# Final Report

## Sex Gene Expression Comparative Study Using RNA-seq

Data Practicum Report for Data2050, Summer2024, Larschan Lab,
Chuning Xiao & Jingyi Lu, Supervised by Mukulika Ray

**Background**

The study of sex differences in gene expression is driven by the need to understand the molecular basis of sex-specific disease mechanisms. Research has shown that sex differences play a critical role in various diseases at the genetic level, such as autoimmune disorders, neurodegenerative diseases, and cardiovascular conditions. For example, women are more susceptible to autoimmune diseases like systemic lupus erythematosus partly due to differences in immune gene regulation which are raised from sex chromosomes and hormones[1]. Sex-based differences in gene expression have also been linked to the distinct prevalence and progression of neurodegenerative diseases such as Parkinson's and Alzheimer's[2], where each gender exhibits different regulatory patterns in brain tissues. Understanding these differences can inform the development of sex-specific therapeutic strategies and improve personalized medicine approaches, ensuring that both men and women receive the most effective treatments.Therefore studying gene expression differences in male and female embryonic stem cells can provide foundational insights into how sex-specific gene regulation emerges and its implications for health and disease.

RNA sequencing (RNA-seq) is a powerful and versatile technology that provides comprehensive insights into the complete set of RNA transcripts produced by the genome. RNA-seq allows for the quantification and discovery of RNA molecules, offering detailed information about gene expression, alternative splicing events and transcript identification.

This report documents the implementation and execution of two variations of the *nf-core/rnaseq* pipeline (version 3.14.0) on four publicly available datasets from the NCBI Sequence Read Archive. This pipeline is designed to provide a standardized, scalable, and reproducible analysis workflow for RNA-seq data, integrating best practices and state-of-the-art tools. Our goal is to explore the two different approaches.

The four datasets used in this implementation are:

1. *SRX17080731* -Female Rep1
2. *SRX17080732* -Female Rep2
3. *SRX131957* -Male Rep1
4. *SRX131956* -Male Rep2

The datasets used in this study focus on gene expression differences between male and female mouse embryonic stem cells (mESCs), specifically from pSM44 mES cells. Investigating

[1] Sharma, S., Eghbali, M. Influence of sex differences on microRNA gene regulation in disease. Biol Sex Differ 5, 3 (2014). https://doi.org/10.1186/2042-6410-5-3

[2]Lopes-Ramos CM, Chen CY, Kuijjer ML, et al. (2020). Sex Differences in Gene Expression and Regulatory Networks across 29 Human Tissues. Cell reports, 31(12), 107795. https://doi.org/10.1016/j.celrep.2020.107795

these differences is crucial for understanding sex-specific developmental processes, gene regulation, and the implications for personalized medicine. Male and female mESCs exhibit distinct gene expression profiles due to the influence of sex chromosomes, affecting cellular functions and developmental outcomes. By analyzing these datasets, we aim to identify key genes and pathways with sex-specific expression patterns, providing insights into the molecular mechanisms underlying early development and contributing to advancements in developmental biology and regenerative medicine.

## Process

To examine the factors of different gene expressions between male and female mouse embryonic stem cells, having a general understanding of the transfer of genetic information is important. In DNA, a specific sequence of nucleotides (adenine, thymine, cytosine, and guanine) arranged in a particular order forms a gene and carries the information required to produce corresponding proteins. Genetic information encoded in DNA is transferred and translated into proteins through gene expression, a multi-step process that includes transcription, RNA processing (such as splicing), and translation. During transcription, RNA polymerase serves as the catalyst of the complementary base-pairing from a single strand of DNA to a precursor mRNA (pre-mRNA), which contains both exons (coding regions) and introns (non-coding regions)[3]. Subsequently, RNA splicing converts the pre-mRNA molecule into a mature mRNA molecule by removing the introns and joining the exons together. However, there can be different junctions for a pre-mRNA molecule to be spliced at and different combinations of exons form different mature mRNA molecules[4]. Such difference leads to the generation of diverse proteins and, therefore, divergent gene expressions between the same type of cells from different individuals.

To further explore the differences between male mouse genes and female mouse genes, we use RNA-seq to study the quality and quantity of RNA molecules in the samples of the mESCs. During the library preparation phase of RNA-seq, unique barcodes—short DNA sequences—are added to the RNA molecules from different samples. This barcoding allows multiple samples to be pooled together and sequenced in a single run. After sequencing, the mixed dataset is processed in the demultiplexing step, where reads are scanned to identify the unique barcodes. Each read is then sorted and assigned back to its corresponding sample, resulting in separate FASTQ files for each sample. These files contain the reads specific to each sample, enabling subsequent analysis.

FASTQ files are generated in the lab through these steps: RNA samples are first isolated and prepared for sequencing, then converting RNA into complementary DNA (cDNA) using reverse transcription. The cDNA is then fragmented, and sequencing adapters are ligated to the fragments. These prepared samples are loaded onto a sequencing platform, such as Illumina, where they undergo sequencing-by-synthesis. During this process, each fragment is sequenced base-by-base, and the resulting data includes both the nucleotide sequences and their quality scores. This data is captured in FASTQ files, with each read represented by four lines: a sequence identifier, the nucleotide sequence, a separator, and the quality scores. This format

[3] Clancy, S. & Brown, W. (2008) Translation: DNA to mRNA to Protein. Nature Education 1(1):101. https://www.nature.com/scitable/topicpage/translation-dna-to-mrna-to-protein-393/

[4] Clancy, S. (2008) RNA splicing: introns, exons and spliceosome. Nature Education 1(1):31.https://www.nature.com/scitable/topicpage/rna-splicing-introns-exons-and-spliceosome-12375

ensures that both the sequence and its quality information are preserved for downstream analysis.

