Log 1:

**Differential Gene Expression in Neural Stem Cells and iPSCs**

(yeah, i think) Shishir, ChatGPT(no, really)

**Abstract**

So, I just wanted to do something to put into my Github, so while this does look like a paper externally, it’s just me writing down things, ok?. I basically found out differences between Neural Stem Cells (NSCs) and induced Pluripotent Stem Cells (iPSCs). This was done because I tried finding out about immortality in *Turritopsis dohrnii*, the immortal jellyfish. It had to do something with transdifferentiation, where the normal cells are converted back into stem cells for redifferentiation. So my dumb idea was to find out what genes are different in both types of cells and see if those genes are the ones really important for transdifferentiation. This may be entirely wrong, but that’s alright it was something to do. Ok let’s go.

**Introduction**

In terms of developmental sequence of neurogenesis (fancy paper words), iPSCs come first and these have the potential to divide into NSCs (Galiakberova et al., 2023), and these NSCs divide into other important neurons and stuff. So naturally finding out what genes make these two types of cell different allows us to see what genes we can potentially target to convert NSCs back into iPSCs (yeah, seems alright on paper, this might have already been done before). Converting them back would allow differentiation to happen again, just like in the immortal jellyfish. Saying that, I should have used jellyfish neurons but I didn't find any, so I used mouse ones instead.

**Methods**

So I used python cause I like programming (kinda) and I have done this kind of thing there before. So, how did I get my data? I used [GEO](https://www.ncbi.nlm.nih.gov/geo/) where I got the SVZ neural stem cells and the iPSCs microarray analysis data. Both used the same platform file as well. Both were from mice.

So I load up these files using pandas and clean it up and put them into two datasets. Next, there was this useful package called statsmodels that I used to calculate a t-test which basically tells me if there is a statistically significant difference between the two types of cells. I then did the FDR correction 💢💢 where I filtered for FDR\_p\_values that are less than 0.05 i.e. 5% . Turns out this was too much so then I filtered it out 0.01 i.e. 1%. Which was still too much. It gave me around 29,500 “significant” genes. But that was something.

Next, I used plotly(great package, btw) to create a volcano plot of the genes. Next I calculated the mean expression level of each gene across the whole sample. This was needed to calculate Log2FC, this number basically tells us how much that gene is different across both the samples, this is important because I wanna use this for the volcano plot.

So, if you’ve never seen a volcano plot before, basically, in this case the further apart the genes are from 0, the more they are expressed (upregulated or downregulated). Next, I filtered out the top 10 most significant genes and their IDs.

I used the Gene symbols from the GPL file, and mapped them onto the most significant genes to get their symbols. I also made a heatmap of these 10 significant genes across the samples to see where these were overexpressed and underexpressed. Good to see there was significant clustering.

For the final bit, I used sklearn to calculate a PCA with 3 components. Using a 3d plotly plot, I plotted the results. Looks really cool.

**Results:**

The biggest disappointment other than me, is the fact that I cannot directly put the 3d plot on this page. Shame.

So, as I mentioned I got 10 significant genes and I mapped these genes to get the symbols and their GO IDs from the GPL file. These genes were : Dcx, Foxg1, Psmb1, Hoxa1, Rps24, Gm5963, Rps10, Dscaml1, Dleu7 and Tmsb10. These names are nice. But what do we know about these genes that's where the GO IDs come in. And here is the other big disappointment. I was too lazy to calculate Enrichment factors and DAVID was down as well, I was becoming impatient. So I just used these significant genes and their IDs to get the potential important differentiating functions.

So the first step was to find out the genes that are different using the volcano plot. The result is in Figure 1.

