**Transcriptional Profile of Mammalian Cardiac Regeneration with mRNA-Seq**

**Individual Project: Liying (Ariel) Xue**

**Group wheeler**

Project 2 - Analyst & Biologist

Github repository: <https://github.com/ariel208x/bf528IndividualProject>

**Introduction**

The capacity for neonatal mice to repair their damaged heart tissue disappears after the first few weeks of birth1. To better understand the underlying genes and pathways that participate in mammalian cardiac regeneration, transcriptomes of mice at different time points after birth were analyzed. After aligning mRNA sequencing reads to the mm9 mouse genome and assembling the transcripts, differentially expressed genes (DEG) between postnatal day 0 and adult mice were obtained with Cufflinks2. Both up- and down- regulated genes during the loss of regeneration ability were identified using their log2 fold changes. Their related enrichment terms and enriched pathways were also analyzed with DAVID3, a functional annotation clustering tool. Biological interpretations were then made by visualizing trends with certain genes of enrichment terms and performing hierarchical clustering of the top DEGs.

**Methods**

**DEG Analysis and Enriched Pathways**

All gene expression analyses were performed by Fragments Per Kilobase of transcript per Million mapped reads (FPKM), which normalizes the reads for gene length and allow the comparison among genes with different lengths. Cufflinks2 was used to compare FPKM values between postnatal day 0 and adult mice, providing the log2 fold changes, p-values, and q-values for each gene. A total of 36,329 observations from Cufflinks were used for downstream analysis. The top 10 DEGs were then identified by q-values, which measures the significance using a false discovery rate. Histograms were made for all genes and for the significant DEGs to visualize their distribution of frequency. The significant DEGs were further separated into up- and down-regulated genes and saved into files based upon log 2 fold change - with a log2FC larger than 0, the gene is considered to be up-regulated, while a log2FC lower than 0 is considered to be down-regulated. The up-regulated and down-regulated DEG names were then imported into DAVID3 using identification of official short names for functional annotation clustering and provided analysis on enriched pathways using gene ontology of BP\_FAT, CC\_FAT, and MF\_FAT.

**Biological Patterns Interpretation**

In order to observe the trend over time and magnitude of enriched pathways, previously obtained FPKM table for postnatal day 0 was combined with other 7 FPKM tables - 2 for postnatal day 4; 2 for postnatal day 7; 2 for adults; and 1 for another sample of postnatal day 0. FPKM values of representative genes for enrichment terms of sarcomere, mitochondria, and cell cycle were made into line plots with the ggplot2 package. Top 1000 DEGs between postnatal day 0 and adults were obtained using lowest q-value from 36,329 observations obtained with Cufflinks. The combined matrix for the 8 samples was then filtered to make a heatmap for the top 1000 DEGs only. Hierarchical clustering was performed for both genes as rows and samples as columns. The results of enriched pathways and enrichment terms from O'Meara *et al*1 were also compared with the results from DAVID analysis3.