Sequencing data can come in two forms: single-end reads and paired-end reads. In single-end sequencing, the DNA fragments are sequenced from one end, resulting in one read per fragment. In paired-end sequencing, both ends of the DNA fragments are sequenced, producing two reads per fragment, often referred to as Read 1 and Read 2. Paired-end sequencing is advantageous because it provides more information about each DNA fragment, which can improve alignment accuracy and the detection of structural variants. The difference in read numbers when using tools like *fasterq-dump* arises from whether the sequencing was performed as single-end or paired-end. For instance, in our datasets:Dataset 1,2 and 3 indicated paired-end sequencing with two reads per spot while Dataset 4 has one read per spot.The difference raised will be discussed later.

## Methods & pipelines

The first approach integrated into the *nf-core/rnaseq* pipeline offers a robust and efficient method for quantifying transcript abundances from RNA-seq data. The analysis begins with configuring the tool Nextflow, which is required to run the pipeline in Linux environment.Then we downloaded raw sequencing data in FASTQ files as well as reference genome (GRCm39) files. Provided with a designer file which specifies the parameters and alignment genome and adjusting memory and CPU allocation with system capabilities, the input data undergoes initial quality control using *FastQC* to ensure the data is reliable. After examining the quality of the data by visualizing the dataset, trimming of adapter sequences and low-quality bases is then performed using [TrimGalore], preparing the reads for further processing. Unlike the second approach that requires alignment, Salmon is free of manual alignment. The quantified transcript data is then aggregated to the gene level using *tximport*, simplifying the data for differential expression analysis. Finally, DESeq2 is used to perform statistical analysis, identifying genes that are differentially expressed between conditions, which does not involve aligning reads to a reference genome before quantification.

As demonstrated below, the initial reading of the four datasets indicates that Dataset 4 ( read SRR960178) generated single-end data instead of paired-end data when processed with *fasterq-dump --split-files* because the sequencing run was performed using a single-end sequencing strategy, where only one end of each DNA fragment was sequenced instead of two, potentially provides less information about the original DNA fragments and reducing the ability to detect structural variations and making it more challenging to map reads to repetitive regions of the genome. However, single-end sequencing is typically faster and more cost-effective than paired-end sequencing.

| |
|---|
| **fasterq-dump --split-files SRR21065599** |
| spots read : 11,983,403 |
| reads read : 23,966,806 |
| reads written : 23,966,806 |
| **fasterq-dump --split-files SRR21065600** |
| spots read : 12,393,752 |
| reads read : 24,787,504 |
| reads written : 24,787,504 |
| **fasterq-dump --split-files SRR960177** |
| spots read : 86,561,969 |
| reads read : 173,123,938 |
| reads written : 173,123,938 |
| **fasterq-dump --split-files SRR960178** |
| spots read : 123,885,746 |
| reads read : 123,885,746 |
| reads written : 123,885,746 |

The designer file which specifies the parameters associated with the pipeline is illustrated below:

| sample | fastq_1 | fastq_2 | strandedness |
|---|---|---|---|
| **female_1** | /mnt/c/Users/heypu/OneDrive/Desktop/Larcshan2024/SRR | /mnt/c/Users/heypu/OneDrive/Desktop/Larcshan2024/SRR21065 | forward |
| **female_2** | /mnt/c/Users/heypu/OneDrive/Desktop/Larcshan2024/SRR | /mnt/c/Users/heypu/OneDrive/Desktop/Larcshan2024/SRR21065 | forward |
| **male_1** | /mnt/c/Users/heypu/OneDrive/Desktop/Larcshan2024/SRR | /mnt/c/Users/heypu/OneDrive/Desktop/Larcshan2024/SRR96017 | forward |
| **male_2** | /mnt/c/Users/heypu/OneDrive/Desktop/Larcshan2024/SRR960178.fastq.gz | | forward |

The second approach focuses on the levels of RNA molecules and the essential tools required are *Bowtie2* and *DESeq2*. The R package *DESeq2* analyzes RNA sequencing data in the form of an object DESeqDataSet, which is constructed from a count matrix and a metadata table. To convert the raw data into the form, Bowtie2 can be used to align the reads from the FASTQ files with genes and the R function featureCounts builds the count matrix.

Alignment is an essential step in RNA-seq analysis. Since the files of the reads are raw data in the form of substantial sequences of RNA molecules, they have to be mapped with a reference genome or transcriptome to determine their original location within the genome. After obtaining the raw data, the pipeline proceeds with the sequencing alignment using Bowtie2 in command line requests. The genome assembly GRCm39, a latest reference of Mus musculus genomes, is built as the index of the alignment using *bowtie2-build*. The output of the *bowtie2* command is in Sequence Alignment/Map (SAM) format, showing the unsorted mapping of each individual read with a genome in the form of a text file. Since the SAM files are too large in size for downstream analysis, they are subsequently converted into binary compressed versions (BAM files) using Samtools. Based on this notion, each SAM file generated by *bowtie2* is piped into *samtools* in command to save the memory of the SAM files. An example of the command request is shown below:

```
bowtie2 -q --local -x mm39 -1 SRR960177_1.fastq -2 SRR960177_2.fastq | \
samtools sort -O bam -o - > SRR960177.bam
```

The next step of data processing is to construct a count matrix from the BAM files in R. Using the *featureCounts* function from the package ‘Rsubread’, the reads in BAM format can be counted and annotated with the corresponding genes. Since the reads from the second repetition of male embryonic stem cells are single-end sequenced, it has to be processed

separately from other samples generated by paired-end sequencing. By merging the two *featureCounts* outputs together, the counts of the single-end reads are attached to the paired-end ones for convenience of analysis. The metadata required for constructing a DESeqDataSet object, also called the colData, is a manually created sample table including the experimental information of the samples.

