Project 2:
Transcriptional Profile of Mammalian Cardiac Regeneration Replication of O'Meara et al.

Carol Muriithi: Analyst
Cory Williams: Biologist
Nicholas Mosca: Curator
Zeyuan: Programmer
TA: Dakota Hawkins

INTRODUCTION

One major cause of heart failure is cardiac myocytes death. Adult mammalian hearts have minimal ability to recover following injury. Hypertrophy of cardiac myocyte cells stimulate growth in newly born mammals. Evidence shows that the ability for cell division to occur is limited after cardiac myocytes have fully differentiated. Other vertebrate species such as zebrafish and newts are able to regenerate hearts throughout adult life[1]. It is proven that neonatal mice can fully regenerate their hearts after sustaining injury to cardiac myocyte cells[2]. Neonatal cardiac myocyte cells reenter the cell cycle after sustaining damage resulting in reversion of damaged cells[1]. The transcriptional mechanisms that underlie this change have not been well studied.

Studying the difference in transcriptional signatures between adult mice, and neonatal mice would help identify the reason behind this reversion at the molecular level. In the paper O'Meara et al., the authors were interested in why cardiac myocyte regeneration changed over time and how it was regulated using transcriptional data [1]. Transcriptional changes and gene expression patterns during the cell cycle process have not been annotated. Within O'Meara et al. study, RNA sequencing was first performed. RNAseq analysis of neonatal cells during regeneration identifies key mechanisms that are being altered compared to adult cardiac myocyte cells.

Following RNA sequencing, Tophat was used to align the sequence reads to the mus musculus genome. The transcript abundances were estimated in fragments per kilobase per million fragments mapped (FPKM), and the most differentially expressed genes were determined using the Cuffdiff module. The statistically significant genes among the differentially expressed genes were then chosen based on the FDR adjusted p-value. Genes with FDR adjusted p-values less than 0.05 were chosen. To ensure that all the biological replicates were robust, the Pearson correlation coefficient was maintained between 0.92 and 0.98. By applying these filters, they obtained 15304 genes that passed the thresholds. Genes expressing a greater than 1 FPKM in at least 1 sample in the two replicates were then filtered. Quantile normalization was then performed using the preprocessCore package .In addition, genes with batch effects were removed. The significant genes were then clustered using Hierarchical clustering in Spotfire. The GO enrichment analysis was performed using DAVID. Within the original paper, principal component analysis (PCA) was also performed using the prcomp function without scaling and the top and bottom 500 genes were obtained [1].

Data

Sample generation

Neonatal cardiac myocyte cells were dissociated from whole mouse hearts at age (P0 and P4)[1]. Additional hearts were dissociated 7 days post surgery using the Neonatal Heart Dissociation Kit (Miltenyi Biotec)[1]. At each time point three biological replicates were isolated from five to ten mice. Heart cells from each time point were pooled for RNA extraction. Total RNA from pooled mouse cells was extracted using the TRIzol kit (Invitrogen)[1]. Polyadenylated RNA was isolated from extracted total RNA using Dynabeads from mRNA Purification Kit (Invitrogen)[1]. After fragmentation first strand cDNA synthesis was conducted on polyadenylated RNA via Superscript III reverse transcription kit (Invitrogen) [1]. Adaptor ligation, A-tailing, and adaptor ligation were conducted using the SRPI-Works system (Beckman Coulter)[1]. Barcodes were added during cDNA amplification. Average library size was roughly 200bp.

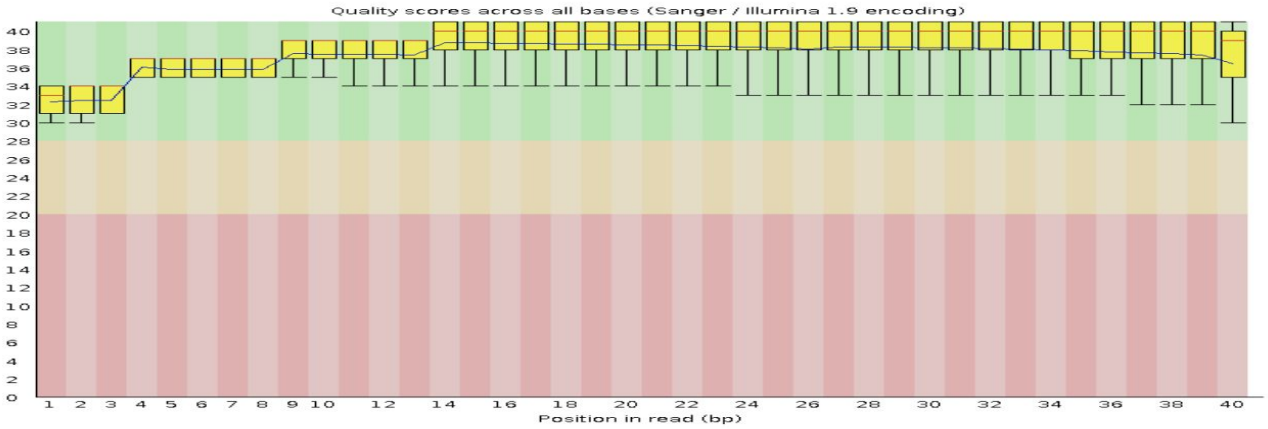
The Neonatal cardiac myocyte samples were downloaded from NCBI functional genomics data repository (GEO). An additional sample was downloaded from the link:(https://trace.ncbi.nlm.nih.gov/Traces/sra/?run=SRR1727914) accession number (GSM1570702).

Quality Control

Cardiac myocyte libraries were sequenced on a Illumina HiSeq 2000 with a run setting of Paired-end 40. For sample P0_1, 21,577,562 reads were generated with a %GC of 49. SRA-tool kit was used to convert the SRA (short read archive) formatted samples into 2 Fastq files. Two Fastq files were generated because of the paired-end sequencing mode. FastQC was run on each file to determine quality.

Figure 1A and 1B display BoxWhisker plots from read one and two from neonatal cardiac myocyte sample P0_1. The y-axis represents a quality score, while the x-axis represents each cycle or read of the transcript. Mean values for each read is represented by the blue line and median is shown with a red line.

A. Per Base Sequence Quality P0_1_1



B. Per Base Sequence Quality P0_1_2

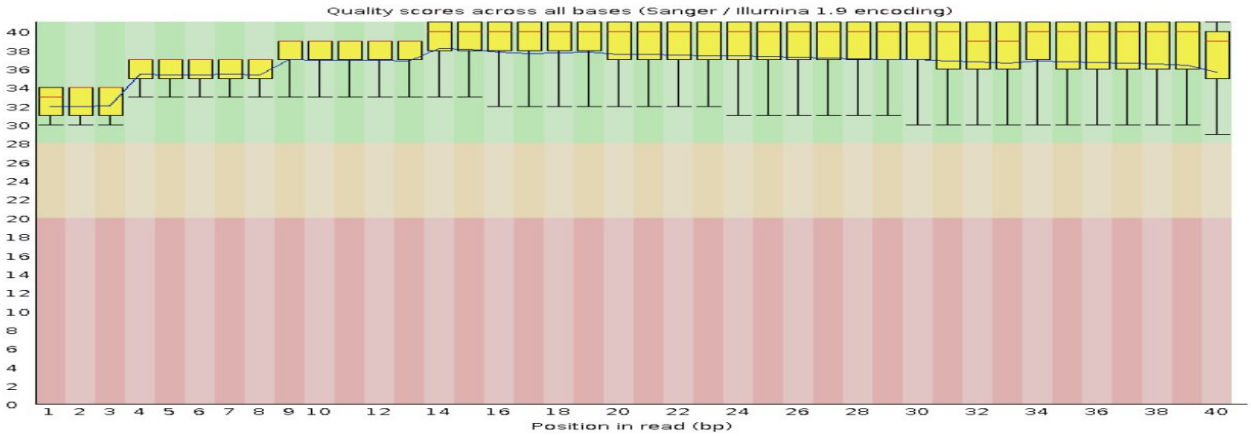
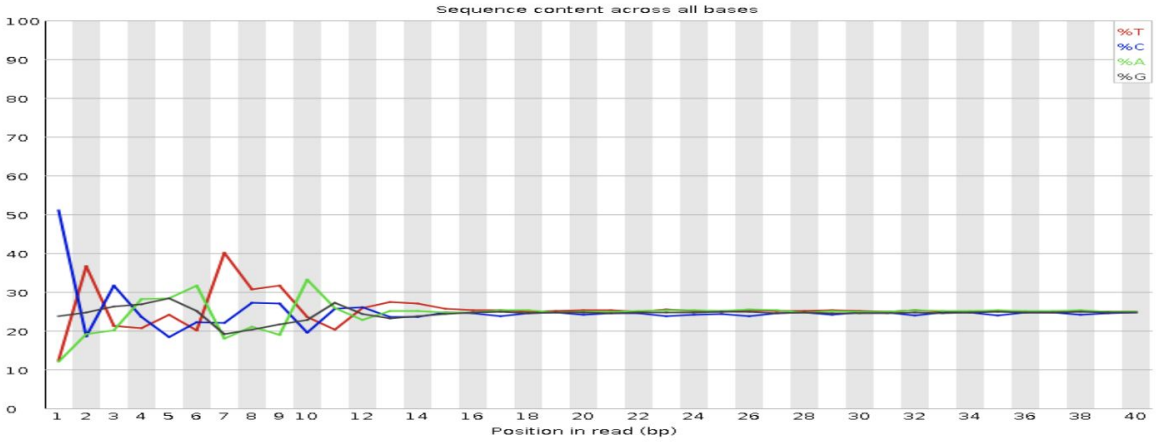


