Bulk RNA-seq Paper

Rika Chan

November 2024

1 Introduction

Brown Adipose Tissue (BAT) functions as a thermogenesis, producing heat and increasing metabolic rate (Marlatt et al., 2017). BAT is originated from the dermomyotome compartment of somites, whose tripotency also differentiates into skeletal muscles and dermal mesenchymes. Being a source of metabolic-related functions, BAT as a therapeutic treatment for obesity and other metabolic disorders is an active area of research. There are three major BAT depots in mice: interscapular (iBAT), scapular (sBAT), and cervical (cBAT). There are many signaling pathways and genes involved in the differentiation and development of BAT in mice, one of which is the *Hoxa5* gene. *Hoxa5* is a part of the *Hox* family genes, whose functions in embryonic development regulates many bodily formations through downstream activation of effectors (Akam, 1987). *Hoxa5*, in particular, has been known to promote BAT differentiation from BAT preadipocytes (Cao et al., 2018). However, while *Hoxa5* is not necessary in BAT differentiation, the lack of its expression in cells reduced BAT formation significantly, specifically in iBAT and sBAT (Holzman et al., 2021).

RNA-seq is a popular tool in analyzing differences in gene expression levels in different samples or groups using traditional statistical methods. Common methods used in RNA-seq analysis are DEseq2 (Love et al., 2014) and edgeR (Chen et al., 2024). RNA-seq is particularly prominent in understanding developmental biology. Our understanding of the field revolved around embryonic growth and organ formations through regulations of specific genes (Potter, 2018). For instance, Takahashi and Yamanaka (2006) induced de-differentiation of adult fibroblasts into stem cells through forced expression of four transcription factors.

Here, we showed a list of differentially expressed genes (DEGs) through RNA-seq analysis using two methods, DEseq2 and edgeR, in *Hoxa5*-/- sBAT cells. We found that edgeR yielded more DEGs with a false discovery rate of less than 0.05 (n=5) than DEseq2 (n=1), excluding the control gene *Hoxa5*, making edgeR a less stringent approach. In both sets of DEGs, only *Ddx3y* gene was present in both, suggesting a high significance despite the different methods. In both sets of DEGs, there was a particular interest in Y-linked genes, so a secondary analysis of DEGs was performed with known Y-linked genes removed. Interestingly, this analysis yielded no significant DEGs besides the control, *Hoxa5* from both methods, suggesting the difference in gene expressions between *Hoxa5*-/- and *Hoxa5*+/+ sBAT cells were mainly driven by the presence of Y-linked genes.

2 Methods

Data Processing. Raw counts data of BAT was retrieved. Data was filtered to sBAT data, and pre-analysis processing consisted of filtering out genes with counts less than 10.

DEG Analysis. Bioconductor packages DESeq2 and edgeR were used to perform DEG analysis on filtered sBAT data in R (version 4.4.1). Bioconductor EnhancedVolcano package was used to generate volcano plots, and pheatmap was used to generate heatmaps. DEseq2 used adjusted p-value (p-adj) and edgeR used false discovery rate (FDR) as their measure of significance. For consistency, FDR is used in this paper to refer to the measures of significance from both methods. Log2FoldChange is abbreviated as log2FC. For analysis of removed Y-linked genes, database was retrieved through BioMart from Ensembl website. Hoxa5-/- cells were denoted as mt samples and Hoxa+/+ cells as WT samples.

Pathway Analysis. Panther Pathway was used to retrieve common pathways the top DEGs were known to be involved in. Panther Pathway database was used in this analysis.

3 Results

Diagnostics plots were generated to visualize the data structures. Volcano plots and heatmaps were generated to visualize DEG analysis from both methods. *Hoxa5* gene was used a threshold guidance by arbitrarily determining the lowest possible threshold that would yield significance for the known downregulated gene.

