# Identifying gene regulatory networks during neuronal differentiation of iPSCs

Dutan Polit, Lucia Margarita

*Awarding institution:*
King's College London

# Identifying Gene Regulatory Networks During Neuronal Differentiation of iPSCs

**Lucía M. Dután Pólit**

**Thesis submitted for the degree of**

**Doctor of Philosophy**

**at King's College London**

**2018**

**Department of Clinical and Basic Neuroscience**
**Institute of Psychiatry**
**King's College London**

# Abstract

Neural induction is the earliest step in the formation of the human nervous system. However, the regulatory signals underlying neural induction are largely unknown due to ethical, technical and legal restrictions that limit access to live human cells during this stage. Dual SMAD inhibition induces ESC/iPSCs cells to acquire neural fate, providing an exceptional *in vitro* system to analyze neural induction. Transcriptome analyses enabled identification of sets of genes down-regulated during the initial 2 days of neural differentiation including pluripotency markers OCT4, NANOG and MYC. Subsequently, numerous genes are activated, including the neuroectodermal markers PAX6, ZEB2 and SOX11 and genes that have not being previously related with neural differentiation. Genes with similar expression profiles regulate biological processes such us cell-cell adhesion, required for the transition from pluripotency to neural competence. Statistical and mathematical approaches enable to infer time delay regulatory interactions of a set of transcription factors with high connectedness and model a gene regulatory network with 9 principal hubs (JUN, MYC, FOS, PAX6, SP3, CDC6, SMAD2, HDAC6, and LEF1). The network modules regulate activation or inhibition of biological processes associated with neural induction, such us cell proliferation (MYC), cell cycle progression (CDC6) and regulation of CNS development (PAX6). Single cell RNASeq demonstrated that neural induction gave rise to a largely homogeneous neuro-ectodermal cells population. Inhibition of WNT signaling during neural induction leads to re-specification of neuroectodermal cells to a placodal fate, which subsequently differentiate into GnRH neurons. Transcriptome analysis revealed a unique set of genes activated and inhibited during neural induction in the presence of WNT inhibition, some of which may be essential for re-specification to a placodal fate. These studies go some way to identifying genes and gene modules that pay a role during human neural induction thereby offering insight into a basic human developmental process and providing a foundation for understanding how specific genetic variations may give rise to neurodevelopmental disorders.

# Acknowledgements

I would like to thank the many people whose support has been essential for completing this project, and for making my PhD such a great experience. I would like to start by thanking my supervisors Noel Buckley and Jack Price for their guidance, advice and confidence during the study, for being inspiring scientists to work with and for encouraging me to pursue my own ideas. Secondly, I wish to thank Deepak Srivastava for his advice and encouragement. I thank our collaborators Shankar Subramanian and Jun Min (UCSD) for their immense contribution to the research by modelling the GRN, generating the STEM profiles and helping with statistical analyses. Equally, Larry Stanton and Steven Havlicek (GIS) deserve a special mention for their fundamental contribution in RNAseq library preparation and sequencing. I am particularly grateful to Leo Perfect and Graham Cocks for teaching me all I know about iPSC tissue culture and RNAseq analyses. I thank them all for their patience, support and friendship. I wish to thank Richard Killick for his time spent helping me analyse XAV393 doses responses. I would like to thank the Jack Price and Noel Buckley lab teams for being actively engaged in the development of my project and being ready to help. I thank my enthusiastic fellow PhD students in the Wohl institute for making the last three years such an enjoyable experience and for always providing good advice. I am particularly grateful to the SENESCYT program "Universidades de excelencia, 2014" from the Ecuadorian government for providing the financial support necessary for this PhD.

Lastly, I would like to thank Berenice Polit, Fausto Dutan and Simon Lee for always encouraging and inspiring me to be better personally and professionally. Without their love and support, this work would not have been possible.

# Table of contents

# Table of Figures

# Table of tables

# Abbreviations

| | |
|---|---|
| iPSC | Induced pluripotent stem cells |
| ESC | Embryonic stem cells |
| TF | Transcription factors |
| NE | Neuroectoderm |
| TE | Trophectoderm |
| PrE | Primitive endoderm |
| Epi | Epiblast |
| ICM | Inner cell mass |
| EGA | Embryonic genome activation |
| Dpf | Days post fertilization |
| CNS | Central nervous system |
| PS | Primitive streak |
| NP | Neural plate |
| NPSC | Neural progenitor stem cells |
| ORF | Open reading frame |
| 2i | Dual SMAD inhibition |
| IF | Intermediate filament |
| ChIP | Chromatin Immuno precipitation |
| NSC | Neural stem cells |
| NESC | Neuroectodermal stem cells |
| RG | Radial glial |
| ICD | Intracellular domain |
| NBZ | Neural border zone |
| NC | Neural crest |

| | |
|---|---|
| PPE | Pre-placodal ectoderm |
| OP | Olfactory placode |
| GRN | Gene regulatory networks |
| RIN | RNA integrity number |
| GO | Gene ontology |
| STEM | Short time series expression miner |
| DE | Differential expression |
| PCR | Polymerase chain reaction |
| GnRH | Gonadotropin releasing hormone |
| EMT | Epithelial mesenchymal transition |
| RA | Retinoic acid |
| NT | Neural tube |
| PCA | Principal component analyses |
| QC | Quality control |

# CHAPTER 1. General Introduction

## 1.1. Neuroectoderm formation

### 1.1.1. Differentiation of the inner cell mass

Before implantation the early mammalian embryo undergoes a series of lineage segregation events that lead to the formation of trophectoderm (TE), primitive endoderm (PrE) and the epiblast (Epi). The TE and PrE give rise to extra-embryonic tissue responsible for the transmission of nutrients while the Epi forms the embryo proper. ESCs can be derived from the blastocyst and represent the Epi cell lineage (Schrode *et al*., 2013).

The first lineages formed are the TE and the inner cell mass (ICM). Several studies suggest that cell position and polarization are implicated in the establishment of the two cell phenotypes. After fecundation the embryo undergoes a transition from transcriptional silence to embryonic genome activation (EGA). This process occurs at 2-cell stage in mice and from 4-8 cell stages in humans (Niakan & Eggan, 2012). Re-localization of E-cadherin and β-catenin complexes during this period results in an increase of intracellular adhesion and compaction (morula stage). In parallel, intracellular components such as cytoskeletal elements, endosomes, microtubules, microvilli and EZRIN or aPKC/PAR protein complexes localize in the apical pole (Schrode *et al*., 2013). During the 8-16 and 16-32 divisions, cells divide symmetrically in a perpendicular plane generating two polar cells, or asymmetrically in a parallel plane giving rise to one polar and one apolar cell. The polar cells are maintained in the outside and will specify into the TE, whereas apolar cells are directed to an inner position and will form the ICM around the 32 cells stage (Figure 1.1) (Lanner, 2011, Artus *et al*., 2014). The apolar/polar division frequency appears to be dependent on different factors such as the cell-to-cell contact, the nuclear size and the apical

domain size. The number of cells dividing with different mechanisms and the number of cells destined to become the ICM is compensated between the 8 cell division and the 16 cell division (Chazaud & Yamanaka, 2016). It has been suggested that the cell division plane and consequent arrangement of polar and apolar blastomeres is related to the position of the cell in the morula. Other studies propose that down-regulation of polarity proteins such aPKC increase the cortical tension and directs the cells to an inner position and segregation of the prospective TE cells (Bruce & Goetz, 2010; Lanner *et al.*, 2014). Additionally, recent studies suggest that the differential blastomeres contractility potentially results from asymmetric divisions. Apolar cells internalize as a consequence of their higher amplitude of contraction compared with polar cells (Menchero *et al.*, 2016).

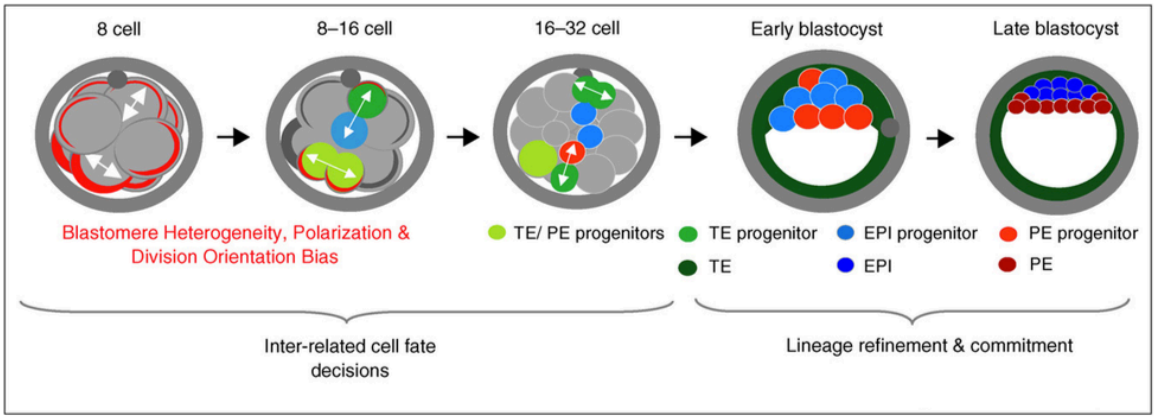



**Figure 1.1. Blastomeres asymmetrical division and trophectoderm, primitive endoderm and Epibalst differentiation, in pre-implantation mouse.** At 8 cells stage, the cells acquire polarity and divide asymmetrically in a perpendicular plane or symmetrically in a parallel plane. The polar cells (green) are maintained in the outside and give rise to the TE. The apolar cells are going to form the ICM, which later differentiate into the Epi (blue) or PE (red) (Bruce & Goetz, 2010).

During the 3-3.5 days post-fertilization (dpf) in mice and 4-5 dpf in humans a fluid cavity called the blastocoel is formed as consequence of a continuous water and ion flux mediated by NA+/K+ ATPase and aquaporins (Artus *et al.*, 2014). At this period the embryo

comprises the TE encapsulating the ICM and the blastocoel cavity (Niakan & Eggan, 2012). Besides polarity, the inner and outer blastomeres present differential gene expression patterns. Mutant mouse blastocyst lacking CDX2 failed to specify the TE, indicating that the transcription factor CDX2 is required for the formation and maintenance of the TE. Additionally, GATA3 and EOMES form part of the TE gene regulatory network. The ICM is marked by the expression of OCT4. It has been proposed that OCT4 and CDX2 interact by mutually repressing each other. For instance, studies with mouse ESC demonstrated that overexpression of Cdx2 leads to the reduction of Oct4, leading to differentiation towards a TE fate. Initially, zygotic OCT4 is expressed in all cells and becomes restricted to the ICM during the blastocyst formation (Lanner *et al*., 2014). OCT4 regulates expression of CDX2 preventing cells from adopting a TE fate. In mouse, CDX2 is expressed in most 16 cells stage blastomeres. Later it is restricted to the outer cells prior to blastocyst formation and is necessary to repress OCT4 and NANOG expression (Schrode *et al*., 2013). In the human embryo it remains unclear when OCT4 and CDX2 expression initiates. Human OCT4 inhibition in the TE begins from 6-10dpf while in mouse it occurs at day 4.5dpf showing temporal species differences of transcription factor expression (Niakan & Eggan, 2012).

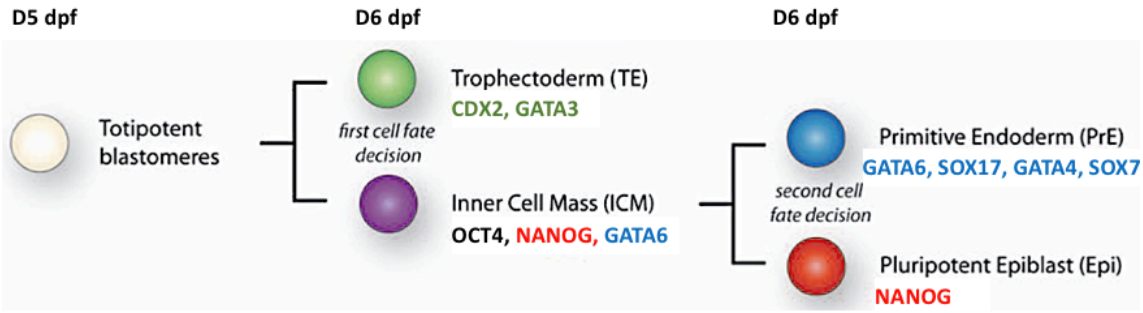



**Figure 1.2. Overview of the blastocyst differentiation in humans.** Blastocyst lineage differentiation initiates with the segregation of the TE (CDX2 and GATA3) and the ICM (OCT4, NANOG and GATA6) at D6 dpf. In parallel, the TF GATA6 and OCT4 become mutually exclusive in a salt-and-pepper manner in the ICM cells to give rise to the PrE and the Epi, respectively (adapted from Schrode *et al*., 2013).

Null TEAD4 mouse embryos lack expression of TE markers like CDX2 while ICM markers such OCT4 and NANOG are maintained. It has been shown that TED4 acts upstream to CDX2 possibly through interactions with the co-activators YAP and TAZ (Lanner *et al*., 2014). TED4 is expressed in all blastomeres but only promotes CDX2 expression in the outer cells. The current model proposes that, in the ICM, YAP and TAZ are phosphorylated and degraded by the Hippo pathway affecting TEAD4 ability to activate its targets. In the outer cells the Hippo pathway is inactive allowing YAP/TAZ to bind TEAD4 and to activate target genes like CDX2 (Schrode *et al,* 2013). On the other hand, it has been suggested that nuclear exclusion of TEAD4 instead of YAP/TAZ in the ICM is responsable for ICM-TE patterning. However, Further experiments are necessary to confirm the model (Lanner *et al*., 2014). Null Tead4 embryos cultivated in low $O_2$ generate TE suggesting that there might other signaling cascade or alternative YAP DNA-binding partners that compensate Tead4 function (Chazaud & Yamanaka, 2016).

Besides transcription factor levels, TE and ICM patterning has being associated with differential TF kinetics. In the 8 cells stage mouse, Oct4 and Cdx2 exhibit differential nuclear fusion kinetics. Cells with slow Oct4 kinetics undergo asymmetric division and therefore mainly contribute to the ICM, whereas cells with slower Cdx2 kinetics principally contribute to the TE. The disparities in fusion kinetics have being related to differences in the chromatin accessibility of the TE and ICM but the precise mechanisms remain unclear (Schrode *et al*., 2013; Lanner *et al*., 2014).

### 1.1.2. Epiblast and primitive endoderm patterning

The earliest event of ICM differentiation is the segregation of the PrE and the Epi. ICM cells lining the blastocoelic cavity will commit into the PrE (hypoblast) lineage, precursor

of the extra embryonic tissue (yolk sac), whereas the remaining ICM cells will form the pluripotent Epi (primitive ectoderm) with the potential to give rise to the three embryonic germ layers during gastrulation (Pelton *et al*., 2001, Roode *et al*, 2012).

The PrE/Epi lineage determination was initially attributed to the position of the cells among the ICM. Later studies demonstrated that ICM cell fate is determined before the prospective PE lines the blastocyst cavity. Nanog and Gata6 are the core TF involved in differentiation of the Epi and PrE respectively. The PrE and Epi genetic programs start at the 8-cell stage where GATA6 and OCT4 are co-expressed in all blastomeres. Around the 20 cells stage the PrE and Epi markers are down-regulated in specific cells of the late morula and become mutually exclusive (Artus *et al*., 2014; Schroter *et al*., 2015). After the mid blastocyst, cells expressing GATA6 and OCT4 are distributed along the ICM in a salt-and-pepper manner. Later, the ICM cell population commits to specific linages as the prospective Epi express pluripotency genes such as Nanog and Sox2 required to maintain the pluripotency; and the PrE precursors express a series of TF including Sox17, Gata4, and Sox7 (Ohnishi *et al*., 2013). It has been shown that the Nanog and Gata6 expression levels fluctuate among ICM cells leading the cells to specify at different rates. It has been proposed that cells expressing both TF represent a transient phenotype that balances between the PrE and Epi lineages. A recent model defines three stable states where cells express Gata6 or Nanog exclusively or express both TF. The mechanism of differentiation into a specific lineage might be related to reduction of mRNA transcripts rather than selective increase of transcription rates. However, chromatin immunoprecipitation studies demonstrated that Nanog and Gata6 increase and maintain their own expression by a direct positive feedback and interact by repressing each other (Chazaud & Yamanaka, 2016).

It has been proposed that the fate choice between cells co-expressing Nanog/Gata6 is mediated by FGF/MAPK signaling. Specifically, Fgf4 modulates the PrE/Epi proportions by promoting Gata6 expression and inhibiting Nanog (Ohnishi *et al*., 2013; Schroter *et al*., 2015). Interestingly, Gata6 expression remains until the early blastocyst stage in Fgf4 mutants, meaning that other genes are necessary to trigger initial expression of Gata6. Fgf, Nanog and Gata6 are the main components of a gene regulatory network (GRN) in which specific levels in individual cells drive the ICM specification process. A recently developed GRN mathematical model predicts that early Nanog expression in the ICM promotes Fgf4 secretion. Neighbor cells with specific extracellular concentration of Fgf4 are induced to differentiate into Epi or PrE depending on the Fgf4 local abundances (Schroter *et al*., 2015). Heterogenic populations result from the differential propagation of Fgf4 paracrine signaling (Artus *et al*., 2014; Chazaud & Yamanaka, 2016). These studies indicate that Fgf4 is expressed in Nanog positive cells only for the establishment of the salt and pepper array. Subsequently, Epi cells insulate themselves from Fgf4 signaling by inhibition of the Fgfr2 (Fgf4 receptor). Additionally, FgF4 activates ERK to down-regulate Nanog in the prospective PrE (Figure 1.3) (Artus *et al*., 2014).

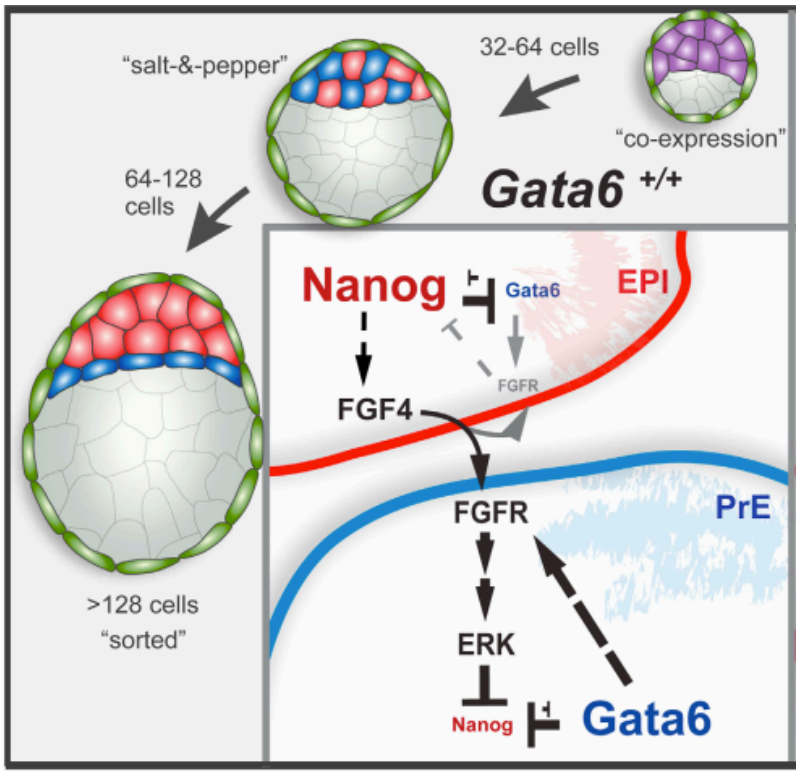



**Figure 1.3. Differentiation of the primitive endoderm and epiblast from the inner cell mass salt and pepper array in mouse.** At the 32-64 cell stage, the ICM Gata6 and Nanog positive cells undergo a transition after which the prospective PrE cells (blue) express exclusively Gata6, whereas the prospective Epi cells (red) express exclusively Nanog. It has been proposed that inhibition of ERK induces Nanog expression in the Epi, which in turn induces Fgf4 expression and secretion to neighboring cells. Fgf4 induces ERK signaling, which inhibits Nanog expression in the PrE cells (Schrode *et al*., 2014).

The first difference between ICM cells is differential expression of Fgf4. Nevertheless, the initial events that promote heterogenic expression of the ICM cells remain unclear. Single cell studies have revealed gene expression heterogeneity at two cells stage before zygotic transcription in mouse. It is though that cell heterogeneity results from unequal distribution of mRNAs, which occurs during early cell division. However, this mRNA heterogeneity relation with differential expression has not been tested. Among the hypotheses that have not been rejected to explain the establishment of the salt and pepper pattern are: early cell division bias; stochastic gene expression and reduced phosphorylated ERK that induces

Nanog expression and promote FgF4 secretion (Chazaud & Yamanaka, 2016). In humans, exclusive expression of GATA and NANOG has being shown at the mid blastocyst stage. However, inhibition of FGF4 signaling in human embryos does not deeply impact the ICM patterning, highlighting possible regulatory differences between species. Potential candidates for driving ICM heterogeneity in humans are members of the transforming growth factor family TGFβ, which have been proved to maintain self-renewal in human TE stem cells. Nevertheless, the mechanisms of PrE derivation in humans and the alternative pathways involve in the ICM segregation remain unknown (Roode *et al*, 2012; Chazaud & Yamanaka, 2016).

At the 64-cell stage PrE cells migrate and form a layer between the Epi and the blastocoel. The PrE cells that fail to locate within the PrE layer either change to an Epi fate or undergo apoptosis (Schroter *et al*., 2015). Prior to implantation in the uterus, the embryo exhibits three different linages resulting from different and sequential cell fate decisions: TE, PrE and Epi. The lineages have specific positions within the embryo, are molecularly distinct and have restricted differentiation potential (Artus *et al*., 2014).

During the past few decades, several groups have extensively studied human development to the point of implantation at day 7 when the embryo must attach to the uterus to survive. *In vitro* studies have used mouse as a model to considerably increase the understanding of early post-implantation development but fail to uncover particular human development landmarks due to specie specific differences. Magdalena Zernicka-Goetz's group has recently developed an outstanding technique that allows growth of human embryos until 13dpf (Deglincerti *et al*., 2016; Morris, 2017). The study validates the expression of TE markers CDX2 and GATA3 and suggests GATA6 as a better marker for human TE due to its high level of nuclear expression. The Epi marker NANOG and the PrE marker GATA6

were shown to be exclusively expressed in a salt and pepper pattern at 6dpf, in contrast with mouse where PrE and Epi markers are sorted by the late blastocyst stage. Interestingly, the results depicted that hESCs closely resemble Epi cells at 10dpf rather than 6-8dpf as previously accepted. The next step in human central nervous system (CNS) development corresponds to the formation of the primitive streak, gastrulation and neurogenesis. Studying the human embryo *in vitro* is legally permitted until day 13 or until the formation of the primitive streak (Deglincerti *et al*., 2016; Shahbazi *et al*., 2016a). iPSCs genetically and phenotypically resemble to ESC *in vivo*. Thus, iPSCs provide a unique *in vitro* system to analyze the molecular events that regulate human development after this period (Lohle *et al. 2012)*.

### 1.1.3. Neuroectodermal induction by BMP, TGFβ and WNT antagonists

During gastrulation the Epi cells initiate a complex developmental program that gives rise to the three germ layers of the mammalian embryo: the endoderm, mesoderm and ectoderm. In mouse E6.5 Epi cells located in the posterior side of the embryo undergo to epithelial mesenchymal transition and migrate through the primitive streak (PS). The cells that ingress through the PS to the interior of the embryo form the mesoderm and the endoderm. The Epi cells that are not recruited through the primitive streak differentiate into the ectoderm. Specific regions of the ectoderm differentiate into the epidermis, CNS cells, neural crest or placodes. In vertebrates the first event of neural differentiation is the ectoderm commitment into the NE. Differentiation of the NE depends on termination of the ESC transcription program, the initiation and maintenance of the NE transcription network and the suppression of other lineages fate. NE differentiation results from a series of regulatory events during the blastula and gastrula stages (Tresdell & LaBonne, 2006; Gaur *et al*., 2016).

The NE is formed in the dorsal region of the embryo in response to signaling from the adjacent mesoderm region called the "node" in chick, mouse and humans and the "organizer" in *Xenopus*. The node cells secrete small molecules that diffuse into the extracellular space of the adjacent ectoderm and bind BMP, TGFβ and Wnt ligands to inhibit them from occupy and activate their receptors. Ectodermal cells where BMP, TGFβ and Wnt have not being inhibited will adopt a non-neural ectodermal fate and subsequently an epidermal phenotype. Inhibition of BMP, TGFβ and Wnt allows expression of several dorsal ectoderm genes, which promote the induction of the NE and prevent the expression of non-neural genes. Additionally, signaling mediated by TGFβ family induces ectodermal cells to adopt a mesodermal fate (Andoniadou & Martinez-Barbera, 2013; Klein & Moody, 2015). BMPs bind and activate type I and type II receptors and induce their hetero-dimerization. Activation of BMP signaling induces phosphorylation of SMAD1, which in turn forms a complex with SMAD4b/SMAD4 and translocates to the nucleus to affect the expression of its gene targets. In the nucleus the complex binds to different transcription factors (TF) such as OAZ in in *Xenopus* to activate the epidermal transcription network. In the *Xenopus* embryo the BMP pathway is inhibited at an extracellular level by the secretion of several ligands that occupy BMP receptors; in the cytosol by Smad6 and Smad7 which prevent Smad4 from binding Smad1 and by Smurf1 and Smurf2, which target the Smad1-Smad2 complex for degradation by the proteasome; and in the nucleus by the transcriptional repressor Ski (Figure 1.4) (Muñoz-Sanjuán & Brivanlou, 2002; Ozair *et al*., 2013). Additionally, it has being demonstrated that in the chick ectoderm, BMP signaling is initially down-regulated at a transcriptional level and is further inhibited by BMP antagonists (Klein & Moody, 2015).

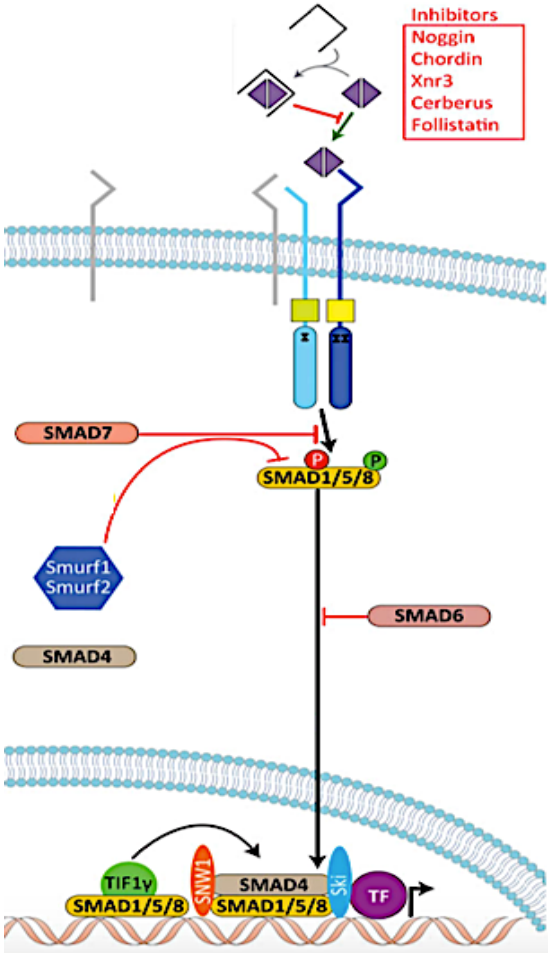



**Figure 1.4. BMP pathway and SMAD inhibition.** Molecules secreted from the node block extracellular BMP binding to type I/II receptors. In the cytosol, SMAD7/6 and Smurf1/2 prevent the formation of the SMAD complex. As a result of these interventions, SMADs cannot translocate to the nucleus to promote expression of their target genes (adapted from Ozair *et al*., 2012).

In vitro, ESC and iPSCs can be induced to differentiate to neural precursors by inhibiting BMP signal with small molecule BMP antagonists or by diluting endogenous BMP signalling by growing cells at low density. Nevertheless, the mechanism of neural precursors differentiation and maintenance by BMP inhibition is not well understood in ESC or in the embryo. Hence, the resemblance of the molecular events that drive neuronal induction between ESC and the embryo is not clear (Moody *et al*., 2013).

The Xenopus default model suggests that BMP inhibition is sufficient to induce neural differentiation in the competent Epi. However, studies in the chick embryo indicate that BMP inhibition is necessary to suppress epidermal phenotype, but an additional signal might be required to induce a neural phenotype. Potential candidates are FGF family members whose expression is up-regulated in the ectoderm and then restricted to the NE (Delaume *et al*., 2005; Marchal *et al*., 2009). In humans 22 FGF members have being reported, 18 correspond to secreted proteins that interact with 4 receptors (FGFR). In Xenopus embryo, neural tissue differentiation is inhibited chemically by inhibiting FGF. However, it has been demonstrated that chemical inhibition of FGF signaling reduces dorsal mesoderm tissue formation and therefore reduces the BMP inhibitory signals secreted from it. Additionally, overexpression of dominant-negative FGFR1 and FGFR2 receptors, which block FGF3 and FGF4 signals leads to loss of pan-neural markers in Xenopus neural ectoderm explants. Over expression of dominant-negative FGFR4α receptor blocks FGF8 signaling and prevents anterior neural marker expression. Posterior neural development is disrupted by inhibition of the FGF pathway through disruption of MAP kinase signaling but has no effect on anterior neural tissue formation (Stern, 2005; Rogers *et al*., 2009). In chick, FGF genes are required to inhibit expression of BMP genes and in Xenopus and zebrafish to inhibit BMP downstream signaling by phosphorylation and inactivation of Smad1. Several studies propose that FGF is necessary for anterior-posterior patterning of the neural plate (NP) and for inhibition of BMP signaling but any FGF independent role to initiate the nervous system development remains in question (Delaume *et al*., 2005; Stern, 2005). It has been suggested that FGF facilitates neural induction through its downstream effector Churchill (Churc), which inhibits the expression of mesodermal genes and promotes the expression of neural genes (Lee *et al*., 2014).

Canonical WNT signalling has also being implicated in multiple mechanisms of embryonic development including anterior NE induction, NP patterning, and neural crest and placode differentiation (McDonnlad *et al*., 2009). During Xenopus early development, maternal Wnt/β-catenin directly induces the expression of the closely related activators Siamois (Sia) and Twin. These signals combined with zygotic Nodal and TGF-β in the dorsal region of the embryo induce expression of the organizer genes. In mammals the Sia/Twin homologue DUXO has being demonstrated to be required for formation of the organizer. Simultaneously, the organizer secretes BMP and TGFβ antagonists like Chordin, Noggin and Cerberus, which specify the embryonic dorso-ventral regions and NE formation; and the Wnt antagonists Dickkopf1 (Dkk1), Frizzled 1 (Frzb1), Crescent, and Cerberus (Ding *et al*., 2017).

WNT zygotic expression in the ventro-lateral mesoderm generates a β-catenin gradient elevated in the posterior region where it activates regulatory signals that inhibit transcription of anterior NE genes or activity of the maternal NE activators and induce expression of genes encoding for the posterior NE. On the other hand, extracellular WNT antagonists in the anterior region inhibit NE posteriorization and induce expression of TF that promote anterior neural fate and will give rise to the forebrain (Range *et al*., 2013). For instance, inhibition of WNT in Xenopus ectoderm generates a significant increase in head size and up-regulation of the signals leads to microcephaly or loss of anterior structures. Thus, the ability of WNT signalling to induce anterior-posterior axis formation has being extensively demonstrated. However, there is not sufficient evidence to demonstrate a direct role of WNT in neural induction (Ciany & Salinas, 2005; Rogers *et al*., 2009).

Additionally, it has being suggested that the transcription the factor Zic1 stabilizes neuronal fate by sensitizing the ectoderm for neural induction, however the mechanism remains unknown (Lee *et al*., 2014).

### 1.1.4. Transcription factors involved in neuroectodermal differentiation

NE differentiation is orchestrated by sets of TFs that are sequentially activated or inhibited in a specific spatiotemporal manner. These TF maintain NE precursors and regulate their transition to NE cells that constitute the NP and then to neural progenitors (neural stem cells). Several TF are required during neural fate acquisition. However, the exact mechanisms and the GRNs that control neuronal development remains largely unknown (Moody *et al*., 2013).

In Xenopus, activation of the organizer genes via Sia and Twin is necessary for the secretion of BMP and WNT inhibitors, and to directly up-regulate the expression of NE genes. Among the organizer the TFs, Sia, Twin, Goosecoid, and Foxa4 are expressed in organizer precursor cells in the dorsal blastula and later in the organizer proper. Other TFs such as Hesx1, Lim1, Xnot, Otx2 and XlPOU2; and the genes Chordin, Noggin, Cerberus are expressed only once the organizer is formed. Ectopic expression experiments demonstrated that Sia and Twin directly activate expression of Foxd4 in the blastula before the organizer and NE are formed during gastrulation. Additionally, Sia and Twin directly induce expression of Geminin (Gmnn) and Zic2 in the blastula. Blocking Foxd4 translation in Xenopus blastomeres leads to inhibition of Sox11 expression, suggesting that Foxd4 directly regulates Sox11expression. Foxd4 is widely expressed in NE precursor cells and later is restricted to the anterior and midline region of the NP comprised of NE cells. Subsequently, Foxd4 expression is restricted to a small region of the midbrain. Once the

organizer has differentiated during gastrulation it secretes factors that inhibit BMP and WNT from the adjacent mesoderm and ectoderm inhibiting reversion to a non-neural fate. In the gastrula stage Gmnn, Zic2 and Sox11 expression is facilitated through Foxd4 and other organizer TFs (Klein & Moody, 2015, Gaur *et al*., 2016) (Figure 1.5). Expressing the human homologous FOXD4 and FOXD4L1 in Xenopus blastomeres explants demonstrated that the human proteins affect neural gene expression similarly to Foxd4, indicating possible functional conservation (Gaur *et al*., 2016). In the mouse embryo, Gmnn and Sox2 expression initiates in the inner cell mass and is up-regulated in NE progenitors during gastrulation. By day E7.5-8 Gmnn and Sox2 are widely expressed in the NE together with Zic1 and Sox3 (Sankar *et al*., 2016). Foxd4, Gmnn, Zic2 and Sox11 block non-NE induction by inhibiting BMP/WNT pathway genes and their targets (Klein & Moody, 2015).

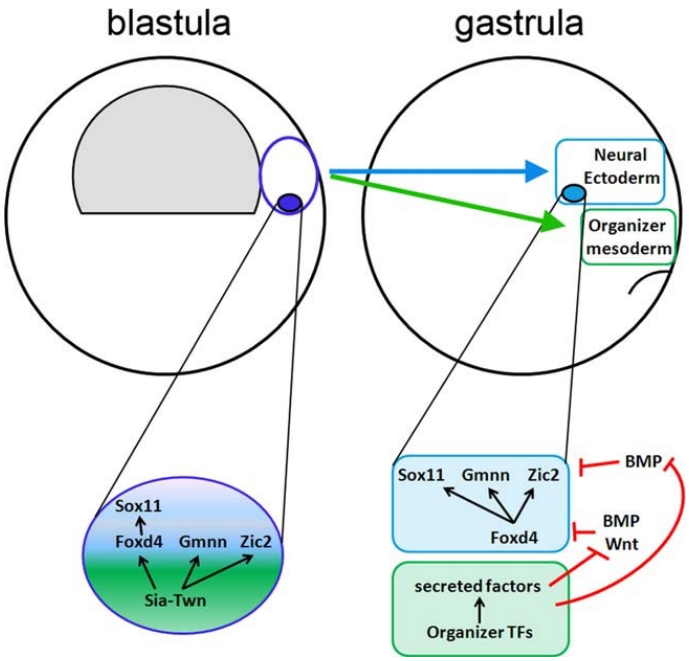



**Figure 1.5. Sia and Twn induction to gene targets.** At blastula stage, in cells fated to give rise to the ectoderm and mesoderm express Sia and twin, which activate Foxd4, Gmnn and Zic2. At this stage, FoxD4 promotes activation of Sox11. Once the mesoderm and endoderm are formed during gastrulation, NE genes expression is maintained by factors secreted by the organizer and enhanced by FoxD4 (Klein & Moody, 2015).

Additionally, these TFs maintain the NE in a proliferative state by inducing NE progenitor's genes facilitating NE expansion. Zic2 regulates ESC development by recruiting the co-factor Mbd3-NuRD to enhancer regions to regulate chromatin state and gene expression. Zic2 knockout in mouse ESCs cultured with N2B27 media to promote neuronal differentiation leads to cell death indicating the important role of Zic2 during neural differentiation (Luo *et al*., 2015).

Sox11 is up-regulated during the transition from NE to neural progenitor, and hence has being implicated in maintenance of pan-neural neural progenitors. In mouse and chick embryo, Sox11 is only expressed in neuronal progenitor cells and in developing neurons (Klein & Moody, 2015). In contrast, Gmn, Zic2 and Foxd4 have being implicated in the temporary inhibition of bHLH neural differentiation genes NeuroG and NeuroD. For instance, Gmnn and Foxd4 are down-regulated during the transition of cells from a proliferative NE to neuronal progenitors (Moody *et al*., 2013; Lee *et al*., 2014).

Gmnn interacts directly with TFs and with chromatin modifying complexes in a specific spatiotemporal manner to activate or repress genes expression. It has been proposed that Gmnn interacts with SWI/SNF and Polycomb chromatin modifying complexes to bind specific chromatin locations (Lim *et al*., 2011). Gmnn expression is required to maintain NE multipotency by promoting deposition of repressive histone modifications at neural genes. Additionally, during neural progenitor fate acquisition Gmnn facilitates the deposition of histone marks associated with open chromatin histone-activating marks in TFs gene targets in mouse ESCs promoting neuronal differentiation (Sankar *et al*., 2016). It has been demonstrated that Gmnn promotes acetylation of several genes involved in neuronal differentiation such as Bcl1, Pou6f1 and Pax4. Moreover, Gmnn could have a role in repressing mesodermal genes expression by enhancing Polycomb repressor complex.

Gmnn expression is also required earlier in development as Gmnn-/- mouse embryos fail to form ICM and generate TE in its place (Yellayoshyula *et al*., 2010; Caronna *et al*., 2011).

The conversion from NE progenitors to NE cells and eventually to neural progenitors requires cells to leave cell cycle and initiate expression of bHLH TFs. To promote this transition a different set of genes is activated. In Xenopus, the TFs Sox1, Sox2 and Sox3 are up-regulated downstream of Foxd4, Gmnn and Zic2. Sox1, Sox2 and Sox3 TF are members of the SoxB1 family and their expression is required to maintain self-renewing progenitors during the CNS development. High expression levels of SoxB TFs reduce the expression of bHLH neural TFs and inhibits neuronal differentiation. Thus, the transition of NE cells to differentiating neurons requires down-regulation of the SoxB family members (Holmberg *et al*., 2008) (Figure 1.6). SoxB TFs function in parallel to maintain cell proliferation and to establish spatial identity of neural cells at the same time and possibly in the same cells. The balance between SoxB1 TFs activity determines whether cells are maintained as progenitors or differentiate into neurons. The factor's binding partner might determine SoxB1 regulatory function, but the exact mechanism of functional diversity is not well understood (Neriec & Desplan, 2014). Additionally, SoxB1 TFs maintain NE fate by inhibiting expression of Vent2, a BMP target necessary for epidermis differentiation and by down-regulating Foxd4 (Rogers *et al*., 2009; Lee *et al*., 2014).

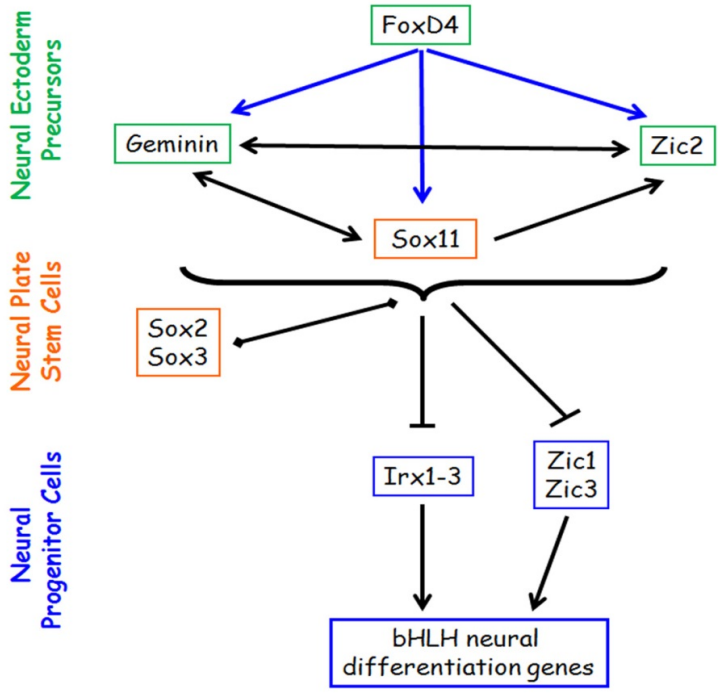



**Figure 1.6. Sets of transcription factors sequentially activated during early neuronal differentiation**. NE precursors (green) express FoxD4, which directly induces the expression of Gmnn, Zic2 and Sox11 and transiently inhibits the expression of SoxB1 and neural progenitor genes. Neural plate stem cells (orange) differentiation requires down-regulation of FoXD4, Gmnn and Zic2 TF and high levels of expression of SoxB1 and Sox11 genes. The transition between neural progenitor stem cells (NPSC) and differentiating neuronal progenitors (blue) depends on the inhibition of Sox family members and up-regulation of Zic and Irx TF (Lee *et al*., 2014).

SoxB1 binding results in bivalent chromatin remodeling, hence Sox proteins might play a role regulating the accessibility of subsequently expressed TFs. The proposed mechanism of regulation indicates that Sox members bind to target promoters/enhancers to deposit activate chromatin marks and prepare them for activation by another family member (Bergsland *et al*., 2011). In the mouse embryo and ESC cells Sox2, Sox3 and Sox11 bind to the promoters of the same target genes in a temporal sequence. Sox2 binds NE cells genes and marks the sites for latter Sox3 binding. Subsequently, Sox3 binds neuronal progenitor's genes to mark Sox11 binding sites (Nishimura *et al*., 2012; Moody *et al*., 2013). Sox2 expression initiates in the mouse ICM and is up-regulated during late

gastrulation. Sox2 expression inhibits activation of proneural genes and is down-regulated during the final cell cycle before neural differentiation proceeds (Wegner, M., 2011; Niu *et al*., 2015). Interestingly, Sox2 is part of the set of TFs capable of directing reprogramming of differentiated cells to pluripotent stem cells (Takahashi and Yamanaka, 2006). Sox3 and Sox1 are activated once mammalian cells have committed to a NE fate and overlap Sox2 expression at stage E7.5-8. Sox2 and Sox3 share high homology and are expressed in overlapping domains. Furthermore, Sox3 activity is compensated during diencephalon development when the Sox3 (open reading frame) ORF is replaced with the Sox2 ORF suggesting that their expression is redundant (Graham *et al*., 2003; Niu *et al*., 2015). Sox2 and Sox3 are necessary to maintain neuronal progenitors but not sufficient to induce neuronal differentiation whereas Sox1 has the ability to induce neuronal fate acquisition in Xenopus ectodermal explants and mouse ESCs. However, Sox1 knockdown in mouse showed mild phenotype defects, indicating functional compensation from other SoxB1 members (Suter *et al*., 2009; Archer *et al*, 2011; Adikusuma *et al*., 2017). Sox1 is the earliest NE marker in mouse and its expression correlates with NP differentiation, thus it is broadly accepted that Sox1 plays a key role driving the transition from pluripotent cells to NE. However, it has been reported that Sox1 expression is not detected in the initial NE cells derived from human ESC revealing profound species-specific differences during NE differentiation (Zhang *et al*., 2010; Chung *et al*., 2012).

Subsequently, the TF Zic1, Zic3, Irx1-3 are up-regulated upstream of early bHLH proneural TFs. Zic1/Zic3 are both required to maintain neural progenitors in a proliferative state and for temporarily inhibit neural differentiation genes. Additionally, Zic and Irx genes inhibit expression of FoxD4 allowing posterior expression of neural differentiating genes. In Mouse, Zic3 is activated in the early gastrula NE at E6.75-E7.0 and might play a role in the formation of the primitive streak. Zic1 is enriched in the NE at stage E7.5-7.8

overlapping Gmnn and Sox3. Zic1 promotes neural progenitor differentiation; possibly in collaboration with Gmn and stabilizes neural fate by attenuating BMP and WNT (Aruga & Mikoshiba, 2011; Moody *et al*., 2013; Sankar *et al*., 2016). However, the mechanism by which Zic1/Zic3 TF promote neural progenitor maintenance and differentiation is not well understood (Inoue *et al*., 2007; Lee *et al*., 2014). Likewise, Irx1, Irx2 and Irx3 are activated during Drosophila, Xenopus and mouse NE differentiation prior to the expression of bHLH genes. Irx genes share 95% of amino acid identity and are highly conserved among vertebrates leading to the suggestion that Irx functions are redundant. Irx genes are expressed in overlapping and also unique spatial-temporal patterns during the development of the CNS where they are necessary for regional specification and patterning of neural progenitors, possibly by directly repressing BMP4 signaling (Cohen *et al*., 2000). Irx loss of function assays in Xenopus, result in posterior displacement of the forebrain and reduction of the midbrain and hindbrain demonstrating their requirement for proper anterior-posterior and dorso-ventral regionalization later in development. Additionally, studies have revealed that Irx3 and Sox3 interact directly by mutually repressing each other. However, the precise function of Irx TFs during NE cells regionalization is not well understood (Gomez- Skarmeta & Modolell, 2002; Rodriguez- Seguel *et al*., 2009; Sankar *et al*, 2016).

NE differentiation has been analysed in a reduced number of studies using mouse and Xenopus as animal models. However, limitations to acquire NE differentiating cells *in vivo* have prevented studing the NE development in humans. Hence, it has been demonstrated that the expression of regulatory markers such us PAX6, NESTIN, ZNF521 and N-cadherin is conserved between species during NE differentiation (Nandadasa *et al*., 2009; Park *et al*., 2010; Kamiya *et al*., 2011). However, specie specific differences in the temporary activation of these markers and regulatory signals unique for each species are largely

unknown. It has been shown that PAX6 and SOX1 are activated in a different temporal manner during NE differentiation between human and mouse. For instance, in human, PAX6 is the earliest NE marker identified in ESCs undergoing neural induction, whereas in the mouse embryo the paired box TF Pax6 is expressed in specific NE regions only before the closure of the neural tube (NT) (Nat e*t al*., 2007; Suter *et al*., 2009). In mouse, Sox1 is the principal candidate for NE formation and fate commitment while Pax6 plays a later role in differentiation of radial glial cells (Figure 1.7). Furthermore, PAX6 is sufficient and necessary for neural induction in human but not in mouse. PAX6 over-expression induces hESCs to adopt a NE fate whereas PAX6 knockdown fail to differentiate into NE even in the presence of dual SMAD inhibition (2i), conditions that favour neural differentiation. It has been suggested that the PAX6 neural inductive effect is achieved by repressing pluripotency genes such as OCT4, NANOG and MYC. However, OCT4, NANOG and SOX2 knockdown in ESC induces the cells to acquire trophoblastic fate. Hence, inhibition of pluripotency genes is a pre-requisite but not sufficient for NE differentiation. It has been proposed that the transition between ESC and NE cells is initiated by pluripotency inhibition and is potentiated by PAX6 activation of NE and neuronal progenitor genes including SIX3, LHX2, FGF8, NR2F2, TBR2 and WNT5b (Zhang *et al*., 2010; Blake & Ziman, 2014). In the mouse embryo, Pax6 and Sox2 occupy several common promoters including Nestin. Hence, these two TF possibly cooperate during neural induction of hESCs. Additionally, mouse studies have shown that PAX6 is required to maintain NP differentiation by inhibiting the expression of genes that promote mesodermal and endodermal differentiation such as Brachiury, Hnf1a and Myf (Thakurela *et al*., 2016). However, the transcriptional regulatory events involved in human NE commitment remain to be investigated.

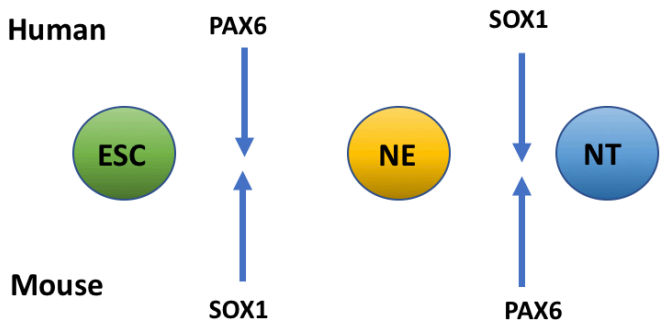



**Figure 1.7. Differential expression of PAX6 and SOX1 in human and mouse.** In humans, PAX6 activated during NE differentiation from ESC (green) and is considered the earliest neural marker, whereas SOX1 is first detected in the neural tube in foetuses (Zhang *et al*., 2010). In the mouse embryo, SOX1 favours the formation and maintenance of the NE, whereas PAX6 is detected in the neural progenitors of the differencing forebrain before the closure of the NT (Suter *et al*., 2009).

### 1.1.5. Additional neuroectodermal markers

Nestin is an intermediate filament (IF) protein type IV that together with microtubules and microfilaments constitutes the cytoskeleton. During the development of the CNS different IF types are express in a cell specific manner, marking distinct differentiation steps. Nestin expression initiates during NE commitment and is dramatically reduced during the transition from neural progenitor cells to post-mitotic cells. Hence, Nestin has been widely employed as a NE marker (Wiese *et al*., 2004; Mahler & Driever, 2007). A recent study conducted with mouse neurospheres demonstrated that Nestin knockdown results in inhibition of cell proliferation and reduction of EGFR signalling, indicating that Nestin in collaboration with EGFR signalling regulate cell proliferation in mouse neurospheres (Hu *et al*., 2016). Furthermore, Nestin deficient mice show a reduction of the neural tube cell number caused by increased apoptosis and reduced self-renewal ability, leading to embryonic lethality. However, microarray data analyses depicted similar gene expression patterns between the mutants and controls. Additionally, in the mutant mouse the

microtubules, microfilaments and IF networks structure and integrity was not affected. Thus, Nestin may exert its regulatory effects through interactions with other non-structural proteins, rather than through its role in cytoskeleton maintenance. For instance, Nestin knockdown in rat neuronal progenitors induces Cdk5 mediated apoptosis under oxidative stress. Hence, it has been proposed that Nestin serves as a scaffold to directly bind and sequester Cdk5 to regulate its activity (Park *et al*., 2010).

Another factor that plays a key role during NE differentiation is ZEB2, a member of the ZHFX1 family that directly interacts with activated SMAD 1-3 proteins, ZEB2 knockdown hESCs initially differentiate into NE cells but eventually fail to express early and late neural markers, whereas ZEB2 over expression enhances NE differentiation. Thus, ZEB2 is necessary for NE progression and maintenance but is not required to initiate NE commitment. Additionally, ZEB2 is required to maintain NE fate commitment over mesendoderm differentiation. ZEB2 antagonizes the expression of Activin/Nodal and BMP signaling by direct interaction with SMAD proteins (Conidi *et al*., 2012; Tang *et al*., 2015). Activin/Nodal signaling maintains pluripotency in mouse Epi and hESCs thought direct up-regulation of NANOG mediated by its intracellular effectors SMAD2/3. In return SMAD, NANOG and OCT4 inhibit expression of ZEB2 and prevent differentiation of hESCs. Concomitantly, ZEB2 is directly up-regulated by SOX2 during this stage. Thus, ZEB2 is in parallel up-regulated and inhibited at the same time, resulting in reduced levels of ZEB2 expression, which are sufficient to reduce ability of SMAD to induce mesendodermal fate. Moreover, studies have proposed that NE differentiation is triggered as Nodal, Lefty and Cerberus inhibit SMAD from the node in the embryo Epi reducing the expression of NANOG and OCT4, which results in an increase in expression of ZEB2 (Chng *et al*., 2010). Another member of the Zeb family, the TF Zeb1 is up-regulated in proliferative neural progenitor cells and down-regulated in post-mitotic neurons. Zeb1

plays an important role during mesenchymal-ephitelial transition by repressing E-cadherin expression and concomitantly inhibiting epithelial markers (Zhang *et al*., 2015; Aiger *et al*., 2017).

Likewise, the zinc-finger protein Znf521 is considered to play a role during NE differentiation as its expression pattern correlates with NE fate acquisition. In the mouse embryo, Znf521 is absent in the ICM and is strongly up-regulated at E7.0-E8.0 in the NE. As gastrulation proceeds Znf521 expression is preferentially located in the rostral neural tube. Overexpression assays in mESCs demonstrated that Znf521 is sufficient to induce expression of NE markers even in presence of BMP4, whereas Znf521 down-regulation resulted in sustained expression of Epi markers and lack of NE markers. NE development was rescued by Znf521 induced expression, demonstrating that Znf521 is necessary and sufficient for the Epi-NE transition (Kamiya *et al*., 2011; Shen *et al*., 2011). Furthermore, chromatin immunoprecipitation (ChIP) assays depicted enrichment of Znf521 factor in the DNA regions of Sox3, Sox1 and Pax6, indicating possible direct regulation of the NE markers by Znf521. Additionally, mouse and human fibroblast transfected with a Znf521 expressing lentivirus have been successfully induced to adopt rostral NSC characteristics. The reprogramed cells express NE markers such as Sox1, Sox2 and Pax6, were stable up to 60 passages and have the ability to differentiate into neurons and astrocytes (Shahbazi *et al*., 2016b). Kamiya *et al*, reported lethality in Znf521 mouse deficient embryo. However, a different study indicated that the deficient embryos formed NE and developed a CNS, suggesting the possibility of functional compensation by another gene *in vivo*. However, the mice showed reduced number of Sox1 positive cells, schizophrenic like behaviour and die after 10 weeks of birth. Together these studies demonstrated that Znf521 is necessary for proper neuronal differentiation (Ohkubo *et al*., 2014).

Finally, N- cadherins are highly up-regulated at NE stage and are commonly used as NE markers. Cadherins are cell-cell adhesion molecules necessary for a number of development processes. Initially the ectoderm express E-cadherins and they are progressively replaced by N-cadherins during NE differentiation while the non-neural ectoderm retains E-cadherin expression. Trans-membrane N-cadherins attach to cortical actin on the cell to maintain adherent junctions necessary for intracellular adhesion in NE cells (Kadowaki *et al*., 2006; Nandadasa *et al*., 2009). Down-regulation of N-cadherin in mouse ESC-derived NE cells caused disruption of the normal architecture of the NE cells and inhibition of rosette like structure formation. In Xenopus neural progenitors N-cadherin down-regulation results in loss of movements of the neural folds. Hence, N-cadherin plays a pivotal role maintaining the normal cellular architecture and movement during NE development (Nandadasa *et al*., 2009; Su *et al*., 2013).

### 1.1.6. Neuroectoderm specification

Neuronal progenitors or neural stem cells are generated from NE and undergo a limited number of symmetrical divisions and then asymmetrical divisions during neural differentiation (Dhanesh *et al*., 2016). As development proceeds, neural progenitors acquire some glial features, and become elongated to form radial glial cells (RG) with the cell body at the ventricular surface of the neural tube and radial fibers extending to the pial surface around E9-E10 (Bergstrom & Forsberg-Nilsson, 2012; Okano & Temple, 2016). Multipotent RG divide asymmetrically generating one basal progenitor cell and an immature neuron (intermediate progenitor cell). Subsequently, immature neurons migrate along the radial fibers to the cortical zone and become deep layer and later superficial layer neurons. Basal progenitors constitute transit amplifying cells that continue producing neurons from the subventricular zone. After generation of neurons RG cell subsequently

give rise to oligodendrocytes, astrocytes and ependymal cells (Kageyama *et al*., 2014; Imayosgi *et al*., 2015) (Figure 1. 8).

Members of the bHLH family, Hes1, Hes5, Ascl1, Neurog2, Olig1 and Olig2 play an important role in maintaining neural progenitors and regulating fate choices. Hes1 and Hes5 are transcriptional repressors that promote neural progenitor self-renewal and inhibit expression of proneural genes. Ascl1 and Neurog2 are the best-studied proneural activators that promote neural fate commitment and neuronal subtype specification. Additionally, Olig1 and Olig2 regulate oligodendrocyte differentiation (Imayosgi *et al*., 2015).

