Gene expression differences between Neural Stem Cells and human Embryonic Stem Cells

Mikk Puustusmaa[1], Märt Roosaare[1], Elena Sügis[2]

[1]Department of Bioinformatics, Institute of Molecular and Cell Biology, University of Tartu
[2]Institute of Computer Science, University of Tartu, Quretec OÜ

Contact: mart.roosaare@ut.ee

Introduction

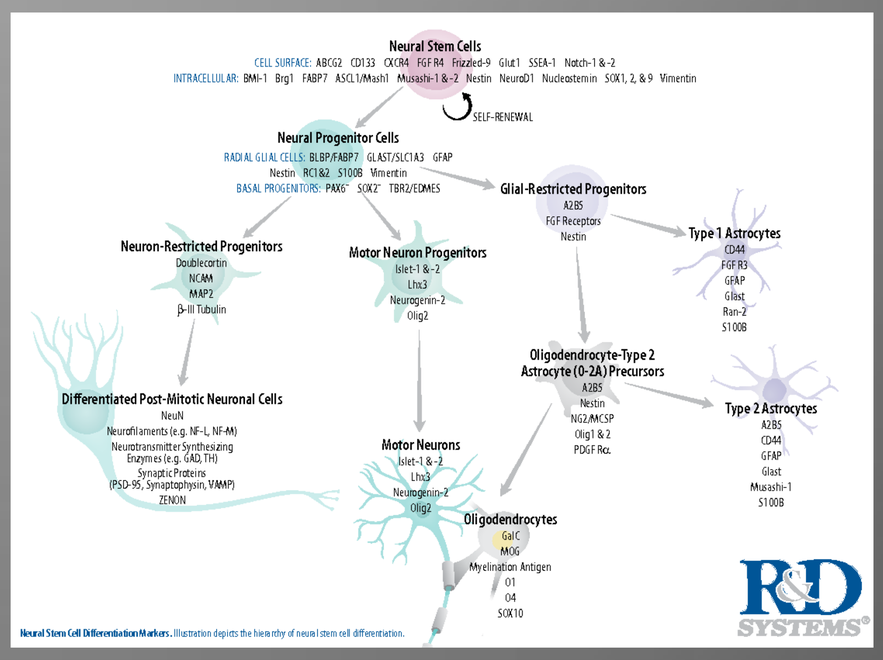
Neural stem cells (NSC) are multipotent cells that are capable of differentiating into the main phenotypes of the nervous system - producing a diversity of neurons, astrocytes and oligodendrocytes. Given their ability to differentiate into various types of neuronal lineages, these cells can be applied to cure neurodegenerative diseases. In vitro NSCs can currently be derived from human embryonic stem cells (hESC). In order to successfully apply NSCs in regenerative medicine it is important to understand the underlying mechanisms of hESC differentiation towards neuronal lineages.

Some possible future uses for NSCs could be:

- Myelin restoration (using oligodendrocytes)
- Mixed pools of progenitors to cure complex populations of neural cells (spinal cord injuries)
- Laboratory experiments where neural cells are required

The main aim of the project is to understand what genes separate hESCs and NSCs originated from different experiments and microarray platforms. In addition, we investigate how much does the technology and human factor influence the derived cell lines.

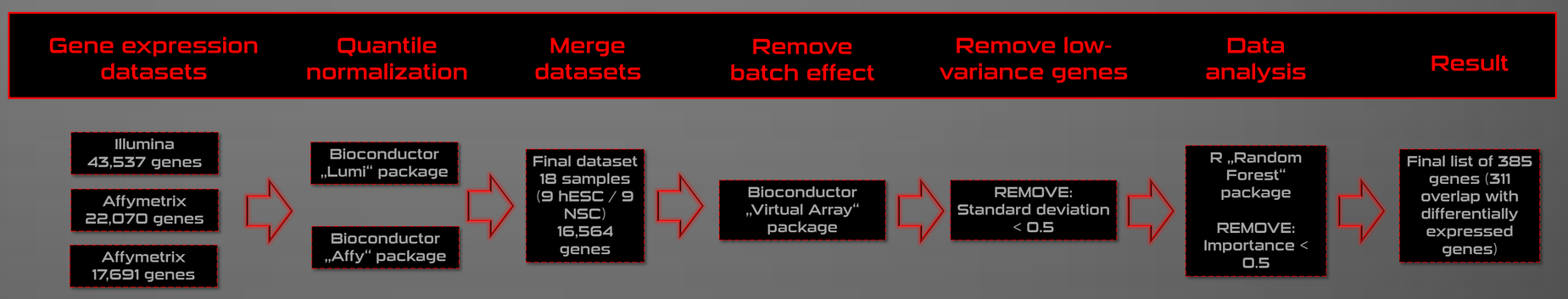
Neural Stem Cell differentiation



Methods

We used datasets from two Affymetrix and one Illumina microarray. Data preprocessing and analysis was done in R, mainly using tools from the publicly accessible Bioconductor repository. Data analysis was done using the Random Forest algorithm (RF) that uses multiple decision trees to classify the data. We picked out the genes which had highest importance score - these were able to separate hESC and NSC datasets the best. We also used *pheatmap* package to visualize differences in gene expression between hESC and NSCs.

