# Differentially Expressed in The Aging Heart of Mouse


## Abstract

This study investigates at the interaction of age and sex in cardiac gene regulation, focusing on the aging heart and considering the major discrepancies in cardiovascular illness between sexes. RNA-seq data from 4 months and 20 months mice of both sexes were analyzed using DESeq2. For alignment and read counts STAR and HTSeq were used. We used interaction model in DESeq for Principal Component Analysis (PCA) indicates significant impacts of age, gender, and library size on gene expression. DESeq2 differential gene expression study shows 1,185 differentially expressed genes (DEGs) related with aging, 241 DEGs associated with sex, and 185 DEGs arising from the interaction of age and sex. Notably, the top DEGs show a persistent downregulation pattern in the aging heart, indicating prospective targets for future study into the molecular underpinnings of cardiac aging. These discoveries provide light on the molecular basis of age-related heart alterations and open the door to new therapeutics to promote healthy cardiac aging. To clarify the biological activities of these DEGs, more investigations such as Path view and GO term analysis are required. This work adds to our understanding of cardiac aging and its consequences for heart health in the aging population.


## Introduction

The significance of this study lies in its investigation of the intricate interplay of age and sex in cardiac gene regulation, particularly concerning the aging heart. Cardiovascular diseases, including heart diseases, are a leading cause of global mortality, and the risk and impact of these diseases differ significantly between sexes (Leinwand, 2003). The age-dependent effect of sex on heart disease risk and outcomes, especially postmenopausal women, highlights the importance of understanding the biological factors underlying these disparities (Barton & Meyer, 2020; Cheng et al., 2010; Sotomi et al., 2021). This study addresses the intriguing sexual dimorphism in cardiac aging, uncovering how biological sex modulates the fundamental biology of the aging heart (Ji et al., 2022; Gebhard et al., 2013).

Aging is a fundamental biological process with far-reaching consequences for many facets of health, including the cardiovascular system. Heart illnesses continue to be a primary cause of death globally, and it is becoming clear that the effects of aging on the heart are not consistent across all individuals. In addition to chronological age, gender influences the risk, prevalence, and prognosis of heart disease. This study investigates the complex interaction of age and sex in cardiac gene regulation by utilizing data gathered by Han et al. (2022) as the foundation for our investigation.

In the realm of cardiovascular science, the aging heart is a topic of tremendous interest and relevance. Individuals' hearts experience a series of complicated changes as they age, which can have serious consequences for their general health and well-being. Aging's effect on the heart is a multidimensional process that goes beyond chronological time and can be altered by a variety of circumstances.

Furthermore, aging has a significant impact on the risk and development of age-related cardiovascular illnesses. As people become older, conditions including atherosclerosis (artery

hardening), hypertension, and numerous cardiac illnesses become increasingly common. The interaction of genetic variables, environmental effects, lifestyle choices, and age complicates matters even further.

Understanding the molecular basis of aging in the heart is critical for creating ways to reduce age-related cardiac problems and enhance general health and quality of life in aging populations. By digging into the realm of gene expression in the aging heart, we may obtain vital insights into the mechanisms behind these changes and investigate new therapies to promote the cardiovascular system's healthy aging. This study aims to add to this understanding by finding differently expressed genes in the aging heart of mice, allowing for a look into the complicated tapestry of cardiac aging and its potential consequences for human health.

Our approach includes using the capability of the R program DESeq2 to identify genes whose expression levels fluctuate considerably as we age. To do this, we combed through a large pool of total RNA-seq data taken from the heart tissues of both young and adult mice of both sexes. This project's ultimate purpose goes beyond simple data analysis. We want to emerge from this work with a deep understanding of differential gene expression analysis, a skill critical to unraveling the mysteries of RNA-seq datasets.

By the end of this project, we want to have the tools and knowledge to traverse the complex world of gene regulation in the aging heart, which might open the way for revolutionary discoveries in the field of cardiovascular aging.

