SINGLE CELL TRANSCRIPTOME ANALYSIS REVEALS DISCRETE EVENTS

DURING NEURAL PRECURSOR CELL DIFFERENTIATION INTO NEURON

By

SATYA PAVITRA RANI

A thesis submitted to the

School of Graduate Studies

Rutgers, The State University of New Jersey

In partial fulfillment of the requirements

For the degree of

Master of Science

Graduate Program in Biomedical Engineering

Written under the direction of

Li Cai, Ph.D.

And approved by

_____

_____

_____

New Brunswick, New Jersey

October 2018

ABSTRACT OF THE THESIS

SINGLE CELL TRANSCRIPTOME ANALYSIS REVEALS DISCRETE EVENTS

DURING NEURAL PRECURSOR CELL DIFFERENTIATION INTO NEURON

By SATYA PAVITRA RANI

Thesis Director:

Dr. Li Cai

Neurogenesis is a complex process which is controlled by intrinsic programs and exogenous signals among the cells resulting in formation of progressively committed neurons. The complex processes governing neurogenesis can be elucidated by determining the developmental stages and compositions of the cells along with their molecular determinants. The single cell RNA sequencing (scRNA-seq) study of undifferentiated neural precursor cells ultimately differentiating into neurons is an effort to understand these mechanisms. This study is performed on a dataset (GEO accession #GSE102066) consisting of four cell types, i.e., neural precursor cell, differentiating neural cell, immature neuron and neuron. A comprehensive analysis examining the trajectory path of neural precursor cells to neurons, the highest expressed genes and marker genes of each cell type and the gene ontology (GO) terms associated with these genes has revealed critical observations. Lack of similar intrinsic properties revealed heterogeneity among immature neurons and neurons through trajectory analysis. The genes which were expressed highest among all the cell types was associated with GO term for housekeeping genes which are crucial for maintaining cell functions. The marker genes of each cell type were associated

with many relevant GO terms such as neurogenesis, neuron projection, olfactory lobe development, forebrain generation of neurons and central nervous system development. Key factors having discrete gene expression activity as cells transitioned from one state to another were revealed in this study. Genes which drastically decreased their expression levels over a period of two days (day 5 to day 7) were SFRP1 and SFRP2, genes which gradually decreased their expression activity from day 0 to day 30 were WNT5A and EFNA5, known neuronal marker genes DCX, MAP2 and STMN2 were upregulated gradually as the cells reached mature neuron stage. This study has helped in revealing transcriptional profiles of cells differentiating through a critical period of time with the key genes responsible for maintaining cell states and determining cell fates.

Acknowledgements

I would like to thank my advisor Dr. Li Cai for his guidance during the entire period of my thesis. I would also like to thank Dr. Mark Pierce and Dr. Jay Sy for being a part of my defense committee and helping me review my thesis. I want to thank my lab members, especially Alex for always helping me out.

I want to thank my mother, father and sister for their ongoing support, constant encouragement and for always believing in me. Without them this would not be possible. I also want to thank my grandmother and uncle for their support. I would like to thank my cousin Teja for always encouraging me and my close friend Aishwarya for always being there for me. Special thanks to Sravanthi and Alekhya for their help and patience.

I would also like to thank Sujatha aunty and her family, Duri uncle and his family for their time. Lastly, I want to thank dear friends and family who helped me in any way throughout this entire journey.

Dedications

This thesis is dedicated to my Mother, Father and Sister.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

DNC	Differentiating Neural Cell
GO	Gene Ontology
hESC	Human Embryonic Stem Cells
IN	Immature Neuron
N	Neuron
NPC	Neural Precursor Cells
NSC	Neural Stem Cell
scRNA-seq	Single cell RNA sequencing

1. Introduction

1.1 Human Embryonic Stem Cells

Human embryonic stem cells (hESCs) are derived from pre-implantation or peri-implantation embryos having the ability for self-renewal, which is the ability to proliferate without differentiation [1]. The hESCs cells are pluripotent and have the potential to give rise to three germ layers- ectoderm, mesoderm and endoderm, which undergo organogenesis[2]. The ectoderm germ layer forms the nerve cells within the brain, spinal cord, hair, skin, teeth, sensory and pigment cells. The mesoderm forms connective tissues, blood vessels, and muscles in the body. The endoderm forms the gut, lungs, bladder and germ cells [2].

The paper by James A. Thomson, et al. describing the generation of hESCs stated that 'these cell lines should be useful in human developmental biology, drug discovery and transplantation medicine'[3]. The hESCs can show great promise in areas of tissue repair and regeneration by generating a clinically relevant number of cell populations and also patient specific ESC equivalents to study disease mechanisms, establish screens for drug discovery and mechanisms of the immune systems [4].

Studies have been conducted to understand the biology of hESCs to achieve neural differentiation as obtaining the samples from postmortem mammalian brain, especially human brain can be challenging due to the presence of highly dense and myelinated extracellular tissue [5]. Also the dissociation of a whole intact cell is difficult due to the conditions in which the postmortem sample are stored [5]. The therapeutic potential of

hESCs derived dopaminergic neurons in Parkinson's rat model showed improved functionality [6]. Motor defects in hemi Parkinson mice model showed improvement in motor skills through dopamine neuron engraft by ESC-derived neuronal progenitor cell transplantation [7].

1.2 Neural Stem Cells and Brain

Molecular clues and developmental processes can be studied by differentiating hESCs into neural lineage. This can help in understanding mechanisms underlying early neurogenesis involved in the development of the human nervous system [8]. Neurogenesis is defined as the transition of neural precursor cells (NPCs) or neural stem cells (NSCs) to differentiated mature neurons [9]. The coordinated action of multiple cues and gene expression pattern act on NSCs ultimately giving rise to a population of neuronal cells and glial cells present in the mature brain [10]. During the brain development, NPCs and neurons form a complex neural network along with the glial precursor cells, astrocytes, and oligodendrocytes which support the neural function and metabolic activities [11].

It is necessary to track dynamic cell differentiation guided by expression and repression of specific genes at the experimental and computational level in order to understand and reveal discrete events between cell states [12], which can be linked to cell fate specification[10]. In vivo, clonal analysis of individual NSCs show they have only neuronal lineages, whereas the in vitro time lapses analysis of NSCs show that they either generate neurons or astrocytes [13].

Throughout the years there has been significant evidence that the NSCs are persistent in the adult life of all mammals including humans. They have the capacity to proliferate and produce glial cells and neurons in the canonical regions (subventricular

zone, sub granular zone) and non-canonical regions (neocortex, striatum and hypothalamus) of the brain [11].

Understanding these processes and how neurons integrate themselves into complex neural networks can help in developing novel strategies for brain repair. Current studies focus on taking advantage of the NSCs regenerative capacities by examining and manipulating the regulatory mechanisms. Future studies will benefit by focusing on identification of subpopulation and state-specific markers [13]. This can be achieved by performing single cell analysis to reveal the properties of various neural stem cell or precursor cell properties within their sub-populations and dynamic states.

1.3 Single cell RNA sequencing analysis of mammalian brain

Traditionally, the cells within the brain are classified based on morphology, molecular markers, location and electrophysiological characteristics [14]. The limitations of markers for the identification and isolation of rare neurons from nervous tissues makes it challenging to capture rare dynamic processes in adult neurogenesis [15]. The neurons belonging to a subtype have variance in their properties which constantly change with time [16]. These challenges make it difficult to identify the dynamic changes within the neuron types and elucidate the heterogeneity properties of cells within the brain.

Single-cell RNA sequencing (scRNA-seq) is one of the high throughput sequencing methods that provides us with insights into the transcriptome of a cell [17]. Recent studies have demonstrated that the scRNA-seq helps in transcriptional profiling of each individual cell revealing the rare cell types and short-lived cell populations as the cells transition through different regulatory mechanisms and biological processes [14], it has been used to

study the cell diversity of many tissues such as the spleen, embryonic brain and the lung epithelium [18]. ScRNA-seq analysis is a useful means to elucidate the differences between different cell subpopulations during their development and the mechanisms which have helped to attain these cell fates.

The number neural subtypes present in the brain is not clear [19]. There have been recent studies single cell profiling studies of the brain, while substantial progress has been made in mice a full understanding of adult human neurons on basis of single cell transcriptomics is yet to be realized [20]. Recently, single cell studies characterizing the dopaminergic neurons (DA) in embryonic and post-natal mouse revealed gene expression profiles specific to certain DA subtypes such as postnatal neuroblast population and substantia nigra DA neurons [21]. Previously, scRNA-seq analysis of region-specific cells, such as the mouse hypothalamus have revealed 11 non-neuronal and 34 neuronal cell clusters having distinct transcriptional profiles [22]. The heterogeneity of the cell population within the regions of the brain were understood by examining the cerebral cortex transcripts of the mouse brain revealing similar interneurons in different regions of the brain and that the oligodendrocytes differentiated into many cell types due to changes in their molecular signatures [18]. Gokce, et al. focused on analyzing 1,208 mouse striatum single cells which revealed 10 differentiated cell types including neurons, astrocytes, vascular and immune cells, oligodendrocytes and ependymal cells, apart from examining cell diversity they were able to observe medium spiny neurons (MSNs) which overexpressed genes linked to cognitive disorders [23]. ScRNA-seq analysis has been helpful in studying the process of neurogenesis in dentate gyrus which is a part of the hippocampus region of the mouse [24, 25]. Shin, et al. studied the transcriptional activity

of quiescent NSCs in adult dentate gyrus, their activation and initiation of neurogenesis in a transgenic mouse model revealed a continuum in adult neurogenesis [24]. Recently, Hochgerner, et al. published a similar study, developing a map of granule cell development in a perinatal attempt to elucidate neurogenesis from early precursor cells to mature neurons revealed that neurogenesis is preceded by distinct cell states and the cell properties are conserved from E16.5 to P132 [25]. Other animal models such as the zebrafish [26] and drosophila [27, 28] have been studied to examine their neuron subtypes. A recently published paper by Pandey, et al. attempted to create a comprehensive gene expression atlas of the habenula region of the zebrafish brain. Their study revealed that the habenula which is a small region of the forebrain had 18 distinct habenular subtypes which were stable across larval and adult stages [26]. In case of drosophila, one study aimed at classifying the olfactory projection neuron transcripts [27] and another study focused on developing a complete atlas of the adult drosophila through its lifespan for examining cellular diversity [28]. This gives us an idea about the complexity of the mammalian brain and the need for gene expression-based classification of cell types in all the regions of the brain.

Attempts have been made to elucidate the adult human brain by transcriptome analysis. A single cell transcriptome analysis of the human olfactory system revealed difference between olfactory progenitor cells (OPCs) and mature olfactory sensory neurons (OSNs) cells, it was governed by different signaling pathways that were upregulated in progenitor cells in early developmental stages [29]. ScRNA-seq has helped to validate the properties of cells developed from hESCs. A recent study showed medial ganglionic eminence (MGE)-like cells generated from in vitro hESCs mimicked the in vivo cells derived from

human fetal interneurons and glial cells in both early and mature stages [30], making it the first transcriptome analysis comparing in-vivo human fetal-derived interneurons and hESCs derived interneurons. Yao, et al. studied the early embryonic brain development by analyzing hESCs transcripts differentiating into distinct neural cell types revealing Wingless-Type MMTV Integration Site (WNT)- signaling governing the lineage of forebrain and mid/hindbrain neuron types which were validated by already established regional differences in neurogenesis [31]. Recent study by Fiddes, et al. by computational analysis revealed that the gene NOTCH2NL secretes NOTCH-like proteins that can prolong cortical neurogenesis by enhancing NOTCH signally and hence inhibit the differentiation of NPCs [32].

Here, we have attempted to elucidate the processes of neurogenesis by analyzing the dataset (GSE102066), consisting of undifferentiated NPCs which go through the main stages of differentiation in a 30-day period, previously this dataset has been analyzed by Wang, et al. by single cell RNA sequencing analysis. The scRNA-seq pipeline by Wang, et al. for processing, mapping and estimating gene expression was as follows—the data was processed and mapped using TopHat, the gene counts were obtained using HTSeq-count, after applying a quality criterion to filter low expressed genes and cells a total of 8957 expressed genes and 483 cells were normalized in DEseq2. A differential analysis using single-cell differential expression with significant cutoff criteria gave 528 differentially expressed (DE) genes which were used to identify subpopulations within the cells as they transitioned through states [12].

Their study revealed that there are three cell subpopulations lineages, "a" and "b" in NPCs and a third lineage "c" arises from "b" when the cells enter the differentiating state

then eventually the lineage "a" and "b" combine to form "ab". Hence at the end of day 30 there are two subpopulations "ab" and "c". The lineages were determined by hierarchical clustering analysis at each time point for 528 dynamic classifier genes determining by bootstrapping. They performed a Pearson correlation analysis for DE genes which were common between two subpopulations at neighboring time points, a gene ontology analysis using David software [33, 34] revealed association with neuron function. The dynamic changes of well-known gene markers were also analyzed within the subpopulations and finally a regulatory network analysis of subpopulation "c" revealed POU3F2, MIAT and PBX1 were critical for neurogenesis and REST is a repressor of neurogenesis [12].

This study has revealed key points which were not observed previously. Here a computational analysis is performed for four cell types as the NPCs differentiated into neurons. The trajectory path from NPCs to neurons revealed heterogeneity among immature neurons and neurons. The GO analysis of highest expressed genes present across all the cell types were associated with GO term for house-keeping genes, genes which were highly expressed in cells once they entered differentiating state are associated with GO term for central nervous system development. The genes with discrete gene expression patterns unique to this study were EFNA5 whose expression gradually decreased as cells reached neuron state, SFRP1and SFRP2 both these genes decreased their gene expression activity from day 5 to day 7 and turned off. The marker gene analysis of the cell types reveal association with GO terms for neurogenesis, olfactory lobe development, forebrain generation of neurons, axonogenesis, synapse formation, neuron part, neuron projection and other terms related to neuronal function. The pipeline and results have been discussed in detail in later sections. The important genes which are highlighted in this study are

summarized in **Table 1** below. The fate of NPCs and neurogenesis are controlled by dynamic processes related to time. The main goal of the project was to study what are the key genes involved in the transition of neural precursor cells to neurons in a time course profile.

Gene Symbol	Gene name	Functional Roles
WNT5A	Wnt Family Member 5 A	Regulates neural stem cell expansion, differentiation.
EFNA5	Ephrin 5A	Regulates proliferation of neural precursor cells.
UGT8	UDP Glycosyltransferase 8	Synthesizes galactosyl ceramide present in the myelin membrane of central nervous system (CNS) and peripheral nervous system (PNS).
SALL3	Spalt like Transcriptional Factor 3	Regulates neurogenesis and highly expressed in progenitor cells.
SFRP1	Secreted Family related proteins	Regulates growth, differentiation and axon guidance.
SFRP2		Regulates proliferation, migration and differentiation.
DCX	Doublecortin	Regulates neuron migration and dendritic growth
MAP2	Microtubule associated protein-2	Structural protein which maintains the cytoskeletal structures of neurons.
STMN2	Stathmin-2	Regulates microtubule destabilization and cytoskeletal network in neurons.

Table 1 Gene names and functional roles of some key genes highlighted in this study.