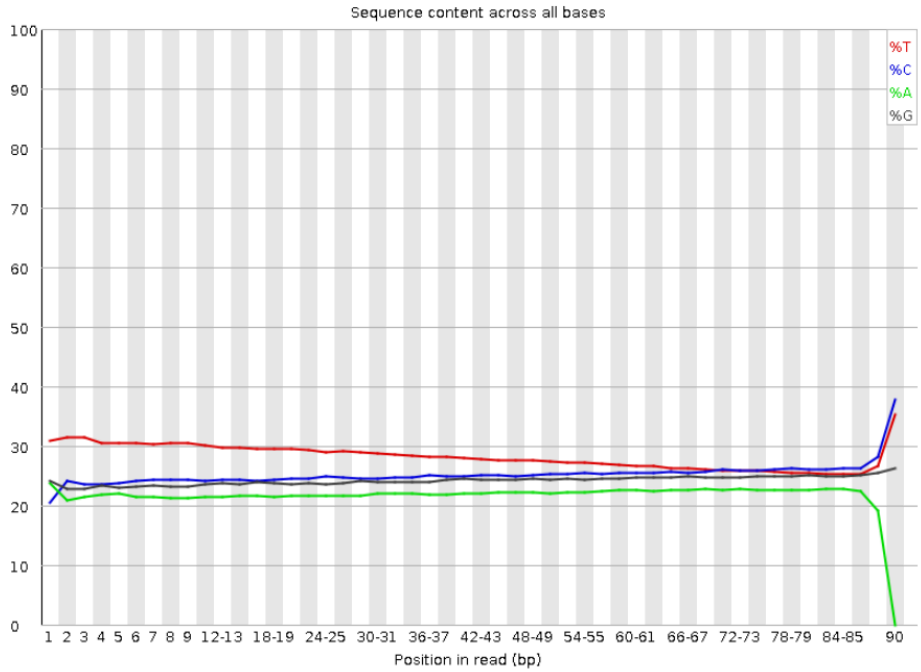
| | SampleName | Cell | Sex | Repetition | Run | avgLength | Experiment | Sample | BioSample |
|---|---|---|---|---|---|---|---|---|---|
| SRR21065600 | GSM6461743 | pSM44 mES | female | 1 | SRR21065600 | 108 | SRX17080731 | SRS14661908 | SAMN30321938 |
| SRR21065599 | GSM6461744 | pSM44 mES | female | 2 | SRR21065599 | 108 | SRX17080732 | SRS14661909 | SAMN30321937 |
| SRR960177 | GSM903663 | pSM33 mES | male | 1 | SRR960177 | 80 | SRX131957 | SRS304714 | SAMN00840798 |
| SRR960178 | GSM903662 | pSM33 mES | male | 2 | SRR960178 | 80 | SRX131956 | SRS304713 | SAMN00840797 |

A DESeqDataSet (dds) object is then created with the count data and the metadata by the *DESeqDataSetFromMatrix* function. As we are interested in comparing the differences of genes between male and female, rows with 0 counts for all 4 samples are filtered out because they are meaningless to the analysis. Moreover, the factor Sex is constructed with levels of female and male. Finally, the *DESeq* function does a differential expression analysis to the dds object and the results can be visualized through MA plots, volcano plots and heatmaps.

## Results

The visualized results from *Salmon* suggest that the data quality of the 7 reads are generally good with some unsatisfactory aspects that vary between individuals, rather than genders. Most of the quality checks (like Per Base Sequence Quality and Per Sequence Quality Scores) have passed, indicating that the overall quality of the sequencing data is high and is reliable for downstream analysisThere's a consistent failure in the Per Base Sequence Content check, suggesting a systemic issue in the sequencing or laboratory preparation, such as an imbalance in the nucleotide composition across reads. A failure in this check often indicates that there is a systematic bias in the nucleotide composition across different positions in the reads.This can affect how accurately the sequences can be aligned and quantified.

## Per base sequence content




The overrepresented sequences consistently pass 0.1% with some rare instances passing 1%. This is considered not satisfactory with the FastQC quality check standard because it means that some individual sequences within the dataset are appearing alarmingly frequently, suggesting a potential contamination from an external source, such as leftover primer dimers, adaptor sequences, or even environmental DNA from bacteria and human skin. This suggests that lab procedures need to be examined for errors. Overrepresented sequences can skew the quantification of gene expression, leading to inaccurate estimates of transcript abundance. Also, sequences that are overrepresented might align multiple times to the reference genome or transcriptome, which can complicate the interpretation of which genomic regions are truly being transcribed at high levels.