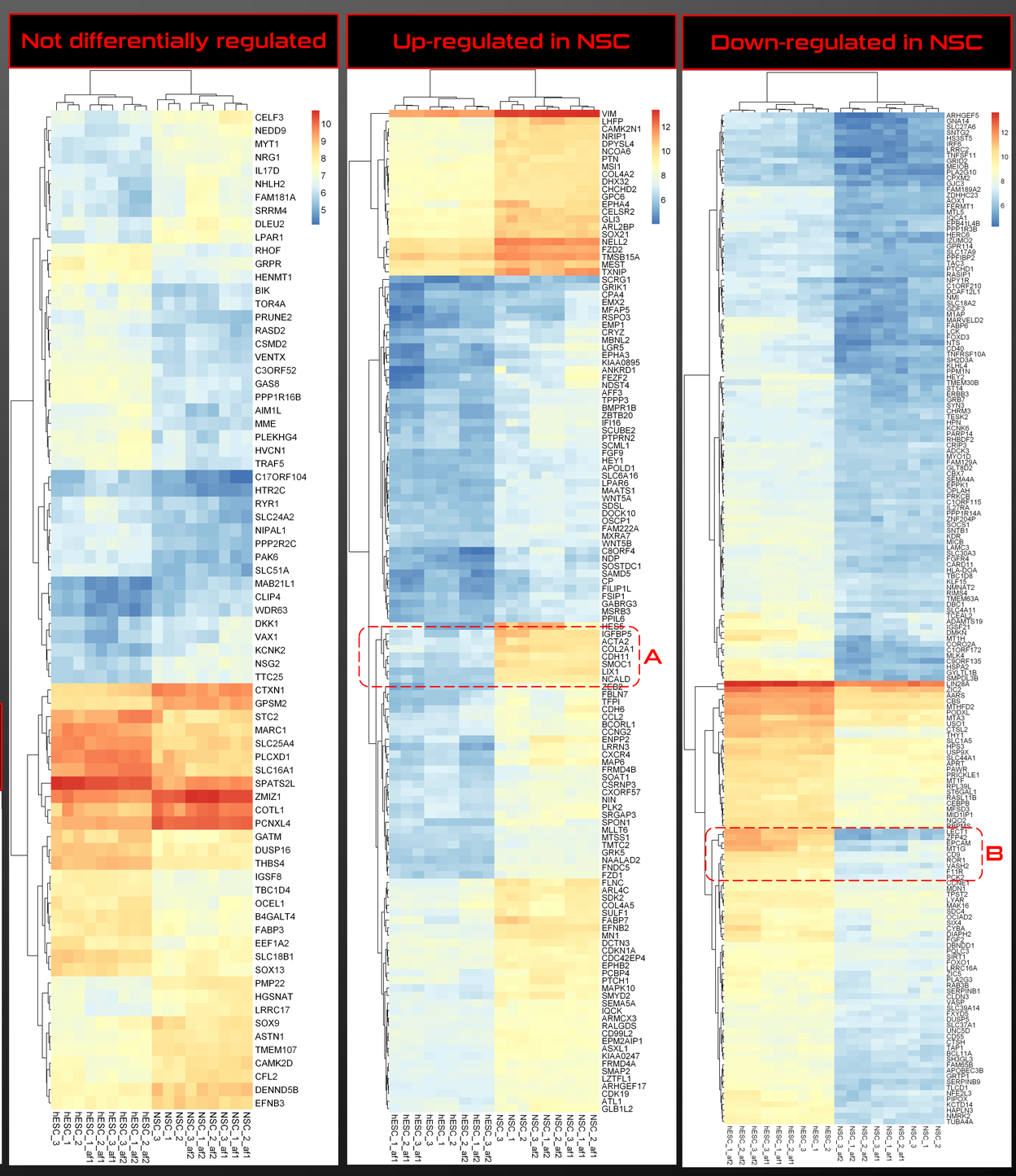
Our pipeline - from three initial gene expression datasets to final results



Results

Using the Random Forest algorithm (RF), we got a list of **386** genes that separate two cell lines the best.
We found a significant overlap between the differentially expressed and 386 selected genes: **178** genes were also down-regulated and **133** genes up-regulated in NSCs. Remarkably, **75** of the selected genes were not differentially expressed, but were still important to discern between hESC and NSC cell lines according to RF.
When comparing the expression of the selected genes of the NSC cell lines, we could not detect any bias that could have resulted from using different technologies. We observed similar results for hESCs as well (**figures on the right**).
We also categorized selected genes by Gene Ontology (GO) terms (**figure below, *GOsummaries* package**). Among the down-regulated NSC genes, epithelium development and cell adhesion stood out. Up-regulated genes were largely associated with tumorigenesis and neuronal development. Non-differentially expressed ones were associated with ErbB signalling pathway.
A few gene groups have large expression differences, among them **group A** consists of genes tied to connective tissue development, whereas in **group B,** no GO-terms stand out significantly (**figures on the right**).

Selected gene expression heatmaps from three different experiments (each consisting of three hESC and three NSC lines)



Conclusions

- In the down-regulated geneset of NSCs, **epithelium development** and **cell adhesion** stood out
- In the NSC cell line, up-regulated genes were connected to **cancer pathways, vasculature and neuronal development**
- Among the genes that were not differentially expressed, **ErbB signalling** stood out. ErbB is an epidermal growth factor, overexpressed in many cancers

Overrepresented Gene Ontology terms