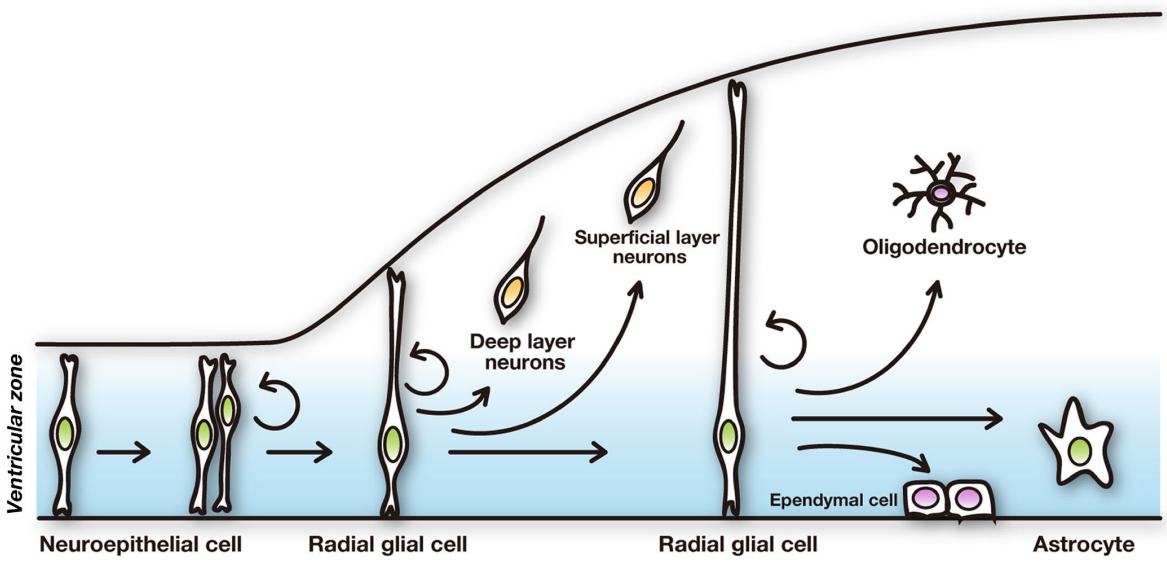



**Figure 1. 8. Differentiation of neuroectoderm cells in the embryonic brain.** NE cells divide symmetrically to form the NP and thicken the NT. NE cells acquire regional identity and become neural progenitors that initially renew by symmetrical division. Neural progenitors from the ventricular zone elongate to the apical zone and differentiate into RG cells. RG cells divide asymmetrically to give rise to an immature neuron that migrates to the cortical zone and differentiate into a deep layer neuron and then subsequently generate superficial layer neurons and basal progenitors. Basal progenitors proliferate in the subventricular zone to generate more neurons. Later dividing RG cells differentiate into oligodendrocytes, astrocytes and ependymal cells (Kageyama *et al*., 2014).

Hes1 and Hes5 are down-stream effectors of Notch, a trans-membrane protein that regulates neural progenitor cell self-renewal and inhibits neural differentiation. Proneural TF induce the expression of Notch ligands Delta and Jagged from neighboring differentiating neurons. After activation Notch undergoes several cleavages and releases its intracellular domain (ICD). The ICD is transported to the nucleus and binds Rbpj DNA binding protein and the co-activator Mam1. The complex activates the expression of Hes1 and Hes5, which inhibits neural differentiation and maintains RG cells. Hence, Notch signalling regulates neural differentiation of neighboring cells trough cell-cell lateral inhibition (Kageyama *et al*., 2005; Dhanesh *et al*., 2016). Hes1 and Hes5 have a WRPW domain that interacts with the co-factor TLE/Groucho to recruit histone deacetylases and repress the expression of target proneural genes including Ascl1 and Neurog2. For instance, Hes1 binds to the Ascl1 promoter to directly inhibit its expression and regulates Ascl1 activity by binding E47 an Ascl1 co-activator to inhibit their protein-protein interactions. Later in development Hes1 expression induces acquisition of astrocyte fate. In contrast, another bHLH family member Hes6 promotes neuronal differentiation by binding Hes1 and inhibiting its activity. Proneural bHLH TFs antagonize Hes1 by activating Hes6 expression (Kageyama *et al*., 2005; Imayosgi *et al*., 2015).

The bHLH activator type TFs such as Ascl1 and Neurog2 control early and late neurogenesis by regulating expression of pan-neural genes and expression of genes involved in neuronal commitment (Kageyama *et al*., 2015). Proneural genes Ascl1 and Neurog2 form a heterodimer with the co-factor E47 to regulate the expression of target genes. Particularly, Ascl1 regulates the expression of several TFs involved in differentiation of neurons, including Sox4, Gli3 and Dlx2. Additionally, Ascl1 activates expression of genes involved in cell cycle arrest such as Fbxw7, Gadd45g and Ccng2 and cell cycle progression such as Cdk1, Cdk2, and Skp2. Thus, Ascl1 regulates opposing

functions, inducing both cell cycle exit and proliferation. Furthermore, Ascl1 is sufficient to induce mouse and human ESCs to differentiate into neurons and to reprogram human fibroblast to acquire neuronal characteristics (Chanda *et al*., 2014). Interestingly, Ascl1 promotes chromatin accessibility to its own target genes regulating temporal progression of its transcriptional program (Castro *et al*., 2011; Raposo *et al*., 2015). Additionally, proneural genes inhibit the expression of glial specific genes such as GFAP by sequestering their activators Smad1-p300 complex from glial promoters (Kageyama *et al*., 2005).

bHLH TFs are broadly expressed in the population of actively dividing neural progenitors, overlapping the expression of NE markers like Sox2. However, Sox2 is homogeneously expressed, whereas bHLH TFs expression is variable among the cells. Imayoshi and colleagues found that Hes1, Ascl1, Ngn2 and Olig2 are expressed in an oscillatory manner in early neural progenitors differentiated from mouse ESCs. Hes1 and Ascl1 proteins expression oscillate every 2-3h while Olig2 proteins oscillate in a 5-8h period. It has been proposed that Hes1 oscillations initiate as Notch is activated and induce up-regulation of Hes1. Subsequently, Hes1 binds and repress its promoter inhibiting its own expression. Hes1 protein and mRNA are highly unstable and rapidly disappear leading to another expression cycle (Shimojo *et al*., 2008). Hes1 oscillatory pattern is in opposite phase to oscillations of Ascl1, Neurog2 and its ligand Delta-like1 (Dll1). Thus, is possible that Hes1 induces Ascl1, Neurog2 and Dll1 oscillations by periodically inhibiting their expression. Subsequently, during neural differentiation one of the three bHLH TFs becomes expressed in a sustained manner while the others are down-regulated. Hence, multipotent actively dividing neural progenitors express various bHLH TFs in an oscillatory manner, while sustained expression of a specific factor drives commitment to a particular fate. It has been suggested that transient inhibition of Hes1 and transient up-regulation Ascl1 or Neurog2 before cell division bias neural progenitors to acquire neuronal fate. Neuronal fate is then

determined by a sustained expression of Ascl1 or Neurog2 and down-regulation of Hes1. On the other hand, up-regulation of Hes1 and Olig2 bias the neural progenitors to commit into astrocytes and oligodendrocytes fate respectively (Kageyama *et al*., 2014; Imayosgi *et al*., 2015).

### 1.1.7. Formation of the neural border zone

During gastrulation the ectoderm is regionalized into neural, non-neural ectoderm and the neural border zone (NBZ) between the two regions. Interactions among neural and non-neural ectoderm and signals from the underlying mesoderm and endoderm induce the differentiation of the NBZ, which will give rise to the neural crest (NC) and the pre-placodal ectoderm (PPE) (Saint-Jeannet & Moody, 2014; Shigetani *et al*., 2016). Members of the Dlx, Msx, Pax and Zic families and TFAP2α, Gata and Foxi1 genes are widely expressed in the non-NE and overlap the expression of NE TFs SoxB1 and Zic in prospective NBZ (Dincer *et al*., 2014; Moody & LaMantia, 2015; Shigetani *et al*., 2016). As the NE cells divide to form the NP, SoxB1 TFs become confined to the NP, Zic members are confined to the NP and the medial NBZ, Dlx, Gata, Foxi1 and TFAP2α are restricted to the non-NE and Msx1 and Pax3 are expressed only in the NBZ.

Dlx genes are required to promote the expression of NC and PPE genes and to inhibit the expression of NE genes. Dlx TF interact with Msx TF by inhibiting each other's expression. Differential expression of these genes bias cell fate towards NC when Msx is up-regulated and Dlx expression levels are low and towards PPE when Dls is up-regulated and Msx levels are low (Phillips *et al*., 2006; Moody & LaMantia, 2015). Dlx5 activates expression of the PPE marker Six1, while Msx1 represses it. Msx1 induces expression of Pax3 and Zic1, which interact to activate NC genes. Alternatively, lack of expression of Pax3 and

expression of Zic1 induce the differentiation of the PPE. Foxi1 and TFAP2α genes are also necessary for NBZ formation. In Xenopus explants, Foxi1 loss of function leads to reduction in expression of Dlx, Six1 and FoxD3 (NC marker) and expansion of the Sox2 positive NP. Likewise, inhibition of TFAP2α results in down-regulation of Msx1, Pax3, FoxD3, Six1 and Eya1 (PPE marker). Thus, maintained expression and cross-regulation between Dlx, GATA, Foxi and TFAP2α are required for the formation of the NBZ, the PPE and the placodes (Sato *et al.*, 2005; Leung *et al.*, 2016 Moody & LaMantia, 2015) (Figure. 1.9).

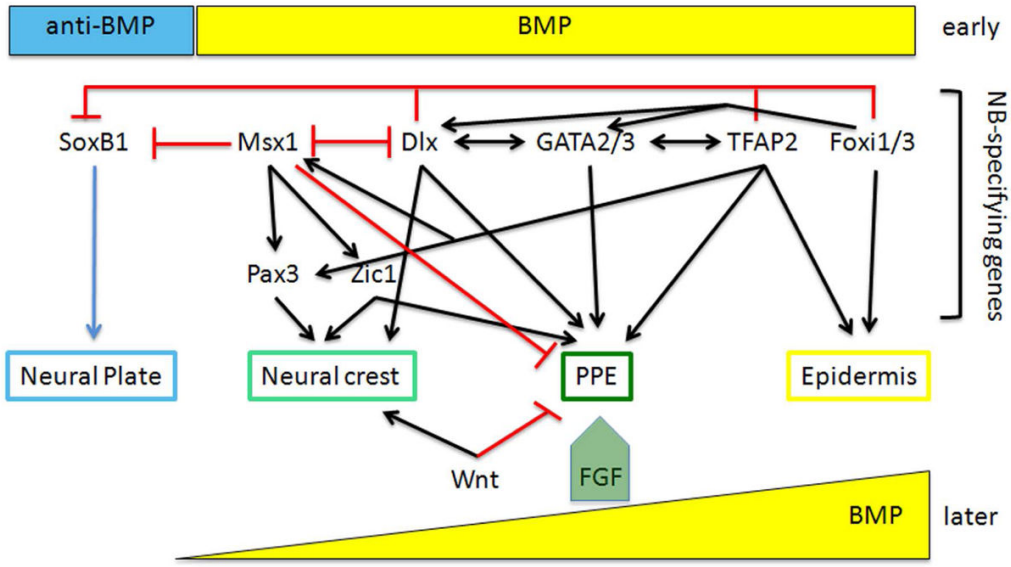



**Figure 1.9. Scheme of the genetic interactions during the formation of the ectoderm sub-domains.** During gastrula stage, inhibition of BMP signaling allows activation of NE genes such us SoxB1 TF. NE cells proliferate to form the NP (blue). High levels of BMP promote the expression of several NBZ genes like Msxl1, Dlx, Gata1/3, Pax, Zic, Tfap2 and Foxi1/3, which interact with each other to repress NE TF (red edges) and to activate and maintain the NBZ genes (black edges). Once the NBZ has been formed, BMP is down-regulated to allow expression of the NC and the PPE, whereas high levels of BMP are required for epidermis differentiation. Differential expression of NBZ genes bias NBZ to differentiate into NC or PPE. The PPE is specified by expression of Dlx, Zic1, Gata and Tfap2 TF and down-regulation of Msx1 and Pax3. Additionally, low levels of FGF signaling from the underlying mesoderm and inhibition of WNT are required for PPE formation. The NC is differentiated by expression of Msx1, Pax3, Zic1 and WNT signaling and requires down-regulation of Dlx (Moody & LaMantia, 2015).

### 1.1.8. Differentiation of the placodes

Once the NBZ has been formed, the PPE regulators Six and Eya are activated. Additionally, low levels of FGF signaling from the underlying mesoderm are required for PPE formation. FGF signals inhibit expression of PPE repressing TFs and induce expression of PPE markers. However, the underlying mechanisms by which FGF genes regulate the expression of PPE inhibitors and markers remain to be elucidated. Similarly, BMP signaling is necessary for the differentiation of the NBZ and it must be reduced to permit formation of the PPE (Grocott *et al*., 2012; Moody & LaMantia, 2015; Shigetani *et al*., 2016). Another signal that needs to be reduced to allow PPE differentiation is WNT, which directly represses genes that promote PPE fate. Hence, the PPE is restricted to the NBZ surrounding the anterior neural plate where WNT is inhibited (Leung *et al*., 2016). In chick, PPE genes are activated in the NBZ in presence of low Fgf8 signaling accompanied with low BMP and WNT levels, whereas if WNT signal persists NC markers are activated (Moody & LaMantia, 2015; Shigetani *et al*., 2016). Additionally, retinoid acid (RA) is expressed in a U-shaped area similar to the PPE in the anterior NBZ. Attenuation of RA signaling increases the posterior limit of the PPE, suggesting a role in the control of the PPE boundary possibly through regulation of RA target genes Tbx1 and Ripply3. In regions where both targets are expressed, PPE genes are inhibited, while if only Tbx1 is activated PPE genes are induced. Thus, RA might restrict the PPE limits by inducing the expression of both targets in the posterior ectoderm (Janesick *et al*., 2012; Moody & LaMantia, 2015).

Six1, Six2 and Six4 are highly expressed in the PPE domains and have been implicated in differentiation of the PPE and placodes. Six1 plays a central role in PPE development, since loss of function studies have demonstrated reduction in the expression of other PPE markers in Xenopus and Drosophila. Six1 loss in mouse affects the formation of the

olfactory placode but has no effect on formation of the PPE, possibly due to compensation from another factor (Chen *et al*., 2009; Grocott *et al*., 2012; Moody & LaMantia, 2015). However, induction of Six1 expression is not sufficient to induce PPE markers outside of the NBZ. Additionally, Six2 and Six4 are widely used as PPE markers, but their regulatory mechanism has not yet been described. Six activities can be modulated by binding of co-factors that lack a DNA binding domain (Moody & LaMantia, 2015).

The TF Eya1 and Eya2 are among the best-studied Six co-factors. Eya1 and Eya2 expression patterns resemble the expression of Six1 in Xenopus, suggesting that they possibly have roles driving the differentiation of the PPE and the placodes. Furthermore, Eya1 directly binds Six1 and possibly other TFs to translocate into the nucleus and regulate expression of PPE target genes. Eya loss of function in Xenopus leads to abnormalities in the placode derivatives but the effects on PPE formation have not been addressed. Six1 can also function as a repressor when it interacts with other TFs such as Groucho to inhibit expression of epidermal and NC genes (Silver *et al*., 2003; Chen *et al* 2009; Moody & LaMantia, 2015). Several other genes that might bind to Six1 during PPE and placode development have been identified, but their regulatory functions and interactions have not yet been delineated. PPE transcriptional regulators including Six and Eya genes set the transcriptional landscape for the development of the distinct placodes and maintain the boundaries between PPE and the adjacent NC and non-NE. However, the molecular pathways that establish and maintain the ectoderm sub-domains are not fully understood (Moody & LaMantia, 2015).

Initially, the PPE has the ability to differentiate into any placode and later acquires anterior-posterior (A-P) patterning in response to local signaling from the mesoderm and ectoderm. The anterior PPE will give rise to the adenohypophyseal, olfactory and lens

placodes, the middle PPE will become the trigeminal placode and the posterior PPE will differentiate into otic and epibranchial placodes. In vertebrates, the placodes will give rise to the majority of the peripheral sensory nervous system (Grocott *et al*., 2012). It is generally accepted that BMP, FGF, WNT and RA contribute to specification of placode identity. Differential expression of Six TFs expression might also contribute to the formation of the A-P patterning. Additionally, the TF Otx and Gbx2 are differentially expressed in the A-P domains, respectively, and might play a role in the differentiation of the PPE (Steventon *et al*., 2012). As a result of the A-P regionalization, different Pax members are induced in specific regions. Pax6 is expressed in the anterior placodes, Pax3 in the ophthalmic placode whereas Pax2 and Pax8 are activated in the posterior placodes. Additionally, genes broadly expressed in the PPE become restricted to specific areas: Six3, Six6, Pitx3 are expressed in the most anterior placodes and Dlx, Irx, Tbx, Foxi3 in the posterior placodes (Grocott *et al*., 2012; Moody & LaMantia, 2015). Subsequently, specific combinations of additional genes are activated in each placode such as Emx2 and Ebf2 in the olfactory placode. Furthermore, genes that were expressed in the PPE are segregated to specific areas. Pax6 and Dlx5 are restricted to mutually exclusive domains comprising cells fated to the lens placode and to the olfactory placode, respectively. The transcriptional networks implicated in the differentiation of each placode will subsequently drive differentiation of specific neurons such as GnRH, which are derived from the olfactory placode (OP) and subsequently migrate to the hypothalamus during early embryogenesis (Provenzano *et al*., 2010; Moody & LaMantia, 2015).

## 1.2. Induced pluripotent stem Cells

The initial steps of mammalian development have been studied with detail up to the point of implantation primarily based on data obtained from the mouse model. A recent study has reported in vitro implantation platforms for mouse and human that allowed insight into human early development. Many key events of development were recapitulated *in vitro* and expression of cell-type specific markers for the ICM, Epi, PE and TE in pre and post implantation embryos were analyzed and compared with those occurring in mouse. The study ended at 14 dpf or before the initiation of the gastrulation according to the international bioethical guidelines (Bedzhov *et al*., 2015; Deglincerti *et al*., 2016). Such studies evidenced the profound differences in architecture, cell diversity and tissue organization between mouse and human, highlighting the necessity to develop an *in vitro* system to study human development. The earliest step of the human CNS differentiation is the segregation of the NE from non-NE during gastrulation. However, the inability to sample live human live cells during this developmental stage limits our knowledge about the key TFs and developmental programs that govern human NE fate acquisition and differentiation and early aberrant regulatory events that might be involved in neurodevelopmental disorders (Marchetto *et al*., 2010).

Spemann and Mangold first discovered neural induction when they reported that NE was formed from non-NE cells exposed to signals from the dorsally specified mesoderm in the amphibian embryo. Over the subsequent decades the molecular interactions and TFs that drive transition from pluripotent Epi cells to NE fated cells have been investigated. The use of Xenopus and mouse systems has led to the identification of a number of TFs (Moody *et al*., 2013). However, the human and mouse brains have profound differences, most notably their size and complexity. Humans have a larger and more complex brain, which is

accompanied by an elongated period of brain development (van de Leemput *et al*., 2014). Disparities are also evident at a molecular level, such as the distinct times of expression of the NE markers Pax6 and Sox1 during NP differentiation (Suter *et al*., 2009). Hence, the gene regulatory events that drive fate decisions during early neuronal development and particular cues and cell types that might be unique to humans are not fully recapitulated in the mouse model (La Manno *et al*., 2016; Yao *et al*., 2017). Thus, the TFs and genetic pathways involved in the inhibition of the pluripotency gene network and the activation of the NE program in humans remains largely unknown.

ESCs and iPSC provide powerful *in vitro* system to analyze the TF and regulatory events that orchestrate human development and have an enormous potential for therapeutic applications. ESCs are derived from the ICM of the blastocyst, have competency to differentiate into all cells from the three germ layers and are capable of indefinite replication (Takahashi & Yamanaka, 2006). Human ESCs are similar to EpiESCs in their developmental potential. ESCs have the ability to replicate the progression during neuronal development from pluripotency to functional neurons. Differentiating ESCs express dynamic genetic programs that regulate pluripotency maintenance networks and activation of the TFs required for differentiation into specific cell lineages (Ardhanareeswaran *et al*., 2017). Loss and gain of function approaches have been widely used to analyze the role of particular TFs during directed ESC differentiation to a specific lineage. Hence, ESCs provide an opportunity to identify specific TF that might be involved in human embryonic development *in vitro* (Hong *et al*., 2016). However, embryos with known genetic defects are rarely accessible for the analyses of neurodevelopmental disorders and ESCs are not optimal for tissue transplantation in patients due to the possibility of rejection (Shi *et al*., 2016). Importantly, the use of human embryos for research purposes has been the center of several ethical and political controversies. iPSCs are capable of differentiating into any cell

type desired, have unlimited capacity to replicate and overcome many of the ethical concerns associated with ESCs.

In 2006, Takahashi and Yamanaka reported that somatic cells could be reprogramed to acquire gene expression profiles and differentiation potential similar than ESC by the introduction of Oct4, Sox2, Klf4 and Myc. The TF Oct4 and Sox2 were previously reported to play a role in the maintenance of pluripotency in the ICM and in ESCs Klf4 and Myc were thought to enhance ESC proliferation and to maintain their phenotype (Takahashi & Yamanaka, 2006). Yamanaka and colleagues used retroviral integration systems to reprogram mouse fibroblasts and a year later they reported the reprograming of human fibroblast. iPSCs technology has vastly improved since then, and in order to avoid integration, different delivery methods are widely used, including adenovirus, Seindai virus, small molecules and RNAs, among others. These advances make iPSC technology more suitable to analyze human development, to model diseases, to test drug toxicity and particularly for patient specific cell therapy (Suh, 2017).

The use of iPSC technology makes possible to obtain large amounts of different relevant cells that were previously unavailable, such as cells in different developmental stages including iPSC-derived neuro-ectodermal stem cells (NESCs). iPSCs can be obtained from easily accessible somatic cell types and from individuals with different genetic background including patients with a known disease phenotype. Patient specific cells with a mutation or multiple mutations can be derived and differentiated into disease relevant cell types to identify the pathological mechanism and the disease etiology (Shi *et al*., 2016; Suh, 2017). Furthermore, the development of gene editing technologies such as CRISPR–Cas9 enables introduction of disease-causing mutations in control iPSC and to remove disease-associated mutations from patient's derived iPSC. Comparing the mutated iPSC with their isogenic

controls allows determining the possible phenotypic effect of disease-associated genes (Shi *et al*., 2016) (Figure 1.10).

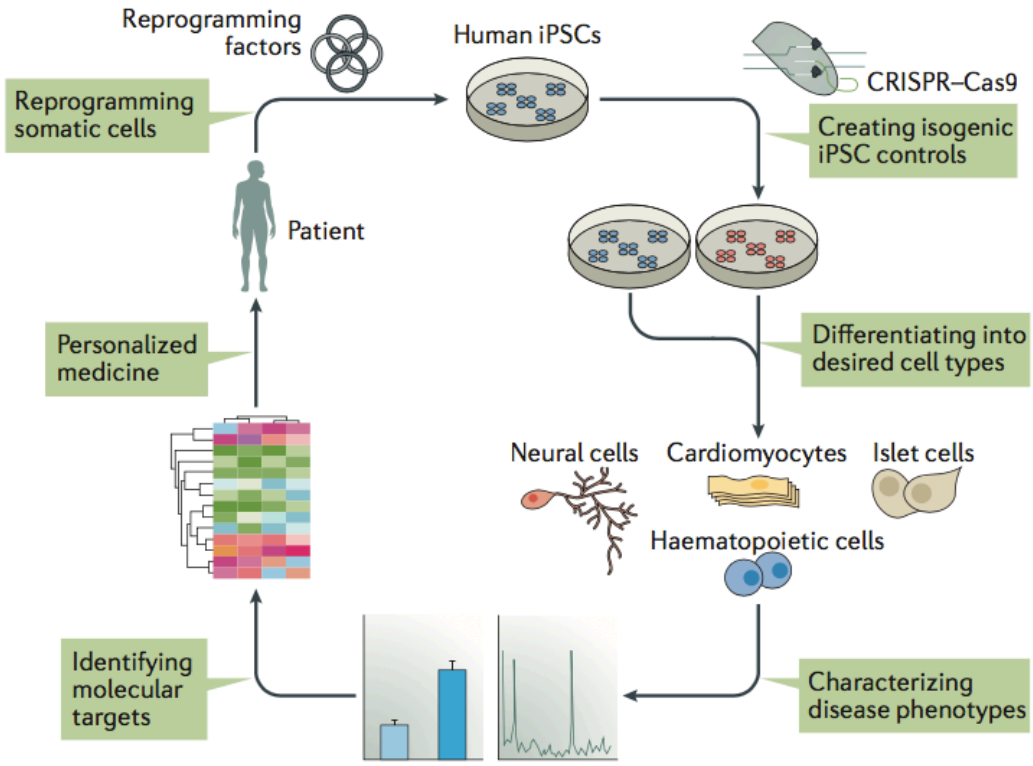



**Figure 1.10. Human disease modeling scheme.** Human somatic cells are harvested from patients and are reprogramed by specific TF. An isogenic control is generated by CRISPR-Cas9 genome editing technology. Derived iPSCs are induced to differentiate into specific cell types. Differentiated cells are used to analyze and identify specific gene expression patterns and cell phenotypes relevant for the disease (Shi *et al*., 2016).

iPSC reprograming is recognized as a major epigenetic remodeling process and it is necessary to adjust the epigenetic state of the parental cell to a state compatible with pluripotency (Gao *et al*., 2017). However, many epigenetic marks in iPSCs have being found to differ from those in ESCs. These differences represent epigenetic memory, which may have a negative effect in the reprograming efficiency and capacity to differentiate into specific lineages (O'Malley *et al*., 2013; Brix *et al*., 2015). However, minor epigenetic and transcriptional memory and unbiased differentiation potential have being shown by genome

wide analyses performed with iPSCs derived from different somatic tissues from multiple individuals (Gao *et al*., 2017). Such studies have also demonstrated that the major source of variability between iPSC lines is the donor's genetic background. For instance, genetic and epigenetic variability between iPSC and ESC might reflect the differences in the genetic background of the individual donors. It has been reported that epigenetic memory effects on gene transcription are reduced over multiple passages. Hence, the major cause of heterogeneity among iPSCs is the genetic differences between the donors, while the epigenetic memory or intrinsic variability of the iPSC system have minimal contributions. For experimental procedures, the incorporation of iPSC derived from multiple individuals is recommended to average and separate the genetic background effect of each cell line from global transcription (Rouhani *et al*., 2014; Suh, 2017). Additionally, incomplete silencing of the reprograming TFs has being a concern about using the iPSC system to model development and disease. However, several gene expression analyses have indicated that the reprogramming transgenes are inactivated during propagation (Xia *et al*., 2007; Nakanishi & Otsu, 2012).

Currently, iPSCs and ESCs are broadly used as *in vitro* systems to gain insight into human neuronal development and to study neurodevelopmental and neurodegenerative conditions such as autistic spectrum disorder, Rett syndrome, Alzheimer's disease and Parkinson's disease (Begum *et al*., 2015). Such studies aim to recapitulate the multistep process of neuronal differentiation to confirm the identity of the TFs and genes involved in normal and aberrant development and to reveal novel TF that might play a role during differentiation. ESC and iPSC neural induction can be achieved in monolayer cultures, which are a reproducible and efficient means of obtaining a large number of highly homogeneous neural progenitors cells. Neural progenitors can then be differentiated into various neuronal subtypes such us GABAergic, glutamatergic and dopaminergic neurons,

whose morphology is relatively easy to analyze in adherent cultures. For instance, RG cells derived from ESC and iPSC display molecular and morphological characteristics of human embryonic RG and follow the RG cell intrinsic mechanism to differentiate first into cortical neurons and subsequently into late progenitors that can potentially differentiate into astrocytes (Figure 1.11) (Neely *et al*., 2012; Duan *et al*., 2015; Ardhanareeswaran *et al.,* 2017).

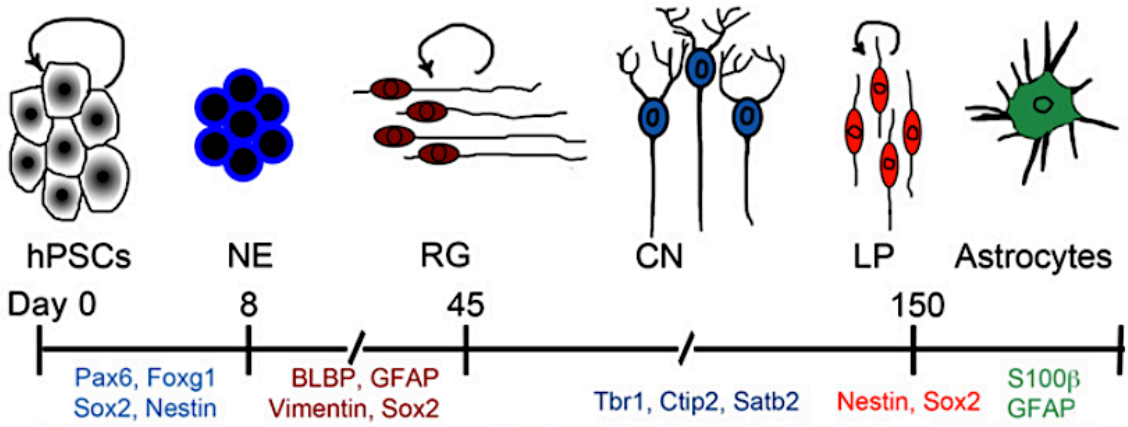



**Figure 1.11. Human pluripotent stem cell differentiation to cortical neurons and astrocytes**. Pluripotent stem cells are induced to differentiate into NE by treatment with 2i over 8 days. NE cells express Pax6, Foxg1, Sox2 and Nestin TF, which are commonly used as NE and forebrain markers. NE cells adopt RG characteristics and express BLBP, GFAP, Vimentin and Sox2 markers after approximately 12 days of culture without inhibitors. Depending on the culture conditions, RG cells differentiate into cortical neurons (CN) expressing Tbr1, Ctip2 and Satb2 TF around 45 days of differentiation. With further culture the remaining RG adopt late progenitor (LP) phenotype that generate astrocytes, positives for S100 $\beta$ and GFAP markers (adapted from Duan *et al*., 2015).

In Xenopus, BMP and TGFβ inhibition via Noggin, Chordin and Follistatin from the node is necessary for neural differentiation. Studies have shown that inhibition of BMP and TGFβ pathways (2i) leads to highly efficient conversion of ESCs and iPSCs to neural progenitors in adherent cultures (Muratore *et al*., 2014; Ardhanareeswaran *et al.,* 2017). SB431542 and LDN193189 small molecule inhibitors are widely used to inhibit TGFβ and BMP signaling respectively. SB431542 inhibits kinase activity of ALK4 ALK5 and ALK7

receptors, which are necessary for SMAD2 and SMAD3 phosphorylation whereas LDN193189 blocks SMAD1, SMAD5 and SMAD9 phosphorylation by inhibiting ALK1, ALK2, ALK3 and ALK6 receptors (van Caam *et al*., 2017). The potential mechanisms that contribute to the generation of highly pure neural progenitors population are the inhibition of trophoblast lineage that is otherwise promoted by BMP; suppression of mesodermal fate that requires NODAL/Activin and BMP signaling; and destabilization of NANOG-pluripotency program mediated NODAL/Activin inhibition (Chambers *et al*., 2009; Muratore *et al*., 2014). Down-regulation of NANOG via SMAD2/3 inhibition leads to the up-regulation of ZEB2, which promotes acquisition of NE fate. Likewise, down-regulation of OCT4 promotes expression of NR2F2, which induces NE markers expression. Additionally, it has being proposed that BMP inhibition promotes the expression of cell intrinsic NE determinants (Ozair *et al*., 2012).

iPSCs have similar gene expression patterns during neuronal differentiation as ESCs, with the same differentiation time and under the same cell culture conditions. In principle, ESCs and iPSCs differentiation programs resemble the embryo developmental program and underlying gene regulatory signals. Thus, the iPSCs model system is widely used to recapitulate the progression of neuronal development and to identify the transcriptional programs that govern neuronal differentiation (Hu *et al*., 2010).

## 1.3. Study aim and objectives

The overall aim of the study is to discover the gene expression profiles and the GRN tranversed during neural induction of human iPSCs to NESCs. Ethical, legal and technical limitations prevent study of *in vitro* development of the human embryo after 13 dpf or the formation of the primitive streak. Following formation of the primitive streak, neural induction of the dorsal ectoderm during gastrulation is the first step in the formation of the CNS. However, restricted access to NE cells has limited the ability to obtain suitable transcriptome data to enable identification of genes and pathways that drive transition from pluripotency to neural competence. iPSCs due to their molecular, morphological and functional similarity to ESC *in vivo* provide a unique experimental system to analyze the molecular events and the lineage-biased states that regulate human development. Several molecules implicated in the ectoderm neural induction have been described in animal models but fail to uncover specific human development landmarks due to species-specific differences.

Due to these limitations, a complete transcriptome analyses and a description of the GRNs that orchestrate NE differentiation in humans have not been reported. With this study we aim to provide a complete view of the genes that are differentially expressed during human NE induction from iPSCs. This approach allows us to validate previously described pathways and to detect novel genes with a potential role during NE differentiation. We aimed to use these data to construct a GRN, which topology would enable to identify the gene regulatory interactions at the initial stages of NE induction; and the potential biological processes that emerge from the specific gene expression patterns during this stage. In parallel, colleagues from the Price laboratory have reported the differentiation of cortical and GnRH neurons by using 2i and '2i plus WNT' inhibition differentiation

protocols, respectively. We aimed to identify the early molecular differences that might bias iPSC neural induction to cortical and GnRH fates.

The specific objectives of the project are:

1. To use iPSCs as an *in vitro* system to uncover gene expression profiles activated during human neural induction.

2. To identify the effect of WNT inhibition on gene expression profiles generated during neural induction.

3. To derive a GRN that comprises the first gene regulatory interactions underlying NE fate commitment.

4. To assess NESC heterogeneity thought single cell RNAseq analyses.

# CHAPTER 2. General Materials and Methods

## 2.1. Neural Induction

For the experiments in this study 3 iPSC lines and 1 ESC line provided by the Price laboratory were used. iPSC lines were derived from keratinocytes from neurotypic males from an age ranging between 35 to 55 years. The cells lines CTR M3 36S, CTR M2 42, CTR M1 04 are referred as iPSC1, iPSC2 and iPSC3 respectively, whereas the ESC SA001++ is referred as ESC. The iPSC lines were reprogramed by introducing C-MYC, KLF4, OCT4 and SOX2 TF in a polycistronic excisable vector. CTR M2 42 and CTR M1 04 lines were transformed with a Human STEMCCA Cre-Excisable Constitutive Polycistronic (OKSM) Lentivirus Reprogramming Kit (Millipore, SCR545) while CTR M3 36S line was transformed with CytoTune-iPS 2.0 Sendai Reprogramming Kit (Thermo Fisher, A16517).

ESC and iPSC plated in 6 well Nunc treated multidishes (Thermo Scientific, 140675) were washed once 1ml/well of HBSS (Life technologies, 14170-070) at room temperature. Cells were incubated for 5 min at with 1ml of Versene (Lonza, BE17-711E) for dissociation. The Versene was removed and cells were detached with a cell lifter and suspended on 1ml of E8 media (Life technologies, A1517001). Cells were counted with a TC10 automated cell counted (Bio-rad) and approximately 3,000,000 cells per well were seeded on Nunc multidishes with 3ml of E8. The dishes were coated previously to seeding with 1ml of 2% Geltrex (Life technologies, A1413302) in DMEM: F12 (Sigma; D6421) for 4 hours at 37°C. Cells were incubated at 37°C, 5% CO2 and 5% O2 for 48 hours. The cells were examined in a microscope for quality control and the media was replaced every 24 hours. After this period a cell monolayer reached approximately 90% confluence and media was

substituted with N2:B27 medium: 1:1 ratio of N2 supplement (Life Technologies, 17502-048) 1X final concentration in DMEM Medium (Sigma; D6421) + B27 supplement (Life Technologies; 17504-044) 1X final concentration in Neurobasal Medium (Life Technologies; 21103-049). The N2:B27 was used as base media and was supplemented with TGFβ and BMP small molecules inhibitors (2i) for neural induction and a second treatment with 2i plus WNT inhibition. For the treatments, either SB431542 (Sigma-Aldrich, S4317) 10µM and LDN193189 (Sigma-Aldrich, SML0559) 0.1µM or SB431542 10µM; Dorsomorphin (Sigma-Aldrich, F5499) 1µM and XAV939 (Sigma-Aldrich, X3004) 2µM were used. During neural induction cells were incubated at 37°C, 5% CO2 and 20% O2 and checked for quality control and media replacement every 24h for 8 days (Figure 2.1).

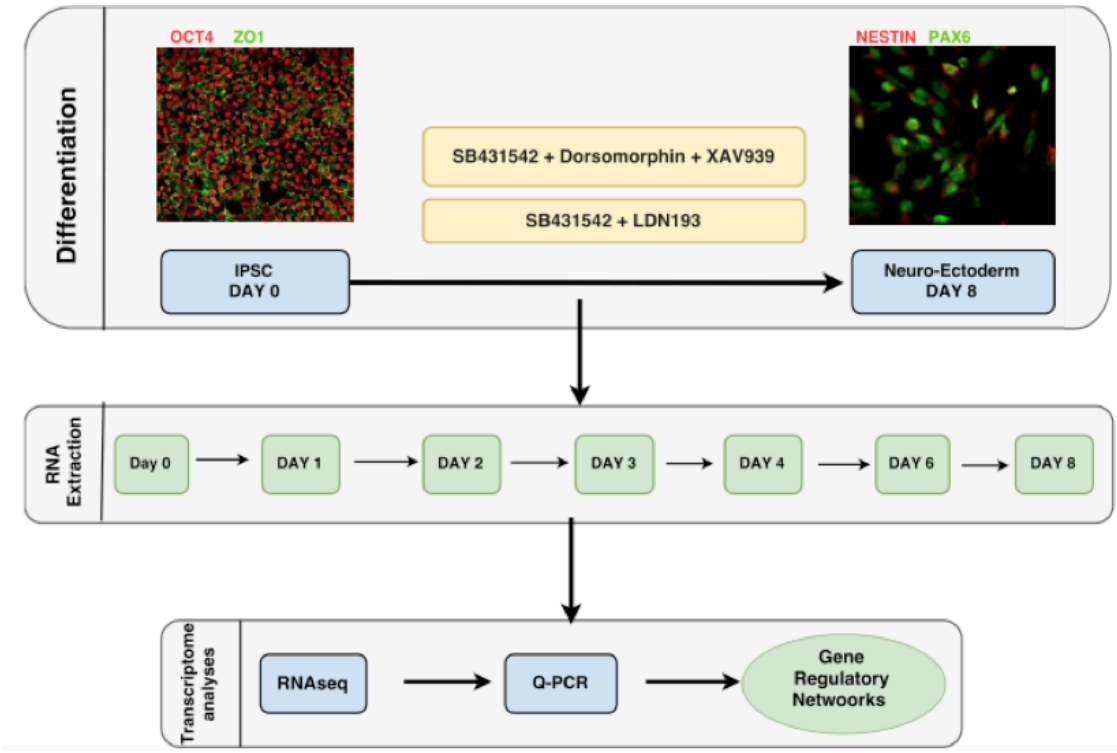



**Figure 2.1. Scheme of neuroectoderm differentiation, RNA extraction and transcriptome analyses.** iPSC and ESC lines were culture with 2i (SB431542, dorsomorphin and XAV939) or 2i-WNT (SB431542 and LDN193) media for 8 days. RNA was extracted at time points d0, d1, d2, d3, d4, d6 and d8 and used for transcriptome analysis through RNAseq, Q-PCRs and to model a GRN.

## 2.2. RNA extraction

RNA was extracted at 7-time points during the first 8 days of neural induction. The first time point was day 0 just before the addition of neural induction media. Subsequently, RNA was extracted at days 1, 2, 3, 4, 6 and 8 after neural induction. The media was removed, and the RNA was harvested by adding 1ml of Trizol reagent (Life technologies, 15596026) at room temperature to each 6 well. Cells were dislodged with a cell scraper, transferred to 1.5ml tubes and lysed by pipetting 15 times. The samples were stored at -80°C until RNA extraction.

RNA was isolated from each sample by centrifugation (12000g, 5 minutes) with 200µl of 100% Chloroform. The top aqueous layer was placed in new 1.5ml tube with 500µl of 100% isopropanol. Samples were mixed 10 times by pipetting and incubated 15 minutes at room temperature. After incubation, the RNA was precipitated by centrifugation (12000g, 15 minutes). The supernatant was removed, and the pellet was suspended in 1ml of 75% ethanol followed by centrifugation (12000g, 5 minutes). The ethanol was removed, and the pellet was air dried for 15 minutes at room temperature. The RNA was dissolved in 100µl of nuclease free water and purified with a Qiaquick RNA purification kit (Qiagen, 28106) according to manufacturer's specifications. RNA was quantified with a NanoDrop 1000 Spectrophotometer (Thermo scientific). Quality control was assessed by analysing each sample with the Agilent RNA 6000 nano Kit (Agilent technologies, 5067-1511) in combination with the Agilent 2100 Bioanalyzer system according to the manufacturer protocol to obtain the RNA integrity number (RIN) (Figure 2.2).

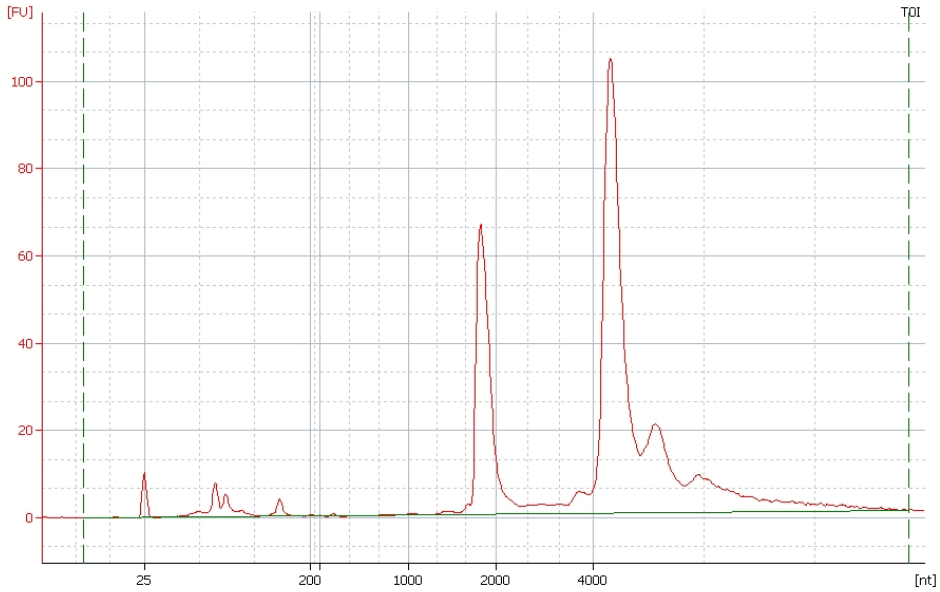



**Figure 2.2. Electropherogram of the RNA extraction from the iPSC1 line at the time point d0.** This figure shows an example of the Electropherograms obtained through the bionalayser RNA analyses with a RIN number of 10. The main peaks represent the 18S (left) and 28S (right) ribosomal subunits. The small peaks observed at the left represent 5S ribosomal and tRNA. The presence of these peaks is indicative of a good quality sample with no degradation.

## 2.3. RNAseq

RNAseq libraries were prepared with the Truseq RNA Kit (standard methods by The Wellcome Trust Centre for Human Genetics of Oxford University). Briefly, mRNA fraction was purified from total RNA followed by cDNA synthesis. Subsequently, the cDNA was end-paired, A-tailed and custom indexing adapters were ligated. Samples were size selected and pooled for sequencing.

Libraries were sequenced in collaboration with Larry Stanton and Steven Havlicek from the Genome Institute Singapore (GIS). Libraries were multiplexed at 8 samples per lane and sequenced in a Hiseq 2500 Sequencing System (Ilumina) to a depth of 20-30 million reads per sample. FASTQ files were aligned to a reference genome by our collaborators

Shankar Subramanian and Jun Min (UCSD). For additional statistical analyses I used the raw data obtained from our collaborators to identify differentially expressed genes and gene profiles through various "R" console programs and Prism platform. Detailed statistical analyses are described in the relevant chapters.

## 2.4. Gene ontology

Significantly DE genes that depicted ≥ 1 log2 fold changes between consecutive time points, between D0 versus D8 or between the two different neural induction treatments at any time point were used for gene ontology analyses (GO). Additionally, genes comprised in each of the gene expression profiles 22, 31, 32, 65 and 99 from a short time series expression miner (STEM) analyses were used for GO. Gene sets were interrogated for enrichment of specific gene classes and functional terms using DAVID Gene Functional Classification Tool. DE gene lists were tested against a reference gene set comprised of all expressed genes between the conditions analyzed.

## 2.5. cDNA synthesis

A set of 14 primers was designed with Primer3Plus (v. 0.4.0) software. Gene sequences were extracted from the NCBI GenBank and the University of California and Santa Cruz (UCSC) genome and bioinformatics browser (Table 1). Genes were selected on the basis of their known role during neural induction.

Complementary DNA (cDNA) was synthetized from 1μg of RNA using M-MLV Reverse Transcriptase enzyme (Promega, M1701) following manufacturer's instructions. Briefly, RNA was incubated with Oligo (dT) primers (Promega, C1101) to a final concentration of 0.02μg/μl for 5 minutes at 70$^{o}$C and then cooled on ice. The following reagents were added:

M-MLV buffer to a final concentration of 1.5X; 0.75mM of dNTP mix (Promega, U1511); 25 units of recombinant RNasin ribonuclease inhibitor (Promega, N2511); 200 units of M-MLV reverse transcriptase and nuclease free water to a final volume of 20µl. Reactions were carried out at 50ºC for 50 minutes. cDNA was diluted to a 1:5 ratio with nuclease free water.

### 2.6. Real-time Polymerase Chain Reaction assays

Real-time polymerase chain reaction (Q-PCR) assays were conducted by using 4µl of cDNA per reaction and adding the primers mix to a final concentration of 0.3µM and iQ Sybr Green supermix (Bio-rad, 178880) to a 1X final concentration. Amplifications were performed in a Bio-Rad PTC-200 Peltier thermal cycler detection system. Q-PCR reaction conditions were: 95ºC for 15m for the initial denaturation and then 95ºC for 30s, 60ºC for 30s and 72ºC for 30s during 35 cycles. The melting curve analyses were performed from 60ºC to 95ºC with readings every 1ºC.

The housekeeping gene GAPDH was used to normalize the genes expression levels between technical and biological samples. A number of housekeeping genes including ACTB, B2M, GAPDH, HMBS, HPRT1, RPL13A, SDHA, UBC and YWHAZ have been tested for efficiency and specificity in the iPSC lines by my colleagues at Price laboratory. The genes were assayed by determining the relative quantification of the genes expression between iPSC samples by Q-PCR to assess the gene stability. The gene GAPDH was select as reference gene since it shows low variability in expression levels between cells. Additionally, the RNAseq results indicated that the expression of this gene was maintained in similar levels during the 8 days of NESC differentiation and between the biological replicates (iPSC lines).

The Pfaffl comparative method of relative quantification was used to quantify gene relative expression of each sample at different time points.

$$\text{Ratio} = \frac{(E_{target})^{\Delta CT_{target}\ (control - sample)}}{(E_{ref})^{\Delta CT_{ref}\ (control - sample)}}$$

ΔCTtarget = target gene cycle threshold

ΔCTref = mean of reference gene CTs

E = PCR efficiency

The Pfaffl method enables relative quantification of a target gene in comparison to an endogenous standard comprised of the reference genes. Neural induction and time course RNA extraction procedures were replicated 3 times for each cell line, which was considered a technical replicate. The technical replicates CT values were averaged and used to generate a biological replicate (iPSC line). The Q-PCR statistical analyses were performed with the averaged relative gene expression values of 3 biological replicates. The reference sample used to compare the gene expression of all cell lines was randomly designated as the time point 0 of the iPSC1 line. The means between samples were compared by a two-way ANOVA test with Bonferroni correction with 95% confidence interval with Prism package of GraphPad software.

**Table 1. Table of primers used for Q-PCR.**

| Gene | Primer Sense 3'-5' | Primer Antisense 5'-3' |
|---|---|---|
| CDH1 | TCC TGG GCA GAG TGA ATT TT | GGCGTAGACCAAGAAATGGA |
| FGF8 | CAG GTC CTG GCC AAC AAG | CTC CTC GGA CTC GAA CTC TG |
| GAPDH | AGC CTC AAG ATC ATC AGC AA | CTG TGG TCA TGA GTC CTT CC |
| GnRH1 | CTT CTG CCC AGT TTC CTC TTC | TTG ATT GAT TCT TTC CAA GAG ATA GT |
| HESX1 | TGC TTT TAC TCA AAA CCA GAT TGA | CCA AAT CTG GAT TCT GTC TTC C |
| NESTIN | AGG ACA CCA TGA GGA ACA GC | GCC ATG TTC TTG CTC ACG TC |
| NPTX1 | GTG ATA GGG CGC CAA GTT CT | ATC AAT GAC AAG GTG GCC AAG |
| OCT4 | TTG GGC TCG AGA AGG ATG TG | GTG AAG TGA GGG CTC CCA TA |
| PAX6 | GCC AGA GCC AGC ATG CAG AAC A | CCT GCA GAA TTC GGG AAA TGT CG |
| SIX3 | AGC AGA AGG ACC GAG TTC TG | CAA GAA CAG GCT CCA GCA C |
| SIX6 | GCT GCA GCC AAG AAC AGA CT | CTG GAC GTG ATG GAG ATG G |
| SMARCA2 | CGC TGA GAA ACT GTC ACC AA | CTGTCGCCCTGAACTGTTTC |
| SOX11 | TAC AGC CCC ATC TCC AAC TC | CTC CGA CTT CAC CAG AGA GC |
| TFAP2C | AGA TGG ACG AGG TGC AGA AT | CAGGGACTGAGCAGAAGACC |

## 2.7. Immunofluorescence

Approximately 100,000 cells were plated on each well of Nunclon delta surface 96 well plates (Thermo Fisher148761) following the methods described in the section 2.1. After confluence, cells were induced with either or both neutralization media depending on the experiment. Cells cultures were fixed at different time points with 50µl/well of 4% paraformaldehyde (PFA) (Thermo Fisher, 28906) in PBS 1X for 15 minutes at room temperature. The PFA was removed and the wells were washed 3 times with 150µl/well of 1X PBS. The cells were permeabilized and blocked by incubation with 50µl/well of 4% normal donkey serum (Sigma Aldrich, D9663) in 1X PBST for 1h at room temperature in

a nutator. Cells were incubated with specific dilutions of mouse monoclonal and rabbit polyclonal primary antibodies together (Table 2) in 50µl/well of blocking buffer at 4°C overnight. Wells were washed three times with 100µl/well of 1x PBS. The dilution of the primary antibodies was determined by following the providers recommendations and references in published literature. 5 serial dilutions were performed based on the datasheet and literature recommendations for each antibody. The optimal dilution was determined by identifying the concentration that provided the best staining and lowest background. Validation of the antibodies binding specificity was assessed by replacing primary antibodies with the same dilution of purified mouse IgG (Merck Millipore, CS200621) or rabbit IgG as controls (Thermo Fisher, 02-6102). Immunoreactivity was imaged using Alexa Fluor 594 conjugated donkey anti-mouse IgG (Invitrogen, A-21203) and Alexa Fluor 488 conjugated donkey anti-rabbit IgG (Invitrogen, R37118) both diluted to 1:250 ratio in 50µl/well of blocking buffer. The nuclei were stained by incubation with Hoechst 33342 (Thermo Fisher, H3570) diluted to a final concentration of 5µg/ml in 50µl/well of 1x PBS for 3 minutes at room temperature. Stained cells were washed 3 more times with 150µl/well of 1x PBS.

Images were acquired with a 20X objective with the Cell insight CX5 High Content Screen Platform (Thermo Fisher, CX51110). Expression was quantified using the bioapplication Cell Health Profiling from the iDev software package (Thermo Fisher). Hoechst staining was used to assess cell viability. Specific staining intensity, shape and size parameters were established to identify positive and negative labeled cells. A total of 3 wells were analyzed per primary antibody pair with 61 acquired fields per well. The means of the percentages of positive cells at each time point were statistically compared by a two-way ANOVA test with 95% confidence interval with Bonferroni correction with Prism package of GraphPad software.

**Table 2. List of antibodies used for immunofluorescence.**

| Antibody | Type | Host | Dilution | Supplier | Catalog Number |
|---|---|---|---|---|---|
| LEF1 | Monoclonal | Mouse | 1:200 | Millipore | Cs200635 |
| NESTIN | Monoclonal | Mouse | 1:500 | R&D Systems | MAB1259 |
| NR2F2 | Monoclonal | Mouse | 1:100 | R&D Systems | PP-H7147-00 |
| OCT4 | Monoclonal | Mouse | 1:200 | Santa Cruz | Sc5279 |
| PAX6 | Polyclonal | Rabbit | 1:200 | Proteintech | 12323-1-AP |
| REST | Polyclonal | Rabbit | 1:200 | Millipore | Cs200555 |
| SOX11 | Polyclonal | Rabbit | 1:50 | Santa Cruz | Sc20096 |
| ZEB1 | Monoclonal | Mouse | 1:300 | Atlas Antibodies | AMAb90510 |
| ZEB2 | Polyclonal | Rabbit | 1:150 | Atlas Antibodies | HPA003456 |
| ZNF521 | Polyclonal | Rabbit | 1:100 | Atlas Antibodies | HPA023056 |
| ZO1 | Polyclonal | Rabbit | 1:500 | Thermo Fisher | 40-2200 |

# CHAPTER 3. Neuroectodermal differentiation of iPSC

## 3.1. Introduction

Human development initiates with the asymmetric division of the fertilized oocyte to generate an 8-cell blastula. Subsequently, cells divide symmetrically and asymmetrically and form the TE and the ICM, which ultimately segregates into the PrE and the Epi. Following implantation, gastrulation occurs with the Epi cells migrating through the PS and node (mouse E6-E7, humans 13-17 dpf) (Artus *et al*., 2014; Zirra *et al*., 2016). During gastrulation, cells that do not invaginate through the PS remain as the most exterior layer that will give rise to the ectoderm. The consecutive steps of neural development proceed with induction of the ectoderm to form the NP in response to inhibitory BMP, TGFβ and WNT signaling from the adjacent mesoderm. Subsequent cell proliferation increases the thickness of the NP, which then folds to form NT at approximately 21 dpf in the human embryo (Klein & Moody, 2015). The NT is later separated into spatially and functionally distinct regions in response to specific concentrations of signaling molecules. Depending on their location and exposure to morphogens, the NT progenitors ultimately differentiate into specific post mitotic neuronal subtypes or glial cells (Imayosgi *et al*., 2015).

Due to legal constraints, the human embryo can be studied in vitro only until day 13 or the formation of the PS. Thus, our knowledge of neural induction in mammals is based on studies based in model organisms, largely on mouse. Analyses of mouse development have significantly contributed to the identification of genes and pathways activated during NE development that play important roles directing fate commitment. Importantly, mouse studies have enabled identification of developmental cues that can be used to manipulate

*in vitro* differentiation of human ESCs and iPSCs to relevant phenotypes (Klein & Moody, 2015; Zirra *et al*., 2016).

The neuronal differentiation protocol used in Price laboratory includes 8 days of neural induction with 2i inhibitors. Further differentiation leads to formation of rosette structures (d14) followed by differentiation of neuronal progenitors and acquisition of cortical neuronal phenotype (d50). Inclusion of WNT inhibitors during 2i (Dorsomorphin, SB431542 and XAV939) induction leads to NESCs specification toward gonadotropin releasing hormone (GnRH) neurons. Importantly, the only step that differs between differentiation of cortical and GnRH neurons is the initial 8 days of neural inhibition; all subsequent steps of cortical and GnRH neuronal induction are identical. This presents an ideal experimental system to identify the common and unique molecular events that underlie neural induction of NESC specified to distinct cortical and GnRH neurons.

### 3.1.1 Human neuro-ectodermal and neural progenitors gene expression patterns

Up-regulation of NE markers PAX6, SOX1, NCAM, ZIC1 and ZEB2, and loss of pluripotency/self-renewal markers OCT4, NANOG and KLF4 occurs at day 5-7 after hPSCs exposure to neural induction media (Hu *et al*., 2010; Kamiya *et al*., 2014; Leemput *et al*., 2014; Huang *et al*., 2016). Additionally, the mesodermal markers CD4, GATA, RUNX ant T and the endodermal marker SOX17 retain similar expression levels in ESC compared with NE derived cells. Thus, analyses of gene expression patterns in human NE cells suggest that several genes and pathways are conserved across animal models (Huang *et al*., 2016). Interestingly, it has being demonstrated that PAX6 expression precedes SOX1 during neural development in humans. However, Chambers *et al* (2009) described SOX1

as the earliest NE marker in their hESCs 2i induced culture system. Direct regulation of SOX1 by SMAD signaling is a possible explanation for the differential expression of SOX1 in specific culture conditions (Chambers *et al*., 2009). Additionally, SOX1 expression levels in neural progenitors might change in response to different concentrations of BMP inhibitors in the media (Neely *et al*., 2012).