Figure 1. Per Base Quality scores for P0_1

Figure 1A from P0_1_1 has mean values all greater than 32 indicating high quality scores at each cycle in read one. Quality trended upward during the first 10 cycles and dropped slightly as the read continued. The lower read quality in the first 10 cycles is likely a result of low base diversity shown in Figure 2. Read two from sample P0_1 is shown in Figure 1B (P0_1_2). Both quality and trends are very similar to Figure 1A. This is to be expected when dealing with paired-end sequencing. The median value for read two was slightly lower than read one, this can be a result of enzyme degradation during read 2.

A. Per Base Sequence Content P0_1_1



B. Per Base Sequence Content P0_1_2

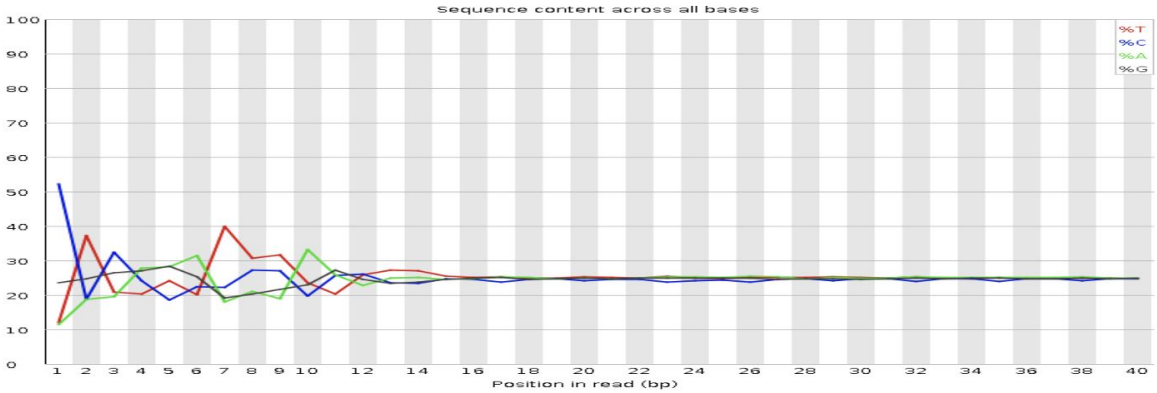
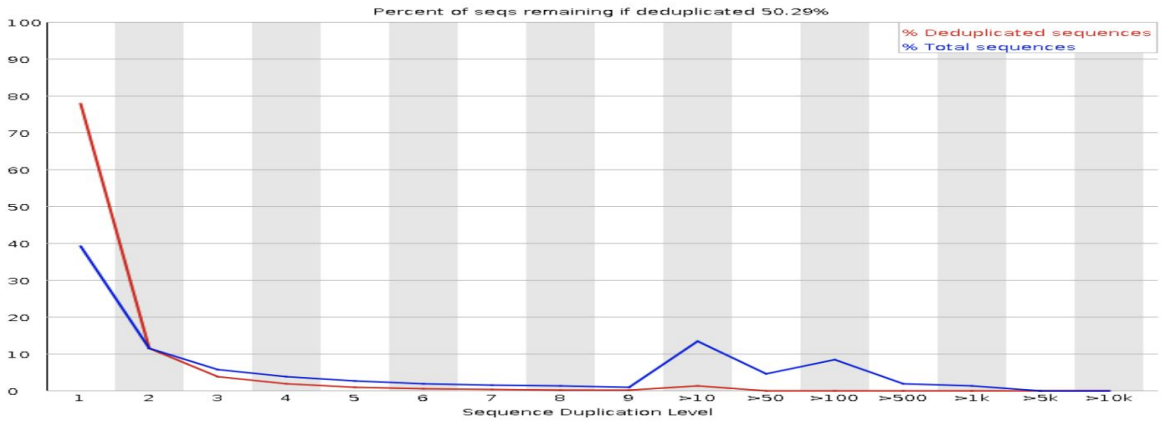


Figure 2. Per Base Sequence Content P0_1

Figure 2 represents the sequence content at each base during both reads. The y-axis shows a percentage score of each possible nucleotide, where the x -axis is the position in read. Diverse and clean libraries should have a uniform percentage of base around 25%. Figure 2A and 2B show a lack of base diversity in the first ten cycles /positions of both reads. This could be an attribute of the transcript or the machine possibly adjusting as the read continues. After cycle ten the diversity increases. This is to be expected according to the increase of quality score shown from Figure 1. Quality scores were also lower in the first ten cycles shown in Figure 1.

A. Sequence Duplication Levels P0_1_1



B. Sequence Duplication Levels P0_1_2

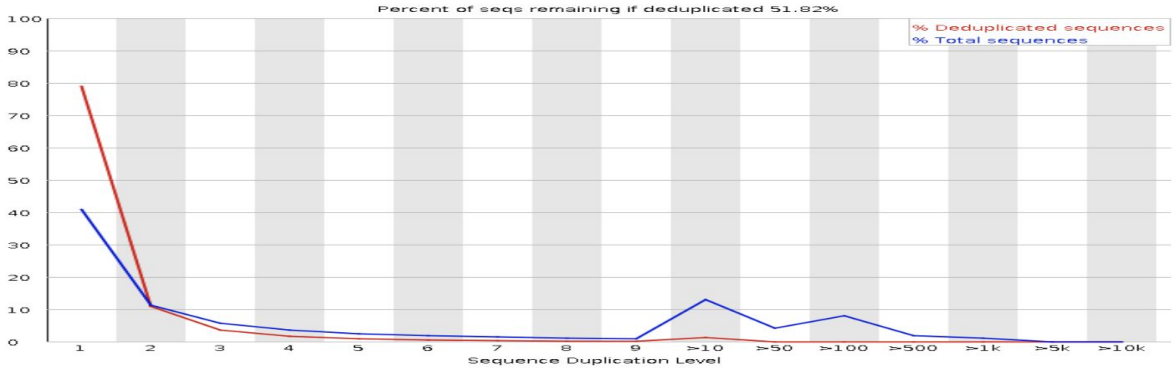


Figure 3. Sequence Duplication Levels P0_1

Figure 3 is a representation of the diversity for library P0_1. The blue line represents the distribution of library duplication within the full sequence set. The red line represents the different duplication levels in sample P0_1. Sequences that show minimal duplication should show no peaks after the initial starting peak. Both figure 3A and 3B show slight duplication of unique sequences. Having a higher total sequence(blue line) in areas can be a result of slight over application during library construction.