## Overrepresented sequences

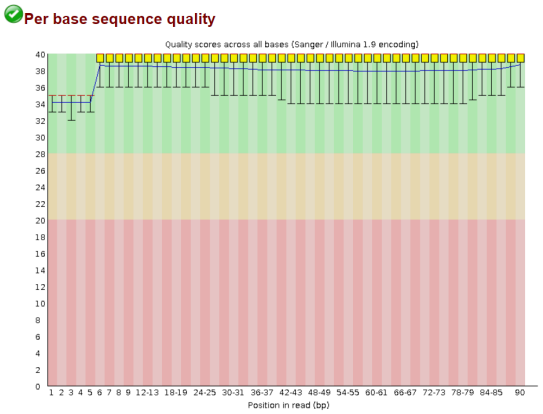
| Sequence | Count | Percentage | Possible Source |
|---|---|---|---|
| CGCGACCTCAGATCAGACGTGGCGACCCGCTGAATTTAAGCATATTAGTC | 13878 | 0.11375628956787281 | No Hit |
| AACACGGACCAAGGAGTCTAACGCGTGCGCGAGTCAGGGGCTCGTCCGAA | 12712 | 0.10419872841812935 | No Hit |
| TTACCTGGTTGATCCTGCCAGTAGCATATGCTTGTCTCAAAGATTAAGCC | 12387 | 0.10153474267742044 | No Hit |

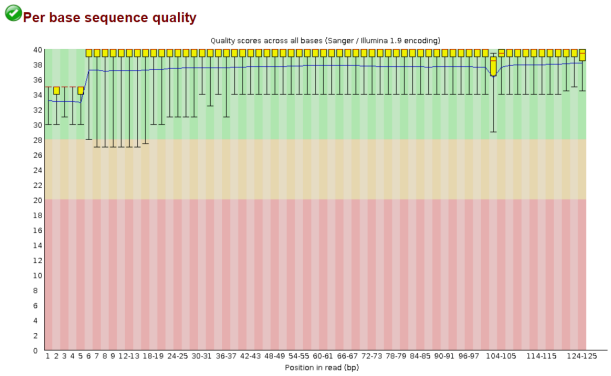
## Overrepresented sequences

| Sequence | Count | Percentage | Possible Source |
|---|---|---|---|
| TTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT | 144428 | 1.1838588694126486 | No Hit |
| AGCGACGCTCAGACAGGCGTAGCCCCGGGAGGAACCCGGGGCCGCAAGTG | 23099 | 0.18933971269118707 | No Hit |
| ACAAACCCTTGTGTCGAGGGCTGACTTTCAATAGATCGCAGCGAGGGAGC | 14018 | 0.11490385265617821 | No Hit |

The distribution of quality scores across all reads in a sequencing run is consistently passing across most reads, suggesting that the majority of our sequencing reads are of high quality,

making the baseline calling accurate and reading results reliable for downstream analysis. Despite the underlying overrepresentation issues, the data reads remain reliable. From the 10 outputs generated by 7 data reads, the variance of overall performance is stronger among individual reads rather than that among gender, undermining the potential difference in sex-specific gene expressions.




The alignment result from *bowtie2* was generally satisfactory with all 4 alignment rates exceeding 85%.

| Experiment | Sex | Repetition | Reads | Overall alignment rate |
|---|---|---|---|---|
| SRX17080731 | Male | 1 | 86561969 | 86.32% |
| SRX17080732 | Male | 2 | 123885746 | 97.54% |
| SRX131957 | Female | 1 | 12393752 | 95.90% |
| SRX131956 | Female | 2 | 11983403 | 96.57% |

The differential expression analysis done by *DESeq* showed that the levels of 83 genes were significantly changed in male pSM43 mES cells compared to female pSM44 mES cells ($\partial$ = .05) from 28603 genes analyzed in total[5].

```
res05 <- results(dds, alpha=0.05)
summary(res05)
```

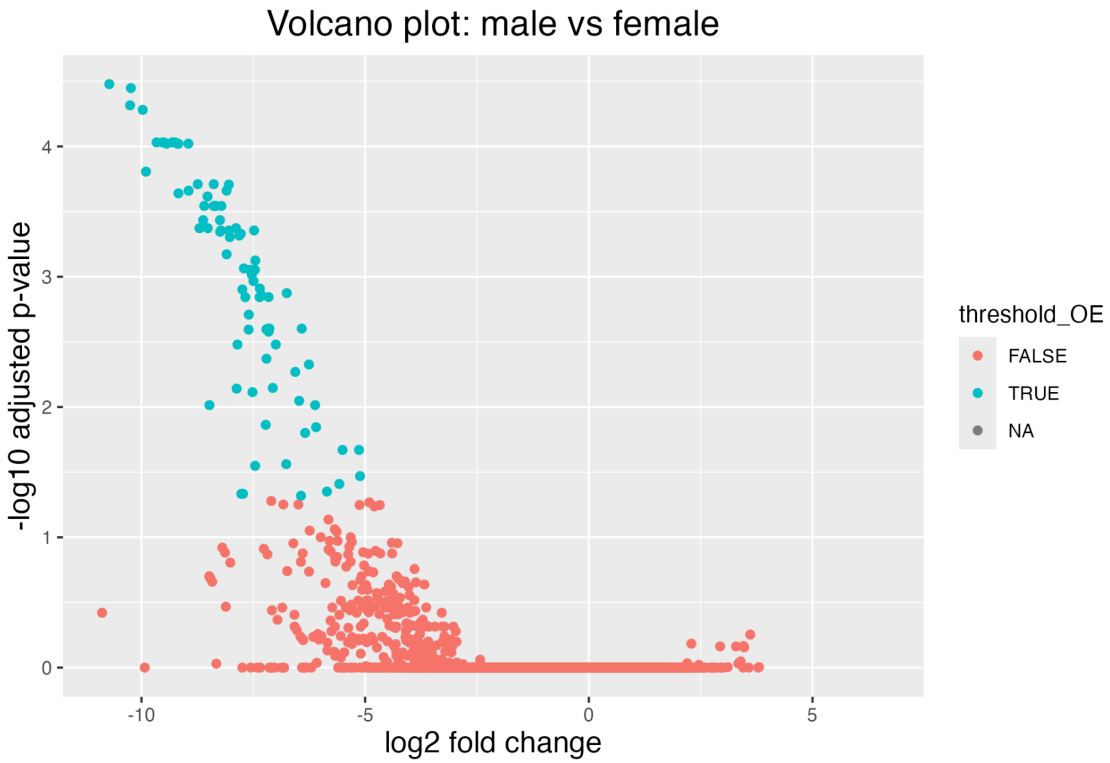
```
out of 28604 with nonzero total read count
adjusted p-value < 0.05
LFC > 0 (up) : 0, 0%
LFC < 0 (down) : 83, 0.29%
outliers [1] : 0, 0%
low counts [2] : 11646, 41%
(mean count < 13)
```