Recently, a microarray study by Huang *et al* (2016) revealed up-regulation of genes that have not been previously linked to neural differentiation, such as HEY1, BCL11A and SP8 after 6 to 10 days of hESC neural induction. The authors speculated that the novel genes might have roles in less well-studied pathways implicated in neural differentiation. A number of genes involved in chromatin modifications such us TET2, KATB2 and SIRT1 were differentially expressed after 6 and 10 days of neural induction. The dynamic expression of epigenetic modifiers might be related to gene expression changes during neural differentiation; however, the role of these genes and their regulatory interactions is unknown. Additionally, the study showed up-regulation of ZEB1 and SNAI2, TF that are necessary for epithelial mesenchymal transition (EMT), and down regulation of their epithelial gene targets including OCLN, CLDN7 and F11R, suggesting a potential role of EMT during neural differentiation. Likewise, ROR2 and WNT5B genes associated with WNT non-canonical pathway WNT/$Ca^{2+}$ were up-regulated during neural progenitor differentiation. Inhibition of CaMKII a kinase down-stream WNT/$Ca^{2+}$ pathway resulted in reduction of PAX6 expression and increased K18 epithelial marker expression. The authors suggest that WNT/$Ca^{2+}$ pathway enhances neural differentiation through CaMKII phosphorylation by inhibiting epidermal fate acquisition (Huang *et al*., 2016).

During chick and mouse development, NE cells adopt dorsal forebrain fate by default in absence of patterning signals, while maintaining potential to respond to morphogens and

differentiate into distinct regional types of neurons. Explant studies have shown that Wnt signaling plays a key-patterning role in inducing dorsal identity, whereas Shh defines ventral forebrain fate. Concomitantly, rostro-caudal patterning is orchestrated by WNT, retinoic acid (RA) and FGF signaling, which induce neural progenitors to adopt a posterior fate (Maden, 2007; Chi *et al*., 2016). Similarly, human and mouse PSCs induced to NE cells adopt a forebrain fate in absence of patterning morphogens. For instance, hPSCs derived NE cells induced by the 2i media express forebrain markers including PAX6, OTX2, HESX1, SIX3, RAX, LHX2 and FOXG1 at day 6 after induction (Huang *et al*., 2016; Zirra *et al*., 2016). However, the progenitors preserve their potency and can be caudalized by addition of WNT or RA agonists to the media (Chi *et al*., 2016).

As differentiation proceeds, NE cells derived from iPSCs and ESCs acquire cell polarity and reorganize to form rosette like structures that resemble a transverse section of the neural tube in the embryo. In humans, the rosette formation time line depends on the tissue culture conditions and generally ranges between 7 and 14 days after initiation of 2i induction. Studies have demonstrated that human and mouse progenitors express PLAGL1, DACH1 and ZBTB16 rosette markers and pan neural progenitor markers including PAX6, NESTIN, SOX1 and SOX2, which are also expressed in the NT progenitors. Additionally, the rosette luminal cella express proteins such us N-CAD and ZO1 that are also present in the apical domain of embryonic NE cells in the NT (Elkabetz *et al*., 2008, Abranches *et al*., 2009). For instance, it has been shown that human neural progenitors within the rosettes represent RG with similar gene expression profiles and differentiation potential as neural progenitors from the NT, which are capable of differentiating into neurons, oligodendrocytes and astrocytes (Malchenko *et al*., 2014). The newly born neurons migrate to localize in the periphery of the rosette, resembling the differentiation of the NT neurons that localize outside of the proliferative ventricular zone (Abranches *et al*., 2009). Human early rosettes

derived from ESC are enriched in highly proliferative neural progenitors that exhibit a propensity for high self-replication and low differentiation. Later, these human neural progenitors progress to a mid RG rosette stage characterized by high propensity to differentiate into neurons and intermediate progenitors due to an increase in asymmetric division. Further culture leads to reduction of neural progenitors number and loss of the rosette integrity (Wilson & Stice, 2006; Ziv *et al*., 2015). Human RG progenitors derived from PSCs adopt a forebrain fate and differentiate into cortical neurons when induced in absence of morphogens by the default system. These neural progenitors predominately give rise to glutamatergic neurons but can also differentiate into inhibitory striatal medium spiny neurons (MSNs) and GABAergic neurons (Yuan *et al*., 2015; Chi *et al*., 2016). Early and mid-radial glial rosette stages recapitulate many aspects of cortical radial glial development in the human embryo (Ziv *et al*., 2015). For instance, human PSCs derived RG cells differentiate into cortical neurons corresponding to deep and superficial cortical layers and subsequently adopt late progenitor phenotypes that give rise to astrocytes. Furthermore, human RG transplanted into mice lateral ventricles differentiate into neurons and astrocytes that migrate and integrate into the appropriate layers of the mouse cortex (Duan *et al*., 2015).

Subsequently, human neural progenitors differentiate into diverse cell types in response to concentration gradients of specific morphogens including WNT, RA, FGF, and SHH *in vivo* (Wilson & Stice, 2006; Ulloa & Briscoe, 2007). Accordingly, the patterning potential of these morphogens has been exploited to generate specific neuronal subtypes from human PSCs. For instance, treatment of hPSCs with SHH and RA at specific time points during neural development results in efficient induction of motor neurons (Hu *et al*., 2009; Chambers *et al*., 2009). Likewise, early exposure to SHH and FGF8 leads to generation of

functional dopaminergic neurons capable to integrate into mouse brain after transplantation (La Manno *et al*., 2016; Li *et al*., 2016).

Accordingly, human PSCs have enormous potential and have been widely used to model diseases and to analyze development of neural progenitors to specific neuronal phenotypes. However, fewer attempts have been made to identify TFs and genes involved in NE differentiation. Hence, the molecular pathways and gene expression patterns that drive NE differentiation remain to be investigated. In this chapter, we report the first complete transcriptome analysis at different stages during iPSC NE differentiation.

### 3.1.2. ESC/iPSC differentiation to GnRH neurons

The developmental pathways that govern the development of the GnRH neurons in humans are poorly understood. However, Xenopus and mouse models and ESC studies have enabled insight into the molecular mechanism that drive formation of the NBZ, placodes, olfactory placodes and GnRH neurons (Ealy *et al*., 2016; Shigetani *et al*., 2016; Matsuoka *et al*., 2017). In Xenopus, expression of NBZ marker genes is promoted by BMP signaling. Interaction of these TFs markers leads to the formation of the NBZ and concomitant inhibition of neural factors. Subsequently, low levels of FGF family genes induce formation of the PPE fated to form placodes in the absence of WNT signaling, whereas, activation of WNT signaling during this stage leads to the formation of neural crest (Moody & LaMantia, 2015; Shigetani *et al*., 2016). Subsequently, different TFs become restricted to the anterior or posterior PPE. Possibly, these TFs interact with each other to promote the specification of placodal subtypes, which express different combination of these factors. The olfactory placode ($Emx2^{+}$ / $Ebf2^{+}$) is derived from the anterior PPE and gives rise to GnRH cells that

migrate to the hypothalamus at around E.11 to E.16.5 in mouse (Wray, 2010; Moody & LaMantia, 2015).

It has been recently reported that spatial temporal recapitulation of these developmental signals leads to the sequential generation of non-neural ectoderm and PPE in both human and mouse ESCs. Inhibition of TGFβ (using SB431542) increases the efficiency of induction of ectodermal derivatives, which in response to recombinant BMP4 and FGF2 leads to the formation of non-neural ectoderm ($DLX3^+$ /$GATA3^+$ / $AP2A^+$). Differentiation of PPE ($EYA1^+$ /$SIX4^+$) is achieved by subsequent inhibition of BMP and WNT signaling with small molecule inhibitors (Matsuoka *et al*., 2017). Different approaches used inhibition of TGFβ and WNT to derive non-ectoderm cells that acquire cranial placodal fate by addition of FGF (Ealy *et al*., 2016). Further differentiation with both protocols leads to acquisition of early otic neural progenitors (Dincer *et al*., 2013; Leung et al., 2013; Ealy *et al*., 2016). However, induction of GnRH through PPE has not been reported. GnRH generation from ESC/iPSC was shown for the first time by inducing neural progenitors differentiation through 2i induction and subsequent treatment with FGF8 (Lund *et al*., 2016; Poliandri *et al*., 2017). The transformation of neural progenitors to GnRH is directed by FGF8, however the molecular mechanism underlying this conversion remains to be elucidated. Overall, these studies demonstrate that placodal and GnRH neuronal differentiation both depend on the temporal modulation of WNT, BMP and FGF signaling.

### 3.1.3. WNT signaling

WNT signaling regulates a broad variety of biological processes during embryonic development and disease. Due to its critical importance in development WNT signaling is tightly spatiotemporally regulated. Deregulation of WNT leads to severe defects including

neurodegenerative diseases (Komiya & Haas, 2008; Munoz-Descalzo *et al*., 2015; Stanganello & Scholpp, 2016). There are 19 WNT glycoproteins in humans, which are modified in the endoplasmic reticulum and then transported and secreted into the extracellular milieu (Komiya & Habas, 2008; Huang *et al*., 2009a). A hallmark of WNT signaling is the formation of concentration gradients, which regulate fate by diffusing across cells that respond in a particular manner to specific WNT thresholds (Harrison &Pleasure, 2012; Stanganello & Scholpp, 2016). WNT morphogenetic activity regulates formation of distinct functional regions in the CNS such us regionalization of the PPE at the expense of neural crest with high WNT concentrations or ventralizing of forebrain progenitors (Harrison & Pleasure, 2012; Chi *et al*., 2016). Differentiation of distinct neuronal subtypes by regulating WNT concentrations has also been reported in *in vitro*. For instance, differentiation of ESCs in a chemically defined culture media without extrinsic signals leads to the generation of dorsal telencephalic glutamatergic and GABAergic neurons. However, the derived cells are ventralized and give raise to subpallial derivatives if exposed to WNT antagonist prior to terminal differentiation (Chi *et al*., 2016; Zirra *et al*., 2016).

#### 3.1.3.1. WNT/ β-catenin canonical pathway

WNT/ β-catenin signaling is involved in many processes during embryogenesis such us cell adhesion, neural patterning and cell proliferation. In the absence of WNT, β-catenin is anchored by the destruction complex comprised of APC, GSK-3β, PP2A, CK1 and AXIN. Within the complex, the scaffolding proteins CK1 and GSK-3β phosphorylate and target β-catenin for ubiquination and degradation in the proteasome. WNT signaling prevents formation of the β-catenin destruction complex and subsequent β-catenin degradation by sequestering AXIN and GSK-3β (Nuse, 2012; Peng *et al*., 2017). Signaling is activated as

WNT proteins bind the trans-membrane receptor Frizzled (FZD) and its co-receptors LRP6 or LRP5. This complex recruits the scaffolding protein Dishevelled (DVL), which along with FZD regulate phosphorylation and activation of LRP6/LRP5 cytoplasmic tail. Subsequently, AXIN and GSK-3β are recruited to LRP6/LRP5 tail and translocated into the cell membrane. In the presence of WNT, stable β-catenin accumulates in the cytoplasm and migrates to the nucleus where it binds to TCF/LEF TFs to activate or inhibit expression of target genes (Heeg-Truesdell and LaBonne, 2006; McDonnlad *et al*., 2009) (Figure 3.1).

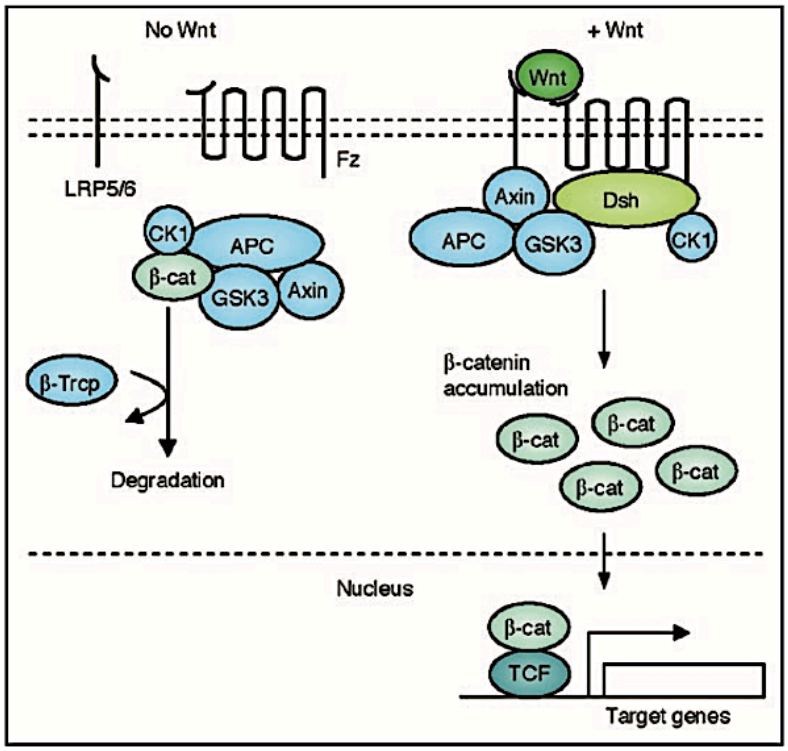



**Figure 3.1. Scheme of WNT canonical pathway.** In the absence of WNT (left), the degradation complex recruits β-catenin in the cytoplasm. β-catenin is phosphorylated by GSK3 and CK1 and then targeted for degradation. Binding of WNT proteins to the FZD receptor and LRP5/6 co-receptors leads to phosphorylation of DVL and recruitment of AXIN and GSK-3β to the membrane. Thereafter, β-catenin dissociates from the destruction complex, accumulates and localizes in the nucleus where it regulates expression of target genes (Komiya & Habas, 2008).

## 3.2. Methods

Three iPSC1 lines and an ESC were culture with 2i and 2i-WNT media for 8 days according to the protocols described in section 2.1.

RNA extraction and RNAseq methods were carried as described in sections 2.2 and 2.3 respectively.

### 3.2.1. RNAseq statistical analyses

Principal component analyses (PCA) and a heatmap depicting the distance between the samples were produced by our collaborators Shankar Subramanian and Jun Min (UCSD) using DESeq2 and ggplot2 packages for effective clustering and comparison between multiple genes. Any gene with the total raw read counts of zero across all conditions were removed.

For the following analyses I used the raw data matrix provided from our collaborators (UCSD). I regularized-logarithm (rlog) the count data using DESeq2 in RStudio console. I preformed the differential expression analyses by using the iPSC lines rlog mean with the DESeq2 function DESeqDataSet (dds) in RStudio to extract the log2 fold changes and Wald test p value Benjamini and Hochberg corrected. I used Dds analyses to compare gene expression between consecutive time points during neural induction and between the time points D0 and D8. I create the Heatmaps with the rlog-transformed values using the package pheatmap and rcolourbrewer in RStudio.

I generate the volcano plots with DE genes from the 2 neural induction protocols at time points D8 and D4 with the calibrate package in RStudio. I created the gene plots by using

the ggplot2 package in RStudio. I used the Dds to compare gene expression between the 2i and 2i-WNT treatments at each time point. These analyses demonstrated that 146 genes are significantly differentially expressed between the 2 neural induction treatments at time point D8. I obtained a mean RPKM value of the 3 iPSC for each gene at each time point. I plotted the RPKMs of the set of DE genes for cells induced either with 2i or 2i-WNT media at d8 at each time point to determine the stage when the expression of these genes starts to diverge between both treatments. I compared the RPKMs with Prism package of GraphPad software.

Additionally, I plotted the mean of the RPKMs values from the three-iPSC lines with Prism package of GraphPad software for a number of selected genes.

### 3.2.2. STEM clustering

Sets of genes with similar temporal expression profiles were cluster with short time series expression miner (STEM) tool by our collaborator Jun Min (UCSD). Briefly, the RNAseq count values were rlog normalized with the DESeq2 package in R and the resulting data were used as input for STEM. STEM identifies significant gene expression profiles and then, by a clustering algorithm assigns genes that match these profiles determined by a correlation coefficient cutoff of 0.9 with respect to the significant model profile. The profiles with statistically significant number of genes are identified with a permutation test with Bonferroni correction.

### 3.2.3. Q-PCR assay

Q-PCR assays were performed as described in section 2.8. The genes analyzed in this chapter were PAX6, NESTIN, OCT4, SOX1, SIX6, SIX3, HESX1, NPTX1 and FGF8 and

the housekeeping gene GAPDH was used for normalization. Gene primer sequences can be found in table 1.

A Pearson's correlation analysis was performed with the RPKMs obtained from the RNAseq data and the Q-PCR relative quantification values. For these analyses, the gene PAX6 was randomly designated as reference to obtain linear expression values of all genes with the Pfaffl comparative method. The Pearson's correlation was calculated with Prism package of GraphPad software.

### 3.2.4. Immunofluorescence

Immunofluorescence assays were carried out with the iPSC1 line differentiated with 2i. Cells were fixed at time points d0, d4 and d8. PAX6, NESTIN, OCT4, ZEB1, ZEB2, NR2F2, ZNF521, REST and LEF1 expression patterns were analyzed with the antibodies described in the Table 2. Statistical analyses were performed as described in section 2.8.

### 3.2.5. Super TOPflash reporter assay

A XAV939 (Sigma-Aldrich, X3004) dose-response curve was performed in collaboration with Dr. Richard Killick laboratory (King's College London). We used Super TOPflash assay to monitor the expression levels of WNT signaling in cells cultured with different concentrations of XAV939. TOPflash is a luciferase expression plasmid that contains 2 sets of 3 copies of TCF binding regions. Cells transfected with the plasmid show increase of luciferase activity when Wnt-β-catenin signalling is active and binds to TCF binding regions. Colleagues from Price laboratory have previously found that inhibition of WNT/ β-catenin signaling during 2i induction leads to the differentiation of GnRH neurons at expenses of cortical neurons (Kathuria *et al*., 2017). Hence, this experimental procedure

enabled to determine the appropriate concentration of XAV939 (2μM) necessary to inhibit WNT/ β-catenin signaling and induce GnRH neuronal fate.

HEK239A cells were plated in a 96 well plate until they reached around 80% confluence. A transfection reaction was prepared for each well by adding 0.3μl of FuGENE transfection reagent (Promega, E2311), 100ng of M50 Super TOPflash luciferase reporter (Addgene, 12456) and 100ng of a (plasmid cytomegalovirus promoter DNA) pcDNA Wnt3A expression vector (Addgene, 35908) to 50μl Opti-MEM Reduced Serum Medium (Gibco, 31985062). An empty vector, pcDNA3.1 (Addgene, 52535), was added instead of the Wnt3A plasmid as a negative control. The transfection reaction was incubated at room temperature for 15 minutes. 50 μl of the transfection reaction were added to each 96 well and HEK293A cells were incubated at 37$^{o}$C with 5% CO2 and 20% O2. After 7-hour post transfection, media was changed with 100ul of DMEM/F-12 supplemented with different doses of XAV939 (1nM, 3nM, 10nM, 30nM, 100nM, 300nM, 1μM, 3μM, 10μM and 30μM) and incubated at 37°C, 5% CO2 and 20% O2 for 12 hours. DMEM/F-12 media containing 0.01% dimethyl sulfoxide (DMSO) was used as XAV939 negative control. After incubation, 100μl of ONE-Glo luciferase assay system (Promega, E6110) were added to each well and incubated for 3 minutes at room temperature. Luciferase reporter gene activity was detected in a GloMax Navigator luminometer (Promega, GM2000). Luciferase activity means were obtained from 3 wells per XAV939 dose. The empty vector control was used to normalize the luciferase activity values to 1 to obtain the fold increase of each XAV939 dose and the DMSO control.

## 3.3. Results

### 3.3.1. Overall similarity between iPSC lines

The scope of this chapter is to identify the gene expression patterns generated during NE differentiation of iPSC in humans. It has been shown that NE markers such us PAX6, SOX1 and ZEB2 are actively transcribed after 8 days of 2i from iPSC (Hu *et al*., 2010; Kamiya *et al*., 2014; Leemput *et al*., 2014; Huang *et al*., 2016). Hence the experimental approaches were carried on during this differentiation period to uncover the molecular signals that initiate and maintained NE differentiation in humans. Further cell culture drives the formation of neural rosettes at D12 approximately and subsequent differentiation of GnRH (2i-WNT) or cortical neurons (2i) at around D50. The differentiation of human neural progenitors has been extensively studied using pluripotent stem cells as and foetal tissue. However, the molecular mechanisms that regulate the formation of NE cells is widely unknown in humans. Thus, analysing this developmental stage was the focus of this study.

Accordingly, 3 iPSC lines derived from different individuals as biological replicates were used to ensure the robustness and reproducibility of the experiments and results. Additionally, transcriptome data from an ESC line was obtained to estimate the variation between the gene expression profiles of the iPSC and ESC lines during differentiation. To determine if the genetic background and technical variability represent main sources of gene expression variation among the samples, a principal component analyses (PCA) was performed with the samples from all time points, cell lines and both treatments (Figure 3.2). The PCA results depicted the formation of 7 clusters along the PC1 on the x-axis, which captures 59% of the variance across the samples. Each cluster predominantly comprises samples from the same time point including iPSC lines and the ESC line. Clusters on PC1

are organized according to developmental time during neural induction. Samples collected at the same time point are tightly clustered in the PCA map independently of the cell line and the induction treatment. The second principal component (PC2) in the y-axis accounts for 10% of the variance. PC2 appears to separate the samples according to their degree of differentiation; i.e. iPSCs and NESCs cluster together and are separated from their intermediates states.

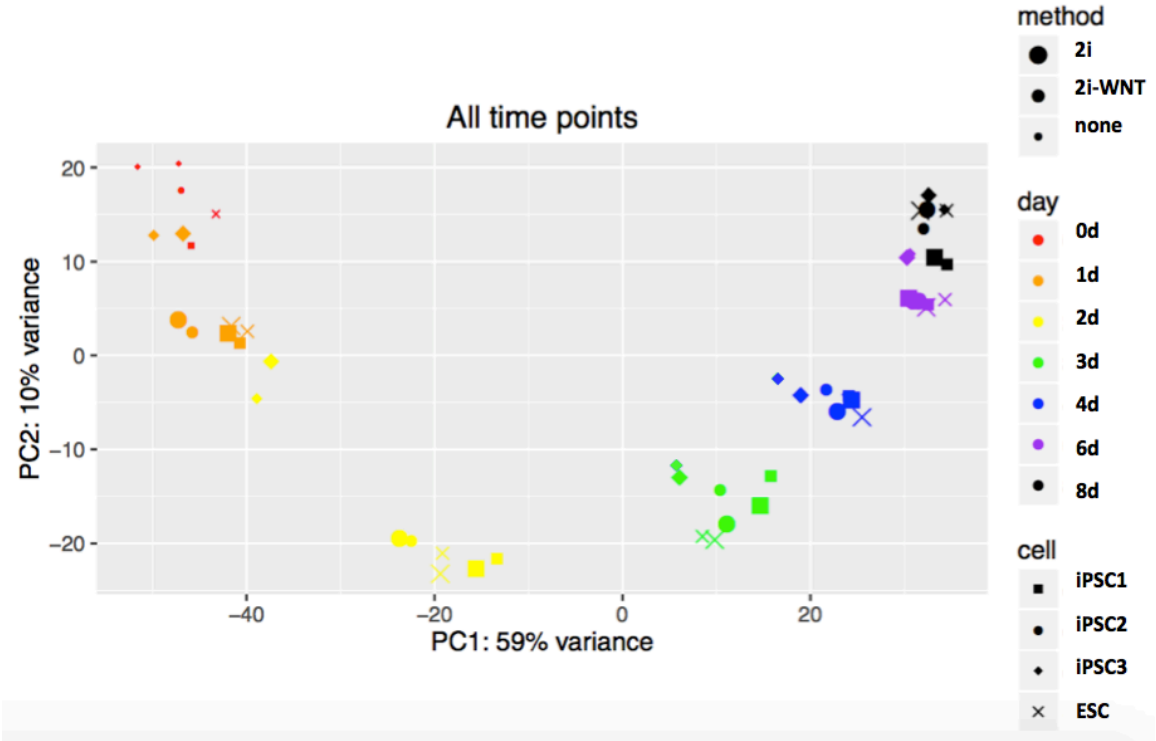



**Figure 3.2. Principal component analyses of all samples at all time points.** PC1 represents 59% of the variance across all samples and is plotted in the axis-x while PC2 represents 10% of the variance and is plotted in the axis-y. Samples from individual PSCs are show with different symbols. The size of the symbol indicates the method used for differentiation while the time points used for RNA extraction during NESC differentiation are shown in different colours.

The samples distance was calculated as another approach to assess the overall similarity and differences among the samples; the distance is visualized in a heatmap (Figure 3.3). The results indicate that the samples collected at time points d0, d1 and d2 have similar

gene expression patterns and are distant from the samples at time points d3, d4, d6 and d8, which form another cluster. For instance, the early samples (d0, d1 and d2) are tightly clustered with other samples from the same time point including the iPSC and ESC lines. Similarly, among the latter cluster (d3, d4, d6 and d8), samples from d3 and d4 are closer to each other and samples d6 and d8 are also close together. Among the clusters the samples are clustered by time point independently of the cell line or treatment.

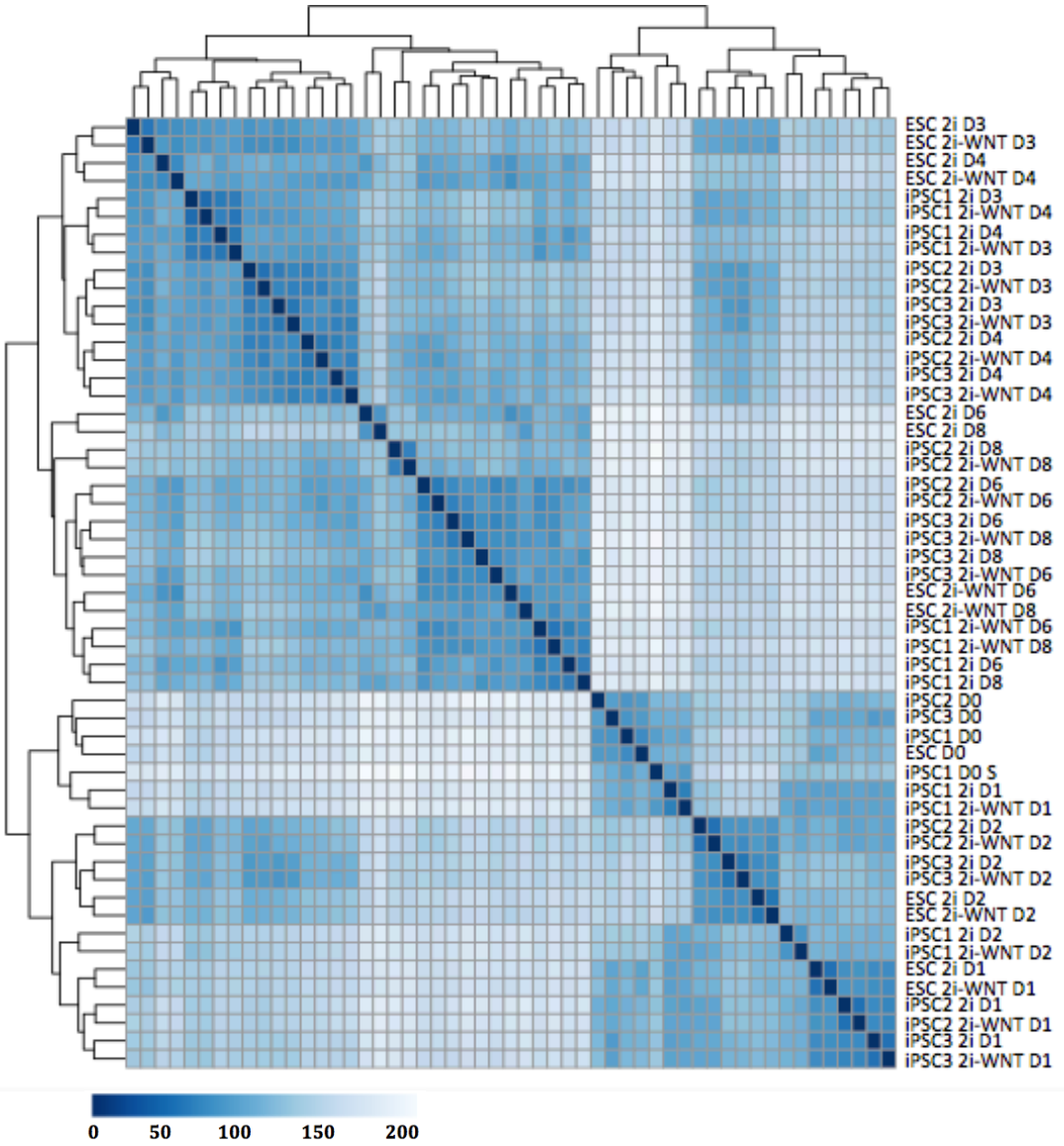



**Figure 3.3. Heatmap of sample-to-sample distances.** On the heatmap the samples are located in the same order from the top to bottom as from left to right. According to the colour key and the scale the distance between identical samples is depicted by dark blue (0) whereas the distance between different samples is represented in white (200).

### 3.3.2. Validation of neuroectodermal fate

Expression of known NE makers was analysed to validate differentiation of iPSCs to NESCs after 8 days of neural induction. iPSCs were treated either with 2i or 2i-WNT media to promote cortical or GnRH neuronal fate acquisition respectively. Analysis of GnRH differentiation in animal models have demonstrated that modulation of BMP, FGF2, TGFβ and WNT signals leads to the formation of PPE and GnRH neurons (Dincer *et al*., 2013; Leung et al., 2013; Ealy *et al*., 2016; Poliandri et al., 2017). For instance, colleagues in Price laboratory have shown that GnRH neurons can be derived from iPSC by the inhibition of WNT signalling during 2i (Kathuria *et al*., 2017). Interestingly, inhibition of WNT during the 8 days of 2i induction is the only step that differs between the differentiation of cortical and GnRH neurons. Hence, we aimed to identify common and unique regulatory signals that might prime cells during NE induction to acquire GnRH or cortical identity.

Gene expression of NE and neural markers was represented in heatmaps for visualization (Figure 3.4). The results indicate that these markers depicted highly similar gene expression patterns between cells differentiated with 2i and 2i-WNT. The results indicate that WNT inhibition does not significantly alter the expression of neural markers during 2i NE induction. For both treatments, the potency/self-renewal markers POU5F1 (OCT4), NANOG and MYC are down-regulated whereas KLF4 is up-regulated. The genes FOXD4, GMNN and ZIC2 are considered early NE progenitors since their expression initiates in the inner cell mass and is up regulated during NE differentiation in mouse and Xenopus. The data show that these genes depict a peak of expression at d2 and d3. Subsequently, the TF IRX1, IRX2, IRX3, ZIC1 and ZIC3 are expressed up-stream bHLH pro-neural markers and are required for the maintenance of neural progenitor cells in a proliferative state in Xenopus. These early neural progenitor markers show a peak of expression at d2. Human

and mouse NE markers PAX6, NESTIN, SOX2, SOX11, CDH2, ZNF521, ZEB2 and NR2F2 are up-regulated between d3 to d8. The bHLH proneural TF NEUROG2 and ASCL1 are up-regulated at d6 and d8 and the neural inhibitors HES1 and HES5 are up-regulated from d3 and d4 respectively. Conversely, OLIG1 and OLIG2 are expressed in iPSC and d1 and are subsequently down-regulated.

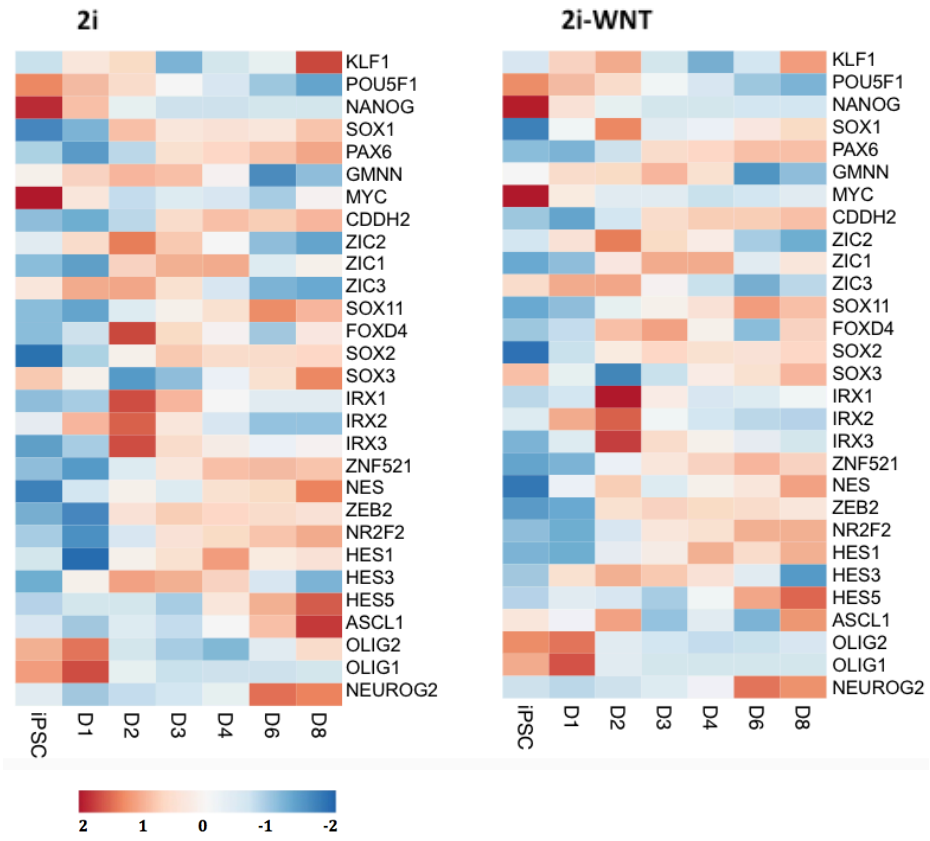



**Figure 3.4. Heatmap of neuroectoderm markers expression.** The heatmap depicts expression of NE markers in iPSC cells induced with 2i (left) and 2i-WNT (right) media at the days d0, d1, d2, d3, d4, d6 and d8 during differentiation. The colour key is scaled from 2 to -2 by row, where 2 (dark red) represents high gene expression and -2 (dark blue) represents low expression relative to the mean of each gene.

### 3.3.3. RNAseq validation

Several studies have demonstrated that RNAseq and Q-PCR techniques are accurate and reliable approaches for the quantification of gene expression. However, the sensitivity of RNAseq method might vary depending on the reads depth, whereas Q-PCR sensitivity can be affected by the quality of the mRNA, the amplification efficiency and the choice of internal control genes among others (Pombo *et al*., 2017). Hence, we validated the transcriptome results obtained through RNAseq by comparison with data from Q-PCRs analyses. The results show that the temporal expression patterns of the analysed genes are similar between the Q-PCR data and the RNAseq assay (RPKMs) for 2i and 2i-WNT treatments (Figure 3.5). The pluripotency gene OCT4 is down-regulated, whereas the NE makers PAX6, NESTIN, SOX1, SIX6, SIX3, HESX1, NPTX1 and FGF8 are all up-regulated during induction similarly to the results from the RNAseq assay. Additionally, the Q-PCR data indicated that SIX6 depicted higher expression values for the time points D4, D6 and D8 compared with the values obtained through RNAseq analyses. SIX6 is among the genes that depicted low expression levels (< 20 PRKMs). The results might indicate that Q-PCR technique show higher sensitivity of the for genes with low expression levels such us SIX6, whereas the expression patterns of genes with high expression values like PAX6 closely resemble between both data sets.

**PAX6**

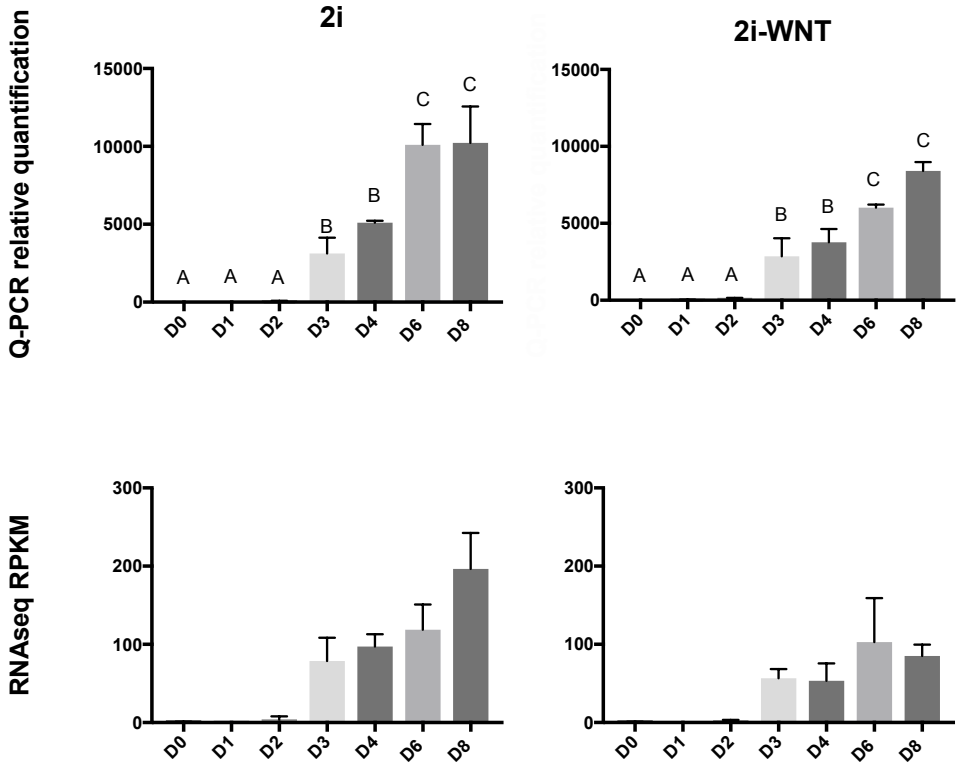
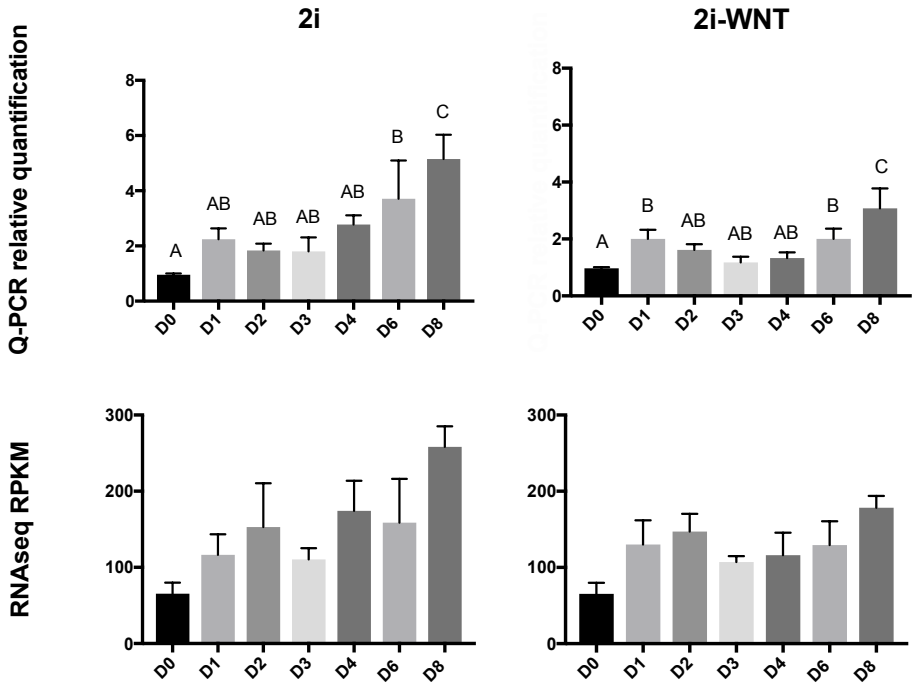



**Figure 3.5. Comparison between real-time PCR and RNAseq data during neuroectoderm differentiation.** Q-PCR relative quantification data and RPKM values are shown in the Y-axis top and bottom respectively. The X-axis indicates the time point of RNA extraction. iPSCs induced with 2i treatment for Q-PCR and RNAseq assays are located on the left, whereas cells treated with 2i-WNT inhibitors are on the right. The statically significant differences are illustrated with letters to simplify visualization of the results. The difference between the means of the variables with the same letter is not significant. The variables with different letter are significantly different. Details of the statistical test and p-values are included in the Appendix 3.1 (CD-ROM).

**NESTIN**




**OCT4**

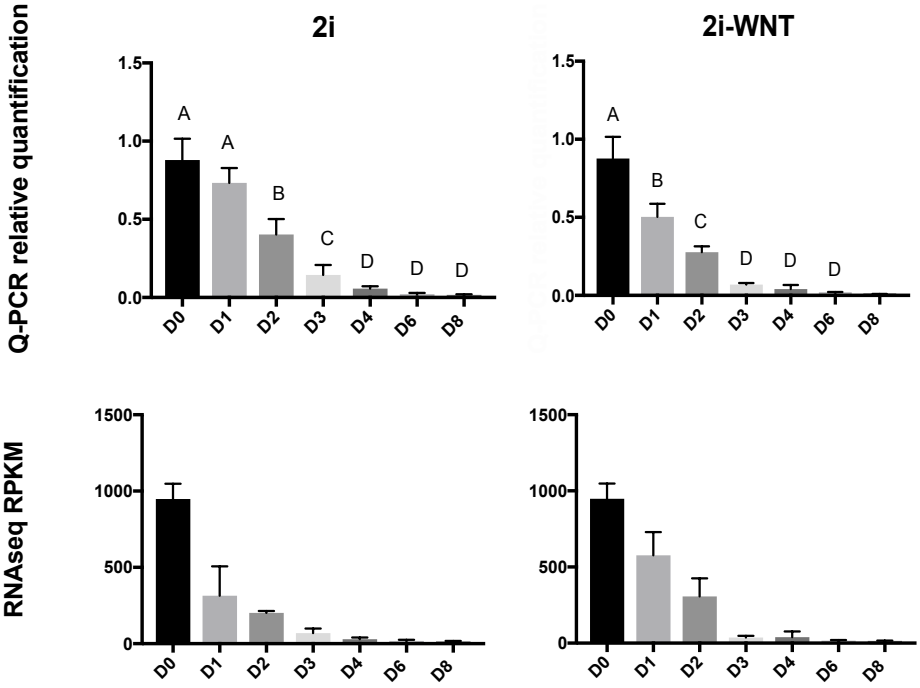
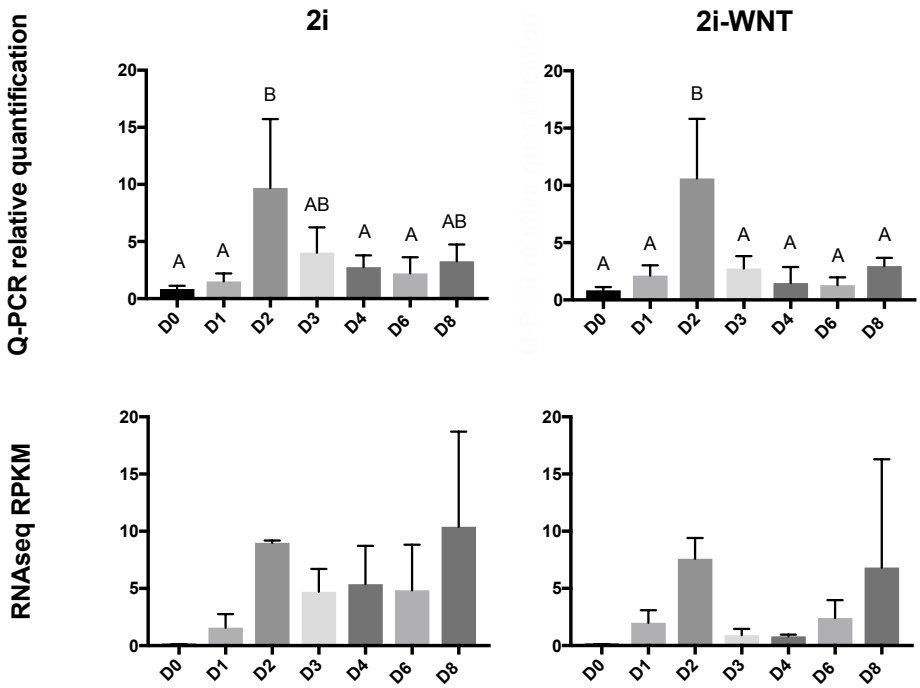



**Figure 3.5. Continued**

**SOX1**




**SIX6**

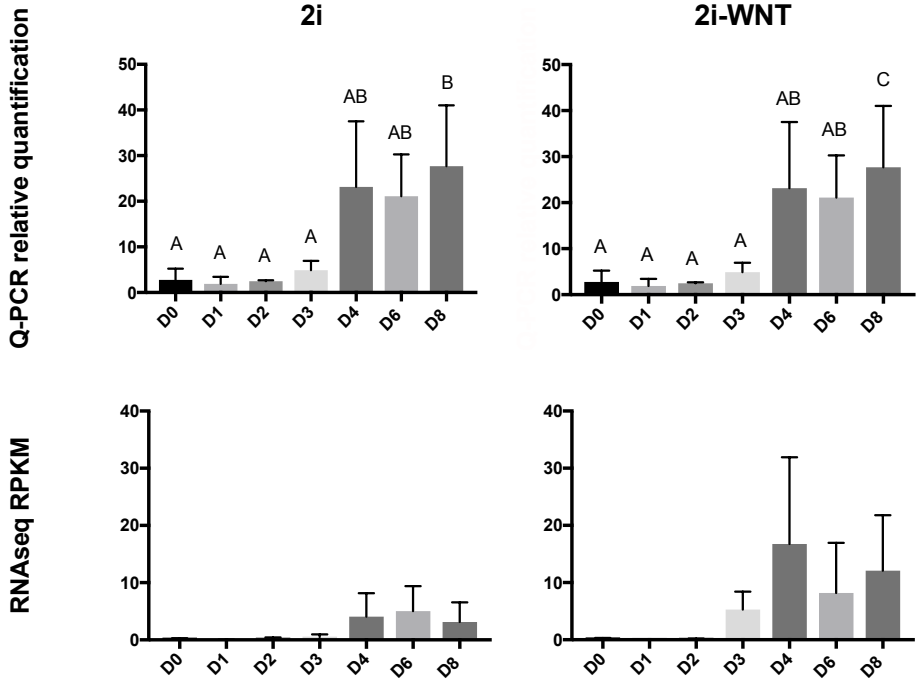
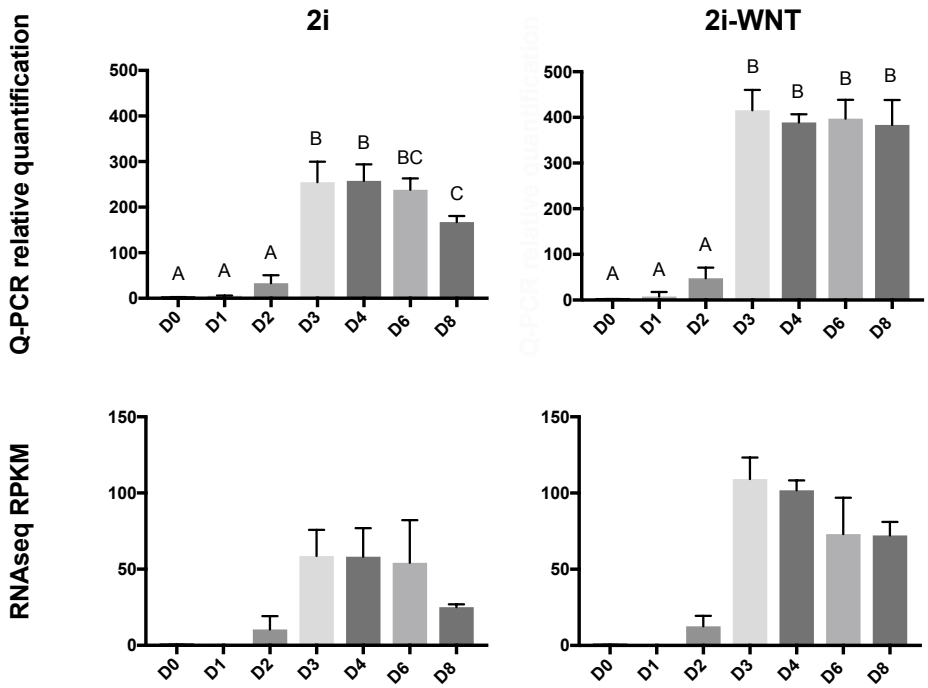



**Figure 3.5. Continued**

**SIX3**




**HESX1**

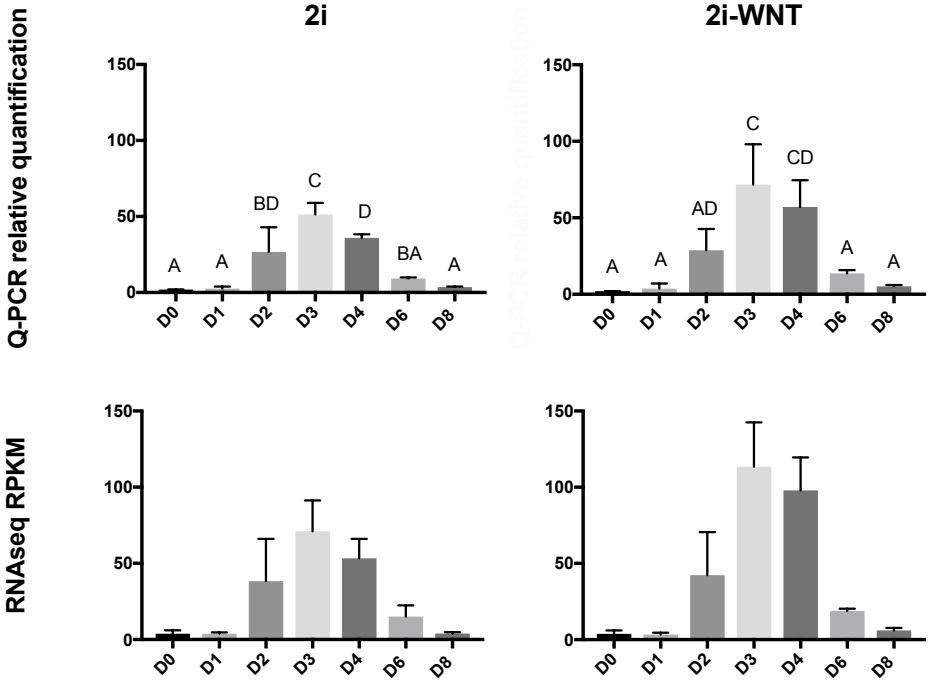
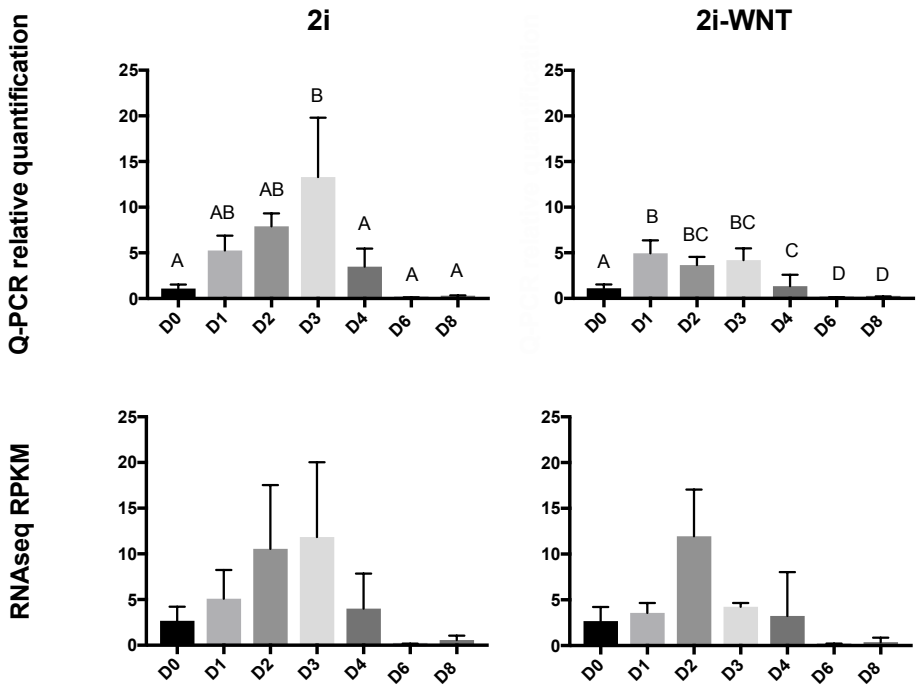



**Figure 3.5. Continued**

**NPTX1**




**FGF8**

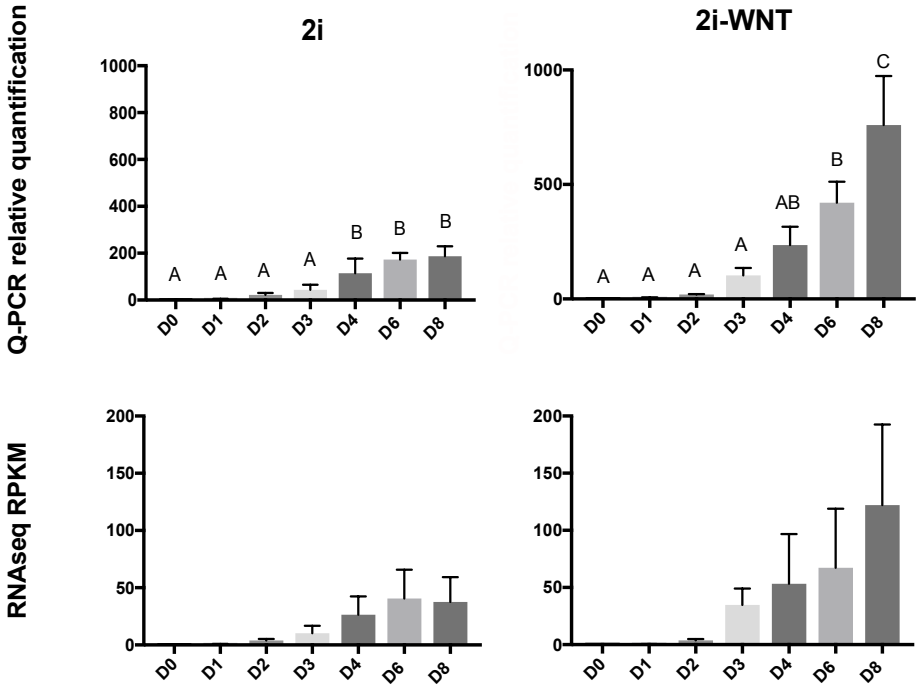



**Figure 3.5. Continued**

Additionally, values obtained from RNAseq data were compared with Q-PCR relative expression values through a Pearson's correlation to determine the accuracy of the results from both methods. To obtain linear values in the same scale from the Q-PCR data, PAX6 was randomly assigned as a reference gene to quantify the expression values for all genes through the Pfaffl method. The Pearson correlation was calculated by comparing the Q-PCR expression values of each gene and the RPKM values obtained through RNAseq of the same gene with Prism software. A significant positive correlation ($R=0.9$, $p < 0.0001$) was observed between the 2 data sets (Figure 3.6). The correlation results demonstrate that the gene expression patters obtained by RNAseq data are linearly significantly associated to the patterns obtained by Q-PCR. Importantly, the high validation between assays demonstrates the stability of gene expression patterns among the three biological replicates.

