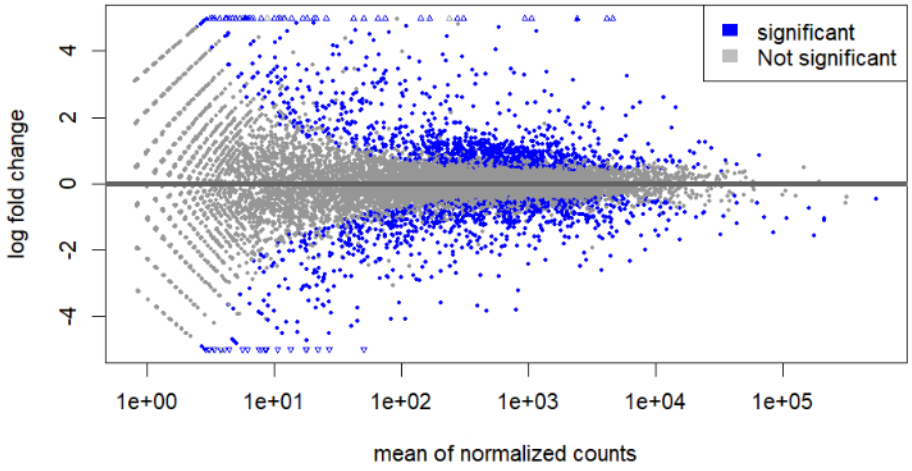


Figure 7: MA plot representing log fold change versus mean expression between treatments.

Figure 8.

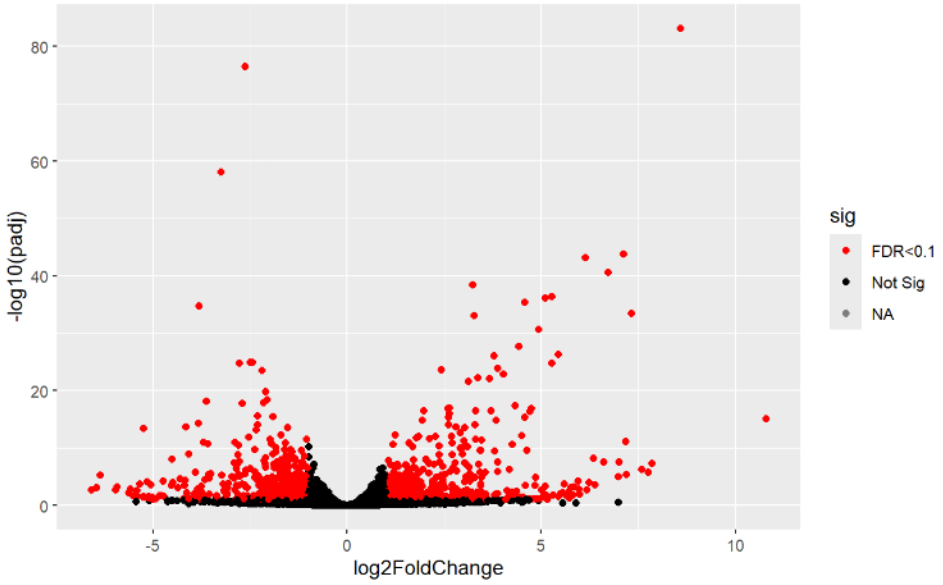
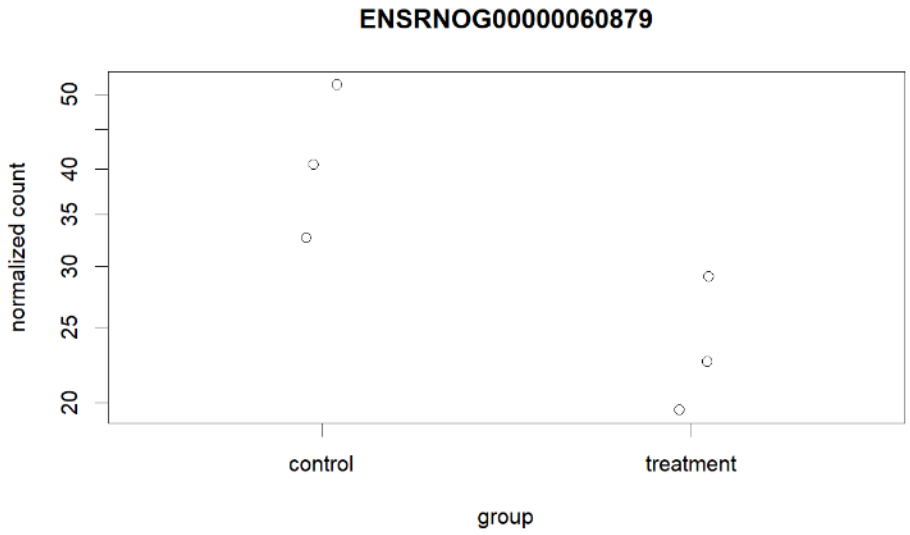


Figure 8: Volcano plot summarizing the statistical significance of the difference relative to magnitude of change in expression level of genes. The red dots represent statistically significant difference whereas black dots represent non-significant fold changes.