2. Materials and Methods

2.1 Dataset

The dataset was obtained from Gene Expression Omnibus (GEO), NCBI which is a public repository of functional genomics data. The dataset selected is GSE102066 which was published as a part of the paper "Single-cell gene expression analysis reveals regulators of distinct cell subpopulations among developing human neurons" by Wang, et al. at the Genome Institute of Singapore [12].

The data consists of 957 single cells at different time points over a 30-day period of non-differentiated human Neural Precursor Cells (NPCs) which were generated from human Embryonic Stem Cells (hESCs).
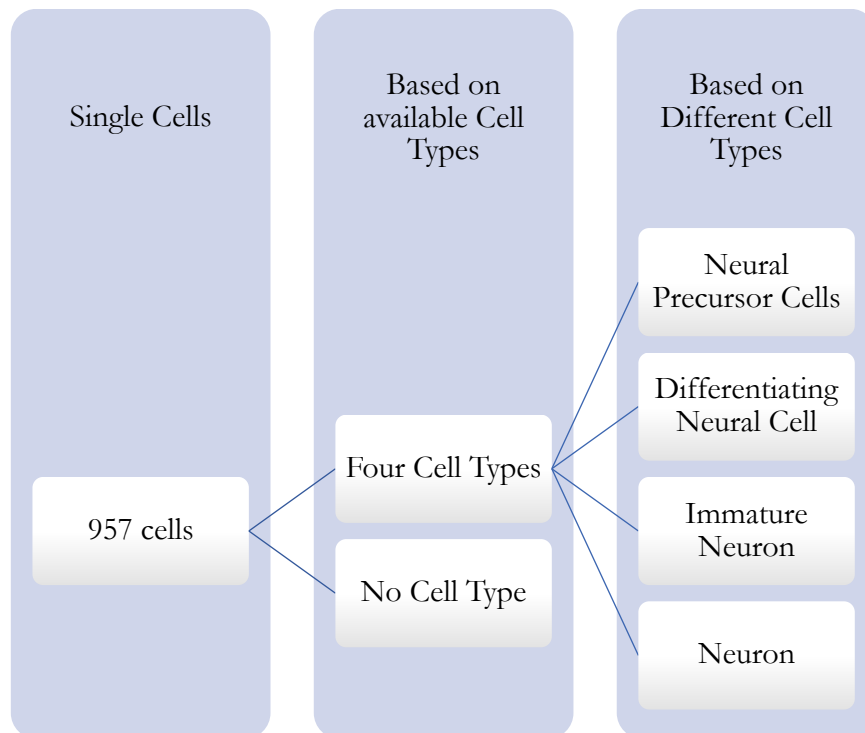
Figure 1 The four cell types present in the dataset.

Based on the presence of cell type information the cells were filtered out. A total of 573 single cells which consists of four cell types - Neural Precursor cell (NPC), Differentiating Neural Cell (DNC), Immature Neuron (IN) and Neuron (N) were obtained as the remaining cells were not classified into any cell types (**Figure 1**). The 573 single cells are spread across six-time points, the samples collected at each time points varied from 94-96 cells. The dataset has 96 NPCs at day 0, 94 DNCs at day 1, 288 IN cells at day 5, day 7, day 10 with 96 cells each and 95 N cells at day 30 (**Table 2**).

Cell Type	Number of cells in each type	Time points (days)
Neural Precursor Cell (NPC)	96	Day 0
Differentiating Neural Cell (DNC)	94	Day 1
Immature Neuron (IN)	288	Day 5, Day 7, Day 10
Neuron (N)	95	Day 30

Table 2 Number of cells in each cell type along with the time points at which they were collected.

2.2 Single cell RNA sequencing pipeline

In order to perform Single cell RNA Sequencing analysis to extract relevant biological information the following protocol was followed (**Figure 2**):

1. Unix Environment:

   The SRR files are fetched from GEO NCBI using the command "fastq-dump" for single ended data. The dumped files in ".fastq" format are aligned to the human genome using HISAT2.

2. <u>Python Environment</u>:

 The ".sam" format files generated from HISAT2 are inputted into HTSeq to obtain the counts of each gene for each cell. At this step, we generate ".csv" format files which each consists of all gene counts for that particular cell. In python environment, we further join all the .csv files in order to obtain a single file with gene names as the row names and "n" cells as the column names. Five joined ".csv" files are generated one containing the expression matrix for all the cell types which is for trajectory analysis and remaining four is one for each cell type which is used for generating highest expressed genes and marker genes.

3. <u>R environment</u>:

Various tools and packages are then used in R for preprocessing, filtering low quality cells and genes, data visualization and obtaining relevant gene information. The section 2.3 discusses the types of analysis performed.

Unix Environment
- Fetch SRR Files using fastq-dump
- Alignment to genome using HISAT2

Python Environment
- Counting of mapped reads using HTSeq
- Merging of count files of all cells

R Environment
- Data visualization and analysis
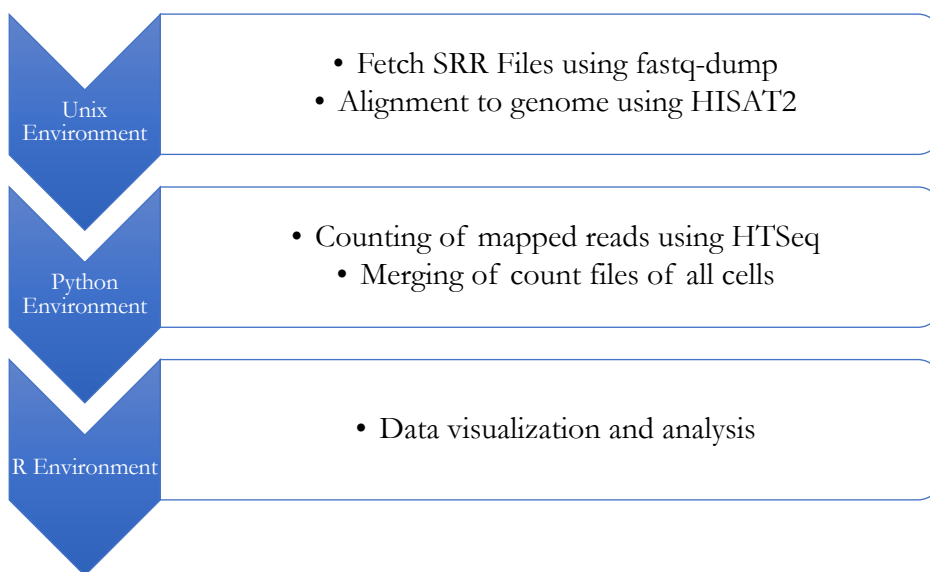
Figure 2 Flowchart of scRNA-seq pipeline.

2.3 Analysis

2.3.1 Trajectory Analysis

The Monocle [35-37] package was used in order to perform an unsupervised analysis and obtain the trajectory of all the 573 cells. The monocle package orders the cells as they progress through a biological process with time. The "CellDataSet" is the main class of monocle which will hold three input files, the expression data consisting of counts for all the single cells, the phenoData is an annotated data frame with cell names, cell types batch numbers etc., the featureData consists of gene names.

The expression matrix consists of genes expressed in at least 10 cells and the low expressed genes were filtered by using a minimum threshold of 0.1. Now, monocle constructs a trajectory by choosing genes which defines the cells progress during differentiation. This is done by determining the differentially expressed (DE) genes using "differentialGeneTest". These DE genes are then ordered and stored in a vector that will be used for ordering the cells later. The function "setOrderingFilter" marks the ordered DE genes from a total of 6241 DE genes. In **Figure 3** the black dots represent the DE genes used to order the cells.

The next step is to reduce the dimensionality of space in which the cells are present by using "reduceDimension" function. As a result, the cell data is transformed from a high-dimensional space to a low-dimensional space. The reduced graph embedded algorithm is used to perform dimensionality reduction, it assigns each cell a position in the reduced space, here the data is reduced to two- dimensions for easy visualization. Finally, the cells are ordered by using "orderCells" function, and a trajectory of the cells is obtained. The

trajectory path is tree-like, where the cells start at the roots (starting point) and moves along the trunk finally reaching the leaf (ending point).
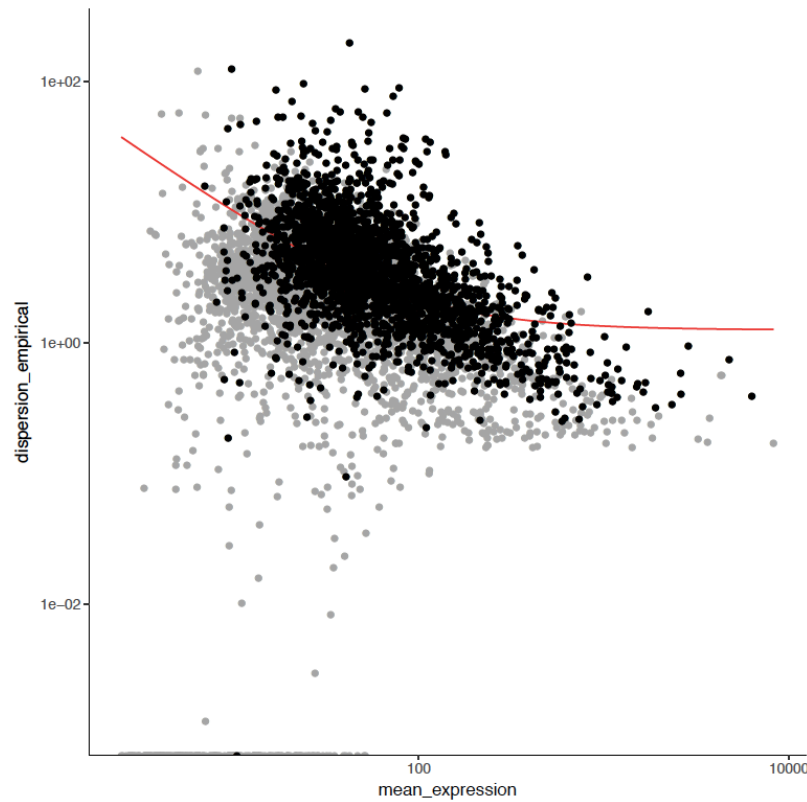
Figure 3 Plot of genes by mean expression value vs dispersion from 573 cells.

Each dot in the graph represents a differentially expressed (DE) gene. A total of 6241 DE genes are present out of which genes represented by **black** dots were selected to order the cells by monocle. Red curve is the mean-variance model learning curve. The X- axis represents the mean expression and Y-axis represents the dispersion of the genes.

2.3.2 Preprocessing of the dataset and obtaining highest expressed genes

The preprocessing for the dataset was performed using the Scater package [38]. The main class of Scater is "SingleCellExperiment", which inputs the counts expression matrix along with metadata file. This file has information about the cell names, batch number, days etc. After constructing the class, pre-processing of the data is performed. The dataset was preprocessed for low counts of genes and cells, mitochondrial genes, ERCC genes and

housekeeping genes. Criteria for filtering the low expressed genes was to exclude genes with counts less than 3 in a particular cell and expression of a particular gene across all cells less than 35-65 counts (different for each cell type). The outliers and spike-ins caused due to presence of mitochondrial and ERCC genes were filtered out using "isOutlier" which helps detect outliers within the data. The total number of expressed features varied from 4000-8000 counts after preprocessing. The first 50 highest expressed genes were thus obtained by using plotQC with "highest-expression" type and the gene lists obtained were further analyzed.

2.3.3 Marker Genes Analysis

Marker genes are computed using SC3 package [39]. It also uses "SingleCellExperiment" as the main class. Pre-processing of the data is not required as this was already performed for obtaining the highest expressed genes for each cell type. The command "sc3_plot_markers" assigns a p-value to each gene by using Wilcoxon signed rank test and the genes which are under the ROC curve (>0.85) and with a p-value < 0.01 are selected to obtain top 10 marker genes for each cell type. These genes are tested for significant overlap with gene ontology terms.

2.3.4 Gene Ontology (GO) Analysis

Gene Set Enrichment Analysis software (GSEA) was used to perform gene ontology analysis [40-42]. It is a computational tool which helps interpret statistically significant difference within gene sets by comparing it to the collection of annotated gene sets in Molecular Signature Database (MSigDB) [41, 42]. The MSigDB consists of 8 major

collections out of which the Gene Ontology gene sets, denoted as C5, was selected to test for significant overlaps with FDR q-value 0.05. The GO gene sets are divided into three sub-collections based on the GO ontologies: GO biological processes (BP), GO cellular component (CC) and GO molecular function (MF). All the gene lists obtained from top 50 highest expressed genes and marker genes for each cell type were tested for significant overlap with GO terms.

2.3.5 Statistical analysis

Key genes obtained from scRNA-seq analysis were tested for significant difference in mean expression between neighboring time points. All statistical analysis was performed using GraphPad Prism software [43]. To determine the choice of statistical test, it is necessary to know whether the gene expression has a Gaussian distribution or not. For this, a normality test was performed using D'Agostino-Pearson omnibus normality test which is recommended by GraphPad Prism software. The normality test was performed at alpha= 0.05 and a confidence interval of 95%. Genes did not pass the normality test thus confirming that the scRNA-seq data has a Non-Gaussian distribution.

Non-parametric Kruskal-Wallis test was the statistical method of choice used to determine statistically significant difference in mean expression of genes between neighboring time points from day 0 to day 30. The post-hoc analysis is performed by using Dunn's multiple comparisons test by comparing the mean rank of each time point with mean rank of every other time point. The graphs plotted are a single bar column for each time point with standard error from mean for each bar. The X-axis represents time in days and the Y- axis represents gene expression in counts.

3. Results

3.1 Trajectory Analysis reveals heterogeneity among neuronal cell types

Heterogeneity can be observed among different species, but it is also present in living cells belonging to the same cell type or the same multicellular organism. This behavior is a result of molecular interactions within each cell giving rise to heterogenous cell phenotypes [44]. Cell-to-cell variances are present in a population of cells and these differences have relevant biological importance. It is important to note that the behavior of a population cannot be a true representation of the behavior of individual cells [45]. Studies have revealed that the NSCs or the NPCs exist in a continuum when they get activated and differentiated. Their cell population studies of neural cell types have identified heterogeneity and rare intermediate states with discrete transcriptional profiles [46].