**Results**

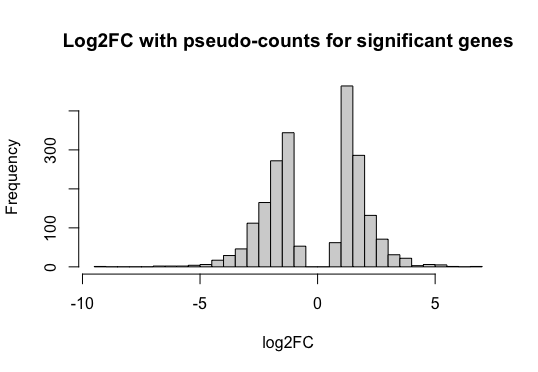
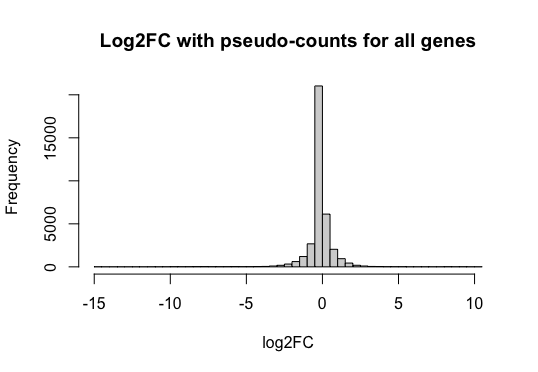
The top 10 DEGs with lowest q-value are shown in **Table 1** together with their corresponding FPKM values at postnatal day0, log 2 fold change, p-value, q-value. Histograms for all genes and for up- and down-regulated genes are shown in **Figure 1**. In order to avoid infinite log 2 fold change values resulting from FPKM values of 0, pseudo-counts were implemented by adding 1 to all FPKM values, and then transformed them into log 2 fold change. Both **Figure 1A** and **Figure 1B** are plotted with pseudo-counts. From **Figure 1A**, we can see that the majority of genes are concentrated around a log 2 fold change of 0 with a range from -5 to 5. From **Figure 1B**, it can be observed that up-regulated genes (log2FC > 0) are more concentrated around log 2 fold change of 0 and having a smaller range of values compared to down-regulated genes (log2FC < 0), while down-regulated genes are more relatively spread out with larger range of log 2 fold change values. There shows no log 2 fold change of 0 in **Figure 1B** because these genes are not significantly differentially expressed with a log2FC equals 0.

With a significance level of *p* < 0.01, a total of 1710 genes were observed as differentially expressed. Within the significant DEGs, 814 are up-regulated and 896 are down-regulated.

Results from DAVID were reported in **Table 2** and **Table 3**3. The top 4 clusters with the highest enrichment scores for up-regulated genes are shown in **Table 2**, and for down-regulated genes are shown in **Table 3**. Comparing the common up- and down- regulated gene enrichment terms reported by O'Meara *et al*’s Figure 1C1, I observed that the terms related to mitochondria and metabolism in up-regulated genes and cell cycle in down-regulated genes are consistent with the results reported by O’Meara *et al*3.

**Table 1.** Top 10 differentially expressed genes between postnatal day 0 mice and adult mice.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Gene Name** | **FPKM Value**  **at P0** | **Log2FC** | **p-value** | **q-value** |
| *PLEKHB2* | 21.59 | 1.70 | 5e-05 | 1e-03 |
| *MRPL30* | 42.54 | 1.52 | 5e-05 | 1e-03 |
| *COQ10B* | 8.65 | 2.27 | 5e-05 | 1e-03 |
| *AOX1* | 1.77 | 2.58 | 5e-05 | 1e-03 |
| *NDUFB3* | 121.14 | 1.40 | 5e-05 | 1e-03 |
| *SP100* | 3.09 | 5.56 | 5e-05 | 1e-03 |
| *CXCR7* | 5.16 | 2.70 | 5e-05 | 1e-03 |
| *LRRFIP1* | 196.05 | -2.27 | 5e-05 | 1e-03 |
| *RAMP1* | 18.19 | -4.26 | 5e-05 | 1e-03 |
| *GPC1* | 56.60 | 1.86 | 5e-05 | 1e-03 |

**(A)** **(B)**

**Figure 1**. Histogram for log 2 fold changes of all genes with breaks of 50 using pseudo-counts (**A**); histogram for both up-regulated (log2FC > 0) and down-regulated (log2FC < 0) significant genes with breaks of 50 using pseudo-counts (**B**).