435 **Figure 9.**



436

437 Figure 9: Scatter plot illustrating normalized count of expression of the top gene
438 identified in analysis.

439

Tables

Table 1: Representative snapshot of the log 2 fold change values among treatment and control conditions with wald test p-values.

ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSRNOG00000060879	32.14714	-0.8193083	0.427382	-1.91704	0.055233	0.230443
ENSRNOG00000055506	2.12359	1.0790471	1.877767	0.574644	0.565532	0.785987
ENSRNOG00000061597	4.14977	-2.0683942	1.265955	-1.63386	0.102288	0.33407
ENSRNOG00000039592	1.84025	-0.8425574	1.933644	-0.43574	0.663029	0.844372
ENSRNOG00000055769	2.34818	-1.3980512	1.546567	-0.90397	0.366011	0.641711
ENSRNOG00000052424	234.35948	0.0527244	0.337949	0.156013	0.876023	0.950412

Table 2: Representative snapshot of the log 2 fold change values among treatment and control conditions with wald test p-values, after removing outliers and null values.

ID	baseMean	log2FoldChange	lfcSE	stat	pvalue	padj
ENSRNOG00000061379	165.3274	10.79785	1.237454	8.71778	2.84E-18	8.21E-16
ENSRNOG00000001469	4573.199	8.59577	0.431745	19.88623	5.36E-88	7.28E-84
ENSRNOG00000051905	42.0038	7.85389	1.267877	6.18663	6.15E-10	5.93E-08
ENSRNOG00000061294	20.2024	7.76361	1.396627	5.55167	2.83E-08	1.88E-06
ENSRNOG00000052159	18.0466	7.60123	1.309274	5.79804	6.71E-09	5.01E-07
ENSRNOG00000011824	273.7817	7.33912	0.574416	12.75927	2.77E-37	3.14E-34

449 **Table 3: Summary of upregulated genes obtained from hypergeometric test with**
450 **biological process ontology.**

	GOBPID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
1	GO:0006260	6.97E-11	4.9545	23.385	50	74	DNA replication
2	GO:0006261	2.65E-09	5.187	18.329	40	58	DNA-templated DNA replication
3	GO:0006259	2.66E-09	2.8412	47.086	80	149	DNA metabolic process
4	GO:0002376	3.84E-07	2.0498	80.584	115	255	immune system process
5	GO:0030097	5.88E-07	2.7377	34.13	58	108	hemopoiesis
6	GO:0002682	1.86E-06	2.1922	54.354	82	172	regulation of immune system process
7	GO:0032508	3.84E-06	9.5252	6.6363	17	21	DNA duplex unwinding
8	GO:0071103	4.70E-06	8.0789	7.2683	18	23	DNA conformation change
9	GO:0006950	8.42E-06	1.6787	155.48	192	492	response to stress
10	GO:0006974	9.44E-06	2.2523	42.03	65	133	DNA damage response

451

Table 4: Summary of upregulated genes obtained from hypergeometric test with cellular components ontology.

	GOCCID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
1	GO:0005694	2.09E-07	2.2578	61.242	93	194	chromosome
2	GO:0005576	4.17E-06	1.8836	86.18	118	273	extracellular region
3	GO:0030312	8.98E-05	2.305	29.674	47	94	external encapsulating structure
4	GO:0031012	8.98E-05	2.305	29.674	47	94	extracellular matrix
5	GO:0042555	9.44E-05	Inf	2.5254	8	8	MCM complex
6	GO:0005615	1.48E-04	1.7757	65.661	89	208	extracellular space
7	GO:0032993	3.26E-04	2.115	30.936	47	98	protein-DNA complex
8	GO:0071162	6.14E-04	17.64	2.8411	8	9	CMG complex
9	GO:0000228	1.78E-03	2.4317	15.784	26	50	nuclear chromosome
10	GO:0031261	2.22E-03	8.8109	3.1568	8	10	DNA replication preinitiation complex

Table 5: Summary of upregulated genes obtained from hypergeometric test with molecular functions ontology.

	GOMFID	Pvalue	OddsRatio	ExpCount	Count	Size	Term
1	GO:0003678	1.60E-06	17.022901	5.327189	15	17	DNA helicase activity
2	GO:0003677	2.67E-06	2.137487	57.345622	85	183	DNA binding
3	GO:0008094	4.58E-06	7.228792	7.834101	19	25	ATP-dependent activity, acting on DNA
4	GO:0140097	1.05E-05	3.654971	15.354839	30	49	catalytic activity, acting on DNA
5	GO:0004386	2.65E-05	7.257143	6.580645	16	21	helicase activity
6	GO:0017116	1.05E-04	12.357683	4.073733	11	13	single-stranded DNA helicase activity
7	GO:0001216	1.79E-04	3.582531	11.281106	22	36	DNA-binding transcription activator activity
8	GO:0001228	1.79E-04	3.582531	11.281106	22	36	DNA-binding transcription activator activity, RNA polymerase II-specific
9	GO:0030545	3.68E-04	2.448864	20.682028	34	66	signaling receptor regulator activity
10	GO:0003697	4.80E-04	3.589974	9.714286	19	31	single-stranded DNA binding