Single cell analysis considers asynchronous activity of individual cells by profiling the transcriptional activity of complex biological processes within the heterogeneous cell populations. Cells do not progress in perfect synchrony. Among cells which are captured at the same time some might be far along in their biological progress, and other might not have started that particular process yet. Taking this into consideration Monocle [35-37] orders the cells by their biological progress rather than by the time at which the cells were collected. Monocle constructs a trajectory by identifying the longest path measured in pseudo time value, which is the distance any a particular cell will have to travel in order to get back to the root.
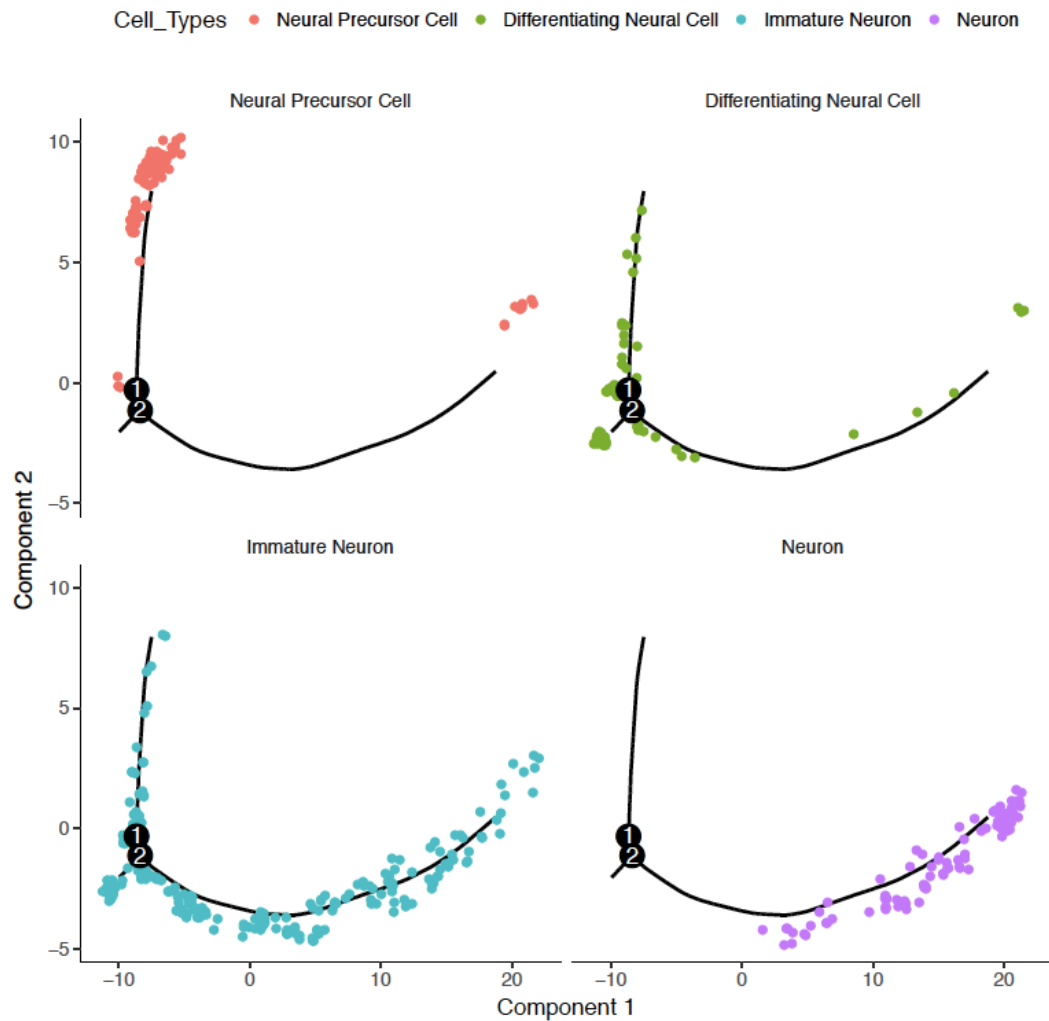
Figure 4 Trajectory path of NPCs differentiation into Neurons

A visual representation of the transcriptional profile of all cells progressing through different stages of differentiation i.e., NPCs to neurons. Each dot in the plot represent one cell and their positions assigned is based on differentially expressed genes used to order the cells. The trajectory path of 573 single cells is a tree-like structure starting at the root (top of branch 1) and ending in the leaf (end of branch 2). Branches 1 & 2 are the two branch points of the tree. The x-axis and the y-axis represents the two principal components of the reduced space.

The 573 single cells occupy an assigned position in the low-dimensional space. The position occupied by each cell depends on the DE expressed genes used to order the cells. The cells are distributed along a trajectory path which shows their transition from a starting state, which is NPCs to an ending state of neurons through differentiation. The trajectory

is a tree-like structure which starts at the roots and ends at the leaves progressing along the trunk. The cells start off at the top of branch 1 (roots) on the left side. As the time progresses the cells travel along the trajectory path and reach the end of branch 2 (leaf) (**Figure 4**).

The cells do not have any alternative cell fate and with time they will differentiate to form mature neuron at the end point of branch 2. The NPCs (red) are present predominantly at the root (day 0). The DNCs (green) occupy their positions at the end of branch 1 and starting of branch 2 as cells progress on the trajectory path (day 1). It is also seen that some NPCs and DNCs are present at the end of branch 2 (leaf) which is a possible indication for the presence of outliers within the scRNA-seq data which must be filtered out during preprocessing for accurate results (**Figure 4**).

The cells of immature neurons (blue) occupy positions all along the trajectory indicating heterogeneity in the cell population (**Figure 4**). Since the immature neurons are present at three-time points day 5, day 7 and day 10, the cells at each time point are evaluated by their biological progress in terms of pseudo time (**Figure 5**). Here, pseudo time measures the progress of individual cell through differentiation. While immature neurons have progressed and are present on branch 2 many of these cells at day 5 and day 7 are present at the end of branch 1 and starting point of branch 2 where the DNCs were present at day 1. At day 10 all the immature cells are present at branch 2 indicating that the cells have reached immature stage. The results clearly indicate heterogeneous behavior of the cells at earlier time points.
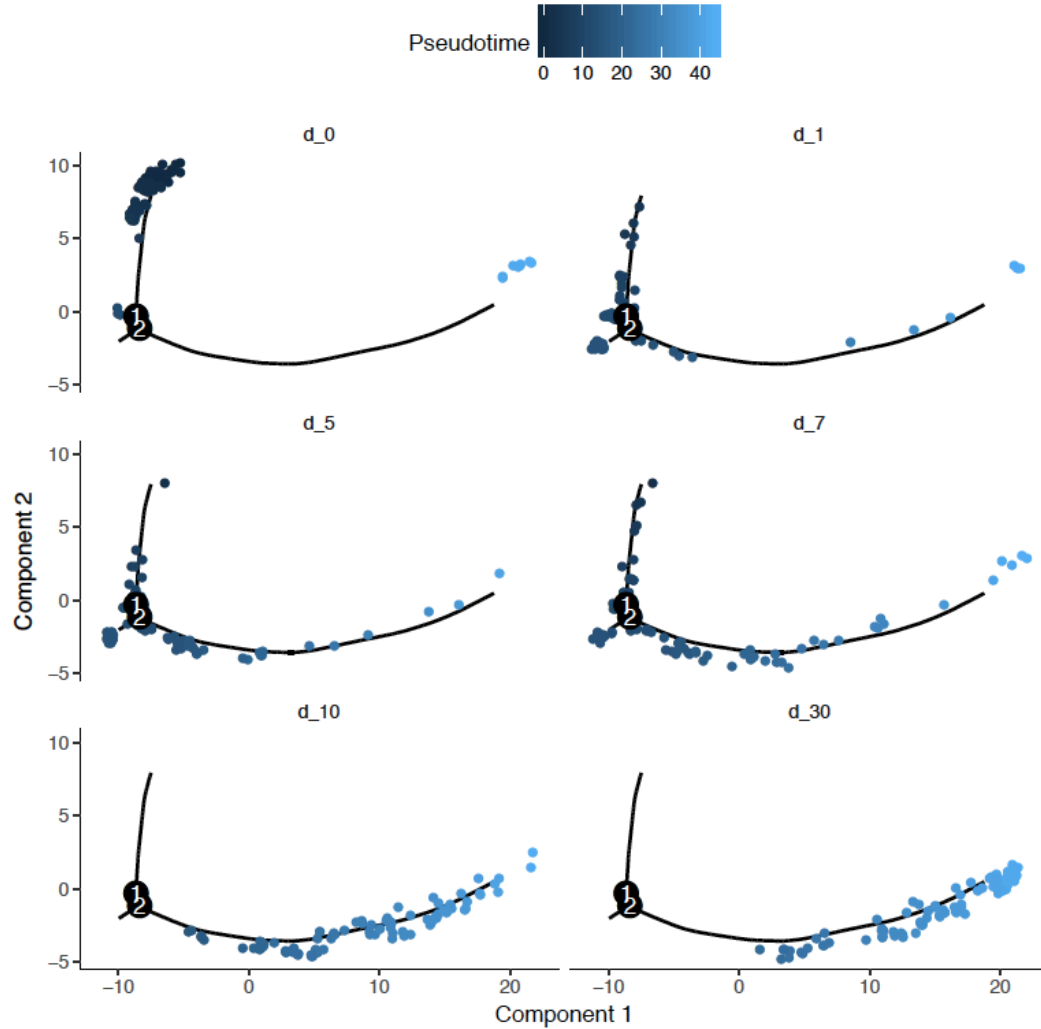
Figure 5 Trajectory path of NPCs to neurons in Pseudo time

Pseudo time shows the progress of the cells during a biological process. Each point represents a cell and the color gradient shows the biological progress of the cells at each time point. The cells travel along the trajectory path to form neurons (day 30). The results indicate heterogeneity among immature neurons and neurons. The X-axis and Y-axis represent the principal components of the reduced space.

The neurons (purple) also exhibit some degree of heterogeneity based on the positions occupied by cells on the trajectory, this is because all the cells have not progressed in synchrony (**Figure 4**). The pseudo time analysis of neurons also shows that even though at day 30 a large number of cells have reached the end point (light blue) on branch 2 many

cells are still in the transitioning state to form neurons. The scRNA-seq analysis by using monocle shows that the neuronal types have heterogeneity in their cell populations. While heterogeneity is observed mostly in immature neurons and to some extent in neurons, this behavior is not observed widely among NPCs and DNCs in pseudo time (**Figure 5**). This indicates that all the cells do not differentiate at the same time. These variances at the molecular level are concealed when a population-based identification of cell types are employed [46]. These results clearly indicate that neurogenesis is a continuous ongoing process and the classification of cells into only four cell types overlooks the cell subtypes and transitioning states arising during differentiation.

3.2 Top 50 Highest Expressed Genes reveals common and unique genes among the four cell types

Top 50 highest expressed genes of each cell type are generated and the list of all these genes is present in the appendix in **Table a**. Firstly, the initial results were filtered for the presence of housekeeping genes taken from Human housekeeping genes, revisited published in Trends in Genetics [47]. A total of 20 housekeeping genes were filtered out manually from all cell types (refer **Table b** in Appendix). Top 50 highest expressed genes thus obtained were analyzed for the presence of common and unique genes between cell types. These gene lists are checked for association with relevant Gene Ontology (GO) terms.

A Venn diagram is generated from the gene lists using a user-friendly web tool which can detect intersection of genes within the four cell types, giving a graphical output

for easy visualization along with their gene lists. The tool used to generate these results was developed by Bioinformatics Evolutionary Genomics [48]. There were an overall of 79 unique genes from a list consisting of 200 genes corresponding to the four cell types. The Venn diagram and table with gene list of common and unique genes are represented by **Figure 6** below and **Table c** in Appendix respectively. These specific set of genes were investigated for statistically significant overlaps with gene ontology terms. The genes were not investigated if there were no significant overlap with any gene ontology term.
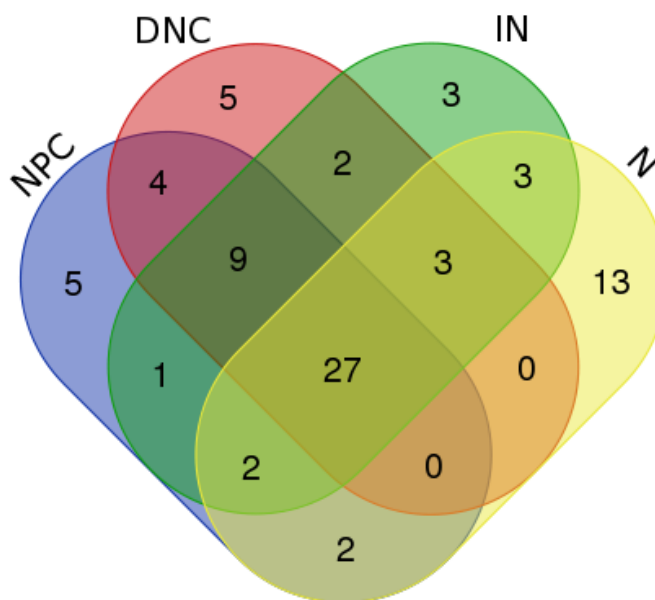
Figure 6 Venn diagram with intersection of common genes for all cell types

The 27 genes which were common to all cell types were tested with curated pathway database and gene ontology terms. Out of the 27 genes, 15 genes were associated with GO term for housekeeping genes which are expressed by 19 normal tissue [49] within the body responsible to maintain basic cellular function. This is an expected result as they

are expressed in all the cells under normal physiological conditions [47]. Apart from this

gene were associated with roles in translation and protein metabolism (**Figure 7(a)**).

Figure 7 GO results for highest expressed genes common among all cell types and three cell types- DNC, IN and N.

(a) represents the GO results for 27 common genes present in all cell types and **(b)** represents the GO results for genes calm2, sox4 and stmn1 common in DNC, IN and N. The relevant overlap and genes are highlighted by red parenthesis.

The genes CALM2, SOX4 and STMN1 were common to three cell types- DNC, IN and N. They were associated with GO term for central nervous system development (**figure 7(b)**). Previous studies have indicated that these three genes have important biological roles in the development of neurons and mamalian brain. Calmodulin (CALM2) is found in abundance within the human brain and is principally responsible for regulating calcium signals, it also participates in proliferation and cell cycle progression [50], SRY-box 4 (SOX4) acts as transcriptional factors which activate the neuronal gene promoter along with other roles in maintaining the central nervous system [51, 52]. In vivo analysis of Stathmin-1 (STMN1) during adult hippocampus neurogenesis revealed upregulation with maturity of cells [24], it is a microtubule destabilizer maintaining cell cycle progression motility and survival [53].

The genes WNT5A, EFNA5, ADNP out of the 5 genes unique to NPCs are associated with GO terms for positive regulation of neuron development which is a process resulting in neuron development over time. Other GO terms were regulation of synapse assembly, synapse organization and axonogenesis, positive regulation of neuron projection development and regulation of synapse structure (**Figure 8(a)**).

WNT family member-5A (WNT5A) is a non-canonical wnt-molecule which is involved in neural stem cell expansion, differentiation and are highly expressed in proliferative cells [54]. The Activity Dependent Neuroprotective Protein (ADNP) gene regulates the expression of multiple genes which are responsible for organogenesis, neurogenesis, lipid transport etc. The deletion of ADNP is shown to have effects on neurogenesis and embryogenesis which was based on ADNP knockout mouse models [55].

