Characterizing alternative first exons from stem cells and cardiomyocytes

Miguel Yanez[1,2], Maritere Uriostegui-Arcos[3], Ana Fiszbein[3]

[1]Whittier College, [2]Boston University Bioinformatics BRITE REU Program, Summer 2021,[3] Department of Biology, Boston University

Abstract

Being able to further understand how RNA processing events can affect gene expression is essential to recognizing how to alter transcriptional output of certain genes. It is known that errors in splicing can lead to disease, and splicing mutations are estimated to cause as much as sixty percent of human genetic diseases, so understanding how messenger RNA molecules are spliced and transcribed is important to understanding how diseases occur. The goal of this project was to classify the splicing characteristics of exons into one of five categories: first, hybrid internal first, internal, hybrid internal last, or last exons. Furthermore, we will use this classification to observe if cell differentiation changes the pattern of exon usage. Samples from stem cells and differentiated cells at different time points post-differentiation were used to analyze first exon frequency and to compare their AFEPSI (Alternative First Exon Percent Spliced In) values, a measurement that indicates an exon's relative usage as first exon compared to the other alternative first exons in the gene. An analysis using an FIHL (First-Internal-Hybrid-Last exon) index will allow us to classify the exons as first, last, internal or hybrid exons.

Background

Many genes contain alternative first exons, i.e., exons within a gene that can serve as first exons depending on how the gene is transcribed. They can be termed hybrid exons, since they can serve as first exons or internal exons depending on how the gene gets spliced and transcribed. There have been exons termed "hybrid" exons that have made the human transcriptome a more complex topic. With errors in splicing regulation being tied to disease, further knowledge on how splicing events and transcription occur will be critical to human health.

One of the goals was to determine if cells under different differentiation states show different patterns of first exon usage

Figure 1: A diagram that shows that the first two exons can work as first exons. The second exon (AFE2 hybrid) can also be internal based on splicing and transcription

Methods

- Data from different cell samples gave a description of their exons. From there we constructed violin plots to visualize how many possible first exons were found in each gene
- The same data came with an AFEPSI measurement that allowed us to build heat maps to show the relative usage of the exons as first exons compared to the rest in the same gene
- Created the plots described above for three sets of data
- The data files came with a FIHL index that will allow us to characterize the exons into first, last, internal or hybrid exons in the form of a heatmap.

cell line	day	cell identity
DCMwt	0	pluripotent
DCMwt	0	pluripotent
DCMwt	0	pluripotent
DCMwt	1	CHIR 24h
DCMwt	1	CHIR 24h
DCMwt	1	CHIR 24h
FN2.1wt	0	pluripotent
FN2.1wt	0	pluripotent
FN2.1wt	0	pluripotent
FN2.1wt	3.5	mesoderm progenitor
FN2.1wt	3.5	mesoderm progenitor
FN2.1wt	3.5	mesoderm progenitor
FN2.1wt	21	cardiomyocyte
FN2.1wt	21	cardiomyocyte
FN2.1wt	21	cardiomyocyte

Table 1: This table shows the sample cell lines and the differentiation states of the cells. The cells start off as Embryonic stem cells.

Results

Graphs for the first six samples:

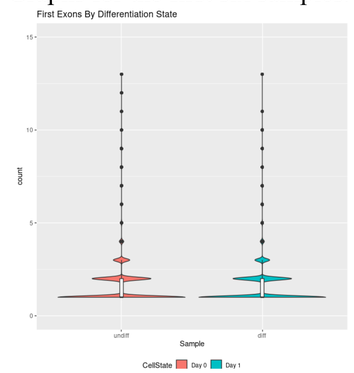

Figure 2: Violin plots showing the distribution of how many exons are first exons in each of the genes between undifferentiated samples (left and samples after 24 hours (right).

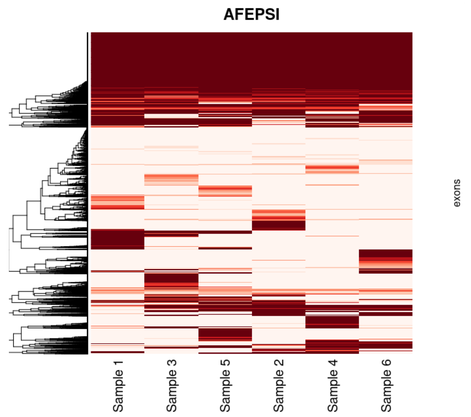

Figure 3: Heat map showing all the exons that were present in the samples. Samples 1,3, & 5 correspond to undifferentiated and samples 2,4 & 6 correspond to differentiated samples after 24 hours

Graphs for the nine samples:

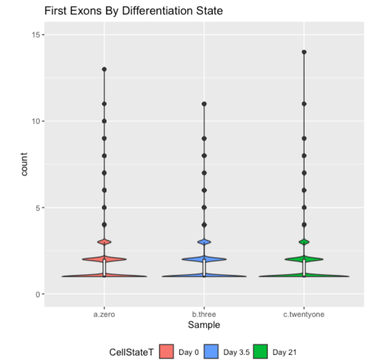

Figure 4: Violin plots similar to Figure 2. Samples from undifferentiated cells, differentiated cells after 3.5 days and after 21 days were used