METHODS

Pipeline for mapping of RNA-Seq libraries

The mapping process of the two given FASTQ files containing the RNA-seq paired end reads needed index files for the barcodes used in the RNA-Seq, in addition to the annotation file of the reference sequence. Due to the computationally demanding nature of this process, a batch job was submitted to utilize a number of tools available on the Shared Computer Cluster (SCC). Bowtie2 is a package used for genomic reads alignment based on the Barrows-Wheeler algorithm, which allows a margin of mismatch and randomization that reduces the memory footprint. Therefore it is fast and described as a greedy algorithm. Bowtie is not Reads splice aware [6]. Tophat is a transcript mapper for RNA-Seq reads based on the aligner bowtie. Samtools is a package of tools to view, sort, merge and get statistics on alignment files, that are kept in SAM format.

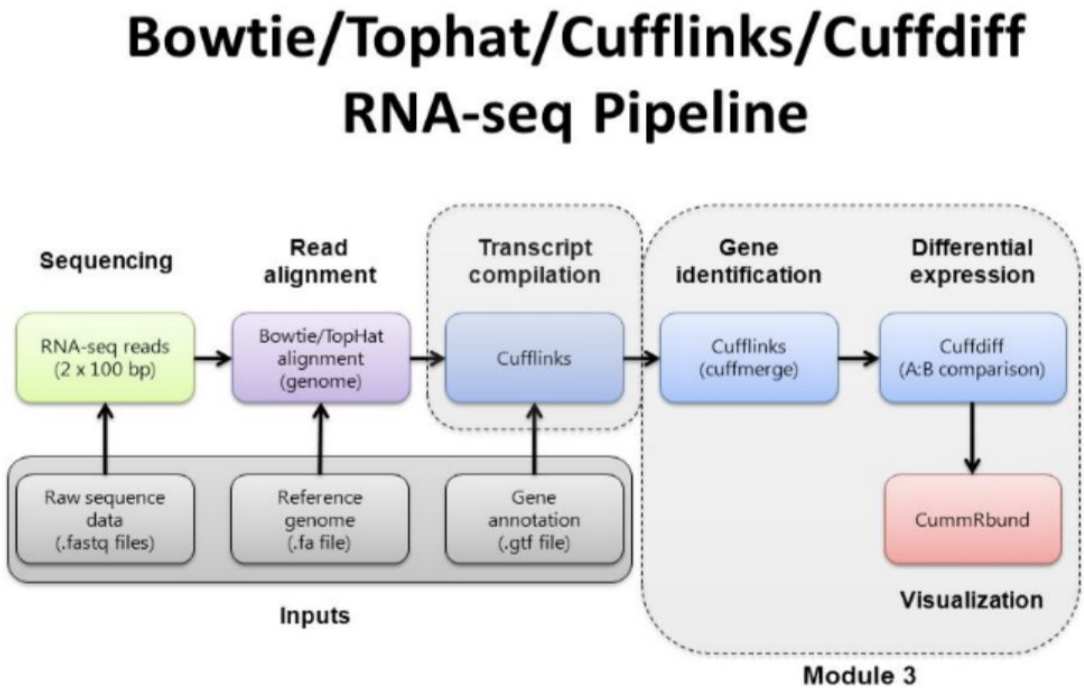


Figure 3 : Pipeline used for analysis of RNA-Seq. [4].

The P0_1 and P0_2 fastq files were aligned against the mm9 reference genome using TopHat. The file accepted_hits.bam was generated from Tophat. The flagstats module from samtools was applied to accepted_hits.bam to obtain summarized statistics of alignment result . There were 21577562 left or right reads in total. For the left reads, 20878784 of them were mapped, accounting for 96.8% of total reads. Therefore, there were 698778 unaligned reads, making up the remaining 3.2%. 1468843 of the aligned reads have multiple alignments, which account for

7.0% of them. For the right reads, 20510550 of them were mapped, accounting for 95.1% of total reads. And the rest 4.9% were 1067012 unaligned reads. Out of the total aligned reads, 1431111 of them have multiple alignments, which account for 7.0%. Next, 3 modules from RseQC, namely geneBody_coverage, inner_distance and bam_stats were run. geneBody_coverage was submitted as a batch job while the other 2 were directly run on the command line.

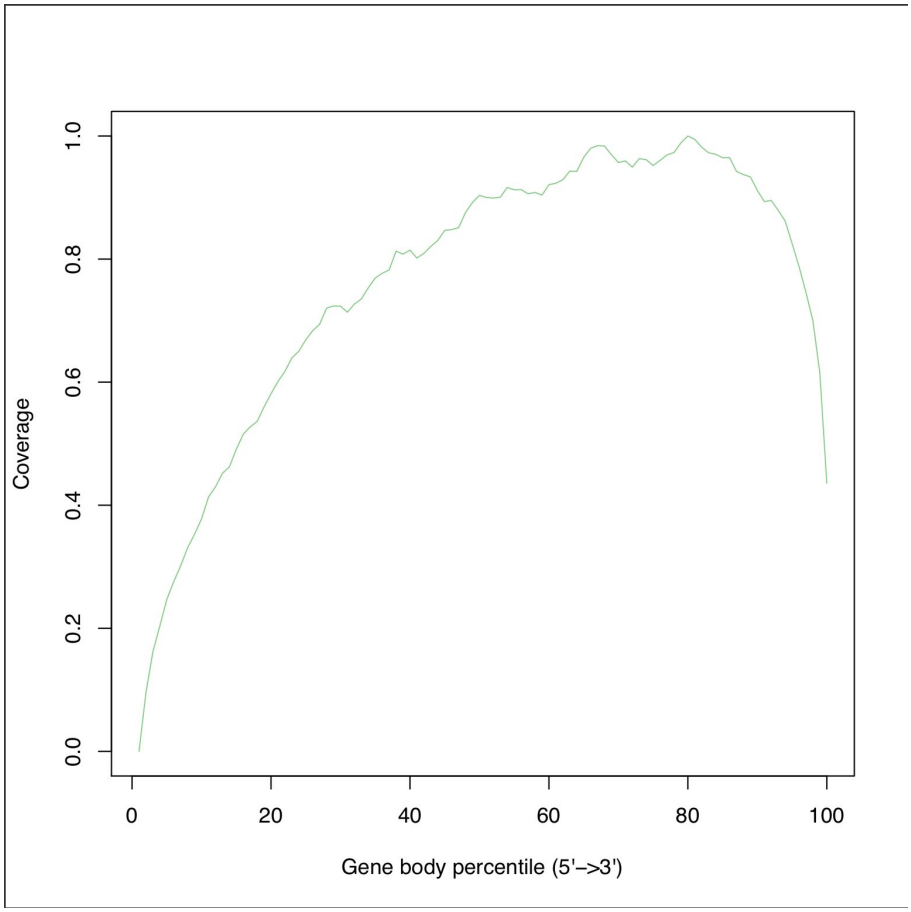


Figure 4 Coverage of gene body by reads

Figure 4 is the output of geneBody_coverage, which calculates the percentage of gene body covered by reads, starting from 5'end. It shows that the reads provided good coverage overall, with slight loss in the first half of the gene since 80% coverage was not achieved until around 40th percentile.