(a)   **(b)**
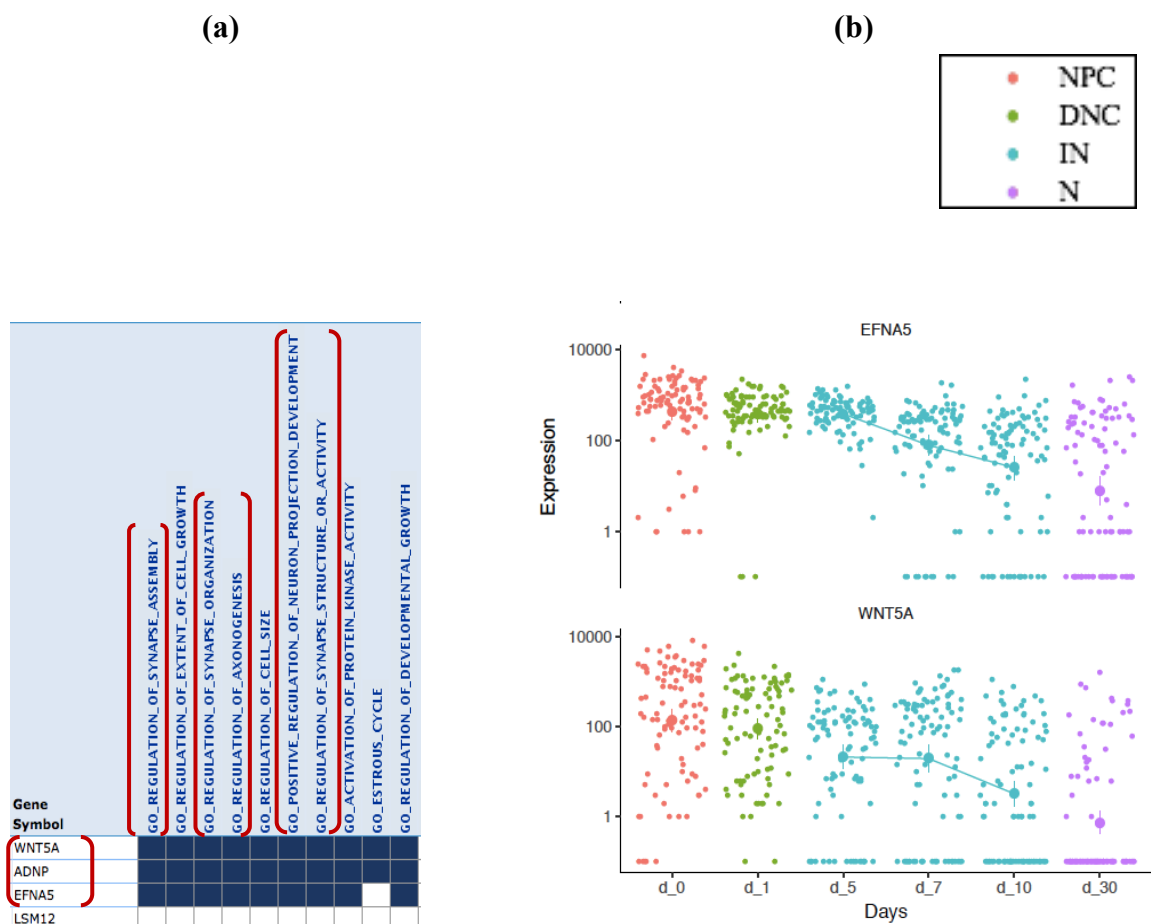
Figure 8 GO results for highly expressed genes unique to NPCs and gene expression of efna5 and wnt5a.

(a) represents the GO results for genes wnt5a, efna5 and adnp which are highly expressed in NPCs alone. The relevant overlap and genes are highlighted by red parenthesis. **(b)** represents gene expression trend for genes efna5 and wnt5a. Both genes show a decreasing trend from day 0 to day 30.

Ephrin A5 (EFNA5) plays role in embryonic and adult neurogenesis signaling but the number of publications which can clearly identify the role of ephrin family in self-renewal, proliferative and non -proliferative cells is still unclear [56]. These genes play critical roles in the development of the nervous system. Examining the expression of these genes showed that there is a decrease in the levels of WNT5A and EFNA5 as the cells

mature while they are highly expressed at proliferative stages (**Figure 8(b)**), their statistical significance is studied in section 3.4

3.3 Gene Ontology analysis of marker genes reveals discrete events among cell types

The marker genes obtained for each cell type using SC3 package [39] are checked for association with significant GO terms and the results are analyzed. Refer **Table d.** of Appendix for list of marker genes of all cell types.

3.3.1 Neural Precursor cells

The marker genes of NPCs are associated with GO terms for forebrain neuron development, neurogenesis and forebrain generation of neurons. These GO terms imply generation of cells within the forebrain, ultimately resulting in its progression into a fully developed differentiated cell.

The genes WNT5A, SLIT2 and SALL3 are associated with GO terms for olfactory bulb lobe development (**Figure 9(a)**). This shows that the cells have the potential to differentiate into sensory neurons within the olfactory bulb which is located in the forebrain. There are six genes- WNT5A, SLIT2, SALL3, MEIS1, B2M and UGT8 associated with neurogenesis (**Figure 9(a)**). Expression of these genes is responsible for generation of cells within the nervous system through cell differentiation. In the absence of genes SALL3 and UGT8, remaining four genes are associated with GO terms for negative regulation of neuron differentiation and negative regulation of nervous system development (**Figure 9(a)**).
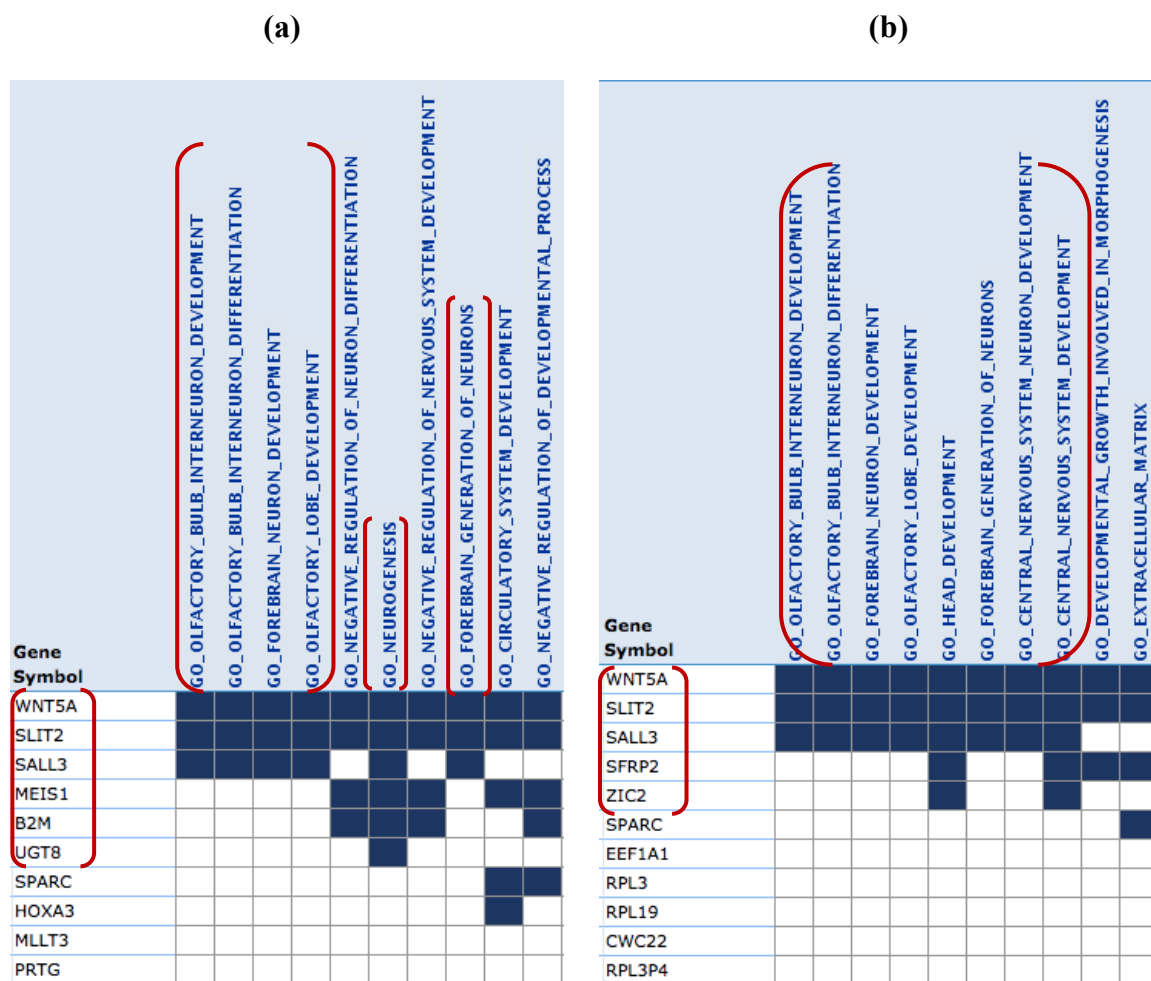
Figure 9 GO results for marker genes of NPCs and DNCs.

(a) represents the GO terms for marker genes of NPCs and (b) represents the GO results for marker genes of DNCs. The relevant overlap and genes are highlighted by red parenthesis.

UDP Glycosyltransferase 8 (UGT8) is involved in regulating enzyme activity in case of nerve injuries, it synthesizes glycosphingolipids in the myelin sheath [57]. UGT8 is mainly responsible for the production of galactosyl ceramide (GalCer), a major glycosphingolipid present in the myelin membrane of central and peripheral nervous systems [58]. GalCer importance was studied in UGT8 knockout mice, where a breakdown of axon insulation was observed [59]. Splat like transcription factor 3 (SALL3) is involved

in development of the olfactory bulb in dopaminergic neurons which was observed by examining sall3-null mice and mutations in vertebrate homologues of splat genes are linked with several human disorders [60]. Other studies have found that SALL3 and its related family are expressed in cortical neural progenitor cells, interacting with each other to regulate neurogenesis [61]. The functions of these two genes indicate significant roles in the development of neurons and nervous system. Future research focusing on their regulatory pathway and interaction with other proteins could help understand unexpected GO results obtained here.

3.3.2 Differentiating Neural Cell

There are four marker genes which are common in both NPCs and DNCs- WNT5A, SLIT2, SALL3, and SPARC hence we observe association with similar GO terms. The marker genes of DNCs were associated with GO terms for forebrain neuron development, forebrain generation of neurons and olfactory bulb development (**Figure 9(b)**).

The marker genes of NPCs were associated with GO terms involved in negative regulation of nervous system development these terms were not observed as cells reached day 1 which is the differentiating state. Instead, they were associated with central nervous system development involving five genes WNT5A, SLIT2, SALL3, SFRP2, and ZIC2.

3.3.3 Immature Neurons

The marker genes of immature neurons are associated with GO terms for neuron projection, neuron part and neurogenesis which are involved in the generation of cells within the nervous system (**Figure 10**).

There are seven genes- DCX, MAP1B, KIF5A, MAP2, STMN2, SFRP2 and SFRP1 associated with GO term for neurogenesis (**Figure 10**). As mentioned earlier neurogenesis is a cell differentiation process which will help in the formation of neurons within the nervous system. Neuron projection represents a process or a projection from a nerve cell, which can either be an axon or a dendrite [62], and the neuron part represents any constituent part of a neuron, a part of the nervous system, showing the cells have begun to differentiate into neurons [63]. Neuron projection and neuron part overlapped with the same set of seven genes out of with five genes were common with neurogenesis.
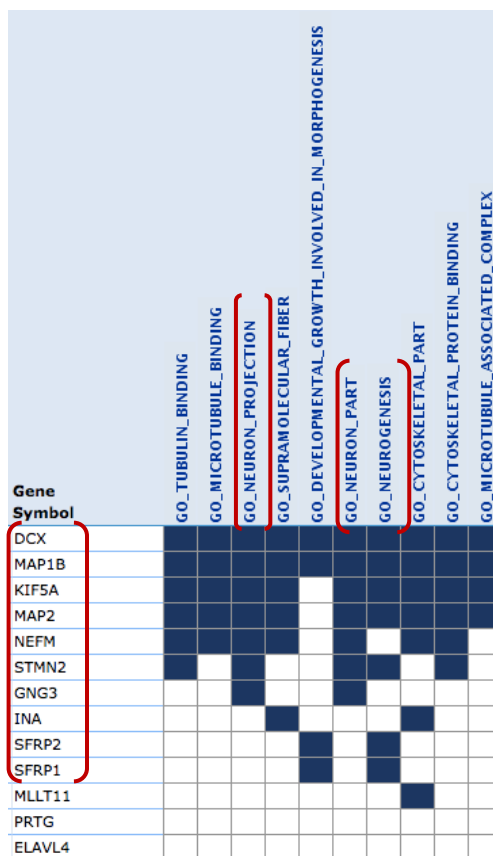

**Figure 10** GO results for marker genes of INs.

The figure to the left represents the GO terms for marker genes of INs. The relevant overlap and genes are highlighted by red parenthesis.

The immature neurons were collected at three-time points- day 5, day 7 and day 10, the trajectory of the cells suggest that the immature state is heterogeneous in nature and all the cells do not exhibit the same transcriptional activity at all three points, this behavior is also reflected from the GO terms observed.

The association of genes with neuron projection and neuron part was observed only in the absence of genes SFRP1 and SFRP2, which were among the seven genes which were associated with neurogenesis. As a result, genes SFRP1 and SFRP2 were further investigated and it is observed that the gene expression levels of SFRP1 and SFRP2 decreases sharply from day 5 to day 7 after which the cells have low expression as they reach day 30 (**Figure 11(a)**).

The members of SFRP (Secreted Frizzled Related Proteins) family are antagonists for WNT-signaling pathway but certain members of SFRP family positively regulate WNT pathway through WNT diffusion [64]. In humans, among the five members of SFRP family - SFRP1, SFRP2 and SFRP 5 belong to one subfamily. SFRP1 is expressed in anterior neural plate and in early developmental stages [65], and SFRP2 is involved in various functions of proliferation, migration and differentiation in vertebrates [66]. A study in 1998 showed that SFRP2 is turned off as cells attain cell differentiation indicating that these genes are only involved in early developmental processes in mice during nephrogenesis [67], but here we observe similar expression during in vitro development of neurons derived from hESCs.
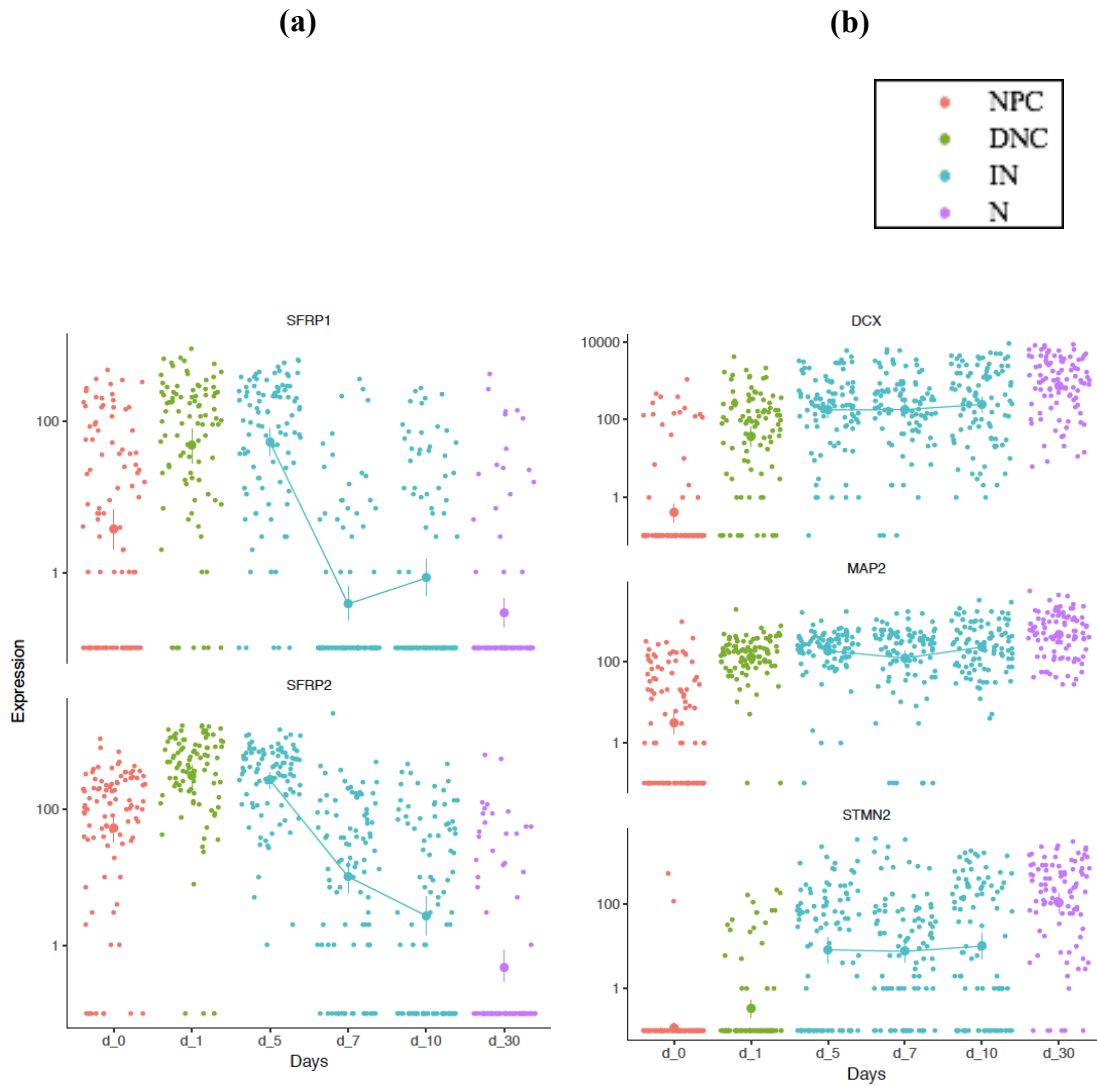
Figure 11 Gene expression pattern for certain marker genes of INs.