The volcano plot demonstrated the overall differences between male and female genes. The x-axis was the levels of log2 fold change, while the y-axis showed the -log10 transformation

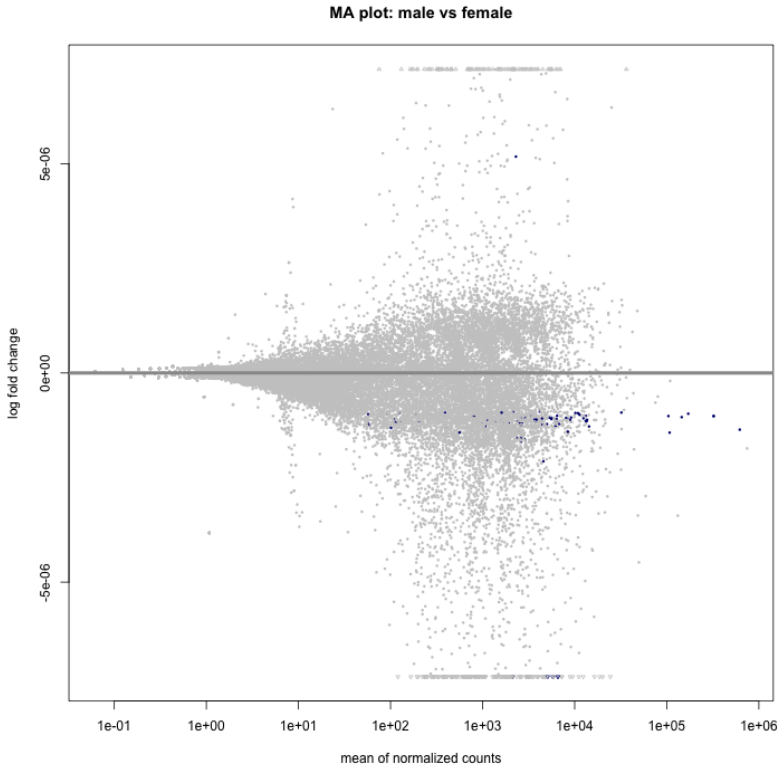
[5] Results appended to the end of the report.

of the adjusted p-values. The threshold was set to interpret genes with adjusted p-values below 0.05 and log fold change levels over 2. Since we were comparing male genes to female genes, the shape of the plot was generally one-sided.

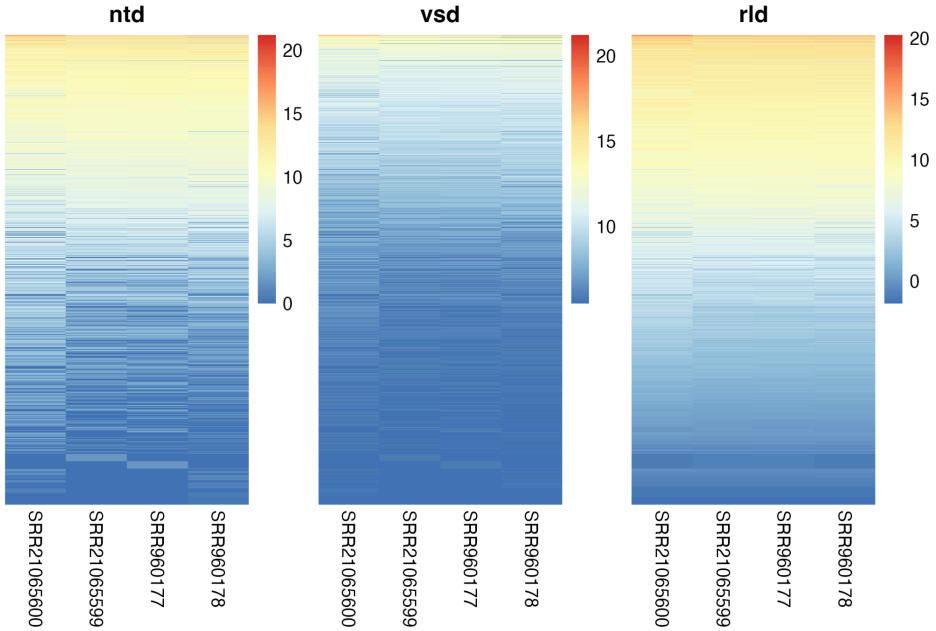
```
res_vol <- res %>% data.frame() %>%
 dplyr::mutate(threshold_OE = padj < 0.05 & abs(log2FoldChange) >= 1)
```




The following MA plot showed the mean of normalized counts on the x-axis and log fold change on the y-axis. The genes that have a p-value of less than 0.05 or a log fold change of over 2 are highlighted in dark blue. The negative values of the fold change levels illustrated that male mouse genes are less expressed than female mouse genes.