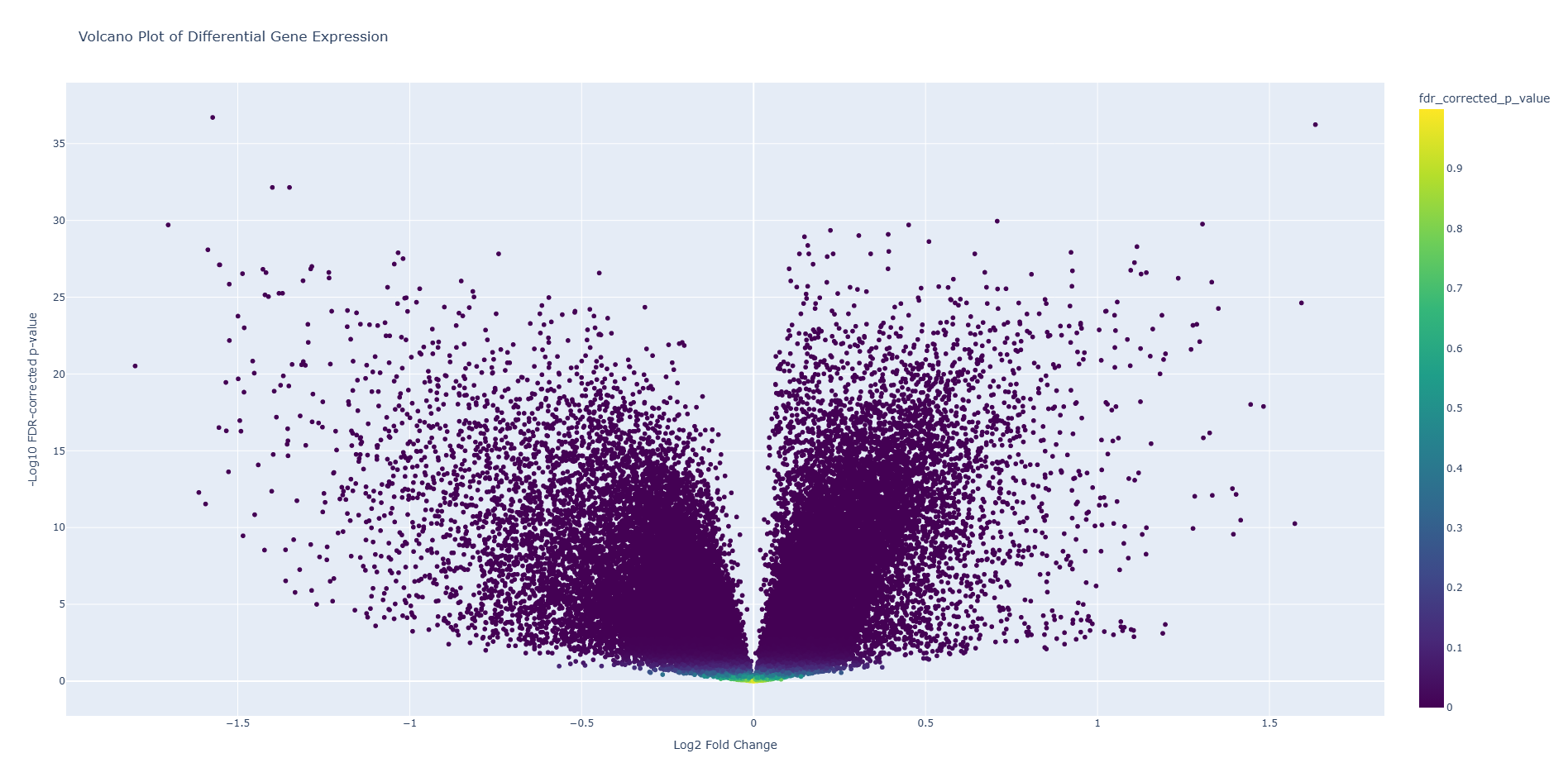


Figure 1. Volcano plot of the most significantly different genes in iPSCs and NSCs. The points at the top are the most significant. The points on the right are upregulated , the points on the left are downregulated.



Random Trex image.

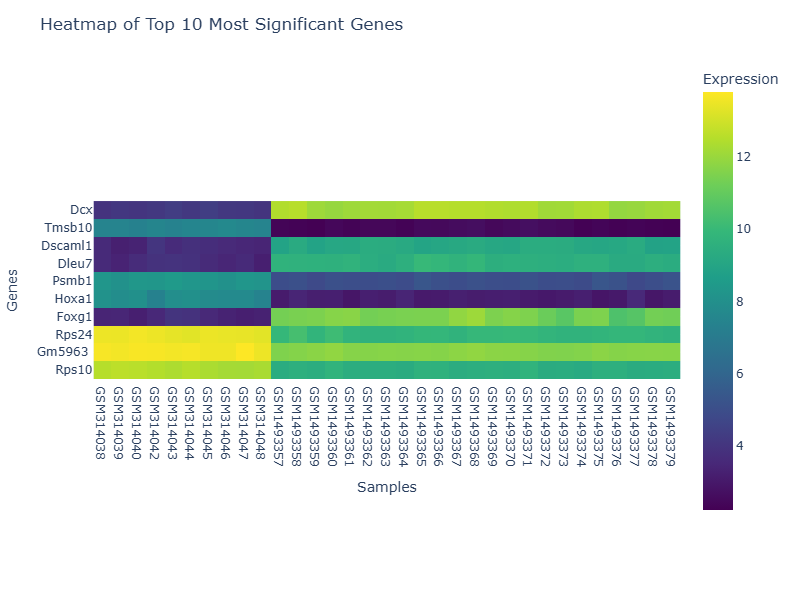
So from Figure 1. The points at the top and the ones that are far apart are the most significant of the genes, these were those significant genes that I mentioned. Next was the heatmap, to see in what samples were these genes upregulated and downregulated. The results are in Figure 2. 