(a) represents the gene expression trend for genes sfrp1 and sfrp2 which drastically decreases after day 5 and **(b)** represents the gene expression for genes dcx, map2 and stmn2 which increases as the cells reach day 30.

Doublecortin (DCX) is a known immature neuron marker which is highly expressed by migrating neurons of developing and adult nervous system [68]. It is a neuronal microtubule binding protein and studies show defects in its expression affects neuronal migration, axon and dendritic growth [69]. This gene shows increase in expression as the cells change from one state to another. At day 0 maximum cells have zero expression of DCX. It is expressed at day 1 and is upregulated as cells reach day 30. There are no cells which have zero expression of DCX from day 10 to day 30 (**Figure 11(b)**), implying to its critical role in developing neurons.

Microtubule associated protein (MAP2) is found in neurons, they are crucial for neuromorphogenic processes and reorganization of microtubule and F- actin [70]. They are predominantly expressed in the cytoskeletal structures of neurons and act as a substrate for protein kinases and phosphatases present in neurons. Its other functions include organelle transport in axons and dendrites [71]. As expected the expression of MAP2 gradually increases as the cells form mature neuron. Stathmin-2 (STMN2) is a neuronal protein which regulates the cytoskeletal and microtubule network [72]. Previous study revealed that STMN2 is expressed in immature olfactory neurons and has supporting role during neurogenesis [73]. The expression of STMN2 is low at days 0 and day 1, it is upregulated in cells at day 30 (**Figure 11(b)**). The statistical analysis is performed to determine whether there is significance difference in expression of these genes between neighboring time points in section 3.4.

3.3.4 <u>Neurons</u>

The five marker genes- SOX4, SOX11, SALL1, POU2F2 and DCX of neurons are associated with GO term for cell development. It is a biological process representing progression of any cell through time to form a mature state. This process does not involve any steps of committing cell to a specific fate [74], which shows that the cell fate was attained at previous stages (**Figure 12(a)**).

Neurons have the ability to mature into different types- Glutamatergic, GABAergic, Dopaminergic, Serotonergic and Cholinergic neurons. Well-known marker genes for these types of neurons were tested, and their expression levels were observed. The only markers which had basal level expression till day 30 were Gamma-Aminobutyric Acid Type B Receptor Subunit 1 (GABBR1), marker for GABAergic neurons, Glutaminase (GLS) and Glutamine Synthase (GLUL), markers for glutamatergic neurons.

Glutamate is an excitatory and GABA is an inhibitory neurotransmitter in the cortex and controls cortical excitability [75]. Studies have shown that within the mature brain glutamate and GABA regulate activation of ionotropic and metabotropic receptors. Apart from these roles they also influence proliferation, migration, differentiation and survival of neural cells during early development [76]. The is no visible change in expression levels of marker genes for glutamatergic and GABAergic neurons as the cells transition from one state to another state (**Figure 12(b)**). To critically analyze their expression levels the scRNA-seq data after day 30 will be required.

(a)

(b)

Figure 12 GO results for marker genes of Ns and Gene expression for marker gene of neuron types

(a) Represents the GO terms for marker genes of Ns, the relevant overlap and genes are highlighted by red parenthesis and **(b)** gene expression of marker genes for Glutamatergic and GABAergic neurons (gls, glul and gabbr1).

3.4 Statistical Analysis

The analysis was performed for key genes to observe whether there is any statistically significant difference in mean expression as cells transitioned from one cell type to another with time. Statistical analysis is done for neighboring time points as the cell types changed from NPC→N (Day 0 vs Day1, Day 1 vs Day 5, Day 5 vs Day 7, Day 7 vs Day 10, Day 10 vs Day 30). Kruskal-Wallis non-parametric analysis and Dunn's Multiple Comparisons post hoc test was performed for downregulated genes SFRP1 and SFRP2 (sharp), WNT5A and EFNA5 (gradual) and upregulated genes DCX, MAP2 and STMN2.

- Genes whose expression decreases sharply -SFRP1 and SFRP2
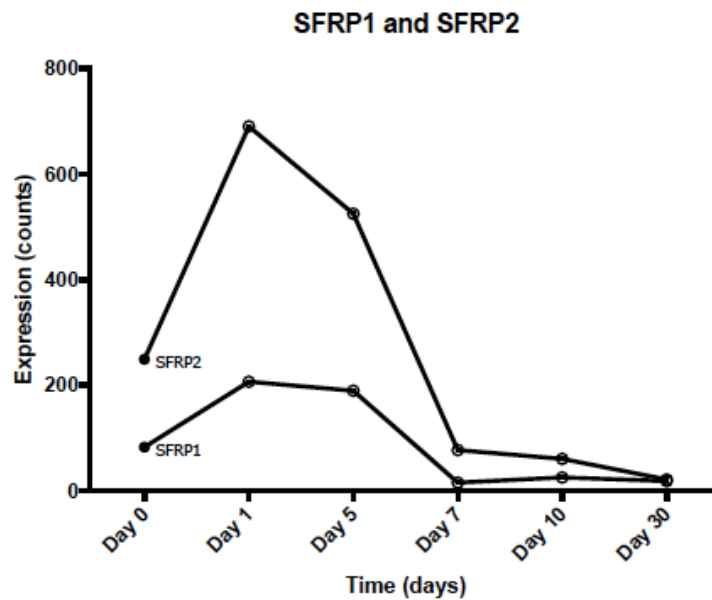
Figure 13 Plot showing mean expression of sfrp1 and sfrp2 with time in days on x-axis and expression in counts on y-axis

- Genes whose expression decrease gradually - WNT5A and EFNA5
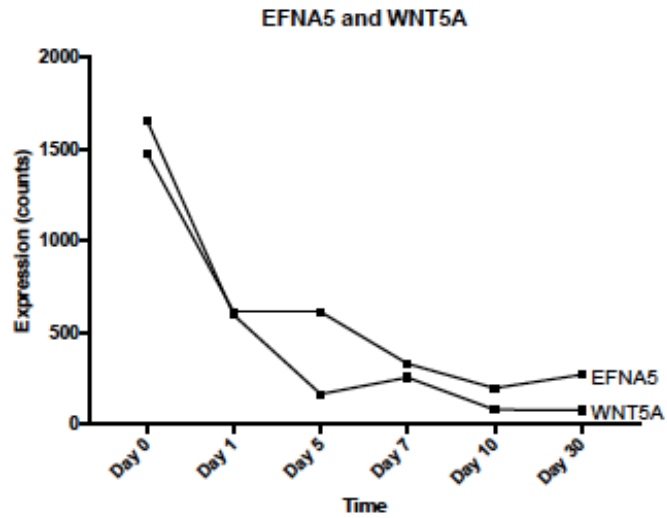
Figure 14 Plot showing mean expression of efna5 and wnt5a with time in days on x-axis and expression in counts on y-axis

- Genes whose expression increase gradually- DCX, MAP2 and STMN2.
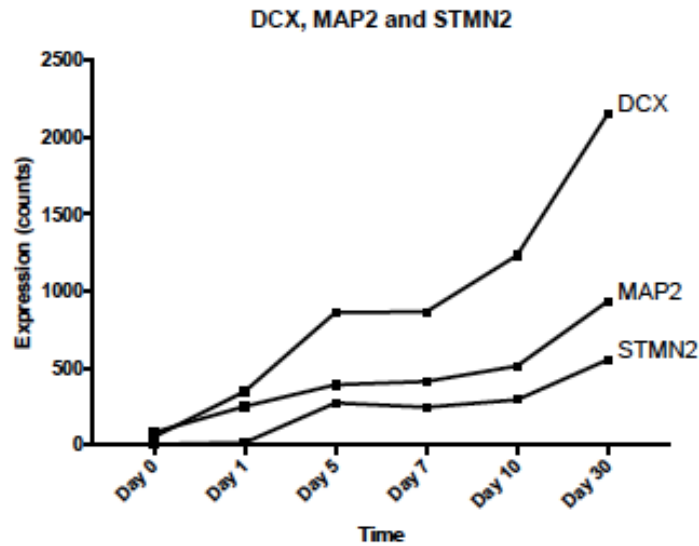
Figure 15 Plot showing mean expression of dcx, map2 and stmn2 with time in days on x-axis and expression in counts on y-axis

3.4.1 <u>SFRP1 and SFRP2</u>

The Kruskal-Wallis non-parametric test for SFRP1 and SFRP2 shows there is significant difference in both the genes at the same neighboring time points. The Dunn's post hoc test showed significant difference from day 0 to day 1 and day 5 to day 7. From day 0 to day 1 SFRP1 has a p-value <0.0001 and SFRP2 has a p-value 0.001. Both the genes showed a significant drop in gene expression from day 5 to day 7 with p-value <0.0001. The cells show very low expression levels after day 7 (**Figure 16**).
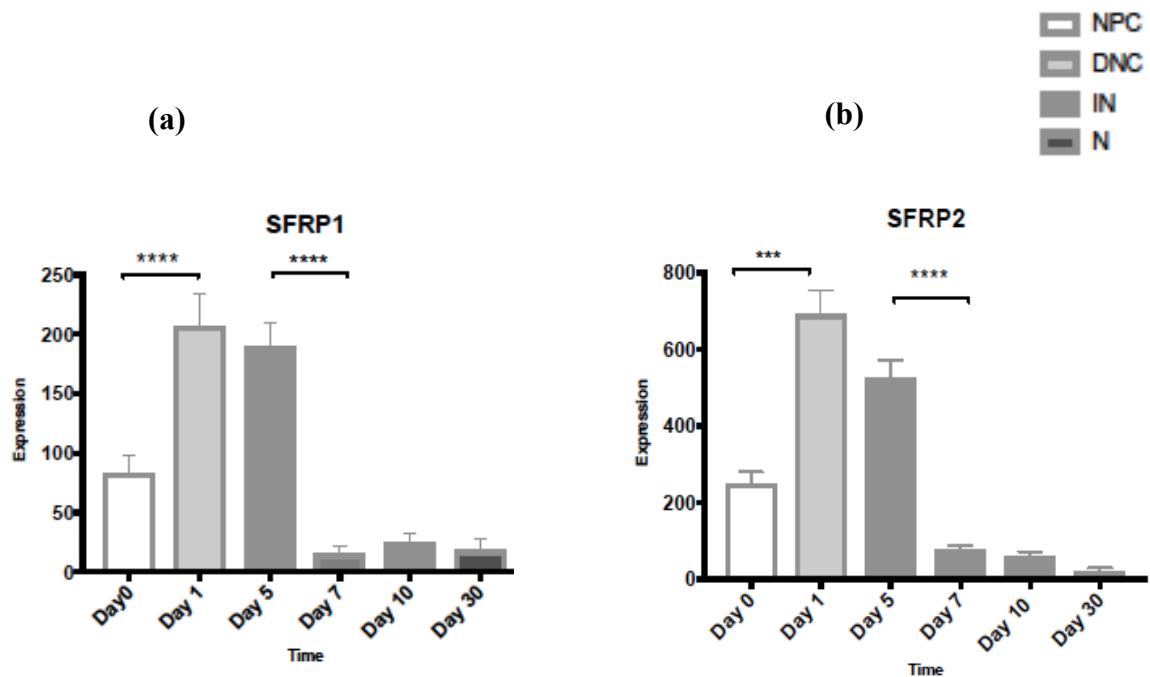
Figure 16 Plot showing statistical analysis of sfrp1 and sfrp2.

The statistical difference in gene expression of (**a**) sfrp1 and (**b**) sfrp2 as cells transition from NPC→ N is obtained by Kruskal-Wallis non-parametric test followed by Dunn's test for multiple comparisons. * p≤0.05; ** p≤0.01; *** p≤0.001; **** p≤0.0001.