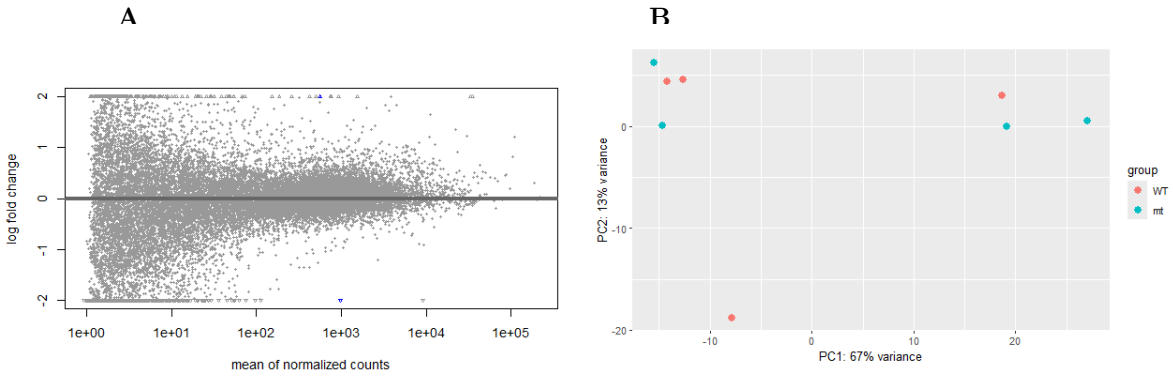

Figure 1: **Diagnostic plots.** (A) MA plot of all samples. Significant genes were marked in blue. (B) Principle Component Analysis (PCA) plot of the samples color coded by groups.

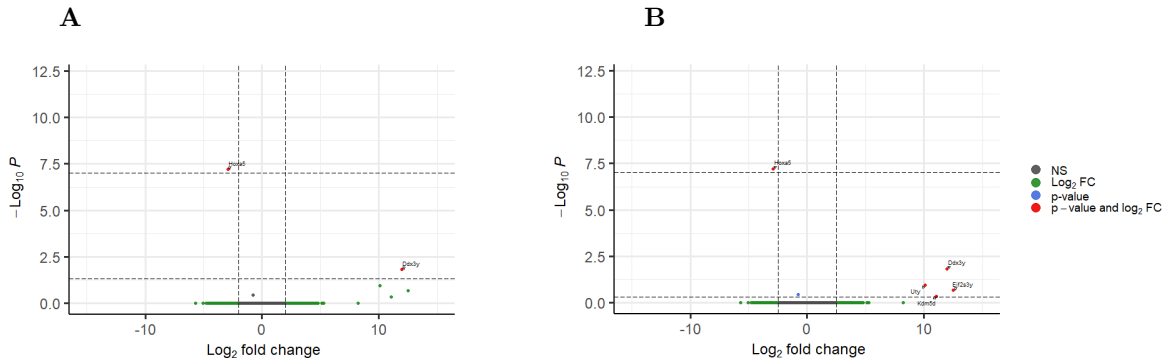

Figure 2: **Differentially Expressed Genes Analysis from DEseq2.** (A) Volcano plot of DEGs with FDR cutoff of 0.05 and log2FC cutoff of 2. (B) Volcano plot of DEGs with FDR cutoff of 0.5 and log2FC of 2.5.