Figure 2. Heatmap of the significant genes. The first 10 samples from the left are from the NSCs and the remaining are from the iPSCs. There is significant clustering seen between the two groups. Mainly in genes like Dcx, Tmsb10, Dscaml1 and Dleu7.

Finally, I made a 3 component PCA plot, which I felt was the most interesting part. This is a 3d plot and since I can’t put a 3d plot here there is just 2D visualization in Figure 3. The actual html files are there in the repo.

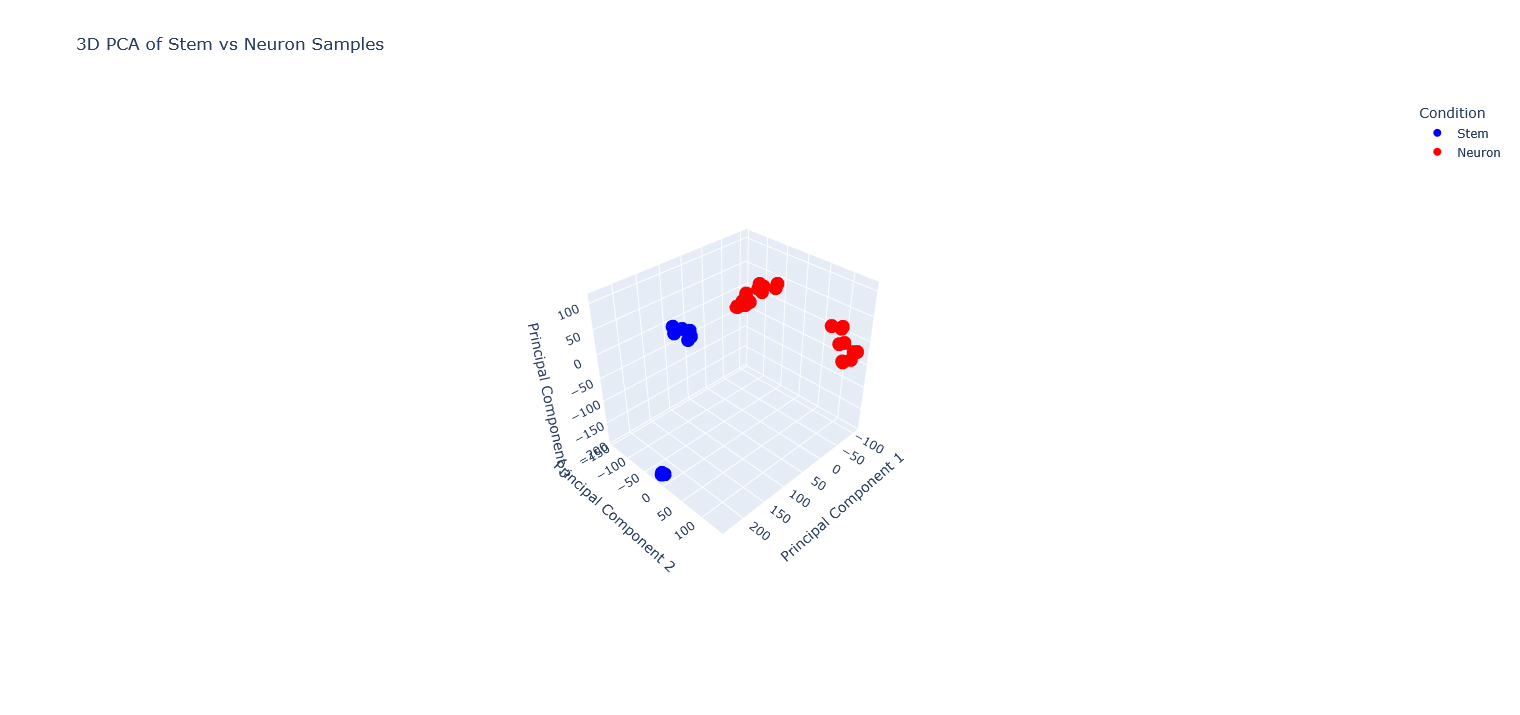


Figure 3. A 3 component PCA plot that captures variances in the samples. The ones in blue called stem are the NSCs and the red ones called neuron are the iPSCs

In the PCA, where PC1 captures the most variance, there is a difference between the blue and red ones, which is obvious because these are two different types of cells. Along PC2 however, there is a separation within the iPSCs this is because the original data consists of two different iPSCs one from the dorsal and the other from the lateral subventricular zone (idk what these mean, but it has to do something with early development). PC3 is interesting because it shows a separation in the NSC samples and I am not entirely sure as to why. I am assuming it is because these were collected at different times or during different stages of development.