3.4.2 <u>WNT5A and EFNA5</u>

The gene expression levels of WNT5A and EFNA5 decrease as cells change state from NPC to mature neurons. The Dunn's post hoc test reveals that there is a significant decrease in expression of WNT5A from day 7 to day 10 with a p-value of 0.039 and in EFNA5 from day 5 to day 7 with a p-value of 0.0001. There is no significant difference between any other time points for both the genes (**Figure 17**).
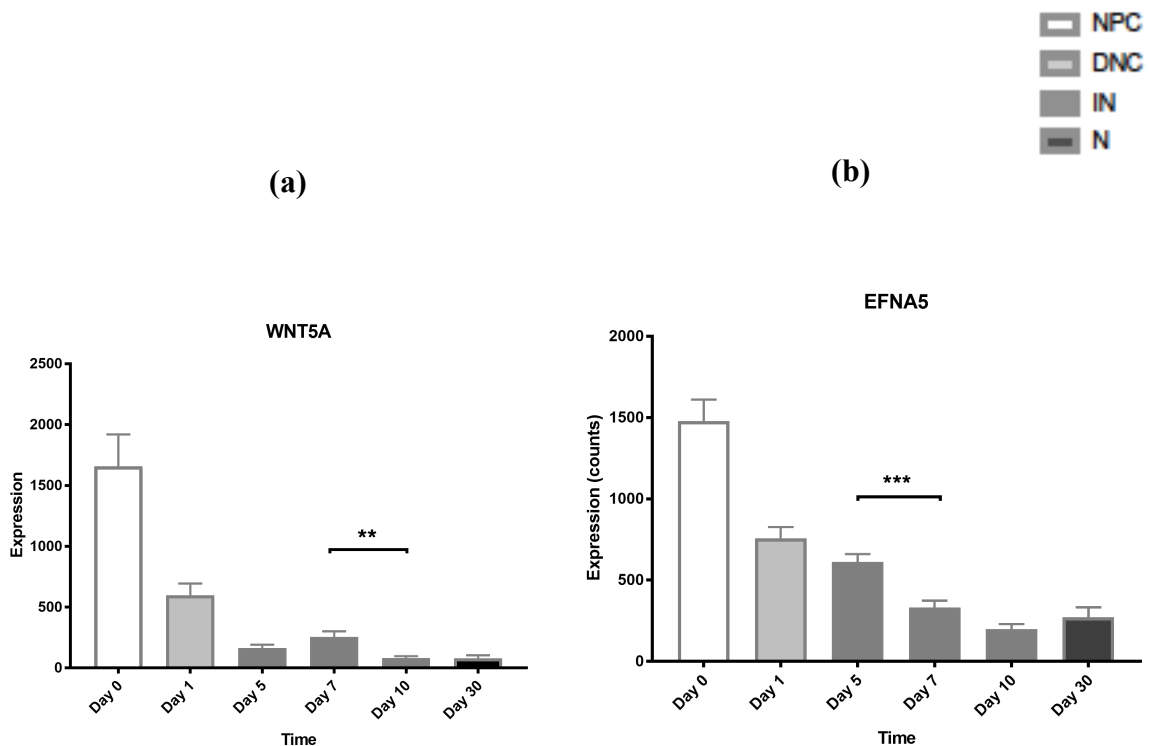
Figure 17 Plot showing statistical analysis of wnt5a and efna5.

The statistical difference in gene expression of (**a**) wnt5a and (**b**) efna5 as cells transition from NPC→ N is obtained by Kruskal-Wallis non-parametric test followed by Dunn's test for multiple comparisons. * p≤0.05; ** p≤0.01; *** p≤0.001; **** p≤0.0001.

3.4.3 <u>DCX, MAP2 and STMN2</u>

Among the upregulated genes, expression of DCX is highly significant from day 0 to day 1 with p-value of <0.001 and from day 1 to day 5 with p-value of 0.0047 which is relatively lower, there is no significant difference between other time points. MAP2 has highly significant difference in gene expression from day 0 to day 1 with p-value <0.001 and small significant difference from day 10 to day 30 with p-value 0.0258. STMN2 has significant difference as cells transition from day 1 to day 5 with p-value <0.001 and from day 10 to day 30 with p-value 0.001 (**Figure 18**).

(c)
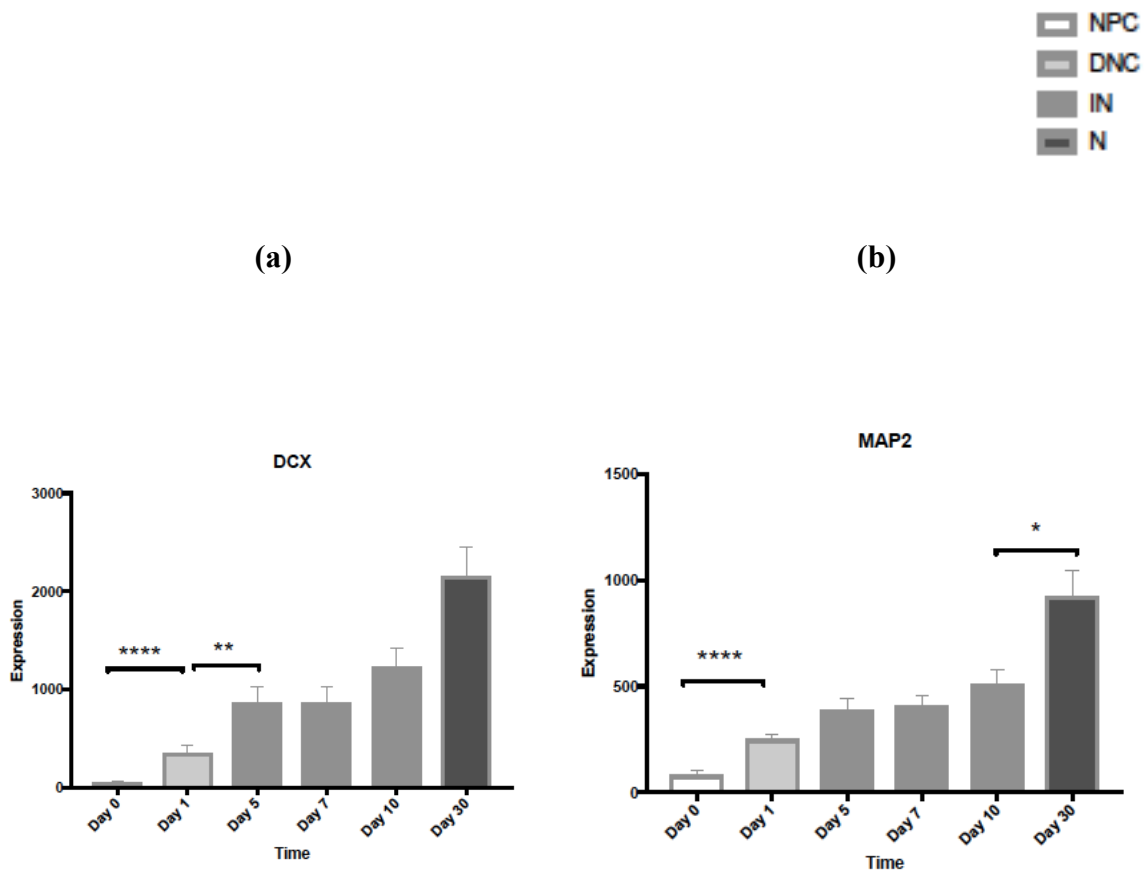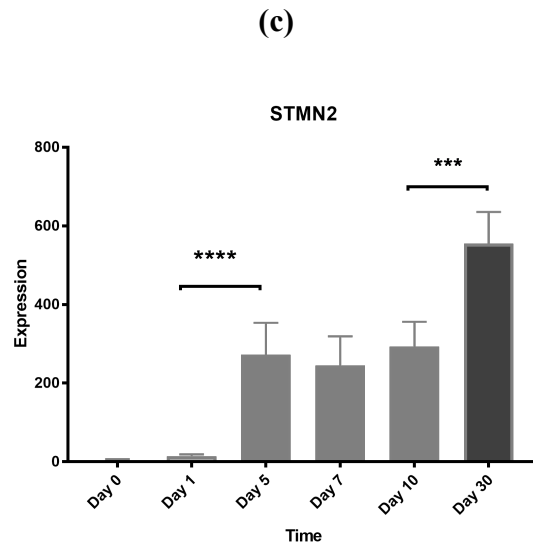
Figure 18 Plot showing statistical analysis of dcx, map2 and stmn2.

The statistical difference in gene expression of (**a**) dcx, (**b**) map2 and (**c**) stmn2 as cells transition for NPC$\rightarrow$ N is obtained by Kruskal-Wallis non-parametric test followed by Dunn's test for multiple comparisons. * p$\leq$0.05; ** p$\leq$0.01; *** p$\leq$0.001; **** p$\leq$0.0001.

4. Discussion

The single cell analysis of neuronal subtypes has helped elucidate the mechanisms underlying neurogenesis. These mechanisms at molecular and cellular level govern heterogeneity within the cell types which are concealed by population-based cell identification approaches. Trajectory analysis revealed that the cells have a single fate of forming neurons but the rate at which all the cells will attain this state is not the same. Neurogenesis is a continuous dynamic process and previous studies have shown presence of rare cell types within a cell population. As expected trajectory analysis indicates heterogeneity because cells do not progress in synchrony during differentiation. Trajectory analysis of immature neurons in pseudo time revealed that some cells at day 5 and day 7 showed biological activity similar to DNCs. None of the cells at day 10 exhibited this behavior clearly indicating presence of a transitioning population at earlier time points. Neurons are predominantly present at the end point of the trajectory path, but some cells are still behind in terms of their biological progression which has resulted in a heterogenous population. Heterogeneity has suggested presence of transitioning cell subpopulations. Critical insights during neurogenesis were concealed due to classification of neuronal cells into only four cell types based on time stamp at which they were collected. Previously, Wang, et al. performed hierarchical clustering for these cell types to detect presence of subpopulations within them. It is based on p-values assigned to genes or cells used to determine the clusters. They revealed presence of two subpopulations "a" and "b" in NPCs and a third subpopulation "c" in DNCs. Later the subpopulations "a" and "b" merged and two distinct populations "ab" and "c" were present at day 30. While their analysis aims at

detecting presence of subpopulations at different time points, it fails to track individual cell progress, variability and their expression profile at a particular time point by examining the pattern of gene expression. One of its disadvantages is that clustering is sensitive to noise and outliers within the dataset while monocle find noise or variability as useful biological information and effectively identifies presence of outliers as seen in NPCs and DNCs.

Key regulators obtained from highest expressed genes and marker genes were evaluated for unique gene expression dynamics. Our observations revealed genes which were down regulated and upregulated as cells formed differentiated neurons. The members of the SFRP family act as WNT signaling pathway inhibitors [77], in addition to this they have specific biological function. SFRP2 plays important biological proliferation, migration and differentiation [66], and SFRP1 is expressed in embryonic brain [67], and helps in axon guidance regulation [77]. But studies have indicated that SFRPs also have WNT- independent function [78], where SFRP1 interacts with SFRP2 to promote axon guidance [77]. Another study revealed that SFRP1 and SFRP2 positively regulated the WNT- signaling pathway necessary for the development of mouse optic cup [64]. Previous studies have also revealed that the gene SFRP2 turned off when the cells attained terminal differentiation in mice during nephrogenesis [67]. SFRP1-/- and SFRP2 -/- mice indicated increased proliferation and generation of early born neurons in retinal neurogenesis study [78]. Low-medium concentrations of SFRP1 and SFRP2 has facilitated in in-vivo differentiation of dopaminergic neurons [79]. In this study, genes SFRP1 and SFRP2 showed decreased level of gene expression over a period of two days (day 5 to day 7) and very low expression at days 7, 10 and 30. The heterogeneous nature of immature neurons and this gene activity suggests that the cells have attained differentiation after the decrease

in expression level of SFRP 1 &2. But given the wide range of biological functions of SFRPs with other factors their functional roles during in vitro and in vivo neurogenesis must be evaluated in the future. The gene expression of these two genes were not studied by Wang, et al., instead they showed downregulation of neuronal differentiation repressor REST as cells reached neuron stage.

Genes which downregulated gradually over the time course profile were WNT5A and EFNA5. WNT5A has a non-canonical mode of action by activating wnt/planar cell polarity pathway [80]. It regulates the proliferation of neural precursor cells, the downregulation of WNT5A is due to decrease in the number of proliferating neural progenitor cells [81]. Wang, et al. also tested for WNTs among which only WNT5A was found to be highly expressed among precursor cells and then was downregulated. In our analysis, it is also observed that only WNT5A is expressed among all the other WNTs which was gradually downregulated indicating crucial role of WNT5A during proliferation of NPCs. Eph receptors are responsible for cell-to-cell signaling and are involved in neurogenesis, axonogenesis, neural migration, synapse formation and axon guidance by ephrin-eph signaling [56, 82]. EFNA5 (EphrinA5) is a neurogenesis factor [82], studies in mice have revealed lack of EFNA5 can severely reduce neuron survival and reduced cell proliferation [83]. The main function of EFNA5 is to modulate proliferation of neural progenitors cells [56]. Here we observed that EFNA5 is highly expressive at day 0 and is gradually downregulated as the cells reach neuron state. The downregulation is probably due to the decrease in the number of proliferative cells and this gene was not studied by Wang, et al.

The genes DCX, MAP2 and STMN2 are well established neuronal markers. They were tested by Wang, et al. who observed upregulation of all the three genes. In contrast to their

approach of testing well-known neuronal markers our method of determining these genes was by performing marker gene analysis using SC3 [39]. DCX is a microtubule associated protein which is expressed in immature neurons during migration and dendritic growth and turned off as neurons reach a matured state [68, 84]. Here we observe that there is a gradual increase in the expression levels of DCX as the cells reach day 30. MAP2 is a structural protein necessary to maintain the cytoskeletal structure, mainly present in the dendrites of the neurons [85]. The expression levels of MAP2 also increases gradually as the cells reach neurons. STMN2 is upregulated from day 5 to day 30, it is a neuronal-growth associated protein (nGAP) and its expression is highly correlated to neuronal process, regeneration and outgrowth, other functions of this gene include regulation of neuronal cytoskeleton by microtubule destabilization [86]. Previous studies have shown that NPCs after differentiation showed expression of neuronal markers MAP2 (day 14) and STMN2 (day 50) [87]. The upregulation of DCX, MAP2 and STMN2 is expected based on their biological functions and roles in neuronal development.

The GO analysis by Wang et al. was performed using David software [33, 34] for two gene sets that distinguished the subpopulations. These gene sets were associated to anterior/posterior pattern, axon, neuron differentiation and other relevant terms. In contrast to their approach here the GO analysis was performed for highest expressed genes and marker genes for each cell type using GSEA software [40-42, 48]. Some of the key terms include neurogenesis, synapse formation, forebrain generation of neurons, olfactory lobe development, central nervous system development and others discussed in earlier sections. Some unexpected GO terms included negative regulation of neuron differentiation and nervous system development in the absence of two genes- UGT8 and SALL3. Both genes

have important biological functions, they are involved in regulation of synthesis of myelin sheath and olfactory bulb development respectively [57, 60], but the exact reasons for these GO terms observed is unclear and their regulatory pathways need further studies to elucidate the results.

Despite current research and new findings, our understanding of neurogenesis is still limited in terms of molecular mechanisms which dictate cell proliferation, differentiation and survival [88]. ScRNA-seq analysis is a better tool to elucidate heterogeneity during development and differentiation by providing information regarding the genomic, epigenomic and transcriptomic states of cells during biological processes [89]. Even though this study provides new insights regarding in vitro neurogenesis many questions still remain unanswered. Based on previous studies, the decrease in gene expression of SFRP1 and SFRP2 is due to differentiation of the cells. But the SFRP proteins have a range of functions and they regulate the WNTs in both canonical and non-canonical pathways [90]. Whether they act as antagonists or protagonist of the WNT-signaling pathway is unclear. Previously it is observed that SFRP1 and SFRP2 do not directly regulate WNT5A [90]. In-depth understanding of the molecular interactions between SFRPs and WNTs can help answer why the only WNT being highly expressed during initial proliferative stages is WNT5A. Whether SFRP1 and SFRP2 interacts with WNT5A in a direct or indirect manner or has an independent mode of action is not understood yet. EFNA5 maintains the neural stem cell survival and proliferation [91], which explains high expression levels of EFNA5 in NPCs. Studies have showed that Eph:ephrin regulates axon guidance and synaptogenesis in the CNS [91]. It promotes or inhibits neurogenesis depending on specific Eph and ephrin interactions in forward or reverse manner [56]. While the GO analysis

revealed that EFNA5 is associated with axonogenesis, synapse formation and assembly, it is unclear how its decrease regulates these biological processes. Whether it has a forward or reverse action during neurogenesis is still unanswered. Future studies focusing on answering some of these questions can help in understanding the complex mechanisms and dynamic states of the cells during neurogenesis.