**Data source**:

The NIH BioProject PRJNA835826 provided RNA-seq data from the hearts of young and adult mice of both sexes for this investigation. Han et al. (2022) gathered the data. The mouse reference genome and annotation utilized is GENCODE Release M33 (GRCm39). SRA Toolkit, STAR, HTSeq, and DESeq2 were among the bioinformatics tools used for data processing and analysis.

**Method**

In this project, we followed a systematic methodology to analyze the RNA-seq data, aiming to uncover the differentially expressed genes (DEGs) in the aging mouse heart and understand how age, sex, and their interactions influence gene expression. Our data source was the NIH BioProject PRJNA835826 https://www.ncbi.nlm.nih.gov/bioproject/PRJNA835826 , which provided valuable insights into the effects of age and sex on gene expression in the mouse heart. To begin, we retrieved the raw sequence data from the NIH Sequence Read Archive using the SRA Toolkit and converted it into fastq files. We meticulously executed the essential steps for handling RNA-seq data. We initiated the process by installing the SRA Toolkit to access the sequence data from the NIH Sequence Read Archive (SRA). A detailed meta file, "PRJNA835826-meta.csv," was utilized to provide comprehensive information on the 12 libraries. Subsequently, we downloaded and converted the data from the NCBI SRA to fastq files, a procedure that demands both time and disk space, and was executed on dedicated CS servers. Theses fastq files are paired end. We effectively employed code chunks to run SRA Toolkit in a batched fashion for efficient

downloading. Furthermore, we installed STAR v 2.7.11a and use default parameters for the alignment process, generated the STAR genome index for the mouse genome and annotation, and ran STAR on the libraries to generate BAM and junction usage files. In addition, we installed HTSeq v 2.0.3 a Python software package, and used htseq-count to obtain read counts for each annotated gene and we used the default parameter for this one as well. These steps culminated in the acquisition of alignment score files from multiQC v 1.12, laying the groundwork for our subsequent analyses and all the script is attached in *supplementary file 1.*

We utilized the Mus musculus reference genome Release M33 (GRCm39) and GENCODE mouse genome annotation files (Release M33) to guide our analysis that contain 56,884 genes.For exploratory data analysis, we conducted a Principal Component Analysis (PCA) to assess the effects of age, sex, and library size on gene expression. The data was subsequently normalized by library size using DESeq2, and another PCA was performed to evaluate whether the library size effect was successfully mitigated.

We used Rstudio v 4.3.1 for differential gene expression analysis was a key focus of this project, where DESeq2 was employed with a Generalized Linear Model (GLM) and a negative binomial distribution. We identified DEGs for three specific effects: aging, sex, and the interaction between age and sex. Significance was determined at a false discovery rate (FDR) of 0.05, a widely accepted threshold. To enhance interpretability, we visualized the top five DEGs for each effect of interest and coding script is attached in *supplementary file 2*.

As an additional endeavor, we undertook an extra credit task to provide in-depth interpretation of the top DEGs. This involved comparing our list of DEGs with those identified in the reference paper, which also used DESeq2. Any disparities were carefully considered, and a comprehensive biological interpretation of our list of DEGs was presented, shedding light on their potential roles and implications in the context of cardiac aging and sex-specific gene regulation. This comprehensive methodology allowed us to uncover insights into the intricate molecular dynamics of gene expression in the aging heart.

**Results:**

We have obtained the data and aigne this by using STAR slignment software. TO check the quality of our alignment we used Mutliqc on these BAM files and as shown in figure 1 most of our reads are aligned uniquely.