**The genes:**

I'll be honest, most of the things here I’ve taken from wikipedia and ChatGPT

**DcX** Neuronal migration protein doublecortin, this gene is downregulated in the NSC samples and upregulated in iPSCs. According to the wiki, that checks out, these genes are expressed in early stages and downregulated in later stages. Heck, this is used as a marker for neurogenesis.

Doublecortin is a protein made by the DcX gene, and this protein binds to the microtubules and helps in bundling, which basically means it acts as a glue that binds microtubules together ensuring stability for the growing neurons. Without this, it will make you smooth brained,Wikipedia's words, not mine.

**Tmsb10**: Apparently you can buy this protein [here](https://www.abbexa.com/human-thymosin-beta-10-tmsb10-protein). It’s called Thymosin Beta 10. This is upregulated in the NSCs and downregulated in the iPSCs. Wikipedia tells me that this is mainly for cytoskeleton formation and remodelling. Now, I haven't read much on this because I'm lazy. I assume this is upregulated in the NSCs cause they have to be highly flexible and motile and have to form connections, like regular ass neurons. Unlike the iPSCs who haven't made a choice to commit to a cell type, so they don’t need it as much.

**Dscaml1:** downregulated in NSCs and upregulated in iPSCs, the main function is to help cells stick together. Now I assume this is because these iPSCs require other cells of its type to be together so that they can coordinate as a team and decide what they as a group would become, but in NSCs they are egoists, like in blue lock. They wouldn't wanna affect each other by sticking so close to each other. That’s my thinking.

**Dleu7**: is special because this is a gene regulator, this tells us other genes what they can or cannot be, like a dictator. These are downregulated in NSCs and upregulated in iPSCs. These guys mainly control tumor suppressor genes. Now I think, these are more strict and more upregulated in iPSCs because these cells have more of a potential to become cancerous, since they are so pluripotent and any leniency from Dleu7 could break that cycle and make these guys cancerous. Essentially these correct the iPSCs(💢💢😭) and prevent them from over proliferation.

**Psmb1**: This protein forms a part of a proteasome, which is this thing that just breaks down bad, unwanted proteins in the cell. These are downregulated in the iPSCs because I assume they need all the proteins they produce and get to grow but in NSCs these need to be cleared because they are less hardy than iPSCs and over accumulation of these bad proteins would be bad for these delicate boys.

**Hoxa1:** These control other Hox genes and are mainly involved in further segmentation, as in when expressed in NSCs they can form more specialised cells, particularly cells in the hindbrain. Since iPSCs have not committed to any single lineage, they wouldn't need this as they first have to become NSCs .

**Foxg1:** Now, this is pretty important as a lack of this gene product is associated with the aptly named Foxg1 syndrome.(**MedlinePlus** (Internet) *FOXG1 syndrome****)*** . Its most notable condition is microcephaly. This means that what determines whether someone gets this syndrome is decided by how this gene is regulated in the iPSCs itself. In the sample that I tested, it is upregulated in iPSCs and no microcephaly was reported in the mouse, I think. So, you could tell someone would get this syndrome by looking for this gene activity in iPSCs itself.

The other genes were not that different, so I didn’t bother writing them up.

**Conclusion:**

I just wanted to try and do something that I did in my assignments but of my own free will(philosophy). There are a lot of errors and probably violations of the scientific method, but eh. If you read this, thank you. Again, if it isn’t obvious this is not an actual paper(shocking, I know). This is just my thoughts and a remainder of what I did and something to add to my Github. Might be a mistake to do that actually. I might do something with bear genomes next time, idk.

**References:**

**MedlinePlus** (Year) *FOXG1 syndrome*. Available at:<https://medlineplus.gov/genetics/condition/foxg1-syndrome/> (Accessed: 2 March 2025).

Galiakberova, A.A., Brovkina, O.I., Kondratyev, N.V., Artyuhov, A.S., Momotyuk, E.D., Kulmukhametova, O.N., Lagunin, A.A., Shilov, B.V., Zadorozhny, A.D., Zakharov, I.S., Okorokova, L.S., Golimbet, V.E., Dashinimaev, E.B., 2023. Different iPSC-derived neural stem cells show various spectrums of spontaneous differentiation during long term cultivation. *Frontiers in Molecular Neuroscience*, 16. Available at:<https://www.frontiersin.org/journals/molecular-neuroscience/articles/10.3389/fnmol.2023.1037902> [Accessed 2 Mar. 2025]. DOI: 10.3389/fnmol.2023.1037902.

Data: <https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE12499>

<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE60905>