The heatmaps of the normal transformation, variance stabilizing transformation, and regularized log transformation of the counts compared the gene levels of the 4 samples altogether. The following plot revealed a significant difference of gene levels between individuals from the same sex, while the difference between the two sex was hardly observed. These results were in accordance with the quality check outputs from *FastQC* to some degree.



## Comparison & conclusion

Similarities:

Both approaches start with the same foundational data—FASTQ files containing raw sequencing data with nucleotide sequences and quality scores. Quality control is a critical early step in both pipelines, performed using [FastQC] to visualize and ensure that the sequencing data is free from low-quality data reads or contamination. Both pipelines use *TrimGalore!* to remove adapter sequences and low-quality bases. These cleaning steps are crucial to the accuracy of later quantification, regardless of the method used.

Differences:

*Bowtie2* requires full alignment of reads to a reference genome, producing SAM and BAM files, which are necessary for creating a count matrix using *featureCounts*. In contrast, Salmon bypasses the alignment step entirely, then uses a quasi-mapping method to directly quantify transcript abundances.

Salmon can seamlessly process both paired-end and single-end reads within the same workflow, while Bowtie2 requires separate processing of these read types before merging them for downstream analysis.

In the *Bowtie2* approach, gene-level counts are generated from the aligned reads using featureCounts, while in the *Salmon* approach, transcript-level data is aggregated to the gene level using *tximport* before differential expression analysis.

In the *Bowtie2* approach, *DESeq2* analysis begins with gene counts that are derived from aligned reads. In the *Salmon* approach, it starts with gene counts derived directly from the quasi-mapping quantification process, bypassing the alignment step entirely.

In general, RNA sequencing analysis proceeds from quality check of the raw data to the alignment of the reads with the genome assembly and then to the differential expression analysis. FastQC provides a comprehensive quality check during the pre-processing phase. Both Salmon and Bowtie2 are effective and productive data processing tools to prepare data for analysis in DESeq2. Gathering outputs from these tools, we are able to conclude that the difference of gene expression is observable between individuals, but more effort is still required to reach a confident conclusion of different gene expressions between different sex. Future research could take a closer look at the genes that have significant fold changes and include more samples of different experiments on the same type of cells into the analysis.

**Contribution**

Background: Chuning
Process: Jingyi & Chuning
Methods & pipelines, Results:
    FastQC & Salmon – Chuning
    Bowtie2 & DESeq2 – Jingyi
Comparison: Chuning
Conclusion: Jingyi

## Appendix

Table: Results of Differential Expression Analysis (∂ = .05)