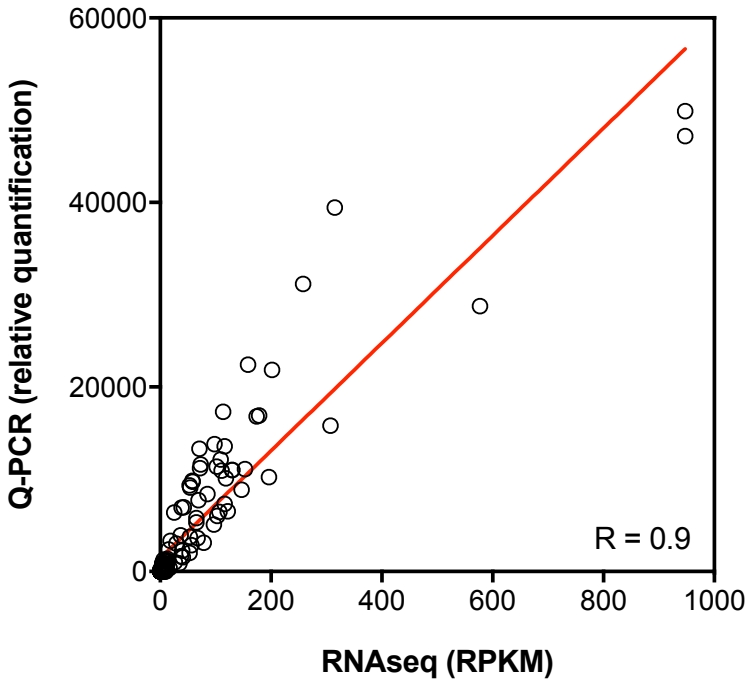



**Figure 3.6. Correlation between RNAseq data and real-time PCR values.** Relative quantification and RPKMs values for each gene at each time point are depicted with circles and represent the mean of the three iPSC lines.

The principal mechanism of modulating protein synthesis is the regulation of mRNA transcripts. However, the mRNA levels might not mirror the protein amounts due to regulatory mechanism that modify the mRNA fate. Hence, immunofluorescence analyses were performed to demonstrate that changes in gene expression were accompanied by comparative changes in protein expression. The results indicate that the pluripotency marker OCT4 is highly expressed at d0 and is not detected at d4 and d8 (Figure 3.7). Additionally, the expression of the NE markers PAX6, ZNF521, NESTIN, ZEB2, ZEB1, NR2F2 and LEF1 increased during the 8 days of neural induction (Figure 3.7). The results indicated that the pluripotency and NE markers are heterogeneously expressed since they are not positively stained in all the cells of the population. Mouse and rabbit IgG at high concentrations (1:50) were used as negative control to detect non-specific binding of the secondary antibodies, for the immunofluorescence assay. The threshold for positive cells detection was set considerably above the background to ensure no false positive cells were accounted as positive. Similarly, for the images acquisition, the exposure time was reduced to eliminate the background. Due to the stringency of the assay, a number of positive cells with low staining intensity might have been considered negative. Accordingly, the heterogeneous expression of the cell population might have been accentuated due to this artifact. Additionally, it is possible that the cell population might include cells individual cells that are transversed in different stages of the differentiation trajectory or acquiring different phenotypes, which might lead to differential expression of these genes.

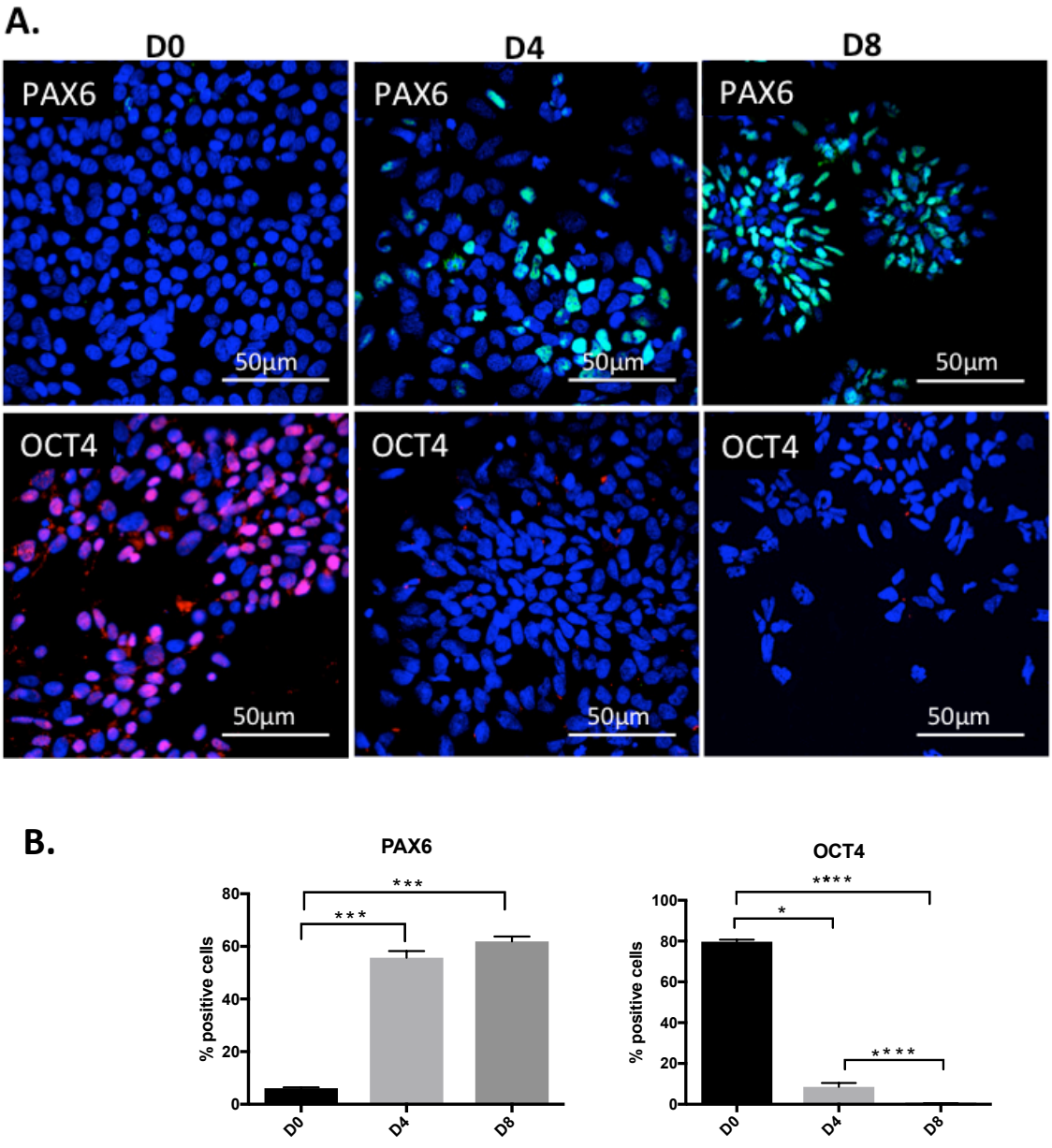



**Figure 3.7. Immunofluorescence analyses. A.** Expression of NE markers and OCT4 pluripotency marker at d0, d4 and d8. The cells nuclei are stained with Hoescht 33342. **B.** Results of the immunofluorescence statistical analyses.

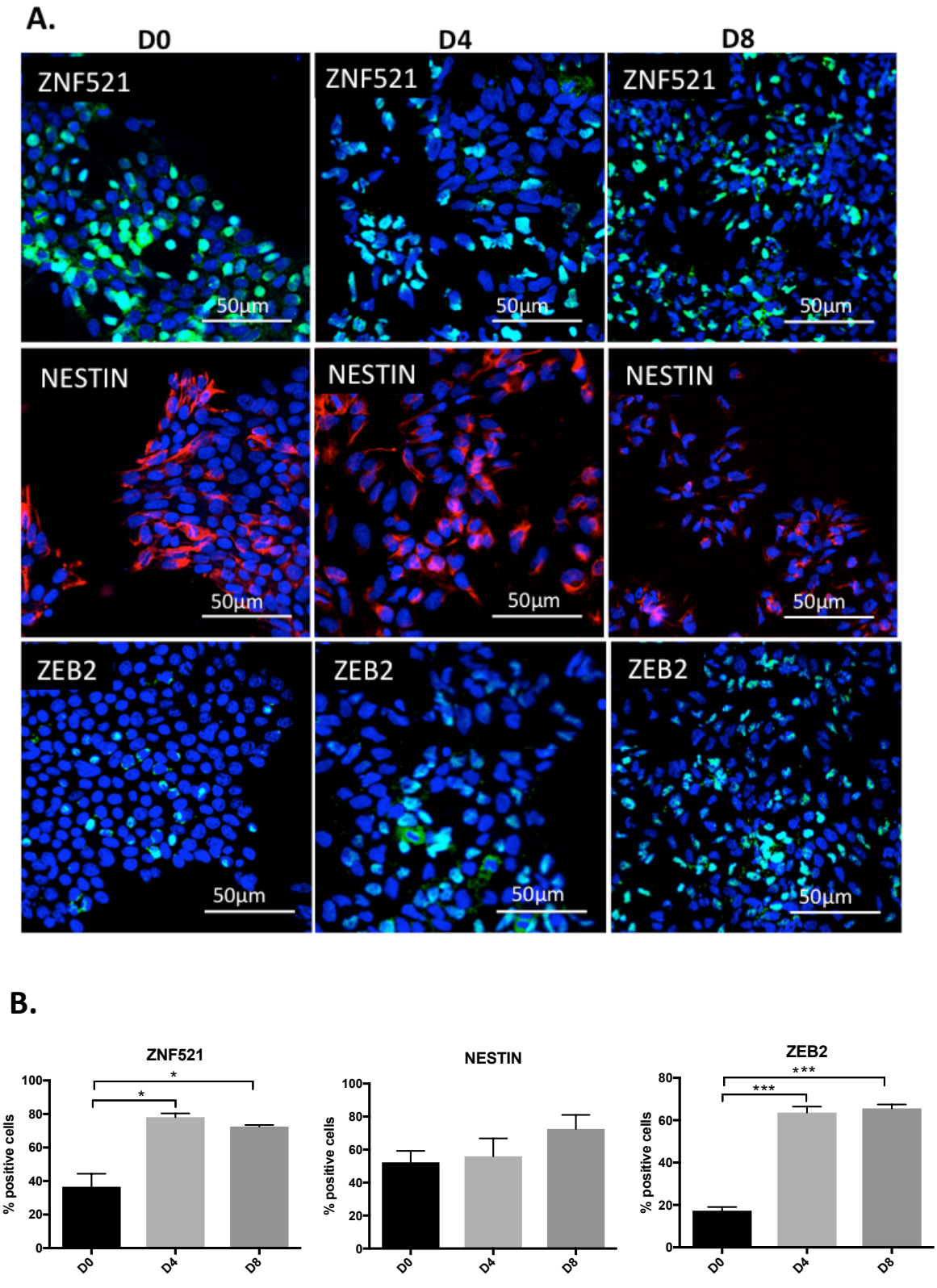



**Figure 3.7. Continued**

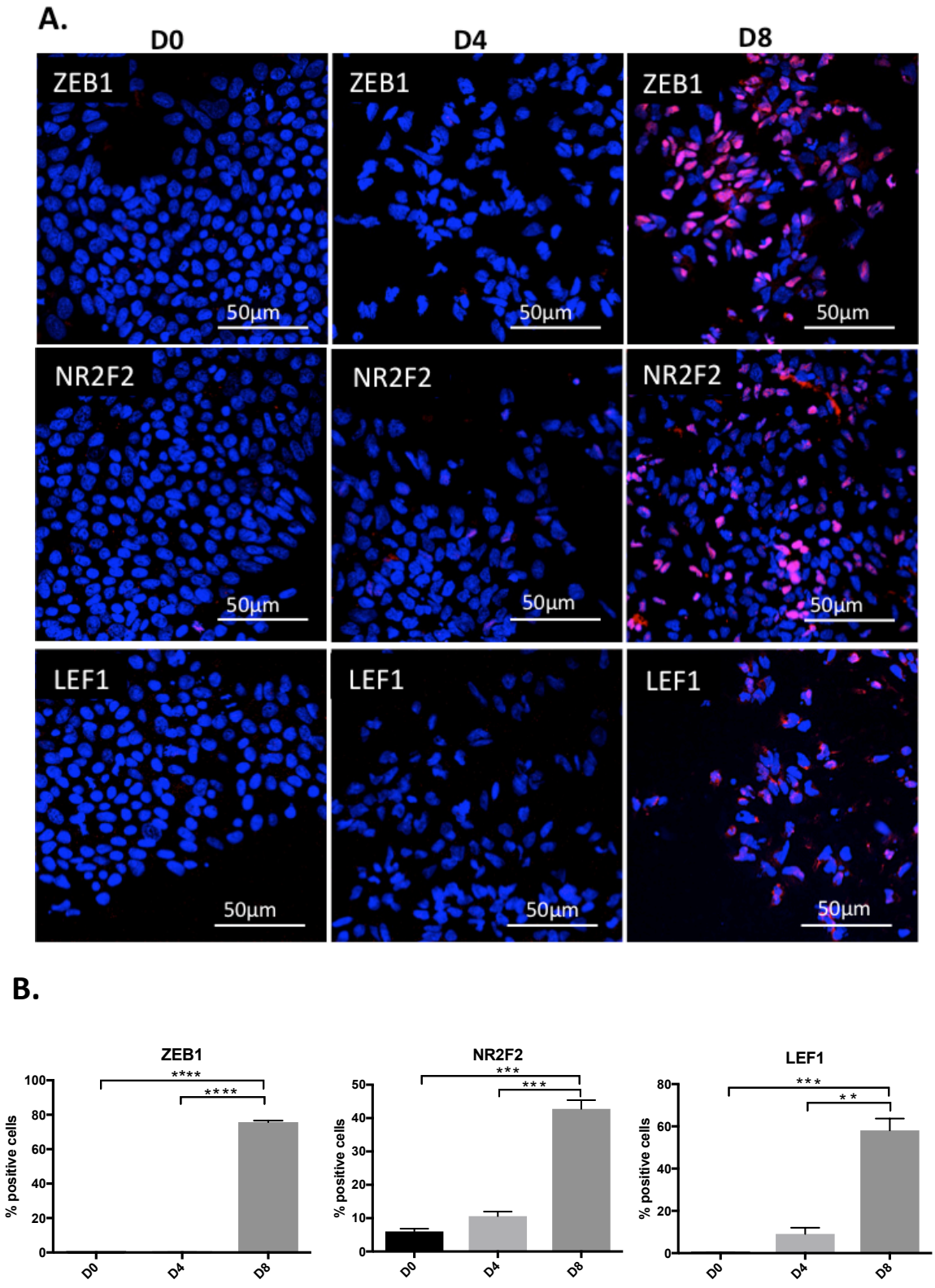



**Figure 3.7. Continued**

### 3.3.4. Differentially expressed genes and transcription factors

Statistical analyses were performed to identify TF and genes differentially expressed between consecutive time points or between time points d0 and d8. Significantly differentially expressed ($p < 0.05$) genes were defined as those with > 1 log2 fold change expression levels between the 2 data sets. The analyses detected 4360 up-regulated genes and 4858 down-regulated genes during 2i NE induction. Among these genes, 459 correspond to TFs differentially expressed between consecutive time points and 108 unique TFs differentially expressed comparing d0 with d8. The identities of the differentially expressed TFs, time points, fold changes and p-values are described in the Appendix 3.2 (CD-ROM) for up-regulated and Appendix 3.3 (CD-ROM) for down-regulated genes.

4650 up-regulated genes and 5180 down-regulated genes were identified during differentiation of iPSC with 2i-WNT. 434 TFs are included in the number of differentially expressed genes from sequential time points and 134 unique TF differentially expressed between d0 with d8. Statistical analyses are detailed in Appendix 3.4 (CD-ROM) for up-regulated and Appendix 3.5 (CD-ROM) for down-regulated.

476 common TFs are differentially expressed in 2i media versus with 2i-WNT media. TFs expression patterns from d0 to d8 are depicted in a heatmap (Figure 3.8). The heatmaps show that several TFs highly expressed during the initial 2 days of differentiation are subsequently down-regulated, whereas initially low expressed TF are up-regulated after the initial 2 days of induction. Additionally, a smaller number of TF have a peak of expression at d2, d3 and d4. Names of individual TF were omitted from the heatmaps due to the large number of genes contained in each data set.

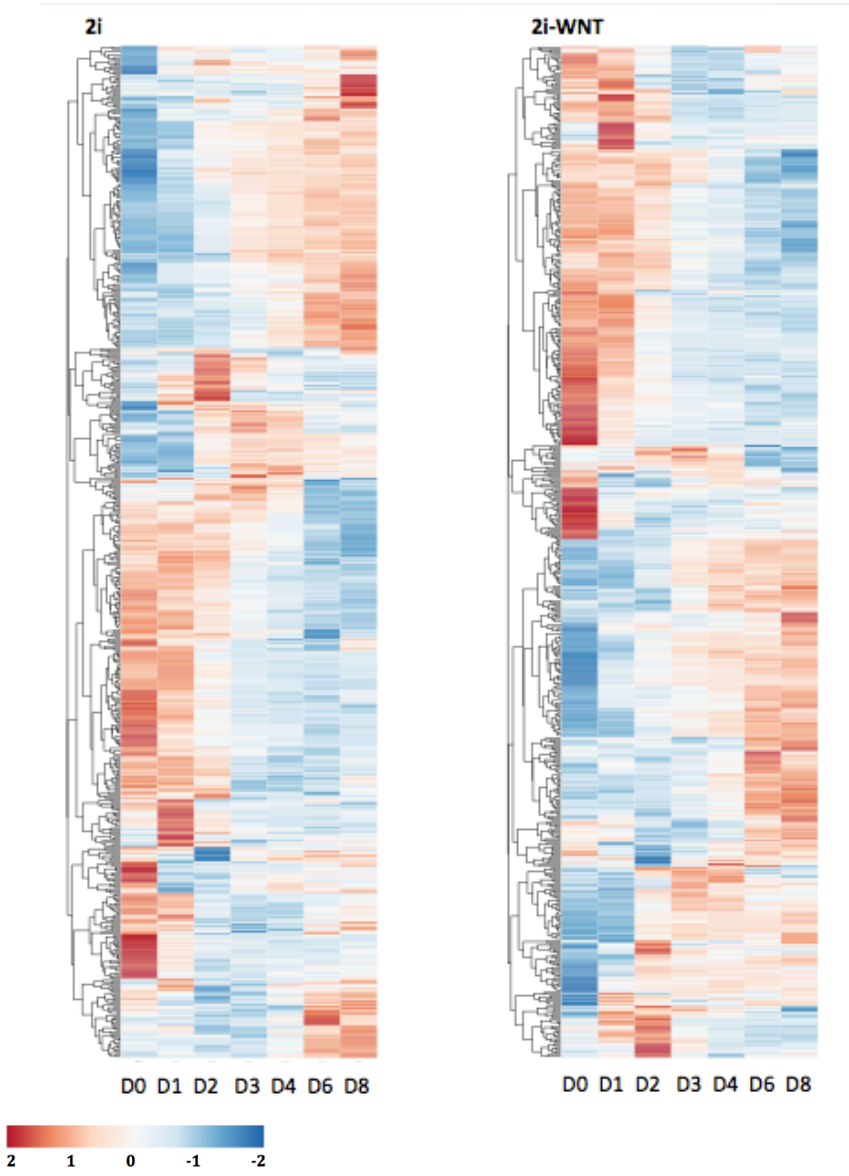



**Figure 3.8. Heatmap of transcription factors differentially expressed during neuroectoderm differentiation with 2i and 2i-WNT media.** The heatmaps illustrate gene expression patterns of TF differentially expressed during neural induction at d0, d1, d2, d3, d4, d6 and d8 with 2i and 2i-WNT induction media. Due to the large number of differentially expressed genes, the names of individual TF have been omitted. Gene names and statistical analyses details can be found in the appendix 3.2. to 3.5 (CD-ROM) . The key colour scale indicates high 2 (dark red) or low -2 (dark blue) expression levels compared with each gene mean.

Significantly up-regulated TFs with > 1.5 log2 fold changes between consecutive time points and > 1 RPKM values are depicted in heatmaps to facilitate visualization of their expression patterns. TFs that depicted peaks of expression during the different stages of

differentiation are represented in the Figure 3.9. Among these TFs are NE progenitor markers such as IRX2, IRX3, IRX5, ZIC1 and HESX1. TFs whose expression levels increase between consecutive time points to d8, are shown in a separate heatmap (Figure 3.10) and comprise NE markers such as NESTIN, ZNF521, SOX1, SIX3, LEF1, NR2F2 and forebrain progenitor markers including PAX6, OTX1, RAX, LHX2 and FOXG1.

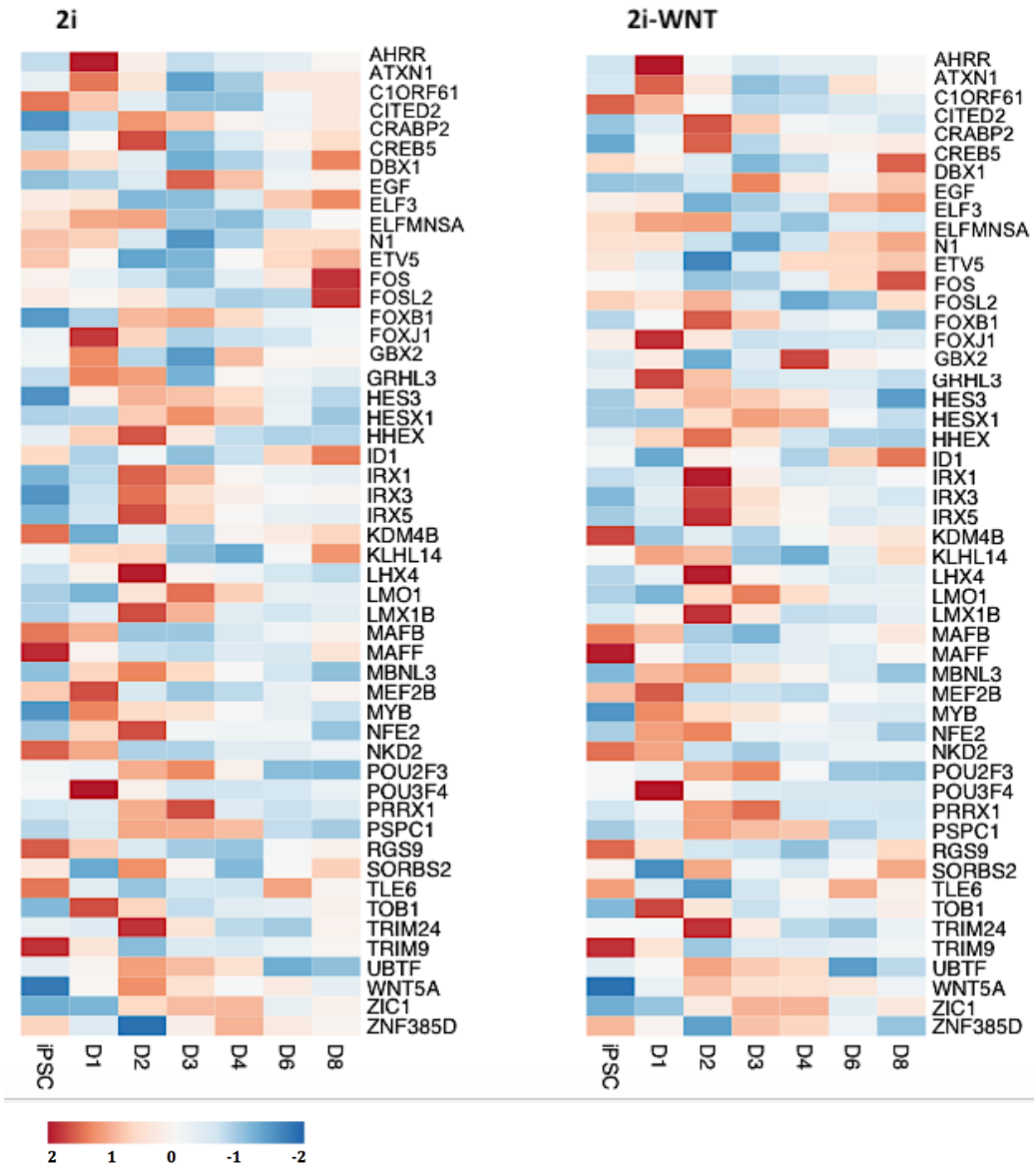



**Figure 3.9. Heatmaps of transcription factors differentially up-regulated with > 1.5 log2 fold changes between consecutive time points that show peak of expression levels during neuroectoderm differentiation.** The heatmap illustrates the expression of 49 up-regulated TF during neural induction with both 2i and 2i-WNT induction media. The key colour scale indicates high 2 (dark red) or low -2 (dark blue) expression levels compared with each gene mean.

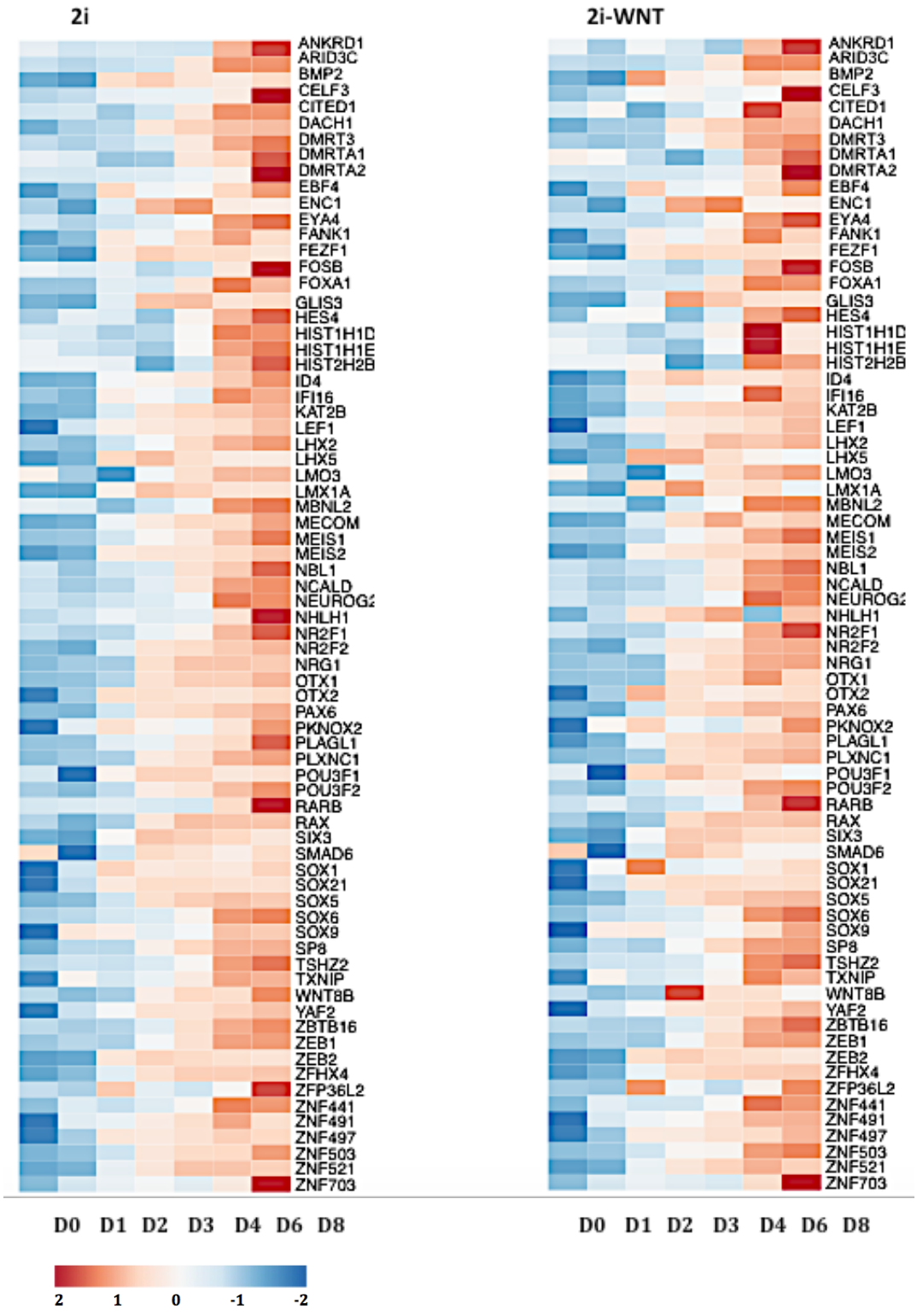



**Figure 3.10. Heatmaps of transcription factors differentially up-regulated with > 1.5 log2 fold changes from consecutive time points.** The heatmap illustrates expression of 73 TFs up-regulated during neural induction with 2i induction media and 59 TFs induced with 2i-WNT media. The key colour scale is the same as described in the heatmaps above.

Several TF depict reduction in gene expression between consecutive time points but are significantly differentially expressed between d0 and d8. These TF (not included in the consecutive time points data) are illustrated in a heatmap (Figure 3.11) and include the NE markers SOX2, SOX11 SIX6 and HES5 TF.

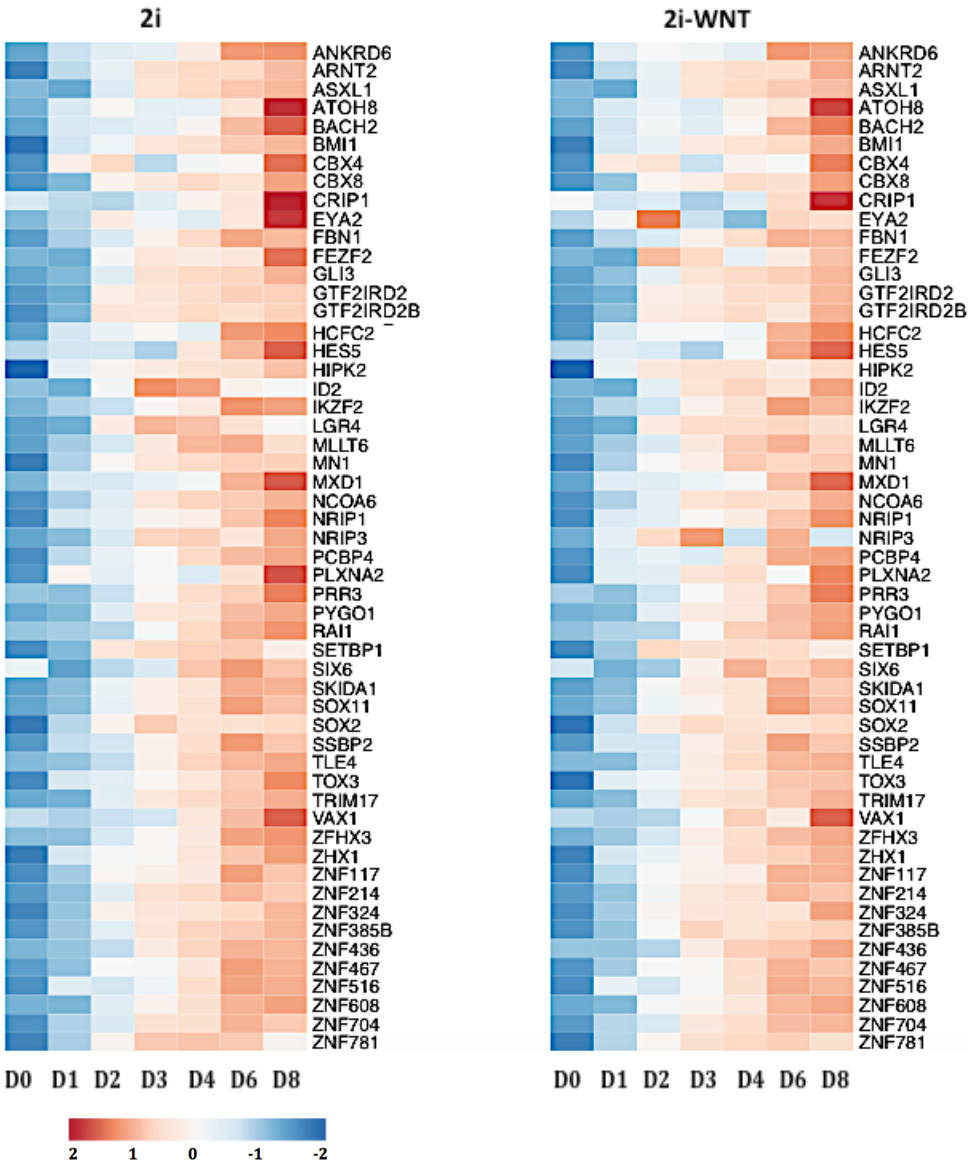



**Figure 3.11. Heatmaps of transcription factors differentially up-regulated with more than 1.5 log2 fold changes from time points d0 to d8.** The heatmap illustrates the expression of 54 and 64 TF up-regulated during neural induction with 2i and 2i-WNT media respectively. The key colour scale is the same as described in the heatmaps above.

Transcription factors significantly down-regulated between consecutive time points with > 1.5 log2 fold changes and > 1 RPKM values are represented in a heatmap (Figure 3.12) and include the pluripotency/ self-renewal markers NANOG and MYC.

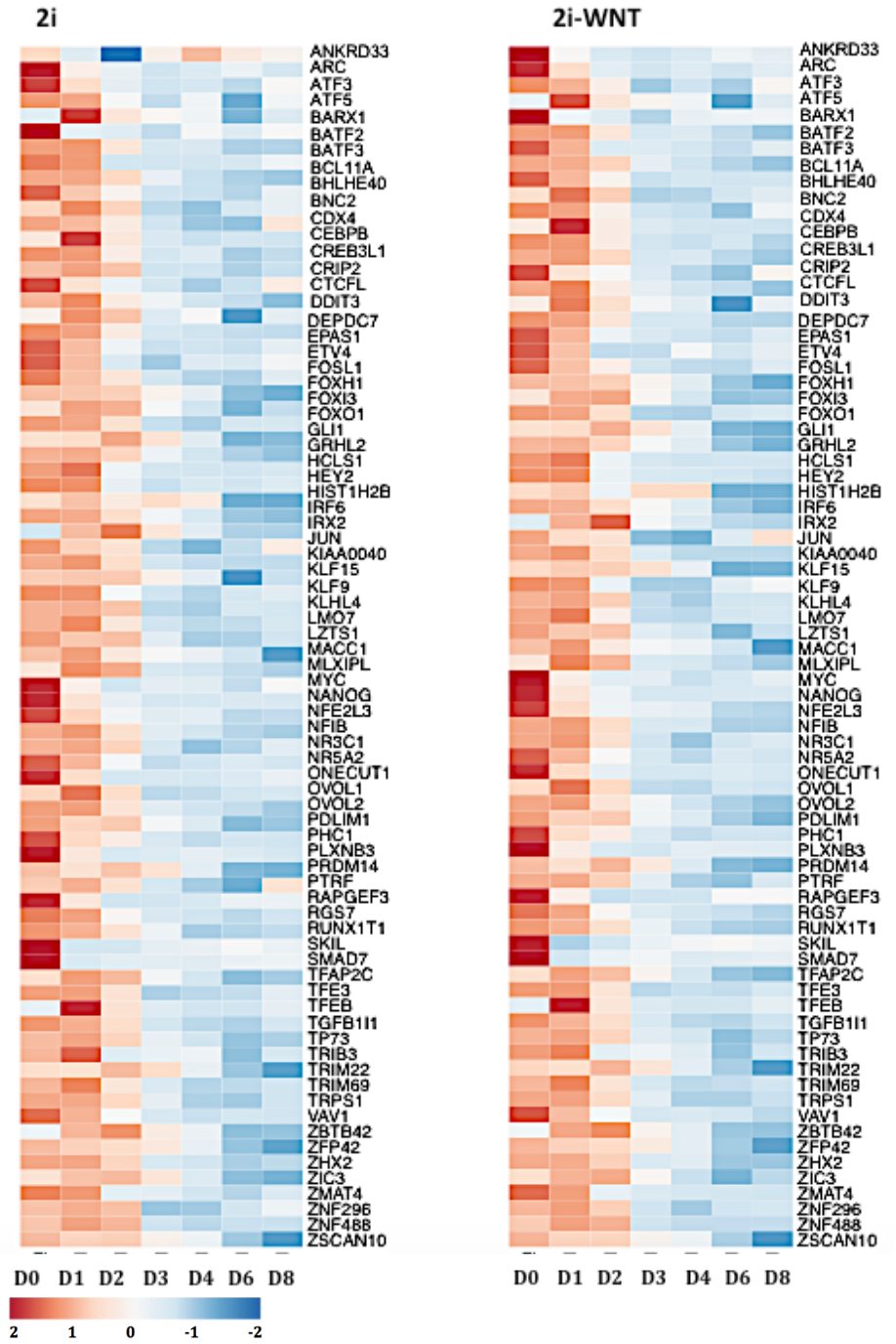



**Figure 3.12. Heatmaps of transcription factors differentially down-regulated with > 1.5 log2 fold changes between consecutive time points.** The heatmap shows the expression of 77 and 80 down-regulated TF during neural induction with 2i and 2i-WNT induction media respectively. The key colour scale is the same as described in the heatmaps above.

Down-regulated TFs, whose expression does not change between consecutive time points but are nevertheless significantly differentially expressed between d0 and d8, are represented in a heatmap (Figure 3.13). The pluripotency marker OCT4 (POU5F1) is included in this map.

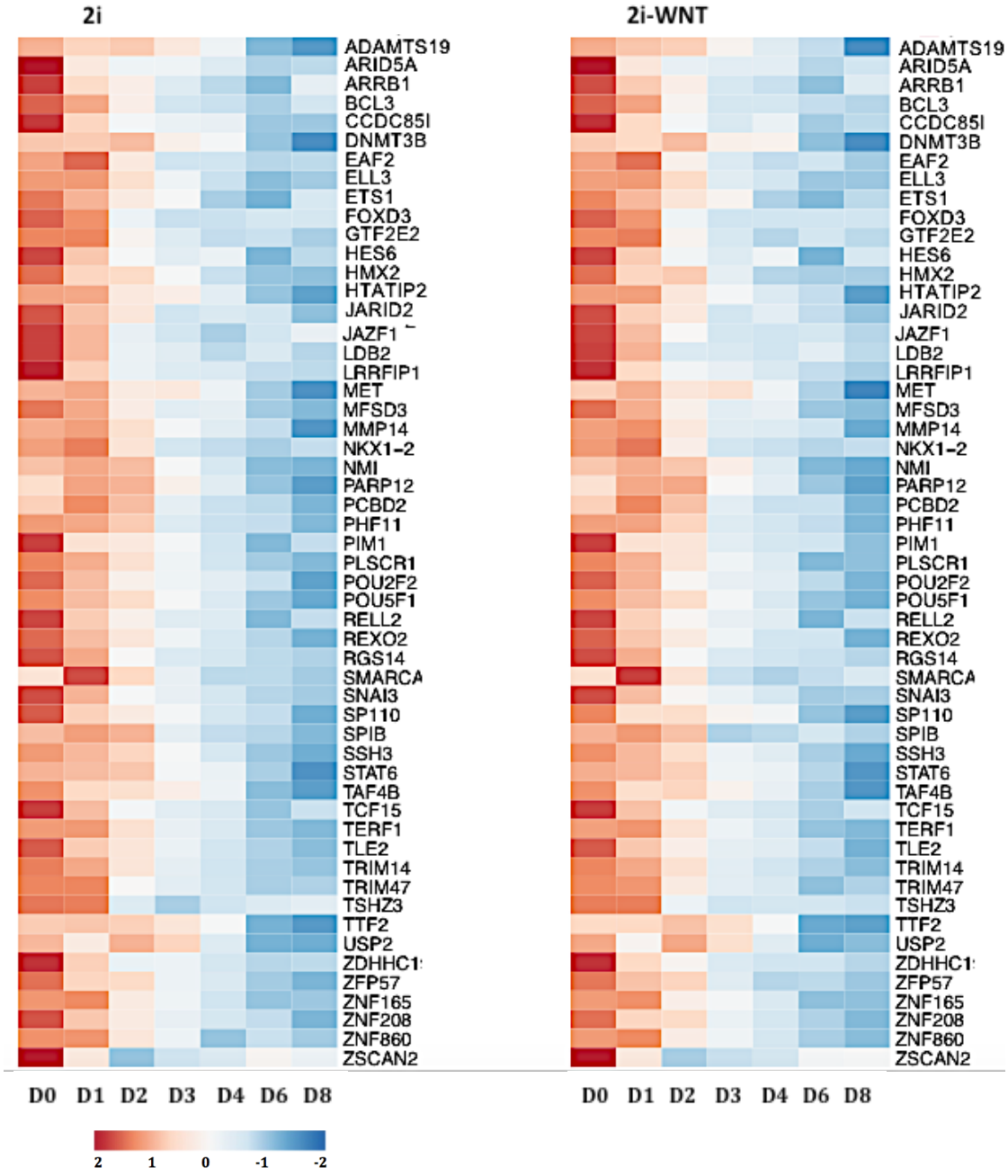



**Figure 3.13. Heatmaps of transcription factors differentially down-regulated with > 1.5 log2 fold changes from time points d0 to d8.** The heatmap illustrates expression of 54 TFs down-regulated during neural induction with 2i induction media and 71 TF with 2i-WNT media. The key colour scale is the same as described in the heatmaps above.

### 3.3.5. Gene ontology of differentially expressed genes

The differentially up-regulated genes with > 1 log2 fold changes in expression between consecutive time points or between time points d0 and d8 were interrogated for enrichment of specific biological process. Gene ontology (GO) analyses were carried out with the complete list of genes and TFs significantly up-regulated during differentiation with 2i (4360 genes) and 2i-WNT (4650 genes) induction media. Biological process annotations are depicted in Figure 3.14. Consistent with previously described results, GO analyses show significant enrichment (fold enrichment > 1.7, p = < 0.05) of the biological process: “diencephalon development”, "forebrain development”, “telencephalon development”, “nervous system development”, and “neurogenesis and tube development”. The biological process terms, fold enrichment and p-values are similar between NESC differentiated with 2i and 2i-WNT induction media. The sub-ontologies, p-values, number of genes for the corresponding term, fold enrichment and false discovery rate are described in the Appendix 3.6 and Appendix 3.7.

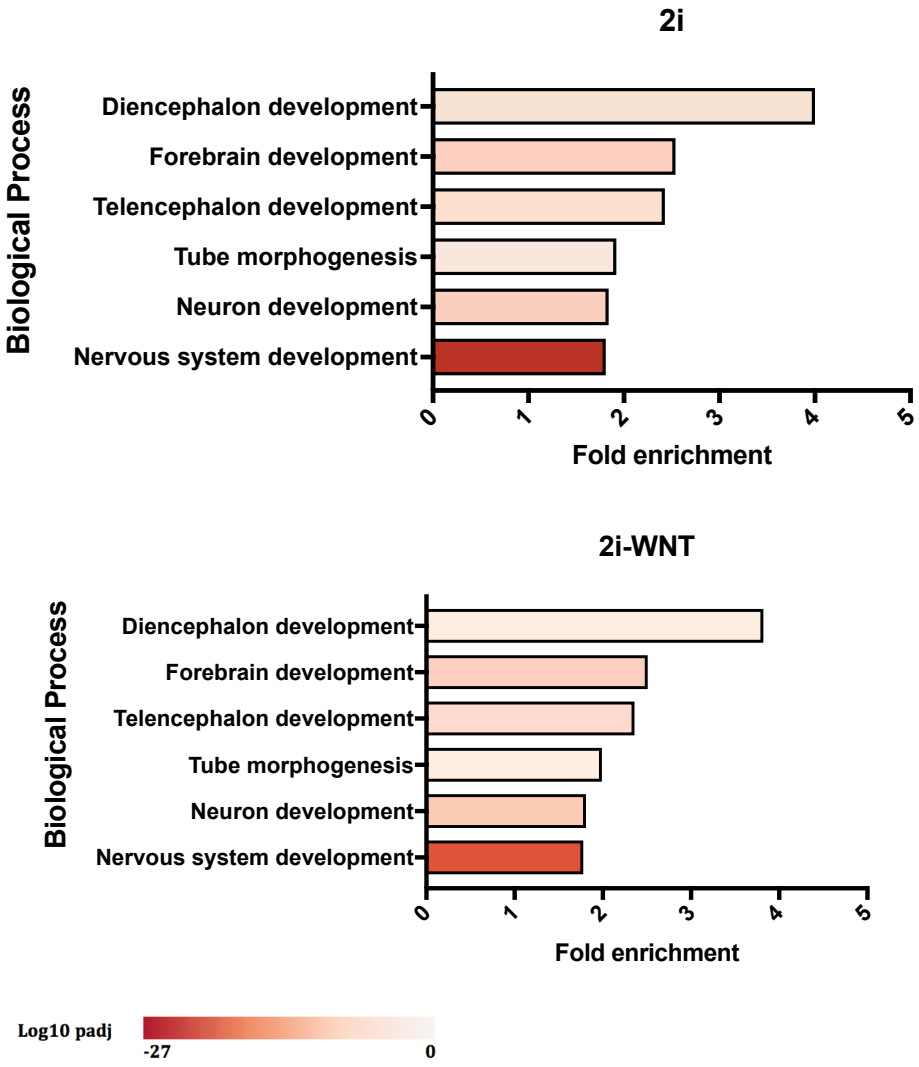



**Figure 3.14. Gene ontology of genes differentially up-regulated during neuroectodermal induction.** GO terms enrichment of genes up-regulated during NE differentiation with 2i (top) and 2i-WNT (bottom) induction media. The key colour scale indicates the log10-adjusted p-value, where the lower p-value is represented in dark red.

Similarly, GO analyses were carried out to interrogate the biological process enrichment in genes differentially down-regulated during NE differentiation with 2i (4858 genes) and 2i-WNT (5180 genes) induction media (Figure 3.15). GO analyses showed significant enrichment (fold enrichment > 1.3, p = < 0.05) for "vascular development", "ion transport",

"mesenchyme development", "cell motility", "MAPK cascade" and "epithelium development". Details of GO results are described in the Appendix 3.8 and Appendix 3.9.

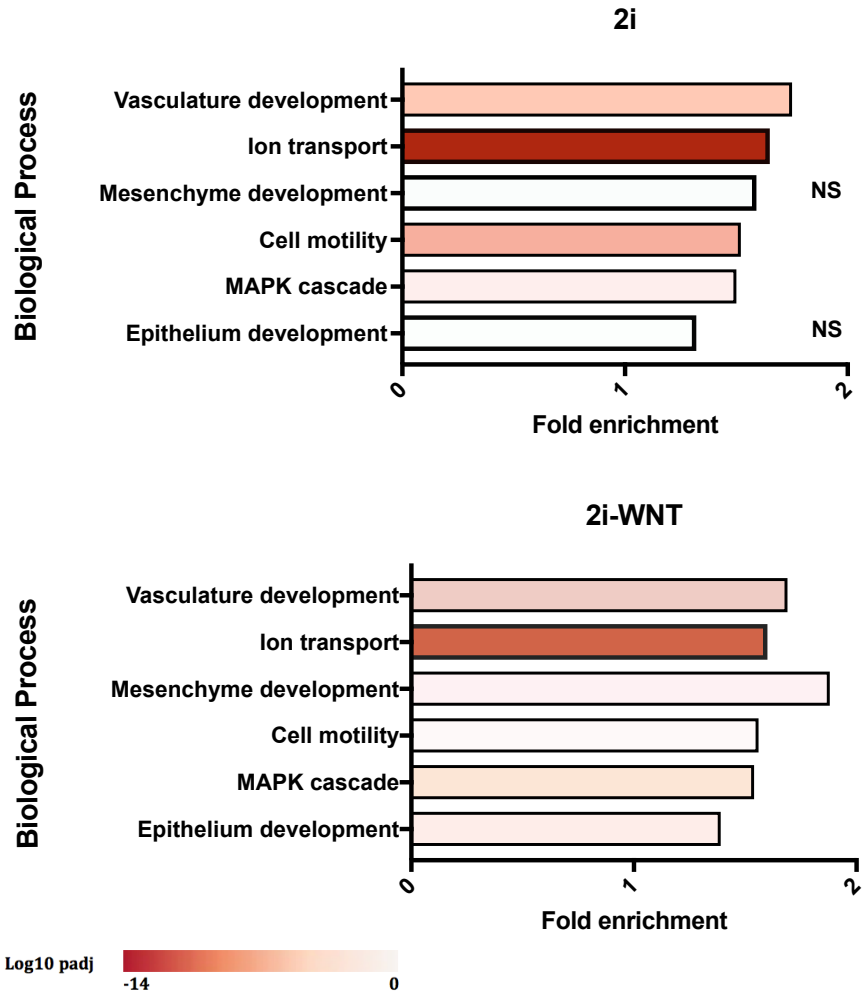



**Figure 3.15. Gene ontology of the genes differentially down-regulated during neuroectodermal induction.** GO terms enrichment of genes down-regulated during NE differentiation with 2i (top) and 2i-WNT (bottom) induction media. The key colour scale indicates the log10 adjusted p-value. The not significant terms are denoted as NS.

### 3.3.6. Short time series expression miner (STEM) analyses

The STEM analyses generated 100 gene expression profiles with a minimum of 2-fold difference between at least 2-time points, including 18 profiles with significant number of assigned genes (Figure 3.16). The profile 65 showing up-regulation (850 genes), profile 31 down-regulation (1315 genes) and the profile 99 depicting biphasic expression (75 genes) were considered of interest since they might include genes necessary for pluripotency, neural differentiation or for coupling of these stages. Additionally, profile 22 was selected due to the high significance (359 genes) matched with this profile.

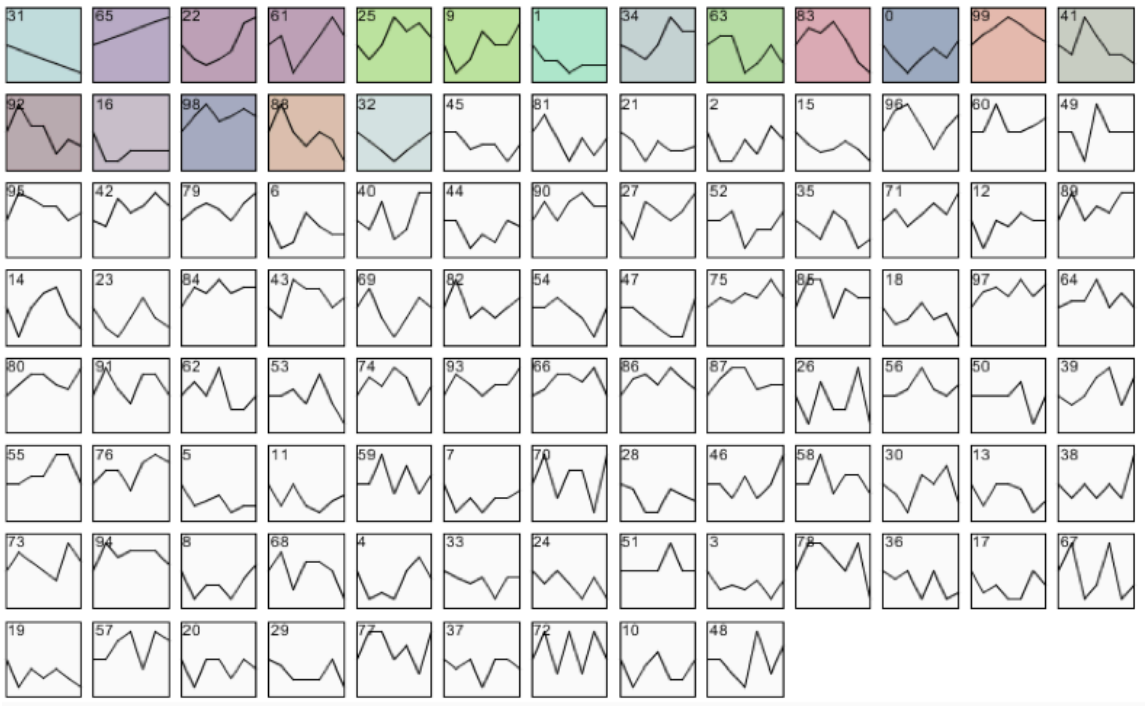



**Figure 3.16. Gene expression profiles generated by STEM for 2i and 2i-WNT treatments.** The profiles are ordered from top left to bottom right by the number of genes in the profile. The significant profiles are depicted with colour.

### 3.3.7. Gene ontology of selected STEM profiles

Genes comprising STEM genes expression profiles 22, 31, 32, 65 and 95 were interrogated for biological process enrichment (Figure 3.17). Gene ontology analyses of profiles 22, 32 and 99, which have a reduced number of genes assigned (<75), did not show significant enrichment for any biological process. 1315 genes were assigned to the STEM profile 31 during NE differentiation with 2i media and 1447 with 2i-WNT media. GO analyses show significant enrichment (fold enrichment > 1.3, p = < 0.05) for the biological processes "transport", "regulation of cell proliferation", "defense response", "angiogenesis", "regulation of cell adhesion" and "hypoxia" (Figure 3.19\8). The biological process terms that are common to 2i and 2i-WNT neural induction show similar fold enrichment (with differences in respective p-values). The GO term 'response to hypoxia' appears to be specific to cells induced with 2i-WNT, since it does not appear enriched in profile 31 in cells induced with 2i media. The sub-ontologies, p-values, number of genes for the corresponding term, fold enrichment and false discovery rate are described in Appendix 3.10 and Appendix 3.11.

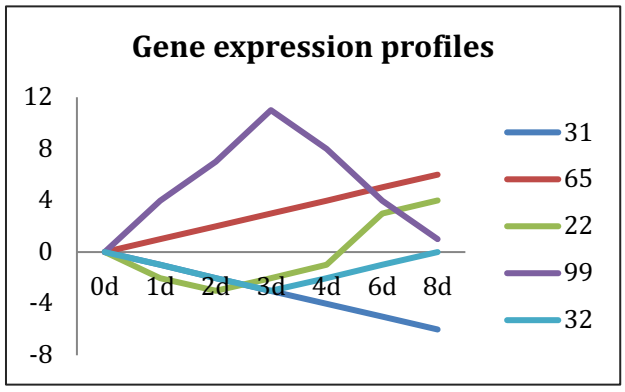



**Figure 3.17. Selected gene expression profiles.** Gene expression profiles of 5 significant gene clusters selected for gene ontology analyses.

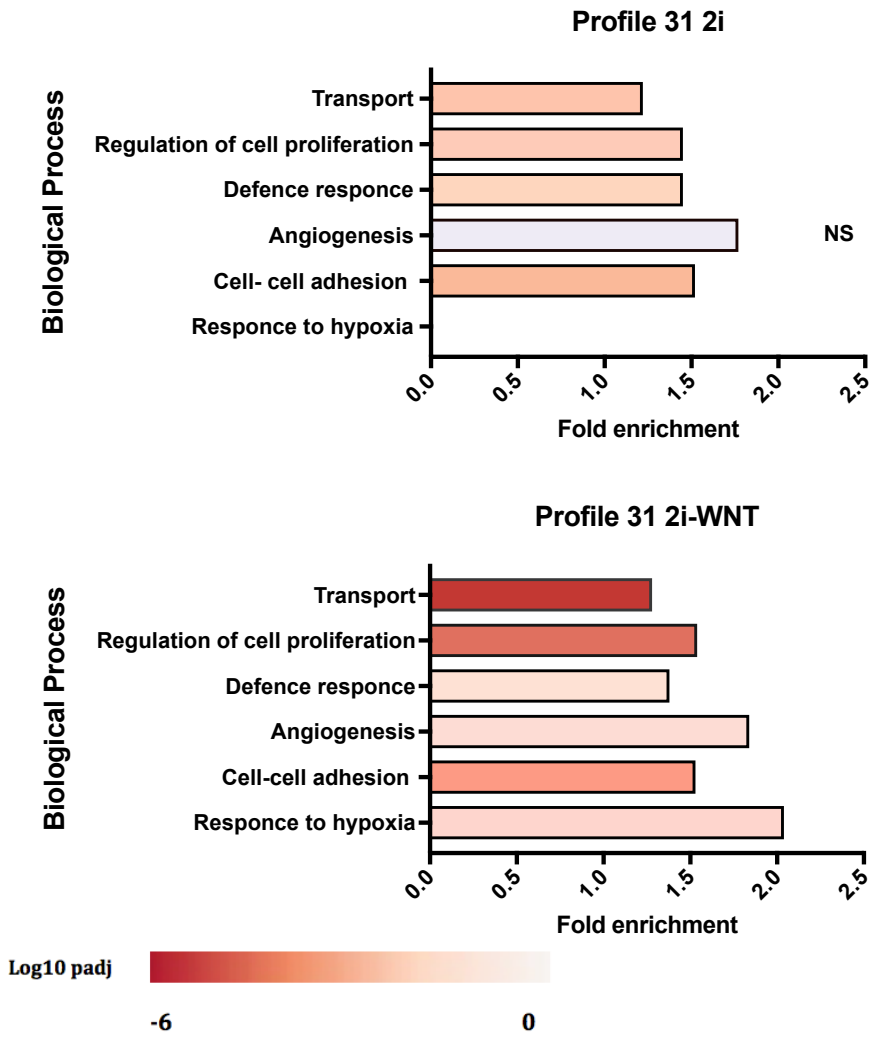



**Figure 3.18. Gene ontology of the genes assigned to STEM profile 31.** GO terms enrichment of genes that matched the profile 31 during NE differentiation with 2i (top) and 2i-WNT (bottom) induction media. The key colour scale indicates the log10-adjusted p-value, where the lower p-value is represented in dark red. The terms that are not significant are denoted with NS.

