From Department of Neuroscience

Karolinska Institutet, Stockholm, Sweden

SINGLE CELL ANALYSIS OF IPS DERIVED NEURAL STEM CELLS AND NEURONS

Matti Lam

Karolinska Institutet

Stockholm 2019

Published by Karolinska Institutet.
Printed by E-Print AB 2019

ISBN 978-91-7831-643-4

Single cell analysis of IPS derived neural stem cells and neurons

THESIS FOR DOCTORAL DEGREE (Ph.D.)

ACADEMIC DISSERTATION

Public defense at Atrium lecture hall, Nobels väg 12b, Solna

Thursday 12th December 2019

By

Matti Lam

Principal Supervisor:
Docent Anna Falk
Karolinska Institutet
Department of Neuroscience

Co-supervisor(s):
Jonas Muhr
Karolinska Institutet
Department of Cell and Molecular Biology

Ana Teixeira
Karolinska Institutet
Department of Medical Biochemistry and Biophysics

Julian Walfridsson
Karolinska Institutet
Department of Medicine, Huddinge

Opponent:
Associate Professor Marius Wernig
Stanford University
Institute for Stem Cell Biology and Regenerative Medicine

Examination Board:
Assistant Professor Qiaolin Deng
Karolinska Institutet
Department of Physiology and Pharmacology

Assistant Professor Patrik Ståhl
Kungliga Tekniska Högskolan
Genteknologi, ScilLifeLab

Professor Lena Gunhaga
Umeå Universitet
Centre for Molecular Medicine

Chair:
Professor Ola Hermansson
Karolinska Institutet
Department of Neuroscience

I studied biomedicine at Lund University, received my bachelor's and master's degrees and moved on to PhD studies at Karolinska Institutet in 2015. During my time at Karolinska Institutet, I worked on cells, human neural stem cells and neurons derived from IPS cells. At a defining point during my first year of studies I realized I needed to acquire an inclusive perspective on how gene expression affects the behavior and function of cells and I learned how to look at data. In my second year I was accepted into the Swedish bioinformatics advisory program and was mentored in the "art" of single cell analysis. Now, at the end of my PhD studies, I consider myself competent in single cell analysis and neurobiology in order to merge two fields of scientific study.

Data + Biology = Truth

“… there are too many coincidences, this is too strange to not be true…”

- recurring thought

If I have seen further it is by standing on the shoulders of Giants.

- Isaac Newton

ABSTRACT

In vitro derived neural stem cells can be used to model events of neurogenesis and neurodevelopmental dysfunction. Formation of stable neural stem cell environments relies on cellular components to stabilize cell-to-cell connections. Cell-to-cell connections are required to stabilize cell-to-cell communication and cell-signaling events that are required to maintain cell identity and cell integrity. Identifying gene expression profiles and protein expression profiles aid in discovery of cellular events governing functional neurogenesis and events leading to neurodevelopmental dysfunction.

In Paper 1, we used single cell RNA-seq and identified cell adhesion molecule neurexin-1 alpha (*NRXN1-a*) as important for establishing functional efficient neural stem cells and for establishing functional neurons.

In Paper 2, we used single cell RNA-seq and identified presence of inherent neurogenic progenitors and gliogenic progenitors in functional efficient neural stem cells and the neuronal or glial differentiation outcome to be linked to presence of predetermined progenitor cell types.

In Paper 3, we identified microtubule stabilizing molecule doublecortin (*DCX*) as important for neuronal dendrite elongation, neuronal differentiation, and neuronal migration.

In summary, this thesis demonstrates the potential of modeling neurodevelopment by using neural stem cells as a tool for investigation.

Publications not listed in thesis

I. **Transcription factor programming of human ES cells generates functional neurons expressing both upper and deep layer cortical markers**. Giedre Miskinyte, Marita Grønning Hansen, Emanuela Monni, **Matti Lam**, Johan Bengzon, Olle Lindvall, Henrik Ahlenius, Zaal Kokaia (2018) *PLoS ONE* 13(10).

II. **Protein receptor-independent plasma membrane remodeling by HAMLET: a tumoricidal protein-lipid complex**. Aftab Nadeem, Jeremy Sanborn, Douglas L. Gettel, Ho C. S. James, Anna Rydström, Viviane N. Ngassam, Thomas Kjær Klausen, Stine Falsig Pedersen, **Matti Lam**, Atul N. Parikh & Catharina Svanborg. (2015) *Scientific Reports* 5.

LIST OF SCIENTIFIC PAPERS

I. **Matti Lam**, Mohsen Moslem, Julien Bryois, Robin J Pronk, Elias Uhlin, Ivar Dehnisch Ellström, Loora Laan, Jessica Olive, Rebecca Morse, Harriet Rönnholm, Lauri Louhivuori, Sergiy V Korol, Niklas Dahl, Per Uhlén, Britt-Marie Anderlid, Malin Kele, Patrick F Sullivan, Anna Falk
Single cell analysis of autism patient with bi-allelic *NRXN1-alpha* deletion reveals skewed fate choice in neural progenitors and impaired neuronal functionality
Experimental Cell Research, 2019

II. **Matti Lam,** Tsukasa Sanosaka, Anders Lundin, Kent Imaizumi, Damla Etal, Jonathan Cairns, Ryan Hicks, Jun Kohyama, Malin Kele, Hideyuki Okano, Anna Falk
Single cell study of neural stem cells derived from human iPSCs reveal distinct progenitor populations with neurogenic and gliogenic potential.
Genes to Cells, 2019

III. M Shahsavani, R J Pronk, R Falk, **Matti Lam,** M Moslem, S B Linker, J Salma, K Day, J Schuster, B-M Anderlid, N Dahl, F H Gage, A Falk
An in vitro model of lissencephaly: expanding the role of DCX during neurogenesis
Molecular Psychology, 2017

CONTENTS

1 Introduction and observations 8
1.1 Induced pluripotent stem cells and neural stem cells 9
1.2 Neural stem cell rosette structure and cell proliferation 9
1.3 NOTCH signaling in neural rosette 11
1.4 Neurogenic progenitors and gliogenic progenitors 12
1.5 Modeling Neurogenesis 13
1.6 Modeling Gliogenesis 14
2 NRNX1 expression and cell adhesion during early mammalian development 15
2.1 Neurexin-1 protein structure 15
2.2 NRXN1 in Human cell models and fetal development 17
2.3 Nrxn1 in early Mouse development 21
2.4 NRXN1 and CDH8 in early Frog development 22
2.5 NRXN1 Evolutionary conserved across species 22
2.6 CDH2, CDH8 and CDH6 in early Chick development 25
3 Revisiting doublecortin mutation study 28
4 Single cell RNA seq 29
4.1 Single cell isolation protocols 29
4.2 Methods of sequencing 30
4.2.1 Single-cell Tagged Reverse Transcription (STRT)-Seq 30
4.2.2 Random Displacement Amplification (RamDA)-Seq 32
4.3 Methods of analysis for single cell data 33
4.3.1 Number of single cells and sequencing depth 33
4.3.2 Quality Control, RNA-seq data and single cell quality evaluation 34
4.3.3 BACKSPIN – Detecting highly variable genes 35
4.3.4 Monocle – Pseudotemporal ordering and trajectory projection 35
4.3.5 Seurat – Single cell data analysis package 36
4.3.6 Data Visualization for single cell data 36
5 Aims of Papers 39
5.1 Paper 1 39
5.2 Paper 2 39
5.3 Paper 3 39
6 Summary, Discussion and Future 41
7 Acknowledgements 42
8 References 43

LIST OF ABBREVIATIONS

RNA	Ribonucleic acid
Seq	Sequencing
iPS cells	induced pluripotent stem cells
NSCs	neural stem cells
ES cells	embryonic stem cells
NES cells	neuroepithelial stem cells
NOTCH1	Notch receptor 1
DLL1	Delta Like Canonical Notch Ligand 1
JAG1	Jagged Canonical Notch Ligand 1
CDH2	Cadherin 2
NRXN1	Neurexin 1
CDH8	Cadherin 8
CDH6	Cadherin 6
ITGA1	Intergrin Subunit Alpha 1
FGF	Fibroblast growth factor
EGF	Epidermal growth factor
DCX	Doublecortin
STRT-seq	Single-cell Tagged Reverse Transcription sequencing
RamDA-seq	Random Displacement Amplification sequencing

1 INTRODUCTION AND OBSERVATIONS

Living organisms consist of single cells and seemingly uniform cells at a glance are upon close inspection very different in their composition and can vary much in transcriptional state, epigenetic state and intracellular microenvironment.

Although genetically identical, cells exhibit stochastic gene expression as a result of alterations in extrinsic and intrinsic cues driving cellular mechanisms of transcription and translation (Elowitz et al., 2002; Raj and van Oudenaarden, 2008). In a general sense, cellular homeostasis and development drive the underlying heterogeneity in cells (Altschuler and Wu, 2010) since different cell types are performing different functions (Arendt, 2008).

Consider the brain, a complex organ composed of hundreds of millions of neuronal and non-neuronal cells. The massive number of cells displays a large variation in morphology, connectivity and function (Bota and Swanson, 2007; Nelson et al., 2006). Further, location, morphology, functionality, connective specificity, molecular markers and gene expression profiles classify the cells of the brain (Custo Greig et al., 2013; DeFelipe et al., 2013; Molyneaux et al., 2007).

RNA sequencing (seq) has become standard for interrogating gene expression and offers a global view of the averaged gene expression profile in a bulk cell sample (Lister et al., 2008; Mortazavi et al., 2008; Nagalakshmi et al., 2008). However, sampling thousands or millions of cells in bulk sample and assessing the average of gene expression will mask single cell gene expression heterogeneity. Therefore, gene expression profiles contributed by rare single cells in sampled bulk of cells will be indistinguishable from overall gene expression noise. Experimental methods for isolation of single cells and acquiring cDNA from low amounts of RNA have allowed single cell RNA-seq to overcome the limitation of studying averaged gene expression in cell samples (Kurimoto et al., 2007; Tang et al., 2009). Advances in single-cell transcriptomic technologies allow for profiling of hundreds to thousands of single cells, returning high-resolution profiles of unique cell types widens and deepens the understanding of tissue complexity.

Using single cell RNA-seq to study gene expression profiles in induced pluripotent stem (iPS) cell derived neural stem cells (NSCs) and neurons allow for in-depth investigation of modeling neurodevelopment and neurodevelopmental disorders in the laboratory in vitro setting.

1.1 INDUCED PLURIPOTENT STEM CELLS AND NEURAL STEM CELLS

The discovery of induced pluripotent stem (iPS) cells opened a new field of scientific study into biological processes governing development and establishment of cells. Fibroblasts reprogrammed into iPS cells contain the capacity to differentiate into all three germlayers, ectoderm, endoderm and mesoderm (Takahashi et al., 2007; Takahashi and Yamanaka, 2006).

For generating cells of the ectoderm, embryonic stem (ES) cells (Thomson et al., 1998) and iPS cells can be patterned toward enrichment of ectodermal cells by neural induction. This process mimics gastrulation and neurulation of the developing embryo and involves blocking signaling pathways for maintaining propagation of the endoderm and the mesoderm lineage. Neural induction efficiently establishes tri-potent neural stem cells holding the potential to generate neurons, oligodendrocytes and astrocytes (Chambers et al., 2009; Falk et al., 2012; Koch et al., 2009) (Figure 1).

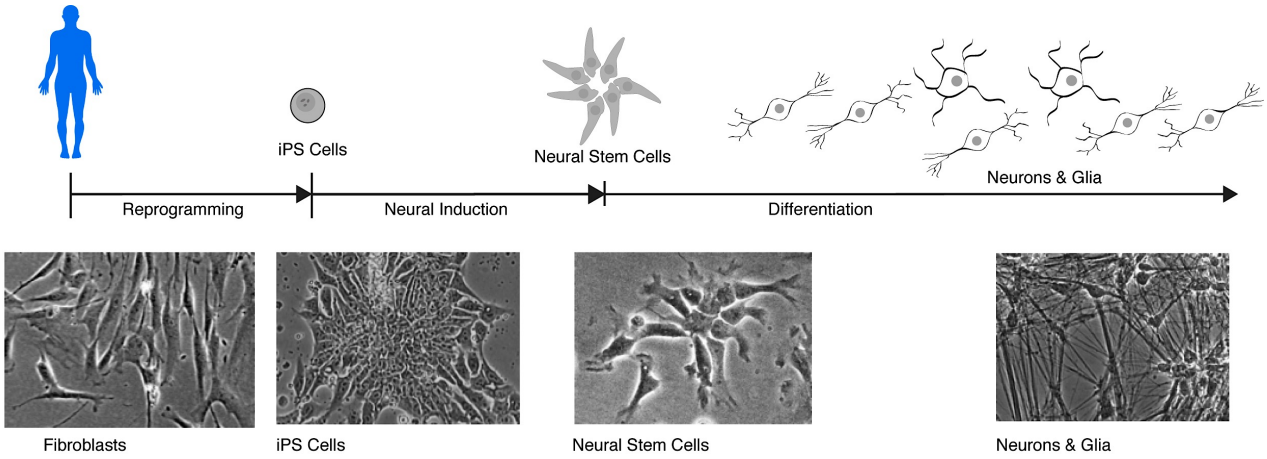


Figure 1. Overview of iPS cell derivation, NES cell derivation and neuron differentiation. Fibroblasts taken from skin biopsy reprogrammed into iPS cells using *OCT4*, *KLF4*, *SOX2*, *MYC*, iPS cells induced to neural stem cells by endorsing enrichment off ectoderm meanwhile blocking pathways signaling for mesoderm and endoderm, NES cells differentiate into neurons and glia by removal of growth factors.

1.2 NEURAL STEM CELL ROSETTE STRUCTURE AND CELL PROLIFERATION

Neural stem cells self-organize into neural rosette structures mimicking cell composition of proliferating cell layer lining the cavity of the neural tube and proliferating cells lining the ventricular cavity in the neocortex. The apical central ring presents an organizing structure, in which cells undergoing mitosis are positively stained for phosphor-histone 3 (pH3) indicating condensed chromatin right before chromosomal segregation. The basal peripheral structure representing cells in non-proliferative state (Banda et al., 2015; Haubst et al., 2006; Horner and Caspary, 2011; Kadowaki et al., 2007; Le Dreau et al., 2014) (Figure 2).