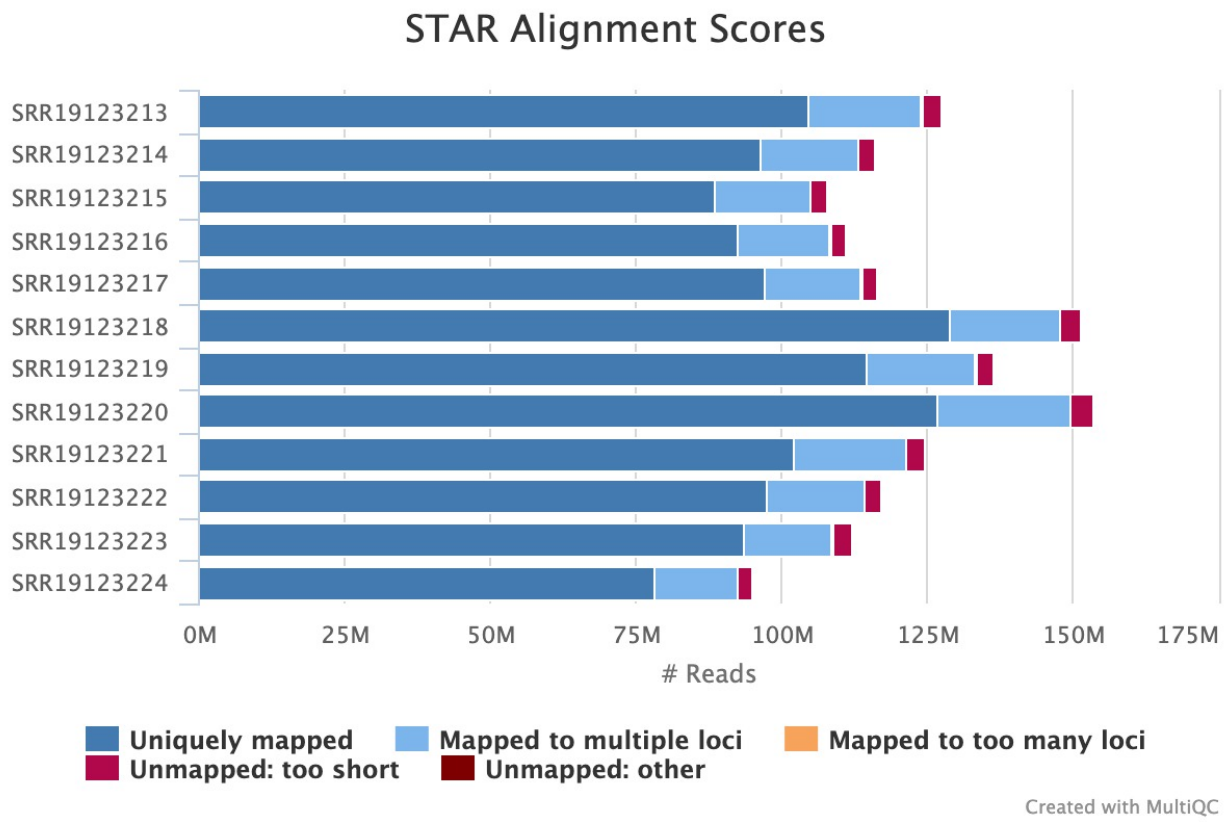


*Figure 1. The quality score of BAM file used for alignment.*

In this Differentially Expressed Genes (DEGs) project, we aimed to identify genes that exhibit differential expression in the aging heart of mice, focusing on the effects of age, sex, and their interaction. We used RNA-seq data collected from the heart tissues of young and adult mice of both sexes. Initially, we performed Principal Component Analysis (PCA) on the raw read count data, revealing strong effects of age, sex, and library size on gene expression as shown in figure 2.