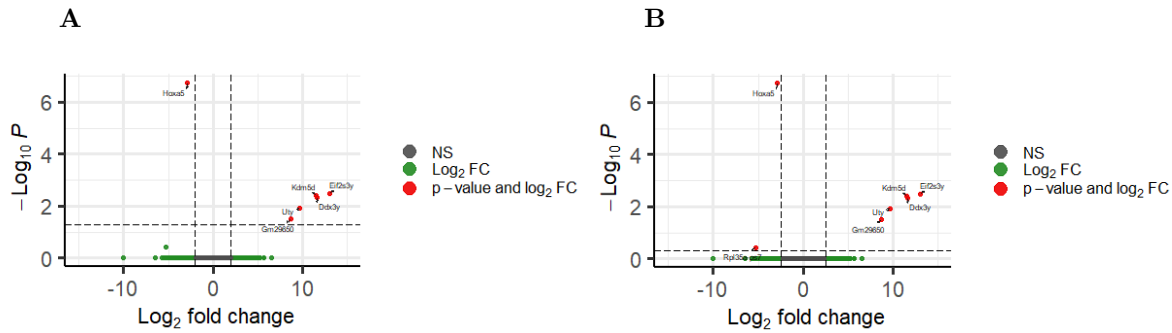

Figure 3: **Differentially Expressed Genes Analysis from edgeR.** (A) Volcano plot of DEGs with FDR cutoff of 0.05 and log2FC cutoff of 2. (B) Volcano plot of DEGs with FDR cutoff of 0.5 and log2FC of 2.5.

The volcano plots in figures 1 and 2 looked out of ordinary compared to a typical volcano plot due to the rank-based nature of the p-value correction method, which caused genes with similar raw p-values to collapse together and retain the same p-adj or FDR, as seen in genes marked in green. For genes marked in red, which signified that it passed the threshold set for both FDR and log2FC, there were less in numbers resulted from DEseq2 than edgeR, suggesting that DEseq2 is a more conservative method, even with a more lenient threshold of FDR and log2FC (Fig. 2B).

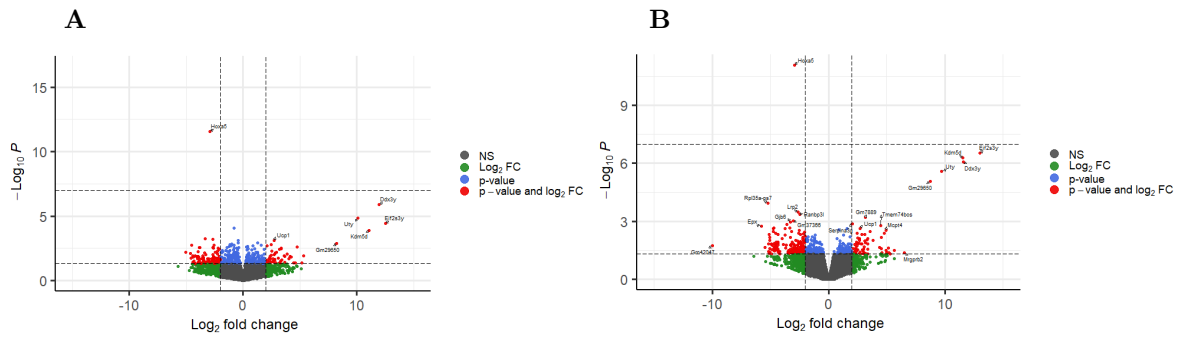

Figure 4: **DEGs Analysis from DEseq2 and edgeR.** (A) Volcano plot of DEGs with raw p-value cutoff of 0.05 and log2FC cutoff of 2 from DEseq2. (B) Volcano plot of DEGs with raw p-value cutoff of 0.05 and log2FC of 2 from edgeR.

Volcano plots generated using the p-value instead of the FDR yielded a more dynamic representation of the genes, though not necessarily a more reliable measure of significance. However, in the p-value plots, it was consistently observed that DESeq2 was more stringent than edgeR, as indicated by the number of genes marked in red (Fig. 3).