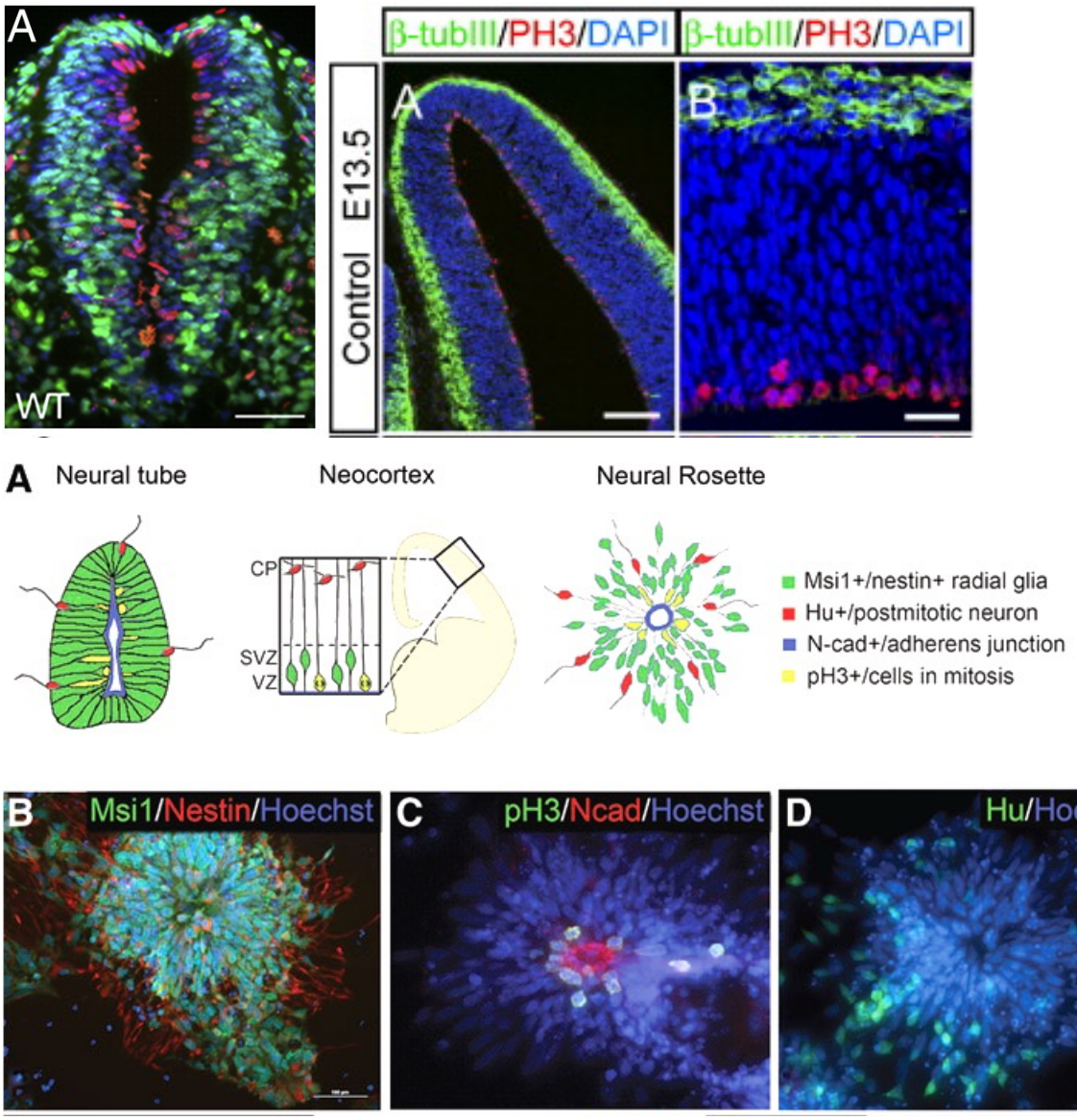


Figure 2. Upper left A. Cell proliferation in the mouse E10.5 wildtype neural tube. Section stained with antibodies against phospho-histone H3 (in red) and BrdU (in green). All cells in the neural tube are stained with Hoechst (blue) (adapted from Fig. 3 (Horner and Caspary, 2011)). Upper right A-B. E13.5 wildtype neocortex at midneurogenesis staining for β-tubulin III (β-tub III), phospho-histone H3 (PH3) and nuclear marker DAPI (adapted from Fig. 3 (Kadowaki et al., 2007)). Middle panel A. Mature human embryonic stem cell (hESC)-derived rosettes resemble the developing neural tube and neocortex and establish a basal and an apical domain. Lower panel B-D. Neural rosette expression of radial glia markers Musashi (Msh1) and Nestin, postmitotic neuron marker Hu, adherens junction marker Ncad (Cdh2) , and M-phase marker pH3 (adapted from Fig. 1 (Banda et al., 2015))

1.3 NOTCH SIGNALING IN NEURAL ROSETTE

Tight cell contact is essential for cell-to-cell NOTCH signaling required for maintaining neural stem cells (Hatakeyama et al., 2014). The cell-to-cell signaling mediated through NOTCH signaling is strengthened by the concentric neural rosettes structure with the cells lining the apical ring structure receive the most efficient NOTCH signal while cells located further away from the center are subject to weaker NOTCH1 signaling (Woo S-M, 2009) (Figure 3).

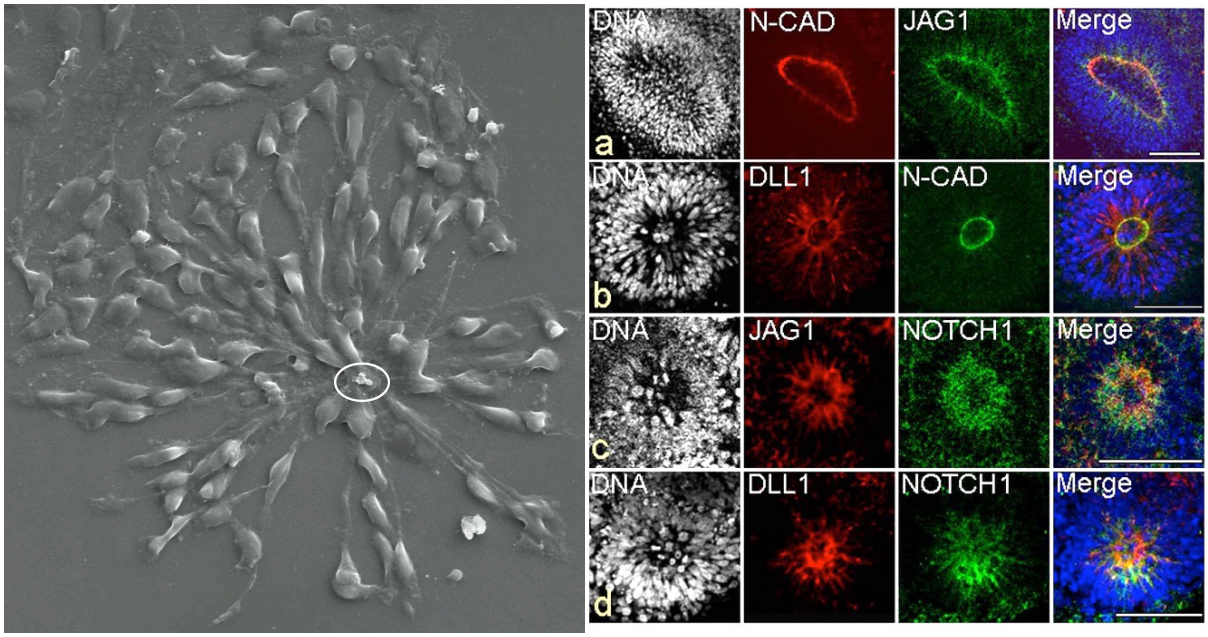


Figure 3. Left, scanning electron microscopy image of neural rosette, highlight for apical ring structure, (image from Grabel lab used with permission) Right. Notch signaling activity n neuroectodermal spheres derived from human embryonic stem cells. Immunostaining of neural rosettes for JAG1 and DLL1 (Notch ligands) and NOTCH1 (receptor). Rosettes structures derived from hESC-derived NESs were visualized using apical rosette marker N-cadherin (N-CAD or CDH2) and shown located to inner rims of rosettes. Both JAG1 (a and c) and DLL1 (b and d) transmembrane ligands are localized in the regions where N-cadherin signals exist. Anti-NOTCH1 antibody, which recognizes both NICD and C-terminal cytoplasmic domain of membrane-bound NOTCH1 receptor simultaneously, generates a rather diffused signal (c and d); nevertheless, its location at the luminal side of rosette is evident. Scale bars, 100 mm. (adapted from Fig. 2 (Woo S-M, 2009))

1.4 NEUROGENIC PROGENITORS AND GLIOGENIC PROGENITORS

The presence of predisposed lineage identity exists within established neural stem cells. This condition manifests the presence of neurogenic progenitors holding potential for differentiation to neuronal cells and gliogenic progenitors holding potential for differentiating to non-neuronal cells (Paper 2, Lam et al 2019).

NOTCH signaling plays an important part for dictating neural stem cell potential, the expression of *DLL1* indicating neurogenic progenitor identity and in contrast low expression of *DLL1* indicating gliogenic progenitor identity. Similarly, expression of distinguished cell adhesion molecules can be observed to differentiate between neurogenic progenitors (*NRXN1*, *CDH8*) and gliogenic progenitors (*ITGA1*, *CDH6*) (Figure 4).

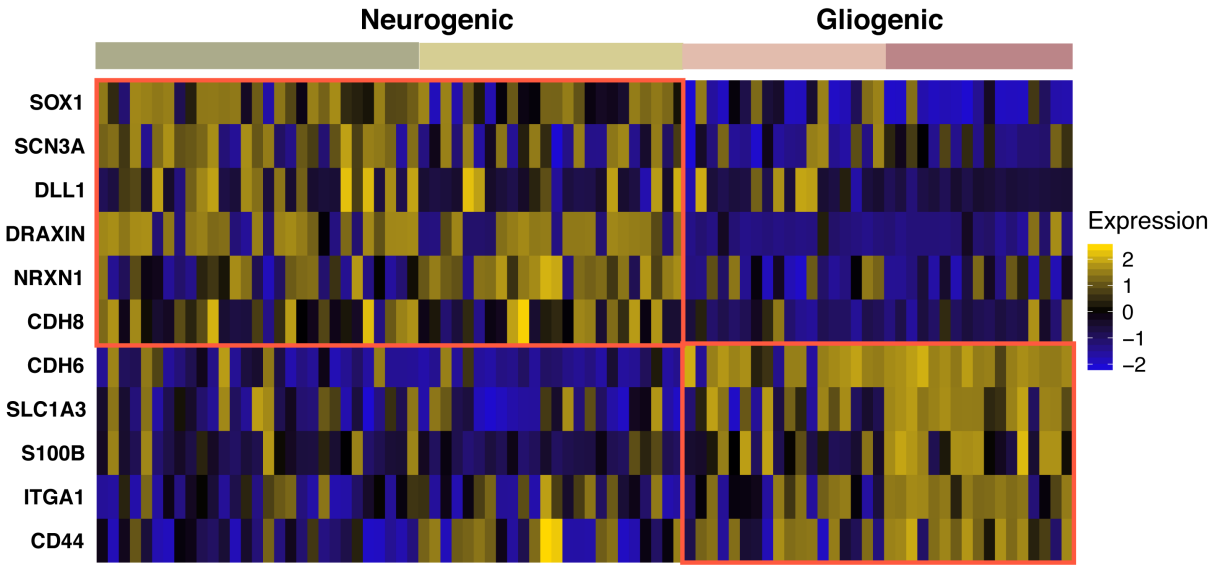


Figure 4. Single cell total RNA RamDA-seq data for AF22 neural stem cells. Heatmap for enriched genes selected for neurogenic progenitors and gliogenic progenitors (subset of data for Figure 1E used in Paper 2, Lam et al 2019).

1.5 MODELING NEUROGENESIS

Differentiation of neural stem cells can be induced by removing growth factors FGF and EGF from the cell culture condition ((Falk et al., 2012; Lam et al., 2019; Sun et al., 2008; Tailor et al., 2013)). Growth factors promote cell proliferation and maintain optimal stem cell environment (Tropepe et al., 1999). Established neural stem cells are primed ectodermal progenitor cells and a sequence of events following growth factors withdrawal occur when initiating undirected differentiation. Initially an increase in cell proliferation is followed by massive cell death, in sequence the emergence of neuroblasts and neurons becomes visible at end of the first week of differentiation. Standard outcome proportion for neural stem cell differentiation averages around 80% neurons and 20% glial cells and can be used as a measure of well-established neural stem cells (Figure 5).

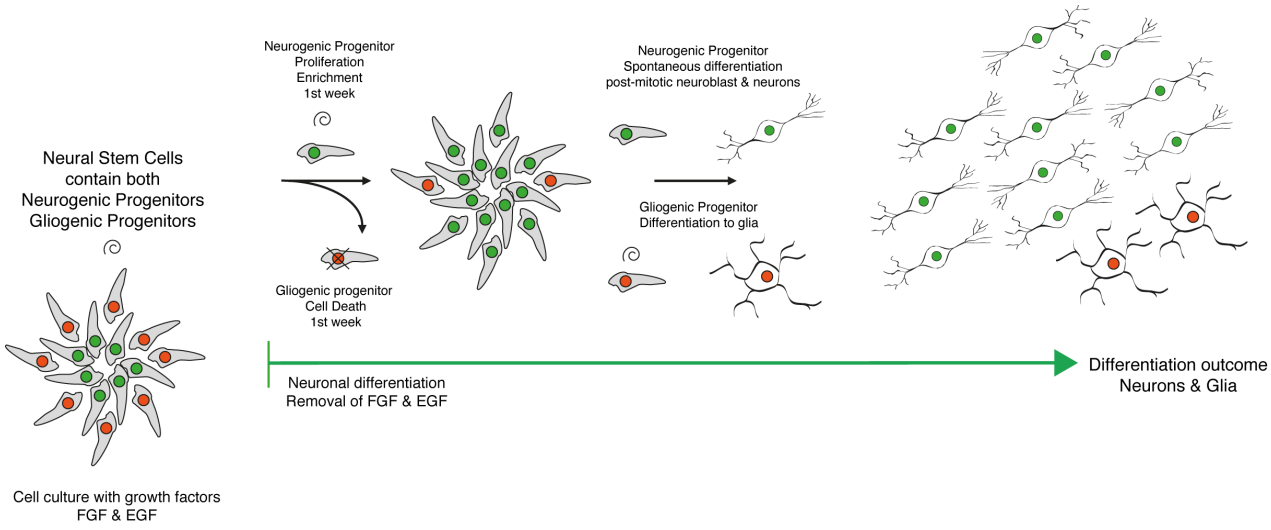


Figure 5. Illustration of neural stem cell model system for neurogenesis (adapted from Fig 5 Paper 2, Lam et al 2019).

1.6 MODELING GLIOGENESIS

A second aspect of differentiation potential for neural stem cells is the presence of gliogenic progenitors that have the potential to differentiate to glial cells. Glia patterning culture protocol enrich for glial cells, maintains the cells in a proliferative state and continuous culture passaging of cells removes spontaneously differentiating neuroblasts (Lundin et al., 2018; Shaltouki et al., 2013) (Figure 6).

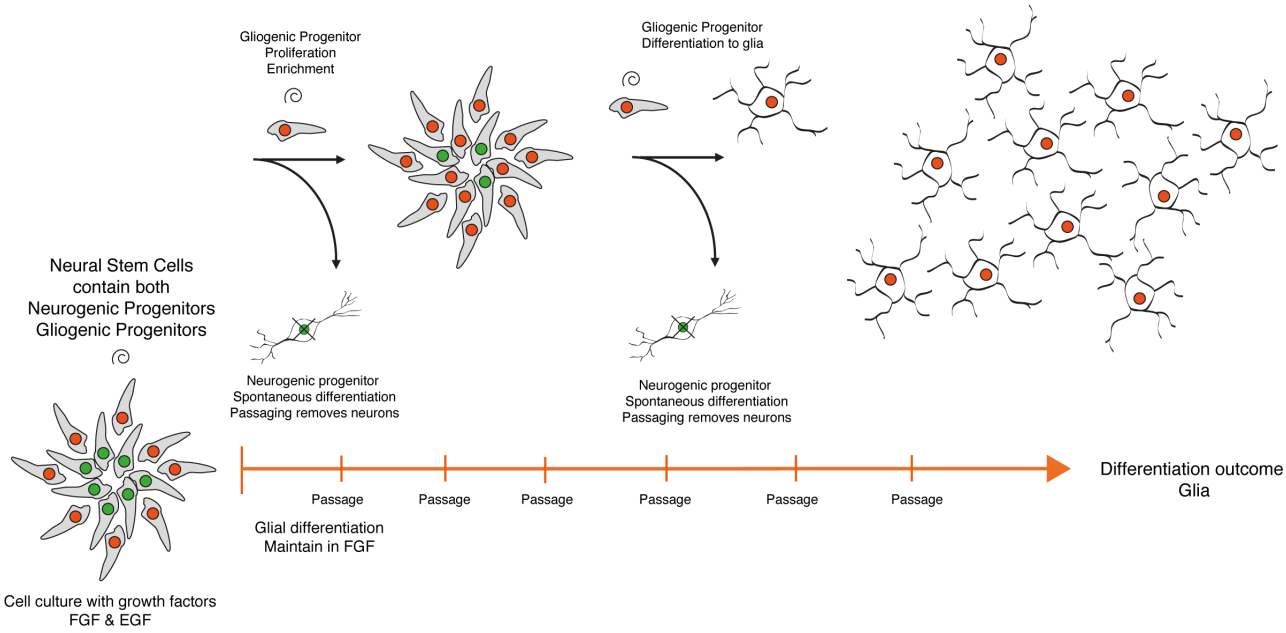


Figure 6. Illustration of neural stem cell model system for gliogenesis (adapted from Fig 5 Paper 2, Lam et al 2019).