STEM profile 61 matched the gene expression levels of 850 genes during neural induction with 2i and 950 genes with 2i-WNT. GO analyses identified significant enrichment (fold enrichment > 2, p = < 0.05) for the biological processes "brain", "forebrain" and "telencephalon development", "synapse organization" and "neural precursor cell proliferation" (Figure 3.19). GO terms and fold changes are similar between cells

differentiated with both neural induction protocols. GO analyses details are described in the Appendix 3.12 and Appendix 3.13.

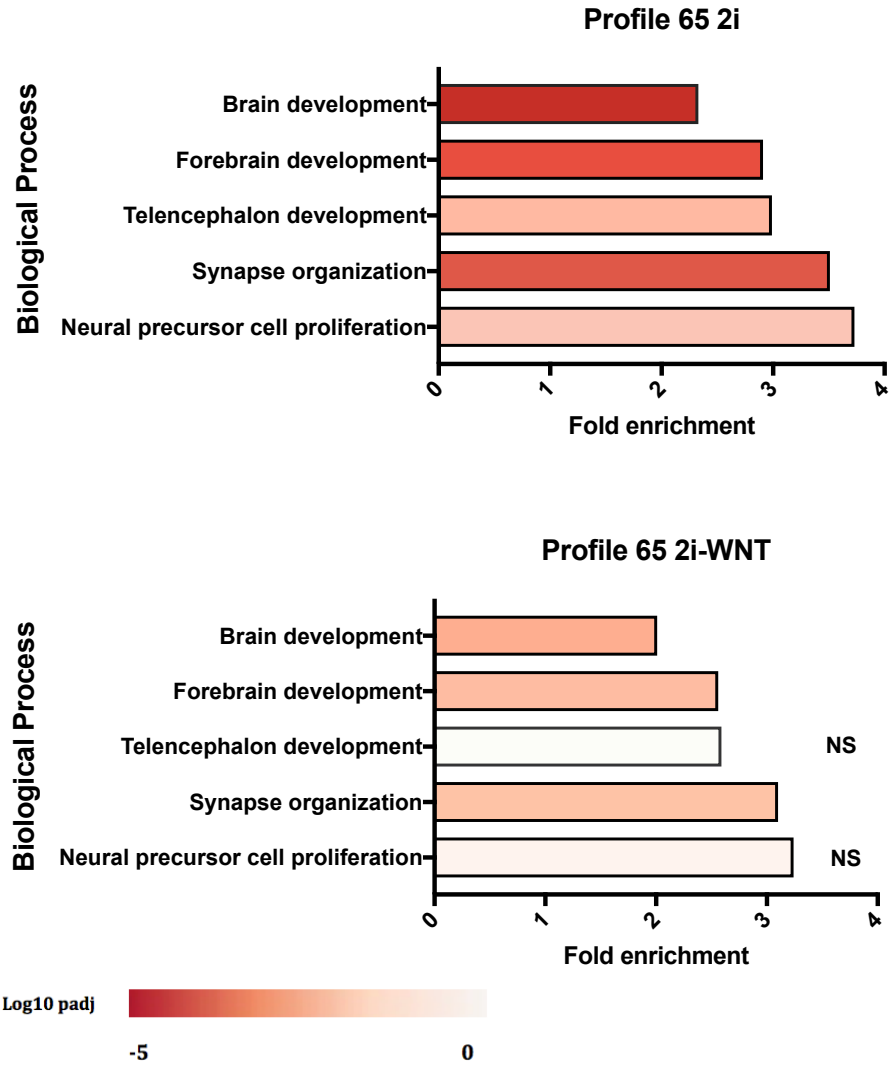



**Figure 3.19. Gene ontology of genes assigned to the STEM significant profile 65.** GO terms enriched during NE differentiation with 2i (top) and 2i-WNT (bottom) induction media are shown. The key colour scale is the same as above.

### 3.3.8. Gene expression profiles

We examined the expression levels of mesoderm, endoderm and non-neural ectoderm markers, to determine if they are actively expressed in the iPSC cells acquiring NE fate. (Figure 3.20). NBZ markers were analysed since in the mouse embryo, inhibition of WNT

signalling in the region between the neural and non-neural ectoderm leads to the differentiation of the NBZ, which subsequently forms the PPC that gives rises to the GnRH neurons. Hence, NBZ markers expression was examined to determine if the differentiating cells depicted NBZ characteristics.

The genes markers were selected since they are expressed in the different germinal layers of the embryo and in the NBZ. However, the function of these sets of genes during the formation, maintenance and regionalization of the mesoderm, endoderm, non-neural ectoderm and NBZ greatly differs. Hence, the selected genes might not be adequate to determine the formation of the 3 germ layers and the NBZ. For instance, the mesodermal markers T, MIXL1 and MESP1 are expressed during the initial stages of mesodermal differentiation from human ESC (Evseenko *et al*., 2010). Whereas, RUNX1 and GATA2 are expressed in the later mesoderm and play a role as endothelial and hematopoietic promoters in human embryoid bodies (Shi *et al*., 2014). Furthermore, the regulatory function of several genes is context dependent. In fact, the mesodermal marker T is also able to induce the expression of endodermal markers FOXA2 and SOX17 if activated along with SMAD2/3 (Faial *et al*., 2015). The endodermal markers Gata4 and Foxd3 promote endodermal differentiation of ES in mouse, whereas Sox7, Sox17, Afp and Foxa2 are expressed at latter stages of endodermal formation (Zaret & Carroll., 2011; Wang *et al*., 2013b; Kinoshita *et al*., 2015). The genes Tfap2a, Gata3, Msx2 and dlx3 are rapidly activated in mouse ESC induced to acquire a non-neural ectoderm fate by the inhibition of Wnt and Tgfβ and are among the genes that promote the separation of the non-neural ectoderm and the NE (Ealy *et al*., 2016). The NBZ markers were described in Xenopus and Tfap2a, Msx1 and Pax3 have been validated in human NBZ lines induced from iPSC/ESC, whereas Foxi1 and Dlx6 have been validated in chick and might depict differences with the expression in the human NBZ cells (Leung *et al*., 2016; Shigetani *et al*., 2016).

Expression of the selected number of genes activated in specific germ layers or NBZ were plotted to depict their gene expression patterns. Since the expression of these genes was similar between cells induced with 2i and 2i-WNT media, the expression showed is an average of the RPKMs results from the 3-iPSC induced with 2i and 2i-WNT media. The results indicated that the gene expression of the majority of the genes analysed was maintained at low levels (< 1 RPKM) during the time course of NESC differentiation, which might indicate that the majority of these genes are not expressed in the cells or are expressed in a reduces number of cells in the population. However, the endodermal markers FOXD3 and AFP show an initial expression higher than 2 RPKM at d0 , but the expression levels was rapidly subsequently down-regulated during NESC induction.

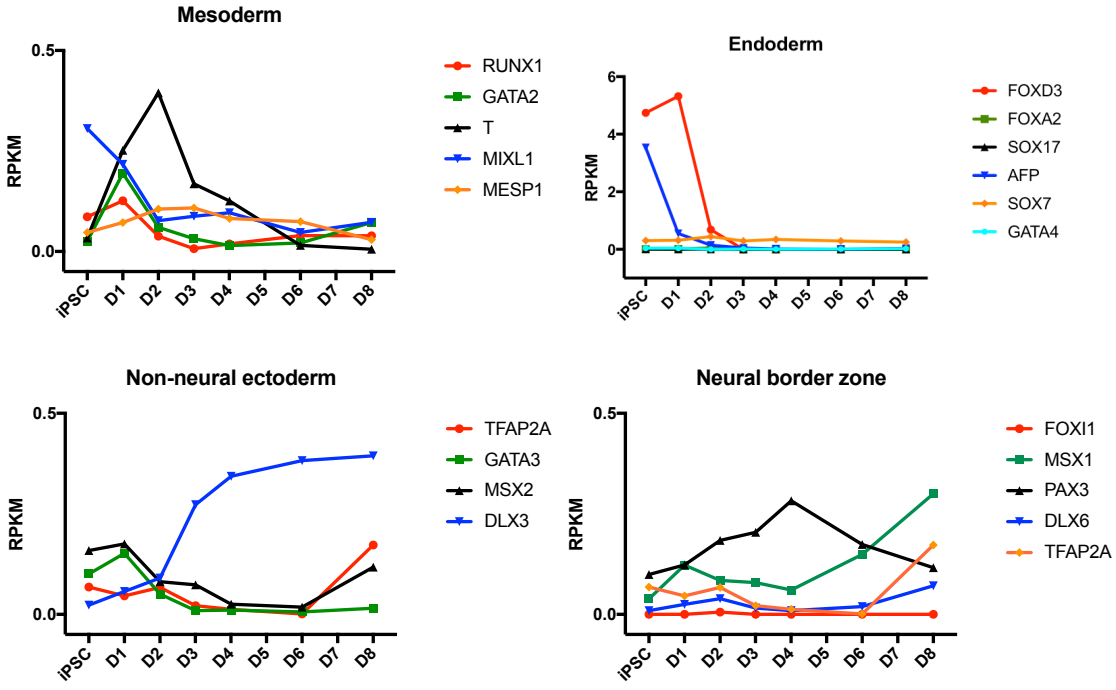



**Figure 3.20. Gene expression profiles of mesoderm, endoderm, non-neural ectoderm and neural border zone markers**. Expression of each gene (RPKM) is represented with a different colour.

Similarly, the expression levels of genes that are activated in cell undergoing apical-basal polarity, rosettes, radial glia cells and cells acquiring forebrain fate were analysed to determine if these TF were precociously expressed during NESC induction (Figure 3.21).

Gene expression profiling studies have demonstrated that a group of genes including PLAG1, DACH1, MMRN1, DMRT3, ZBTB16, LIX1 and LMO3 genes are highly expressed in rosettes derived from human iPSC/ESC, embryoid bodies and ESC in mouse (Elkabetz *et al*., 2008; Abranches *et al*., 2009; Koch *et al*., 2009; Lukovic *et al*., 2017). However, the function that some of these factors play during rosette formation is not well understood. For instance, LIX1 gene has been implicated in motor neuron survival in felines and proliferation of stomach mesenchymal progenitors in the chick embryo, but its function in neural rosettes has not been defined (Fife *et al*., 2010; McKey *et al*., 2016). Hence, mechanism and stages in which these genes regulate rosette formation and their expression levels in rosettes might vary, indicating that they are not ideal rosette markers. The profile analyses of rosette gens indicated that DACH1, ZBTB16, LMO3 showed an increase on their expression from day 2 to day 8 after neural induction. Particularly, the gene LIX1 is highly up-regulated after day 2. Expression of these genes during NESC might indicate that they play a role this early stage of neural development or that a reduced subpopulation of cells might be precociously acquiring rosette identity. However, higher quantities of these markers might be necessary to induce rosette formation in the population. Neural differentiation of iPSC/ESC initiates with the formation of NESC that subsequently acquire cell polarity and reorganize to give rise to rosette like structures (Kageyama *et al*., 2014; Marchenco *et al*., 2014). The genes PARD6B and CDC42 form part of apical polarity complexes responsible for the apico-basal axis formation and construction and maintenance of apical junctions, whereas CRB2 localizes in the apical side of rosettes and interact with these apical-basal complexes and stabilise them.

Experiments in human ESC demonstrated that these genes are necessary for the differentiation of neural rosettes, since knowing down PARD6B, CDC42 and CRB2 leads to the disruption of rosette formation (Boroviak & Rashbass, 2011; Harding *et al*., 2014; Kedziora *et al*., 2016; Deglincerti *et al*., 2017). Similarly, studies have suggested that Uspx9 is necessary for embryonic progenitors in the mouse embryo, since ablation of the genes resulted in transient disruption of cell adhesion and apico-basal polarity (Premarathne *et al*., 2017). Our analyses indicated that the majority of these cell polarity markers are down-regulated during NESC induction, suggesting that the differencing NESC are not undergoing apical-basal polarization during this early stage. During rosette formation, neuro-ectodermal cells acquire glial features and become RG cells, which are able to form neurons, ependymal cells, astrocytes and oligodendrocytes (Kageyama *et al*., 2014; Marchenco *et al*., 2014). The expression of FABP7 and SLC1A3 is required for the formation, proliferation and migration of RG cells and are widely used as RG markers (De Rose *et al*., 2012; Pollen *et al*., 2016). The genes TNC and S100B were selected as a RG marker since its expression is detected in RG cells of the mouse neural tube (Hachem *et al*., 2007; Brozzi *et al*., 2009; Faissner & Reinhard, 2015). The expression of Hes1 and Hes5 was analysed since these factors play a role in neural progenitors self-renewal and inhibition of proneural genes (Kageyama *et al*., 2005; Dhanesh *et al*., 2016). The results indicated that the genes FABP7, HES1, HES5 and TNC are up-regulated during different stages of NESC induction. Importantly, cell polarization, formation of the neural rosettes and radial glia differentiation are mechanism that are tightly regulated during neural development. Hence, the heterogeneous expression of the selected markers during NESC differentiation, reflected the variability in the functions that these genes play during cell polarization and rosette and radial glia formation. Thus, heterogeneous expression of the

selected genes, indicates thatthey are not optimal markers for these neural developmental stages.

The expression of forebrain markers during neural induction was analysed to validate previous observations that indicated that NESC are primed to acquire forebrain fate during early stages of neural development. We analysed the expression of OTX1, OTX2, HESX1, SIX3, RAX, LHX2, FOXG1 and EMX2. The results indicate that most forebrain markers are highly up-regulated after d2 and the TFs HESX1 and SIX3 depicted a peak of expression at d3. The selected markers play different roles during the formation and regionalization of the forebrain. For instance, OTX1 and OTX2 regulate the differentiation of the caudal forebrain, whereas FOXG1 is required for the differentiation of the ventral telencephalon (Larsen *et al*., 2010; Manuel *et al*., 2010). Hence, the high variability in the expression of these markers during NESC induction is expected.

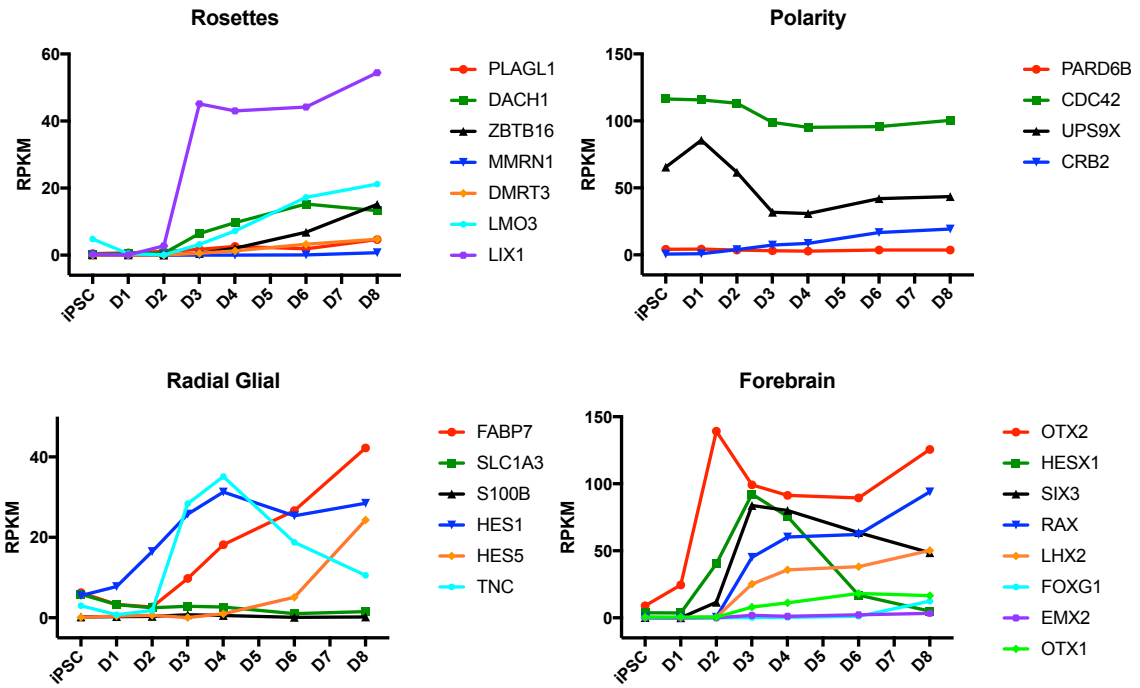



**Figure 3.21. Gene expression profiles of rosette, polarity, radial glial and forebrain markers**. Expression of each gene (RPKM) is represented with a different colour.

A recent microarray study reported up-regulation of the EMT marker ZEB1 and down-regulation of its epithelial targets OCLN, F11R, MARVELD2, PPL, DSG2, EPPK1, SH3YL1, DMKN, TMEM30B and MAL2s during neural induction of hESCs (Huang *et al*., 2016). Hence, expression of EMT markers and ZEB1 targets is depicted in Figure 3.22. The results show that ZEB1 ad N-cadherin (CDH2) are highly up-regulated after d3, whereas E-cadherin (CDH1) is down-regulated. Additionally, ZEB1 epithelial targets DMKN, F11R, MAL2, CDH3, DSG2, OCLN, PPL, SH3YL1, MARVELD2, EPPK1, SYTL1, PKP3, SFN, TMEM30B are down-regulated and its mesenchymal targets FN1, COL4A1ID1, ID2 and VIM are up-regulated.

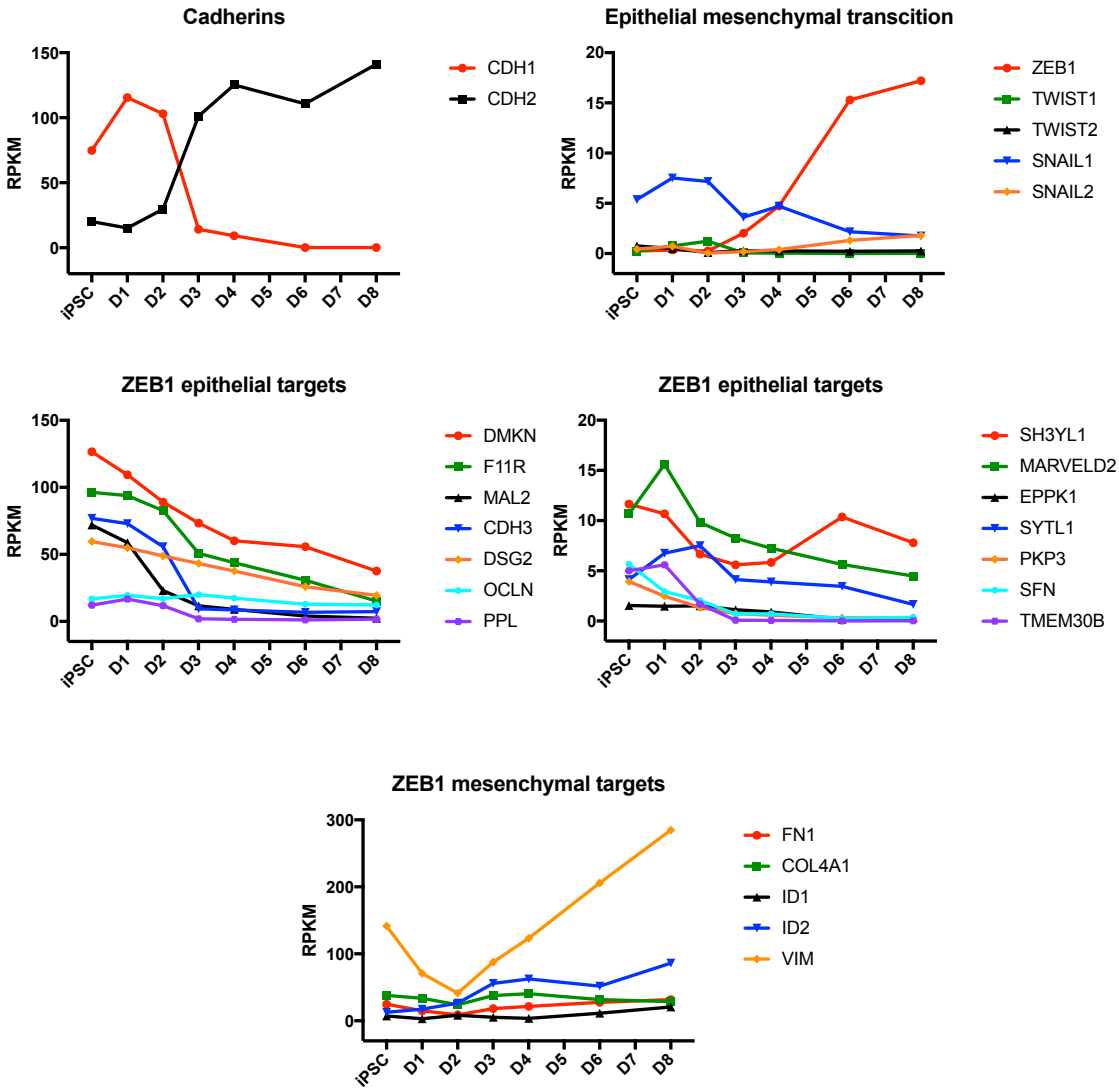



**Figure 3.22. Gene expression profiles of cadherins, epithelial mesenchymal transition and ZEB1 epithelial and mesenchymal targets**. The expression of each gene (RPKMs) is represented with a different colour.

Additionally, Huang *et al*., 2016 microarray study reported that the expression levels of genes WNT5B and RAR2, which are part of WNT/$Ca^{2+}$ signalling, were up-regulated during neural differentiation of human ESC. Likewise, the study shows significant changes in the expression of a number of DNA and histone modifiers. The authors suggested a potential role of WNT/$Ca^{2+}$ signalling and the analysed epigenetic modifiers during neuronal differentiation of ESC. However, the mechanism in which these WNT/$Ca^{2+}$

signalling and modifiers regulate neural differentiation remained to be studied. We analysed the expression of these genes to determine if it reflected the results reported by Huang *et al*, 2016. The results indicated that WNT5B and RAR2 are up-regulated during the course of NESC differentiation and that the epigenetic modifiers depicted similar changes in expression of those previously reported, validating Huang *et al*., 2016 data. Additionally, the gene expression profiles of the NE markers PAX6 and SOX1 were plotted (Figure 3. 23).

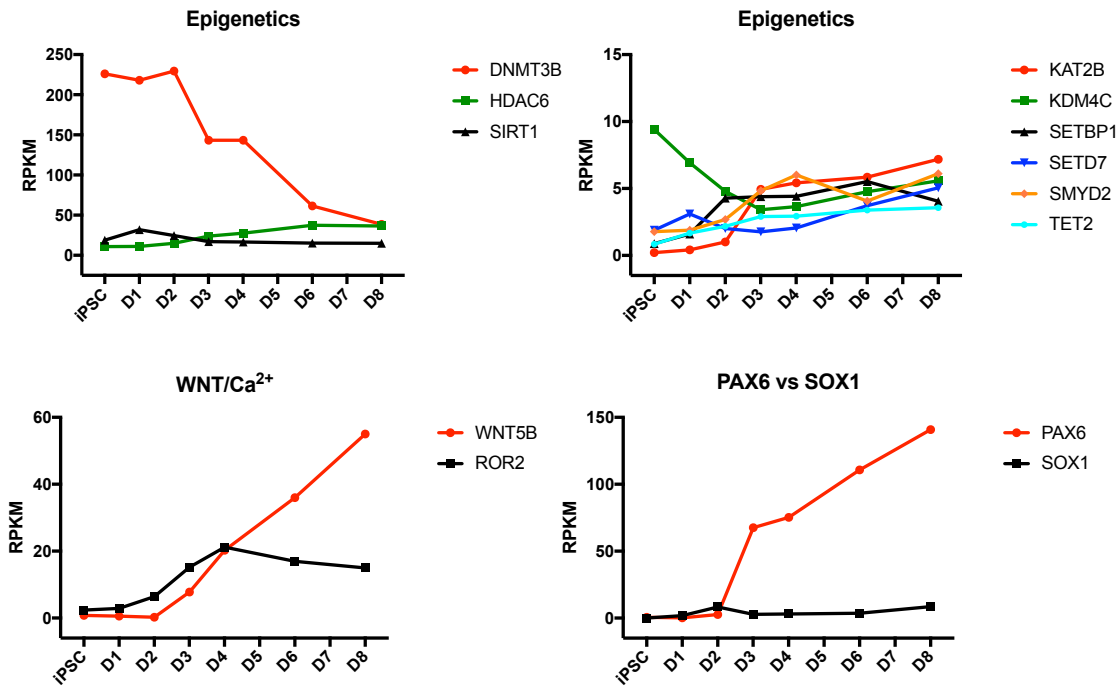



**Figure 3.23. Gene expression profiles of epigenetic modifiers, WNT/Ca2+ genes and PAX6 and SOX1**. The expression of each gene is represented with a different colour.

### 3.3.9. Dose response of XAV939 WNT signalling inhibition

β-Catenin concentration in the cell cytoplasm is tightly regulated by the destruction complex comprised of APC, GSK-3β, PP2A, CK1 and AXIN. The small molecule

inhibitor XAV939 blocks tankyrase 1 and tankyrase 2 enzymes that interact with the axin conserved domain and target it for ubiquitination. Hence, XAV939 selectively inhibits of β-catenin by stabilizing axin, which is the concentration-limiting component of the destruction complex (Huang *et al*, 2009a). A dose response curve was performed with 10 different concentrations of XAV939 ranging from 1nM to 10uM, to determine the optimal concentration of the small molecule that most effectively inhibits WNT signalling. HEK939 cells were transfected with a WNT expression plasmid and a luciferase reporter TOPflash vector, which contains TCF binding regions. As a negative control, cells were transfected with an empty plasmid and TOPflash. The negative control and WNT transfected cells were tested with 10 different doses of XAV939 inhibitor during 24h. Additionally, WNT transfected cells were tested in the absence of WNT inhibition as a positive control. The results indicate that the negative control does not induce any change in luciferase activity, whereas the positive control shows elevated levels of activity. Importantly, luciferase activity is reduced in WNT transfected cells with increasing concentrations of XAV939 doses so for instance, luciferase activity is not detected in cells treated with doses ≥ 2uM of XAV939 (Figure 3.24).

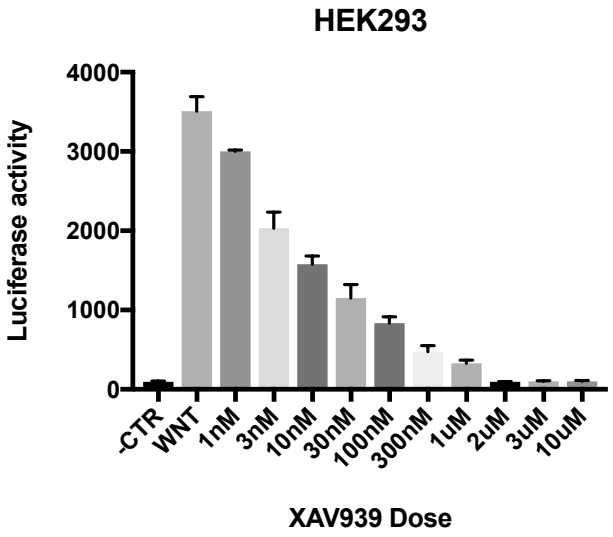



**Figure 3.24. Luciferase activity in HEK939 cells transfected with a WNT expression plasmid and a TOPflash vector**. From left to right, the negative control (-CTR) shows low levels of luciferase activity and the positive control (WNT) shows high levels of activity. WNT transfected cell luciferase activity is reduced in response to XAV939 doses from 1nM to 10μM.

### 3.3.10. Differential gene expression between iPSC derived NESC induced with 2i and 2i-WNT inhibitors.

Differential gene expression analyses were carried out to compare expression of all genes between iPSC induced with 2i and 2i-WNT at each time point during NESC differentiation. The results show that no genes were significantly differentially expressed (p-value < 0.05) and >1 log2 fold changes between the 2 treatments at time points d1, d2, d3 and d6. We identified 3 genes differentially up-regulated for the cells induced with 2i and 1 gene differentially up-regulated for cells induced with 2i-WNT media at time point d4 (Figure.3.25). The identities of the differentially expressed TF, fold changes and p-values are described in the Appendix 3.14 (CD-ROM) for 2i up-regulated and Appendix 3.15 (CD-ROM) for 2i-WNT up-regulated.

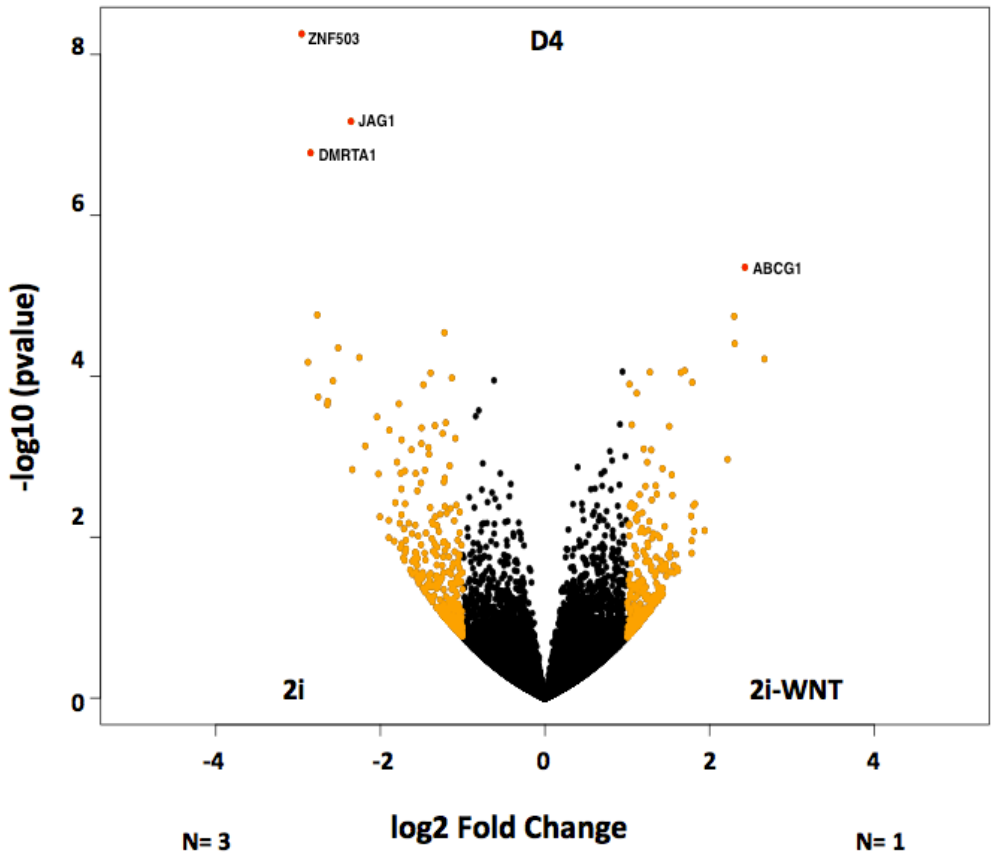



**Figure 3.25. Volcano plot of the differential expression of genes induced with 2i and 2i-WNT neutralization media at time point d4.** The X axis depicts the log2 fold change and the Y axis the –log10 p-value. The significantly DE genes with more than 1 log2 fold changes between conditions are represented with red dots. Genes with more than 1 log2 fold changes that do not reach significance (p-value < 0.05) between conditions are represented with orange dots. Genes whose expression levels are not significantly DE and have < 1 log2 fold changes in expression between the 2 treatments are shown with black dots. Genes that are up-regulated when differentiated with 2i media are located on the left and those that are up-regulated with 2i-WNT are shown on the right.

Additionally, 148 genes were differentially expressed (>1 log2 fold; p-value < 0.05) at d8. Among these genes (excluding TFs), 79 were up-regulated in cells induced with 2i media and 39 in cells induced with 2i-WNT media. To determine if the 2i-WNT up-regulated genes has been previously related with WNT signalling, enrichment for "pathways", "protein domain" and "protein-protein iteraction" was determined through GO analyses.

The results indicated that the genes up-regulated in NESC induced with 2i-WNT media, have not been previously associated with WNT signalling. The DE genes are depicted in the Figure 3.26. TF are show in a different figure to facilitate visualization of the gene names. Only the name of the most significant differentially expressed genes with higher fold changes is included in the figure to facilitate their visualization. Names of the entire set of genes and statistical analyses details are included in the Appendix 3.14 (CD-ROM) for 2i up-regulated and Appendix 3.15 (CD-ROM) for 2i-WNT up-regulated.

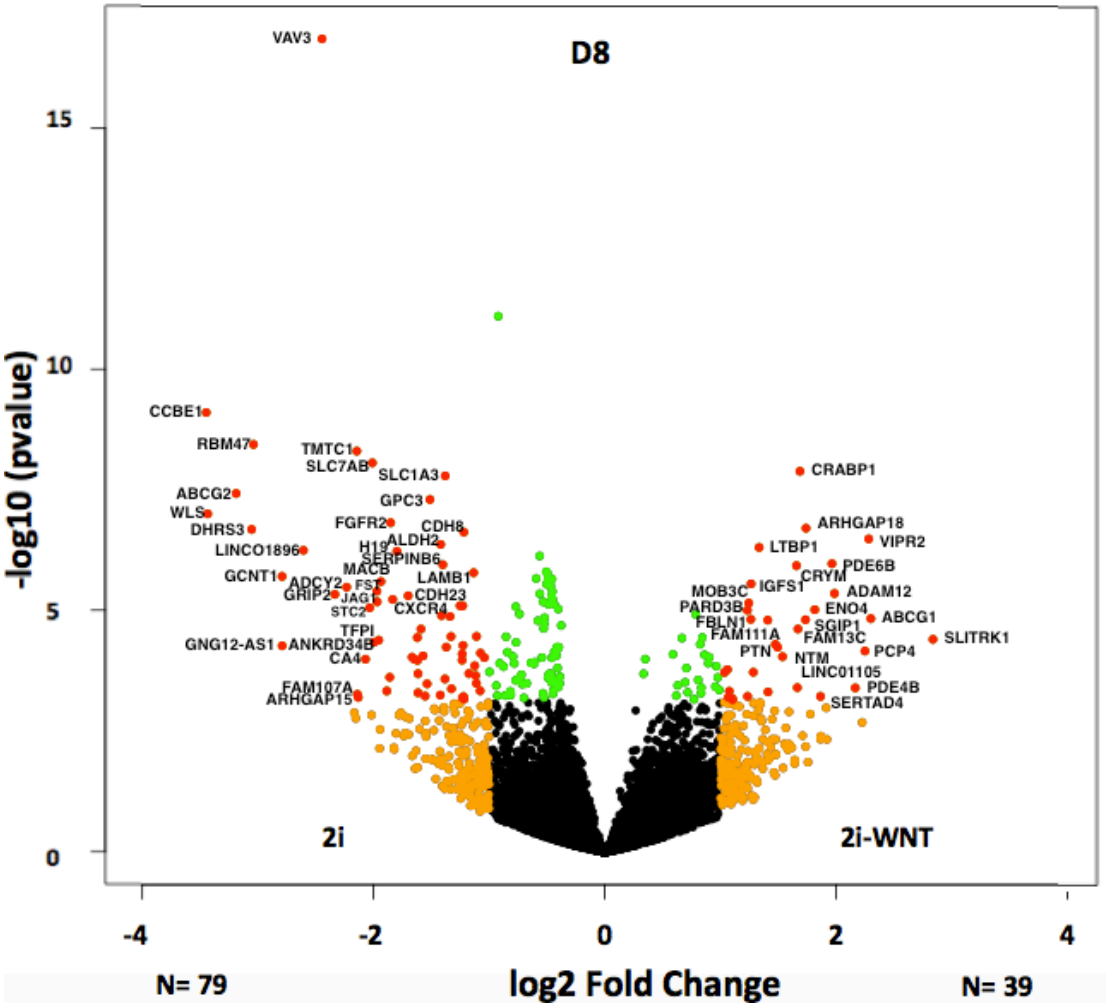



**Figure 3.26. Volcano plot of differentially expressed genes (no TFs included) induced with 2i and 2i-WNT neutralization media at time point d8.** The volcano plot is organized as described above.

Additionally, we aimed to identify the sets of TF included in the list of DE genes between 2i and 2i-WNT induced cells. For this, the ensemble gene ID of the genes DE was changed to Entrez ID with the "biomaRt" package in Rstudio. The list of DE genes was compared with a list of human TFs obtained from “http://www.tfcheckpoint.org” using the “VLOOKUP” formula in Microsoft Excel to identify TF. The data comprising only TF was separated from the data containing all genes sets and the TF volcano plots were generated with “calibrate” package in Rstudio. However, the curated list of TF included in tfcheckpoint database might include genes are not TFs, such us BMP4 and WNT8B, which are incorrectly included in the TF volcano plots generated in this study. Among the genes differentially expressed at d8, 29 TFs are included. 24 TFs are differentially up-regulated in cells induced with 2i media and 5 TFs in cells induced with 2i-WNT media (Figure 3.27). Statistical analyses details are included in the Appendix 3.14 (CD-ROM) for 2i up-regulated and Appendix 3.15 (CD-ROM) for 2i-WNT up-regulated.

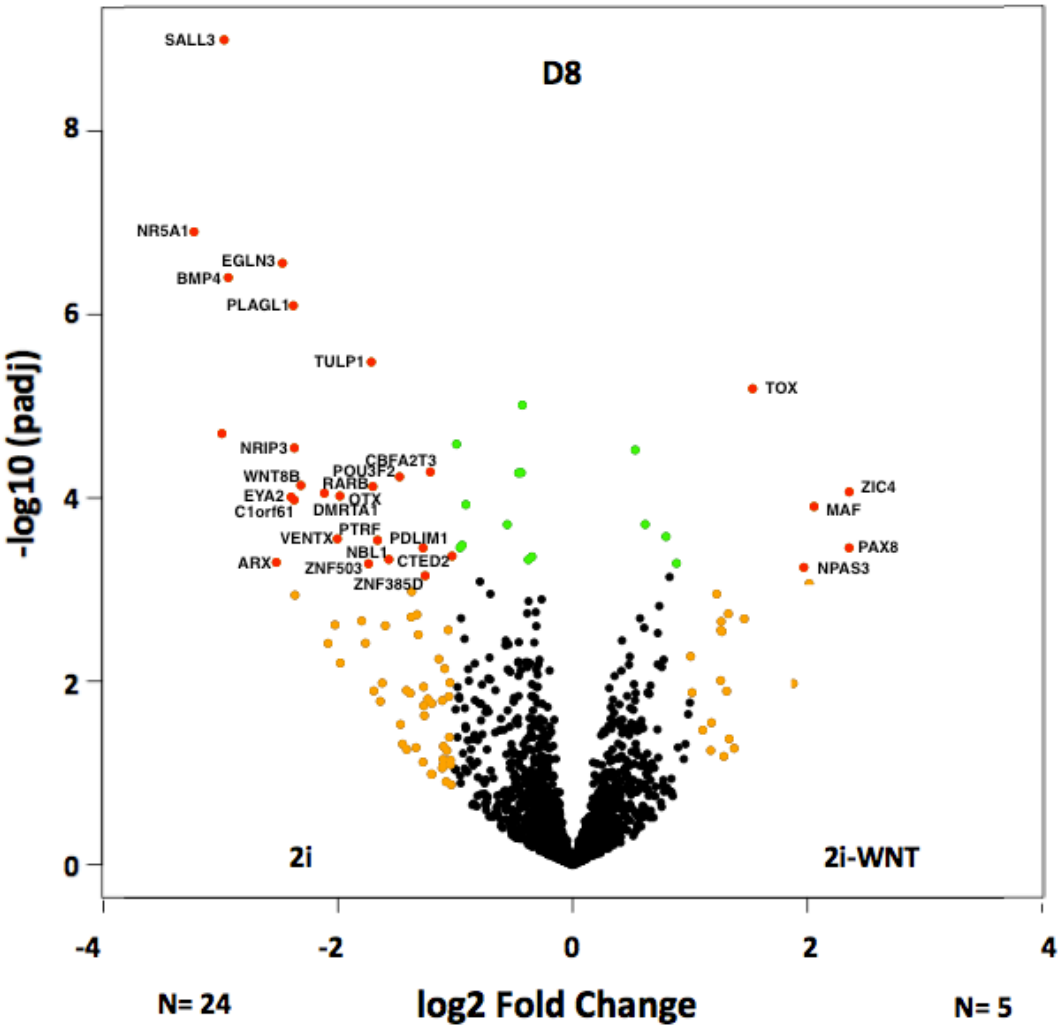



**Figure 3.27. Differential expression of transcription factors induced with 2i and 2i-WNT neural induction media at time point d8.** The volcano plot is organized as described above.

## 3.11. Bulk expression analyses of the genes differentially expressed between cells induced with 2i and 2i-WNT media.

The expression of all differentially up-regulated genes for the cells induced with 2i media were compared with the same genes expressed in cells induced with 2i-WNT media to identify any divergence in expression at earlier time points (Figure 3.28). Similar analyses

where performed for genes differentially up-regulated in cells induced with 2i-WNT compared with 2i media. The results indicate that differences in gene expression emerge after d2.

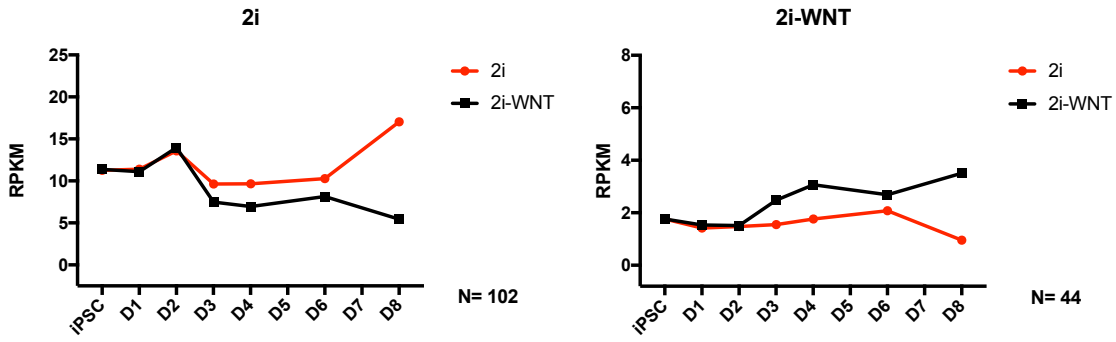



**Figure 3.28. Bulk expression patterns of the genes differentially expressed in cells induced with 2i and 2i-WNT neural induction media.** The gene expression profiles of cells induced with 2i media are shown in red and those of genes of cells induced with 2i-WNT media are depicted in black.

## 3.12. Gene otology of the genes differentially express between cells induced with 2i and 2i-WNT media.

The differentially up-regulated genes with > 1 log2 fold changes in expression between cell differentiated with 2i and 2i-WNT induction media were interrogated for enrichment of specific biological process. Gene ontology (GO) analyses were carried out separately for the up-regulated genes for each condition including 103 genes for 2i and 44 genes 2i-WNT conditions. Biological process annotations are depicted in Figure 3.29. GO analyses of genes up-regulated for 2i treatment show significant enrichment (fold enrichment > 2, p = < 0.05) of the biological process: “anatomical morphogenesis”, “tissue development”, “organ development” and “system development”. The p-values, number of genes for the

corresponding term, fold enrichment and false discovery rate are described in Appendix 3.16 GO analyses of genes up-regulated in cells induced with 2i-WNT showed no significant enrichment for any term.

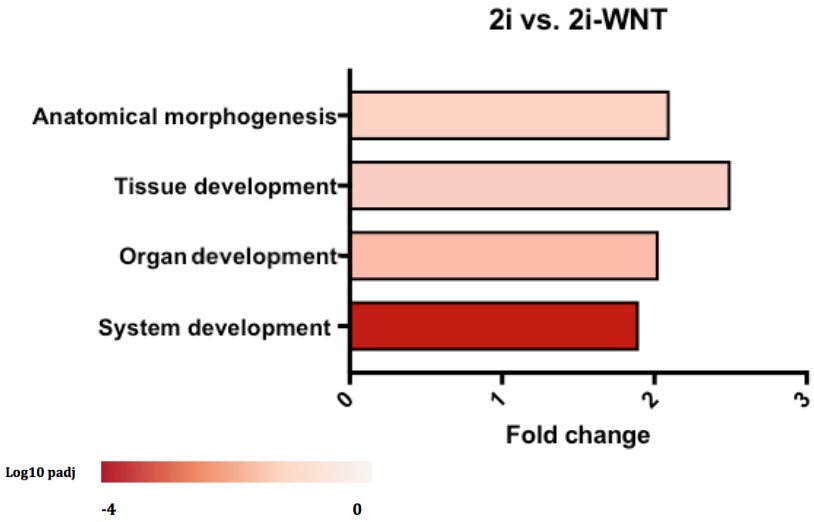



**Figure 3.29. Gene ontology of genes differentially up-regulated in cells induced with 2i media compared with cells induced 2i-WNT media.** The key colour scale indicates the log10-adjusted p-value, the lower p-value is represented in dark red.

## 3.13. Expression profiles of placode and olfactory placode markers

During the development of the embryo, the olfactory placode originates from the NBZ formed around the NP during gastrulation as a result of interactions of neural ectoderm with non-neural ectoderm. GnRH cells develop from the olfactory placodes and latter migrate to the hypothalamus. The canonical WNT signalling has been associated with the regionalization of the NBZ. For instance, absence of WNT in the NBZ leads to the formation of the PPE that gives rise to the olfactory placode, at expenses of the neural crest. We analysed the expression of NBZ and placode markers to determine if inhibition of WNT

signalling during NESC differentiation induces the differentiation NBZ and placodes like cells, which give rise to GnRH neurons at day 50 after neural induction. The expression of NBZ and placodal markers was analysed in cells induced with 2i and 2i-WNT neural induction media (Figure 3.30). The results indicate that the gene expression profiles of these markers are very similar between cells induced with both neural induction media.

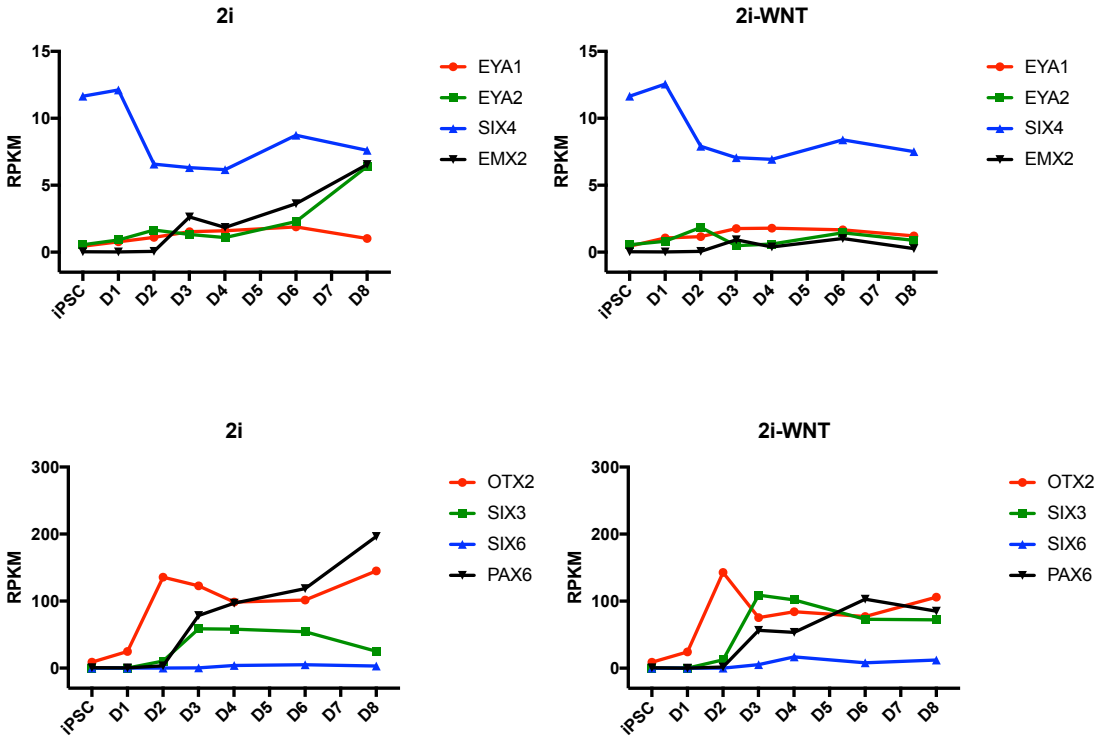



**Figure 3.30. Temporal profiles of placodal marker expression.** The expression of each gene is represented with a different colour. Profiles on the left represent gene expression patterns of cells differentiated with 2i media, whereas profiles on the right show the patterns in cells induced with 2i-WNT media.

## 3.4. Discussion

### 3.4.1. Gene expression patterns underlying neuro-ectoderm stem cells differentiation

The implementation of iPSC and RNAseq technologies enable study of *in vitro* differentiation of a number of human cell types of the CNS. For instance, several studies have analyzed the molecular events that drive the development of NE cells to neural progenitors and to specific neuronal subtypes. However, the events underlying human neural induction, which constitutes the initial step of CNS development, is largely unknown. This study provides a unique insight into the initial molecular events that underlie the transition of PSCs to NESCs and contributes to the knowledge of gene expression patterns and signalling pathways accompanying neural induction.

For this study, we used 3 iPSC lines generated from different individuals and 1 ESC line to account for the technical and genetic background differences among cell lines. PCA analyses revealed that the genetic background of each cell line and means of neural induction do not account for a significant component of the transcriptome variation during neural induction (Figure 3.2). The largest principal component of variance is correlated with the state of differentiation across all cell lines, indicating high reproducibility of the induction system and robustness of the expression levels among biological replicates. For instance, changes in gene expression patterns are smaller between consecutive time points than between distal time points. During the initial 48h of NE induction, changes in the expression patterns tend to be small and are clustered together (Figure 3.3). This period is succeeded by profound gene expression changes that continue throughout the remainder of neural induction. Gene product abundances and gene regulatory interactions, which are

reflected in the mRNA transcripts levels, control the biological processes occurring in the cell at specific stages of differentiation. Possibly, these gene expressions changes are associated with inhibition of pluripotency after d2 of induction and activation of the NE pathways that underlie the progressive restrictions in cell fate that drive neuronal differentiation.

Several genes have been identified during neural development in animal models and in ESC studies, including PAX6, NESTIN, SOX11, CDH2, ZNF521, ZEB2 and NR2F2, which are considered the main markers and drivers of neural differentiation. We report that the expression of these markers is first evident at d2 after 2i neural induction and progressively increases until d8. Whereas, expression of NANOG, OCT4 and MYC, which are associated with maintenance of pluripotency and self-renewal in PSCs, is down-regulated after d2 of induction (Figure 3.4). In animal models BMP and TGFβ inhibition by antagonists secreted from the node drive ectodermal cells to acquire NE fate. Following the developmental cues initiated by inhibition of BMP and TGFβ signaling leads to down-regulation of pluripotency/self-renewal markers after 2 days and subsequent activation of NE markers, validating previous ESC/iPSC studies that have demonstrated the same outcome. These results suggest that neural induction in iPSCs occurs in a similar manner to that described in animal models, highlighting the utility of iPSCs system to gain insight into early human development. However, it is important to recognize that 2i induction directs iPSC to acquire NE fate without evidence of prior formation of dorsal ectoderm. Hence, 2i differentiation of NESCs does not entirely recapitulate neural induction *in vivo*.

Animal models have enable to identify genes and pathways that regulate neural and neuronal differentiation. Particularly, studies in mouse model are highly relevant since they provide an insight into the regulatory mechanism that drive the transition from pluripotency to neural fate in humans. However, only a reduce number of studies have focussed in identifying the gene expression patterns transverse during the early stages of neural differentiation in mammals. For instance, the majority of information about the genes activated during NE and their interactions comes from studies has carried on with the Xenopus embryo. These, analyses have considerably increase the knowledge of the mechanism that drive neural differentiation in vertebrates but are not an ideal system for understanding the neural development in mammals. A number of genome wide analyses acquisition *in vivo* in mouse and *in vitro* in mouse and human, have analysed the molecular mechanism that regulate the differentiation of neural progenitors. However, there is a lack of information about the regulatory mechanism that orchestrate NE induction in mammals.

Klein *et al* demonstrated that expression of FOXD4, GMNN and ZIC2 NE precursor markers is necessary to directly activate the transcription of early-actin genes required for neural induction in Xenopus. FOXD4 expression initiates at blastula stage and is maintained during early NE development to inhibit BMP signalling and enhance GMNN and ZIC2 expression. FOXD4 is subsequently repressed during early neural plate formation to enable transition from NE precursors to mature NE (Lee *et al*., 2014; Klein & Moody, 2015). A recent study with ESC suggested that Foxd4 might play an important role during the transition from pluripotent cells to neural cells in mouse. Foxd4 is transiently expressed after neural induction in mouse ESC, depicting a peak in NESC (Gaur *et al*., 2016, Sherman *et al*., 2017). Foxd4 knowdown in mouse ESC lead to an increase in the expressing of pluripotency markers, whereas its overexpression down-regulated the expression of these markers, suggesting that Foxd4 might enhance neural fate acquisition by inhibiting

pluripotency. Additionally, ectopic expression of Foxd4 in the ventral epidermis of the mouse embryo leads to the expression of Gmnn, Zic2 and a decrease in the expression of Sox11. The study suggest that Foxd4 regulate the acquisition of neural competence in ectodermal cells, similarly than is counterpart in Xenopus (Sherman *et al*., 2017). Similarly, Gmm plays an essential role in early mouse embryogenesis since Gmm-nule mouse embryo are early lethal (Lim *et al*., 2011). ChIPseq analyses enabled to demonstrate that Gmnn promotes histone acetylation in neural genes promoters such as Sall3 and Hoxa2. Additionally, Gmnn and Zic1 have associated chromatin locations and might cooperatively activate the expression genes that regulate neuronal development in mouse, including Ascl1, Pax7 and Irx3 (Sankar *et al*., 2016). Consistent with these reports, we show that GMNN and ZIC2 have a high peak of expression at d2 and are subsequently down-regulated but remain expressed until d8 (Figure 3.4). However, the RPKM values of the factor FOXD4 (< 1 RPKM) indicated that the factor is either expressed at low levels or restricted to a subpopulation of cells. Perhaps, the expression of FOXD4 in human iPSC is not necessary for activation of ZIC2 and GMNN. Additionally, expression of FOXD4 might be necessary *in vivo* to promote the transition from pluripotent cells to a neural fate but might not be required for iPSC *in vitro* since inhibition of BMP and TGFβ signalling is sufficient to promote neural-ectoderma induction. However, more studies are necessary to analyse the role of FOXD4 during neural induction in humans.

Subsequently, down-regulation of Foxd4, Gmnn and Zic2 and overlapped activation of SOXB1 and SOX11 factors leads to the acquisition of NE phenotype in Xenopus (Lee *et al*., 2014). Similarly, in chick and mouse Soxb1 are expressed in NE precursors and in the cells of the neural plate. Overexpression of these genes prevents neural precursors to differentiate into neurons, suggesting that these factors are necessary to maintain pluripotency in NE cells. In mouse Sox2 and Sox3 are widely expressed in the neural plate,

whereas Sox11 activated latter during neural progenitors stage and differentiating neurons (Bergsland *et al*., 2011; Lee *et al*., 2014). We demonstrated that SOX2, and SOX11 are highly up-regulated after d3 and SOX3 is up-regulated at d6 and d8, in a similar manner to the chronological activation order described in Xenopus. As NE cells differentiate into neural progenitors, expression of the precursor markers IRX1, IRX2, IRX3, ZIC1 and ZIC3 is up-regulated just prior to expression of bHLH factors (Moody *et al*, 2013; Lee *et al*., 2014). Our results reveal that IRX1, IRX3 and ZIC1 expression is low during NE differentiation (<1 RPKM), whereas, the markers IRX2 and ZIC3 factors are up-regulated during d1, d2 and d3, overlapping expression of GMNN and ZIC2. Possibly, the role of IRX, ZIC1 and ZIC3 factors differ between humans and Xenopus during iPSC neural development due to amphibian- mammal's differences. Furthermore, the expression of these genes could be related with particular development landmarks in humans, which might differ from mouse and Xenopus. Additionally, differentiation of NESC in vitro, does not entirely recapitulate the multistep process of neural differentiation *in vivo*. For instance, the differentiating neural cells do not acquire endodermal phenotype previously to differentiating in neuro-ectodermal cells. Thus, the divergence in the expression of these genes could be related with differences in the neural differentiation program *in vivo* and *in vitro*. Overall, our results indicate for the first time that the temporal expression patterns of GMNN, ZIC2, SOX2, SOX3 and SOX11 factors previously described Xenopus model closely resemble their expression during human iPSC neural induction, whereas FOXD4, IRX, ZIC1 and ZIC3 factors are expressed in a different manner.