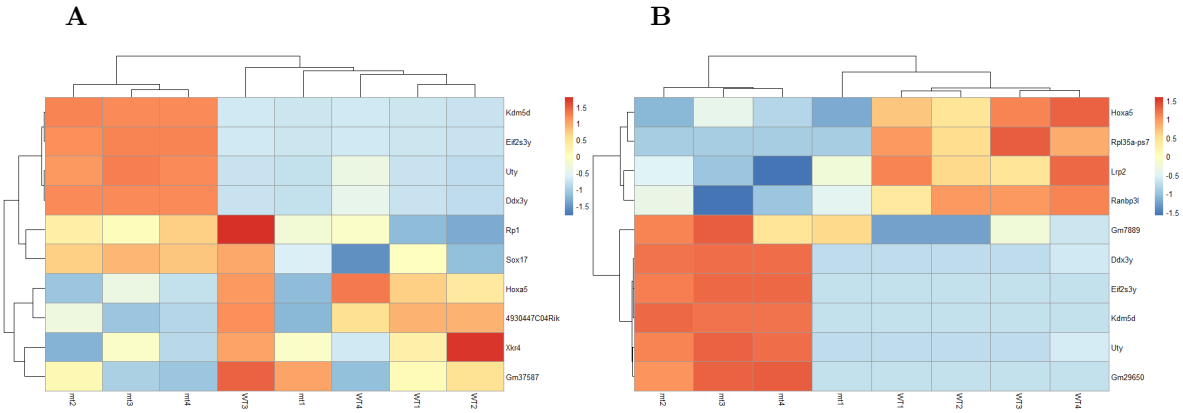

Figure 5: **Differentially Expressed Genes Analysis from DEseq2 and edgeR.** (A) Heatmap of top 10 DEGs from DEseq2 sorted by FDR. (B) Heatmap of top 10 DEGs from edgeR sorted by FDR. Not all 10 genes in both sets were below 0.05 FDR.

Both methods seem to have clustered the samples slightly differently, but still retained the observation that mt1 is more similar to the WT samples (Fig. 4). Genes such as *Ddx3y, Eif2s3y, Uty, Kdm5d, Gm29650* were shown to have upregulated compared to WT samples, and shown to be clustered together. These were also the genes that have shown significance in expression from either DEseq2 or edgeR approach (Table 1 & 2). Between the two methods, edgeR showed a clearer distinction in expression levels of *Hoxa5* between the WT and mt samples (Fig. 4B). edgeR heatmap also showed the expected clustering of samples, which allowed for a better readability between the two groups in terms of expression levels, suggesting that if DEseq2 was performed alone, there could be a misguided inter-

pretation that the two groups were more similar than not. Although, this observation was inconsistent with the PCA plot, which showed a high dissimilarity within groups (Fig. 1B).

Table 1: **Pathway Analysis Result of top DEGs from DEseq2.** $**$ < 0.05 FDR. Non-significant genes were included to show DEseq2 result dynamic as these six genes were the only genes that have unique FDR values; the rest of the genes have FDR value of 9.999×10^{-1}. - represented unknown.

Gene name	**FDR**	**Panther Family/Subfamily**	**Panther Protein Class**
Hoxa5**	5.94×10^{-8}	Homeobox Protein Hox-a5	Homeodomain Transcription Factor
Ddx3y**	1.47×10^{-2}	ATP-dependent RNA Helicase	RNA Helicase
Uty	1.11×10^{-1}	Histone Demethylase	Chromatin/Chromatin-binding/Regulatory Protein
Eif2s3y	2.08×10^{-1}	Eukaryotic Translation Initiation Factor 2 Subunit 3	Translation Initiation Factor
Kdm5d	4.70×10^{-1}	Lysine-specific Demethylase 5D	Histone Modifying Enzyme
4930447C04Rik	3.75×10^{-1}	Protein Six6OS1	-

Table 2: **Pathway Analysis Result of top DEGs from edgeR.** Genes are listed in order of most to least significance. All genes were < 0.05 FDR. - represented unknown.

Gene name	**FDR**	**Panther Family/Subfamily**	**Panther Protein Class**
Hoxa5	1.89×10^{-7}	Homeobox Protein Hox-a5	Homeodomain Transcription Factor
Eif2s3y	3.26×10^{-3}	Eukaryotic Translation Initiation Factor 2 Subunit 3	Translation Initiation Factor
Kdm5d	3.85×10^{-3}	Lysine-specific Demethylase 5D	Histone Modifying Enzyme
Ddx3y	4.63×10^{-3}	ATP-dependent RNA Helicase	RNA Helicase
Uty	1.17×10^{-2}	Histone Demethylase	Chromatin/Chromatin-binding/Regulatory Protein
Gm29650	3.18×10^{-2}	-	-