2 NRNX1 EXPRESSION AND CELL ADHESION DURING EARLY MAMMALIAN DEVELOPMENT

Investigating the presence of Neurexin-1 (*NRXN1*) expression in cell model systems, sources of human fetal development, and during mammalian model system development, *NRXN1* can be found expressed across different neural modeling time-points and early stages of fetal neural development. Separate observations summarized under coming section show pieces of clues that point towards *NRXN1* mode of action. All in all these observations communicate *NRXN1* expression being required and needed to form proper structural neural stem cell niches to enhance integrity for cell-to-cell signaling and establishment of cell identity.

2.1 NEUREXIN-1 PROTEIN STRUCTURE

NRXN1 has mainly been studied specifically in synapse formation, synaptic plasticity during neural development and found to aid in calcium-dependent transmission in the central and peripheral nervous system (de Wit and Ghosh, 2016; Sudhof, 2017).

Neurexin-1 (*NRXN1)* can be transcribed from different promoters, has three predominant forms, the full length *NRXN1-alpha* (*NRXN1-a*) and mid-sized *NRXN1-beta* (*NRXN1-b*) short *NRXN1-gamma* (*NRXN1-g)* (Figure 7). *NRXN1-a* and *NRXN1-b* contain multiple slice sites that can be alternatively spliced and specifically *NRXN1-a* can be combined into thousands of isoforms (Jenkins et al., 2016; Sudhof, 2017; Ullrich et al., 1995). *NRXN1-a* isoforms preferentially bind to various trans-synaptic partners, including neuroligins, cerebellins, neurexophilins, and leucine-rich repeat transmembrane proteins each of which serve different synaptic functions (de Wit and Ghosh, 2016; Huang et al., 2017; Sudhof, 2017).

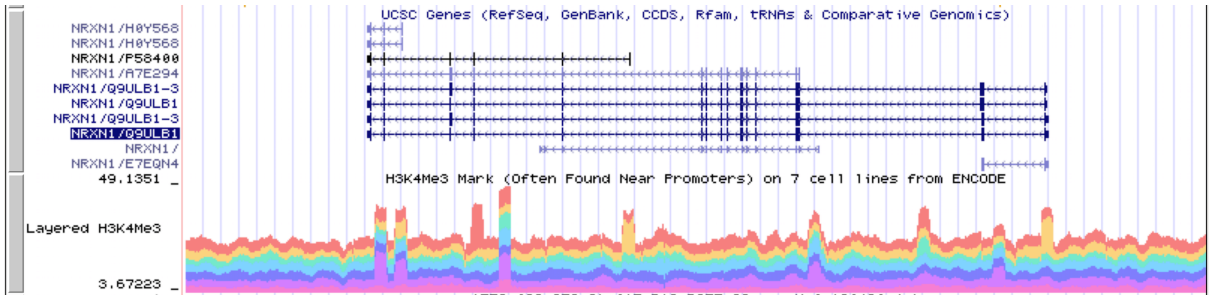


Figure 7. UCSC ENCODE track, *NRXN1* gene region with promoter marker H3K4Me3 shows presence (ridges) of promoter regions mapped across length of gene and different cell lines.

NRXN1-a protein contain six extracellular laminin–neurexin–sex-hormone-binding globulin (LNS) domains that confer adhesive properties, interspersed with three epidermal growth factor (EGF)-like domains and contain a carbohydrate-attachment sequence (CHB) near the transmembrane regiom. The NRXN1-a extracellular domain contains up to six splice sites (SS1–SS6).

NRXN1-b protein contain a short amino-terminal NRXN1-b-specific leader sequence, a single LNS domain and a carbohydrate-attachment sequence (CHB). Their extracellular domain contains two splice sites (SS4 and SS5). The carboxyl terminus of all neurexins contains a PDZ (postsynaptic density 95, Discs large 1 and zonula occludens 1)-binding domain (Figure 8).

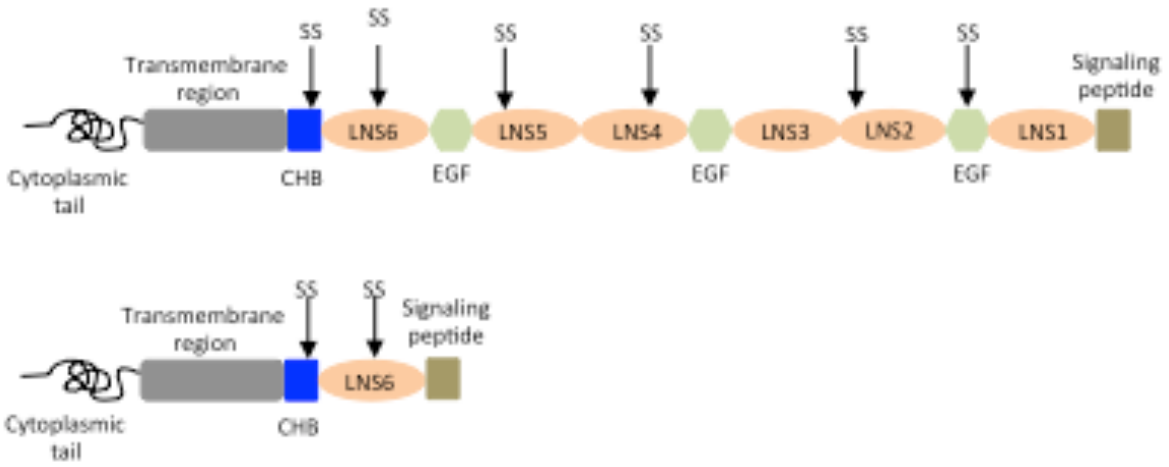


Figure 8. Schematic overview of NRNX1-a (upper) and NRXN1-b (lower) protein structure.

2.2 NRXN1 IN HUMAN CELL MODELS AND FETAL DEVELOPMENT

Probing for *NRXN1* during neural induction stages in iPS cells (Shang et al., 2018), *NRXN1* expression is observed peaking during neural rosette stages. As neural rosettes approximate the structure of developing neural tube, it is reasonable to propose that modes of *NRXN1* cell adhesion action is in effect to establish structural cell-to-cell connections to facilitate the required environment to maintain optimal neural stem cell environment (Figure 9).

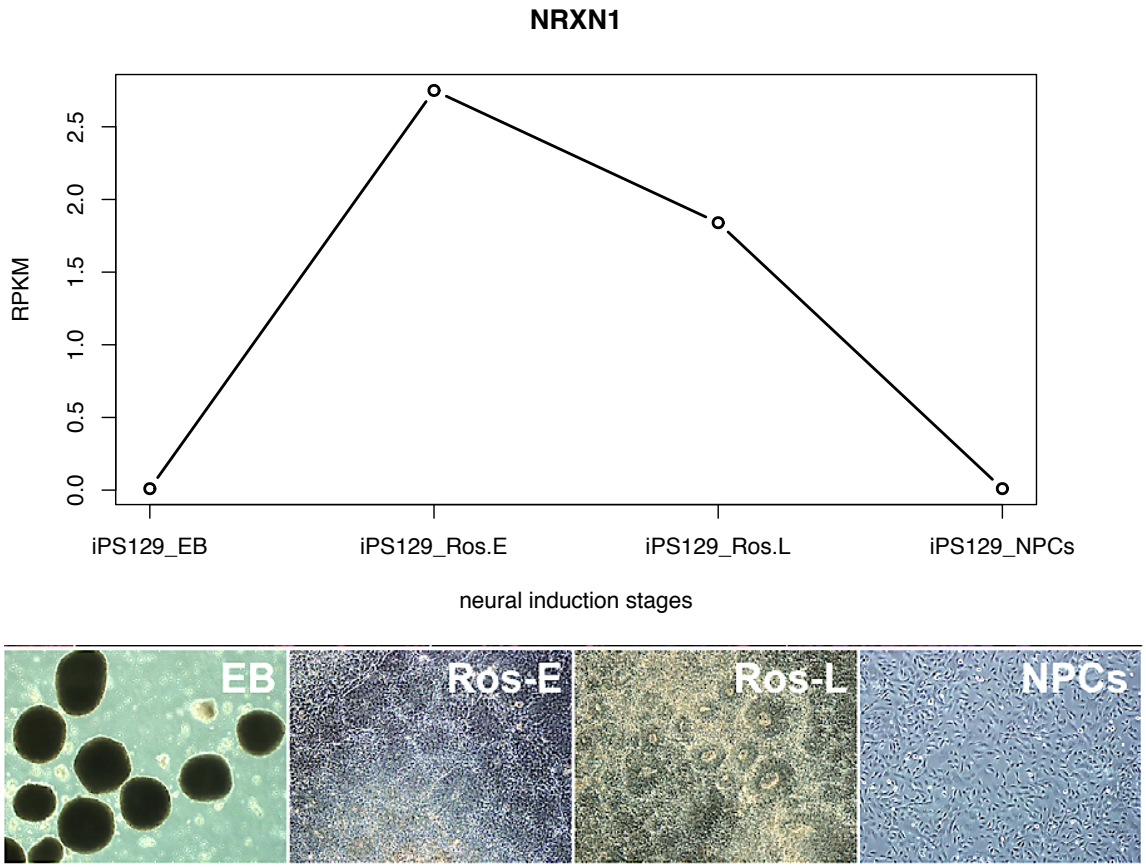
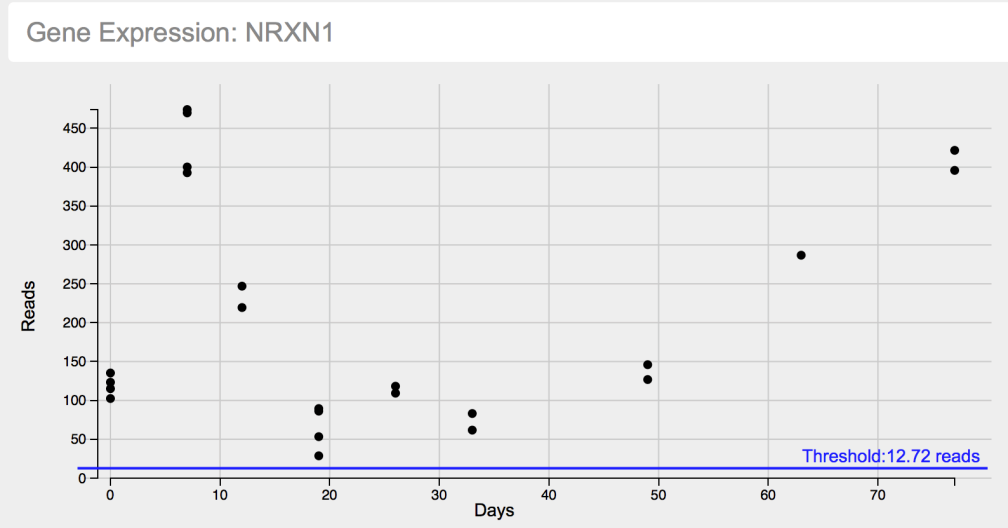


Figure 9. Human iPS neural induction, (Upper) Bulk RNA-seq *NRXN1* expression from embroid body (EB), early rosette (Ros-E), late Rosette (Ros-L) and neural progenitor cells (NPCs). (Lower) Brightfield images over neural induction (adapted from Fig 1b. (Shang et al., 2018))

Screening for *NRXN1* expression in human ES cells induced towards ectoderm and further differentiated to neurons (van de Leemput et al., 2014). Observation of two peaks of *NRXN1* expression can be seen (Figure 8), first peak of *NRXN1* expression in time-point encompassing establishment of neural progenitor cells suggests cell adhesive properties are in action, as neural stem cells require close cell-to-cell proximity to maintain neural rosette structure and stemness niche through signaling pathways like NOTCH signaling (revisit Figure 3), second peak of *NRXN1* expression coincides with time-points for emergence of neurons and synaptic connection maturation (Figure 10).



Neural induction Neural differentiation

Figure 10. Cortecon human ES cell neural induction/differentiation. *NRXN1* expression over neural induction and neural differentiation shows two ascending waves of expression. Screenshot from http://cortecon.neuralsci.org (van de Leemput et al., 2014)

In single cell study of human pre-implantation human embryo (Petropoulos et al., 2016), *NRXN1* can be observed expressed at the embryonic day 3 stage corresponding to cell developmental stage between illustrated 8-cell to compacted 16-cell morula (Figure 11). The presence of *NRXN1* at this early stage of development suggests essential cell adhesion function to establish close cell-to-cell environment to maintain the proper environment for cells to thrive during this delicate developmental stage.

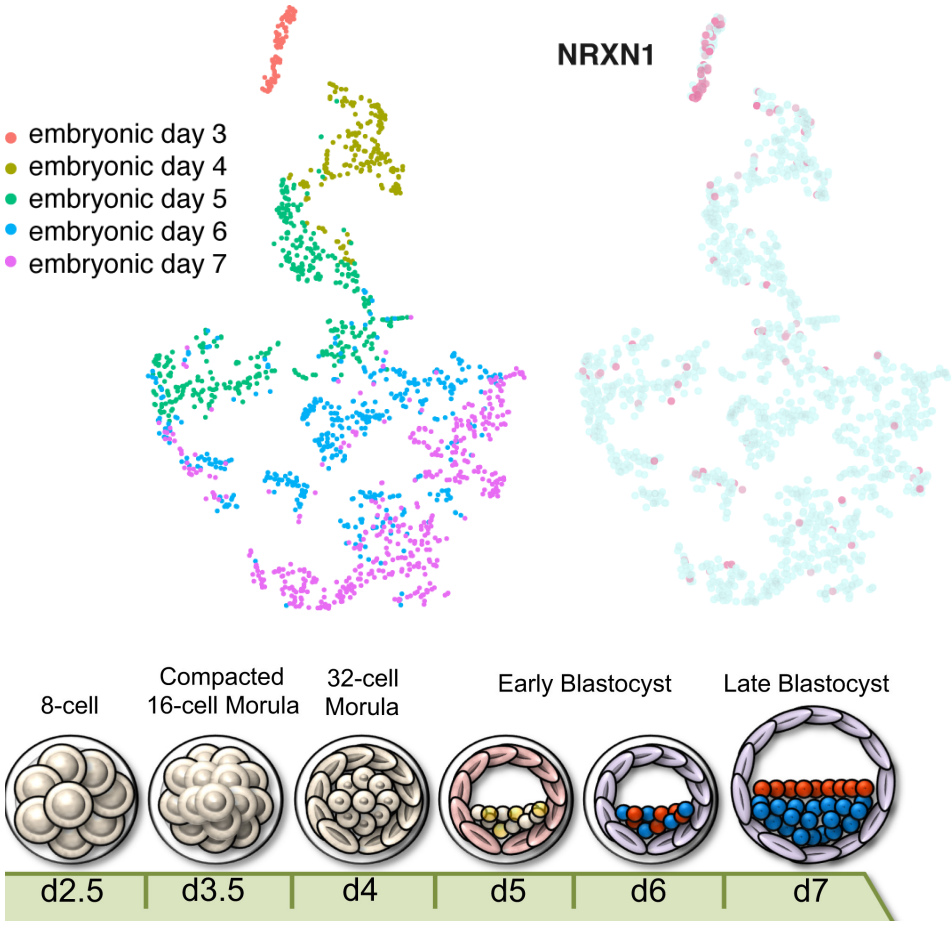


Figure 11. tSNE-plots of time-points in human pre-implantation embryo, *NRXN1* highly expressed at embryonic day 3, bottom illustration adapted from Fig 1 (Sozen et al., 2014).

In single cell study of developmental stages of the human ventral midbrain (La Manno et al., 2016), *NRXN1* expression can be observed in radial glia-like cells, oligodendrocyte progenitors and various types of neurons (Figure 12). Observation of *NRXN1* expression in radial glia-like cells suggests presence of *NRXN1* in neural progenitor type cells in the developing human ventral midbrain and mode of action for cell adhesion.