Subsequently, activation of bHLH family factors is required for transition of NE cells to neural progenitors. ASCL1 and NEUROG2 promote expression of target genes involved in neuronal differentiation and cell cycle arrest, whereas OLIG1 and OLIG2 promote oligodendrocyte differentiation and HES1 and HES5 maintain RG stage by inhibiting

neural differentiation (Kageyama *et al*., 2005; Dhanesh *et al*., 2016). The results indicated that ASCL1, NEUROG2 and HES5 are up-regulated and OLIG1 and OLIG2 are down-regulated during the 8 days of neural induction carried out in this study. However, RNAseq data show a reduced expression value for these genes (<1 RPKM), indicating that the genes are either expressed in a reduced number of cells of the population or that the population express low levels of the genes, which might not be biologically significant. Reduced expression of pro-neural markers after 8 days of 2i induction suggest that the majority of derived iPSC maintain a NESC phenotype and have not acquired neural progenitor's identity during this period. Interestingly, HES1 is highly up-regulated after d3 suggesting a potential function during NESC development. Possibly, high levels of HES1 are required to inhibit the premature expression of proneural markers as described during early RG stage.

Additionally, we identify a large number of differentially up-regulated genes during specific stages of NE induction, including many TFs that could be used as novel NESC markers. Among these genes, a number of TFs such as HESX1, IRX5, HES3, POU3F4, KDM4B, IHX4, FOS, SORBS2, RGS9, CITED1, HES4, SOX6, OTX1, FOXA1, DMRTA2, LEF1, RAX, SOX21, GLIS3, MEIS1, VAX1, FEZF2, TOX3, ATOH8, SMAD6 and GLI3 play important roles during later stages in development of the CNS (Figures 3.9, 3.10 and 3.11). Other differentially up-regulated factors have not been related previously to the development of the CNS including FANK1, MECOM, KAT2B, TRIM24 and ELF3. Activation of these TFs during NESC differentiation suggests potential roles for the maintenance of the neural transcriptional program or inhibition of genes and pathways that lead to differentiation of other lineages. Identification of these factors and their temporal expression pattern facilitates future studies to determine their precise role during early NE differentiation. The GO annotations of the entire set of up-regulated genes were

particularly enriched in terms related to development of the CNS including “diencephalon”, “telencephalon”, “forebrain”, “neuronal development” and “tube morphogenesis” (Figure 3.14). Likewise, we identified a set of up-regulated genes with a similar expression pattern (profile 65) using STEM software (Figure 3.16). This approach enabled clustering, comparison, and visualization of sets of genes correlated with a gene expression profile. The profile 65 was selected since it comprises genes that are highly up-regulated from day 0 to day 8 after neural induction. Hence, activation of these genes might be required for the transition of iPSC to a neural fate. Profile 65 was interrogated for enrichment of biological functions trough GO enrichment terms. Similar to the differentially up-regulated genes, GO analyses identified terms enriched for “brain”, “forebrain” and “telencephalon development”, “synapse organization” and “neural precursor cell proliferation” (Figure 3.18). Individual genes and gene modules are often present in several GO terms reflecting their multiple roles in different biological processes. For instance, GO subcategories such as “neuron projection guidance” and “synapse formation” comprise modules of genes that coordinate structural changes in the actin filaments and microtubules, a biological process that also occurs during NE differentiation. Hence, genes present in GO terms associated with later developmental stages might also be involved in novel biological functions during NE induction.

Additionally, we identified sets of down-regulated genes that form part of the ESC/iPSC transcriptional network, including the TFs ETV4, SKIL, ONECUT1, OVOL2 and HEY2 (Figures 3.12 and 3.13). Other down-regulated TFs such us GLI1, VAV1, BARX1 and BNC2 have not been previously implicated in pluripotency. GO analyses indicated that the down-regulated factors are enriched for the terms “vascular development”, “ion transport”, “mesenchyme development”, “cell motility”, “MAPK cascade” and “epithelium development” (Figure 3.15). Likewise, STEM analyses generated an expression profile

(profile 31) comprising genes that are highly expressed in iPSC and are down-regulated during different stages of neural induction. Thus, the profile 31 was selected for further analyses since it might contain genes necessary for pluripotency maintenance but not for neural differentiation. The profile 31 was interrogated for cell biological functions enrichment by GO analyses. The analyses were enriched for the terms "transport", "defence response" "angiogenesis" "regulation of cell proliferation" "response to hypoxia" and "cell-cell adhesion". Genes and gene modules involved in these biological functions might play a role in the regulation of the pluripotency program. For instance, hypoxia conditions promote the activation of hypoxia inducible factors HIFs, including HIF1 and HIF2, which are necessary to initiate the metabolic switch from oxidative to glycolytic metabolism, which promotes the pluripotency program. For instance, HIF2 has been shown to regulate expression of pluripotency/self-renewal TFs, such as OCT4, enhancing pluripotency. Hence, hypoxic conditions improve the reprograming efficiency and iPSCs maintenance (Yoshida *et al*., 2009; Mathieu *et al*., 2014; Saito *et a*l., 2015). Therefore, iPSC lines are maintained in hypoxic conditions until confluence is achieved and subsequently, the oxygen levels are incremented during NE induction to promote differentiation. Possibly, increase in oxygen levels is accompanied by down-regulation of "hypoxic response" genes included in the GO term. Additionally, cell-cell adhesion molecules such us E-cadherin contribute to the maintenance of pluripotency and cell survival in PSCs (Li *et al*., 2014; Pieters *et al*., 2014). Subsequent down-regulation of E-cadherin and up-regulation of N-cadherin promotes neural differentiation and rosette formation. Hence, the GO term "cell-cell adhesion" comprising down-regulated genes is consistent with these reports. A number of genes included in the GO terms have known functions in pluripotent cells, whereas others are related to different developmental stages and cells functions. Additional studies will enable identification of the function of these genes during pluripotency. Interestingly,

the terms ‘angiogenesis’ and ‘response to hypoxia’ are only significantly enriched for the 2i-WNT induced cells. However, genes enriched in these terms are also down-regulated during NE differentiation with 2i, but the number of genes highly correlated (R=0.9) to the profile was reduced and consequently these terms are not significant for cells undergoing 2i neural induction. These differences are likely to be driven by WNT inhibition, however the differential expression of these genes between treatments at each time point is not significant and the biological meaning of them remains unknown.

### 3.4.2. Gene expression patterns

Analyses of gene expression pathways relevant during differentiation demonstrated low expression levels of mesoderm, endoderm, non-neural ectoderm and NBZ markers, suggesting that the majority of the cells are acquiring NE fate (Figure 3.20). Additionally, we report that the neural rosette and radial glial markers are first up-regulated after d2 of induction (Figure 3.21). However, the expression of cell polarity markers does not increase during the time course of neural induction suggesting that NESCs are not acquiring rosette morphology or RG phenotype, as validated by ZO1 immunofluorescence analyses at d8 (Appendix 3.16). The precociously expressed rosette and RG markers might play a novel role during NE differentiation that has not been studied to date, or perhaps, higher levels of the markers or the expression of other gene products are necessary to trigger rosette formation and RG conversion. Animal and ESC studies have demonstrated that absence of patterning molecules during differentiation bias NE cells to acquire forebrain fate (Huang *et al*., 2016; Zirra *et al*., 2016). We report that the majority of analysed forebrain markers are highly express and continuously up-regulated after time point d3 of 2i differentiation, validating previous findings.

### 3.4.3. Neuroectodermal stem cells expression patterns described by Huang *et al* (2016)

In a recent study, Huang *et al* (2016) demonstrated that the EMT marker ZEB1 was up-regulated after 6 to 8 days of ESC NE differentiation, while ZEB1 epithelial targets were down-regulated, suggesting a novel potential role for EMT during neural differentiation. ZEB1, TWIST1, TWIST2, SNAIL1 and SNAIL2 TFs promote EMT by directly repressing expression of epithelial differentiation regulators and inducing expression of mesenchymal markers (Aigner *et al*., 2007; Kim *et al*., 2014; Lamouille *et al*., 2014). Our analyses show that ZEB1 expression is up-regulated from d3 of neural induction, whereas ZEB1 epithelial targets and mesenchymal targets are subsequently down-regulated and up-regulated, respectively. During EMT, E-cadherin expression is inhibited leading to the disruption of cell-cell contacts (Taube *et al*., 2010; Kim *et al*., 2014; Lamouille *et al*., 2014). Activation of N-cadherin, which is required to maintain NE cells architecture and for proper formation of the neural rosettes during neural induction, is also necessary for EMT (Nandadasa *et al*., 2009). Hence, concomitant down-regulation of E-cadherin and up-regulation of N-cadherin are a hallmark of EMT. Our results show that E-cadherin is down-regulated at time point d3 whereas N-cadherin is up-regulated after this period. The results validate the data presented by Huang suggesting a potential role of EMT during neural induction (Figure 3.22).

Neural induction is accompanied by chromatin structural changes that modulate the accessibility of TFs to specific loci (Broccoli *et al*., 2015; Tang *et al*., 2015; Quiao *et al*., 2016) but the epigenetic mechanisms that regulate neural induction are largely unknown. Huang *et al* (2016) reported significant changes in the expression of DNA and histone modifiers. Our results resemble the expression patterns of the modifiers, validating the data

(Figure 3.23). Mapping chromatin accessible sequences and regions enriched for specific histone modification would enable identification of the role of these modifiers. ROR2 and WNT5B were also up-regulated at d6 and d8 after differentiation (Huang *et al*., 2016). These genes are part of the WNT/$Ca^{2+}$ signaling up-stream of CaMKII kinase and since inhibition of CaMKII caused down-regulation of PAX6 and up-regulation of the epithelial marker K18, then this suggests that WNT/$Ca^{2+}$ signaling might play a role during neural induction by inhibiting epidermal fate. Our results validate these data showing that ROR2 and WNT5B are up-regulated after d2 of neural induction.

### 3.4.4. PAX6 and SOX1 expression

Previous studies have suggested that PAX6 expression precedes SOX1 during human neural induction (Pankratz *et al*., 2007; Suter *et al*., 2009; Zhang *et al*., 2010). However, Chambers *et al* reported that SOX1 was activated and up-regulated after 24h of ESCs neural induction before PAX6 activation. Neely *et al* (2012) suggested that SOX1 expression variability is related to the concentration of BMP inhibitors used in the induction protocol, demonstrating that higher concentrations of DMH1 or Noggin lead to higher expression of SOX1. We show that SOX1 and PAX6 are both significantly up-regulated after d2 of 2i induction (Figure 3.23). Subsequently, PAX6 expression increases rapidly after this stage whereas SOX1 expression is down-regulated and remains low throughout neural induction. We are not able to relate these findings to the directly to the concentration of BMP inhibitor since the other studies used different small molecule SMAD signaling inhibitors. However, we show a novel expression pattern for SOX1, which depicts a peak of expression after d2 of induction. These results highlight the importance of determining the factors that underlay SOX1 expression variability and their potential effect during neural differentiation in humans.

### 3.4.5. Differential expression between genes induced with 2i and 2i-WNT

Embryonic generation of GnRH neurons depends on the sequential formation of the non-neural ectoderm, NBZ, PPE and olfactory placodes, which subsequently gives rise to GnRH neurons that migrate to the hypothalamus. *In vitro* and *in vivo* studies demonstrated that modulation of WNT, BMP, TGFβ and FGF signals leads to the formation of PPE. However, the PPE cells subsequently acquire otic placode fate at expenses of olfactory placode. Hence, to date, PPE have not been effectively differentiated into olfactory placodes that can be induced to GnRH neurons. In the Price laboratory, generation of GnRH neurons is achieved by inducing iPSC/ESC lines with 2i induction media and WNT inhibitors. Since the only difference in the generation of cortical and GnRH neurons resides in the initial 8 days of neural induction, we hypothesize that during this period cells induced with WNT inhibitor are specified to acquire GnRH fate.

The results indicate that there are 44 genes including 5 TF differentially up-regulated at d4 and d8 in 2i-WNT compared with 2i induced cells (Figures 3.25, 3.26 and 3.27). Among the DE genes the factors PAX8, MAF, PDE6B, NPASS3, ZIC4 and PTN have been connected with different process of CNS development but not specifically with differentiation of GnRH neurons. Other genes are related to functions such as transport, proliferation and morphogenesis but GO analyses show that these genes are not enriched for any biological function term. Similarly, there are 103 genes, including 24 TFs differentially up-regulated in cells induced with 2i media compared with 2i-WNT media. Among these, POU3F2, ARX, LAMB1, FGFR2, DMRTA2, RARB, SALL3, CCK, CDH23, PLK2, SAG1and OTX1 play a role in brain development and other genes participate in distinct biological functions. GO analyses of this set of genes shows enrichment for broad terms such as “system development”, which include genes related to

several biological processes (Figure 3.29). Possibly the expression of these genes is necessary for the differentiation of cortical cells but not for the generation of GnRH neurons. However, due to the pleitropy associated with many of these genes, it is difficult to determine which biological process might trigger the acquisition of different neuronal phenotypes. GnRH neuronal differentiation has only recently been achieved *in vitro*; hence the regulatory interactions underlying different stages of development are largely unknown. Inhibition of WNT signalling during neural induction is sufficient to trigger the progressive transcriptional changes that re-specify NESC to differentiate into GnRH neurons. Early activation of the genes exclusively differentially up-regulated in 2i-WNT induced cells possibly prime NESCs to acquire GnRH neuronal fate. However, the molecular mechanism and gene interactions by which these genes drive early GnRH neuronal fate acquisition remain unclear.

### 3.4.6. Limitations

ESC and iPSC have huge potential for elucidating fundamental aspects of early human development even though some specific molecular events might differ from those occurring in the embryo. For example, Shin *et al* (2007) highlight these differences during neuronal differentiation by demonstrating that neural progenitors derived from hESCs and progenitors isolated form a foetal ventricular zone are accompanied by differential expression of a number of genes. Hence, the gene expression patterns obtained in this study give us an insight of the molecular events that underlay human neural induction but might not completely recapitulate all the events that drive development in the embryo.

Importantly, it is critical to repeat the experimental procedures in different micro-environmental conditions to validate transcriptome data presented in this study. It has been

demonstrated that ESC/iPSC respond differently to their microenvironment generating transcriptional signatures associated with the laboratories from where the data were derived. These differences might be related to the reprograming process, the stem cell culture conditions, the progenitors or the handling of the cells (Newman & Cooper, 2010). Hence, repeating the experiment in different laboratories would enable detection of the degree of variance in the gene expression patterns among laboratories and accurately identify the expression signatures traversed during neural induction.

## 3.5. Conclusions

The experimental approaches described here enabled capture of the complete transcriptome at different time points and unwrap the global expression patterns generated during NE differentiation. Accordingly, it appears that the plutipotency program is down-regulated during the initial 2 days of induction, followed by activation of the neural induction program. We demonstrated that the initial steps underlying neural induction are accompanied by activation of thousands of genes including NE markers that have been described in animal and ESC/iPSC studies and several genes and TFs that have not been previously related with neural induction or CNS differentiation. Hence, we have identified sets of genes and TFs that may orchestrate the molecular mechanisms that drive different stages of neural induction in humans. Importantly, we provide sufficient transcriptome data to be of use in future studies to infer specific signaling pathways involved in neural induction. Additionally, the results indicate that the molecular mechanisms that drive NESC specification to GnRH neurons are activated by d8 during neural induction and are accompanied by specific activation of 44 genes and down-regulation of 103 genes compared with NESCs fated to become cortical neurons. However, we did not identify specific gene modules that provided insight into how this specification occurs.

# CHAPTER 4. Gene Regulatory Networks

## 4.1 Introduction

During neural induction, the gene products encoded by the genome coordinate different cell biological functions such as maintaining proliferative state, driving neural commitment and determining neuronal differentiation (Qin *et al*., 2015; Ardhanareeswaran *et al.,* 2017). The transition between cell states depends on changes of gene product abundances at specific developmental times, which are coordinated by various gene regulatory events. Amongst the cell mechanisms of regulation, the most common and best-studied gene regulators are TFs. TFs are proteins with a DNA binding domain that bind to regulatory sequences (enhancers or promoters) of target genes to stimulate or repress their transcription (Qin *et al*., 2015; Yu *et al*., 2017). TFs orchestrate neural induction by activating and stabilizing expression of sets of genes that initiate and maintain neural fate whilst down-regulating expression of genes involved in pluripotency or differentiation toward alternative fates (Sankar *et al*., 2016).

Recent implementation of high-throughput technologies facilitates determination of TF DNA binding sites by ChIP, ChIP-chip or ChIPseq analyses (Wang *et al*., 2015). Additionally, RNAseq analyses allows quantification of changes in expression of a large number of genes at specific time points, enabling determination of the identity of genes and TFs activated or inhibited during specific stages of differentiation (Hecker *et al*., 2009; MacNaeil & Walhout, 2009). The global transcriptome data obtained from these approaches enables prediction of TF gene targets based on a binding site preference and to infer potential regulatory interactions between them through various statistical and mathematical analyses (MacNaeil & Walhout, 2009; Kordmahalleh *et al*., 2017). The gene

regulatory interactions inferred are often captured in a gene regulatory network (GRN), which provides a mechanistic view of how these molecular regulatory events orchestrate the different biological processes required at specific stages during cell fate commitment (Li & Davidson, 2009; Vashishtha *et al*., 2015; Peter & Davidson, 2017). However, the regulatory interactions between genes are not always simultaneous and can occur with temporal changes in gene expression. Time delay interactions can be caused by differences in the transcription and translation completion kinetics related with the gene size and differences in gene product stability. Hence, several methods have being proposed to incorporate time delay in GRN modeling (Xu *et al*., 2007; Kordmahalleh *et al*., 2017).

The GRN comprises nodes that represent genes (TFs can be specifically selected) and edges that represent the regulatory interactions between nodes. Genes that have large number of connections with other genes are named hubs and are frequently TFs that control expression of numerous targets (Blais & Dynlacht, 2005; Yu *et al*., 2017). The sets of genes expressed in the same cells conditions and connected to a common TF can be considered as subunits of the GRN and are referred to as modules. For practical reasons modules are analyzed as autonomous units and are frequently associated with a specific cell regulatory function (Peter & Davidson, 2017; Charney *et al*., 2017; Yu *et al*., 2017). It has been proposed that the distinct biological processes carried out by the cell might require a level of compartmentalization, which is associated with a reduced number of direct connections between the modules in the GRN. Since not all the regulatory functions of the cell are required at all times, the activity of the modules and interconnectivity of the network (topology) changes over time during development (Blais & Dynlacht, 2005). The interactions between genes within the modules ultimately determine cell phenotype and behaviour. Hence, identification of the GRN and the modules that are active during specific developmental stages provides a systematic understanding of the different steps and

molecular mechanisms that underlie neural induction (Vashishtha *et al*., 2015; Yu *et al*., 2017).

As development proceeds, gene expression patterns continuously change in response to different regulatory signals. Changes in gene expression patterns are translated into changes in GRN topology. The gene expression patterns captured at a specific time point in the GRN constitute network states (Huang *et al*., 2009b; Gomez *et al*., 2012). During cell fate acquisition, cells traverse the network state space until they reach a state where the gene product interactions are stable (Figure 4.1.) (Huang *et al*., 2009b; Choi *et al*., 2012). The cell morphology, function and behaviour (phenotype) emerge as a property of these stable molecular interactions within the cell (Huang *et al*., 2009b; Vashishtha *et al*., 2015). The stable state is robust to changes in expression levels of individual genes of the GRN. However, perturbations such as mutations can generate genome-wide changes compromising the stable interactions between nodes and driving the cells to different state and therefore a different phenotype (Huang *et al*., 2009b; Ghaffarizadeh *et al*., 2017). Thus, the GRN inferred relations can be validated by testing the network predictions using specific intervention including small molecule inhibitors, gene knock-down and gene editing (Blais & Dynlacht, 2005; Olsen *et al*., 2014; Olsen *et al*., 2015).

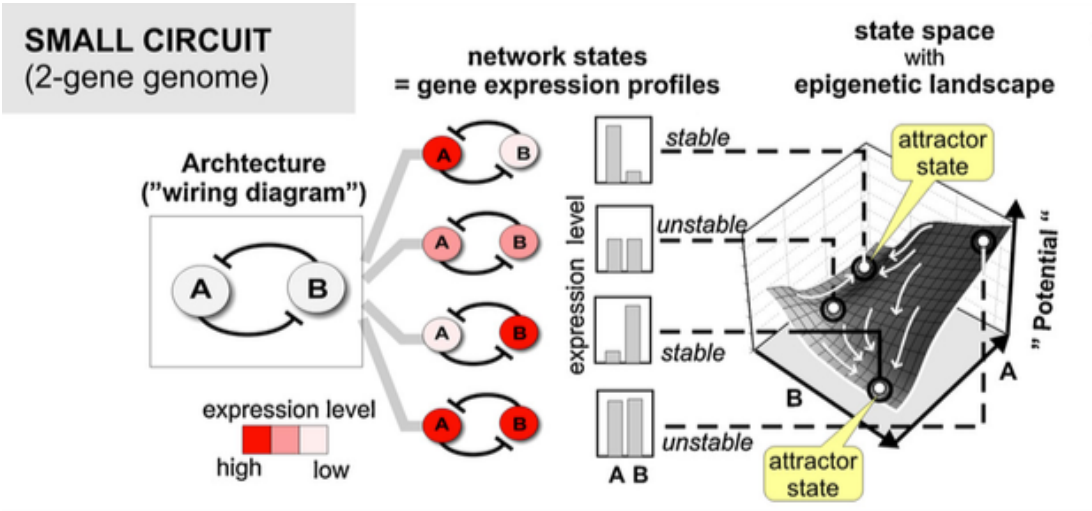



**Figure 4.1. Representation of the regulatory interaction between 2 genes.** The genes A and B interact by inhibiting each other expression, forming 4 possible gene expression profiles. When both genes are equally express the interactions are not stable due to mutual inhibition. When one gene express higher levels inhibits the other gene expression and forms a stable expression pattern. The cell phenotype emerges from the stable pattern (attractor state). The gene expression patterns are often represented as a point in a Waddington epigenetic landscape. Elevation in the landscape represents stability, thus unstable transient states during differentiation are shown as high points, whereas stable regulatory interactions between the genes are represented as basins (Huang *et al*., 2009b).

GRNs enable visualization of the gene regulatory interactions inferred from high-throughput analyses and provide systematic interpretation of the global expression patterns that emerge from these interactions. Furthermore, GRNs modeled from time course analyses allow identification of the consecutive gene expression patterns or states that the cell occupies during its trajectory between distinct cell states (Charney *et al*., 2017; Yu *et al*., 2017). Hence, the GRN can be interrogated to identify TFs that are potentially responsible for the expression of target genes, the active modules potentially related to a particular cell biological process and to predict GRN outputs due to specific interventions of the network topology (Bisguas & Acharyya, 2016). Such information is essential to understand the regulatory interactions that direct human development, the molecular events

underlying aberrant cell phenotypes and to identify potential therapeutic strategies for treating diseases (Choi *et al*., 2012; Lopes *et al*., 2012).

## 4.2. Methods

### 4.2.1. Gene regulatory network

The different experimental approaches that I performed during this study enabled to obtain a large amount of transcriptomic data during different stages of NESC differentiation from iPSC trough RNAseq analyses. I used this data to generate differential expression analyses, co-relation analyses, PCAs, STEM profiling, genes expression heatmaps, gene expression profile and volcano plots. Additionally, the raw data generated from the iPSC1 line induced with 2i media was used by our collaborator Jun Min (UCSD) to model the gene regulatory network. I contribute with the GRN refinement by suggesting the exclusion of several genes with low expression levels (RPKMs) from the initial network, since these genes might not be biologically relevant. However, the GRN was entirely model by our collaborator. Briefly, Jun Ming used the DESeq2 function dds to rlog-transformed the data and performed differential expression analyses. A number of TFs differentially expressed at time points d1, d2 and d3 with respect to d0 were identified. A set of TF were selected due to high number or interactions with other genes determined from TRANSFAC database. If the TF was differentially expressed between more than one time points, the earliest time point of differential expression was selected for further analyses. The set of selected TFs were correlated (Pearson correlation R=0.95) to a subset of potential targets or interacting proteins, which were also differentially expressed at the same time point or with 1, 2 or 3 days time delay. In parallel, our collaborator derived an interactome from the RNAseq data and compared these interactions with the potential interactions of the highly correlated TFs

and targets. Common TF-target or protein-protein interactions from the time delay correlation and the interactome were plotted in a GRN for visualization with GraphViz software and Cytoscape software (Figure 4.2).

I interrogated the genes comprising the GRN for GO terms enrichment with DAVID as described in section 2.4 (Figure 4.5).

Additionally, I plotted the gene expression counts with ggplot2 packages in R (figure 4.6).

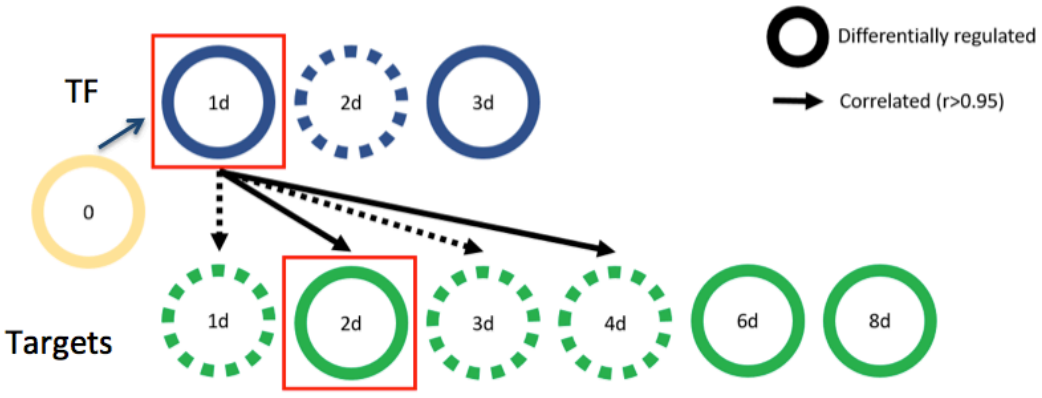



**Figure 4.2. Scheme of the gene regulatory network modeling.** TFs (blue) differentially expressed (complete circle) from d0 (yellow) at d1, d2 or d3 after NE induction were identified and the earliest time point of differential expression was selected (red)**.** The gene targets (green) are mapped to the TF only if their expression is correlated (arrows) at the same time (d1) or with 1, 2, or 3 days time delay (d2, d3, d4) and if the targets are differentially regulated (complete circle) at the correlation time point.

### 4.2.2. MYC inhibition assay

Approximately 100,000 iPSCs were plated on each well of Nunclon delta surface 96 well plates (Thermo Scientific, 167008) with similar conditions as described in the section 2.1. Cells were grown for 24 hours with E8 media until they reached confluence. Confluent iPSCs were treated with doses of 30μM or 10μM final concentrations of the MYC small

molecule inhibitor 10058-F4 (Sigma Aldrich, F3680) or 0.01% of dimethyl sulfoxide (DMSO) in E8 media for 24 hours. After incubation, cells were observed for quality control and survival.

In parallel, 100,000 iPSCs were plated in Nunclon 96 well plates and grown for 24 hours in E8 media until they reached confluence. Confluent cells were treated with 30μM or 10μM of 10058-F4 or 0.01% DMSO in 2i neural induction media, which was changed every 24 hours. After 24, 48 and 72 hours of MYC inhibition treatment, a set of cells was fixed with 4% paraformaldehyde and used for immunocytochemistry analyses following procedures detailed in the section 2.8.

Additionally, 1,500,000 iPSCs were plated in Nunc 6 well multidishes and incubated in E8 media for 24 hours until they reached confluence. Confluent cells were treated with 2i induction media and 30μM or 10μM of 10058-F4 or 0.01% DMSO. Cells were checked for quality control and media was changed every 24 hours. RNA was extracted from 2i-induced cells treated with each dose of 10058-F4 or DMSO after 24, 48 and 72 hours. RNA extraction was as described in section 2.2. cDNA was synthesized from purified RNA and was used to analyze relative expression of SMARTA, CDH1 and TFAP2C through Q-PCR analyses as described in section 2.7. The reference sample was provided by the DMSO treatment at d1.

## 4.3. Results

### 4.3.1. Gene regulatory network

The GRN genes are depicted as nodes and their regulatory interaction as edges. Gene interactions can be DNA-protein or protein-protein. Selected TFs and genes from the TRANSFAC database are represented in green and their targets in yellow. Lines represent

bidirectional protein-protein interaction and lines with arrows TF-gene target interactions. Blue lines indicate negative correlation, whereas red lines show positive correlation. Regulatory interactions without time delay are represented with solid lines. Longer dashed lines and shorter dashed lines show time delay of 1 and 2 days, respectively. Dotted lines show time delay of 3 days with respect of the selected TF (Figure 4.3).

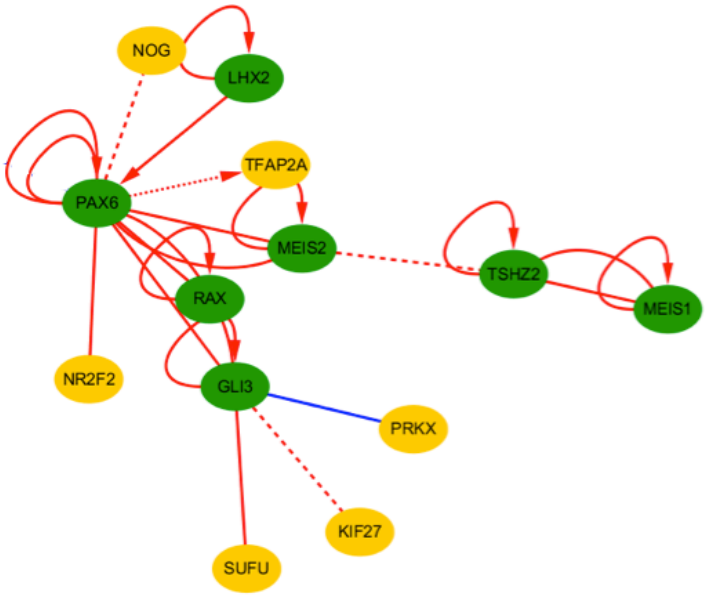



**Figure 4.3. PAX6 gene module** shows: selected TF and genes (green) and targets or interacting proteins (yellow) with positively (red) or negatively (blue) correlations and their protein-protein interactions (lines with arrows) or TF-target regulatory interactions (lines with arrows) with time delay correlation of 0 (solid lines), 1 (long dashes), 2 (short dashes) and 3 days (dotted lines) between them.

The GRN comprises 9 highly connected nodes (hubs) including MYC (40), JUN (32), SP3 (21), SMAD2 (19), FOS (16), CD6 (14), LEF1 (11), PAX6 (8) and HDAC6 (6) - connectedness is shown in parenthesis (Figure 4.4). A full-size image of the GRN has been included in the Appendix 4.1 (CD-ROM).

The genes comprised in the GRN were interrogated for GO enrichment for biological process terms. The results show that the genes were enriched for “localization”, “response to stimulus”, “biological adhesion” and “locomotion” (Figure 4.5). Sub-ontologies, p-values, number of genes for the corresponding term, fold enrichment and false discovery rate are described in the Appendix 4. 2.

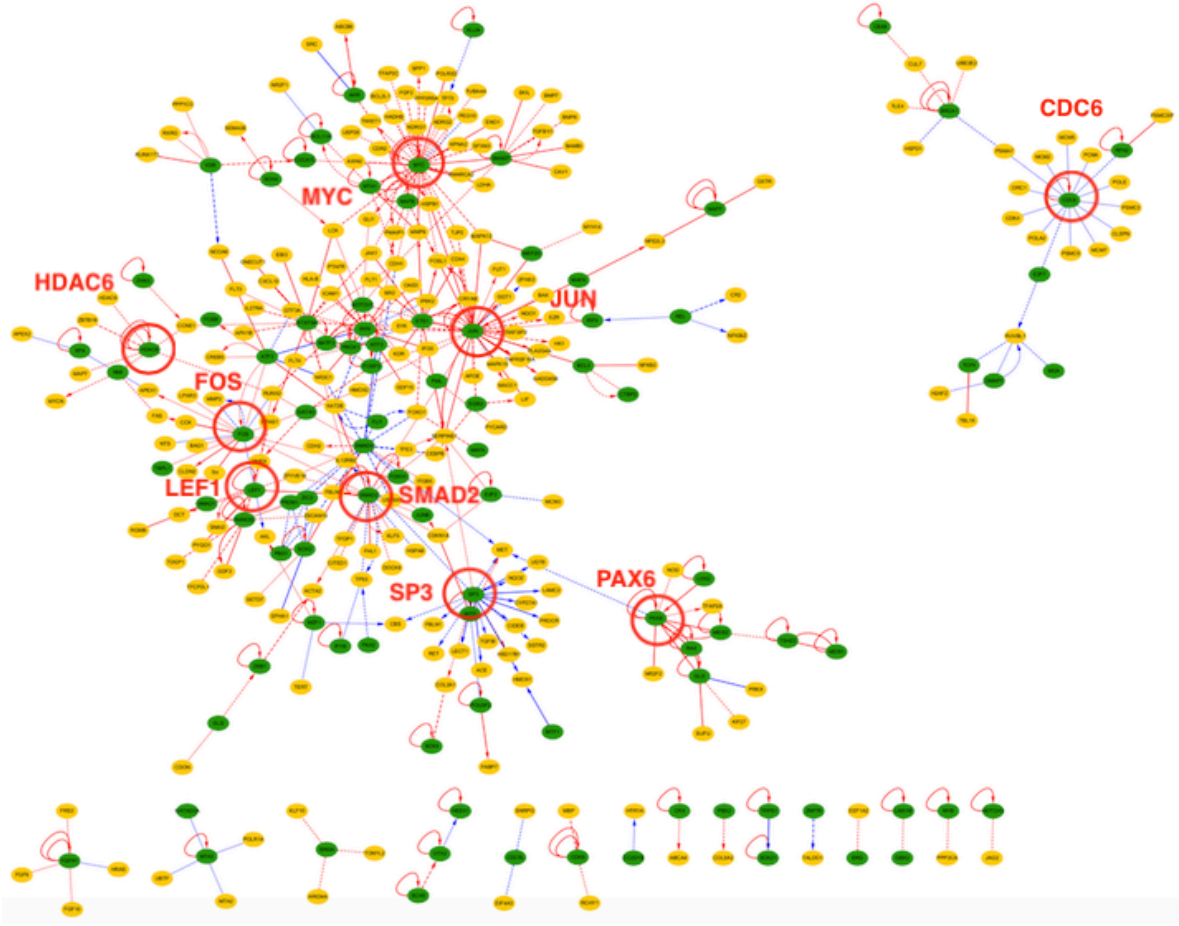



**Figure 4.4. Gene regulatory network from d0 to d6 during neural induction.** The GRN comprises 9 principal hubs (highlighted with a red circle) highly connected to TFs and genes.

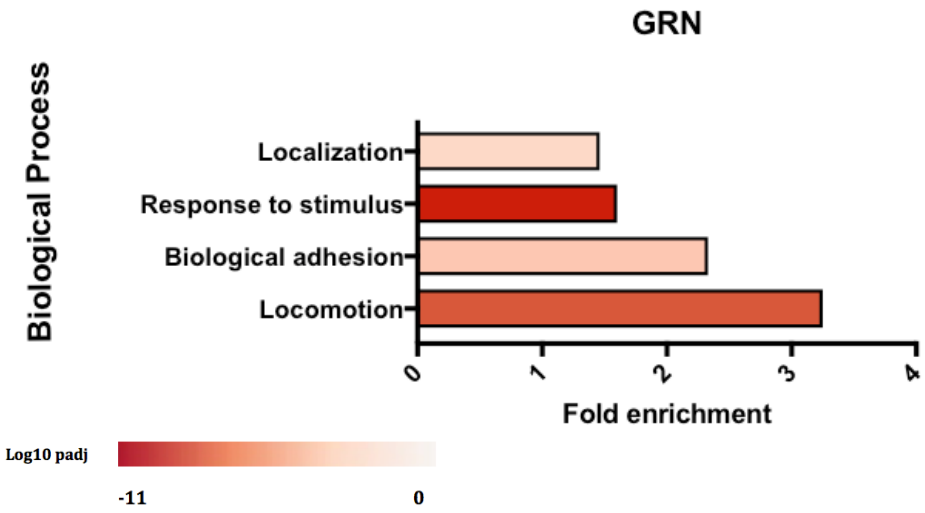



**Figure 4.5. Gene Ontology of the genes comprising the gene regulatory network.** GO terms enrichment of the GRN genes. The key colour scale indicates the log10-adjusted p-value, where the lower p-value is represented in dark red.

Counts of MYC, PAX6, SMAD2 and DCD6 hubs show the expression patterns of these genes among the 3-iPSC lines (Figure 4.6). DE analyses indicate that MYC expression is significantly down-regulated at d1 while PAX6 expression is significantly up-regulated at d2 and expression of SMAD and CDC6 is significantly up-regulated at d3 and d2, respectively.

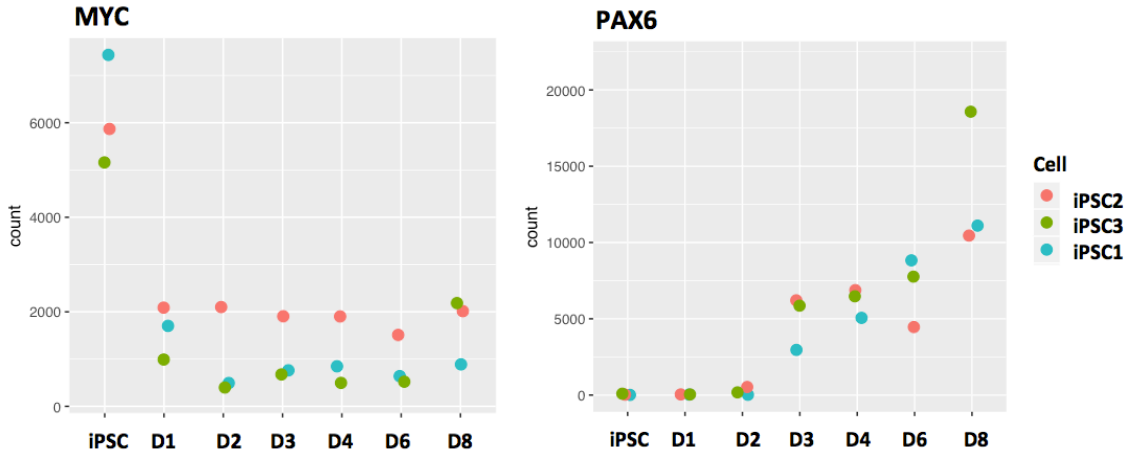



**Figure 4.6. MYC, PAX6, SMAD2 and CDC6 counts from time point d0 to d8.** Expression patterns of iPSC1 (blue), iPSC2 (red) and iPSC3 (green) during 8 days of neural induction.

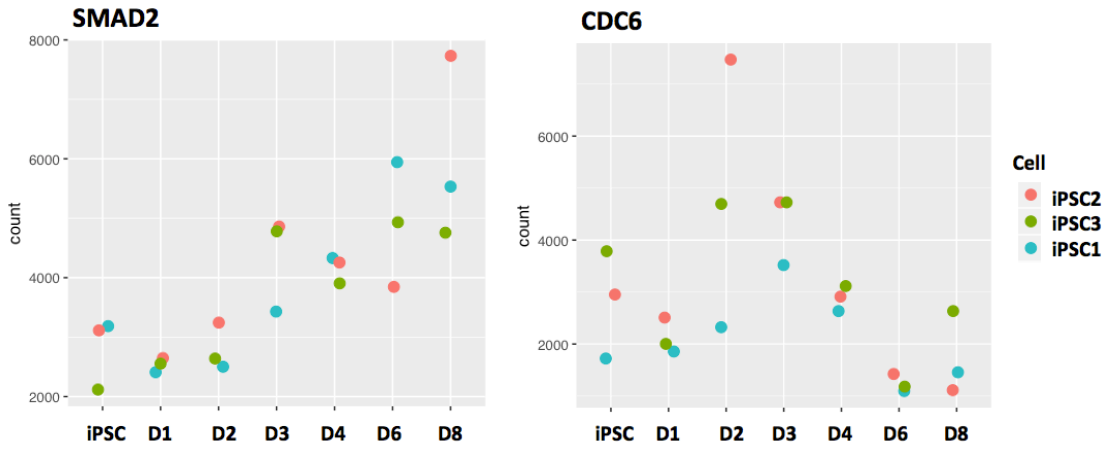



**Figure 4.6. Continued**

The main hubs and genes directly connected to them in the GRN were interrogated for enrichment of GO terms (Figure 4.7). The hubs SP3, HDAC6, FOS, JUN and LEF1 did not present significant enrichment for any biological function term. The MYC hub and associated targets were significantly enriched for the terms "programmed cell death", "cell proliferation" and "regulation of response to stress". PAX6 and associated gene targets in the GRN were enriched for the terms "regulation of RNA biosynthetic process", "regulation of transcription, DNA template", "central nervous system development", "sensory organ development", "neural tube development" and "dorso-ventral pattern formation". SMAD6 hub and GRN targets were enriched for "regulation of gene expression", "regulation of RNA metabolic process", "regulation of RNA biosynthetic process" and "regulation of transcription, DNA-templated". The hub CDC6 and its interacting genes were enriched for the terms "regulation of cell cycle process", "cell cycle phase transition", "G1/S transition" and "DNA replication". Sub-ontologies, p-values, number of genes for the corresponding term, fold enrichment and false discovery rate are

described in the Appendix 4.3 for MYC network, Appendix 4.4 for PAX6, Appendix 4.5 for SMAD6 and Appendix 4.6 for CDC6.

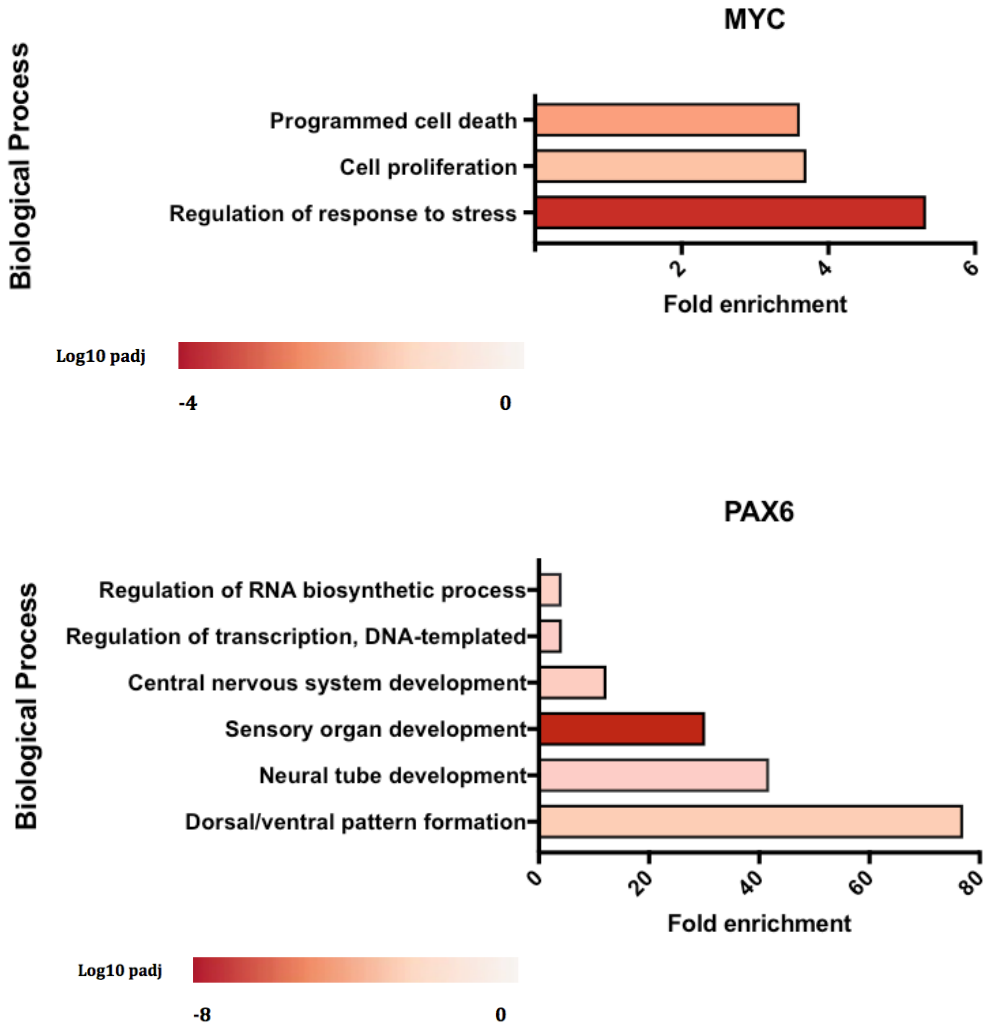



**Figure 4.7. Gene Ontology of the genes interacting with MYC, PAX6, SMAD2 and CDC6 in the gene regulatory network.** The key colour scale indicates the log10-adjusted p-value, where the lower p-value is represented in dark red.

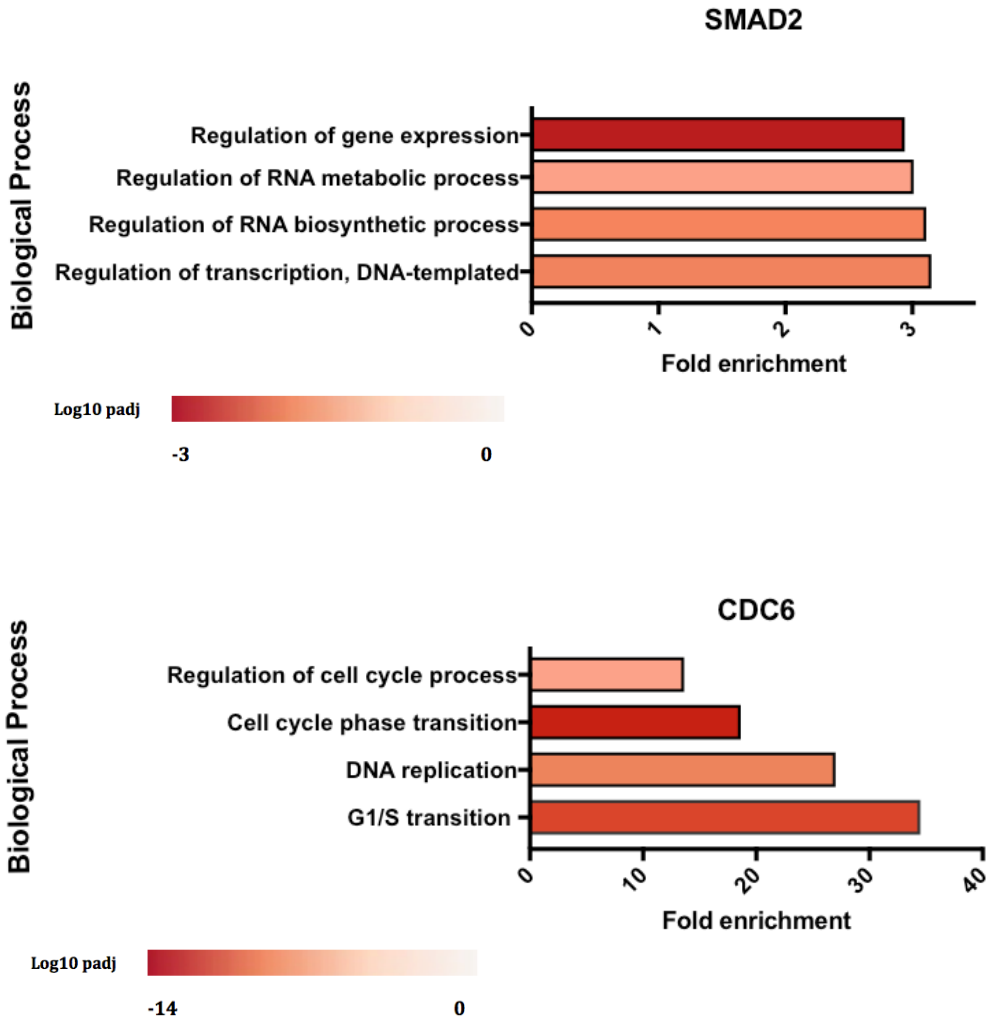



**Figure 4.7. Continued**

### 4.3.2. MYC inhibition assay

The GRN comprising the hub MYC and interacting genes is depicted on Figure 4.8. Importantly, several genes that are not connected directly with MYC in the GRN have been removed to facilitate visualization. Inferred gene interaction of the GRN can be validated by testing the predictions of the model to a specific interference. Intervention of the network can be achieved by different strategies such us CRISPRi, CRISPa and small molecule inhibitors, been the later the most affordable and relatively simpler technique. We selected the MYC hub to test the GNR predictions due to the availability of a small molecule

inhibitor (10058-F) that targets MYC-MAX interaction, preventing the regulation of MYC targets. MYC hub is highly connected with several nodes of the GRN. Hence, inhibition of MYC might affect the expression of the genes directly connected to the hub in the network.

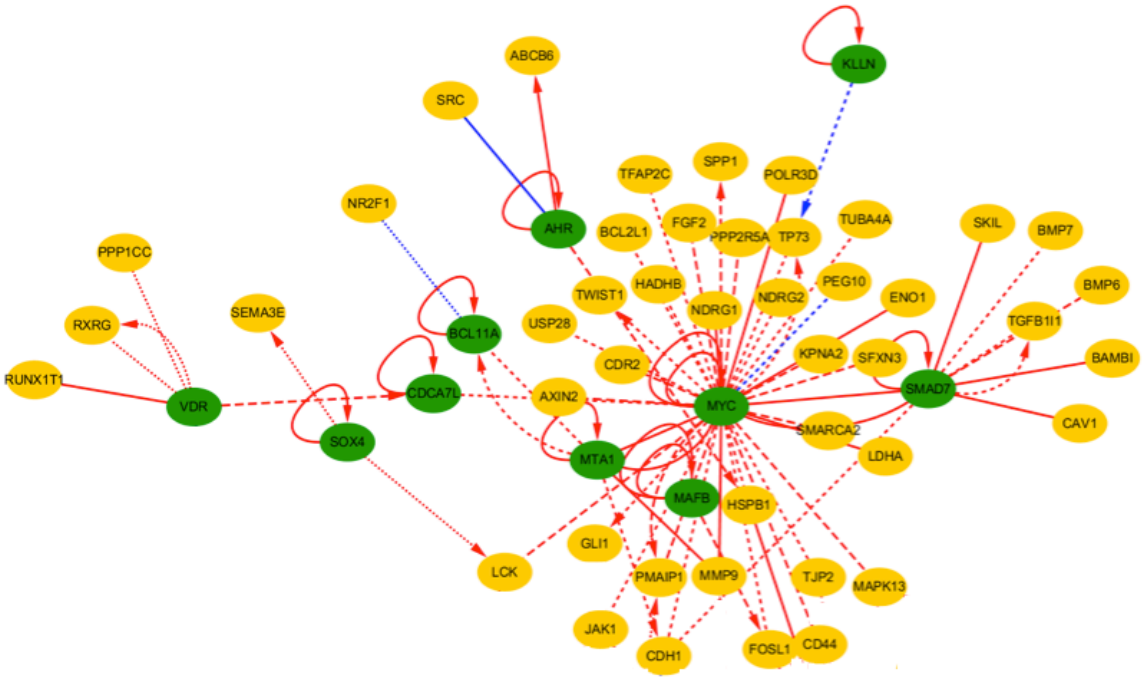



**Figure 4.8. Gene regulatory network of MYC target genes or interacting proteins.** The majority of MYC targets or interactive proteins are positively correlated (red edges) with the hub. 40 MYC interactions (both DNA-protein and protein-protein) can be seen.

Cells undergoing 2i neural induction were treated with 30μM or 10μM 10058-F MYC inhibitor or with DMSO and were fixed after d1, d2 and d3. The expression of pluripotency markers was analyzed by immunofluorescence for the pluripotency proteins, OCT4 and NANOG and for the NE markers, SOX11 and ZNF521. The results demonstrated that inhibition of MYC during d1, d2 and d3 of differentiation does not significantly change expression of OCT4, NANOG, SOX11 and ZNF521, as shown in Figures 4.9, 4.10, 4.11 and 4.12.

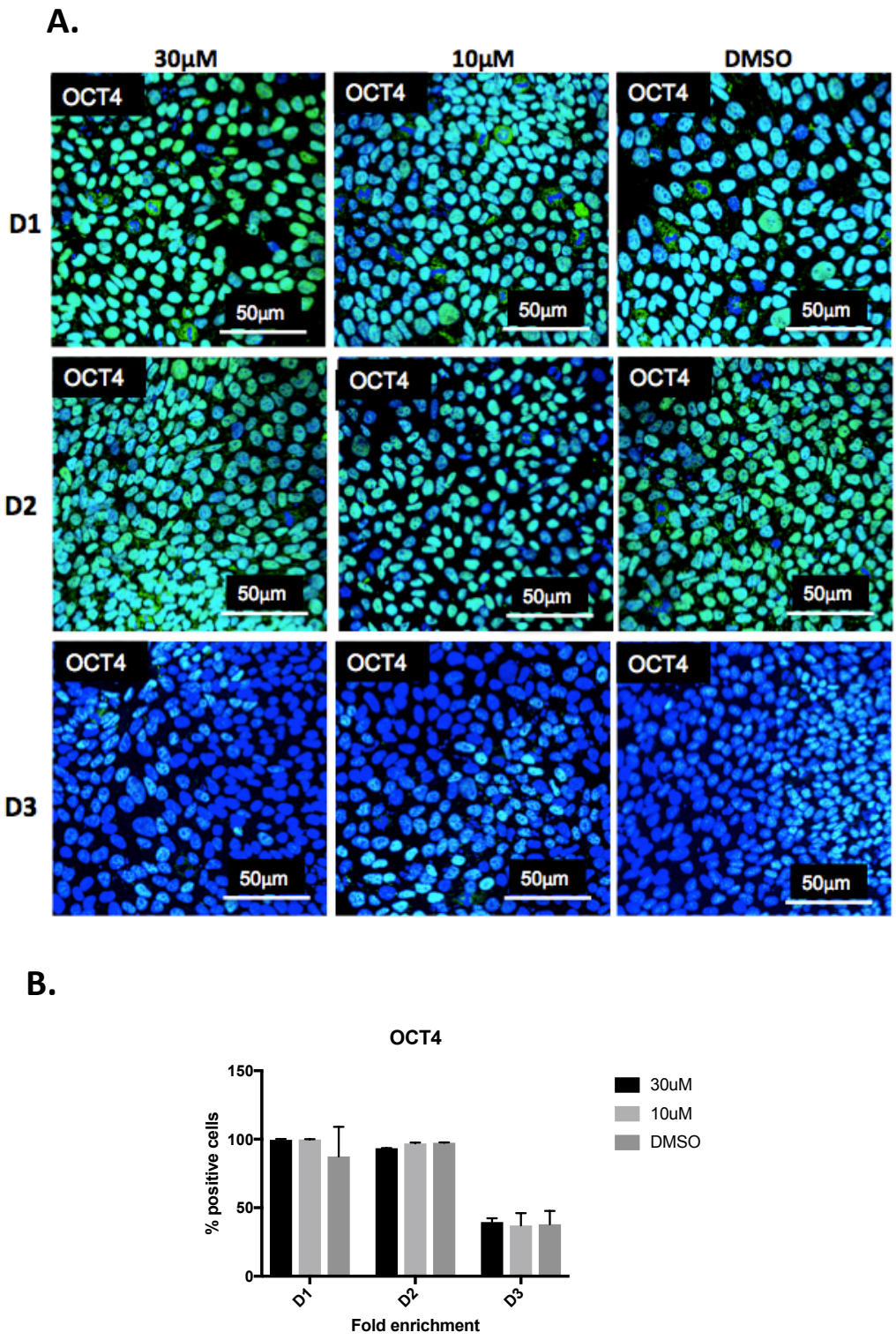



**Figure 4.9. OCT4 immunofluorescence analyses. A.** Expression of pluripotency marker OCT4 at d1, d2 and d3 after 2i neural induction with 30µM, 10µM 10058-F or DMSO treatment. Cell nuclei are stained with Hoescht 33342. **B.** Statistical analysis of immunofluorescence. The error bars represent the standard deviation among 3 technical replicates.