**Table 2.** Enriched pathways with up-regulated genes. \* Represent consistent enrichment terms reported by O'Meara *et al*’s Figure 1C1.

|  |  |  |
| --- | --- | --- |
|  | Enrichment Terms | score |
| Cluster 1 | Mitochondrion\*  Mitochondrial part\*  Mitochondrial inner membrane\*  Mitochondrial envelope\*  Mitochondrial membrane\* | 21.34 |
| Cluster 2 | Generation of precursor metabolites and energy  Energy derivation by oxidation of organic compounds  Purine ribonucleoside metabolic process\*  Purine nucleoside metabolic process\*  Purine nucleoside monophosphate metabolic process\* | 15.29 |
| Cluster 3 | Organic acid metabolic process\*  Oxoacid metabolic process\*  Carboxylic acid metabolic process\*  Monocarboxylic acid metabolic process\*  Fatty acid metabolic process\* | 14.57 |
| Cluster 4 | Extracellular organelle  Extracellular vesicle  Extracellular exosome  Membrane-bounded vesicle  Extracellular region part | 10.79 |

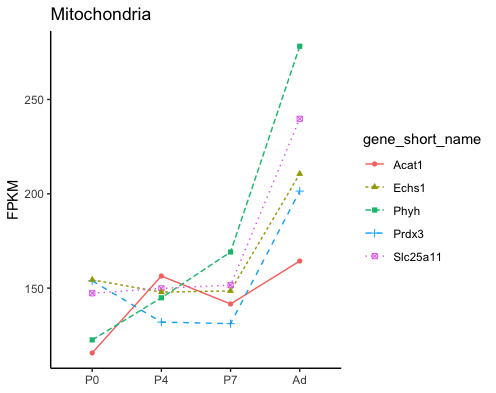
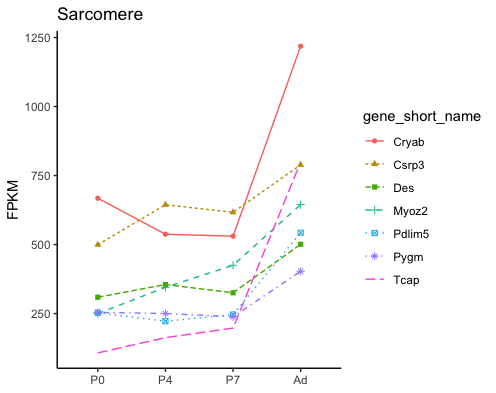
**Table 3.** Enriched pathways with down-regulated genes. \* Represent consistent enrichment terms reported by O'Meara *et al*’s Figure 1C1.

|  |  |  |
| --- | --- | --- |
|  | Enrichment Terms | score |
| Cluster 1 | Cell cycle\*  Cell division  Mitotic cell cycle\*  Cell cycle process\*  Mitotic cell cycle process\* | 11.85 |
| Cluster 2 | Proteinaceous extracellular matrix  Extracellular matrix  Extracellular matrix component | 9.69 |
| Cluster 3 | Cell proliferation  Regulation of cell proliferation  Positive regulation of cell proliferation  Negative regulation of cell proliferation | 9.16 |
| Cluster 4 | Regulation of cellular component organization  Regulation of organelle organization  Positive regulation of cellular component organization  Positive regulation of organelle organization | 8.63 |

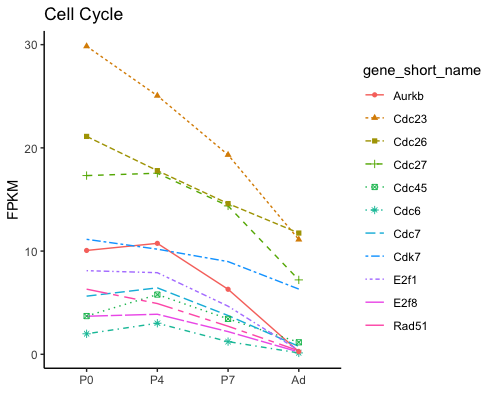
After visualizing the representative genes for enrichment terms of sarcomere, mitochondria, and cell cycle from postnatal day 0 to adult (**Figure 2**), I observed a consistent trend and magnitude with what was reported in O'Meara *et al*’s Figure 1D1 - for FPKM values in genes of sarcomere, they increased intensively from postnatal day 7 to adult, with the highest FPKM value around 1250; mitochondria genes showed similar increase between postnatal day 7 to adult with a highest FPKM value around 300; genes for cell cycles dropped constantly from postnatal day 0 to adult starting from a FPKM value around 30.