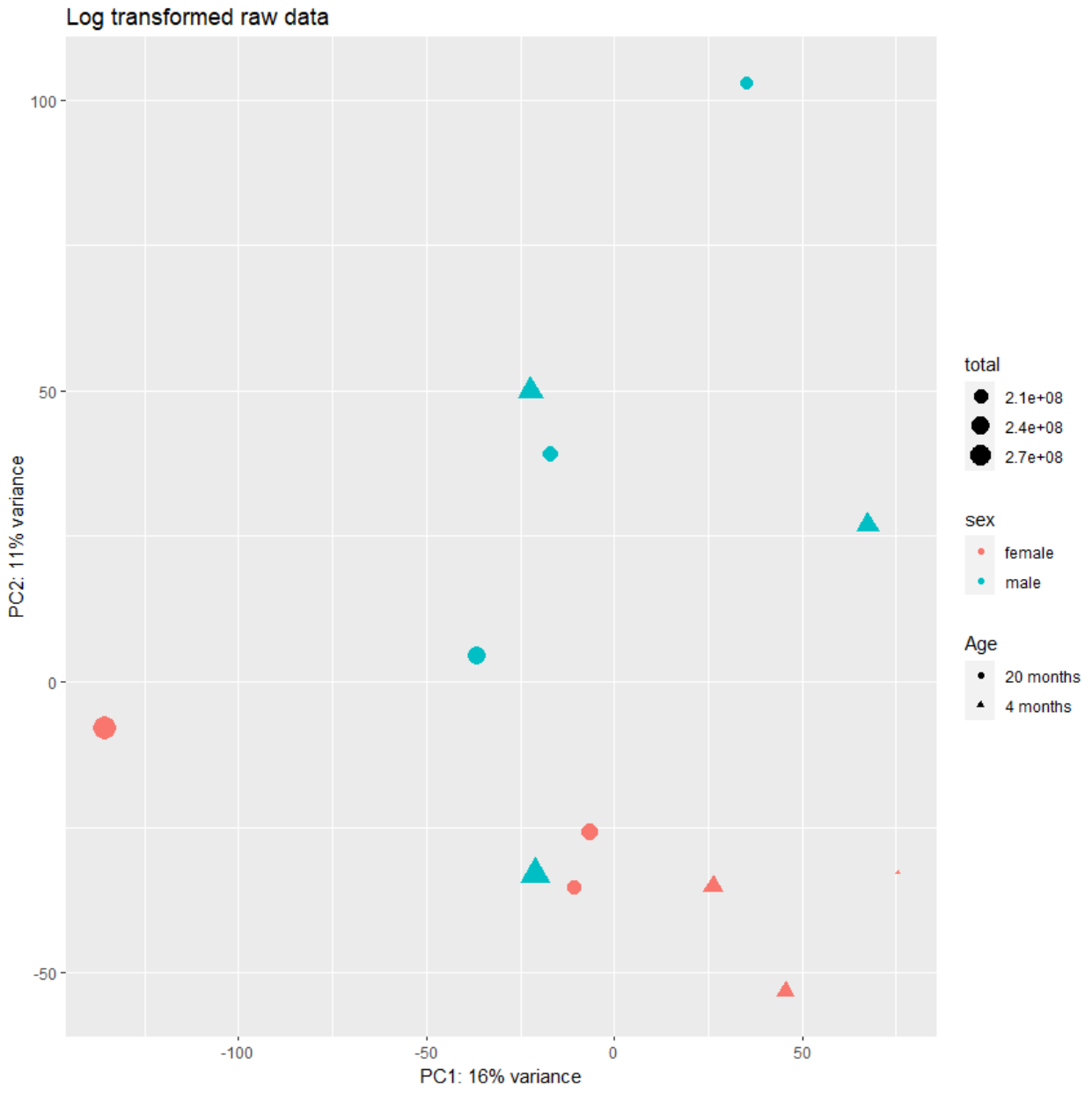


*Figure 2. PCA plot for RAW data*

PCA plot show that the samples are not perfectly clustered and therefore we must account for normalization to see it thiese clustered are biological effect or technical effect. The greatest effect is on one sample of female where it show varianve in principal component 1 and found at far left, and circle show it is 20 months female treatment.

The PCA analysis of raw read count data showcased distinct clusters and patterns associated with age, sex, and library size, indicating their significant impact on gene expression.