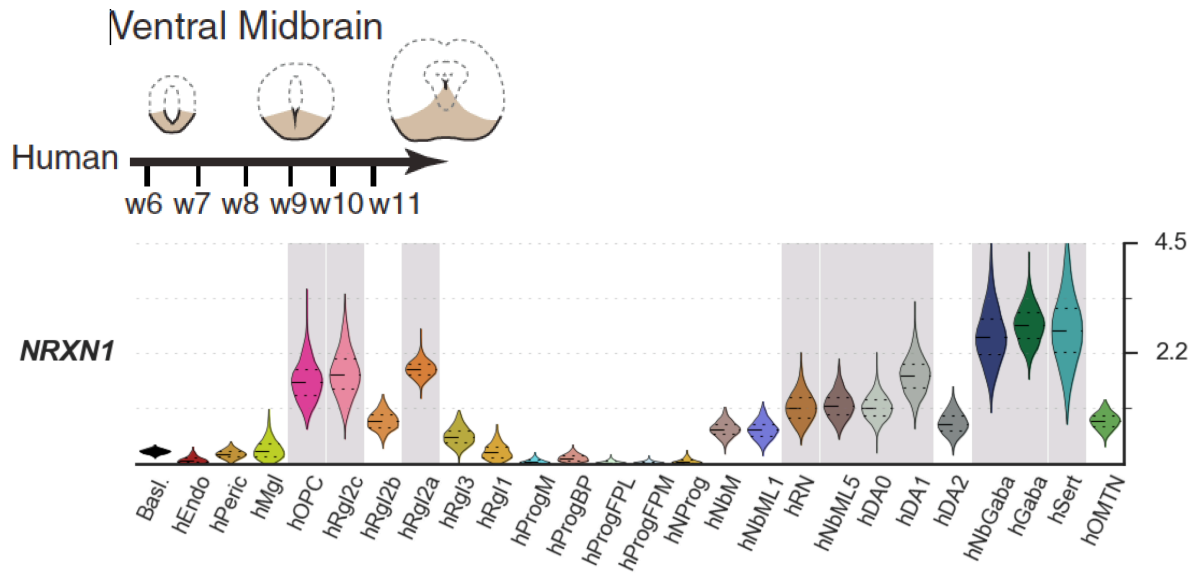


Figure 12. (Upper) Sampling scheme of fetal tissue from ventral midbrain across week 6 to 11. (lower) *NRXN1* expressed across cell types in the developing human ventral midbrain, enrichment observed in radial glia-like cells (Rgls), graph generated from http://linnarssonlab.org/ventralmidbrain

Supporting scenario in human developing cerebral cortex at 8 weeks post-conception (PCW) display NRXN1 enriched in cells of the outermost apical surface layer of the developing ventricular zone (VZ) (Figure 13) (Lauren F. Harkin and Clowry, 2016).

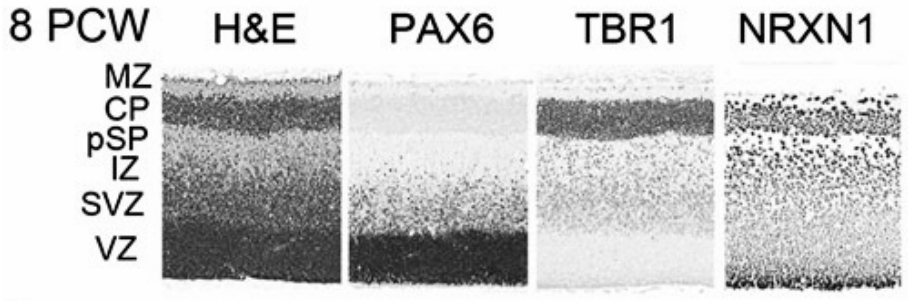


Figure 13. NRXN1 immuno-reactivity with other cell markers across the cortical layers at 8 PCW shows NRXN1 enriched on apical ventricular surface, PAX6 reveals radial glial progenitor cells and TBR1 reveals postmitotic neurons. H&E hemotoxylin and eosinstained, VZ Ventricular zone, SVZ subventricular zone, IZ Intermediate zone, pSP presubplate, CP Cortical plate, MZ Marginal zone (adapted from Fig 3A (Lauren F. Harkin and Clowry, 2016))

2.3 NRXN1 IN EARLY MOUSE DEVELOPMENT

Single cell RNA-seq study of E6.5 and spatial transcriptomic bulk RNA-seq study of E7.0 gastrulating mouse embryo (Cheng et al., 2019; Peng et al., 2016; Peng et al., 2019), display *Nrxn1* preferentially expressed along the posterior late mid-streak stage area corresponding to the developing ectoderm and neural tube (Figure 14 and 15). Observation of *Nrxn1* in mouse embryo suggests cell adhesive properties in action to support emergence and maintenance of the neural tube and neural stem cells lining the central cavity.

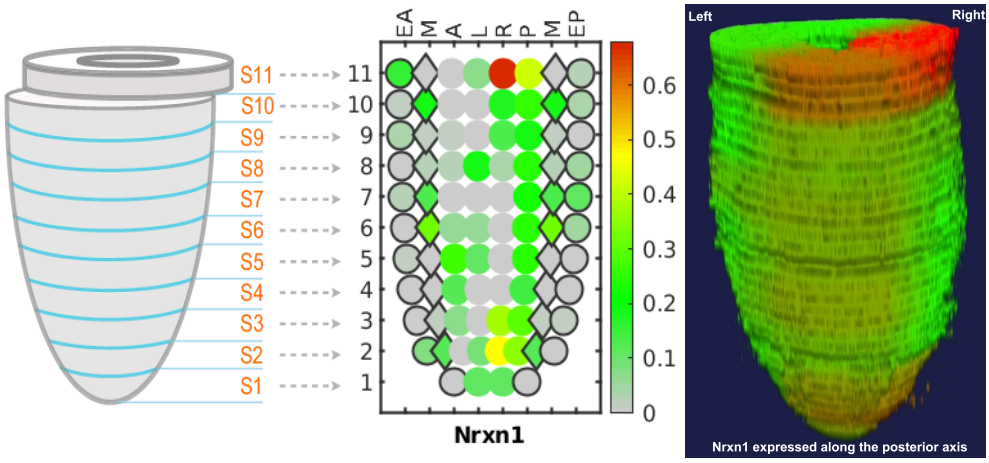


Figure 14. Geo-seq spatial transcriptomics of *Nrxn1* expression along posterior. EA=anterior endoderm, M=whole mesoderm, A=anterior, L=left lateral, R=right lateral, P=posterior, EP=posterior endoderm (left figure adapted from (Peng et al., 2016) middle and right figure generated from http://egastrulation.sibcb.ac.cn/E70/aPatternSearchByGene/).

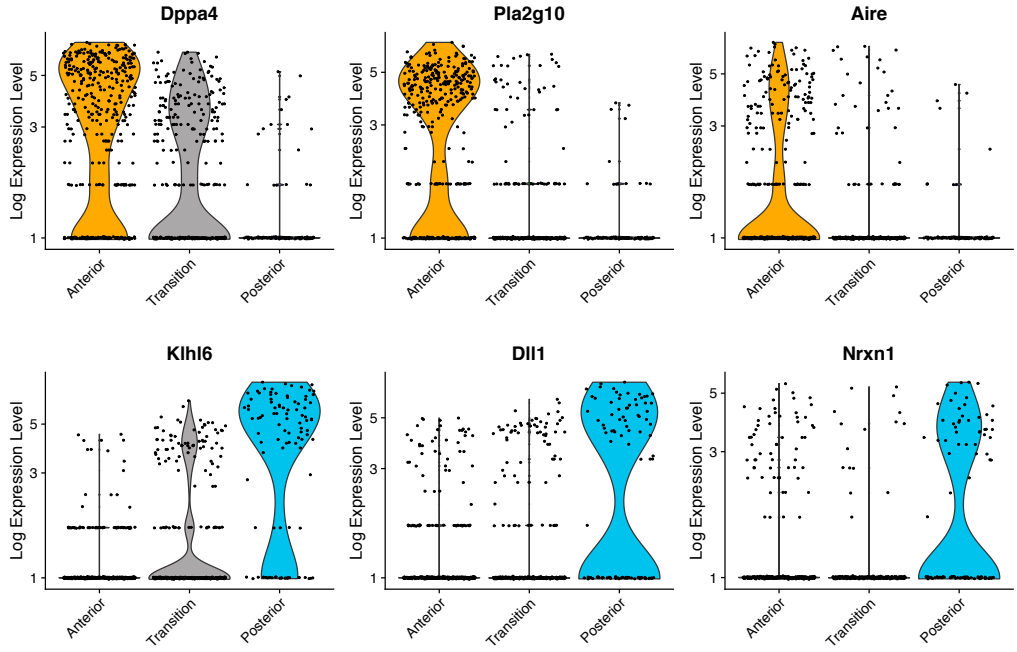


Figure 15. Single cell RNA-seq of E6.5 mouse embryo, display *Nrxn1* preferentially expressed in posterior identity cells over anterior identity cells, reuse of data from (Cheng et al., 2019).

2.4 NRXN1 AND CDH8 IN EARLY FROG DEVELOPMENT

Looking into *Nrxn1* expression presence in studies spanning through frog (Xenopus) development and observations of *Nrxn1* expression can be found in various reference sources. Single cells study of early Xenopus development (James A Briggs, 2018), display *Nrxn1* and *Cdh8* expressed in cells comprising the blastula, organizer, developing ectoderm and marginal zone across a range of time-point samples. Co-expression of *Nrxn1* and *Cdh8* in the identified regions suggest a co-localized expression pattern in cells of the developing tissue structures (Figure 16). Similar observation in bulk RNA-seq of Xenopus developmental stages display *Nrxn1* presence during early development (Owens, 2016) (Figure 17) and in situ hybridization study of *Nrxn1* presence in early oocytes (Zeng et al., 2006) (Figure 18).

2.5 NRXN1 EVOLUTIONARY CONSERVED ACROSS SPECIES

The combined observations suggest NRXN1 cell adhesive mode of action is evolutionarily conserved across species and is important for the establishment and formation of ectodermal structures needed to support the formation of the central nervous system during human, mouse and frog development.

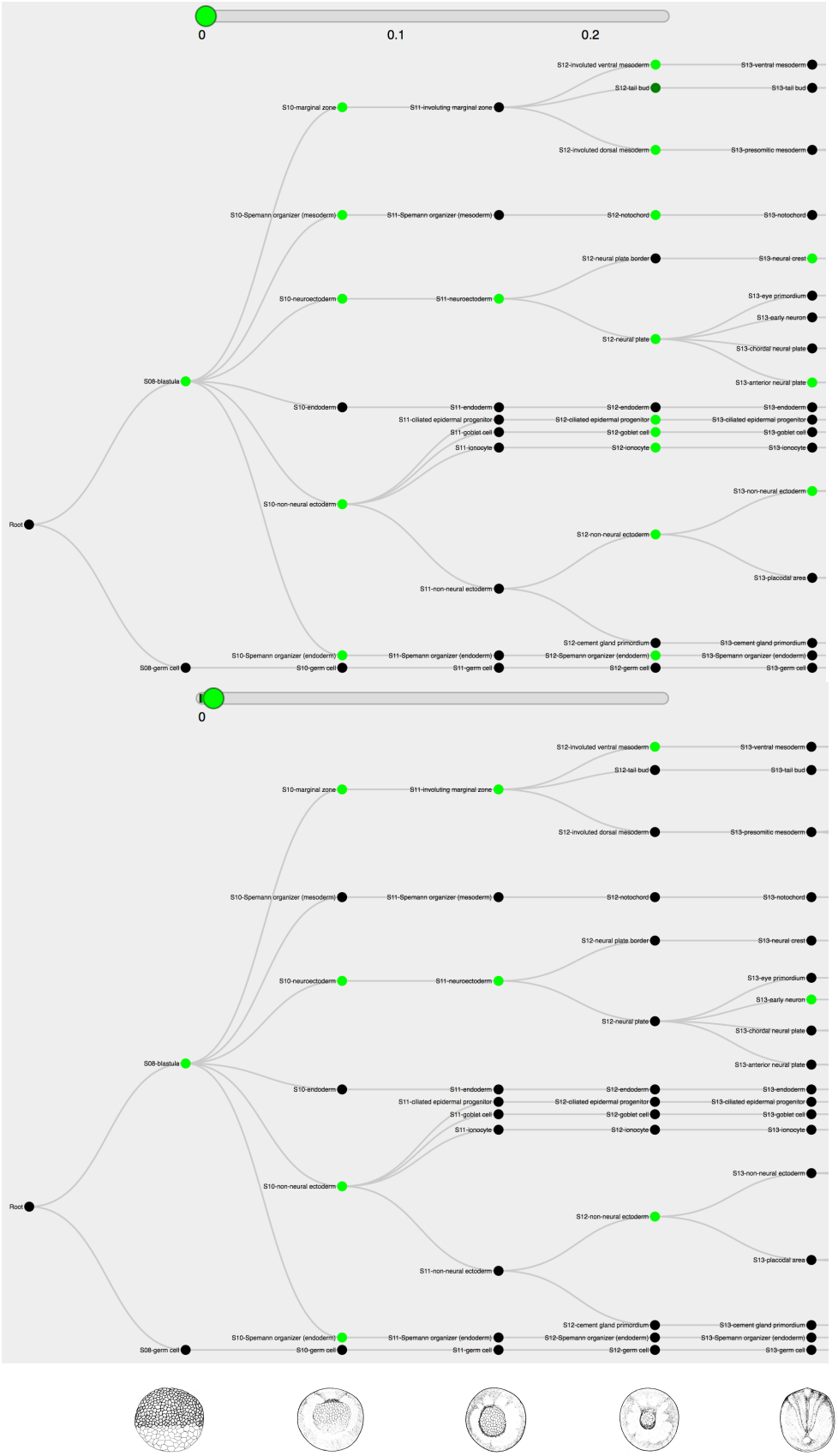


Figure 16. Single cell study in Xenupus. *Nrxn1-like 1* (upper) and *Cdh8* (lower) expressed (green) in blastula, marginal zone, organizer, ectoderm across in early Xenopus developmental stage 8, stage 10, stage 11, stage 12, stage 13. (tree view from https://kleintools.hms.harvard.edu/tools/currentDatasetsList_xenopus_v2.html) (James A Briggs, 2018)

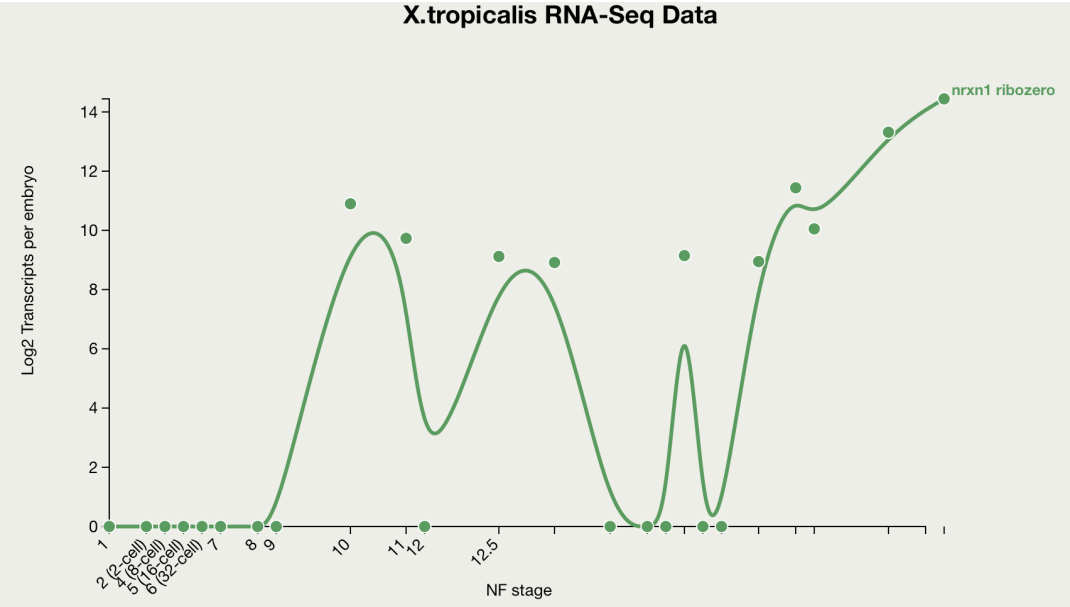


Figure 17. Bulk RNA-seq of Xenopus developmental stage show *Nrxn1* expressed during early stages (9.5-12.5). (Owens, 2016) Graph from gene expression query on http://www.xenbase.org.