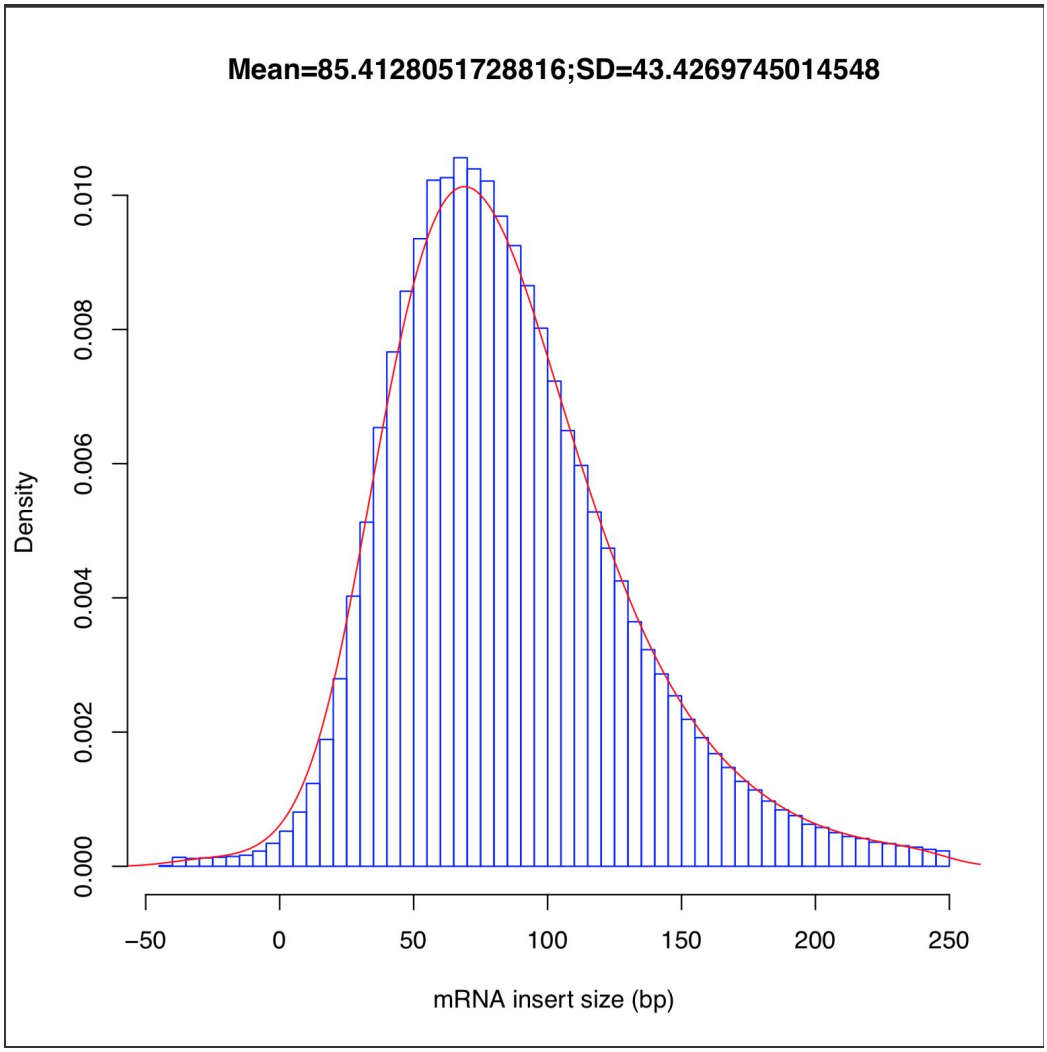


Figure 5 Distribution of insert size between read pairs.

Figure 5 shows the distribution of inner distance, also called insert size, between two paired reads. The distribution shown in the figure is roughly normal with a mean of about 85. Therefore there is little structural variation in the genes.

Bam_stats is a module that calculates the number of uniquely mapped reads in order to assess mapping quality. The result shows that the data is of good quality since there are no QC failed or PCR duplicated reads. Out of the total 49706999 reads, 38489380 or 77.4% are unique.

Both cufflinks and cuffdiff were run as batch jobs. One of the outputs of cufflinks, gene.fpkm_tracking, was loaded into R using the read.delim function. FPKM, as a measure of normalized gene expression value, was extracted from the table. Then a threshold of 1 was set up to select for FPKM values used for plotting. FPKM values between 0 and 1 were discarded as outliers. 14205 out of 37429 FPKM values passed this threshold. These values were then transformed to their log base 10 values before a histogram was plotted.

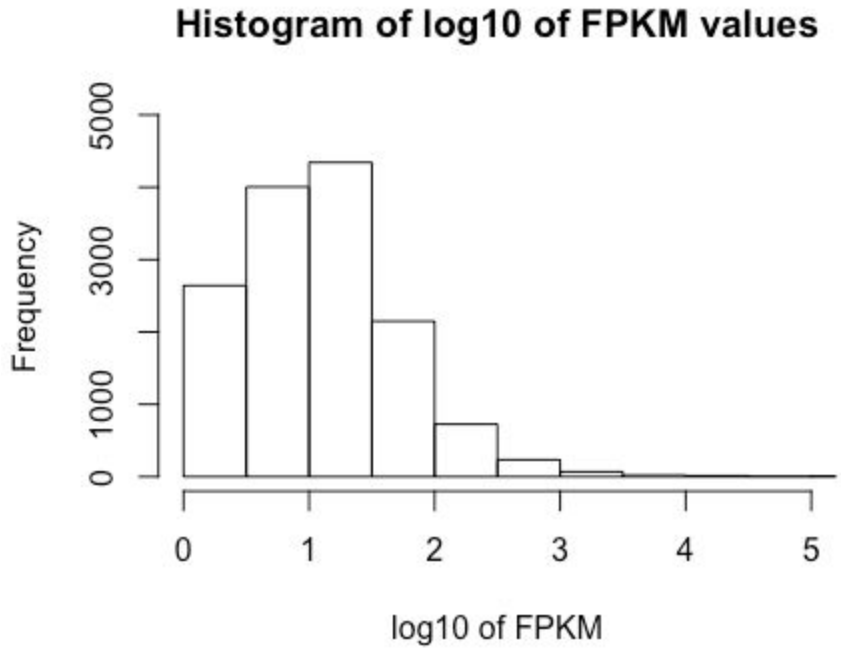


Figure 6 Histogram of log-transformed FPKM values

Figure 6 displays the histogram of FPKM values in log base 10. The majority of data have a log base 10 FPKM value between 0 and 2, meaning that between 1 and 200 fragments were expressed per kilobase million.

Analysis

Following the cuffdiff procedure, the generated file, which contained the differentially expressed genes' statistics, was analyzed to determine an association with myocyte differentiation. Within this file, differential expression statistics were contained for postnatal day 0 (P0) vs Adult (Ad) mice. The R 'order' function sorted the data and created a gene list that contained the smallest q_values at the top (Table 1). In addition, the data frame was subset into genes that were denoted as 'significant'. Two log2 fold change histograms were created: one for the sorted data of all genes and another for significant genes (Figure 8). Log2 fold change was calculated for the entire gene set in order to symmetrically represent the data in a normalized fashion. Note that when the Ad/P0 ratio is closer to 1, this results in Log2 fold change = 0, which indicates that the Ad and P0 FPKM values are equal. Then, this value would indicate that the gene expression level did not change [5]. Two separate csv files were created containing the up-regulated and down-regulated genes. Within DAVID Functional Annotation Clustering, the files were imported. Mus Musculus was selected as the species of interest. In the Gene Ontology group, GOTERM_BP_FAT, GOTERM_MF_FAT, and GOTERM_CC_FAT were selected for further analysis.

RESULTS

Differentially Expressed Genes

The ten top differentially expressed genes based on q-value and p_values. The genes had similar p-values and q-values (Table 1). Given that the p_value and q_value were similar for all genes, there was no difference in using either values for the analysis.