A clustered heatmap was created for the top 1000 DEGs between postnatal day 0 and adult in **Figure 3**. It can be observed that two adult samples are outliers as they are clustered together and have the largest distance with other samples. These two adult samples showed enriched expression in a good amount of genes which are not as much expressed during postnatal day 0 to day 7, and they also showed significantly decreased expression in many genes compared to postnatal day 0 to day 7. Although it is not completely consistent with O'Meara *et al*’s Figure 2A1, it shows a similar trend with clustering that samples from adults are most differentially expressed from other samples from postnatal day 0 to day 7.

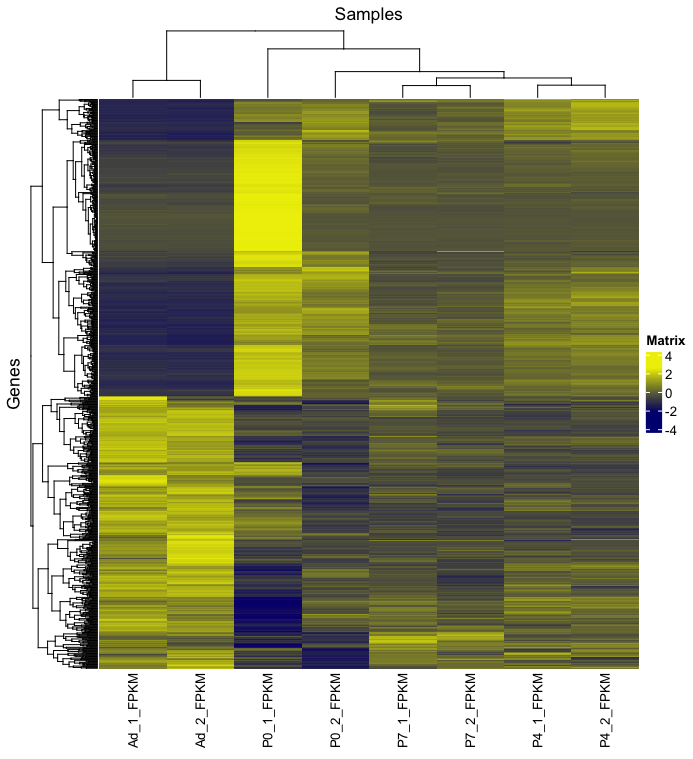
**(A)**  **(B)**



**(C)**



**Figure 2.** FPKM values of representative sarcomere, mitochondrial, and cell cycle genes significantly differentially expressed during in vivo maturation.



**Figure 3.** Clustered heatmap of FPKM values with top 1000 DE genes found in the P0 vs Adult analysis on *In vivo* heart maturation of mice.

**Discussion**

To better understand the process of neonatal mice losing the ability to regenerate damaged heart tissue, mRNA samples from postnatal day 0 to adult mice are aligned, assembled and analyzed. The top DEGs were filtered either by q-values or p-values, and then saved into files. The enriched pathways analysed by DAVID shows good similarities with what was originally reported 1 - the results show significantly enriched expression in representative genes for enrichment terms of mitochondria and metabolism, and greatly repressed expression for representative genes for enrichment terms of cell cycle. It can also be concluded that these differences in gene expression are most likely to happen during postnatal day 7 to adult by finding consistent biological patterns in line plots and clustered heatmaps for samples across different time points compared to the original report1.

**References**

1. O'Meara CC, Wamstad JA, Gladstone RA, Fomovsky GM, Butty VL, Shrikumar A, Gannon JB, Boyer LA and Lee RT. Transcriptional reversion of cardiac myocyte fate during mammalian cardiac regeneration. *Circ Res*. 2015;116:804-15.
2. Trapnell, C., Williams, B., Pertea, G. *et al.* Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol.* 2010; 28, 511–515.
3. Huang DW, Sherman BT, Lempicki RA. Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists. *Nucleic Acids Res.* 2009;37(1):1-13.