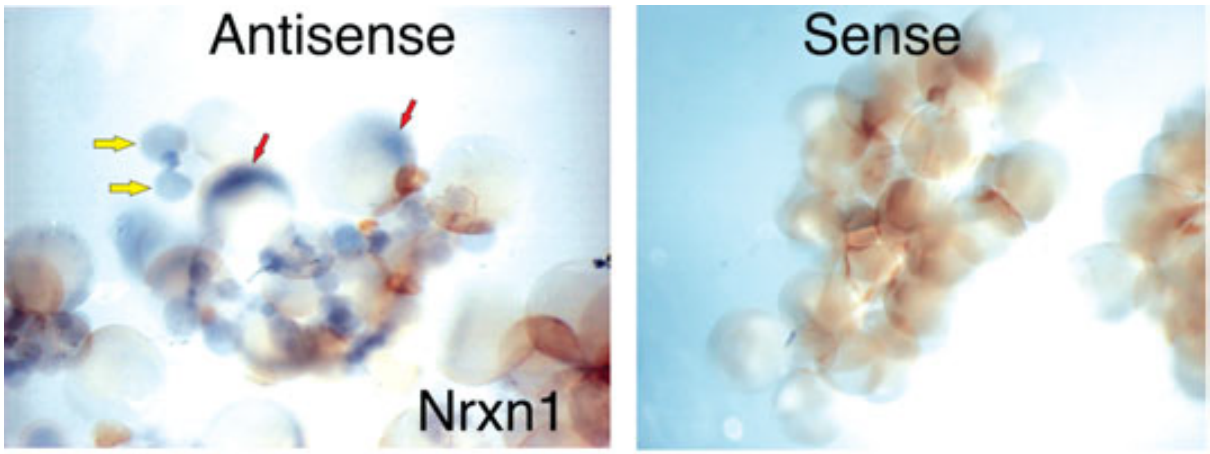


Figure 18. An *Nrxn1* specific probe identifies transcripts equally distributed through small, early stage oocytes (yellow arrows). At later stages, the *Nrxn1-α* transcript is localized to a crescent within the large oocyte (red arrows). Transcripts were not detected in cells other than oocytes. (adapted from Fig 3 (Zeng et al., 2006))

2.6 CDH2, CDH8 AND CDH6 IN EARLY CHICK DEVELOPMENT

Study of cadherins expressed across various structures of the developing chick central nervous system display *Cdh2*, *Cdh8* and *Cdh6* restricted patterns of expression (Lin et al., 2014). At the early stage of chick embryo development at E4 corresponding to development in human week 5 embryo, inclusive and exclusive patterns of *Cdh2*, *Cdh8* and *Cdh6* can be observed in lumbar section of chick spinal cord (Figure 19). At E4 the neural tube display high presence of *Cdh2*, with the neuroepithelial layer displaying exclusive presence of *Cdh8* and roof plate/floor plate/basal plate displaying presence of *Cdh6* (Figure 20).

The expression across the various regions suggest exclusive expression patterns for *Cdh8* coinciding with for neurogenic progenitors of the central lining of the neural tube, *Cdh6* in roof plate/floor plate for emerging neural crest/glia cells and inclusive expression of *Cdh2* for all cells of the developing ectoderm.

In line with this observation in total RNA-seq data on AF22 neural stem cells, *CDH2* is observed expressed across all neural stem cells, *NRXN1* and *CDH8* expression coincides with enrichment in neurogenic progenitors and *CDH6* is enriched in gliogenic progenitors (Figure 21).

These combined observations across species suggest ties between *NRXN1* and *CDH8* in performing cell adhesive modes of action to support in the emerging developing neuroectoderm.

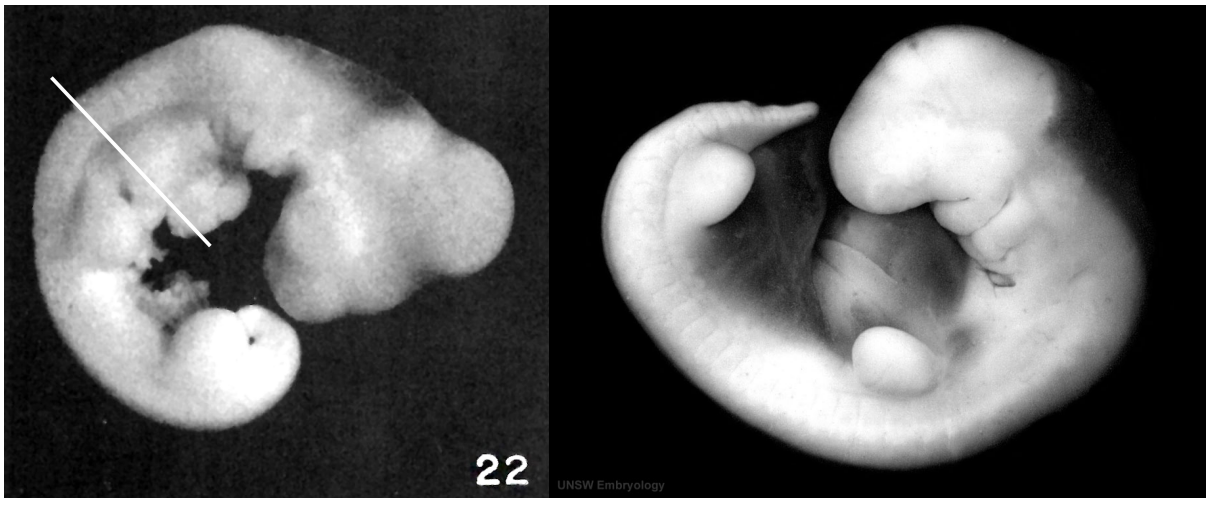


Figure 19. Embryonic development chicken and human (left) chick Hamburger Hill stage 22 ~E4, stroke approximates position of lumbar section, (right) human Carnegie stage 14 ~week 5 display similarity between species (images from https://embryology.med.unsw.edu.au).

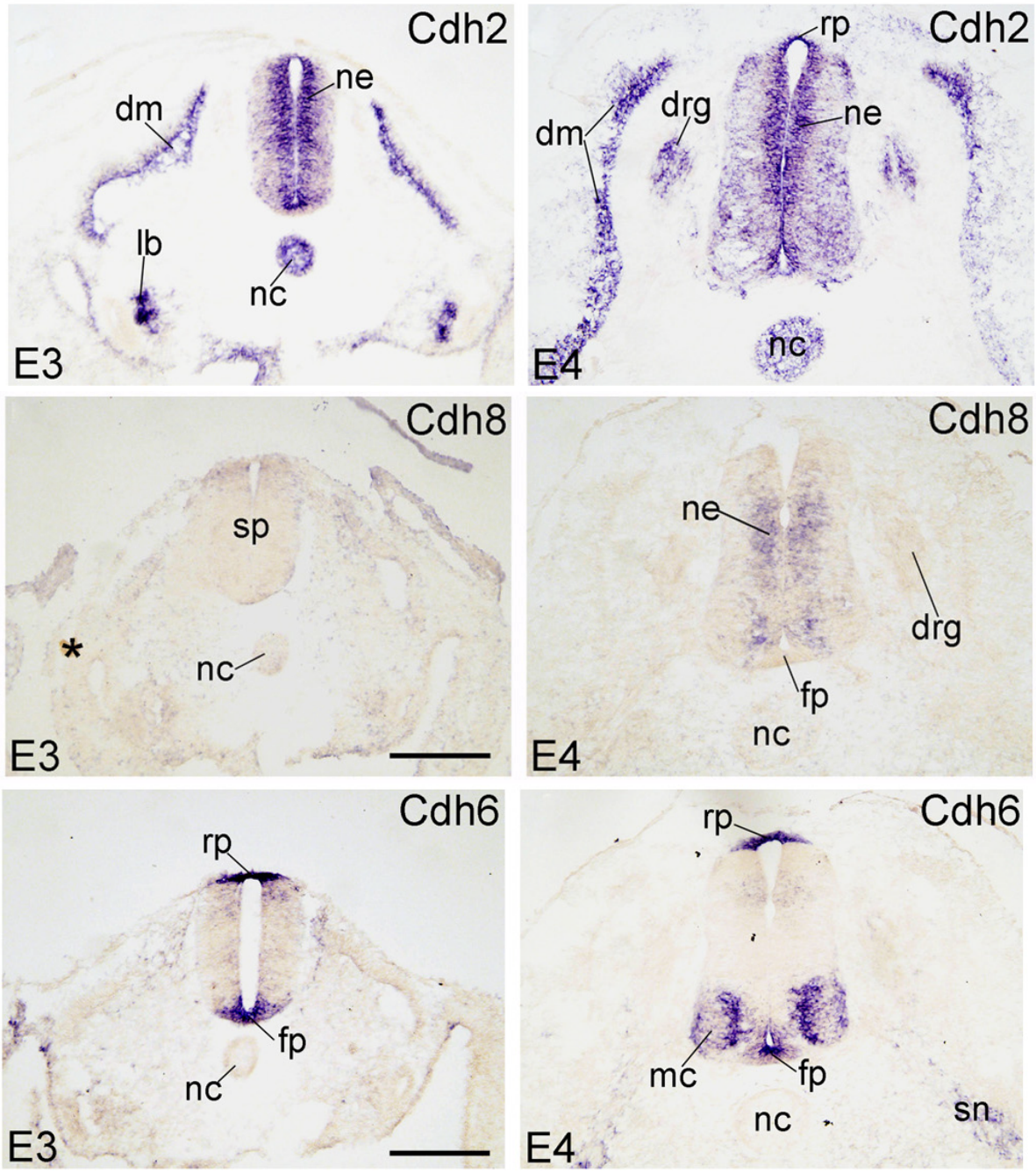


Figure 20. Chick embryo E4 lumbar section, *In situ* hybridization result *Cdh2*, *Cdh8* and *Cdh8* show overlapping expression pattern (*Cdh2*) and exclusive expression pattern (*Cdh8, Cdh6*) across region of neurpepithelia layer of neural tube. Abbreviations: DM, dermomyotome; DRG, Dorsal root ganglion; FP, floor plate; NC, notochord; NE, neuroepithelial layer; RP, roof plate; SP, spinal cord.

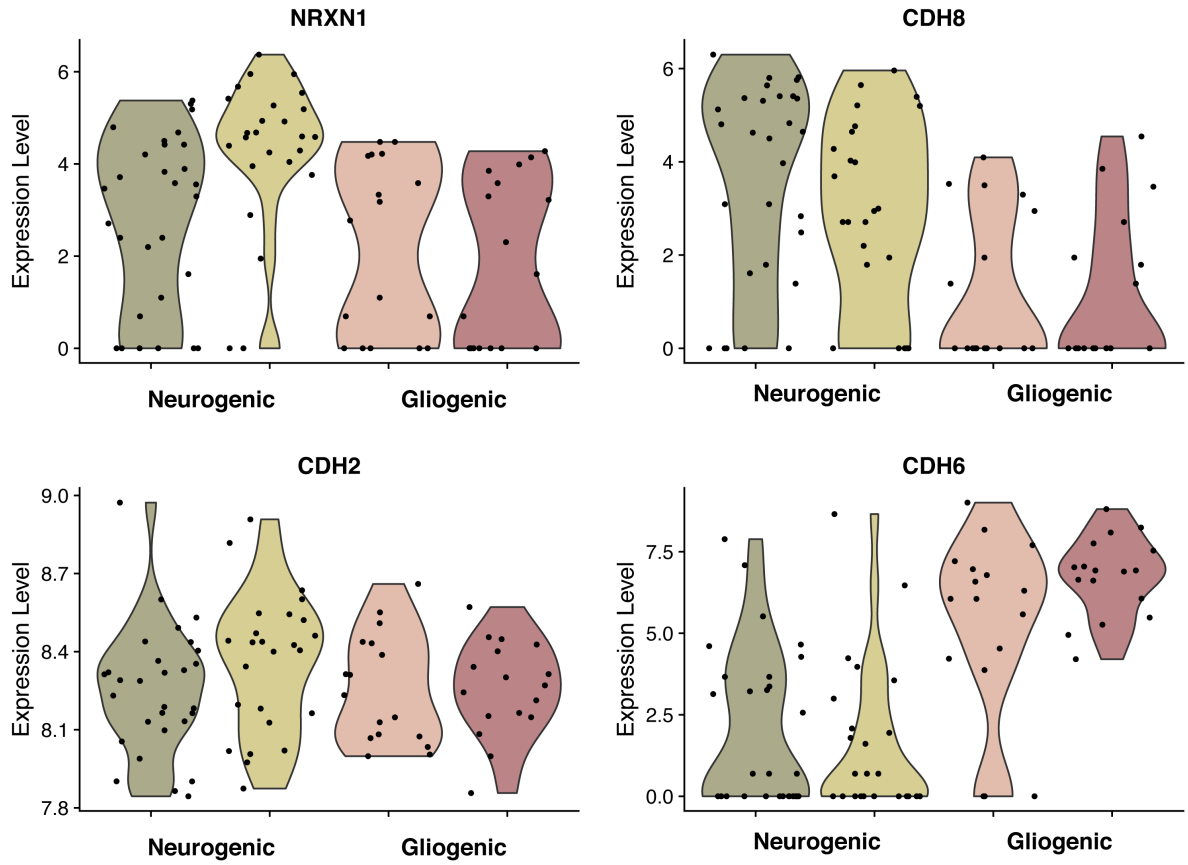


Figure 21. Single cell total RNA RamDA-seq data for AF22 neural stem cells. Violin plots for enriched genes selected for neurogenic progenitors and gliogenic progenitors (subset of data for Figure 1E used in Paper II, Lam et al 2019).

3 REVISITING DOUBLECORTIN MUTATION STUDY

As presented in Paper 3, mutations in *DCX* cause neurite outgrowth defects in neuronblasts, neural stem cell inability to differentiate and structurally stunted axons in neurons. Including the added perspective of presence of neurogenic progenitors and gliogenic progenitors inherently present in the neural stem cell culture (Paper 2), probing for lineage marker genes increases insight into neural stem cell lines used in *DCX* mutation study. Reviewing microarray data at neural stem cell proliferation stage for neurogenic (*CDH8*, *NRXN1*, *ONECUT1*) and gliogenic (*FABP7*, *HES5*, *SLC1A3*) genes provides an added perspective to increase the understanding of the cell model system used to model neurodevelopmental disease (Figure 22).

In addition to *DCX* mutations causing severe disease phenotype on neural stem cell differentiation capacity, an increased presence of gliogenic progenitors shifts the neuron differentiation potential of neural stem cells to yield fewer neurons and more glial cells. Glial cells do not readily form neurites or axonal projections and this might have impacted the perception and interpretation of the disease burden phenotype presented in Paper 3.