Top Ten Differentially Expressed Genes						
	Gene	P0 FPKM	Ad FPKM	Log2 Fold Change	p-value	q-value
1	Plekhb2	22.56790	73.568300	1.70481	5e-05	0.00106929
2	Mrpl30	46.45470	133.038000	1.51794	5e-05	0.00106929
3	Coq10b	11.05830	53.300000	2.26901	5e-05	0.00106929
4	Aox1	1.18858	7.091360	2.57682	5e-05	0.00106929
5	Ndufb3	100.60900	265.235000	1.39851	5e-05	0.00106929
6	Sp100	2.13489	100.869000	5.56218	5e-05	0.00106929
7	Cxcr7	4.95844	32.275300	2.70247	5e-05	0.00106929
8	Lrrfip1	118.99700	24.640200	-2.27184	5e-05	0.00106929
9	Ramp1	13.20760	0.691287	-4.25594	5e-05	0.00106929
10	Gpc1	51.20620	185.329000	1.85570	5e-05	0.00106929

Table 1: Differentially Expressed Genes Associated with Myocyte Differentiation by q-value. P0 FPKM = Postnatal Day 0 ‘s FPKM, Ad FPKM = Adult’s FPKM; log2 Fold Change = log2 (Ad FPKM/P0 FPKM).

Differentially Expressed Genes (p<0.01)	
Up-regulated Genes	1084
Down-regulated Genes	1055
Total Genes	2139

Table 2: Differentially Expressed Genes with significant level of p < 0.01.

Table 2: Shows the number of Differentially Expressed Genes with significant level of p < 0.01. From the generated gene_exp.diff file, we determined the number of differentially expressed genes with p < 0.01 to be 2139. Out of the total 2139 differentially expressed genes, 1084 genes were up-regulated and 1055 were downregulated in Ad mice relative to P0 mice (Table 2). Within the R script, genes were denoted as up-regulated if the log2 fold change > 0 and were down-regulated if log2 fold change < 0.

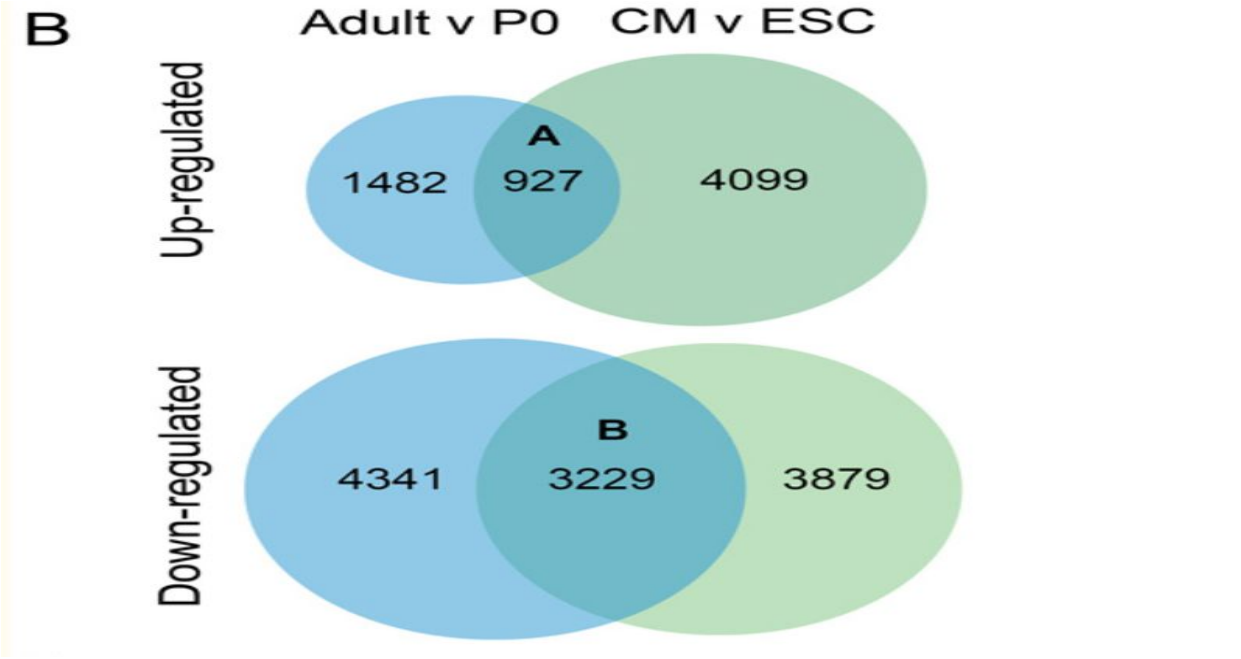


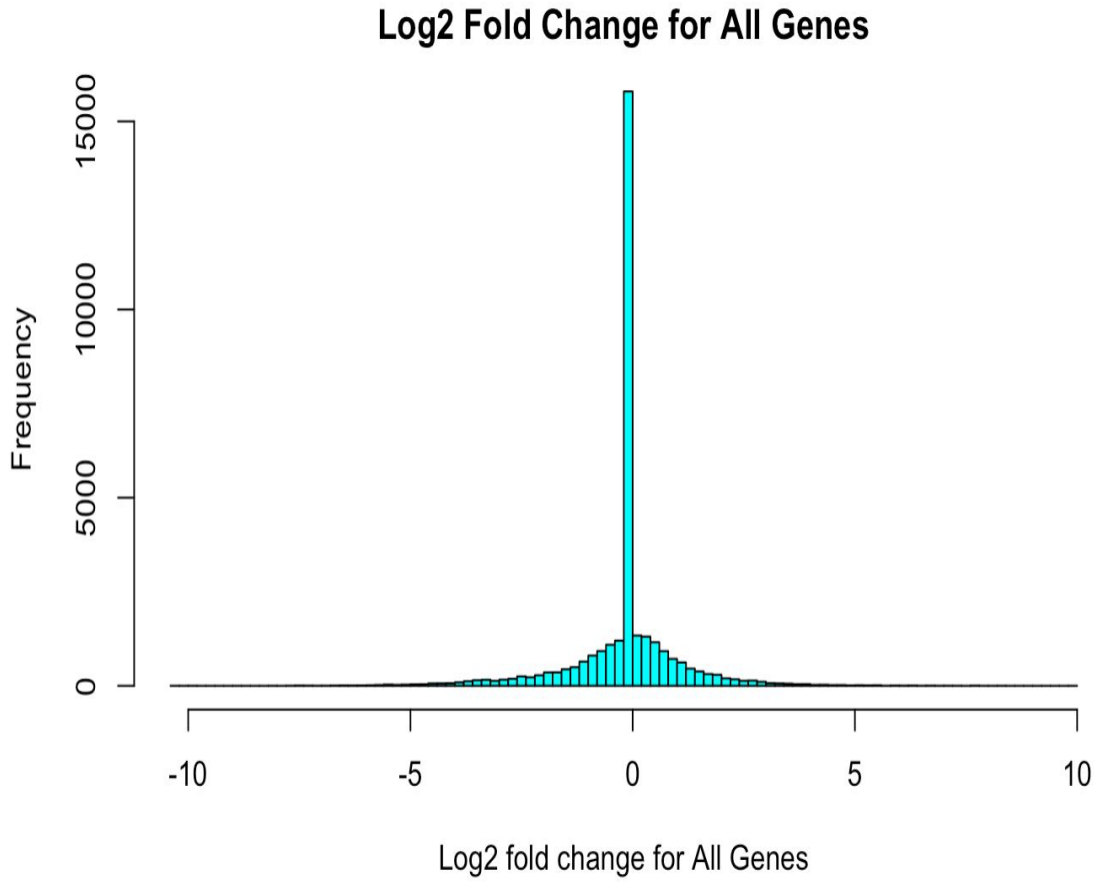
Figure 7: The Venn diagram from Figure 7 is obtained from O'Meara et al. It shows the estimated number of upregulated and downregulated genes from the original study[1].

The up-regulated genes from O'Meara et al were 2409. The number down-regulated genes was 7570. Our number of upregulated and downregulated genes were lower. We were not able to reproduce this part of the analysis.

Log2 Fold Change Histograms

The significant expressed genes histogram displayed log2 fold change values that were split on both sides of 0(Figure 8B). In contrast, the histogram with all the differentially expressed genes has a large peak at log2 fold change of 0 (Figure 8A). This difference observed is due to the fact that all insignificant genes were removed in histogram B, thus the log2 fold change of 0s are not present. The single large peak at 0 in Figure 8A shows the large frequency of differential genes that were determined to not be significant.