4 Discussion

edgeR seemed to be a less stringent method compared to DEseq2, yielding more DEGs (n=5) than DEseq2 (n=1), excluding the control gene *Hoxa5*. In a comparison between eight methods for RNA-seq, edgeR was observed to have a low ranking among all the methods in simulated data of n=6 and n=3 samples per group, while ranking highest with real data in terms of yielding the number of significant genes (Li et al., 2022). They also observed that there was a different yield in the number of DEG identified among the negative binomial models, which both DEseq2 and edgeR deployed. Perhaps, the decision for which one to use may depend on how much researchers are willing to tolerate the result.

Before the removal of Y-linked genes, *Ddx3y* gene was the only gene shown as DEG from both DEseq2 and edgeR. *Ddx3y* is a Y-linked gene located in the AZFa region of the Y chromosome. In a paper identifying key metabolic genes, they observed a sexual dimorphism in metabolic responses between male and female mice (Savva et al., 2022). Male mice were observed to have upregulated Y-linked genes such as *Ddx3y* and *Eif2s3y*, which were involved in embryonic development, while female mice possess a higher expression of *Eif2s3x, Kdm6a, Kdm5c* and *Ddx3x*, which were known as potential regulators of adiposity (Link et al., 2020). Interestingly, yet perhaps unfortunately, the DEGs resulted from both methods were driven by the presence of Y-linked genes. It added another layer of complexity in exploring the RNA-seq data, which requires further experiments and analyses. Although, another avenue that would be interesting to explore is how *Hoxa5* gene interact with these Y-linked genes, if at all.

Although data not shown, DEG analysis with known Y-linked genes removed showed no significant

DEG, except for the control *Hoxa5*, suggesting Y-linked genes were significant contributors to the differential gene expressions initially observed. This could suggest that there are underlying mechanisms of *Hoxa5* downregulation that interacted with these Y-linked genes to induce a sex-specific expression pattern. Interestingly, females mice have shown to contain less triacylglycerol and more phospholipids in certain regions than males, suggesting that there were sex-specific regulation of BAT functions (Hoene et al., 2014). Although this is a knowledge gap, if there are sex-specific responses to BAT functions, it is reasonable to hypothesize that there are also sex-specific responses in BAT differentiation and development, and this could become an active area of research.

Although there were no significant pathways shown when performing pathway analysis, Fgf signaling pathway was chosen to described further because of its significance to BAT differentiation and formation. The Fibroblast Growth Factor (Fgf) is a polypeptide family, that when activated in a signaling pathway, controls many developmental processes (Thisse, 2005). It has a role in processes such as but not limited to cell proliferation, embyrogenesis, and differentiation (Seitz et al., 2012). Although not directly regulating BAT formation, Fgf signaling encouraged browning of White Adipose Tissue (WAT), forcing it to adopt brown fat-like characteristics (Fisher et al., 2012). Dysregulation in the signaling pathway was associated with many cancers, suggesting a therapeutic potential (Turner et al., 2010). The pathway is initiated by the binding of FGF ligands to FGF receptors, which then induces phosphorylation of phosphotyrosine groups. These groups act as sites for various downstream signaling activities, one of the most commonly known is the Ras/MAPK pathway, which controls cell proliferation (Teven et al., 2014).

5 References

1. Akam M. (1987). The molecular basis for metameric pattern in the Drosophila embryo. Development (Cambridge, England), 101(1), 1–22.

2. Cao, W., Xu, Y., Luo, D., Saeed, M., Sun, C. (2018). Hoxa5 Promotes Adipose Differentiation via Increasing DNA Methylation Level and Inhibiting PKA/HSL Signal Pathway in Mice. Cellular physiology and biochemistry : international journal of experimental cellular physiology, biochemistry, and pharmacology, 45(3), 1023–1033. https://doi.org/10.1159/000487343

3. Chen Y, Chen L, Lun ATL, Baldoni P, Smyth GK (2024). "edgeR 4.0: powerful differential analysis of sequencing data with expanded functionality and improved support for small counts and larger datasets." bioRxiv. doi:10.1101/2024.01.21.576131.