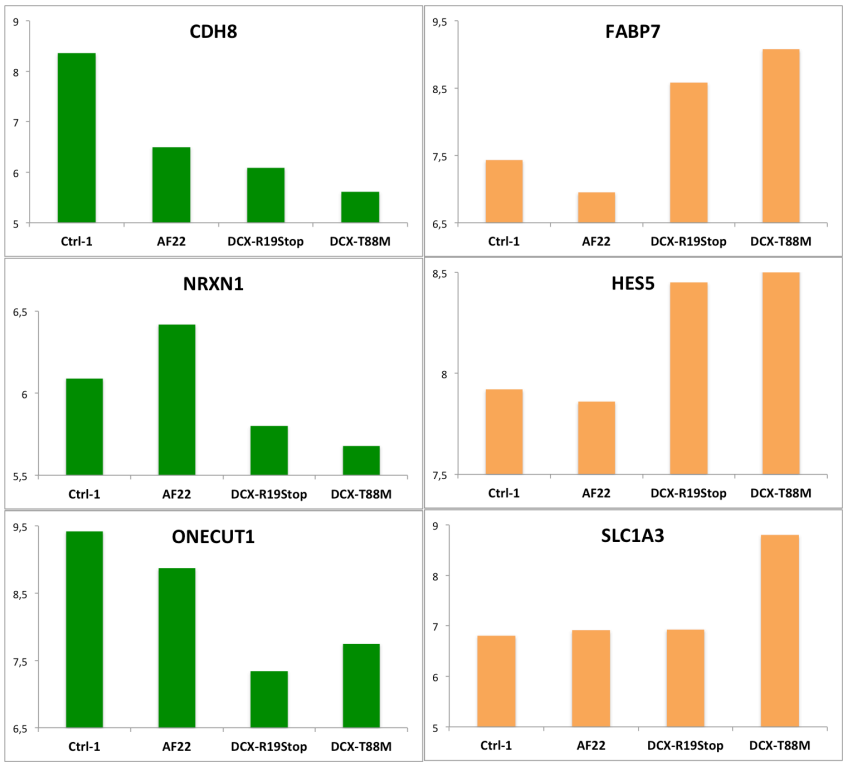


Figure 22. Normalized intensity log2 fold-change values for microarray at neural stem cell stage. Genes corresponding to neurogenic(green) and gliogenic(orange) progenitor identity display preference for neurogenic progenitors in control neural stem cell lines and gliogenic progenitors in *DCX* mutation carrying neural stem cell lines.

4 SINGLE CELL RNA SEQ

4.1 SINGLE CELL ISOLATION PROTOCOLS

The first step in scRNA-seq is to isolate single cells from tissue samples while keeping gene expression pattern as intact as possible. Techniques used for this purpose include manual cell picking, fluorescent assisted cell sorting (FACS), magnetic activated cell sorting (MACS), laser capture microscopy (LCM), microfluidics and droplet-based cell capture (Basu et al., 2010; Brennecke et al., 2013; Emmert-Buck et al., 1996; Macosko et al., 2015; Miltenyi et al., 1990).

Low throughput techniques like manual cell picking by micromanipulators (or mouth pipetting) and LCM requires a high levels of hands-on work since single cells need to be visually identified and captured. LCM presents an advantage in that cells can be studied within the microenvironment with preserved spatial location.

High throughput techniques like FACS and MACS presents a fast and efficient way for sorting out single cells. FACS can be used to sort out cells by size, granularity and surface markers. Fluorescence-tagged monoclonal antibodies aimed at cell surface markers enable for efficient isolation of cell subpopulations within a heterogeneous cell sample. MACS can be used to sort out target cells by surface markers, the cells are isolated by biodegradable iron based nanobeads bound with specific cell surface antibodies.

The Fluidigm C1™ system with 96-well chip has built-in microfluidics architecture for single cell capture. Single cell suspension is flowed through the chip in microfluidic channels and cells are captured in sites that can be inspected optically for single cell capture, cell size and cell integrity (Figure 23).

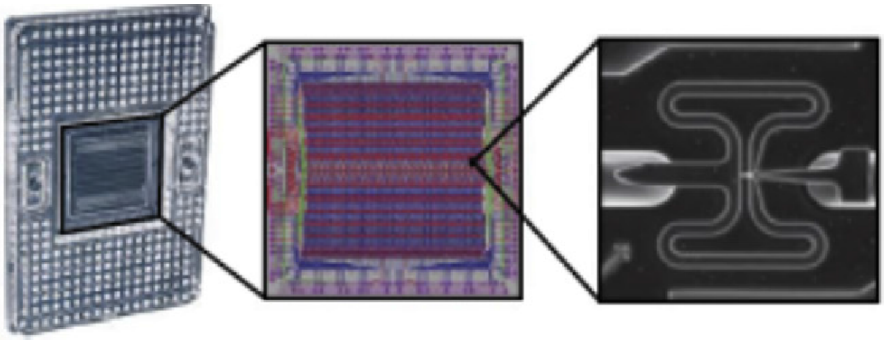


Figure 23. Fluidigm C1 microfluidics cell capture device, showcasing single cell capture site (from (Ouriel Caen, 2017))

Droplet-based single cell capture is an emerging powerful high-throughput microfluidics technique. Microfluidic device is used for co-flowing two aqueous solutions (cell suspension and bead suspension with lysis buffer) across an oil channel to form nano-liter sized droplets. The process can generate more than 100.000 droplets per minute and millions of droplets per hour, of which thousands will contain both a single cell and a bead. Single cell transcriptomes are obtained from the subset of droplets that contain both a cell and a bead (Macosko et al., 2015) (Figure 24).

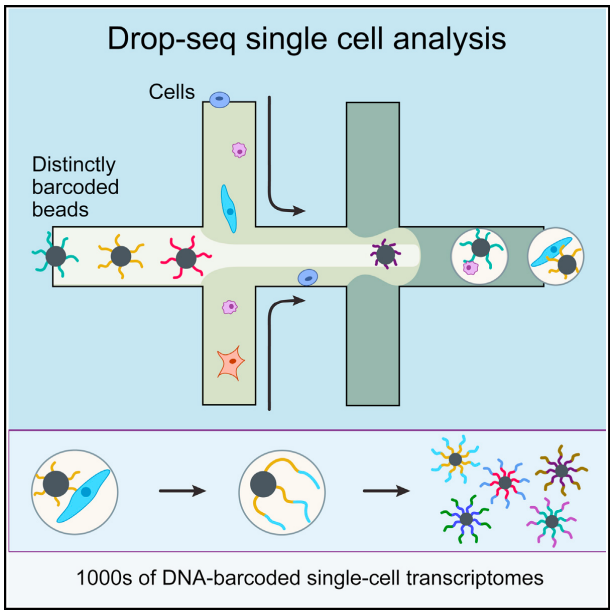


Figure 24. Droplet-based single cell capture, single cell suspension flows into single bead suspension to for micro-droplet forming reaction chamber (Graphical abstract from Macosko et al 2015).

4.2 METHODS OF SEQUENCING

Following single cell capture, cell lysis and reverse transcription of mRNA (≈0.1 pico gram/single cell) to complementary DNA (cDNA), the cDNA needs to be amplified (Liu et al., 2014)

4.2.1 Single-cell Tagged Reverse Transcription (STRT)-Seq

cDNA amplification method based on "template-switching" for generating cDNA libraries have proven efficient for single-cell application and is marketed as "Switching mechanism at the 5' end of the RNA transcript" (SMART) (Zhu et al., 2001). SMART uses the advantage of two intrinsic properties of the Moloney murine leukemia virus (MMLV) reverse transcriptase

(RT in any commonly used RT kits), first, the ability to add non-templated cytosines to the 3'-end to the cDNA first strand and, second, the capability to switch templates.

In the single-cell tagged reverse transcription (STRT)-seq protocol the SMART amplified full-length cDNA transcripts are immobilized on beads at the 5'-end, fragmented and adapter ligated at the free 3'-end. Then the fragmented cDNA is unbound from bead and ligated with second adapter to the 5'-end. The fragmented two-end adaptor ligated cDNA is sequenced and used for library building (Islam et al., 2012)

The SMART protocol introduce some amplification biases in their inherent process, STRT-seq has 5'-end amplification bias. Unique molecular identifiers (UMI) are short randomly generated DNA fragments used to leverage STRT-seq 5'-end bias. UMIs are labeled on the first strand cDNA during reverse transcription before cDNA amplification and allow for absolute transcript/molecule quantification (Islam et al., 2014) (Figure 25)

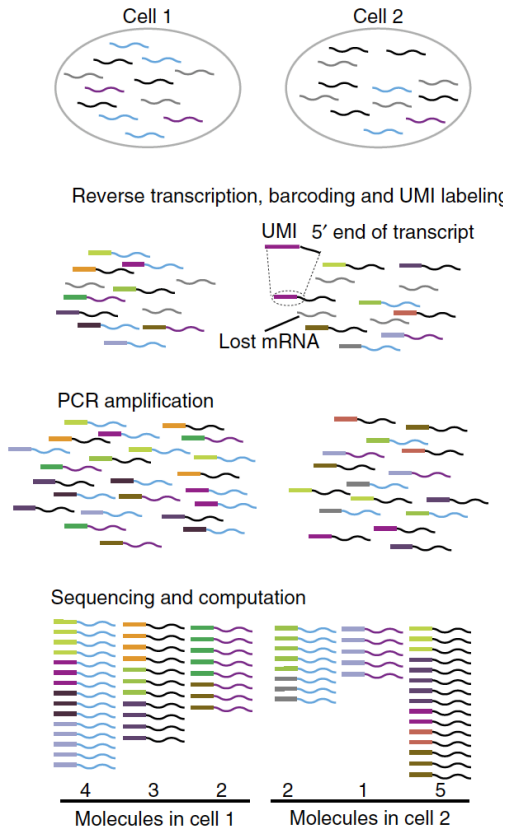


Figure 25. Molecule counting using UMIs, overview of tagging single mRNA molecules with UMIs. Two cells are shown containing mRNAs from different genes represented by distinct colors. UMIs are represented by colored boxes; untagged mRNA molecules were not reverse transcribed (right) (adapted from Figure 1A, Islam 2014).

4.2.2 Random Displacement Amplification (RamDA)-Seq

RamDA-seq is a total RNA sequencing method that can be used to reveal poly-A and non-poly-A RNA expression, RNA processing and enhancer activity. It shows high sensitivity to non-poly-A RNA and near-complete full-length transcript coverage for long transcripts exceeding 10kb.

The method combines reverse transcription (RT) with random displacement amplification (RamDA) and not-so-random primers (NSRs). RT-RamDA provides global cDNA amplification directly from RNA during RT, which benefits RT efficiency and decontaminates genomic DNA. NSRs enable random priming while preventing cDNA synthesis from rRNAs. (Hayashi et al., 2018) (Figure 26).

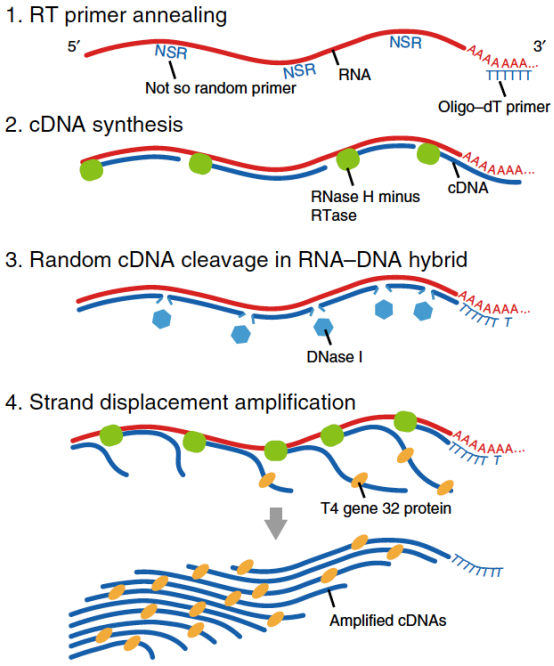


Figure 26. Schematic diagram of RT-RamDA.

1. RT primers (oligo-dT and not-so-random primers) anneal to a RNA template.

2. Complementary DNA (cDNA) is synthesized by the RNA-dependent DNA polymerase activity of RNase H minus reverse transcriptase (RTase).

3. Endonuclease (DNase I) selectively nicks the cDNA of the RNA:cDNA hybrid strand.

4. The 3′ cDNA strand is displaced by the strand displacement activity of RTase mediated by the T4 gene 32 protein (gp32), starting from the nick randomly introduced by DNase I. cDNA is amplified as a displaced strand and protected by gp32 from DNase I (adapted from Figure 1 (Hayashi et al., 2018)).

4.3 METHODS OF ANALYSIS FOR SINGLE CELL DATA

4.3.1 Number of single cells and sequencing depth

Two general considerations for single cell experiments needing to be addressed are, the minimum number of single cells needed and how deep the sequencing needs to be.

Ideally, the number of single-cells needed will depend on study aim as each study have a unique set of questions to be answered and the natural occurrence of cells of interest will limit the abundance of cells that can be isolated. However, the actual real numbers of cells that can be isolated depend on the used single-cell isolation technique.

Taking into account that deeper sequencing depth per single cell provide more data to classify unknown cell types within seemingly homogeneous cell populations, one study reported that more than 90% of all genes detected at 30 million reads were already detected at a sequencing depth of 2 million reads (Figure 27) (Wu et al., 2014) and another study reported shallow single-cell sequencing ~50.000 reads per cell being sufficient for unbiased cell-type classification and biomarker identification (Pollen et al., 2014)

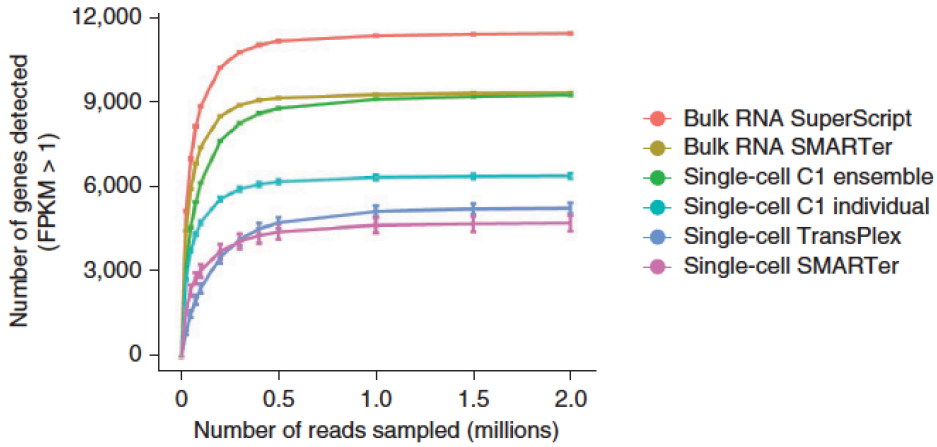


Figure 27. Saturation curves for different sample preparation methods.

As isolation protocols of single-cells result in cells captured in micro-well plates, chambers in microfluidics devices or droplets, quality control of single-cells before sequencing can be estimated by visual inspection, automated imaging and viability dyes.

Parallel and massive processing will generate isolation of no cells, single cell or multiple cells and the isolates of no cells and multiple cells need to be excluded from the experiment. Also single-cells successfully captured may be dying, stressed or broken at time of capture and will yield misleading information. The low quality of these cells can be estimated after sequencing.

4.3.2 Quality Control, RNA-seq data and single cell quality evaluation

FastQC is a quality control tool that can perform checks on raw sequence data (https://www.bioinformatics.babraham.ac.uk/projects/fastqc), and has a user-friendly graphical interface and can include metrics such as: per base sequence content, per base CQ content and sequence length distribution. Comparing quality metrics across all cells included in the experiment will help in exposing low quality outliers. Filtering thresholds for identifying low quality cells can also be used based on the number of mapped reads and/or the proportion of detected genes over mapped reads.

One study reported a set of biological and technical measures for detecting low quality single cells. They proposed broken or dying cells displayed gene expression profiles enriched in "mitochondria encoded genes" and "mitochondria localized proteins" while simultaneously being impoverished in genes related to gene ontology terms such as "cytoplasm", "metabolism" and "membrane". The rationale being that broken cells would be emptied of their cytoplasmic mRNA content while mRNA enclosed in mitochondria would still be intact at time of lysis and cDNA amplification. Technical measures for identifying low quality cells included, that broken cells yield a low number of total sequenced reads resulting in lower number of detected genes. In line with this, when multiple cells were captured the proportion of duplicated reads would increase compared to single cells captured (Ilicic et al., 2016)

Another technical measure practice for filtering out high quality cells with a set interval for detected genes between a low cut-off and a high cut-off point. Wherein, low-cut-off genes in a single cell are likely dead or broken and high cut-off genes presents more than one cell captured. How to apply the cut-off interval depends largely on acquired data set, evaluation of reasonable gene expression profile for cell types being investigated, overall transcript yield from chosen sequencing method and overview of gene enrichment detected for determining unique cell types (own observations).