A



B

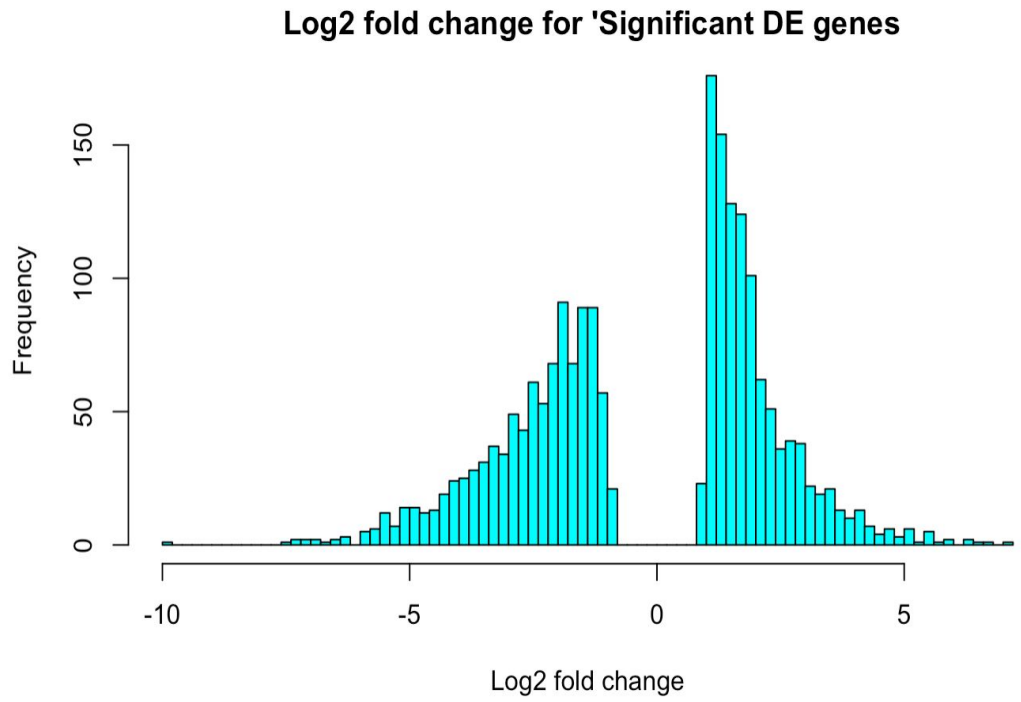


Figure 8: The Histograms of Log2 Fold Change for All Differentially Expressed Genes and Significant Genes. A shows all Differentially Expressed Genes. B shows Significant Differentially Expressed Genes.

In the histogram of log2_fold_change: A. The histogram shows all the differentially expressed genes. 8B. Shows significant differentially expressed genes. The log2 fold change value represents how much the gene is expressed, either under or over expressed. From observing the histogram of significant genes, the genes with log2 fold change < 0 are down-regulated genes, and those genes with log2 fold change > 0 are up-regulated genes in Ad relative to P0. The log2 fold change represents a change in the expression of the gene between Ad and P0. The log2 fold change is the representation of the ratio of the Y/X, which in this study is Ad/P0 ratio.

DAVID Annotation Clusters using David Analysis

The table summarizes the top cluster results after the DAVID annotation clustering (Table 3). Since there were multiple genes within each cluster, in order to condense the data, a specific

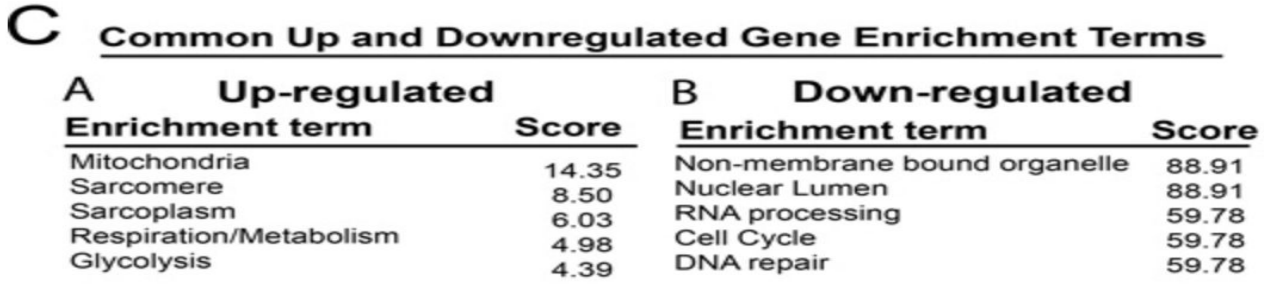
GO Enrichment Term was chosen to summarize the terms associated with the genes within the clusters.

Top DAVID Clusters From David Analysis			
	Cluster Number	**GO Enrichment Term**	**Enrichment Score**
Up-regulated Genes	1	Mitochondrion	21.34
	2	Nucleoside metabolic process	15.29
	3	Lipid metabolic process	14.57
	4	Extracellular organelle	10.79
	5	Glycolysis: Glycolytic process, Cofactor metabolic process, regulation of carbohydrate metabolic process	8.43
	6	Sarcomere	6.99
	7	Respiration/ Metabolism: small molecule catabolic process, carboxylic acid catabolic process, fatty acid catabolic process	6.19
	12	Sarcoplasm, sarcoplasmic reticulum	4.18
Down-regulated Genes	1	Proteinaceous extracellular matrix	9.68
	2	Cell proliferation	9.16
	3	Regulation of cellular component organization	8.63
	4	Organ morphogenesis	8.08

	5	Cardiovascular system development	7.59
	6	Nuclear Lumen: Chromosome, nuclear chromosome, nuclear chromosome part, nuclear chromatin	7.55
	8	DNA repair	7.16
	9	Cell cycle	7.15
	12	RNA processing: regulation of RNA metabolic process, RNA biosynthetic process, negative regulation of macromolecule biosynthetic process, RNA metabolic process	6.27

Table 3: Top DAVID Clusters produced for the up-regulated and down-regulated gene csv files. GO Enrichment Term provides a concise term that encompasses the genes found within that cluster. Enrichment score = -log10(cluster genes' average p-value).

From each cluster from the up-regulated and down-regulated genes csv_files, one gene was chosen from each cluster. The genes with the high count values from the top clusters were chosen for use in the above table(Table 3).
Comparatively, the table below from O'Meara et al.. shows the common up-regulated and down-regulated genes enrichment terms from their analysis.


C Common Up and Downregulated Gene Enrichment Terms

A Up-regulated

Enrichment term	Score
Mitochondria	14.35
Sarcomere	8.50
Sarcoplasm	6.03
Respiration/Metabolism	4.98
Glycolysis	4.39

B Down-regulated

Enrichment term	Score
Non-membrane bound organelle	88.91
Nuclear Lumen	88.91
RNA processing	59.78
Cell Cycle	59.78
DNA repair	59.78

Figure 9: Shows the top gene enrichment terms for common up-regulated and down-regulated genes during differentiation from O'Meara et al.

Compared to analysis results from O'Meara et al. results shown above , the top gene enrichment terms for common up-regulated and down-regulated genes such as Mitochondria, Sarcomere,

Sarcoplasm, Glycolysis and Respiration/Metabolism were reproducible. For the down-regulated gene enrichment terms, DNA repair, Cell cycle, Nuclear Lumen and RNA processing were reproducible.

