Transcriptional Profile of Mammalian Cardiac Regeneration with mRNA-Seq

Analyst: Manas Dhanuka

Introduction

Neonatal mice possess a fascinating ability to fully regenerate their heart after injury [2], but this potential is lost after the first week of life [2]. The reason why mammals cannot fully regenerate their heart after an injury as an adult is due to cardiac myocytes (CM) leaving the cell cycle [3]. To further investigate this O'Meara et. al. 2015, resected the left ventricle apex from neonatal mice, genetic fate mapping showed that already established CMs gave rise to the new and regenerating CMs, which produced a fully regenerated heart [2]. In a broader sense, this study found that identifying the mechanisms in which myocytes undergo cell cycle activity during regeneration will give a deeper understanding of why heart regeneration is limited in adult hearts [2]**.**

The transcriptional changes that cause regenerative heart phenotypes were generally unknown until O'Meara et al. (2015) aimed to identify the genes and gene networks that change throughout the regeneration process. Specifically, global gene expression patterns at different time points of mouse CM differentiation in both *in vivo* and *in vitro*. Additionally in neonatal mice, researchers observed global gene expression patterns in whole heart ventricles and purified CMs after apical resection. This revealed that heart regeneration can be characterized by a transcription reversion of the CM differentiation process.**6** Using RNA-sequencing datasets, researchers were able to accurately predict regulators and associated pathways that control the cell cycle state of CMs.

Overall, O'Meara et al. concluded that cardiac regeneration is a regulated process in which transcriptional reversion of the differentiation process occurs. The goal of this report is to reanalyze just one sample of the data and replicate the results found in the paper using the same bioinformatics tools. Reanalyzing and focusing on just one sample of the data instead of the 36 examined in the paper could provide different insights as to how neonatal mice can regenerate their heart tissue but lose the ability later in life at a transcriptional level.

Methods

Now, to obtain the difference between the samples of P0 mice and Adult mice on the gene level. A gene differentially expressed analysis based on the gene difference table that was generated earlier via cuffdiff was used. The top 10 differentially expressed genes were selected by sorting the table with q-value. Next significant genes were filtered out by p-value and "significant" label.

Two histograms of the log2 fold change were plotted for all genes and only significant genes to visualize the distribution. The gene set was then split into up-regulated and down-regulated genes using log2 fold change as a parameter (positive and negative respectively). These genes were then used for enrichment analysis.

For the enrichment analysis, DAVID functional annotation tool was employed. This was done to cluster the up and down regulated genes with similar functional roles. The parameters selected gave results for *Mus musculus* from the Gene Ontology (GO) terms for biological processes, molecular function, and cellular components. The GO terms from the annotation clusters that showed overlap with the common GO terms identified in O'Meara et al. (2015) were determined using the results from supplementary data from the paper.

Results

For the differentially expressed gene analysis, the gene.diff table was sorted by q-value and the top 10 differentially expressed genes between P0 and Ad are as shown in Table 1. From the table, all these 10 genes had the smallest q-value of 0.000320557 and p-value for 5e-5 which were much smaller than 0.05. The range of FPKM value_1 was from about 0.40 to 500.00, and the range of FPKM value_2 was from about 6.00 to 881.00. One gene Nck2 had a negative log2 fold change value and the other genes had positive values.

gene	FPKM_1	FPKM_2	log2.fold_change	p_value	q_value
Adhfe1	12.7138	25.7368	1.01743	5.00E-05	0.000320557
Tmem70	36.9448	80.9577	1.1318	5.00E-05	0.000320557
Gsta3	0.412723	6.76999	4.03591	5.00E-05	0.000320557
Lmbrd1	6.58776	12.6753	0.944164	5.00E-05	0.000320557
Dst	19.721	51.5969	1.38755	5.00E-05	0.000320557
Plekhb2	25.8529	68.601	1.4079	5.00E-05	0.000320557
Cox5b	505.414	881.798	0.802984	5.00E-05	0.000320557
Mrpl30	56.205	124.292	1.14497	5.00E-05	0.000320557
Tmem182	46.2243	103.517	1.16314	5.00E-05	0.000320557
Nck2	12.174	6.29987	-0.950409	5.00E-05	0.000320557

Table 1: Top 10 differentially expressed genes select from gene.diff table sorted by q-value.