4.3.3 BACKSPIN – Detecting highly variable genes

For finding genes that define cell types, BackSPIN is a powerful bi-clustering algorithm that can identify highly variable genes enriched in cells. By investigating the variable and cell enriched genes, gene expression profiles corresponding to unique cell types can be discovered (Zeisel et al., 2015)

4.3.4 Monocle – Pseudotemporal ordering and trajectory projection

Monocle is an R package with useful features to reconstruct pseudotime on single cell data and plot trajectories to line up cells along trajectory paths of dynamic processes (Qiu et al., 2017). Pseudotemporal ordering is a highly informative method used on single cells to model dynamic processes such as development, differentiation or disease progression. The general idea is that a population of single cells will have a high probability to contain cells at different stages during an ongoing process. In profiling gene expression of the cells, it is possible to reconstruct an order in which changes over gene expression will set cells that are different apart and at the same time bring cells that are similar together. Single cells placed in pseudotemporal order will display cells with similar gene expression aggregated at the beginning of pseudotime and cells having more different gene expression profiles aggregated at the end of pseudotime. Additionally, cells that have differences in gene expression profiles that are intermediate between the cells at beginning and cells at end will be ordered in between (Figure 28)

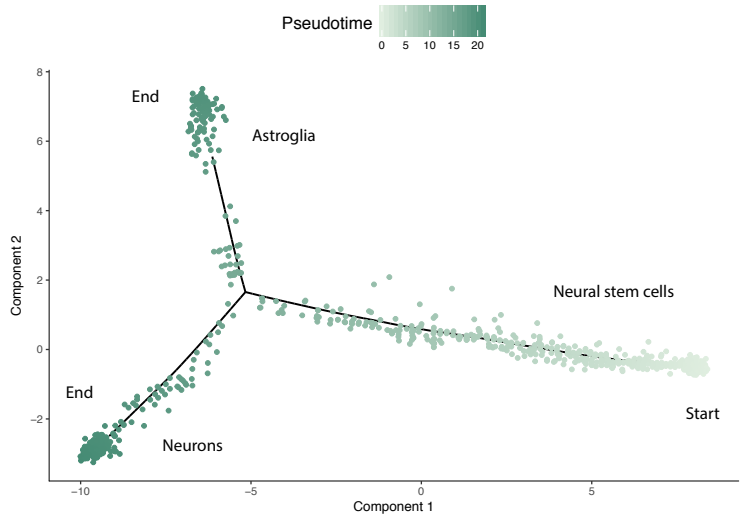


Figure 28. Pseudotemporal ordering of NES cells differentiating into neurons and glia (adapted from Figure 3 Paper 1, Lam et al 2019)

4.3.5 SEURAT – Single cell data analysis package

Seurat is a dedicated R package for single cell data analysis (Seurat version 3 in mid 2019 (Stuart et al., 2019)). Seurat encompasses a wide range of single cell analysis perspectives with well-maintained code and support for work practices commonly encountered when analyzing single cell data (https://satijalab.org/seurat/).

4.3.6 Data Visualization for single cell data

Graphical presentation of single cell data is a necessary for efficient communication of high dimensional output from data analysis.

Clustering algorithms such as t-Distributed Stochastic Neighbor Embedding (t-SNE) (Maaten et al., 2008) and Uniform Manifold Approximation and Projection (UMAP) (Leland McInnes, 2018) are widely used dimensionality reduction techniques for projecting high-dimensional data onto 2-dimensional (2d) space for visual data presentation. The graphical data presentation of t-SNE and UMAP confer interpretation of single cells that are closely aggregated next to each other hold gene expression profiles that are more similar than compared to single cells that are further apart on the 2d space (Figure 29).

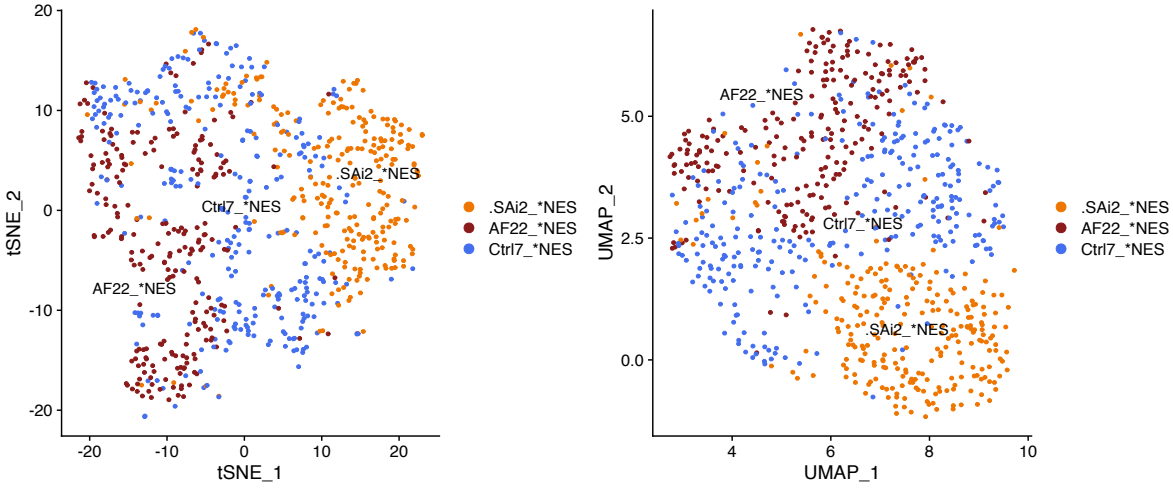


Figure 29. To illustrate t-SNE and UMAP 2d plots, each dot represents a single cell, the same data set (data for Figure 2 used in Paper 2, Lam et al 2019).

The heatmap presents a highly useful way of visualizing single cell data. Gene expression enrichment in clusters of cells or clusters of cells displaying expression of selected genes can be used to efficiently communicate overview of gene to cell relationships in single cell data sets (Figure 30).

The cluster heat map is a resourceful display that simultaneously reveals genes (rows) and cells (columns) hierarchical cluster structure in a data matrix. Consisting of a rectangular tiling, every tile in the heatmap is shaded on a color scale to represent the value of the corresponding element of the data matrix. The genes of the tiling are ordered such that similar genes are near each other. The same principle is applied to the cells, with similar cells being near each other (Leland Wilkinson, 2008).

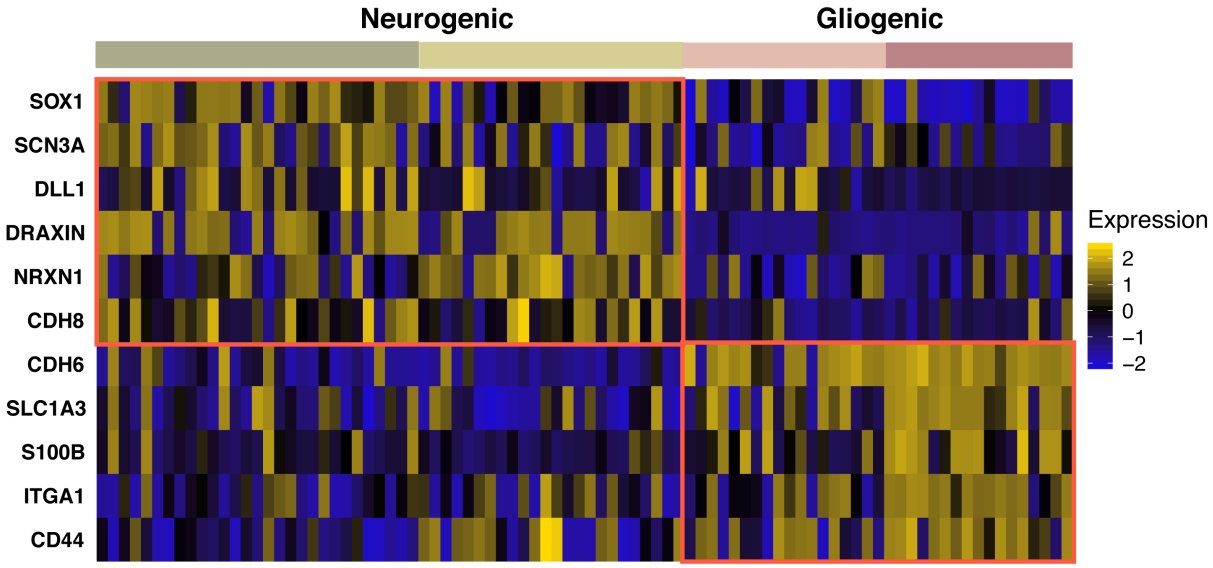


Figure 30. To illustrate, heatmap for enriched genes selected for neurogenic progenitors and gliogenic progenitors same as Figure 4 (subset of data for Figure 1 used in Paper 2, Lam et al 2019).

The violin plot presents distribution of data by synergistically combining the boxplot and density trace into a single display that reveals structure found within the data (Jerry L. Hintze, 1998). The violin distribution display a good overview of differences that can be found between cell types and/or conditions in a data set, e.g gene expression enrichment in neurogenic progenitors *NRXN1* and *CDH8* in contrast to gliogenic progenitors *CDH6* while almost all neural stem cells cells express high levels of *CDH2* (Figure 31).

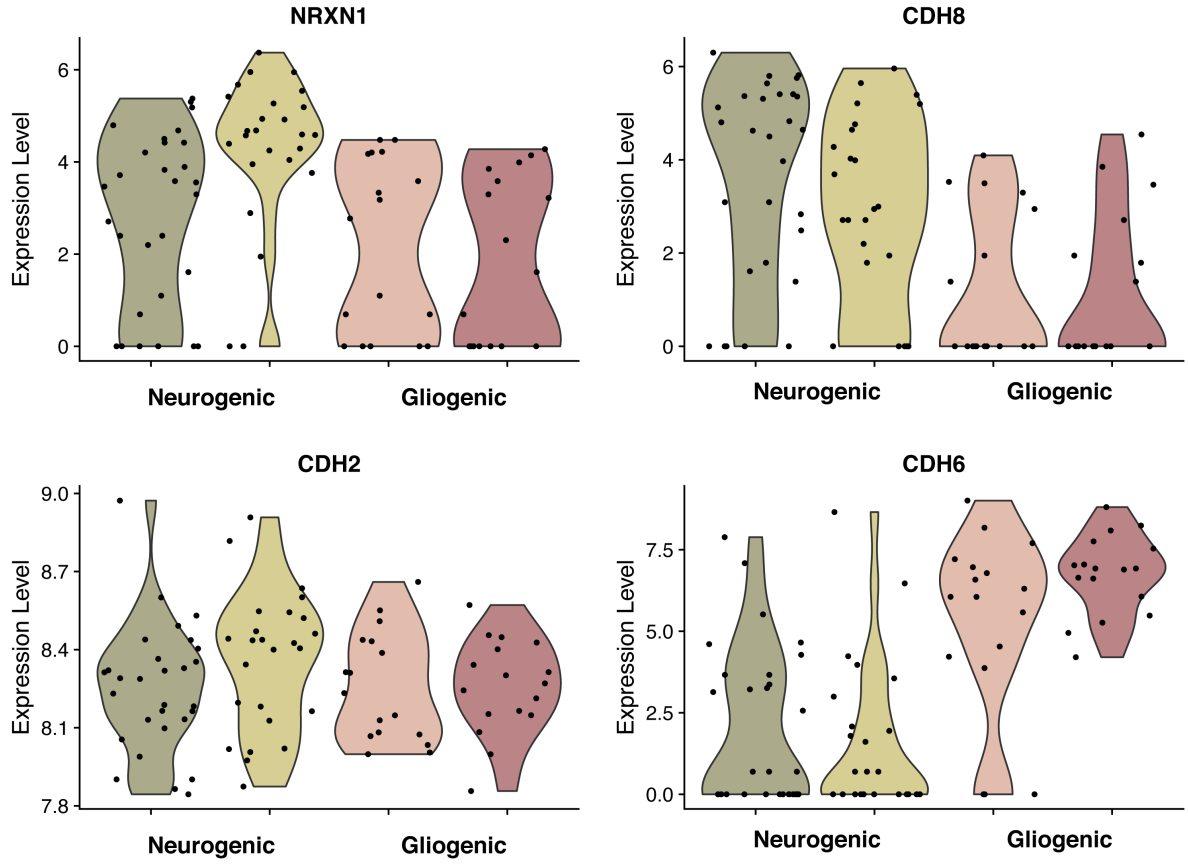


Figure 31. To illustrate violin plots, enriched genes selected for neurogenic progenitors and gliogenic progenitors same as Figure 21 (data from Figure 1 used in Paper 2, Lam et al 2019).

5 AIM OF PAPERS

5.1 PAPER 1

Using single cell RNA-seq to investigate the impact of NRXN1-a deletion on establishment of human iPS derived neural stem cells and functional properties of differentiated neurons.

5.2 PAPER 2

Using single cell RNA-seq to investigate presence of neurogenic progenitor and gliogenic progenitor and potential of differentiation capacity for human fetal neural stem cells and human iPS derived neural stem cells.

5.3 PAPER 3

Investigate DCX mutation impact on differentiation potential of human iPS derived neural stem cells and cell migration properties of differentiated neurons.

6 SUMMARY, DISCUSSION AND FUTURE

In Paper 1, NRXN1-a was observed expressed during neural induction phase and established neural stem cells carrying NRXN1-a deletion being radial-glial-like in cell identity. Further, increase in glial cells was observed in differentiation outcome and neurons exhibiting immaturity and cellular dysfunction.

In Paper 2, neural stem cells were observed to contain subpopulations of both neurogenic progenitors and gliogenic progenitors. The presence of predestined progenitors will determine differentiation outcome of neurons and glia.

In Paper 3, DCX mutation was observed to impair potential of neural stem cell differentiation, capacity of neurite outgrowth, and hamper migratory properties of cells. In addition to DCX mutation causing dysfunction in differentiation of neurons, re-analysis of microarray data for presence of genes indicative of neurogenic and gliogenic progenitors in NES cell stage displayed elevated markers for gliogenic genes in DCX mutation carrying cell lines. This added perspective suggest presence of more gliogenic cells contributing to observation of less migratory and hampered axon elongation characteristics in differentiated cells when interpreting impact of DCX mutation on the reported results in article.

Observations show that estimating the cell identity subpopulation of lineage progenitors in neural stem cells aids in the interpretation of cell differentiation outcomes. Becoming aware of the fine line between estimating cell identity and neurogenic potential in opposition to the gliogenic potential of a given neural stem cell line will help to make a proper readout of differentiation. Understanding the results, whether it is neurons or glial cells being investigated, there should in best of cases be an awareness of preexisting progenitor cell heterogeneity residing inside the neural stem cell culture.

Anticipating heterogeneous progenitor subpopulations residing in neural stem cells should be a factor to be considered as a major contributor of differentiation outcome. For future studies optimized protocols for establishing neural stem cells should also be robust and readily reproducible. Single cell RNA-seq screening could be considered as routine to ensure understanding and knowledge about the starting neural stem cell material used in experiments on studies in neurogenesis and neurodevelopmental dysfunction.

7 ACKNOWLEDGEMENTS

My research has been supported **Karolinska Institutet doctoral funding**, a highly generous funding source initiated by Karolinska Institutet to support start-up Principal Investigators to recruit doctoral students. My single cell bioinformatics work has been supported by participation in the **Swedish Bioinformatics Advisory Program**.

First of all, **Anna Falk**, I thank you for trusting in me and letting me study under your guidance, investing your time, dedication and allowing me to learn and mature into the principles of science and scientific work.

Åsa Björklund, I thank you for your guidance and mentoring, without you I could have never accomplished my goals and solved my research projects using single cell analysis.

To all Falk lab members, past and present, we had a rough ride and yet we survived it all. **Malin**, your positive outlook never waivers, you are a beacon of light. **Mohsen**, your dedication is great, have the best in life and work. **Elias**, there are no limits. **Harriet**, your guidance and experience brought me through it all, I would have failed and quit without your relentless encouragement. Words are not enough to explain what you mean to me.

Ronny, **Kelly**, **Mastoureh**, **Robin** and **Ana**, I wish you all the best in work and life.