| | baseMean | log2FoldChange | lfcSE | stat | pvalue | padj |
|---|---|---|---|---|---|---|
| Rprl3 | 32150.2767 | -10.719348 | 1.7895737 | -5.9898894 | 2.09983848496838e-09 | 3.32803401482639e-05 |
| Snord17 | 6437.17303 | -10.236746 | 1.74548493 | -5.8647005 | 4.49944048236045e-09 | 3.56558161024654e-05 |
| H4c12 | 10175.3681 | -10.254058 | 1.78463142 | -5.7457569 | 9.15107238993471e-09 | 4.83451154360251e-05 |
| Rpph1 | 171109.127 | -9.972466 | 1.75471466 | -5.6832408 | 1.32165938807923e-08 | 5.23674491041694e-05 |
| H4c1 | 5621.4264 | -9.3118136 | 1.68772755 | -5.5173678 | 3.4411472574648e-08 | 9.31400202870624e-05 |
| H4c14 | 1596.61718 | -9.2485734 | 1.68615901 | -5.4849948 | 4.13481651382999e-08 | 9.31400202870624e-05 |
| H4c18 | 13240.9925 | -9.6626591 | 1.76324329 | -5.4800487 | 4.25208810514459e-08 | 9.31400202870624e-05 |
| H4c8 | 14319.0485 | -9.5007104 | 1.74533875 | -5.4434764 | 5.22506512075514e-08 | 9.31400202870624e-05 |
| H4c17 | 11264.6588 | -9.5293099 | 1.7512899 | -5.4413092 | 5.28904147002058e-08 | 9.31400202870624e-05 |
| Rn7s2 | 322399.431 | -9.1977978 | 1.70313616 | -5.4005064 | 6.64530244819817e-08 | 9.52738003113598e-05 |
| H4c6 | 10870.3805 | -9.4388589 | 1.75093565 | -5.3907515 | 7.01636518641812e-08 | 9.52738003113598e-05 |
| Rn7s1 | 322984.858 | -9.1732453 | 1.70518322 | -5.3796244 | 7.46413951965524e-08 | 9.52738003113598e-05 |
| H2bc3 | 5578.26167 | -8.9507753 | 1.66639078 | -5.3713543 | 7.81474795916258e-08 | 9.52738003113598e-05 |
| Scarna10 | 806.723398 | -9.899373 | 1.87915509 | -5.2679915 | 1.37924506392211e-07 | 0.00015614 |
| H3c7 | 4422.04768 | -8.741505 | 1.67649188 | -5.2141649 | 1.8464720021891e-07 | 0.00019484 |
| H3c2 | 4438.25728 | -8.3854282 | 1.61182808 | -5.2024334 | 1.96695836400066e-07 | 0.00019484 |
| H3c3 | 6197.14082 | -8.0473834 | 1.5506168 | -5.1897951 | 2.10525660104027e-07 | 0.00019627 |
| H2bc14 | 3366.68957 | -8.9469121 | 1.73472744 | -5.1575319 | 2.50226252123503e-07 | 0.00021861 |
| Gm26448 | 3114.9722 | -8.0978657 | 1.57274849 | -5.1488625 | 2.62070949576225e-07 | 0.00021861 |
| H4c16 | 4571.79459 | -9.1746057 | 1.78827068 | -5.1304345 | 2.89074096326196e-07 | 0.00022908 |
| H4c3 | 3656.88536 | -8.5221212 | 1.66747113 | -5.1108058 | 3.20787532616268e-07 | 0.0002421 |
| Snord3a | 12449.3503 | -8.3838305 | 1.65923617 | -5.0528253 | 4.35322052030046e-07 | 0.00028638 |
| H4c4 | 8349.68852 | -8.3666896 | 1.65636861 | -5.0512245 | 4.38986831699803e-07 | 0.00028638 |
| Rn7sk | 104109.586 | -8.5968853 | 1.70274874 | -5.0488279 | 4.44528969485246e-07 | 0.00028638 |
| H2bc24 | 5081.53334 | -8.2131641 | 1.6300844 | -5.0384901 | 4.69218741176872e-07 | 0.00028638 |
| H2bc12 | 6354.2386 | -8.336719 | 1.65468398 | -5.0382545 | 4.69796493034952e-07 | 0.00028638 |
| H4c11 | 4181.14286 | -8.6215569 | 1.73121363 | -4.9800653 | 6.35628254031038e-07 | 0.00036795 |
| H2bc15 | 1804.93734 | -8.2453117 | 1.65710838 | -4.9757226 | 6.50046756028562e-07 | 0.00036795 |
| Rmrp | 145206.473 | -8.5184476 | 1.72607911 | -4.9351432 | 8.00919798970545e-07 | 0.00042422 |
| H4c9 | 6340.59271 | -7.8866083 | 1.59842439 | -4.933989 | 8.05669871016404e-07 | 0.00042422 |
| H4c2 | 4839.49735 | -8.704886 | 1.76632947 | -4.9282346 | 8.29759552390111e-07 | 0.00042422 |
| H2bc18 | 6153.99557 | -8.0431188 | 1.63695287 | -4.91347 | 8.94784178487819e-07 | 0.00044155 |
| H2bc22 | 2265.07467 | -7.4832824 | 1.52548566 | -4.9055082 | 9.31857570578799e-07 | 0.00044155 |
| H1f3 | 8063.40102 | -8.2248875 | 1.67776221 | -4.9022963 | 9.47227985111506e-07 | 0.00044155 |
| H1f5 | 3918.13291 | -8.241501 | 1.68461841 | -4.8922064 | 9.97117799763995e-07 | 0.00045152 |
| H2bc11 | 2517.46391 | -7.7803057 | 1.59452055 | -4.8794013 | 1.06408344943051e-06 | 0.00046846 |
| Mir6516 | 3711.43919 | -7.8147294 | 1.60533262 | -4.8679814 | 1.12743919606511e-06 | 0.00048294 |
| H3c10 | 3865.00439 | -8.0255817 | 1.65233379 | -4.8571189 | 1.19106080469772e-06 | 0.00049677 |
| H1f1 | 5308.91076 | -8.0957824 | 1.68959725 | -4.7915457 | 1.65501299376972e-06 | 0.00067257 |
| H3c4 | 2637.89964 | -7.4556587 | 1.56487962 | -4.7643656 | 1.89448721725937e-06 | 0.00075064 |