Two different methods were used to filter out the significant genes among the total of 36329 genes. The first one was using a p-value less than 0.01 as a condition, returning 4862 significant genes. Second, by using the ‘significant label’ in the table, as a condition, giving 5427 significant genes. The results of these filters have been summarized in Table 2.

	Total Genes	Up-regulated Genes	Down-regulated Genes
Labelled Significant	5427	2830	2597
p-value < 0.01	4862	2587	2275

Table 2: Summary of the number of significant genes based on different select methods.

To visualize the distribution of Log2_fold_change, the histograms of Log2_fold_change for genes are shown in Figure.1 and Figure.1. The distribution of log2 fold change for all genes (36329 genes in total) and the distribution of log2 fold change for the significant genes equaled 5427 genes. The histogram for all genes has a peak around the zero tick, whereas the histogram for only significant genes had no values around zero. This is due to the way the log2 fold change is calculated, i.e., log to the base 2 of the FPKM value ratios. A ratio close to one returns a zero, meaning there does not exist a significant difference between the P0 and Ad samples.

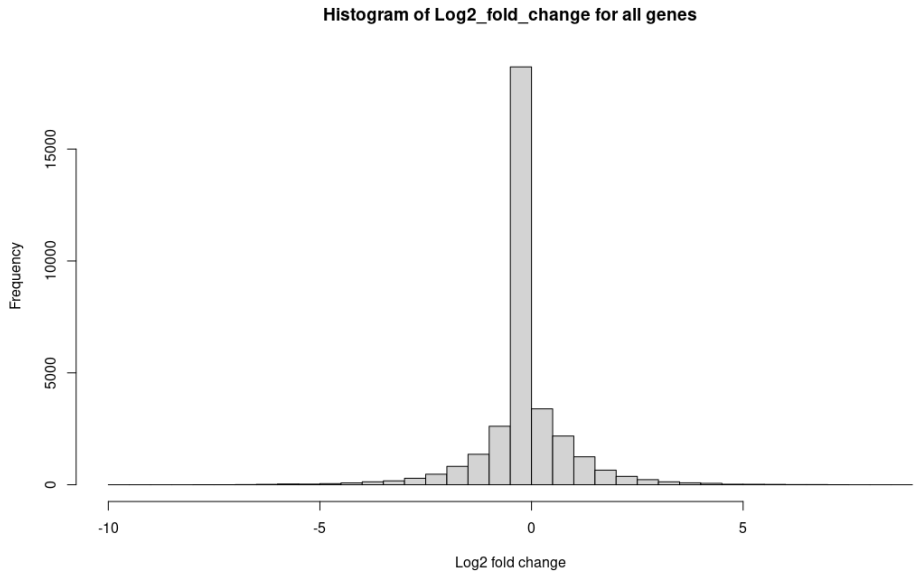

Figure 1: Histogram of the distribution of Log2_fold_change for all genes.

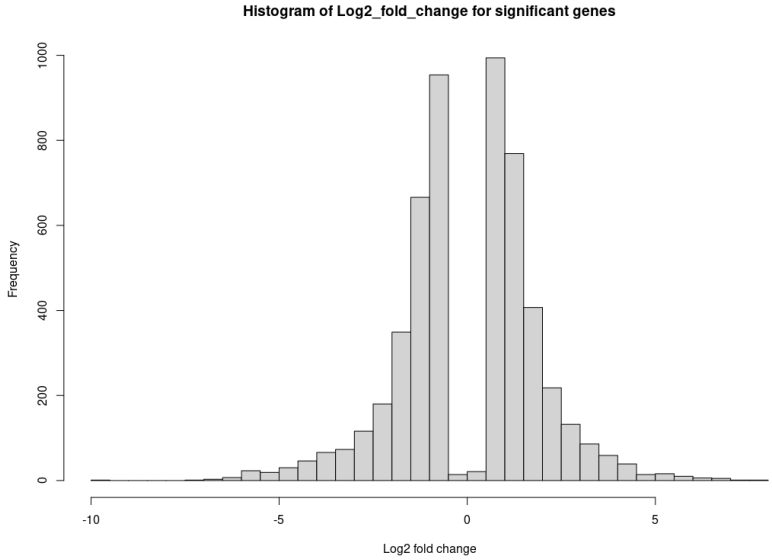

Figure 2: Histogram of the distribution based on Log2_fold_change for significant genes.