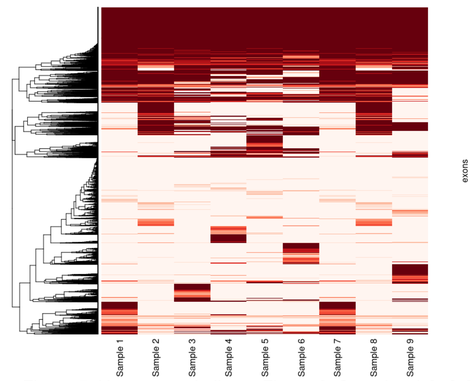

Figure 5: Heap map similar to Figure 3. Samples 1,2, & 3 correspond to undifferentiated cells. Samples 4,5 & 6 correspond to samples after 3.5 days. Samples 7,8, & 9 correspond to samples after 21 days.

Graphs for one stem cell and three differentiated samples:

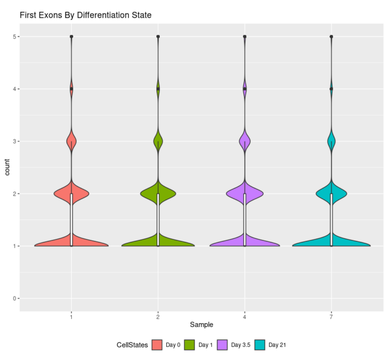

Figure 6: Violin plot for samples that include undifferentiated cells, and differentiated samples after 1, 3.5, & 21 days

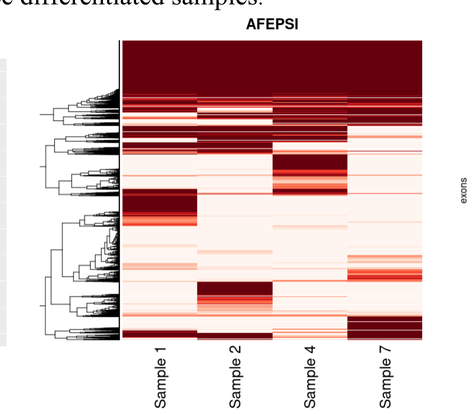

Figure 7: Heat map of exons that appear in the samples of undifferentiated cells and cells after 1 (sample 2), 3.5 (sample 4), & 21 (sample 7) days.

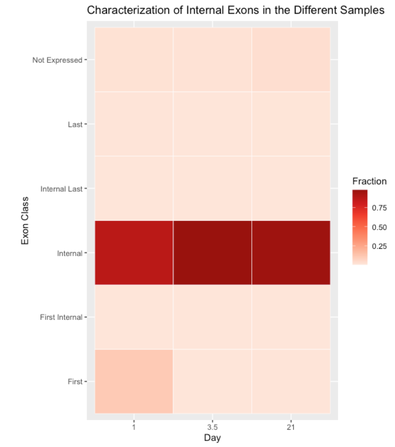

Figure 8: Using the same samples from Figure 6 & Figure 7, the heatmap shows different types of exons for each sample. This figure is supposed to show what happened to the internal exons as differentiation occurred

Discussion

- In the violin plots, we see that most genes only have one exon that acts as a first exon but based on the data in the figures, there are a few genes that have multiple exons that could act as a first exons. Something to look into would be what kind of genes are more likely to have multiple exons that can act as first exons.
- The Heatmaps were done to give us a visualization of when the exons are appearing as differentiation occur. This would help us identify any patterns that occur during cardiomyocyte differentiation. For example, on Figure 3, a bit beneath the middle of the figure, there is a dark red band that appears on samples 1,3 & 5 but in Samples 2,4,& 6 , that same group of exons do not appear.
- In Figure 8, most of the internal exons from the undifferentiated sample stayed as internal exons in the differentiated samples, but we do see that after 24 hours, some of those internal exons do become First exons and after 21 days, there is a portion of internal exons that are not expressed.

For future work: Look into the number of hybrid first internal exons per gene in each of the samples. Also look into first exons that change significantly over differentiation and explore their roles in cardiac differentiation.

References

Ana Fiszbein, Keegan S. Krick, Bridget E. Begg, Christopher B. Burge (2019). Exon-Mediated Activation of Transcription Starts. *Cell*, Volume 179(7), 1551-1565

Ana Fiszbein, Michael McGurk, Ezequiel Calvo-Roitberg, GyeungYun Kim, Christopher B. Burge, Athma A. Pai (2021). Widespread occurrence of hybrid internal-terminal exons in human transcriptomes. BioRxiv

Douglas L. Black (2003). Mechanisms of Alterative Pre-Messenger RNA Splicing. *Annual review of biochemistry*, Volume *72*, 291–336.

Guey-Shin Wang & Thomas A. Cooper (2007). Splicing in disease: disruption of the splicing code and the decoding machinery. *Nature Reviews Genetics*, Volume 8, 749-761.

Acknowledgement

This work was funded, in part, by NSF grant DBI-1949968, awarded to the Boston University Bioinformatics BRITE REU program