| | | | | | | |
|---|---|---|---|---|---|---|
| H3c6 | 6063.13974 | -7.7107499 | 1.62974827 | -4.7312521 | 2.23139242054923e-06 | 0.00086257 |
| H2bc8 | 3855.72599 | -7.5794012 | 1.60693999 | -4.7166673 | 2.39739295087122e-06 | 0.00088604 |
| H3c1 | 2670.494 | -7.4876549 | 1.58786224 | -4.715557 | 2.41050402439827e-06 | 0.00088604 |
| H2bc23 | 5005.16889 | -7.465092 | 1.58446378 | -4.7114312 | 2.45983191774001e-06 | 0.00088604 |
| Snord3b2 | 13482.7862 | -7.5347511 | 1.60669003 | -4.6896109 | 2.73725060219141e-06 | 0.00096406 |
| H3c14 | 1936.41927 | -7.4961612 | 1.6080298 | -4.6617054 | 3.135999017832e-06 | 0.00108049 |
| H2bc13 | 2618.64486 | -7.3539456 | 1.58825606 | -4.6302015 | 3.6531007716856e-06 | 0.00123187 |
| H2ac19 | 9138.51689 | -7.7446523 | 1.67542681 | -4.6224951 | 3.79151701520023e-06 | 0.00125191 |
| mt-Rnr2 | 620233.247 | -6.7518875 | 1.46632367 | -4.6046365 | 4.131868918949e-06 | 0.00133645 |
| H2bc7 | 985.339285 | -7.3208115 | 1.59387769 | -4.5930824 | 4.3674658730823e-06 | 0.0013844 |
| H1f2 | 6750.82897 | -7.1591878 | 1.5625553 | -4.581718 | 4.61171494192646e-06 | 0.00143316 |
| Scarna2 | 396.490495 | -7.3596985 | 1.60883522 | -4.5745508 | 4.7724270240982e-06 | 0.0014351 |
| H3c13 | 2888.86656 | -7.6798459 | 1.67924754 | -4.5733852 | 4.79906574516097e-06 | 0.0014351 |
| H2bc6 | 2922.30944 | -7.600282 | 1.68729045 | -4.5044302 | 6.65512787456764e-06 | 0.00195328 |
| Snord3b4 | 13629.5749 | -7.1762391 | 1.61612071 | -4.4404104 | 8.97874942343967e-06 | 0.00250251 |
| Snora73b | 8435.5169 | -6.4167948 | 1.44522708 | -4.4399907 | 8.9962771637007e-06 | 0.00250251 |
| Snord3b3 | 13170.6219 | -7.1403301 | 1.6082191 | -4.4398989 | 9.00011686793607e-06 | 0.00250251 |
| H2ac18 | 8937.14992 | -7.2084386 | 1.62595171 | -4.4333657 | 9.27732857313764e-06 | 0.00253511 |
| H3c8 | 2300.04159 | -7.6043554 | 1.71700806 | -4.4288408 | 9.47409096219033e-06 | 0.002545 |
| H3c11 | 3361.37075 | -7.1577512 | 1.62035965 | -4.4173843 | 9.9902527607474e-06 | 0.00263893 |
| Terc | 1608.02175 | -7.8543212 | 1.79981532 | -4.3639595 | 1.27729223463082e-05 | 0.00331577 |
| H2ac15 | 2368.10589 | -6.9970272 | 1.60460438 | -4.3605933 | 1.29710230225183e-05 | 0.00331577 |
| H2af-ps2 | 3709.3349 | -7.2058547 | 1.67505118 | -4.3018714 | 1.6936161505134e-05 | 0.00426065 |
| Snora74a | 561.246954 | -6.2563 | 1.4631772 | -4.275832 | 1.90424758826771e-05 | 0.00471569 |
| H2bc4 | 5782.6654 | -6.5616858 | 1.54647064 | -4.243007 | 2.20544511393698e-05 | 0.00537755 |
| H2bc9 | 1117.38433 | -7.0641513 | 1.691811 | -4.1754967 | 2.97336558959352e-05 | 0.00714013 |
| Aspn | 56.5678765 | -7.8728233 | 1.88822624 | -4.169428 | 3.05365115235985e-05 | 0.00722348 |
| H2ac20 | 2162.0272 | -7.5183896 | 1.81087482 | -4.1517997 | 3.29870889708412e-05 | 0.00768842 |
| H3c15 | 1685.95044 | -6.4742645 | 1.57400007 | -4.1132555 | 3.90118187544209e-05 | 0.00896085 |
| Gm34885 | 57.4884367 | -8.4803756 | 2.07363613 | -4.089616 | 4.32087858554413e-05 | 0.00967504 |
| Snora73a | 6577.95125 | -6.118178 | 1.4962888 | -4.0889018 | 4.33420254430454e-05 | 0.00967504 |
| Gm34389 | 59.1529261 | -7.22209 | 1.80359468 | -4.0042755 | 6.22078312850886e-05 | 0.0136935 |
| H1f4 | 1067.69494 | -6.096442 | 1.52755577 | -3.9909784 | 6.58012851944181e-05 | 0.01428609 |
| Gm25360 | 1392.2355 | -6.3408955 | 1.59975718 | -3.9636612 | 7.38089824016791e-05 | 0.01580809 |
| mt-Rnr1 | 107329.605 | -5.5063176 | 1.41690748 | -3.8861518 | 0.00010185 | 0.02135403 |
| H2ac6 | 2085.89992 | -5.1376439 | 1.3224859 | -3.8848383 | 0.0001024 | 0.02135403 |
| Gm24339 | 128.119553 | -6.7649498 | 1.77078906 | -3.8203025 | 0.00013329 | 0.02743485 |
| Ecm2 | 93.380794 | -7.4592606 | 1.95792286 | -3.8097827 | 0.00013909 | 0.02826182 |
| H2ac4 | 2656.3552 | -5.1135169 | 1.35949615 | -3.7613324 | 0.00016901 | 0.03390696 |
| Gm24119 | 192.655963 | -5.5776377 | 1.49796269 | -3.7234824 | 0.00019649 | 0.03892784 |
| Gm41607 | 188.323685 | -5.8535495 | 1.58795544 | -3.6862177 | 0.00022761 | 0.04453601 |
| 2310067E19Rik | 62.7950156 | -7.7302279 | 2.10623571 | -3.6701628 | 0.0002424 | 0.04637262 |
| Gm35566 | 35.0783999 | -7.7679956 | 2.11680195 | -3.6696846 | 0.00024285 | 0.04637262 |
| Lrrtm2 | 108.79967 | -6.4334206 | 1.75877303 | -3.6579027 | 0.00025429 | 0.04797861 |