Then the enrichment analysis was performed based on DAVID Functional Annotation Tool for further functional annotation clustering. By using the up-regulated gene set and down-regulated gene annotation clusters were generated. Up-regulated genes had 828 clusters and down-regulated genes had 710 clusters.

Within the DAVID analysis results for up-regulated genes, several cluster elements were common with O'Meara et al. These included terms of mitochondria, generation of precursor metabolites, and energy derivation. These results correspond to the process of in vivo maturation of the cardiomyocyte from neonatal day 0 (P0) to adult stage (Ad) since as the cardiomyocyte matures, there will be organization of the compositional units of the cardiomyocyte and an increase in energy demand.

In the DAVID analysis results for down-regulated genes, like the upregulated DAVID analysis there were several overlapping GO terms. These included terms of the cell cycle, extracellular matrix,

regulation of organelle organization, epithelium development, and tissue morphogenesis. As cardiomyocytes mature, there is commonly an exit of the cell cycle, so there will be down-regulation of genes associated with the cell cycle.

Discussion

With the above analysis, the obtained upregulated significant genes were 2830 and significant down-regulated genes were 2597. This is close to the results for upregulated genes from O'Meara et. al. they had 2409 genes in up-regulated gene set and 7570 genes in the down-regulated gene set.

From the enrichment analysis with DAVID, the correlation clusters obtained had overlaps as well as differences from the results of O'Meara et. al. The enrichment scores also showed differences in magnitude from the study. The up-regulated set has the greater highest enrichment score for this analysis and the study has a greater highest enrichment score for the down-regulated set. This may be attributed to the difference in the p-value used for calculations.

Overall, this analysis was able to reproduce some similar results from O'Meara et al (2015) using data from *in vivo* maturation models. Some explained differences were observed due to the different sizes of the gene data set and the inclusion of data of only P0 through Ad maturation stages.

References

1. O'Meara, C. C., Wamstad, J. A., Gladstone, R. A., Fomovsky, G. M., Butty, V. L., Shrikumar, A., Gannon, J. B., Boyer, L. A., & Lee, R. T. (2015). Transcriptional reversion of cardiac myocyte fate during mammalian cardiac regeneration. *Circulation research*, *116*(5), 804–815. https://doi.org/10.1161/CIRCRESAHA.116.304269
2. Porrello ER, Mahmoud AI, Simpson E, Hill JA, Richardson JA, Olson EN, Sadek HA. Transient regenerative potential of the neonatal mouse heart. *Science* . 2011;331:1078–1080. doi: 10.1126/science.1200708
3. Bicknell KA, Coxon CH, Brooks G. Can the cardiomyocyte cell cycle be reprogrammed? *J Mol Cell Cardiol* . 2007;42:706–721. doi: 10.1016/j. yjmcc.2007.01.006.
4. Hadley Wickham, Romain François, Lionel Henry and Kirill Müller (2022). dplyr: A Grammar of Data Manipulation. R package version 1.0.9. https://CRAN.R-project.org/package=dplyr
5. Kirill Müller and Hadley Wickham (2021). tibble: Simple Data Frames. R package version 3.1.6. https://CRAN.R-project.org/package=tibble
6. Hadley Wickham, Jim Hester and Jennifer Bryan (2022). readr: Read Rectangular Text Data. R package version 2.1.2. https://CRAN.R-project.org/package=readr
7. Wickham et al., (2019). Welcome to the tidyverse. Journal of Open Source Software, 4(43), 1686, https://doi.org/10.21105/joss.01686