5. Conclusion

The single cell analysis of the undifferentiated NPCs to neurons suggests that there exists heterogeneity within the cell types which is governed by the changes in expressed genes. The trajectory path of NPCs to neurons gives a visual representation of the biological activity of individual cells revealing heterogeneous nature of immature neurons and neurons. The GO results for highest expressed genes revealed that genes common among all the cell types were associated with GO terms for housekeeping. These genes are crucial to all the cells to maintain their proper functioning. Genes CALMN2, STMN1, SOX4 were highly expressed across all cell types except the NPCs. They were associated with GO term for the development of the central nervous system. Among highest expressed genes unique to only NPCs- WNT5A, EFNA5 and ADNP were associated with GO terms related to neuron development. The GO results for marker genes of each cell type revealed association with relevant GO terms such as forebrain generation of neurons, neurogenesis, olfactory lobe development and other related terms. In this study, we observed genes that were upregulated and downregulated as cells transitioned from one state to the other. Genes DCX, MAP2 and STMN2 were upregulated, whereas a gradual downregulation was observed in WNT5A and EFNA5, and SFRP1 and SFRP2 downregulated over a period of two days (day 5 to day 7). In summary, the study reveals critical factors which play important role in neurogenesis. Yet, there are a significant number of unanswered questions which can hinder progress in this area of research. Nevertheless, employing scRNA-seq technology has shown great promise in unfolding some novel insights regarding neurogenesis.

6. Future Directions

The analysis performed in this study and results drawn are purely computational and hence the scRNA-seq pipeline can be completely automated to save computational time. ScRNA-seq platform is constantly improving with new technology and tools for analysis, it is necessary to update the pipeline in order to obtain more robust results. Different tools such as such as Seurat [92], and TSCAN [93] will be used to perform analysis to either validate or improvise the existing results. This study focus on four cell types and overlooks the presence of transitioning states. An analysis using dropClust will be employed for identification of these minor cell subtypes [94]. An unsupervised analysis which does not consider any prior information regarding cell states will be insightful. A better understanding of the biological activity of the cells and gene expression activity will be possible if the scRNA-seq data is generated for nearer time points. In case of genes SFRP1 and SFRP2 transcriptional profile of the cells at day 6 could help us understand functioning of the genes better. While the roles of WNT5A, DCX, MAP2 and STMN2 have been studied excessively, experimental studies to elucidate the functional roles of the genes EFNA5, SFRP1 and SFRP2 and their interactions with other closely related proteins during neurogenesis should be done. This will help to validate results and to reveal their clinical implications. Defects in these factors could play key roles in neurodegenerative diseases. A comparative analysis of neurodegenerative disorders against a healthy brain will elucidate the mechanisms leading to the development of Parkinson's, Schizophrenia and Alzheimer's.

7. Appendix

A.: Supplementary Tables

Table a.: List of top 50 expressed genes in four cell types

NPC (19.3%)	DNC (18.7%)	IN (17.8%)	N (22.9%)
LSM12	TMSB4X	NREP	LDHB
ENO1	YWHAZ	RPL10	SET
RPL7P9	ZNF286A	HNRNPA2B1	SAT1
EFNA5	SCUBE3	ATP5A1	MAP2
YWHAZ	PPIA	YWHAZ	EIF2S3
LDHB	ATP5A1	LDHB	YWHAZ
MARCKS	NR6A1	TBL1XR1	NKTR
ZNF286A	NPM1	DCX	TXNIP
ARRDC3	PAFAH1B2	KPNA2	WSB1
NR6A1	NASP	CAPZA1	FTL
HSP90AA1	ANP32E	H2AFZ	SRSF6
H3F3B	H3F3B	ETNK1	RPS18
ADNP	SOX4	HSP90AA1	HMGCS1
RPL41	LDHB	PPIA	HSP90AA1
SAT1	MAPK1IP1L	ARRDC3	CKB
DDX5	SOX11	H3F3B	RPL7P9
WNT5A	RPL41	PAFAH1B2	RPL10
MALAT1	CALM2	RPL7	CALM2
MAPK1IP1L	MORF4L1	RPL41	RPL41
RACK1	H2AFZ	CALM2	DAAM1
RPL7	RPL7	SOX11	KCNQ1OT1
H2AFZ	HSP90AA1	SRSF3	NREP
SRSF3	PPP1CB	HMGN2	ALDOA
RPL15	CAPZA1	PPP1CB	SOX11
NPM1	RACK1	ENO1	DDX5
HMGN2	HMGN2	EIF2S3	RPL7
CNOT9	DDX5	ANP32E	NEAT1
ETNK1	EIF2S3	RACK1	H3F3B
PAFAH1B2	CBX5	HSP90AB1	RACK1
PPP1CB	RPL15	RPL15	HSP90AB1
SOX11	SRSF3	SET	SCD
GAPDH	HSP90AB1	SOX4	ARRDC3
RPLP0	SET	CBX5	RPL15
ACTG1	RPLP0	DDX5	STMN1
RPL3	GAPDH	RPL3	RPLP0
HSP90AB1	MARCKS	RPLP0	DCX
CAPZA1	RPL3	STMN1	MARCKS
EEF1A1P6	STMN1	GAPDH	RPL3
ANP32E	MALAT1	ACTG1	ACTG1
TUBA1A	PRTG	MARCKS	GAPDH
CBX5	ACTG1	SRSF6	ENO1
EEF1A1P5	MAP1B	ACTB	MTCO1P12
ACTB	DPYSL2	EEF1A1P6	EEF1A1P6
MTCO1P12	SRSF6	DPYSL2	SOX4
DPYSL2	EEF1A1P6	EEF1A1P5	EEF1A1P5
MAP1B	EEF1A1P5	MTCO1P12	MAP1B

EIF2S3	ACTB	MALAT1	DPYSL2
SET	TUBA1A	MAP1B	TUBA1A
SRSF6	MTCO1P12	TUBA1A	MALAT1
EEF1A1	EEF1A1	EEF1A1	EEF1A1

Table b.: The 20 housekeeping genes filtered

ANKDR10	TLK1	ADNP	YWHAE	CHP1
TUBA1C	MARK1IPIL	HNRNPH1	HMGB1	RPL4
UBC	TUBB	C6orf62	HSPA8	TPI1
TUBA1B	MARCH6	FDFT1	MDRF4L2	GPI

Table c.: Gene list for common and unique highest expressed genes among four cell types

Cell type	Number of Genes	List of Genes
NPC ∩ DNC∩ IN ∩ N	27	SET HSP90AB1 MAP1B EEF1A1P5 MTCO1P12 EIF2S3 GAPDH RPL7 DDX5 MALAT1 RPLP0 EEF1A1 SOX11 RPL41 MARCKS DPYSL2 RPL3 LDHB RACK1 HSP90AA1 EEF1A1P6 TUBA1A YWHAZ ACTG1 RPL15 H3F3B SRSF6
N	13	ALDOA SCD CKB DAAM1 NEAT1 RPS18 KCNQ1OT1 TXNIP MAP2 WSB1 NKTR FTL HMGCS1
NPC ∩ DNC ∩ IN	9	PPP1CB ACTB PAFAH1B2 CBX5 SRSF3 HMGN2 H2AFZ ANP32E CAPZA1
DNC	5	NASP TMSB4X PRTG MORF4L1 SCUBE3
NPC	5	WNT5A EFNA5 ADNP LSM12 CNOT9
IN	3	HNRNPA2B1 KPNA2 TBL1XR1
NPC ∩ DNC	4	ZNF286A MAPK1IP1L NPM1 NR6A1
DNC ∩ IN ∩ N	3	CALM2 STMN1 SOX4
IN ∩ N	3	RPL10 NREP DCX
DNC ∩ IN	2	PPIA ATP5A1
NPC ∩ IN ∩ N	2	ENO1 ARRDC3
NPC ∩ N	2	RPL7P9 SAT1
NPC ∩ IN	1	ETNK1

Table d.: List of marker genes of all cell types

NPC	DNC	IN	N
PRTG	ZIC2	STMN2	ADCY2
MEIS1	SLIT2	DCX	POU2F2
HOXB−AS3	WNT5A	MLLT11	SOX4
WNT5A	CWC22	ELAVL4	DCX
HOXA3	SPARC	KIF5A	CADM1
MLLT3	SALL3	NEFM	SOX11
UGT8	SFRP2	MAP2	SALL1
SLIT2	RPL3	GNG3	
SALL3	EEF1A1	INA	
B2M	EEF1A1P6	MAP1B	
SPARC	EEF1A1P5	PRTG	
	RP11−889L3.1	SFRP2	
	RPL3P7	SFRP1	
	RPL19		
	RPL3P4		
	EEF1A1P11		

8. References

1.   Turksen K. and T. Troy, *Human Embryonic Stem Cells Protocols.* Methods in Molecular Biology, 2006. **331**.
2.   Yu, J. and J. Thomson. *Stem Cell Information.* 2016.
3.   Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S. S., Waknitz, M. A., Swiergiel, J. J., Marshall, V. S., & Jones, J. M, *Embryonic stem cell lines derived from human blastocysts. .* Science, 1998. **282**(5391): p. 1145-1147.
4.   Murry, C.E. and G. Keller, *Differentiation of embryonic stem cells to clinically relevant populations: lessons from embryonic development.* Cell, 2008. **132**(4): p. 661-680.
5.   Johnson, M.B., & Walsh, C. A., *Cerebral cortical neuron diversity and development at single-cell resolution. .* Current opinion in neurobiology, 2017. **42**: p. 9-16.
6.   Yang, D., Zhang, Z. J., Oldenburg, M., Ayala, M., & Zhang, S. C., *Human embryonic stem cell-derived dopaminergic neurons reverse functional deficit in parkinsonian rats.* Stem Cells, 2008. **26**(1): p. 55-63.
7.   Ganat, Y.M., Calder, E. L., Kriks, S., Nelander, J., Tu, E. Y., Jia, F, *Identification of embryonic stem cell–derived midbrain dopaminergic neurons for engraftment.* The Journal of clinical investigation., 2012. **122**(8): p. 2928-2939.
8.   Roubal, I., Park, S. J., & Kim, Y. , *Derivation of Neural Precursor Cells from Human Embryonic Stem Cells for DNA Methylomic Analysis.* In Embryonic Stem Cell Protocols, 2014: p. 345-357.
9.   Ming, G.L., & Song, H. , *Adult neurogenesis in the mammalian brain: significant answers and significant questions.* Neuron, 2011. **70(4)**: p. 687-702.
10.   Urbán, N., & Guillemot, F., *Neurogenesis in the embryonic and adult brain: same regulators, different roles. .* Frontiers in cellular neuroscience, 2014. **8**.
11.   Jin, X., *The role of neurogenesis during development and in the adult brain.* European Journal of Neuroscience, 2016. **44**(6): p. 2291-2299.
12.   Wang, J., Jenjaroenpun, P., Bhinge, A., Angarica, V. E., Del Sol, A., Nookaew, I., ... & Stanton, L. W., *Single-cell gene expression analysis reveals regulators of distinct cell subpopulations among developing human neurons.* Genome research, 2017. **27**(11): p. 1783-1794.
13.   Bond, A.M., Ming, G. L., & Song, H., *Adult mammalian neural stem cells and neurogenesis: five decades later.* Cell Stem Cell, 2015. **17**(4): p. 385-395.
14.   Ofengeim, D., Giagtzoglou, N., Huh, D., Zou, C., & Yuan, J. , *Single-Cell RNA sequencing: unraveling the brain one cell at a time.* Trends in molecular medicine, 2017. **23**(6): p. 563-576.
15.   Habib, N., Li, Y., Heidenreich, M., Swiech, L., Avraham-Davidi, I., Trombetta, J. J., ... & Regev, A., *Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons.* Science, 2016. **353**(6302): p. 925-938.
16.   Lengler, J., Jug, F., & Steger, A., *Reliable neuronal systems: the importance of heterogeneity.* PloS onw, 2013. **8**(12).
17.   Kukurba, K.R. and S.B. Montgomery, *RNA Sequencing and Analysis.* Cold Spring Harb Protoc, 2015. **2015**(11): p. 951-69.

18. Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., ... & Rolny, C. , *Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq.* Science, 2015. **347**(6226): p. 1138-1142.

19. Tao, Y., & Zhang, S. C. , *Neural subtype specification from human pluripotent stem cells.* Cell Stem Cell, 2016. **19**(5): p. 573-586.

20. Lake, B.B., Ai, R., Kaeser, G. E., Salathia, N. S., Yung, Y. C., Liu, R., ... & Vijayaraghavan, R., *Neuronal subtypes and diversity revealed by single-nucleus RNA sequencing of the human brain.* Science, 2016. **352**(6293): p. 1586-1590.

21. Hook, P.W., McClymont, S. A., Cannon, G. H., Law, W. D., Morton, A. J., Goff, L. A., & McCallion, A. S. , *Single-Cell RNA-Seq of Mouse Dopaminergic Neurons Informs Candidate Gene Selection for Sporadic Parkinson Disease. .* The American Journal of Human Genetics, 2018. **102**(3): p. 427-446.

22. Chen, R., Wu, X., Jiang, L., & Zhang, Y. , *Single-cell RNA-seq reveals hypothalamic cell diversity.* Cell reports, 2017. **18**(13): p. 3227-3241.

23. Stanley, G., Gokce, O., Treutlein, B., Sudhof, T. C., & Quake, S., *Cellular Taxonomy of the Mouse Striatum as Revealed by Single Cell RNA Sequencing.* Biophysical Journal, 2016. **110**(3): p. 321a.

24. Shin, J., Berg, D. A., Zhu, Y., Shin, J. Y., Song, J., Bonaguidi, M. A., ... & Song, H., *Single-cell RNA-seq with waterfall reveals molecular cascades underlying adult neurogenesis. .* Cell Stem Cell, 2015. **17**(3): p. 360-372.

25. Hochgerner, H., Zeisel, A., Lönnerberg, P., & Linnarsson, S. , *Conserved properties of dentate gyrus neurogenesis across postnatal development revealed by single-cell RNA sequencing. .* Nature Neuroscience 2018: p. 1.

26. Pandey, S., Shekhar, K., Regev, A., & Schier, A. F. , *Comprehensive identification and spatial mapping of habenular neuronal types using single-cell RNA-seq* Current Biology, 2018. **28**(7): p. 1052-1065.

27. Li, H., Horns, F., Wu, B., Xie, Q., Li, J., Li, T., ... & Luo, L. , *Classifying Drosophila Olfactory Projection Neuron Subtypes by Single-Cell RNA Sequencing.* Cell, 2017. **171**(5): p. 1206-1220.

28. Davie, K., Janssens, J., Koldere, D., De Waegeneer, M., Pech, U., Kreft, Ł., ... & Poovathingal, S., *A single-cell transcriptome atlas of the aging Drosophila brain.* Cell, 2018.

29. Tietjen, I., Rihel, J. M., Cao, Y., Koentges, G., Zakhary, L., & Dulac, C, *Single-cell transcriptional analysis of neuronal progenitors.* Neuron, 2003. **38**(2): p. 161-175.