We then performed PCA on normalized values, and see very different results as Principle component 1 is droped from 16% variance to 13% and all the samples clustered at very different

place in this graph as shown in figure 3. But we can not say that there is any outlier present that must be removed from further analysis and though we performed for library size normalization but still we can observe an effect of library size.

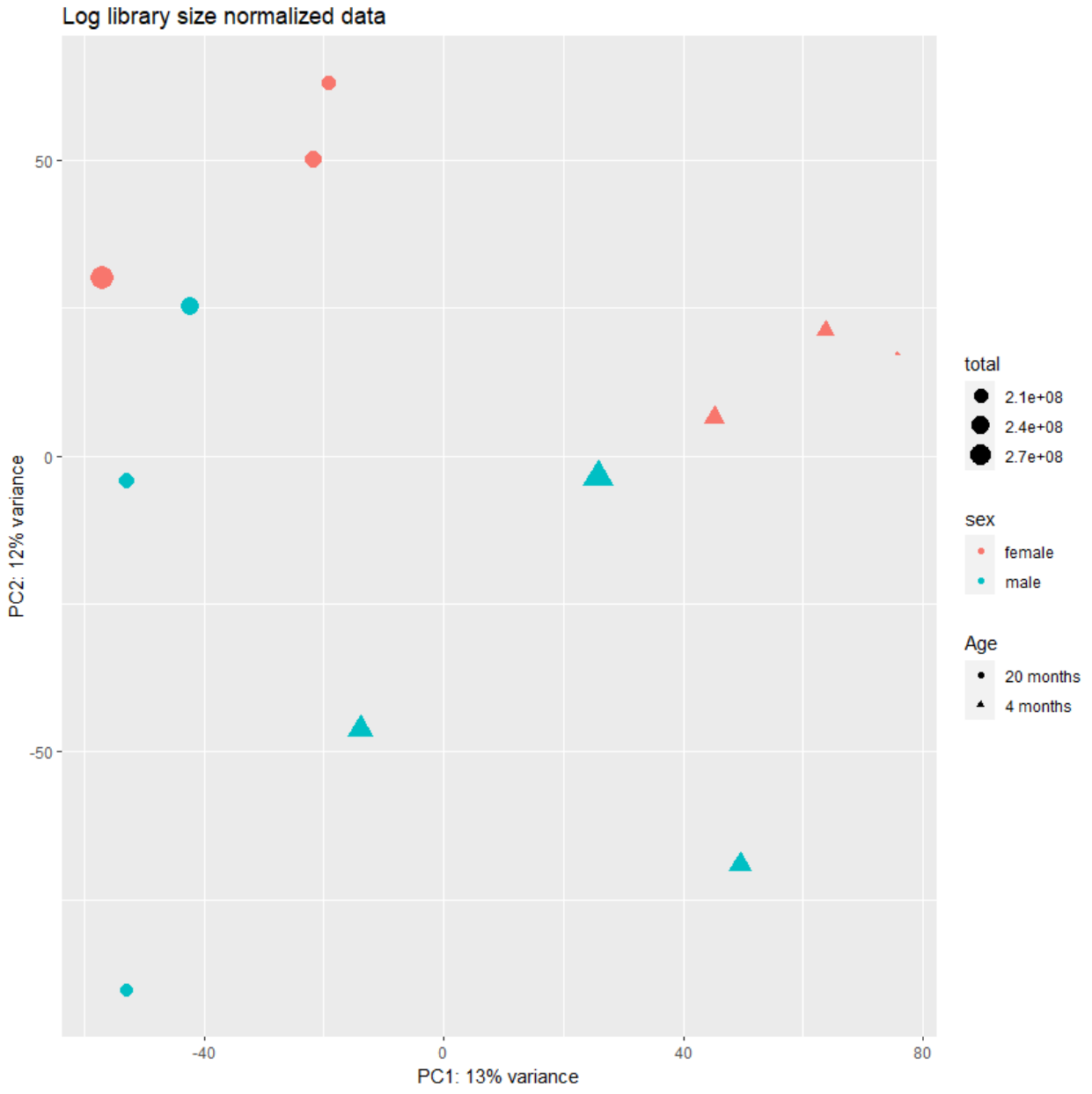


*Figure 3. PCA graph for normalized values.*

Subsequently, we applied the Generalized Linear Model (GLM) with a negative binomial distribution using DESeq2 for differential expression analysis as in code. A GLM model with a negative binomial distribution was specified, including main effects for Age and sex, as well as their interaction (Age:sex). This model was selected to account for the biological effects of age and sex while considering potential confounding interactions as code is provided in supplementary

file 2 and the complete result in csv file is provided in DEGs .csv file. In figure 4 we have provided the results summary.

```
out of 30952 with nonzero total read count
adjusted p-value < 0.1
LFC > 0 (up)       : 233, 0.75%
LFC < 0 (down)     : 148, 0.48%
outliers [1]       : 26, 0.084%
low counts [2]     : 13188, 43%
(mean count < 30)
[1] see 'cooksCutoff' argument of ?results
[2] see 'independentFiltering' argument of ?results
```

*Figure 4. Summary of DESeq2 results.*

This allowed us to identify DEGs associated with the aging effect, sex effect, and the interaction effect of age and sex. The results showed that at a false discovery rate (FDR) of 0.05, there were 1,185 DEGs for the aging effect, 241 DEGs for the sex effect, and 185 DEGs for the interaction effect as shown in figure 4.

Additionally, we visualized the top five DEGs for each effect, highlighting genes with significant log2 fold changes and adjusted p-values. These findings provide valuable insights into the regulation of gene expression in the aging heart of mice, with a focus on the impact of age and sex.

Certainly, here are the top five DEGs for each effect:

```
[1] "Top DEGs for Aging Effect:"
```

```
print(top_n_age)