DAVID Annotation Clusters: Biologist

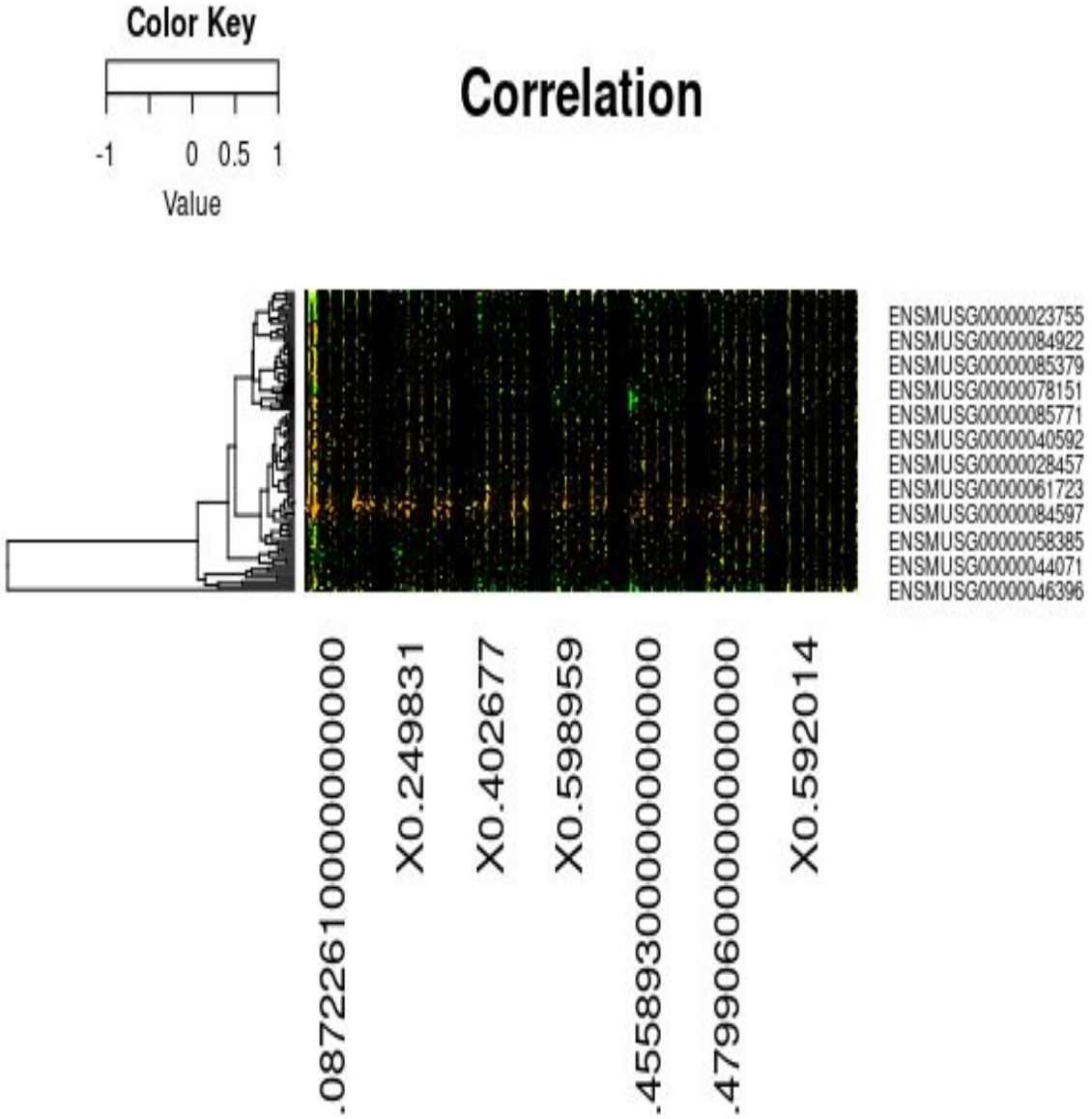


Figure 10: Heatmap of top upregulated and down regulated genes

Examining the complete FPKM matrix, a total of over 4k genes was observed out of those 4k genes we were able to successfully identify 2000 regulated genes with 1k being up-regulated

and 1k being down-regulated. This was taken and placed into a clustered heatmap and showed a correlation that cosigned the results found on the paper regarding what genes are significant compared against the P0 and AD pools.

Differentially Expressed Genes' Functions and Total Count Comparisons

The test data, in contrast to the original data, presented the same overall magnitude of change. The trend appears to follow what is presented in the paper, where the increase in gene expression shows a score of double in the un-normalized values, which resulted in an FPKM normalized enrichment score similar to that of the ones provided in the paper in the P4-P7 sample pools.

DAVID Clustering GO Term Comparison

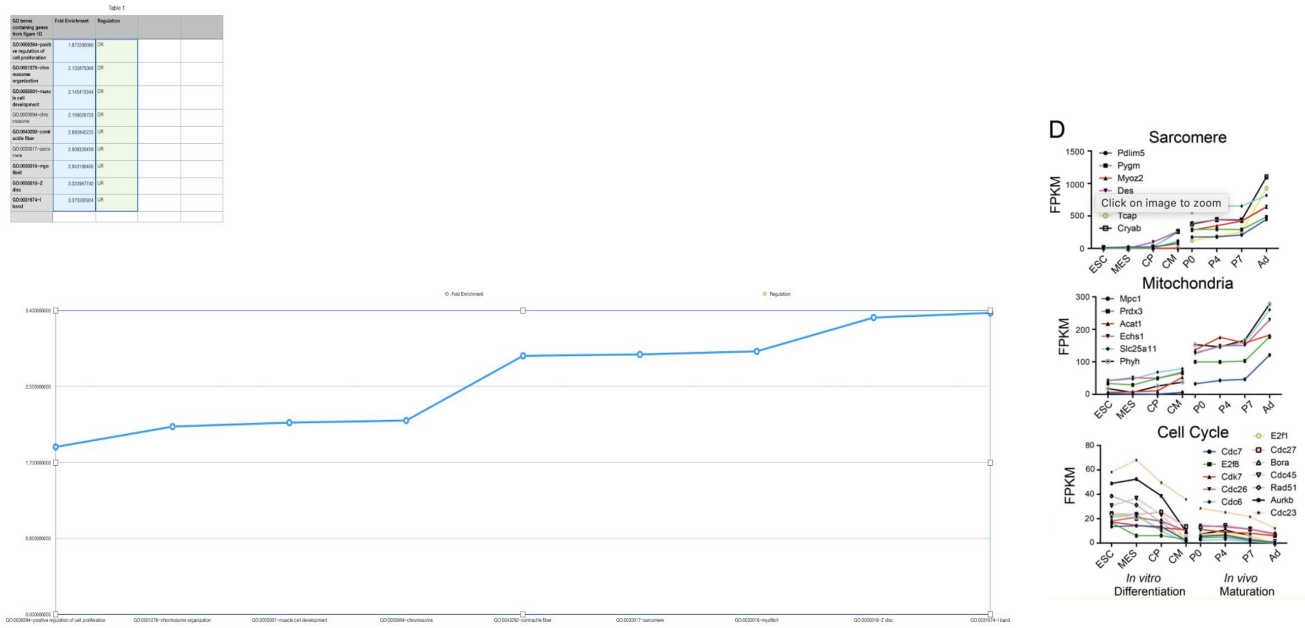


Figure 11: Line plot of GO terms from genes on figure 1D of original project paper. The table above line plot contains the fold change enrichment value of the DAVID analysis GO terms associated with the genes of interest from figure 1D of the original paper.

After receiving the DAVID Clustering values, a find script was used to search for the genes that appeared in figure 1D of the paper. Those genes were then identified and associated with the GO term results from the DAVID analysis. For example, the TCAP gene was associated with GO:0030017~sarcomere, the fold enrichment value of this GO term contained the TCAP gene and other associated genes. The fold enrichment value was extracted from those terms and

placed into a table, and a line plot was generated based on the table values. As you can see above, the line plot has the same trend increase as the line plots on figure 1D, proving that the results are similar with little to slight variation in exact fold enrichment score numbers. The trend line could have been more exact, but the genes that we had data for were limited compared to the original papers, which is explained in more detail in the conclusion section. This caused a trend line that is not exactly like the table in the paper.