30. Close, J.L., Yao, Z., Levi, B. P., Miller, J. A., Bakken, T. E., Menon, V., ... & Nelson, A. M., *Single-cell profiling of an in vitro model of human interneuron development reveals temporal dynamics of cell type production and maturation. .* Neuron, 2017. **93**(5): p. 1035-1048.

31. Yao, Z., Mich, J. K., Ku, S., Menon, V., Krostag, A. R., Martinez, R. A., ... & Gregor, B. W., *A single-cell roadmap of lineage bifurcation in human ESC models of embryonic brain development.* Cell Stem Cell, 2017. **20**(1): p. 120-134.

32. Fiddes, I.T., Lodewijk, G. A., Mooring, M., Bosworth, C. M., Ewing, A. D., Mantalas, G. L., ... & Lorig-Roach, R., *Human-specific NOTCH2NL genes affect Notch signaling and cortical neurogenesis.* Cell, 2018. **173**(6): p. 1356-1369.

33. Huang, D.W., Sherman, B. T., & Lempicki, R. A. , *Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources.* Nature Protocols, 2008. **4**(1): p. 44.

34. Huang, D.W., Sherman, B. T., & Lempicki, R. A. , *Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene lists.* Nucleic acids research, 2008. **37**(1): p. 1-13.

35. Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., ... & Rinn, J. L., *The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells.* Nature Biotechnology, 2014. **32**(4): p. 381.

36. Qiu, X., Hill, A., Packer, J., Lin, D., Ma, Y. A., & Trapnell, C. , *Single-cell mRNA quantification and differential analysis with Census.* Nature methods, 2017. **14**(3): p. 309.

37. Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H., & Trapnell, C. , *Reversed graph embedding resolves complex single-cell developmental trajectories.* BioRxiv, 2017. **110668**.

38. McCarthy, D.J., Campbell, K. R., Lun, A. T., & Wills, Q. F. , *Scater: pre-processing, quality control, normalization and visualization of single-cell RNA-seq data in R. .* Bioinformatics, 2017. **33**(8): p. 1179-1186.

39. Kiselev, V.Y., Kirschner, K., Schaub, M. T., Andrews, T., Yiu, A., Chandra, T., ... & Hemberg, M., *SC3: consensus clustering of single-cell RNA-seq data. .* Nature Methods, 2017. **14**(5): p. 483.

40. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P., *Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles.* Proceedings of the National Academy of Sciences 2005. **102**(43): p. 15545-15550.

41. Liberzon, A., Subramanian, A., Pinchback, R., Thorvaldsdóttir, H., Tamayo, P., & Mesirov, J. P. , *Molecular signatures database (MSigDB) 3.0.* Bioinformatics, 2011. **27**(12): p. 1739-1740.

42. Liberzon, A., Birger, C., Thorvaldsdóttir, H., Ghandi, M., Mesirov, J. P., & Tamayo, P. , *The molecular signatures database hallmark gene set collection.* Cell systems, 2015. **1**(6): p. 417-425.

43. X, G.P.v.f.M.O., *Normality test and Kruskal-Wallis non-parametric test followed by Dunnett's multiple comparisons test.* 2017: La Jolla California USA. p. GraphPad Prism Software.

44. Komin, N., & Skupin, A., *How to address cellular heterogeneity by distribution biology.* Current Opinion in Systems Biology, 2017. **3**: p. 154-160.

45. Altschuler, S.J., & Wu, L. F. , *Cellular heterogeneity: do differences make a difference?* Cell, 2010. **141**(4): p. 559-563.

46. Dulken, B.W., Leeman, D. S., Boutet, S. C., Hebestreit, K., & Brunet, A. , *Single-cell transcriptomic analysis defines heterogeneity and transcriptional dynamics in the adult neural stem cell lineage.* Cell reports, 2017. **18**(3): p. 777-790.

47. Eisenberg, E., & Levanon, E. Y., *Human housekeeping genes, revisited.* Trends in Genetics, 2013. **29**(10): p. 569-574.

48. *Bioinformatics and Evolutionary Genomics.* Available from: http://bioinformatics.psb.ugent.be/webtools/Venn/.

49. Hsiao, L.L., Dangond, F., Yoshida, T., Hong, R., Jensen, R. V., Misra, J., ... & Weng, Z., *A compendium of gene expression in normal human tissues.* Physiological genomics, 2001. **7**(2): p. 97-104.

50. Toutenhoofd, S.L., Foletti, D., Wicki, R., Rhyner, J. A., Garcia, F., Tolon, R., & Strehler, E. E., *Characterization of the human CALM2 calmodulin gene and comparison of the transcriptional activity of CALM1, CALM2 and CALMS. .* Cell calcium, 1998. **23**(5): p. 323-338.

51. Bergsland, M., Werme, M., Malewicz, M., Perlmann, T., & Muhr, J., *The establishment of neuronal properties is controlled by Sox4 and Sox11.* Genes & development, 2006. **20**(24): p. 3475-3486.

52. Cheung, M., Abu-Elmagd, M., Clevers, H., & Scotting, P. J. , *Roles of Sox4 in central nervous system development1.* Molecular brain research, 2000. **79**(1-2): p. 180-191.

53. Machado-Neto, J.A., Saad, S. T. O., & Traina, F. , *Stathmin 1 in normal and malignant hematopoiesis.* BMB reports, 2014. **47**(12): p. 660.

54. Lange, C., Mix, E., Rateitschak, K., & Rolfs, A., *Wnt signal pathways and neural stem cell differentiation. .* Neurodegenerative Diseases, 2006. **3**(1-2): p. 76-86.

55. Mandel, S., Rechavi, G., & Gozes, I., *Activity-dependent neuroprotective protein (ADNP) differentially interacts with chromatin to regulate genes essential for embryogenesis.* Developmental Biology, 2007. **303**(2): p. 814-824.

56. Laussu, J., Khuong, A., Gautrais, J., & Davy, A., *Beyond boundaries—Eph: ephrin signaling in neurogenesis. .* Cell adhesion & migration, 2014. **8**(4): p. 349-359.

57. Guan, Q., Wang, X., Jiang, Y., Zhao, L., Nie, Z., & Jin, L. , *RNA-Seq Expression Analysis of Enteric Neuron Cells with Rotenone Treatment and Prediction of Regulated Pathways.* Neurochemical research, 2017. **42**(2): p. 572-582.

58. Dzięgiel, P., Owczarek, T., Plazuk, E., Gomułkiewicz, A., Majchrzak, M., Podhorska-Okołów, M., ... & Ugorski, M. , *Ceramide galactosyltransferase (UGT8) is a molecular marker of breast cancer malignancy and lung metastases. .* British journal of cancer, 2010. **103**(4): p. 524.

59. Hunter, M., Demarais, N. J., Faull, R. L., Grey, A. C., & Curtis, M. A., *Layer-specific lipid signatures in the human subventricular zone demonstrated by imaging mass spectrometry.* Scientific Reports, 2018. **8**(1): p. 2552.

60. Heng, X., Breer, H., Zhang, X., Tang, Y., Li, J., Zhang, S., & Le, W., *Sall3 correlates with the expression of TH in mouse olfactory bulb. .* Journal of Molecular Neuroscience, 2012. **46**(2): p. 293-302.

61. Harrison, S.J., Nishinakamura, R., Jones, K. R., & Monaghan, A. P., *Sall1 regulates cortical neurogenesis and laminar fate specification in mice: implications for neural abnormalities in Townes-Brocks syndrome.* Disease models & mechnisms, 2012. **dmm-002873**.

62. *Quick GO: Neuron Projection.* European Bioinformatics Institute; Available from: https://www.ebi.ac.uk/QuickGO/term/GO:0043005.

63. *Quick GO: Neuron Part.* European Bioinformatics Institute; Available from: https://www.ebi.ac.uk/QuickGO/term/GO:0097458.

64. Esteve, P., Sandonìs, A., Ibañez, C., Shimono, A., Guerrero, I., & Bovolenta, P. , *Secreted frizzled-related proteins are required for Wnt/β-catenin signalling activation in the vertebrate optic cup.* Development, 2011. **138**(19): p. 4179-4184.

65. Bovolenta, P., Esteve, P., Ruiz, J. M., Cisneros, E., & Lopez-Rios, J. , *Beyond Wnt inhibition: new functions of secreted Frizzled-related proteins in development and disease.* J Cell Sci, 2008. **121**(6): p. 737-746.

66. Lin, X., Dong, R., Diao, S., Yu, G., Wang, L., Li, J., & Fan, Z. , *SFRP2 enhanced the adipogenic and neuronal differentiation potentials of stem cells from apical papilla.* Cell biology international, 2017. **41**(5): p. 534-543.

67. Leimeister, C., Bach, A., & Gessler, M. , *Developmental expression patterns of mouse sFRP genes encoding members of the secreted frizzled related protein family.* Mechanisms of Development, 1988. **75**(1-2): p. 29-42.

68. Gleeson, J.G., Lin, P. T., Flanagan, L. A., & Walsh, C. A. , *Doublecortin is a microtubule-associated protein and is expressed widely by migrating neurons.* Neuron, 1999. **23**(2): p. 257-271.

69. Yap, C.C., Digilio, L., McMahon, L., Roszkowska, M., Bott, C. J., Kruczek, K., & Winckler, B. , *Different doublecortin (DCX) patient alleles show distinct phenotypes in cultured neurons: evidence for divergent loss-of-function and off-pathway cellular mechanisms.* Journal of Biological Chemistry, 2016. **jbc-M116**.

70. Dehmelt, L., & Halpain, S. , *The MAP2/Tau family of microtubule-associated proteins.* Genome biology, 2005. **6**(1): p. 204.

71. Sanchez, C., Dıaz-Nido, J., & Avila, J., *Phosphorylation of microtubule-associated protein 2 (MAP2) and its relevance for the regulation of the neuronal cytoskeleton function. .* Progress in neurobiology, 2000. **61**(2): p. 133-168.

72. Chiellini, C., Grenningloh, G., Cochet, O., Scheideler, M., Trajanoski, Z., Ailhaud, G., ... & Amri, E. Z., *Stathmin-like 2, a developmentally-associated neuronal marker, is expressed and modulated during osteogenesis of human mesenchymal stem cells.* Biochemical and biophysical research communications, 2008. **374**(1): p. 64-68.

73. Camoletto, P., Colesanti, A., Ozon, S., Sobel, A., & Fasolo, A. , *Expression of stathmin and SCG10 proteins in the olfactory neurogenesis during development and after lesion in the adulthood.* Brain research bulletin, 2001. **54**(1): p. 19-28.

74. *Quick GO: Cell Development.* European Bioinformatics Institute; Available from: https://www.ebi.ac.uk/QuickGO/term/GO:0048468.

75. Petroff, O.A., *Book review: GABA and glutamate in the human brain.* The Neuroscientist, 2002. **8**(6): p. 562-573.

76. Lujan, R., Shigemoto, R., & Lopez-Bendito, G. , *Glutamate and GABA receptor signalling in the developing brain. .* Neuroscience, 2005. **130**(3): p. 567-580.

77. Nathan, E., & Tzahor, E. , *sFRPs: a declaration of (Wnt) independence.* Nature cell Biology, 2009. **11**(1): p. 13.

78. Esteve, P., Sandonìs, A., Cardozo, M., Malapeira, J., Ibañez, C., Crespo, I., ... & Shimono, A. , *SFRPs act as negative modulators of ADAM10 to regulate retinal neurogenesis.* Nature Neuroscience, 2011. **14**(5): p. 562.

79. Kele, J., Andersson, E. R., Villaescusa, J. C., Cajanek, L., Parish, C. L., Bonilla, S., ... & Arenas, E. , *SFRP1 and SFRP2 dose-dependently regulate midbrain dopamine neuron development in vivo and in embryonic stem cells.* Stem Cells, 2012. **30**(5): p. 865-875.

80. Andersson, E.R., Prakash, N., Cajanek, L., Minina, E., Bryja, V., Bryjova, L., ... & Arenas, E. , *Wnt5a regulates ventral midbrain morphogenesis and the development of A9–A10 dopaminergic cells in vivo.* PLoS one, 2008. **3**(10): p. e3517.

81. Subashini, C., Dhanesh, S. B., Chen, C. M., Riya, P. A., Meera, V., Divya, T. S., ... & James, J., *Wnt5a is a crucial regulator of neurogenesis during cerebellum development.* Scientific Reports, 2017. **7**: p. 42523.

82. Worku, T., Wang, K., Ayers, D., Wu, D., Ur Rehman, Z., Zhou, H., & Yang, L., *Regulatory roles of ephrinA5 and its novel signaling pathway in mouse primary granulosa cell apoptosis and proliferation.* Cell Cycle, 2018: p. 1-11.

83. Hara, Y., Nomura, T., Yoshizaki, K., Frisén, J., & Osumi, N. , *Impaired Hippocampal neurogenesis and vascular formation in ephrin‑A5‑deficient mice.* Stem Cells, 2010. **28**(5): p. 974-983.

84. Spampanato, J., Sullivan, R. K., Turpin, F. R., Bartlett, P. F., & Sah, P., *Properties of doublecortin expressing neurons in the adult mouse dentate gyrus.* PLoS one, 2012. **7**(9): p. e41029.

85. Johnson, G.V.W., & Jope, R. S., *The role of microtubule‑associated protein 2 (MAP‑2) in neuronal growth, plasticity, and degeneration.* Journal of neuroscience research, 1992. **33**(4): p. 505-512.

86. Mori, N., & Morii, H. , *SCG10‑related neuronal growth‑associated proteins in neural development, plasticity, degeneration, and aging.* Journal of neuroscience research, 2002. **70**(3): p. 264-273.

87. Blair, J.D., Hockemeyer, D., Doudna, J. A., Bateup, H. S., & Floor, S. N., *Widespread translational remodeling during human neuronal differentiation.* Cell reports, 2017. **21**(7): p. 2005-2016.

88. Zhao, C., Deng, W., & Gage, F. H. , *Mechanisms and functional implications of adult neurogenesis.* Cell, 2008. **132**(4): p. 645-660.

89. Duran, R.C.D., Wei, H., & Wu, J. Q. , *Single-cell RNA-sequencing of the brain. .* Clinical and translational medicine, 2017. **6**(1): p. 20.

90. Satoh, W., Gotoh, T., Tsunematsu, Y., Aizawa, S., & Shimono, A., *Sfrp1 and Sfrp2 regulate anteroposterior axis elongation and somite segmentation during mouse embryogenesis. .* Development, 2006. **133**(6): p. 989-999.

91. Jiao, J.W., Feldheim, D. A., & Chen, D. F., *Ephrins as negative regulators of adult neurogenesis in diverse regions of the central nervous system.* Proceedings of the National Academy of Sciences, 2008. **105**(25): p. 8778-8783.

92. Butler, A., Hoffman, P., Smibert, P., Papalexi, E., & Satija, R. , *Integrating single-cell transcriptomic data across different conditions, technologies, and species.* Nature Biotechnology, 2018. **36**(5): p. 411.

93. Ji, Z., & Ji, H., *TSCAN: Pseudo-time reconstruction and evaluation in single-cell RNA-seq analysis.* Nucleic acids research, 2016. **44**(13): p. e117-e117.

94. Sinha, D., Kumar, A., Kumar, H., Bandyopadhyay, S., & Sengupta, D., *dropClust: efficient clustering of ultra-large scRNA-seq data.* Nucleic acids research, 2018. **46**(6): p. e36-e36.