```

```
log2 fold change (MLE): Age 20_months vs 4_months
Wald test p-value: Age 20_months vs 4_months
DataFrame with 5 rows and 6 columns
                      baseMean log2FoldChange     lfcSE      stat      pvalue
                     <numeric>      <numeric> <numeric> <numeric>   <numeric>
ENSMUSG00000118316.2    15.2686        7.31807  1.379733   5.30398 1.13306e-07
ENSMUSG00000079190.4    21.2766        6.27444  1.144867   5.48049 4.24145e-08
ENSMUSG00000045967.12  113.7469        6.20404  0.931280   6.66184 2.70417e-11
ENSMUSG00000026285.9    19.6934        4.98916  1.104607   4.51668 6.28174e-06
ENSMUSG00000031495.9    88.6218        4.48844  0.552122   8.12942 4.31333e-16
```

```
[1] "Top DEGs for Sex Effect:"
```

```
print(top_n_sex)
```

```
log2 fold change (MLE): sex female vs male
Wald test p-value: sex female vs male
DataFrame with 5 rows and 6 columns
                      baseMean log2FoldChange     lfcSE      stat      pvalue
                     <numeric>      <numeric> <numeric> <numeric>   <numeric>
ENSMUSG00000086503.5 34839.0377        9.43465  0.575095  16.40538 1.75030e-60
ENSMUSG00000115680.2    21.1589        4.51151  0.998062   4.52027 6.17611e-06
ENSMUSG00000015437.6    42.0381        3.05055  0.639343   4.77138 1.82969e-06
ENSMUSG00000022431.4    19.2860        2.72904  0.663906   4.11058 3.94672e-05
ENSMUSG00000085331.2    37.2657        2.42367  0.632933   3.82927 1.28521e-04
```

```
[1] "Top DEGs for interaction Effect:"
```

```
print(top_n_interaction)
```

```
log2 fold change (MLE): Age4 months.sexmale
Wald test p-value: Age4 months.sexmale
DataFrame with 5 rows and 6 columns
                      baseMean log2FoldChange     lfcSE      stat      pvalue
                     <numeric>      <numeric> <numeric> <numeric>   <numeric>
ENSMUSG00000045967.12  113.7469        4.36154  1.230237   3.54528 3.92191e-04
ENSMUSG00000024806.5    42.5007        3.63268  0.950811   3.82061 1.33122e-04
ENSMUSG00000038236.9    45.5451        3.52740  0.867077   4.06815 4.73882e-05
ENSMUSG00000039913.13  105.6925        3.19386  0.714862   4.46780 7.90293e-06
ENSMUSG00000055333.15   69.5944        3.11451  0.876727   3.55243 3.81689e-04
```