DISCUSSION

Differentially Expressed Genes' Functions and Total Count Comparisons

Using R for analysis, most of the analysis was doable with minimum error. The top-ten differentially expressed genes were determined based on lowest q-value; Using q-values also yielded similar results because they had similar p-values and q-values. Out of the top-ten differentially expressed genes, 8 genes had had a positive log2 fold change value. 2 genes, Lrrfip1 and Ramp1 had negative log2 fold change value which means they were down-regulated. We can therefore conclude that the majority of the top-ten differentially expressed genes were up-regulated in Adult relative to P0 mice. As for the number of genes that were determined to be differentially expressed, the numbers in our study do not match O'Meara et al. Of the 36329 genes we used in this study(in vivo maturation conditions), we determined the number of differentially expressed genes to be 2139. Out of these genes, we determined 1084 to be up-regulated and 1055 to be down-regulated. In comparison, O'Meara et al. had 2409 up-regulated genes while the number down-regulated genes was 7570. Our numbers of up-regulated and down-regulated genes were lower and we were not able to reproduce this part of the analysis. One of the possible reasons for observing this discrepancy is because of different gene sets being analyzed as well as differences in sub-setting and clustering. Furthermore, the authors in O'Meara et al. had access to samples and cell lines used in both the in vivo maturation and in vitro differentiation models while we only had access to some samples in the in vivo maturation model. The reason for the different conditions for both models allow researchers to dive into details when investigating differences in gene expressions of different cell types and at stages. With this in mind, it is interesting to note that although there were differences in the differentially expressed gene numbers, the overall top clusters GO Enrichment Terms from our study mostly match the ones from O'Meara et al.

DAVID Clustering GO Term Comparison

In this project, the analyst used DAVID to annotate the significantly differential expressed genes in the in-vivo maturation cell lines. The functional annotation results from our DAVID Analysis were mostly similar to those in O'Meara et al. Compared to analysis results from O'Meara et al., the top gene enrichment terms for common up-regulated genes such as Mitochondria, Sarcomere, Sarcoplasm, Glycolysis and Respiration/Metabolism were all reproducible. For the down-regulated gene enrichment terms, DNA repair, Cell cycle, Nuclear Lumen and RNA processing were similarly reproducible. However, it is also important to note that a few of GO

Enrichment Terms were not in the same hierarchy as observed in O'Meara et al. We observed in our study that some of the gene enrichment terms such as Glycolysis and Respiration/ Metabolism did not have exact matches rather groups of genes that had GO terms that fell under glycolysis and Respiration/ Metabolism. Similar observations were made for some of the down-regulated gene enrichment terms such as RNA processing whereby exact matches were not available in our results. Rather we observed groups of genes that had enrichment terms that were involved in RNA processing such as regulation of RNA metabolic process, RNA biosynthetic process, negative regulation of macromolecule biosynthetic process and RNA metabolic process. The same observation was made for the Nuclear Lumen.
The small differences between two DAVID annotation results from both studies could be attributed to several reasons. The first reason is because of version control. DAVID had been updated since O'Meara et al. was published in 2016. We had no access to the older version and had to use the newest version which had a lot of trouble loading. The second reason is that the DAVID annotation results in O'Meara et al. are from the in vitro differentiation conditions and in vivo maturation conditions while we only used the in vivo maturation conditions in our study. The third and last reason is because the authors in O'Meara et al.. provided broad terms for the enrichment terms rather than matching/distinct GO IDs and enrichment scores. We could not precisely match our results to those of the authors in their study.

Studying the differentially expressed gene functions in Ad vs. P0 helps us determine the differences in cell functional processes. Up-regulated genes had GO terms that are related to energy production such as Metabolism, Respiration/Metabolism(fatty acids and lipids metabolism) and Glycolysis. Mitochondria produces the energy for the Ad mice and most cellular processes. Adult mice are more dependent on fatty acids and lipids in order to produce energy in comparison to P0 mice [7]. Genes that are down-regulated in Ad relative to P0 are closely related to growth and development required for P0 as they are developing and producing new cells that are needed for survival. One of the major biological reasons why adult mice and mammals do not undergo tissue regeneration is because of the process of cell cycle exit. Therefore, genes related to cell cycle were down-regulated in adult mice relative to P0 genes.

GO Enrichment Scores

The results of our study did not reproduce the Enrichment Scores in O'Meara et al.. As previously mentioned, this is because only a subset of the genes were used in the clustering that was done in the latter study. [1]. The GO Enrichment Score is the -log10(average p-value). As the overall average p-value of the enriched gene set decreases, the GO Enrichment Score would be expected to increase. When samples are averaged in large gene sets, each gene's p-value slightly contributes to the overall average as compared to a gene in a smaller gene set. When a new gene value is added with a different p-value than the average, it would be expected to have a greater effect on the smaller gene set's average than the larger gene set's average p-value. This caused a change in the overall Enrichment Score.

CONCLUSION

In comparison to O'Meara et al., we were partially able to reproduce their results using our study. We were able to reproduce some similar results and biological interpretations of the study using samples in the in vivo maturation model. Some of the possible reasons for observing this discrepancy is because of different gene sets being analyzed as well as differences in sub-setting and clustering. The total number of differentially expressed genes, including both the up and down regulated differentially expressed genes in our study did not precisely match O'Meara et al. Based on the observations we made from our analysis and results, the GO Enrichment Terms were mostly related to each other as well as the paper. The GO Enrichment Terms that were up-regulated and down-regulated in Ad relative to P0 were related to the differences in functions between Adult and P0 mice. The exact gene counts and top genes in our study were not similar to O'Meara et al. The overall biological functions were able to be observed between our study and O'Meara et al. This final conclusion that we were able to make is supported by the consistent GO Enrichment Terms that were up and down-regulated in Ad relative to P0. However, it is important to note that this conclusion was made in the absence of matching GO IDs and Enrichment scores from O'Meara et al. The genes that were up-regulated were related to energy production and musco-skeletal processes. The genes that were down-regulated in Ad mice were related to cell growth and differentiation and would therefore be expected to be less active in adults and prevent the regeneration of cardiac tissue later on in life.

REFERENCES

1: O'Meara CC, Wamstad JA, Gladstone RA, et al. Transcriptional reversion of cardiac myocyte fate during mammalian cardiac regeneration. *Circ Res*. 2015;116(5):804–815. doi:10.1161/CIRCRESAHA.116.304269

2: Porrello ER, Mahmoud AI, Simpson E, et al. Transient regenerative potential of the neonatal mouse heart. *Science*. 2011;331(6020):1078–1080. doi:10.1126/science.1200708

3: Porrello ER, Mahmoud AI, Simpson E, et al. Regulation of neonatal and adult mammalian heart regeneration by the miR-15 family. *Proc Natl Acad Sci U S A*. 2013;110(1):187–192. doi:10.1073/pnas.1208863110

4. Gilbert, D. (2004). Bioinformatics software resources. Briefings in bioinformatics, 5(3), 300-304.

5. Tusher, Virginia Goss; Tibshirani, Robert; Chu, Gilbert (2001). "Significance analysis of microarrays
applied to the ionizing radiation response". Proceedings of the National Academy of Sciences of the
United States of America. 98 (18): 5116–5121. doi:10.1073/pnas.091062498. PMC 33173. PMID
11309499.

6. Langmead, B., Trapnell, C., Pop, M., & Salzberg, S. L. (2009). Bowtie: An ultrafast memory-efficient
short read aligner. Genome Biol, 10(3), R25.

7. 15. Goldberg, I. J., Trent, C. M., & Schulze, P. C. (2012). Lipid metabolism and toxicity in the heart. Cell metabolism, 15(6), 805-812.

LINKS:

- https://www.miltenyibiotec.com/US-en/products/macs-sample-preparation/tissue-dissociation-kits/neonatal-heart-dissociation-kit-mouse-and-rat.html

- http://david.abcc.ncifcrf.gov/summary.jsp

Shared roles write-up

Introduction: Caroline Muriithi/ Nicholas Mosca
Methodology(Curator, Programmer, Analyst and Biologist): Everyone
Overall Discussion and Conclusion: Caroline Muriithi
Final Edits and submission: Nicholas Mosca/Caroline Muriithi/Zeyuan Cao