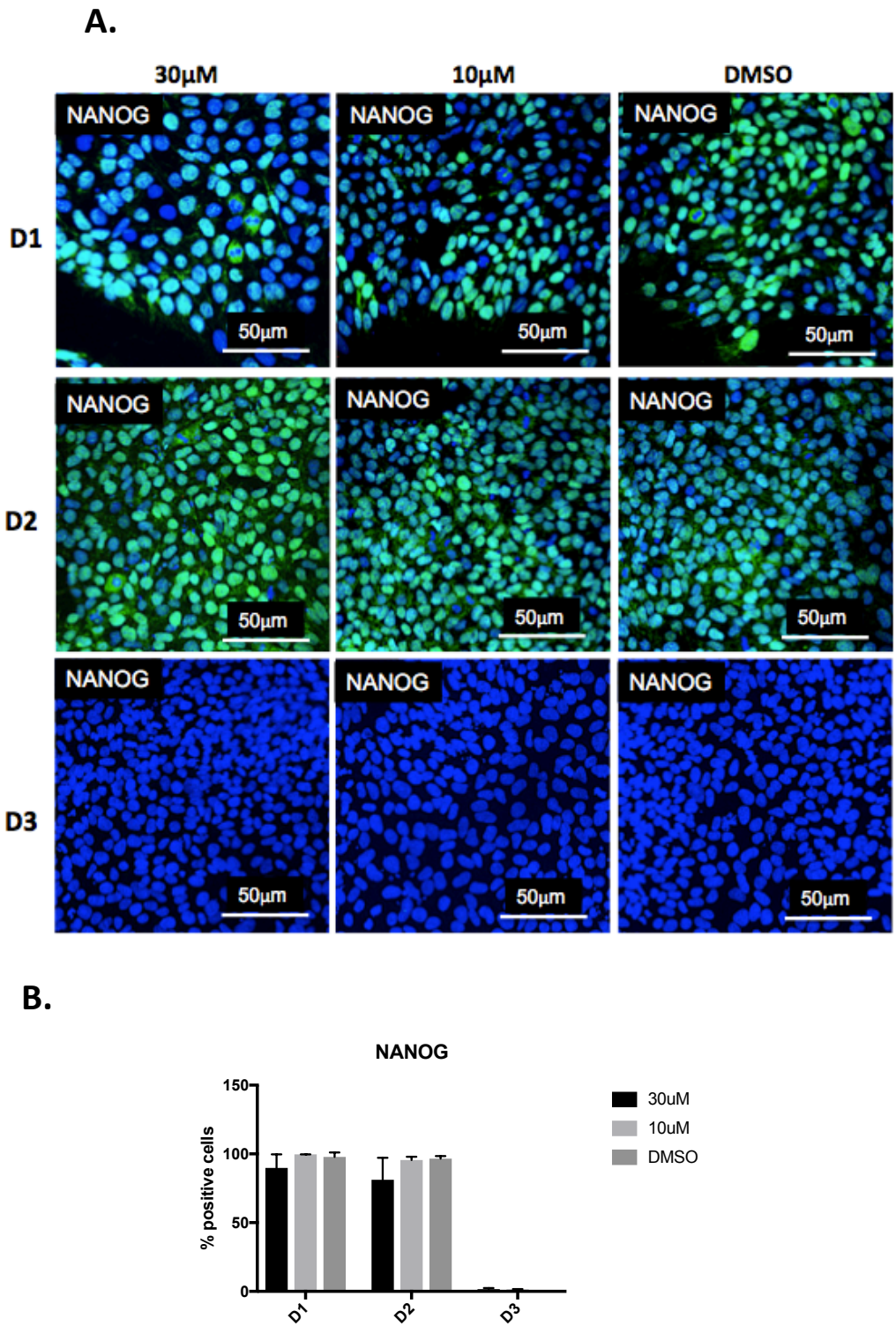



**Figure 4.10. NANOG immunofluorescence analyses. A.** Expression of pluripotency marker NANOG at d1, d2 and d3 of 2i neural induction with 30µM, 10µM 10058-F or DMSO treatment. Cell nuclei are stained with Hoescht 33342. **B.** Statistical analysis of immunofluorescence. The error bars represent the standard deviation among 3 technical replicates.

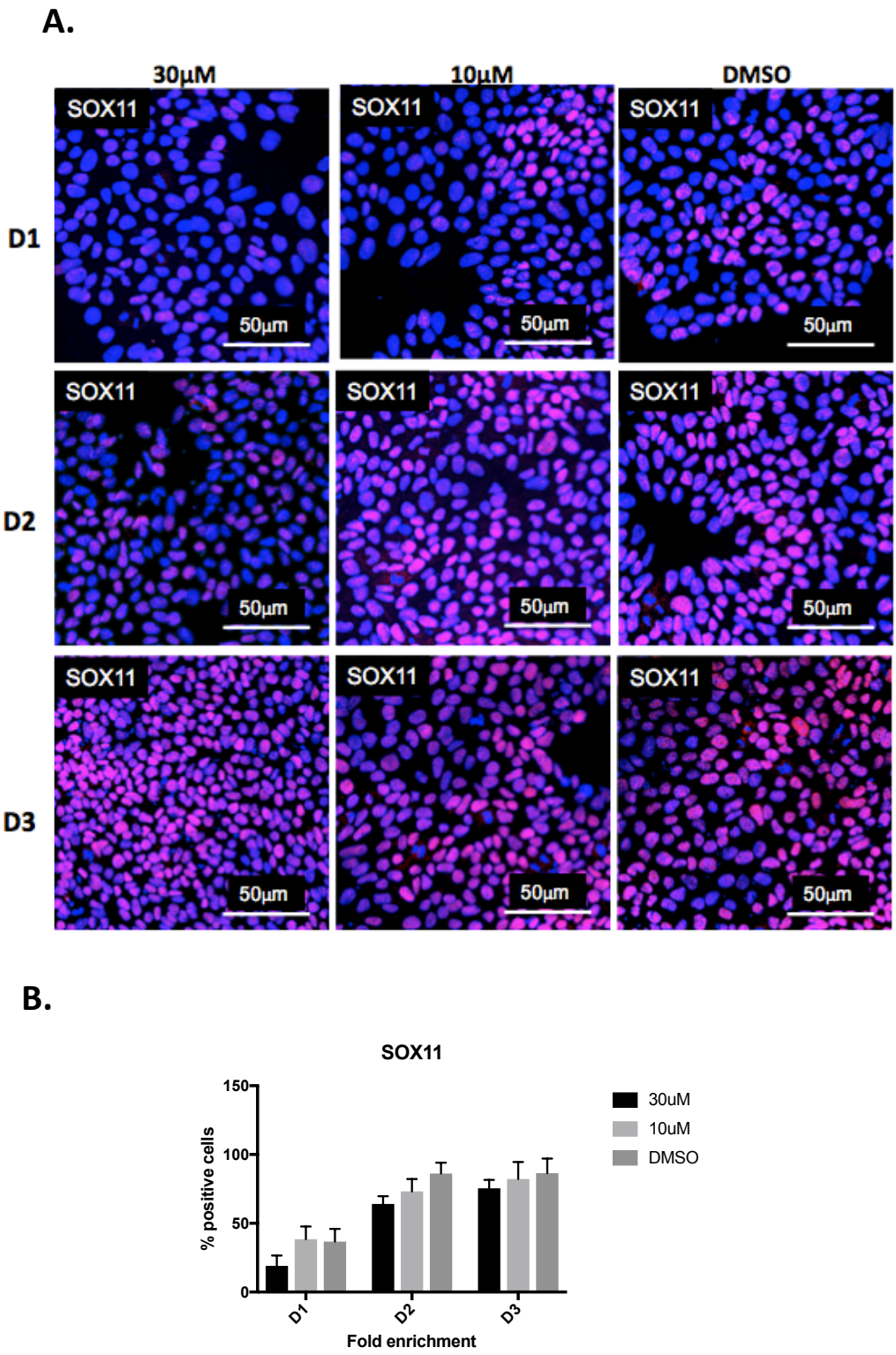



**Figure 4.11. SOX11 immunofluorescence analyses. A.** Expression of NE marker SOX11 at d1, d2 and d3 of 2i neural induction with 30µM, 10µM 10058-F or DMSO treatment. Cell nuclei are stained with Hoescht 33342. **B.** Statistical analysis of immunofluorescence. The error bars represent the standard deviation among 3 technical replicates.

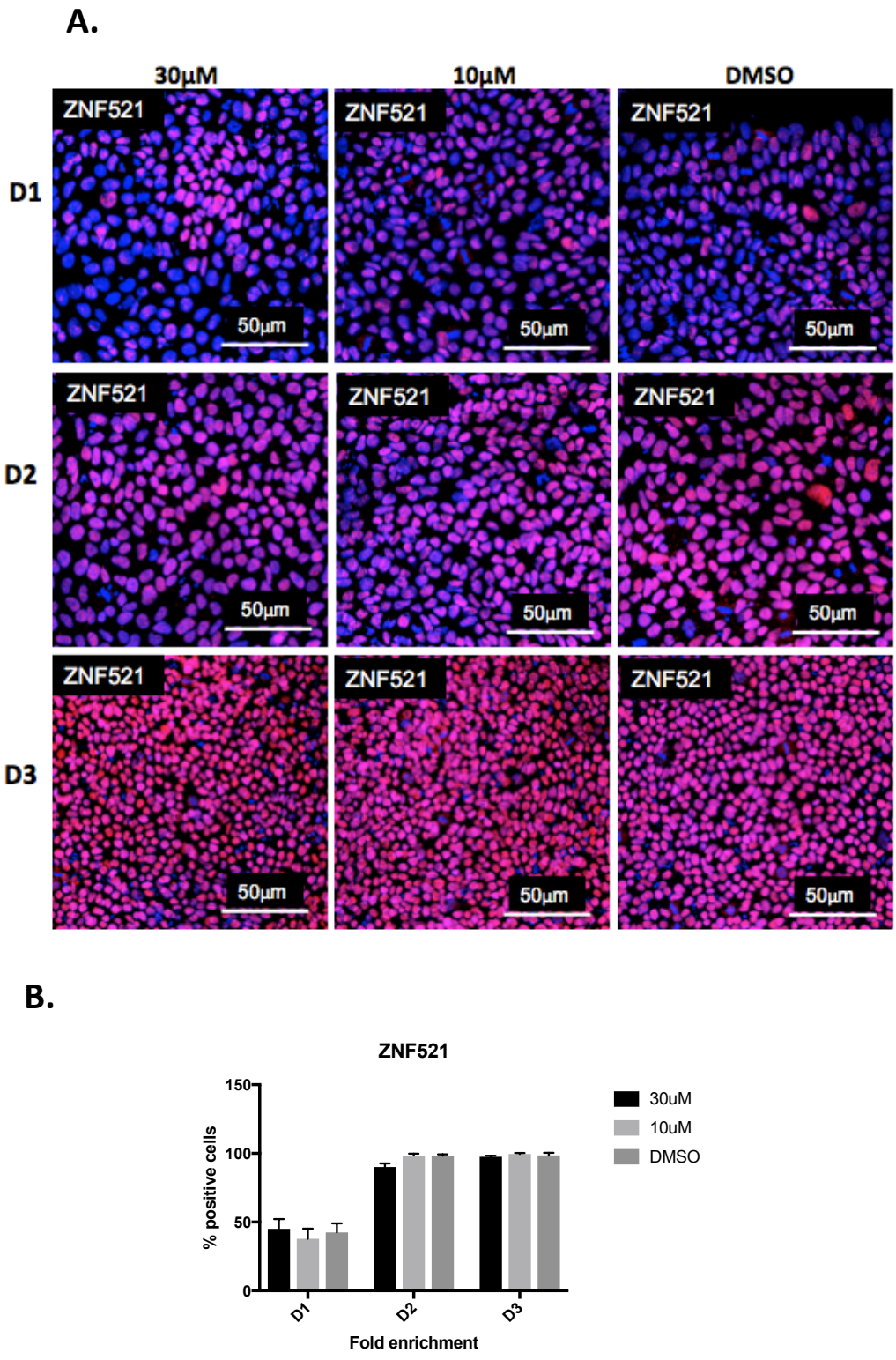



**Figure 4.12. ZNF521 immunofluorescence analyses. A.** Expression of NE marker ZNF521 at d1, d2 and d3 of 2i neural induction with 30µM, 10µM 10058-F or DMSO treatment. Cell nuclei are stained with Hoescht 33342. **B.** Statistical analysis of immunofluorescence. The error bars represent the standard deviation among 3 technical replicates.

The genes SMARCA2, CDH1 and TFAP2C are positively correlated to the hub MYC in the GRN. MYC and SMARCA2 are first differentially down-regulated at d1 with no time delay, whereas CDH1 and TFAP2C are first down-regulated at d3 with 2 days time delay compared with MYC. Q-PCR analyzes of these genes expression after 2i neural induction in the presence of 30µM or 10µM of the MYC inhibitor F10058-F or DMSO indicated that CDH1 and TFAP2C are significantly more highly expressed in cells treated with 10µM 10058-F compared with 30µM and DMSO at d1. SMARCA2, CDH1 and TFAP2C are significantly more highly expressed at in cells treated with 30µM 10058-F compared with 10µM or with DMSO at d2. CDH1 is significantly more highly expressed in cells treated with 30µM 10058-F compared with 10µM and DMSO at d3 (Figure 4.13). Statistical analyses details are included in the Appendix 4.8 (CD-ROM).

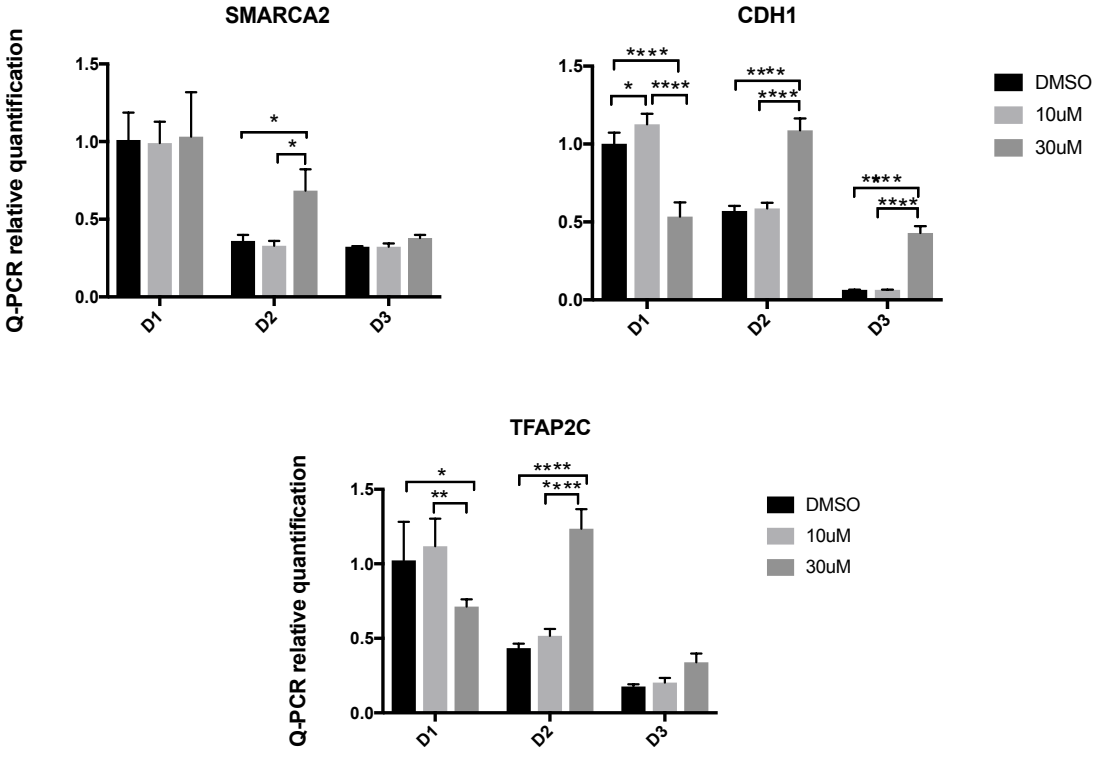



**Figure 4.13. Real-time PCR analyses of SMARCA2, CDH1 and TFAP2C.** Expression of SMARCA2, CDH1 and TFAP2C was quantified by Q-PCR at d1, d2 and d3 of neural induction in the presence of 10µM or 30µM F10058-F or DMSO.

## 4.4. Discussion

### 4.4.1. Gene regulatory network modules

We report for the first time a GRN comprising the gene regulatory interactions transversed during the differentiation of NESC from iPSC. A number of activated and down-regulated genes during different stages of NESC differentiation have been previously described in mammals, particularly using mouse as a model. However, interactions between these factors have not been deeply explored either in human or in mouse. The GRN generated by our collaborator Jun Ming GRN enables visualization of potential protein-protein and TF-target interactions among selected TFs that are differentially expressed during early neural induction (d1, d2 and d3) and a sub-subset of correlated genes that are differentially expressed at the same day or with 1, 2 or 3 days delay. To identify the potential biological functions regulated by the genes comprising the GRN I interrogated them for enrichment of GO terms. The results indicated that the biological processes “localization”, “response to stimulus”, “biological adhesion” and “locomotion” or genes included in these terms might play a role during neural induction.

I identify 9 hubs with high connectedness among the GRN that are considered modules. 7 modules highly interact with each other and include MYC, FOS and JUN, possibly generating feedback circuitry. Expression of these modules might regulate specific or multiple biological functions required by the cell at particular developmental stages. MYC interacts with 44 genes forming the largest module. 40 genes are positively correlated with MYC, reflecting that their expression is also down-regulated during NESC differentiation. MYC and correlated genes such us MTA1 and ENO1 are rapidly down-regulated at d1 after neural induction, whereas others are down-regulated at d2 and d3. GO analyses of the MYC

module indicate that the biological processes “programed cell death”, “cell proliferation” and “regulation of response to stress” are enriched for genes comprised in the module. These biological processes play important roles during different developmental stages, including neural induction. Perhaps, the genes involved in these terms are necessary for the maintenance of the pluripotency state in iPSCs but are subsequently down-regulated during neural induction. A GRN study conducted during the tumorigenic transformation of human primary cells, show that activation of MYC induces the expression of several gene modules, which regulate the biological processes that drive the cell to a cancerous phenotype, including rRNA synthesis and chromatin remodelling (Malysheva *et al*. 2016). Among this GRN, MYC is connected with several genes including SMARCA2 and SMAD7, which are also connected to the MYC hub in our GRN. However, no other genes and hubs are common between these GRN. Possibly, the gene modules and cell biological functions necessary for the differentiation of tumorigenic cells and the maintenance of iPSC are not highly interconnected during this stage of tumour development. Furthermore, the sets of TF and genes and the inferred regulatory interactions between nodes might vary between GRNs due to the variability of the statistical and mathematical methods used by both studies. Additionally, the MYC is essential in human and mouse for the efficient reprograming of iPSC and for the maintenance of pluripotent stem cells. For instance, a GRN modelled from PSC from the mouse embryo depicted MYC as a central hub in a pluripotency GRN that integrates multiple gene modules, which modulate various biological processes such us cell cycle, metabolism and epigenetic remodelling and inhibition of differentiation genes (Fagnocchi *et al*., 2016; Fagnocchi & Zippo, 2017). The genes connected with MYC in the mouse pluripotency GRN are not directly connected with the MYC hub in the network modelled from our data. However, several of the genes interacting with MYC in our GRN, such as FGF2, SMARCA2 and AHR are necessary for

pluripotency, indicating that the MYC module might play a role in pluripotency maintenance of iPSC. It is likely that the genes comprised in the both GRN differ due to the different mechanism used for selecting the genes to model the networks. Hence, the genes included in the mouse pluripotent GRN might have been excluded from our data prior to the netbook modelling. For instance, it is possible that down-regulation of these genes is not necessary for human iPSC differentiation toward NESC lineage. Therefore, these genes were not significantly down-regulated in our data set and were not included in the GRN. However, these genes might also play a role in the maintenance of pluripotency in human iPSC. Additionally, genes included in the mouse pluripotency network, which are not directly connected with MYC are also highly expressed at day 0 in our data set, including NANOG, SOX2 and OCT4. However, the statistical analyses used to infer the interactions of the GRN indicated correlation of these genes with other modules different than MYC.

The PAX6 hub is up-regulated after d2 of neural induction similarly to the majority of genes (8) interacting with the hub. GO analysis of the PAX6 module show enrichment for the terms “regulation of RNA biosynthetic process”, “regulation of transcription, DNA template”, “central nervous system development”, “sensory organ development”, “neural tube development” and “dorso-ventral pattern formation”. These regulatory functions might be required for activation and maintenance of the neural induction program. For instance, the 8 genes interacting with PAX6 in the GRN have been implicated in different steps of CNS development. Possibly, activation and interaction of genes within the module is necessary to initiate the sequential regulatory signals that direct neural induction. PAX6 is highly expressed in NE cells derived from ESCs and directly from human foetuses and it is required for the neural induction, as demonstrated by known-down studies in human ESC (Zhang *et al*., 2010; Blake & Ziman, 2014). It has been suggested that PAX6 promotes

neural induction by inhibiting the pluripotency factors OCT4, NANOG and possibly by activating NE genes. Our results suggest that PAX6 might interact with the TFs RAX, GLI3, MEIS2, LHX2 and NR2F2 at d2 after 2i neural induction to promote the transition between PSC and NESC. Importantly, the gene module might interact with other genes not included in the network to modulate differentiation. Genome editing techniques such as CRISPR/CAS9 and CRIPSPRi would enable silencing of these TFs at specific times during neural induction to test the GRN predictions. A recent study, show that Pax6 is the central hub a GRN necessary for the activation and maintenance of radial glial cells and cortical neurons differentiated from mouse ESC. The results indicated that Pax6 hub activates neuronal genes and inhibits mesodermal and endodermal genes to ensure unidirectionality of the differentiation program (Sun *et al*., 2015; Thakurela *et al*., 2016). However, the genes include in this study were not comprised in our data set and GRN. For instance, neuronal genes included in the radial glial GRN, such as ASCL and NEUROG2, are expressed at low levels during NESC differentiation. Perhaps, these GRN is activated posteriorly to NESC differentiation and proceeds the regulatory interactions of PAX6 module described in this study. Aditionally, PAX6 is the central regulator of the mammalian lens developmental program. Analyses of the regulatory interaction of Pax6 during mouse lens development indicated that the TF directly activates modules of crystalline genes, cell cycle exit genes, Wnt signalling genes and extra cellular matrix genes, biological processes that might be required for lens differentiation (Sun *et al*., 2015; Antosova *et al*., 2016). However, these gene modules might not be required for NESC differentiation from iPSC in humans since the genes comprised in the modules are not interacting with PAX6 in the GRN modelled form our data.

Similarly, the SMAD2 hub is differentially up-regulated at d3 of neural induction. SMAD2 shows a connectedness of 19 and target genes are up- or down-regulated at either d6 or d8.

GO analyses shows enrichment for “regulation of gene expression”, “regulation of RNA metabolic process”, “regulation of RNA biosynthetic process” and “regulation of transcription, DNA-templated” terms. SMAD2 module might modulate changes in RNA processing, which impact DNA transcription after d3 of neural induction. These changes might be necessary for NE induction or for inhibition of genes that induce differentiation to other cell phenotypes such as pluripotency, but the precise mechanism remains to be investigated.

CDC6 is initially up-regulated after d2 and is subsequently down-regulated after d3. CDC6 interacts with 14 genes, all of which are down-regulated at d4 and d6. GO analysis reveals enrichment for “regulation of cell cycle process”, “cell cycle phase transition”, “G1/S transition” and “DNA replication” terms. Several studies have demonstrated that cell fate decisions are tightly synchronized with specific cell cycle profiles (Sakaue-Sawano *et al*., 2008). For instance, in mouse and human ESCs, pluripotency is associated with a truncated G1 phase that lacks G1 checkpoint regulation. Increase in the length of G1 phase through CDK inhibition leads to spontaneous differentiation of ESC/iPSC, possibly by increasing the time of exposure to specification signals (Ruiz *et al*., 2011; Soufi & Dalton, 2016). Moreover, ESC initiates cell fate decisions at G1 state. Pluripotent cells at early G1 differentiate into mesendoderm due to elevated expression of SMAD2/SMAD3 during this phase. Cells at late G1 acquire neural-ectoderm differentiation competence due to accumulation of cyclin D, which activates CDCK4/6 leading to phosphorylation and reduction of SMAD2/3 (Pauklin & Vallier, 2013, Soufi & Dalton, 2016). Additionally, it has been shown that neural progenitor cells have an extended G1 phase and short S phase during neuronal differentiation. For instance, truncated G1 and elongated S1 in neural progenitors leads to reduced production of cortical neurons, indicating that completion of G1 might be necessary for establishing the molecular program that directs neural

differentiation (Lanctot *et al*., 2017). Possibly, down-regulation of gene modules enriched for "cell cycle phase transition" and "G1/S transition" at d4 and d6 enhances neural differentiation by promoting elongation of G1 phase or by a different mechanism through regulation of cell cycle progression. However, the precise means by which this gene module inhibits or promotes neural induction remains to be elucidated.

### 4.4.2. MYC inhibition

Validating gene-gene relationships inferred from the GRN can be achieved by testing the network predictions using specific interventions (Olsen *et al*., 2014). The small molecule inhibitor 10058-F inhibits MYC-MAX interaction preventing regulation of their gene targets. I inhibited MYC activity during the first 3 days of neural induction to identify possible effects in the expression of other genes of the GRN. Since the MYC module is highly connected with other GRN modules, I hypothesized that inhibition of MYC may affect expression of other genes within the GRN. The results indicated that MYC inhibition does not affect expression of other pluripotency markers, NANOG and OCT4, or NE markers, SOX11 and ZNF521. However, inhibition of MYC leads to an increase in expression of SMARCA2, CDH1 and TFAP2C after 2 days and CDH1 by d3. These factors are significantly down-regulated after 2 (SMARCA2) and 3 days (CDH1 and TFAP2C) of 2i induction and are directly connected with MYC in the GRN. Importantly, MYC expression is differentially down-regulated at d1. Hence, initial levels of MYC during the first day of neural induction might be necessary to enhance down-regulation of SMARCA2, CDH1 and TFAP2C. Premature inhibition of MYC with 10058-F during 2i differentiation might prevent these regulatory mechanisms. Hence, the results suggest that MYC exerts a regulatory effect up-stream of these genes, validating the inferred connection between MYC and SMARCA2, TFAP2C and CDH1 in the GRN. It seems that MYC function is

compensated, possibly by other factors, at d3 for SMARCA and CDH1 inhibition but not for CDH1. Induction of MYC expression during the initial 3 days of 2i induction might provide additional data to infer MYC regulatory effects over the genes in the GRN module.

### 4.4.3. Limitations

GRNs are often modelled based heavily on known TF regulatory interactions. However, the gene expression patterns transversed during development are also regulated by post-translational and epigenetic modifications. Incorporating these regulatory events in the GRN would vastly improve the GRN predictions; but requires elevated resources and extended periods of time due to the computational complexity of the models. Accordingly, epigenetic and post-translational modifications are rarely incorporated and are not included in the GRN modelled by our collaborators in this study. Hence, not all regulatory mechanisms that control the formation of specific gene product abundances are included. It is important to realize that this does not invalidate the network, just that certain interactions or edges may obscure undetected nodes.

An enormous number of potential interactions arise from the data generated from high-throughput technologies, increasing the complexity of the GRN topology. Filtering the data according to specific criteria is essential to reduce the dimensionality of the GRN. The GRN inferred by Jun Ming comprises only selected TFs that are highly connected with other genes. These interactions are based on prior knowledge of the regulatory functions of these factors and might exclude interactions with biologically relevant TF that are unknown. Additionally, the GRN nodes might interact with constitutively expressed genes that do not show differential expression during neural induction (these would be an example of undetected nodes – see above). Identification of genes that might be biologically

meaningful without prior knowledge of all relevant interactions remains challenging. Hence, the GRN shows potential causal relations between genes at different states of differentiation, but possible intermediate nodes are not evident in the network. Additionally, visualization of specific spatiotemporal states of the GRN may improve by integrating the time points and direction (up-regulation or down-regulation) of the TF differential expression. For future studies, I will to optimize the GRN topology by testing the GRN predictions through combinatorial gene activation and repression (using CRISPRi and CRISPRa). Additionally, I will perform ChIPseq and proteomic analyses would enable to continue the refinement of the GRN model.

### 4.4.4. Conclusion

Interdisciplinary collaboration between experimentalists and modellers enables computational modelling of GRN from large-scale transcriptional datasets. Accordingly, our collaborators Jun Min and Shankar Subramanian (UCSD) modelled a GRN from the high-throughput transcriptional datasets obtained from our experimental approaches. The GRN shows the potential causal relations of a set of selected TF with a subset of genes during different stages of NE differentiation. The regulatory interactions between these TF and genes lead to distinct network topologies at different stages of differentiation, representing dynamic changes in gene expression levels and connectivity. This network controls the different biological processes carried out during neural induction. I identified gene modules within the GRN and interrogated them GO analysis, which enabled determination of the distinct biological processes operative during different stages of development.

# CHAPTER 5. Single cell RNAseq

## 5.1 Introduction

The development of techniques for single cell isolation, and cDNA synthesis and amplification from a small amount of starting RNA, enable quantitative analyses of the transcriptome at a single cell resolution (Tang *et al*., 2010, Poulin *et al*., 2016). These approaches revealed that isogenic cell populations exposed to identical micro-environmental conditions could depict heterogeneous gene expression levels (Elowitz *et al*., 2002; Tang *et al*., 2010; Fujita *et al*., 2016). It has been proposed that transcriptional bursting is the major driver of gene expression variability. Individual genes show different burst kinetics, resulting in different gene product abundances among cells within a population, which might impact cell fate commitment decisions. Measurement of bulk expression of the population can obscure fluctuations in gene expression of single cells (Koern *et al*. 2005). Hence, considering only the average expression of a population could mask subpopulations with different burst kinetics, which may be fated to acquire different phenotypes or may represent intermediate states of differentiation (Islam *et al*. 2011; Trapnell *et al*. 2014; Buettner *et al*., 2015) (Figure 5.1). Additionally, to identify the regulatory relations among genes underlying different stages of development, it is important to determine that the interacting genes are expressed in the same individual cells (Moignard & Gottgens, 2014).

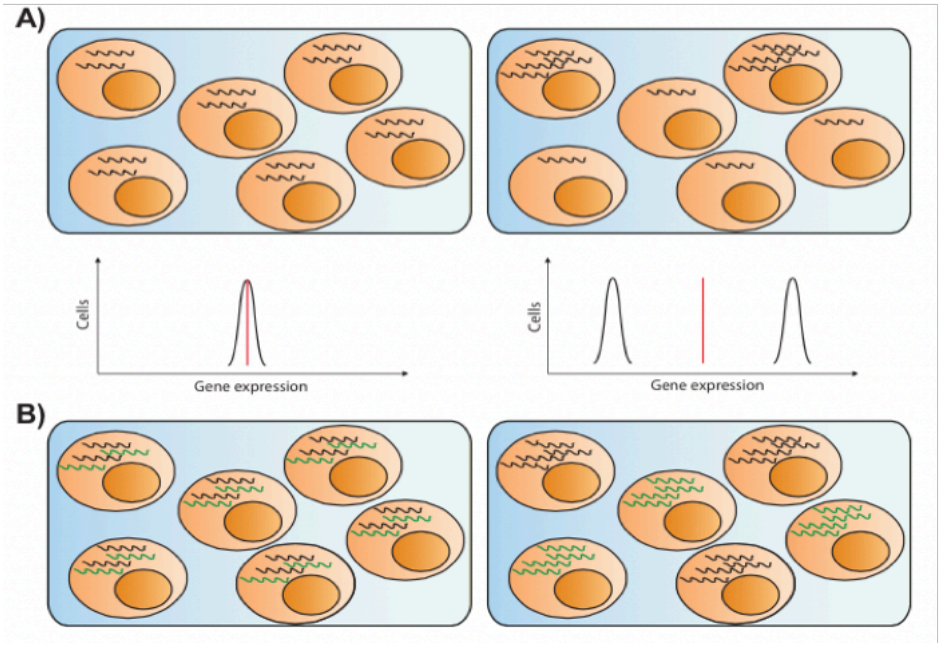



**Figure 5.1. Gene expression profiles from single and bulk cells analyses. A**. Expression analyzes from bulk cells show that all cells in the population express similar levels of a transcript (top left) giving a unimodal distribution (mid left); whereas single cell analyses indicate that some cells among the population show low levels of expression and other high levels of expression (top right), leading to a bimodal distribution (mid right). **B** Single cell expression analyses separate the expression of different transcripts in individual cells (bottom right) from co-expression of these transcripts in the cell population (bottom left) (Moignard & Gottgens, 2014).

### 5.1.2. Transcriptional bursts

Gene transcription comprises a series of biochemical reactions that are determined by molecular concentrations, diffusion rates and TF dynamics. Hence, factors extrinsic to the gene such as cell size, cell cycle phase and the concentrations of molecules required for transcription might add variability to the process (Munski *et al*., 2012; Schoech & Zabet, 2014; Nicolas *et al*., 2017). However, the main source of transcriptional fluctuation among cells is intrinsic transcriptional bursts, which lead to variation of mRNA transcript production due to irregular gene transcription. Integration of quantitative expression of gene transcripts with mathematical models has led to the formulation of 2 main models to

explain transcriptional bursting. The one-state model suggests that mRNA is produced and degraded in a constant rate proportional to the number of mRNA molecules, where a Poisson distribution characterizes the distribution of mRNA among the cell population (Corrigan *et al*., 2016; Nicolas *et al*., 2017). The two-state model of gene regulation better explains the distribution of genes whose expression does not follow a Poisson distribution, which is common in higher eukaryotes. The model assumes that short episodes of RNA synthesis are followed by transcriptional silence at a constant rate (Islam *et al*., 2011; Fujita *et al*., 2016; Nicolas *et al*., 2017). In this model, promoters have 2 states: the "on" state when the promoter is randomly activated, producing a burst of mRNA transcripts, and an "off" state when the promoter is inactive, and no transcription occurs (Munski *et al*., 2012; Corrigan *et al*., 2016; Soltani & Sight, 2016). The two-states model generates variable distribution of mRNA transcripts depending on the burst size (number of transcripts produced in each burst) and frequency (frequency of burst occurrence) at a specific time point. For instance, short "on" and long "off" periods produce occasional mRNA busts with high variance and a long tailed distribution; whereas long periods of activation or inactivation generate bimodal distribution with clearly delineated "on" and "off" populations (Munski *et al*., 2012; Nicolas *et al*., 2017).

The molecular mechanisms that drive transcriptional variability are not well understood. However, several studies have identified a number of factors that might influence the burst shape including local genomic environment, the nucleosome occupancy in the transcription-starting site (TSS), histone marks and DNA regulatory elements (Raj *et al*., 2006; Muramoto *et al*., 2010; Dar *et al*., 2012; Fukaya *et al*., 2016; Nicolas *et al*., 2017; Hendy *et al*., 2017) (Figure 5.2). These mechanisms lead to changes in TF availability at the DNA binding sites, suggesting that TFs availability impacts transcriptional bursting (Nicolas *et al*., 2017). In addition, it has been demonstrated that at lower concentrations of

TFs, the dominant source of variability in transcriptional bursts of gene targets is the frequency, whereas at higher concentrations of TFs the main difference is the size (Carey *et al*., 2013).

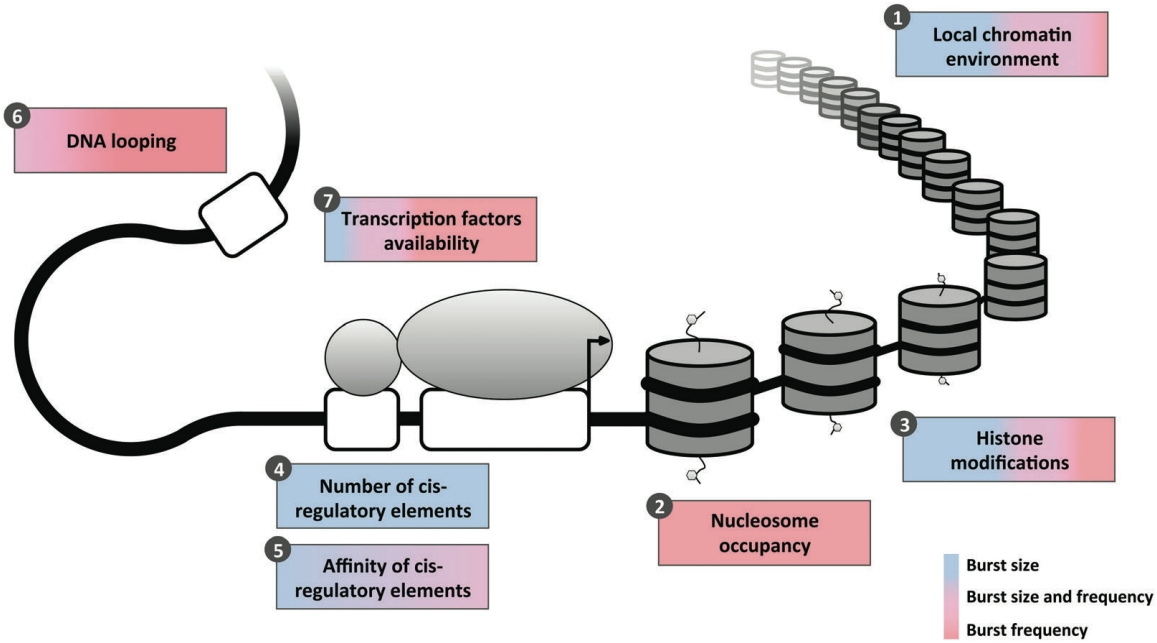



**Figure 5.2. Regulatory mechanism of transcriptional bursting.** Transcriptional burst size and frequency are influenced by the local chromatin environment; histone modifications; number and affinity of promoters (large square) core regulatory elements (*cis*) (small squares); formation of loops with distal regulatory elements; and the availability of TF. These molecular mechanisms are depicted in boxes, where the color represents their role in regulating the size (blue) and the frequency (pink) of transcriptional bursts (Nicolas *et al*., 2017).

Additionally, the transcriptional bursting frequency might fluctuate among cells during distinct cell cycle phases and different sized cells (Soltani & Sight, 2017). During the S phase of the cell cycle the cell duplicates the number of gene copies, which might double the mRNA transcripts produced. A recent study analyzed the transcription kinetics of Nanog and Oct4 in mouse PSCs throughout the cell cycle and demonstrated that frequency of transcriptional burst is reduced after DNA replication to compensate additional DNA

copies of cells in S and G2 phases. These findings suggest that cell cycle progression might influence the frequency of transcriptional bursts (Padovan- Merhar *et al*., 2015; Ben-Moshe & Itzkivitz, 2016; Skinner *et al*., 2016). Similarly, it has been show that larger cells increase the burst size to maintain the appropriate mRNA abundances despite the cell size (Padovan- Merhar *et al*., 2015).

Variability in gene transcriptional bursting among individual cells might have significant biological functionality in relation of cell fate commitment (Chang *et al*. 2008; Yan *et al*. 2013; Trapnell *et al*. 2014). The biological relevance of tissue heterogeneity has been highlighted in a recent study that demonstrated that heterogeneous populations as early as the 4-cell mouse embryo bias cell fate decisions. The TF Sox21 show the highest heterogeneous expression at this stage, with lower Sox21 expressing cells acquiring extra-embryonic fate, whereas cells expressing higher level of Sox21 were maintained in a pluripotent state (Goolam *et al*., 2016).

### 5.1.3. Gene expression changes during the cell cycle

During mitosis (M), several nuclear processes are disrupted by the displacement of multiple components from the chromatin, including several TFs, epigenetic marks and RNA Pol II polymerase (Pol II), and the dissociation of enhancers-promoter interactions. Dissociation of these factors leads to interruption of RNA synthesis (Young *et al*., 2007; Hsiung *et al*., 2016; Liu *et al*., 2017). However, specific regions of the mitotic chromatin remain associated with particular histone marks and TFs. Maintenance of these genes during mitosis is referred as bookmarking. TFs and epigenetic marks might pass to the newly born cell facilitating displacement and recruitment of other TFs, providing an inheritance mechanism (Wang & Higgins, 2013a; Liu *et al*., 2017). Subsequently, during late telophase

Pol II and global TFs are recruited to the newly formed cell nuclei and the cell restarts RNA synthesis (Prasanth *et al*., 2003; Hsiung *et al*., 2016). The transcription of genes expressed during interphase is progressively reactivated during G1 phase. However, a few genes important for cellular functions such as housekeeping are preferentially re-activated earlier during M/G1 transition (Fukuoka *et al*., 2012). Additionally, a recent study reported that during the first rounds of transcription after mitosis at G1, the majority of activated genes are transcribed at their highest levels during cell cycle. M/G1 transcriptional peak might occur heterogeneously among cells and might increase transcriptional differences among the cell population (Hsiung *et al*., 2016).

Heterogeneous gene expression results in variability of mRNA and protein levels within the cell population leading to different rates of cell cycle progression among individual cells (Matson & Cook, 2017). Additionally, unequal distribution of RNA, proteins, metabolites and organelles might contribute to cell cycle differences among cells, which remains for more than two cell cycles. Hence, proliferation or exit from the cell cycle decisions are made individually by each cell of the population, leading to distinct subpopulations with different proliferation dynamics. A cell population might comprise proliferating, quiescent or senescent cells, which are evident only with single cell analyses (Sakaue-Sawano *et al*., 2008; Overton *et al*., 2014; Matson & Cook, 2017).

Several studies have reported that cell cycle progression is strongly related with cell fate choice. The cell cycle machinery might orchestrate maintenance of pluripotency or cell specification to particular cell types during different phases. As mentioned in Chapter 4, transition trough G1 phase establishes a window of opportunity for cell cycle exit and differentiation. hESCs differentiate into mesendoderm during early G1 phase, whereas cells

in late G1 acquire neuroectoderm fate (Pauklin & Vallier, 2013, Soufi & Dalton, 2016). Furthermore, it has recently been reported that during S and G2 phases, the damage checkpoint factors ATR/ATM and cyclin B enhance TGF-β/Nodal activity, establishing an intrinsic propensity towards pluripotency in hESCs. These mechanisms might potentiate DNA damage repair since damage repair proteins are largely expressed in ESCs compared to differentiating cells; additionally, if the damage is not repairable, ESCs undergo apoptosis effectively since they have lower apoptotic threshold than differentiating cells. Restriction of ATR/ATM during G1 phase enables the inhibition of pluripotency, leading to differentiation. Hence, these studies demonstrated that the molecular mechanism underlying specific cell cycle stages direct selective preference toward pluripotency or differentiation in hESCs (Gonzales *et al*., 2015; Vallier, 2015) (Figure 5.3).

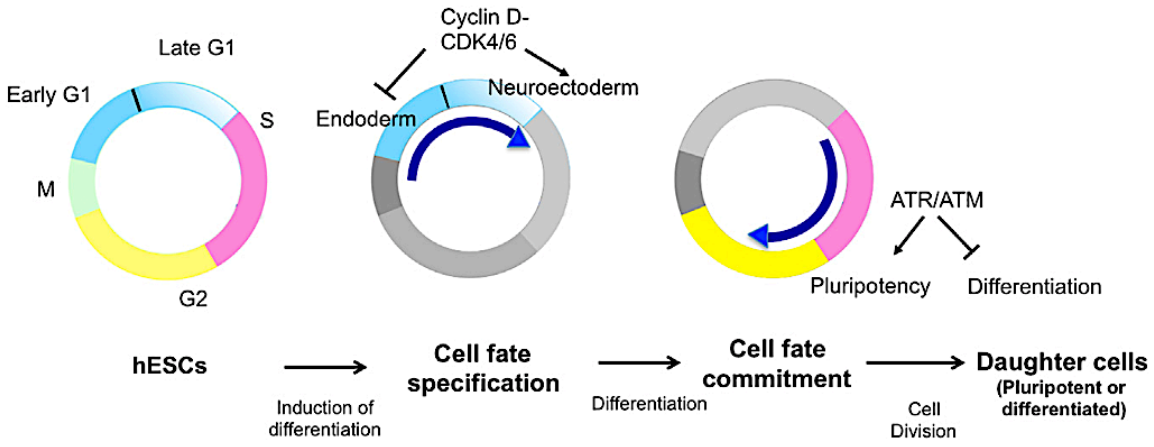



**Figure 5.3. Pluripotency and differentiation during the cell cycle.** During early G1, cells tend to differentiate into endoderm due to the expression of SMAD2/3, whereas at late G1 accumulation of cyclin D and activation of CDK4/6 leads to phosphorylation of SMAD2/3 and acquisition of neuroectoderm lineage. During S and G2 phases ATR/ATM kinases enhance TGF-β/Nodal signaling promoting pluripotency maintenance (Vallier, 2015).

For this study we used single cell analyses to compare the expression patterns of individual cells at d8 after neural induction, to assess population heterogeneity and identify potential

cell sub-populations that may be primed to acquire different cell phenotypes or may represent novel transitional states in the NESCs population.

## 5.2. Methods

### 5.2.1. Single cell RNAseq

A confluent well from a Nunc 6 well plated with around 3,000,000 cells that were neural induced using 2i for 8 days was used for a single cell RNAseq assay. Media was removed, and the cells were incubated for 4 minutes at 37°C with 1ml of accutase. After incubation, cells were detached from the well by pipetting 4 times. The accutase cell suspension was transferred to a 50ml tube containing 6ml of DMEM/F12 without phenol red (Thermo Fisher, 21041025). Cells were centrifuged at 900rpm for 2 minutes and the media was removed. Cells were re-suspended in 1ml of DMEM/F12 and filtered through a 4µm cell strainer (Falcon, 08-771-1). Cells were counted and diluted with DMEM/F12 to obtain approximately 500,000cells/ml. A final concentration of 500ng/ml of DAPI (Sigma Aldrich, D9542) was added to the cell suspension to check for viability. Cell sorting was performed in the cell sorting facilities of the Weatherall Institute of Molecular Medicine (WIMM) at the University of Oxford. Single cells were sorted using a BD FACS ARIA III instrument considering single cells, viability, size and granularity by the WIMM technician. Sorted single cells were collected into 4µl of Smart-seq2 lysis buffer containing 2.5mM dNTPs (Thermo-Fisher) and 2.5µM Oligo-dT30VN (Biomers.net) in a 4°C cooled 96 well plate. Libraries were prepared and sequenced by our collaborator Dr. Neil Ashley according to the standard procedures of the WIMM (Picelli *et al.*, 2914). Briefly, single cell cDNA synthesis was performed using SuperScript II (Invitrogen, 18064-014) and was amplified with the KAPA Hifi Hot Start Ready Mix KAPA (Biosystems, KK2601). Smart-seq2

cDNA was end-paired, A-tailed and commercial adapters and indexes were ligated with the Illumina Nextera XT Kit (Illumina, FC-131-1024). Samples were pooled and sequenced with the NextSeq® 500/550 High Output Kit v2 (FC 404-2005) in a NextSeq 550 system by our collaborator Jerome Nicod in the Wellcome Trust Centre for Human Genetics. 95 Nextera samples were sequenced in one lane at a combined depth of 400,000,000 reads (approximately, 4,000,000 reads per cell). FastQ files were aligned to a reference genome trough the Tuxedo pipeline by our collaborator Leo Perfect (Price laboratory).

### 5.2.2. Single cell analysis

FASTQ files were aligned to the GENCODE human genome release 24 (GRCh38.p5) concatenated to the ERCC RNA spike-in FASTA file using Tophat (ref). The "no novel junctions" option was selected to restrict alignment to genes contained in the GENCODE human transcriptome annotation release 24. Cuffnorm was used to calculate FPKMs. A Pearson's correlation analysis was performed with the RPKMs obtained from the bulk RNAseq data at d8 from the iPSC1 and the FPKMs from the single cell RNAseq data. The correlation was calculated with Prism package of GraphPad software. Raw read counts were extracted from the BAM files in R using the summarizeOverlaps function from the GenomicFeatures package in Bioconductor. The raw counts were adjusted for the size of the library and the gene and were log2 normalized with the dependency EdgeR in the package "scater 1.4.0" in RStudio. Quality control was assessed with the function QC metrics. The quality control figures, gene expression plots and PCAs were performed the package "scater" in RStudio with the functions plotQC, plotPCA and plotExpression.

## 5.3. Results

### 5.3.1. Correlation between bulk RNAseq and single cell RNAseq data

RPKMs obtained from bulk RNAseq analyses from the cell line iPSC1 at d8 were compared with the average FPKMs from the single cell RNAseq analyses from the same time point and cell line using a Pearson's correlation. A significant positive correlation ($R=0.56$, $p < 0.0001$) was observed between the 19218 genes compared between the 2 data sets (Figure 5.4). The results indicate that the gene expression of the genes obtained by bulk RNAseq analyses is significantly associated to the single cell RNAseq data.

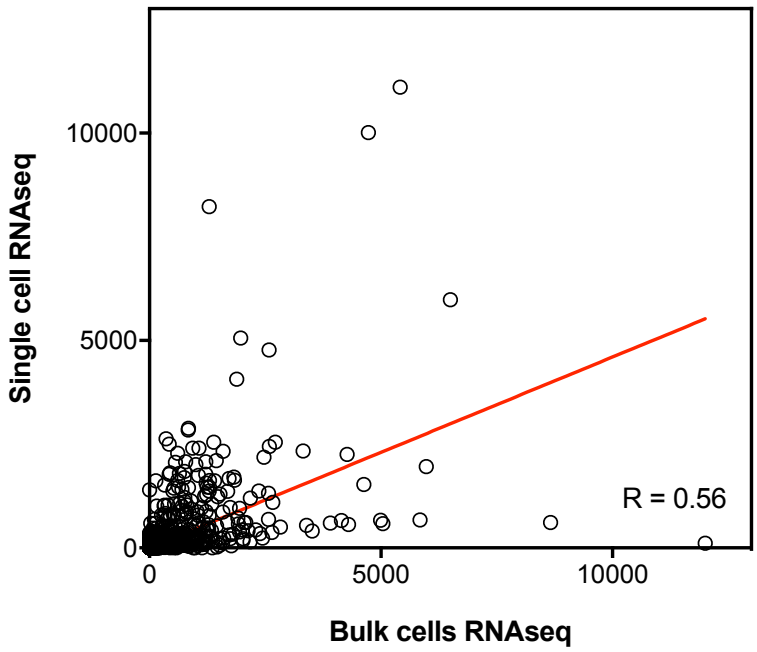



**Figure 5.4. Correlation between bulk RNAseq data and single cells RNAseq data.** RPKMs from bulk RNAseq analyses and average FPKMs from single cells RNAseq for 19,218 genes at are depicted with circles.

### 5.3.2. Single cell RNAseq quality control

Quality control (QC) analyses of single cell RNAseq data were performed to identify low quality cells including cells with a high percentage of reads corresponding to the RNA spike-ins, which suggest capture of a dead cell or a failure during library preparation. The average expression of the gene among all cells, the proportion of cells expressing the gene and the counts of control genes (genes constitutively expressed), mitochondrial genes and spike-ins are computed to determine low abundance genes and dropout rates (MacCarthy *et al*., 2017).

Plotting the distribution of reads among genes identifies libraries where a reduced number of genes contain the majority of counts and libraries where counts are evenly distributed among the genes. The proportion of the library that accounted for the 6000 most highly expressed genes was analyzed. The results indicated that the majority of counts are evenly distributed among genes. However, for one cell the majority of counts corresponded to a reduced number of genes, which possibly are the RNA spike-ins. Hence, this cell was excluded from down-stream analyses (Figure 5.5).

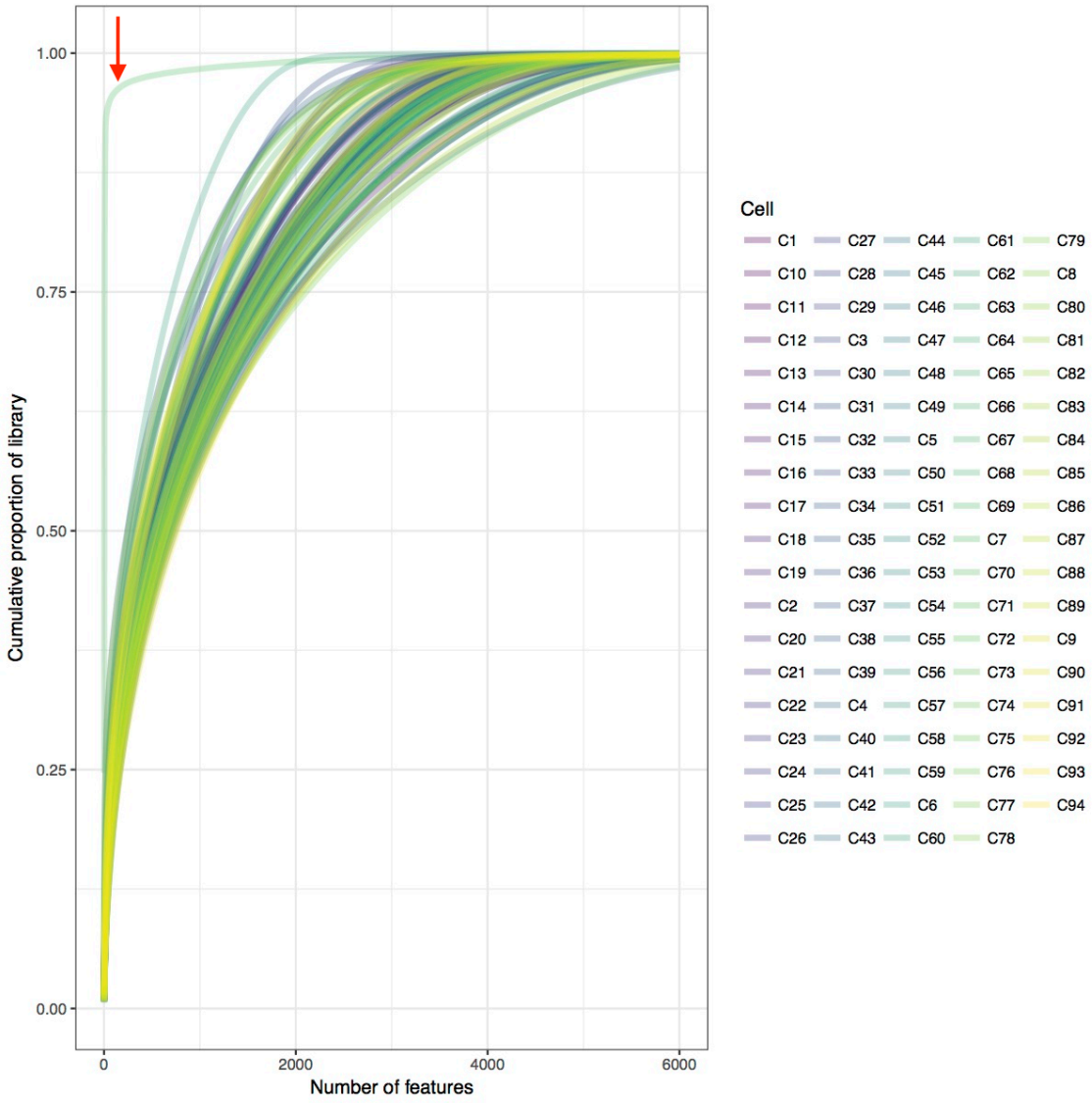



**Figure 5.5. Cumulative proportion of the library accounted for the 6000 most expressed genes.** The cumulative proportion of the library (total counts) is shown in the Y-axis and the 6000 most expressed genes are show in the X-axis. The cells are represented with different colors. The arrow shows the cell where the majority of counts correspond to a reduced number of genes.

The 50 most highly expressed genes in healthy cells are often mitochondrial, ribosomal, constitutively expressed genes and RNA spike-ins (when used for RNAseq). Analyses of the expression of these genes thus enable verification that transcripts levels are behaving as expected. The results indicate that mitochondrial genes such us MT-CO1, MT-CO2 and MT-CO3; ribosomal genes including RPS19, RPS2 and RPL7; constitutively expressed

genes such us ACTB, GAPDH and UBC; and RNA spike-ins are among the highest expressed (Figure 5.6).