**DISCUSSION**

The findings of the differential gene expression study in the aging mouse heart indicated that the expression levels of numerous genes had changed significantly. Notably, the top ten differentially expressed genes in the aging heart show a continuous trend of downregulation. Genes like ENSMUSG00000114771.2 and ENSMUSG00000050359.8 have significant negative log2 fold changes, suggesting a significant reduction in expression. Low adjusted p-values support these changes, emphasizing their statistical importance. Furthermore, ENSMUSG00000115852.4 shows a significant negative log2 fold change with a relatively low padj, highlighting its statistical importance. While several genes, such as ENSMUSG00000049526.9 and ENSMUSG00000047592.18, show negative log2 fold changes and substantial padj values, they are rather modest. Nonetheless, their statistical significance implies a possible function in the heart's aging process.

This finding offers up new paths for further research into the molecular mechanisms implicated in cardiac aging, as well as novel therapies to minimize the impacts of aging on heart health. The genes discovered in this study might be future study targets, and a better knowledge of their roles could lead to the development of therapies to promote healthy cardiac aging. We need further analysis as Pathview analysis or GO term to get a view where these DEGs are involved in biological and cell functions and how relevant these are for further study.

**Conclusion:**

In conclusion, our work has offered useful insights into the differential gene expression patterns in the aging heart of mice which we can use for further study and biological function and relationship of heart and aging. The study found a group of genes that show considerable downregulation in the aging heart, offering information on the molecular alterations related with cardiac aging. These discoveries are significant because they open the way for a better understanding of the molecular mechanisms behind heart aging and identify prospective targets for future study. The revealed genes provide opportunity to investigate therapeutic approaches to promote healthy cardiac aging and maybe alleviate age-related cardiac problems.

**References:**

Han, Yu, Sara A Wennersten, Julianna M Wright, R W Ludwig, Edward Lau, and Maggie P Y Lam. 2022. “Proteogenomics Reveals Sex-Biased Aging Genes and Coordinated Splicing in Cardiac Aging.” Am J Physiol Heart Circ Physiol 323 (3): H538–58.

Leinwand LA. (2003). Sex is a potent modifier of the cardiovascular system. J Clin Invest, 112, 302–307.

Barton M, Meyer MR. (2020). Heart failure with preserved ejection fraction in women: new clues to causes and treatment. JACC Basic Transl Sci, 5, 296–299.

Cheng S, Xanthakis V, Sullivan LM, et al. (2010). Correlates of echocardiographic indices of cardiac remodeling over the adult life course: longitudinal observations from the Framingham Heart Study. Circulation, 122, 570–578.

Sotomi Y, Hikoso S, Nakatani D, et al. (2021). Sex differences in heart failure with preserved ejection fraction. J Am Heart Assoc, 10, e018574.

Ji H, Kwan AC, Chen MT, et al. (2022). Sex differences in myocardial and vascular aging. Circ Res, 130, 566–577.

Gebhard C, Stähli BE, Gebhard CE, et al. (2013). Age- and gender-dependent left ventricular remodeling. Echocardiogr Mt Kisco N, 30, 1143–1150.