To everyone I met during my exchange trip to Tokyo, **Okano-sensei**, you are a source of awe and inspiration, I am truly grateful for the generosity and trust you placed in me. **Yasui-sensei**, thank you for supporting my exchange trip, I am grateful for the experience. **Kohyama-sensei** and **Sanosaka-san**, I am so fortunate to have visited and worked with you, thank you for everything. **Mao-san**, thank you for the good times, you inspire me to reach into the future.

To all collaborators, I thank you for contributing to our work, all the hard work amounted into something great. **Julien**, **Ivar**, **Jessica**, **Rebecka**, **Loora, Lauri**, and **Jens**, thank you for your work on the NRXN1 project. **Anders** and **Kent,** thank you for sharing your knowledge and material on the progenitor project.

Seunghee, my life, my wife, together we accomplish the impossible. I am truly blessed to have your love in my life.

8 REFERENCES

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

Arendt, D. (2008). The evolution of cell types in animals: emerging principles from molecular studies. Nat Rev Genet *9*, 868-882.

Banda, E., McKinsey, A., Germain, N., Carter, J., Anderson, N.C., and Grabel, L. (2015). Cell polarity and neurogenesis in embryonic stem cell-derived neural rosettes. Stem Cells Dev *24*, 1022-1033.

Basu, S., Campbell, H.M., Dittel, B.N., and Ray, A. (2010). Purification of specific cell population by fluorescence activated cell sorting (FACS). J Vis Exp.

Bota, M., and Swanson, L.W. (2007). The neuron classification problem. Brain Res Rev *56*, 79-88.

Brennecke, P., Anders, S., Kim, J.K., Kolodziejczyk, A.A., Zhang, X., Proserpio, V., Baying, B., Benes, V., Teichmann, S.A., Marioni, J.C.*, et al.* (2013). Accounting for technical noise in single-cell RNA-seq experiments. Nat Methods *10*, 1093-1095.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat Biotechnol *27*, 275-280.

Cheng, S., Pei, Y., He, L., Peng, G., Reinius, B., Tam, P.P.L., Jing, N., and Deng, Q. (2019). Single-Cell RNA-Seq Reveals Cellular Heterogeneity of Pluripotency Transition and X Chromosome Dynamics during Early Mouse Development. Cell Rep *26*, 2593-2607 e2593.

Custo Greig, L.F., Woodworth, M.B., Galazo, M.J., Padmanabhan, H., and Macklis, J.D. (2013). Molecular logic of neocortical projection neuron specification, development and diversity. Nat Rev Neurosci *14*, 755-769.

de Wit, J., and Ghosh, A. (2016). Specification of synaptic connectivity by cell surface interactions. Nat Rev Neurosci *17*, 22-35.

DeFelipe, J., Lopez-Cruz, P.L., Benavides-Piccione, R., Bielza, C., Larranaga, P., Anderson, S., Burkhalter, A., Cauli, B., Fairen, A., Feldmeyer, D.*, et al.* (2013). New insights into the classification and nomenclature of cortical GABAergic interneurons. Nat Rev Neurosci *14*, 202-216.

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183-1186.

Emmert-Buck, M.R., Bonner, R.F., Smith, P.D., Chuaqui, R.F., Zhuang, Z., Goldstein, S.R., Weiss, R.A., and Liotta, L.A. (1996). Laser capture microdissection. Science *274*, 998-1001.

Falk, A., Koch, P., Kesavan, J., Takashima, Y., Ladewig, J., Alexander, M., Wiskow, O., Tailor, J., Trotter, M., Pollard, S.*, et al.* (2012). Capture of neuroepithelial-like stem cells from pluripotent stem cells provides a versatile system for in vitro production of human neurons. PLoS One *7*, e29597.

Hatakeyama, J., Wakamatsu, Y., Nagafuchi, A., Kageyama, R., Shigemoto, R., and Shimamura, K. (2014). Cadherin-based adhesions in the apical endfoot are required for active Notch signaling to control neurogenesis in vertebrates. Development *141*, 1671-1682.

Haubst, N., Georges-Labouesse, E., De Arcangelis, A., Mayer, U., and Gotz, M. (2006). Basement membrane attachment is dispensable for radial glial cell fate and for proliferation, but affects positioning of neuronal subtypes. Development *133*, 3245-3254.

Hayashi, T., Ozaki, H., Sasagawa, Y., Umeda, M., Danno, H., and Nikaido, I. (2018). Single-cell full-length total RNA sequencing uncovers dynamics of recursive splicing and enhancer RNAs. Nat Commun *9*, 619.

Horner, V.L., and Caspary, T. (2011). Disrupted dorsal neural tube BMP signaling in the cilia mutant Arl13b hnn stems from abnormal Shh signaling. Dev Biol *355*, 43-54.

Huang, A.Y., Yu, D., Davis, L.K., Sul, J.H., Tsetsos, F., Ramensky, V., Zelaya, I., Ramos, E.M., Osiecki, L., Chen, J.A.*, et al.* (2017). Rare Copy Number Variants in NRXN1 and CNTN6 Increase Risk for Tourette Syndrome. Neuron *94*, 1101-1111 e1107.

Ilicic, T., Kim, J.K., Kolodziejczyk, A.A., Bagger, F.O., McCarthy, D.J., Marioni, J.C., and Teichmann, S.A. (2016). Classification of low quality cells from single-cell RNA-seq data. Genome Biol *17*, 29.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J.B., Lonnerberg, P., and Linnarsson, S. (2012). Highly multiplexed and strand-specific single-cell RNA 5' end sequencing. Nat Protoc *7*, 813-828.

Islam, S., Zeisel, A., Joost, S., La Manno, G., Zajac, P., Kasper, M., Lonnerberg, P., and Linnarsson, S. (2014). Quantitative single-cell RNA-seq with unique molecular identifiers. Nat Methods *11*, 163-166.

James A Briggs, C.W., Daniel E Wagner, Sean Megason, Leonid Peshkin, Marc W Kirschner, Allon M Klein (2018). The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. Science.

Jenkins, A.K., Paterson, C., Wang, Y., Hyde, T.M., Kleinman, J.E., and Law, A.J. (2016). Neurexin 1 (NRXN1) splice isoform expression during human neocortical development and aging. Mol Psychiatry *21*, 701-706.

Jerry L. Hintze, R.D.N. (1998). Violin Plots: A Box Plot-Denstiy Trace Synergism. The American Statistician *52*, 181-184.

Kadowaki, M., Nakamura, S., Machon, O., Krauss, S., Radice, G.L., and Takeichi, M. (2007). N-cadherin mediates cortical organization in the mouse brain. Dev Biol *304*, 22-33.

Koch, P., Opitz, T., Steinbeck, J.A., Ladewig, J., and Brustle, O. (2009). A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for in vitro instruction and synaptic integration. Proc Natl Acad Sci USA *106*, 3225-3230.

Kurimoto, K., Yabuta, Y., Ohinata, Y., and Saitou, M. (2007). Global single-cell cDNA amplification to provide a template for representative high-density oligonucleotide microarray analysis. Nat Protoc *2*, 739-752.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., Borm, L.E., Stott, S.R.W., Toledo, E.M., Villaescusa, J.C.*, et al.* (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. Cell *167*, 566-580 e519.

Lam, M., Moslem, M., Bryois, J., Pronk, R.J., Uhlin, E., Ellstrom, I.D., Laan, L., Olive, J., Morse, R., Ronnholm, H.*, et al.* (2019). Single cell analysis of autism patient with bi-allelic NRXN1-alpha deletion reveals skewed fate choice in neural progenitors and impaired neuronal functionality. Exp Cell Res, 111469.

Lauren F. Harkin, S.J.L., Yaobo Xu, Ayman Alzu'bi, Alexandra Ferrara, Emily A. Gullon, Owen G. James, and Clowry, a.G.J. (2016). Neurexins 1–3 Each Have a Distinct Pattern of Expression in the Early Developing Human Cerebral Cortex. Cereb Cortex.

Le Dreau, G., Saade, M., Gutierrez-Vallejo, I., and Marti, E. (2014). The strength of SMAD1/5 activity determines the mode of stem cell division in the developing spinal cord. J Cell Biol *204*, 591-605.

Leland McInnes, J.H., James Melville (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction.

Leland Wilkinson, M.F. (2008). The History of the Cluster Heat Map. The American Statistician *63*, 179-184.

Lin, J., Wang, C., and Redies, C. (2014). Restricted expression of classic cadherins in the spinal cord of the chicken embryo. Front Neuroanat *8*, 18.

Lister, R., O'Malley, R.C., Tonti-Filippini, J., Gregory, B.D., Berry, C.C., Millar, A.H., and Ecker, J.R. (2008). Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell *133*, 523-536.

Liu, N., Liu, L., and Pan, X. (2014). Single-cell analysis of the transcriptome and its application in the characterization of stem cells and early embryos. Cell Mol Life Sci *71*, 2707-2715.

Lundin, A., Delsing, L., Clausen, M., Ricchiuto, P., Sanchez, J., Sabirsh, A., Ding, M., Synnergren, J., Zetterberg, H., Brolen, G.*, et al.* (2018). Human iPS-Derived Astroglia from a Stable Neural Precursor State Show Improved Functionality Compared with Conventional Astrocytic Models. Stem Cell Reports *10*, 1030-1045.

Maaten, L.v.d., Hinton, G.E., and Bengio, Y. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research *9*, 2579--2605.

Macosko, E.Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A.R., Kamitaki, N., Martersteck, E.M.*, et al.* (2015). Highly Parallel Genome-wide Expression Profiling of Individual Cells Using Nanoliter Droplets. Cell *161*, 1202-1214.

Miltenyi, S., Muller, W., Weichel, W., and Radbruch, A. (1990). High gradient magnetic cell separation with MACS. Cytometry *11*, 231-238.

Molyneaux, B.J., Arlotta, P., Menezes, J.R., and Macklis, J.D. (2007). Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci *8*, 427-437.

Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., and Wold, B. (2008). Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat Methods *5*, 621-628.

Nagalakshmi, U., Wang, Z., Waern, K., Shou, C., Raha, D., Gerstein, M., and Snyder, M. (2008). The transcriptional landscape of the yeast genome defined by RNA sequencing. Science *320*, 1344-1349.

Nelson, S.B., Sugino, K., and Hempel, C.M. (2006). The problem of neuronal cell types: a physiological genomics approach. Trends Neurosci *29*, 339-345.

Ouriel Caen, H.L., Philippe Nizard, Valerie Taly (2017). Microfluidics as a Strategic Player to Decipher Single-Cell Omics? Trends Biotechnol *35*, 713-727.

Owens (2016). Measuring Absolute RNA Copy Numbers at High Temporal Resolution Reveals Transcriptome Kinetics in Development. Cell Reports *14*, 632-647.

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., *et al.* (2016). Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. Dev Cell *36*, 681-697.

Peng, G., Suo, S., Cui, G., Yu, F., Wang, R., Chen, J., Chen, S., Liu, Z., Chen, G., Qian, Y., *et al.* (2019). Molecular architecture of lineage allocation and tissue organization in early mouse embryo. Nature *572*, 528-532.

Petropoulos, S., Edsgard, D., Reinius, B., Deng, Q., Panula, S.P., Codeluppi, S., Plaza Reyes, A., Linnarsson, S., Sandberg, R., and Lanner, F. (2016). Single-Cell RNA-Seq Reveals Lineage and X Chromosome Dynamics in Human Preimplantation Embryos. Cell *165*, 1012-1026.

Pollen, A.A., Nowakowski, T.J., Shuga, J., Wang, X., Leyrat, A.A., Lui, J.H., Li, N., Szpankowski, L., Fowler, B., Chen, P., *et al.* (2014). Low-coverage single-cell mRNA sequencing reveals cellular heterogeneity and activated signaling pathways in developing cerebral cortex. Nat Biotechnol *32*, 1053-1058.

Qiu, X., Mao, Q., Tang, Y., Wang, L., Chawla, R., Pliner, H.A., and Trapnell, C. (2017). Reversed graph embedding resolves complex single-cell trajectories. Nat Methods *14*, 979-982.

Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. Cell *135*, 216-226.

Shaltouki, A., Peng, J., Liu, Q., Rao, M.S., and Zeng, X. (2013). Efficient generation of astrocytes from human pluripotent stem cells in defined conditions. Stem Cells *31*, 941-952.

Shang, Z., Chen, D., Wang, Q., Wang, S., Deng, Q., Wu, L., Liu, C., Ding, X., Wang, S., Zhong, J., *et al.* (2018). Single-cell RNA-seq reveals dynamic transcriptome profiling in human early neural differentiation. Gigascience *7*.

Sozen, B., Can, A., and Demir, N. (2014). Cell fate regulation during preimplantation development: a view of adhesion-linked molecular interactions. Dev Biol *395*, 73-83.

Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius, M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-1902 e1821.

Sudhof, T.C. (2017). Synaptic Neurexin Complexes: A Molecular Code for the Logic of Neural Circuits. Cell *171*, 745-769.

Sun, Y., Pollard, S., Conti, L., Toselli, M., Biella, G., Parkin, G., Willatt, L., Falk, A., Cattaneo, E., and Smith, A. (2008). Long-term tripotent differentiation capacity of human neural stem (NS) cells in adherent culture. Mol Cell Neurosci *38*, 245-258.

Tailor, J., Kittappa, R., Leto, K., Gates, M., Borel, M., Paulsen, O., Spitzer, S., Karadottir, R.T., Rossi, F., Falk, A., *et al.* (2013). Stem cells expanded from the human embryonic hindbrain stably retain regional specification and high neurogenic potency. J Neurosci *33*, 12407-12422.

Takahashi, K., Tanabe, K., Ohnuki, M., Narita, M., Ichisaka, T., Tomoda, K., and Yamanaka, S. (2007). Induction of pluripotent stem cells from adult human fibroblasts by defined factors. Cell *131*, 861-872.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Tang, F., Barbacioru, C., Wang, Y., Nordman, E., Lee, C., Xu, N., Wang, X., Bodeau, J., Tuch, B.B., Siddiqui, A.*, et al.* (2009). mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods *6*, 377-382.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Tropepe, V., Sibilia, M., Ciruna, B.G., Rossant, J., Wagner, E.F., and van der Kooy, D. (1999). Distinct neural stem cells proliferate in response to EGF and FGF in the developing mouse telencephalon. Dev Biol *208*, 166-188.

Ullrich, B., Ushkaryov, Y.A., and Sudhof, T.C. (1995). Cartography of neurexins: more than 1000 isoforms generated by alternative splicing and expressed in distinct subsets of neurons. Neuron *14*, 497-507.

van de Leemput, J., Boles, N.C., Kiehl, T.R., Corneo, B., Lederman, P., Menon, V., Lee, C., Martinez, R.A., Levi, B.P., Thompson, C.L.*, et al.* (2014). CORTECON: a temporal transcriptome analysis of in vitro human cerebral cortex development from human embryonic stem cells. Neuron *83*, 51-68.

Woo S-M, J.K., Hyo-Won Han, Jung-Il Chae, Mi-Young Son, Sunwha Cho, Hyung-Min Chung, Yong-Mahn Han and Yong-Kook Kang (2009). Notch signaling is required for maintaining stem-cell features of neuroprogenitor cells derived from human embryonic stem cells. BMC Neurosci *10*.

Wu, A.R., Neff, N.F., Kalisky, T., Dalerba, P., Treutlein, B., Rothenberg, M.E., Mburu, F.M., Mantalas, G.L., Sim, S., Clarke, M.F.*, et al.* (2014). Quantitative assessment of single-cell RNA-sequencing methods. Nat Methods *11*, 41-46.

Zeisel, A., Munoz-Manchado, A.B., Codeluppi, S., Lonnerberg, P., La Manno, G., Jureus, A., Marques, S., Munguba, H., He, L., Betsholtz, C.*, et al.* (2015). Brain structure. Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. Science *347*, 1138-1142.

Zeng, Z., Sharpe, C.R., Simons, J.P., and Gorecki, D.C. (2006). The expression and alternative splicing of alpha-neurexins during Xenopus development. Int J Dev Biol *50*, 39-46.

Zhu, Y.Y., Machleder, E.M., Chenchik, A., Li, R., and Siebert, P.D. (2001). Reverse transcriptase template switching: a SMART approach for full-length cDNA library construction. BioTechniques *30*, 892-897.