4. Fisher, F. M., Kleiner, S., Douris, N., Fox, E. C., Mepani, R. J., Verdeguer, F., Wu, J., Kharitonenkov, A., Flier, J. S., Maratos-Flier, E., Spiegelman, B. M. (2012). FGF21 regulates PGC-1 and browning of white adipose tissues in adaptive thermogenesis. Genes development, 26(3), 271–281. https://doi.org/10.1101/gad.177857.111

5. Hoene, M., Li, J., Häring, H.-U., Weigert, C., Xu, G., Lehmann, R. (2014). The lipid profile of brown adipose tissue is sex-specific in mice. Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids, 1841(10), 1563–1570. https://doi.org/10.1016/j.bbalip.2014.08.003

6. Holzman, M. A., Ryckman, A., Finkelstein, T. M., Landry-Truchon, K., Schindler, K. A., Bergmann, J. M., Jeannotte, L., Mansfield, J. H. (2021). HOXA5 Participates in Brown Adipose

Tissue and Epaxial Skeletal Muscle Patterning and in Brown Adipocyte Differentiation. Frontiers in cell and developmental biology, 9, 632303. https://doi.org/10.3389/fcell.2021.632303

7. Li, D., Zand, M. S., Dye, T. D., Goniewicz, M. L., Rahman, I., Xie, Z. (2022). An evaluation of RNA-seq differential analysis methods. PloS one, 17(9), e0264246. https://doi.org/10.1371/journal.pone.0264246

8. Link JC, Wiese CB, Chen X, Avetisyan R, Ronquillo E, Ma F, et al. X chromosome dosage of histone demethylase KDM5C determines sex differences in adiposity. J Clin Invest. 2020;130:5688–702. doi: 10.1172/JCI140223.

9. Love MI, Huber W, Anders S (2014). "Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2." Genome Biology, 15, 550. doi:10.1186/s13059-014-0550-8.

10. Marlatt, K. L., Ravussin, E. (2017). Brown Adipose Tissue: an Update on Recent Findings. Current obesity reports, 6(4), 389–396. https://doi.org/10.1007/s13679-017-0283-6

11. Potter S. S. (2018). Single-cell RNA sequencing for the study of development, physiology and disease. Nature reviews. Nephrology, 14(8), 479–492. https://doi.org/10.1038/s41581-018-0021-7

12. Savva, C., Helguero, L. A., González-Granillo, M., Melo, T., Couto, D., Buyandelger, B., Gustafsson, S., Liu, J., Domingues, M. R., Li, X., Korach-André, M. (2022). Maternal high-fat diet programs white and brown adipose tissue lipidome and transcriptome in offspring in a sex- and tissue-dependent manner in mice. International journal of obesity (2005), 46(4), 831–842. https://doi.org/10.1038/s41366-021-01060-5

13. Seitz I.A., Teven C.M., Reid R.R., editors; Neligan P.C., Gurtner G.C., editors. Repair and Grafting of Bone. vol. 1. Elsevier Saunders; Philadelphia, PA: 2012. (Plastic Surgery).

14. Teven, C. M., Farina, E. M., Rivas, J., Reid, R. R. (2014). Fibroblast growth factor (FGF) signaling in development and skeletal diseases. Genes diseases, 1(2), 199–213. https://doi.org/10.1016/j.gendis.2014.09.005

15. Thisse B., Thisse C. Functions and regulations of fibroblast growth factor signaling during embryonic development. Dev Biol. Nov 15 2005;287:390–402. doi: 10.1016/j.ydbio.2005.09.011

16. Turner N., Grose R. Fibroblast growth factor signalling: from development to cancer. Nat Rev Cancer. Feb 2010;10:116–129. doi: 10.1038/nrc2780.