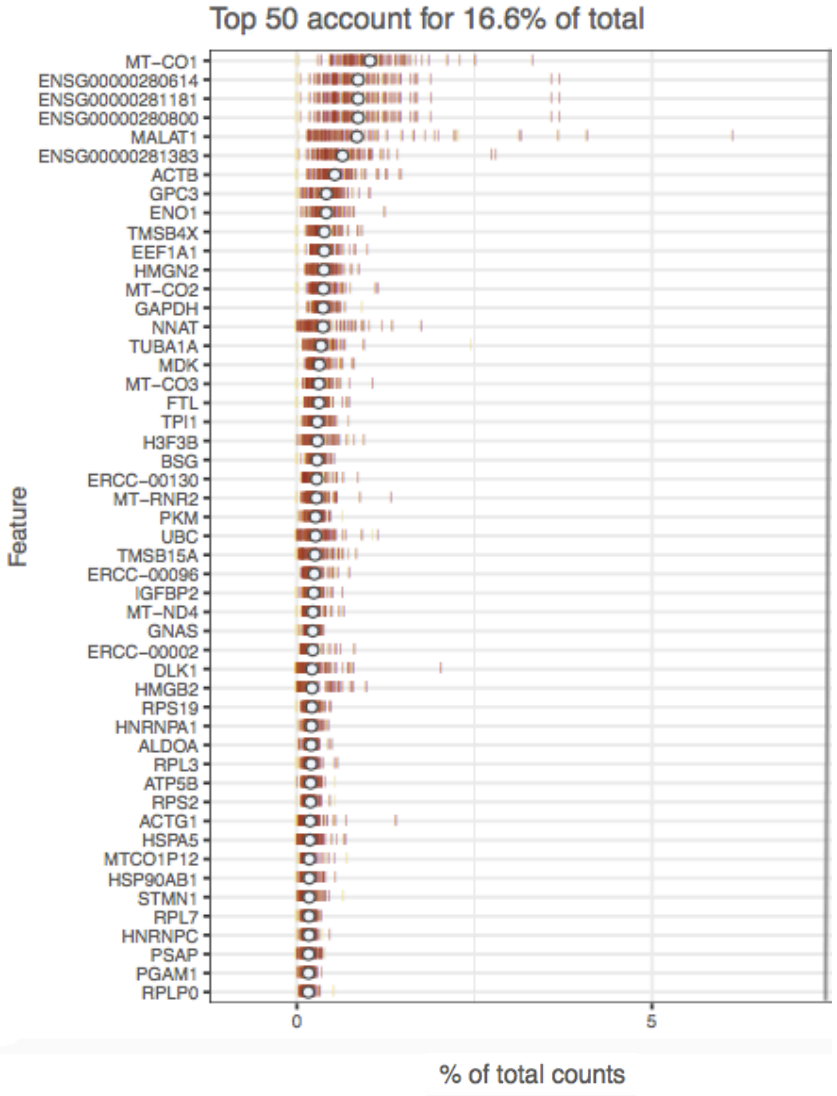



**Figure 5.6. 50 most highly expressed genes.** The percentage of the total counts is indicated in the X-axis and the genes names in the Y-axis. The circles show the percentage of counts among all cells and the bars represent the gene counts for individual cells. The 50 highly expressed genes account for the 16.6% of the total counts across all cells.

### 5.3.3. Single cell RNAseq data analysis

PCA and t-Distributed Stochastic Neighbour Embedding (t-SNE) analyses were preformed to reduce data dimensionality, determine the degree of heterogeneity in the dataset and

identify potential subpopulations. PCA is a linear dimensionality reduction method, whereas t-SNE is a non-linear method, both are widely used for single cell RNA analyses. The gene expression data from 93 high quality cells was used for these analyses. The PCA results indicated that the PC1 corresponds to 10% of the variance among cells and PC2 to 2.5%. Neither component shows segregation of cells into obvious clusters among the population (Figure 5.7).

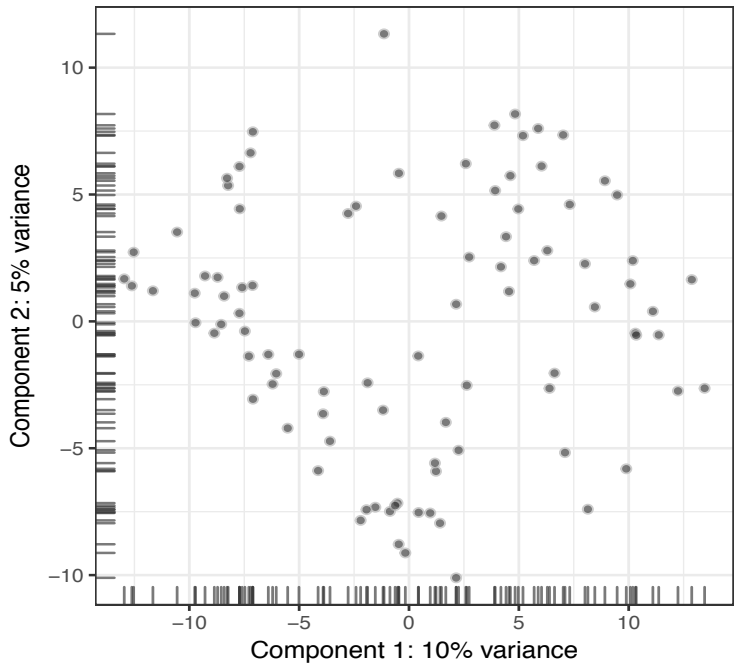



**Figure 5.7. Principal component analyses of the single cell RNAseq data.** PC1 is show in the X-axis and indicates the 10% of variance among cells. The PC2 is show in the Y-axis and represents 2.5% of variance. Each cell is represented as a dot in the plot.

The t-SNE analyses shown the multidimensional data mapped to 2 dimensions, with no obvious clustering among the cells in the population (Figure 5.8).

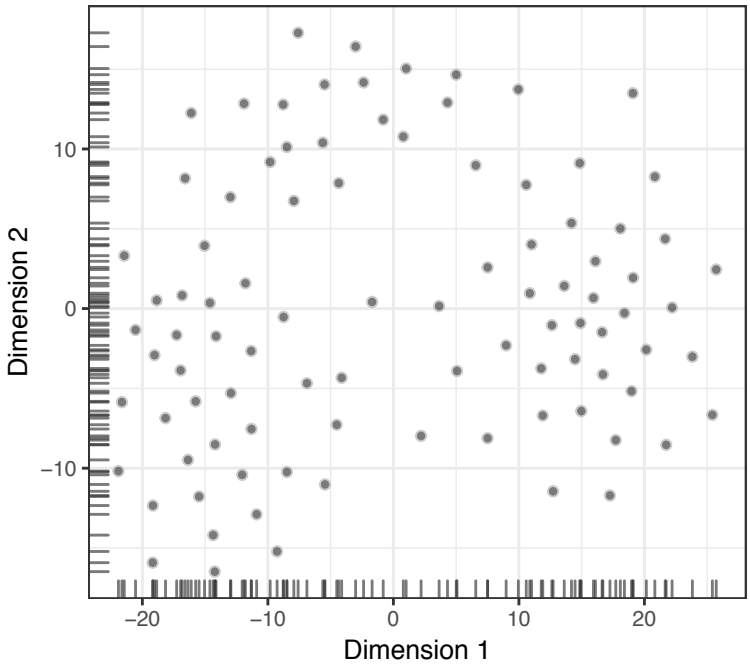



**Figure 5.8. t-SNE analyses of the single cell RNAseq data.** Dimension 1 is show in X-axis and dimension 2 in Y-axis. The cells are represented as dots in the plot

Cells in different cell cycle phases generate different levels of mRNA transcripts and have different gene expression profiles. Hence, we analyzed the expression of genes that regulate the cell cycle to identify potential differential expression across the cells. The results indicate that expression of the genes CCNB2, CDK1, CCNB1, CDC6, CDK2 and GMNN is bimodally distributed (Figure 5.9).

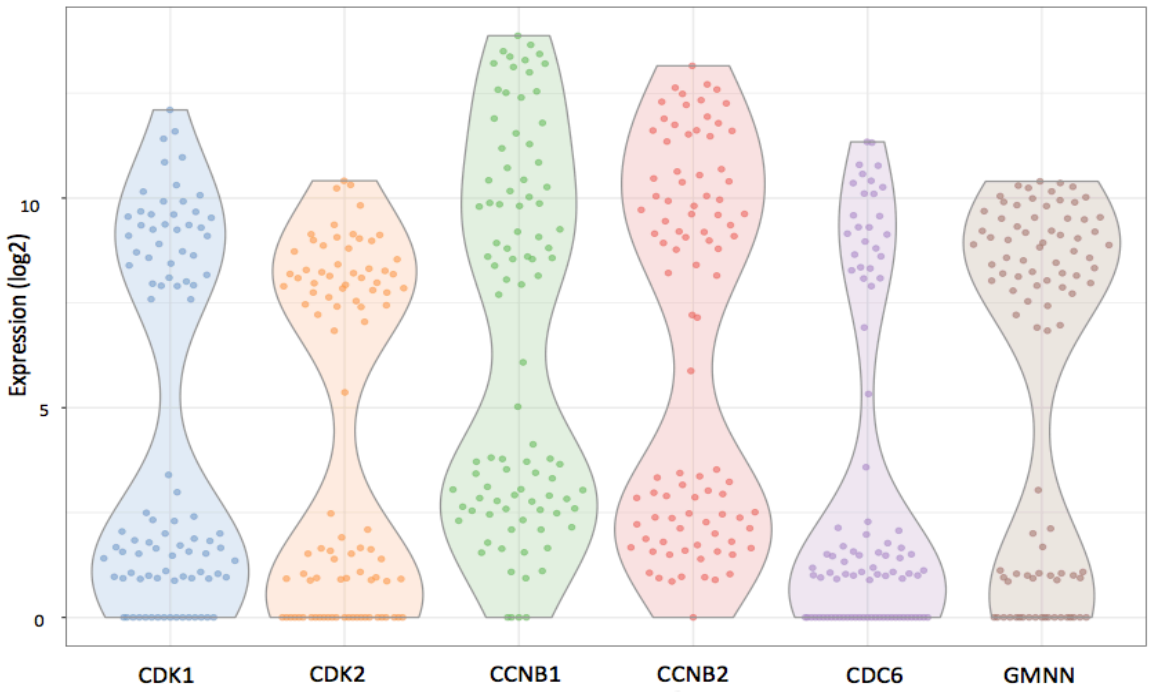



**Figure 5.9. Expression of cell cycle genes in individual cells.** Violin plots show the expression of the cell cycle regulators CCNB2, CDK1, CCNB1, CDC6, CDK2 and GMNN represented with different colors. Individual cells are shown as dots in the plots. The Y-axis shows the gene expression levels (log2 counts adjusted for the library and gene size).

A PCA showing the expression levels of CDK1, CDK2, CCNB1, CCNB2, CDC6 and GMNN cell cycle regulators was generated to determine if the heterogeneous expression of these genes across the cell population is related the to degree of separation of cells across PC1 or PC2 (Figures 5.10, 5.11, 5.12, 5.13, 5.14 and 5.15.). The results suggest that the majority of cells with high expression values of these genes are slightly segregated from the cells expressing lower values of the genes across PC1.

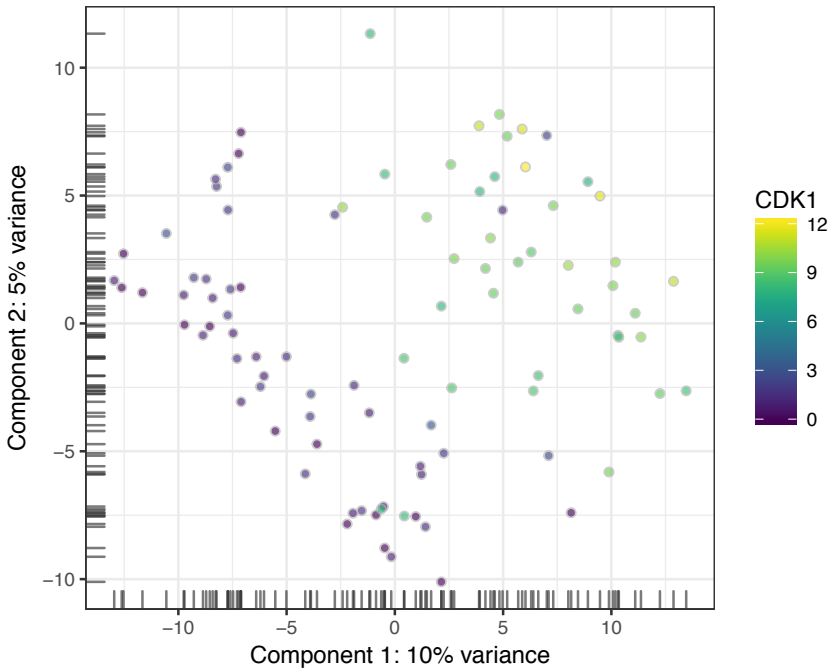



**Figure 5.10. PCA depicting expression levels of CDK1.** The color key shows CDK1 expression, where purple indicates low expression and yellow high expression. Individual cells are represented as dots in the plot. PC1 is show in the X-axis and PC2 Y-axis.

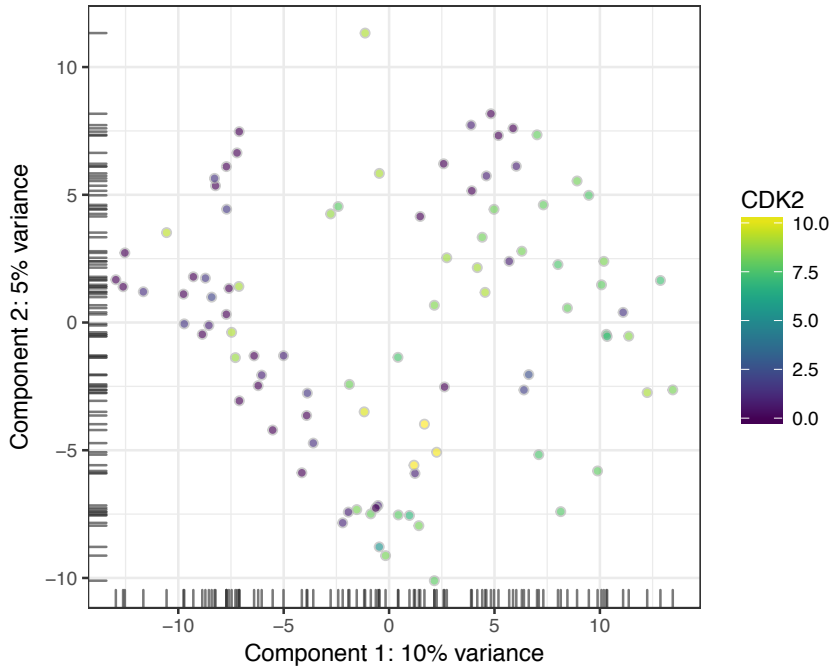



**Figure 5.11. PCA depicting expression levels of CDK2.** The color key shows CDK2 expression, where purple indicates low expression and yellow high expression. Individual cells are represented as dots in the plot

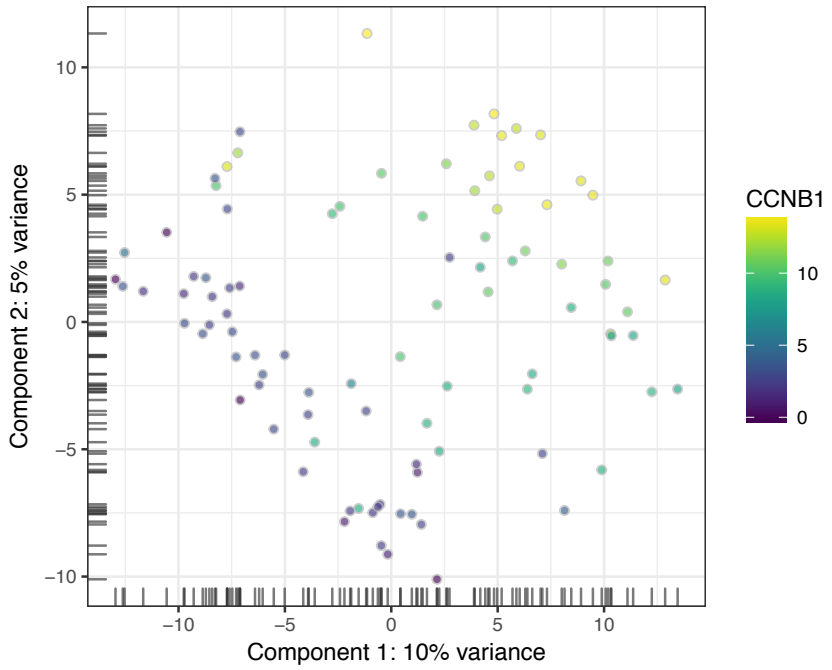



**Figure 5.12. PCA depicting expression levels of CCNB1.** The color key shows CCNB1 expression, where purple indicates low expression and yellow high expression. Individual cells are represented as dots in the plot.

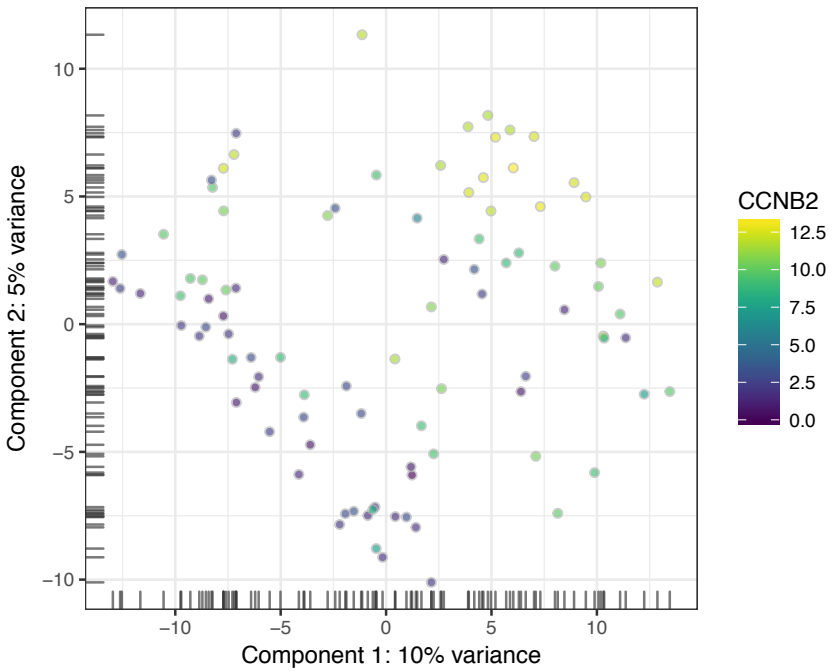



**Figure 5.13. PCA depicting expression levels of CCNB2.** The color key shows CCNB2 expression, where purple indicates low expression and yellow high expression. Individual cells are represented as dots in the plot.

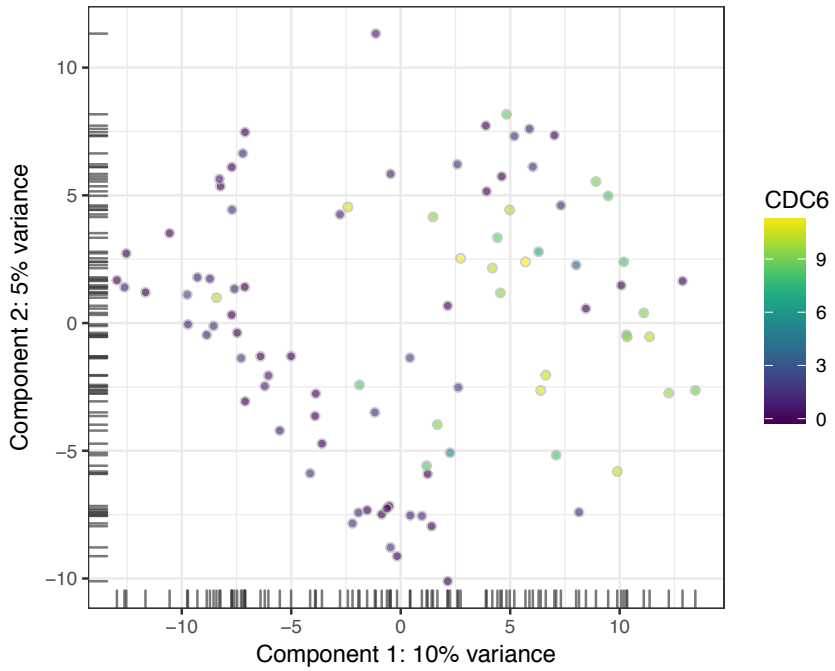



**Figure 5.14. PCA depicting expression levels of CDC6.** The color key shows CDC6 expression, where purple indicates low expression and yellow high expression. Individual cells are represented as dots in the plot.

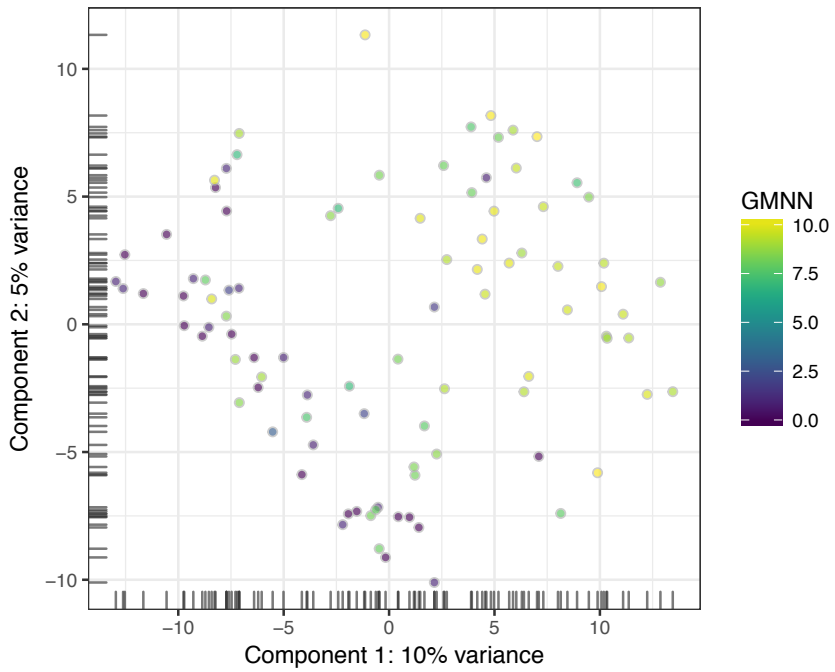



**Figure 5.15. PCA depicting expression levels of GMNN.** The color key shows GMNN expression, where purple indicates low expression and yellow high expression. Individual cells are represented as dots in the plot.

Expression of neuroectodermal markers was analyzed to determine if all cells of the population at d8 express these genes. The results indicated that the majority of cells express high levels of PAX6 and SOX11. However, a number of cells show an absence of SOX11 expression, indicating a bimodal distribution. The expression of ZEB2 and ZNF521 is detected in a range from 0 to high expression levels across the population. The expression of the genes NR2F2 and NESTIN is bimodally distributed, where the majority of cells show high expression levels and another cluster shows low or no expression (Figure 5.16).

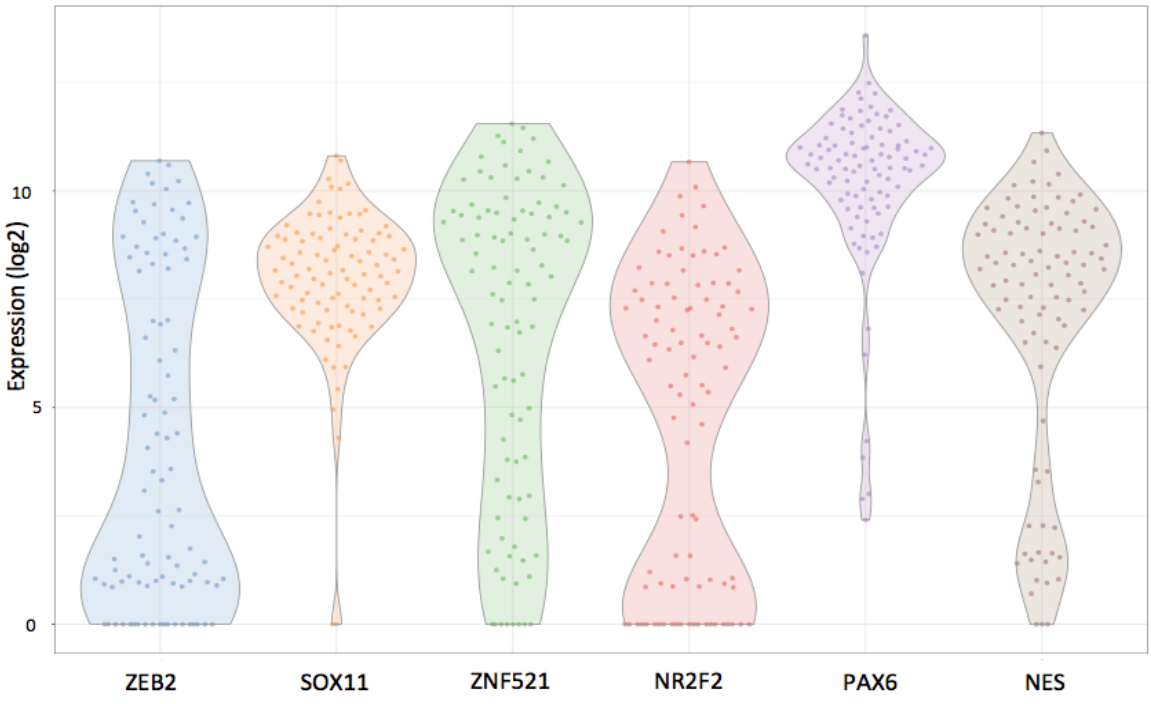



**Figure 5.16. Expression of neuroectodermal markers in individual cells.** Violin plots show expression of the neuroectodermal markers ZEB2, SOX11, ZNF521, NR2F2, PAX6 and NESTIN, represented with different colors. Individual cells are shown as dots in the plots. The Y-axis shows the gene expression levels (log2 counts adjusted for the library and gene size).

Similarly, expression of neuroectodermal precursors was analyzed in individual cells. The results indicated that expression of IRX3, ZIC3 and FOXD4 is not detected in the majority of the cells. Expression of the genes ZIC2 and SOX3 is detected in a range from zero to high, whereas expression of GMNN is bimodally distributed (Figure 5.17).

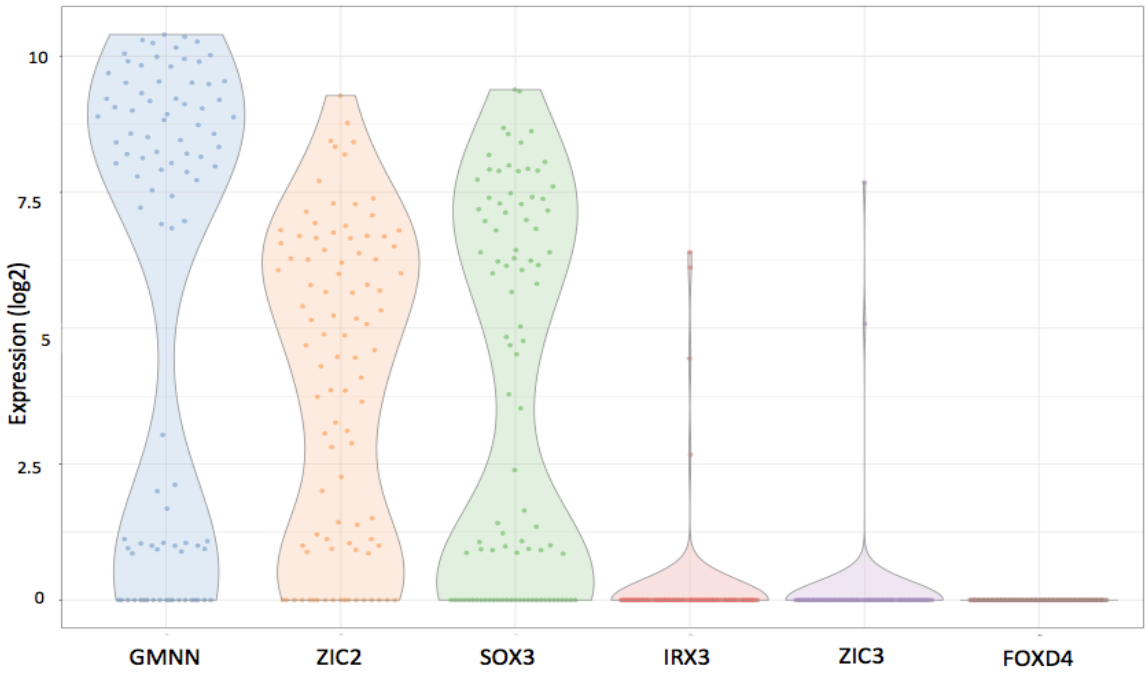



**Figure 5.17. Expression of neuroectodermal precursor markers in individual cells.** Violin plots show expression of the neuroectodermal precursor markers GMNN, ZIC2, SOX3, IRX3, ZIC3 and FOXD4 represented with different colors. Individual cells are shown as dots in the plots. The Y-axis shows the gene expression levels (log2 counts adjusted for the library and gene size).

Analyses of expression of pluripotency markers in individual cells indicated that the genes NANOG, OCT4 and ZFP42 are not expressed or detected in low levels across all cells in the population, whereas no expression was detected for UTF1. MYC expression is bimodally distributed among the cells in the population. The expression of the genes ZIC2

and SOX3 is detected in a range from zero to high expression, whereas expression of GMNN is bimodally distributed (Figure 5.18).

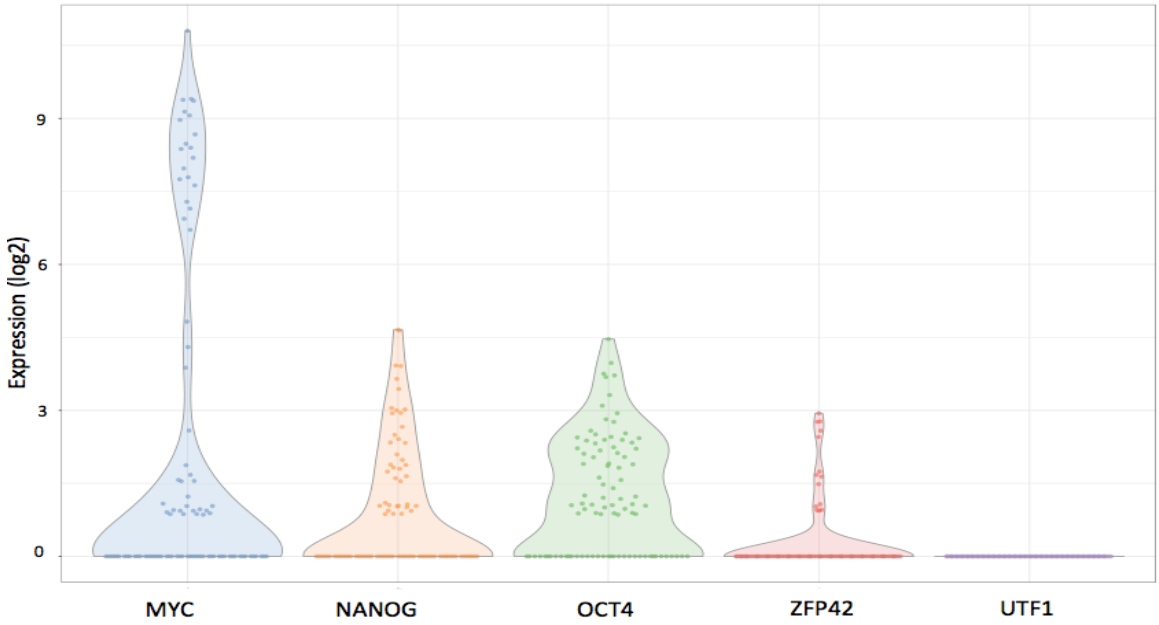



**Figure 5.18. Expression of pluripotency markers in individual cells.** Violin plots show expression of the pluripotency markers MYC, NANOG, OCT4, ZFP42 and UTF1 represented with different colors. Individual cells are shown as dots in the plots. The Y-axis shows the gene expression levels (log2 counts adjusted for the library and gene size).

Additionally, expression of genes most highly up-regulated from time points d6 to d8 in the bulk cells analyses, was analyzed to determine if expression of these genes increased in all cells of the population. The results show that expression of DMRTA2, ABCG2, ALDH1A1 and CCK is not detected in the majority of cells and that high expression levels are detected only in a minority of cells. The gene TNFSF15 is expressed at low levels or not at all, whereas WNT2B expression is bimodally distributed (Figure 5.19).

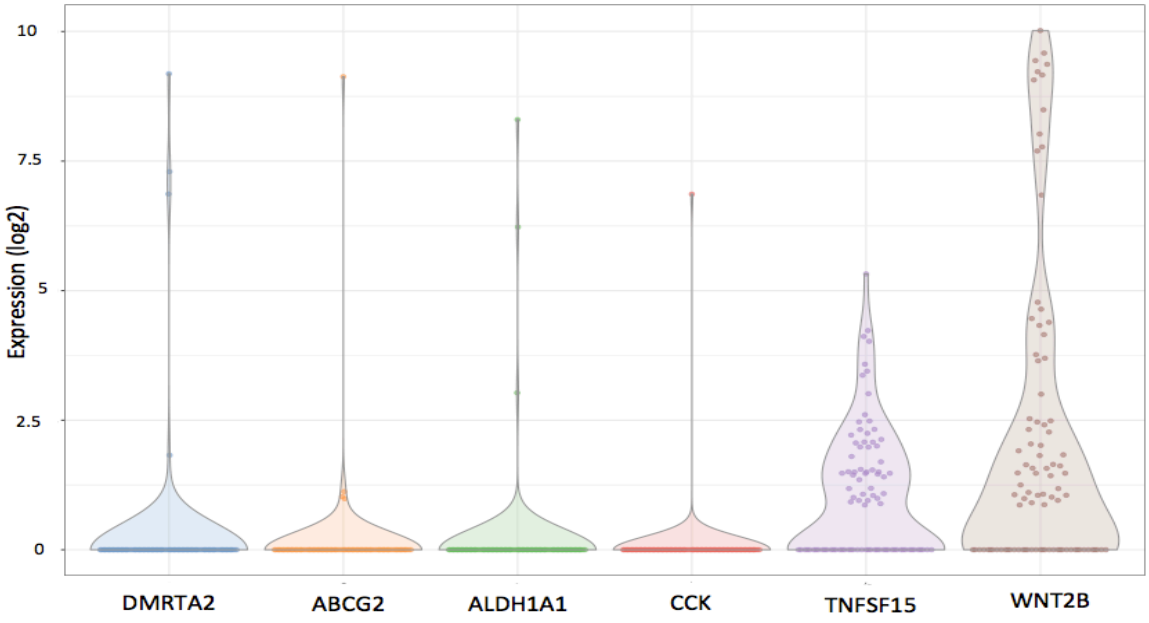



**Figure 5.19. Individual cells expression of the genes most highly up-regulated from time point d6 to d8 in bulk analyses.** Violin plots show expression of DMRTA2, ABDG2, ALDH1A1, CCK, TNFSF15 and WNT2B the genes most highly up-regulated from d6 to d8 in bulk analyses. The Y-axis shows the gene expression levels (log2 counts adjusted for the library and gene size).

## 5.4. Discussion

Single cell RNAseq studies have shown that seemingly homogeneous cell populations comprise cells with variability in gene expression, protein abundances and potency to differentiate into specific phenotypes (Shalek *et al*., 2013). Cells captured at the same time might include subpopulations with intermediate differentiation states or subpopulations of cells fated to distinct phenotypes, which can only be identified by single cell studies. Hence, we performed single cell RNAseq analysis to assess transcriptional heterogeneity in iPSCs derived NESCs, which might be masked in the bulk cells analyses.

Dimensionality reduction techniques such as PCA and t-SNE enable intuitive visualization of similarities and differences among individual cells within a population (McCarthy *et al.*, 2017). The results indicate that after 8 days of neural induction the NESC population is highly homogeneous. Variance across PC1, which explains 10% of the variance across cells, appears to be driven by differential expression of genes that regulate cell cycle progression. The cell cycle regulators CCNB2, CDK1, CCNB1, CDC6, CDK2 and GMNN are either highly expressed or lowly expressed forming 2 adjacent clusters along PC1. Possibly, cells comprising the NESC population are progressing through different phases of the cell cycle, which leads to bimodal distribution of expression levels of these genes. Global transcriptional variability during cell cycle occurs because genes are activated or inhibited in a specific order during different cell cycle phases. For instance, during mitosis global transcription is interrupted leading to a decrease of mRNA transcript abundances. The expression of the majority of genes is re-activated during early G1 depicting a peak of expression (Fukuoka *et al*., 2012; Hsiung *et al*., 2016). As the cell cycle proceeds, three waves of expression during G1/S, G2/M and M/G1 phases regulate the molecular events that drive progression to the next phase (Bertoli *et al*., 2013). During S phase, DNA is replicated, which might lead to the production of double the amount of mRNA transcripts (Padovan- Merhar *et al*., 2015). Hence, variable gene product abundances are produced during particular stages of the cell cycle, and this seems to represent the major partition of the NESC population. Correlation results suggest that the expression of single cells at d8 after neural induction closely resemble that of the bulk transcriptome captured at the same stage. Single cell data depicted high homogeneity among the cell population, aside from that represented by changes in cell cycle phase, thus bulk data can be interpreted without the confound of population heterogeneity.

Similarly, analyses of NE markers revealed variability in gene expression patterns across NESC at d8. Immunofluorescence analyses (section 3.3.3) of ZEB2, ZNF521, NR2F2, PAX6 and NESTIN show that the majority of cells are expressing high levels of these proteins. However, a reduced number of cells in the population are either expressing low levels or not expressing these factors, suggesting that the heterogeneous expression of these specific genes among the population is reflected at a protein level. However, a percentage of cells with low intensity staining levels might have been incorrectly count as negative, due to the high intensity threshold set during the data analyses to discriminate positively stained cells from non-specific staining. Additionally, transcriptional burst and fluctuation of different biochemical processes including transcription, translation kinetics, mRNA and protein stability and degradation rates might contribute to the differential expression levels observed trough immunofluorescence assays. Alternatively, the population might include cells at different developmental stages or acquiring different phenotypes, thus expressing different levels of these genes.

The bimodal distribution of the genes expression at an mRNA levels, determined from single cells RNA sequencing, could be related to variability of the transcriptional bursting frequency. For instance, a recent study demonstrated that the transcriptional burst frequency of a set of genes bound to the polycomb repressive complex (PRC) was lower compared with active genes not bound by PRC. Genes with lower frequency of variation including Klf4, Tbx3 and Lefty2 show bimodal distribution, whereas genes with higher bust frequency such us Oct4 and Sox2 depicted unimodal distribution in mESCs (Kar *et al*., 2017). Possibly, bimodally distributed NE markers have low frequency transcriptional bursts and cells were captured and cluster together during a long “on” activation period or “off” transcriptional inactive periods. It is also feasible that the cell population comprises two sub-populations primed to acquire different fate or transitioning between distinct

neural differentiation stages. Additionally, the majority of single cell RNAseq data presents high frequency of dropout events, where no gene counts are observed in a cell where the gene is actually expressed. Dropout events occur when low starting amounts of mRNA transcripts are "lost" during library preparation and are consequently excluded from sequencing data (Kharchenko *et al*., 2014; McCarthy *et al*., 2017; Vallejos *et al*., 2017). Therefore, high dropout events are more frequent in, but not exclusive to, genes with low expression. Dropout events could also contribute to the bimodal expression of the NE markers observed in the d8 NESCs. However, immunofluorescence assays show that SOX11, NR2F2 and NESTIN proteins are not detected in all cells of the population, suggesting that heterogeneous expression of these genes is not entirely caused by technical noise. Similarly, the NE markers ZEB2 and ZNF521 show variable expression with continuous changes from high to low transcript amounts. Possibly, these genes depict different transcriptional bursts sizes among the cell population. Additionally, expression of these genes might be activated gradually during differentiation, leading to variable expression in cells transitioning through different developmental stages. However, the global gene expression patterns across the population are homogeneous, indicating that cells are not forming defined clusters that might drive differences in cell fate decisions and suggesting that cells are progressing through similar developmental trajectories.

Single cell analyses show that the NE precursor factors IRX3, ZIC3 and FOXD4 and the pluripotency markers NANOG, OCT4, ZFP42 and UTF1 are not expressed in the majority of cells at d8 in human iPSC derived NESCs, as observed in the bulk transcriptome analysis at the same stage. Additionally, the NE progenitors SOX3 and ZIC2 are expressed from high to low levels at d8 across single cells, whereas GMNN and the pluripotency marker MYC showed a bimodal distribution, which might result from differences in the transcriptional burst shape, cell cycle progression or developmental stages. In bulk cell

analyses GMNN, ZIC2, SOX3 are highly expressed at d8, whereas MYC is expressed at low levels.

The RNAseq data from bulk analyses on the cell line iPSC1 show that the genes DMRTA2, ABDG2, ALDH1A1, CCK, TNFSF15 and WNT2B are the most highly up-regulated from d6 to d8. Interestingly, single cell data analyses at d8 from the same cell line show that these transcripts are either not detected or expressed in low levels in the majority of individual cells. The results indicated that the gene expression data from single cells and bulk cell is highly co-related. Thus, the expression of these genes at a single cell level might reflect the gene expression of the individual cells of the population evaluated trough bulk analyses. The results suggest that the averaged gene expression depicted in bulk analyses might mask the transcriptional events occurring in subpopulations of cells. GRN modelling is largely based on bulk expression data, assuming that differentially expressed genes are activated or inhibited in the majority of cells. Hence, these results highlight the importance of determining that the nodes of a GRN are expressed in the same cell, to validate the inferred regulatory connections and refine the GRN. Therefore, single cell RNA sequencing analyses during different time points of NESC differentiation would enable determine if the GRN nodes interacting with no time delay are expressed in the same cell, allowing to greatly refine the GRN interactions. However, due to the elevated cost of single cell RNA sequence assays, these analyses were not performed this study.

## 5.5. Limitations

The development of bioinformatics tools to analyse single cell datasets enables determination of level of gene expression at a specific time point in a single cell, but this level of resolution comes with an increase in technical noise (McCarthy *et al*., 2017). Thus,

quality control and normalization are critical to estimate and reduce the technical noise of the dataset. However, tools to identify and remove dropout events are limited. Hence, the major technological challenge of single cell RNAseq analyses is to improve sensitivity and accuracy to separate technical noise from intrinsic heterogeneous expression. These techniques are being developed and will vastly improve single cell clustering and data visualization analyses. Similarly, an important source of biological noise is the cell cycle, which generates variability in the gene expression as cells progressing through different cell cycle phases, which might obscure other meaningful biological signals in the data. However, current methods to remove the cell cycle effect generate a significant risk of removing other components of interest (Barron & Li, 2016). Additionally, cell isolation, reverse transcription, cDNA synthesis, library preparation and sequencing are expensive techniques, which restricts the number of cells, time points and treatments that can be feasibly analyzed.

## 5.6. Conclusions

After 8 days of neural induction the NESC population is highly homogeneous with no evidence of clustering. The principal source of variability among cells appears to be related to cell cycle progression. Additionally, we show that even if the expression levels of the bulk analyses are highly correlated with the averaged quantification of single cells transcripts, bulk analyses mask substantial differences occurring at single cell level. These differences should be considered for inferring relations among genes in a GRN to confirm that the nodes are expressed in the same cell.

# CHAPTER 6. General Discussion

The scope of this study was to discover the gene expression patterns underlying neural induction in human iPSCs and ESCs; and to identify TFs that orchestrate acquisition of NESC fate. Most emphasis in PSC studies has been on analyzing the differentiation of neural progenitors to specific neuronal phenotypes focusing on neuronal subtypes associated with developmental disorders. Neural induction is the first step of CNS development and there is abundant evidence to suggest that fate specification occurs during this stage. Hence, the gene expression and regulatory interactions generated during NE differentiation control the progressive restrictions in cell fate that drive cell competence towards neural and neuronal differentiation. However, the molecular events that cause human ESC/iPSCs to become neural progenitors have not been deeply studied (Neely *et al*., 2012; Moody *et al*., 2013; Kamiya *et al*., 2014). Knowledge of the initial gene expression patterns that drive neural induction in iPSCs greatly improves our knowledge about this critical stage of development in humans and identifies the first molecular events that drive both normal neurodevelopment and aberrant differentiation leading to neurodevelopmental disorders.

For this study, neural differentiation was induced in 3 iPSC and 1 ESC line by recapitulating the developmental cues identified from animal models. Mouse NE differentiation initiates with inhibition of a BMP and TGFβ from the "node" to the dorsal region of the ectoderm (Klein & Moody, 2015). Similarly, neural induction in ESC/iPSC is initiated by inhibiting SMAD signaling trough the addition of small molecule inhibitors (2i). The resulting data indicated that ESCs and iPSCs lines with different genetic background and iPSC lines reprogramed with distinct methods generated similar transcript levels and gene expression patterns during neural induction (Figure 3.2). These results

demonstrate that the protocol used in this study is highly reproducible for generating NESCs and that iPSC constitute a remarkable tool to recapitulate *in vitro* molecular mechanisms that drive pluripotent stem cells toward a neural fate.

After 2i induction ESC/iPSC, which represent Epi cells during human development, maintain transcription of pluripotency markers and genes that regulate cell biological functions necessary for pluripotency until d2 after induction (Figures 3.12 and 3.13). Following down-regulation of pluripotency genes, thousands of genes are activated including NE markers, genes related with other stages of the CNS differentiation and other genes that have not been previously related with the CNS development and could be used as novel NE markers (Figures 3.10 and 3.11). Activation of this large number of genes is necessary to initiate and maintain NESCs differentiation, inhibit the pluripotency program and inhibit pathways that drive differentiation towards other phenotypes. The neural induction program comprises genes that are up-regulated from d2 to d8 and also genes that have peaks of expression during different time points during neural induction (Figure 3.9). GO analyses of genes associated with specific expression profiles, indicated that these gene modules regulate different cell molecular functions during neural induction that are necessary to promote the transition from pluripotency to neural fate (Figure 3.14). These gene modules are additionally required during other developmental stages such as forebrain development and synapse formation. Among these genes GMNN, and ZIC2 have a peak of expression at d2, SOX11 is up-regulated from d3 and SOX3 is highly expressed at d8 (Figure 3.4). In Xenopus, FoxD4 induces the expression of Gmnn and Zic2 and later the expression of Sox11 (Lee *et al*., 2014). In mouse, Foxd4 shows a peak of expression in NESC and might regulate nueral fate acquisition by inhibiting pluripotency markers and promoting the expression of Gmnn and Zic2, similarly than in Xenopus (Gaur *et al*., 2016; Sherman *et al*., 2017). Similarly, Gmm promotes neural differentiation in mouse by

promoting histone acetylation in neural genes and by activating neural development regulators in association with Zic1 (Lim *et al*., 2011; Sankar *et al*., 2016). The study results suggest that the sequential activation of these genes expression and their function is conserved across species during embryonic development. However, in humans these markers are likely to be induced by different factors since FOXD4 is not expressed during NESCs induction. In Xenopus, transition between NE progenitors and NE cells depends on inhibition of Sox3, Sox2 and Sox11 family members and up-regulation of Zic and Irx factors that promote expression bHLH proneural factors in neural progenitors (Lee *et al*., 2014). Studies in mouse and chick embryos suggest that Sox1, Sox2 and Sox3 factors are expressed in most neural precursors and are necessary to maintain the cells in a progenitor state, whereas Sox11 is activated later in neural progenitors and differentiating neurons (Bergsland *et al*., 2011; Lee et al., 2014). The results show that IRX1, IRX3 and ZIC1 genes were not detected during neural induction in human PSCs, whereas IRX2 and ZIC3 have a peak of expression at d2 and are subsequently down-regulated. Additionally, SOX2, SOX3 and SOX11 are activated during different stages of NE induction and remain active in NESCs at day 8 (Figure 3.4). The results indicate that the sequential expression of IRX, ZIC and SOX family factors are activated in a different manner than in Xenopus, suggesting specie-specific differences and highlighting the necessity to use human cells to analyze human neural development. However, similarly than in mouse and Xenopus, the TF SOX1, SOX2, SOX3 and SOX11 are actively expressed in NESC. Factors required for both pluripotency and neural differentiation such us SOX2 and NESTIN show high levels of expression during the course of neural induction, validating their role in the maintenance of pluripotency and promotion of neural differentiation in humans (Figure 3.4).

Further culture of NESCs leads to formation of rosettes around d12, whose morphology and cell composition resembles that of the neural tube. The gene expression data show that

at d8 cells do not express polarization markers necessary for the formation of rosettes (Figure 3.21). Hence, activation of the genes necessary for polarization and rosette formation are possibly activated subsequent to NESC formation. However, we report for the first time that several rosette and RG cell markers and neuronal sub-types markers are precociously activated during early (d2) neural induction.

During gastrulation in the Xenopus embryo, the area between the NE and non-NE differentiates into the NBZ at the expense of the neural crest due to inhibition of WNT signaling. The NBZ gives rise to the PPE, of which the most anterior region will form the olfactory placode. GnRH neurons differentiate from the olfactory placode and migrate to the hypothalamus (Moody & LaMantia, 2015). In humans, the regulatory events that drive differentiation of GnRH neurons are largely unknown. For instance, only a single alternative study has reported a protocol to generate GnRH neurons from ESC (Lund *et al.*, 2016). In this study we aimed to identify the initial molecular signals that drive NESCs to acquire GnRH fate. Colleagues in Price laboratory have demonstrated that the addition of WNT inhibitors during neural induction leads to the generation of positive GnRH neurons after approximately 50 days (Kathuria *et al*., 2017). We identify a set of 44 genes differentially up-regulated exclusively in NESCs fated to become GnRH neurons, which might play a role in the adquisition of GnRH neuronal phenotype (Figures 3.26 and 3.27). Additionally, 103 genes were up-regulated exclusively in NESCs destined to differentiate in cortical neurons. Inhibition of these genes might also be necessary to initiate the GnRH induction program.

## 6.1. Gene regulatory networks

The genome encodes thousands of genes, whose expression continually changes during different stages of development. Expression of these genes is tightly regulated to generate specific amounts of gene products necessary to control the cellular functions required for each developmental stage (Karlebach & Shamir, 2008). GRNs govern the expression of genes and the product abundances at specific time points and can be modeled from time series transcriptome data by mathematical and statistical analyses to identify regulatory gene interactions during induction. We show the first GRN modeled during NESCs differentiation, which comprises 9 main hubs with high connectedness (JUN, MYC, FOS, PAX6, SP3, CDC6, SMAD2, HDAC6, and LEF1) and incorporates with time delay co-relations (Figure 4.4). The GRN connectivity (topology) continuously changes in response to regulatory signals that control gene expression, generating different expression patterns during differentiation. The genes and gene modules activated during neural differentiation regulate cell biological functions necessary for neural differentiation and pluripotency inhibition. Among these modules, the MYC hub regulates different molecular functions including cell proliferation, CDC6 regulates cell cycle progression and PAX6 regulates the CNS development among others (Figure 4.7). It has been suggested that PAX6 promotes neural differentiation by inhibiting pluripotency markers such as NANOG and OCT4 and possibly by activating TFs required for NE differentiation (Zhang *et al*., 2010; Blake & Ziman, 2014). We identify 8 genes within the PAX6 module, whose activation at d2 potentially initiates the consequent steps that drive initiation of neural induction and inhibition of pluripotency (Figure 4.3).

## 6.3. Single cell analyses

NESCs are positive for known NE markers such us PAX6, SOX11 and ZEB2. Single cell RNAseq analyses enabled identification of expression variability across NESCs at d8. The data indicated that the NESC population is highly homogeneous (Figures 5.7 and 5.8). The largest percentage of variance (10%) appears to relate to the cell cycle phase in which cells were captured (Figures 5.10 to 5.15). Nevertheless, genes did show heterogeneous expression, which is also reflected at a protein level (Figures 5.16 to 5.19 and 3.7). These results highlight the importance of single cell analyses in GRN modeling to determine that those genes proposed to interact within the network are expressed in the same cell. Heterogeneous expression across the cell population might be related to variability in the transcriptional burst shape, the cell cycle phase in which cells were captured, dropout events, variability in the developmental stage across individual cells and due to the generation of 2 sub-populations primed to acquire different fate. However, the global gene expression patterns suggest that cells are progressing through similar developmental states.

## 6.3. Future experiments

Transcriptome analyses during induction enabled identification of thousands of genes differentially activated and inhibited during different stages of neural induction. Identification of these genes permits determination of the potential roles and pathways of TFs of interest by inhibiting their expression trough different techniques such us shRNA gene knockdown, small molecule inhibitors or CRISPR-CAS9, and analysing the subsequent effect of inhibition on gene expression and neural induction. Likewise, it would be interesting to generate transcriptome data under the same conditions of this study from iPSC lines with penetrant mutations associated with neurodevelopmental disorders risk.

For instance, we have a SHANK3 mutated (highest monogenic cause of autistic spectrum disorders) cell line isogenic to the ESC line used in this study. Comparing the expression patterns of the mutated and control lines would enable identification of potential early molecular events that distinguish normal from aberrant neural induction. Additionally, it would be interesting to analyze neural induction in 3D cultures, which provide spatial organization that promotes exposure to the biochemical signals from the microenvironment, extracellular matrix and neighbouring cells thereby offering a more complete recapitulation of *in vivo* development. However, 3D cultures are less reproducible since their microenvironment can vary depending on several factors including the scaffold material mechanics, topography and biochemistry and the delivery and transportation of soluble factors and oxygen (Carletti *et al*., 2011; Hollister, 2013). Cells among the 3D culture are exposed to different spatial and physical constraints and biochemical cues, leading to variability in cell proliferation and differentiation and generating heterogeneous cell populations (Edmondson *et al*., 2014). Hence, it is not suitable to use bulk cells analyses to analyse the transcriptome of cells comprised in 3D cultures and it can be difficult and expensive to preform immunofluorescence analyses and single cell isolation depending on the scaffold material and topography (Antoni *et al*., 2015). Additionally, modelling a GRN from transcriptome data from 3D culture isolated cells isolated remains challenging due to the low read coverage of single cell RNAseq analyses. Nevertheless, as these technologies improve they will offer an exciting alternative to 2D monocellular cultures.

Similarly, it would be important to continue the GRN refinement by testing the GRN predictions by perturbing the network. CRISPRi technologies could be used to repress expression of PAX6 and OTX2 and CRISPa to activate expression of MYC in iPSC derived NESCs and quantitatively assess the response of the perturbation in the expression of their

interacting gene targets. CRISPRi and CRISPRa offer the additional benefit of being well suited to modulation of expression of several genes simultaneously. Similarly, connectivity between TFs and gene targets inferred in the GRN could also be validated by ChIP-seq analyses to determine if their regulatory interactions are direct, and potentially to identify novel interactions of the TFs with other genes that are differentially expressed during differentiation. Additionally, including constitutive expressed genes and less connected genes that might interact with the network modules may improve the GRN model.

High-throughput RNAseq data enables modelling a GRN that provides a global view of the gene regulatory interactions and potential cellular biological functions underlying neural induction. However, RNA transcription is also regulated by epigenetic modifications. Combining, transcriptome and epigenetic data sets to model a GRN would enable a deeper insight into the molecular events that drive NESCs development. Thus, the GRN model would significantly improve by the incorporation of epigenetic regulatory interactions. Post-translational modifications of histones tails and DNA methylated cytosines can be detected by using specific antibodies through ChIP-seq analyses. Additionally, Assay for Transposase-Accessible Chromatin with high-throughput sequencing (ATAC-seq) method could be used to identify open chromatin regions at specific time points during NESCs induction. These data would enable determination of the epigenetic status during neural induction. This may identify distinct cellular states prior to changes in the transcriptome. Alternatively, such studies might illuminate the epigenetic mechanisms that regulate the expression of specific genes among the GRN and determine the sequential epigenetic changes necessary for neural induction. The function of individual genomic elements identified can be screened by using modified CRISPR to deliver specific chromatin modifying activities (Enriquez, 2016).

Additionally, single cell analyses demonstrated that averaging the gene expression values of cell populations could obscure variability in the expression of individual cells across the population. However due to the elevated cost of the RNA sequencing assays we did not performed these analyses during different stages of differentiation. Hence, further experiments should include single cell RNAseq during the 7-time points of NESCs differentiation to verify that the genes interacting in the GRN are expressed in the same cell. It would be interesting to analyse the gene expression profile of NESCs induced with 2i-WNT inhibition protocol at d8 to determine if the 44 differentially up-regulated genes compared with 2i treatment are expressed in the same cells or generate sub-populations that might bias the cell fate. Additionally, inhibition of the expression of the 5 TFs differentially up-regulated in NESCs fated to differentiate to GnRH neurons might enable determination of the mechanism by which these factors promote GnRH fate acquisition.

## 6.4. Conclusion

In this thesis I report for the first time a complete overview of the gene expression patterns underlying neural induction of human iPSC/ESCs. Identification of genes and pathways underlying iPSC transition from pluripotency to neural competence would enable identification of the molecular pathology of neurodevelopmental disorders caused by known mutations and provide an insight into the gene regulatory events that direct normal and aberrant human neurodevelopment. Inhibition of the pluripotency program occurs during the initial 2 days of 2i neural induction, followed by activation of thousands of genes that orchestrate NE fate acquisition. Among these genes several NE precursors, NE and forebrain markers were identified, validating previous findings from animal and ESC studies. Additionally, we report activation of genes that have not been previously related with CNS development and that could be used as novel NE markers. The potential

regulatory interactions between genes activated during NESC induction were captured in a GRN, which enables clear visualization of the potential causal relations between genes at different stages of induction. We identified GRN modules and with similar expression profiles, which potentially regulate the biological processes required for the transition from pluripotency to NE fate. Additionally, we report that WNT inhibition during neural induction leads to the specific activation of 44 genes, which may prime NESCs to differentiate into GnRH neurons. Down regulation of 103 genes activated during 2i (cortical) differentiation might also be necessary for cells to acquire GnRH phenotype. Additionally, we demonstrated that the differentiation protocol generates a highly homogenous population of NESC at d8. However, a number of genes depicted heterogeneous expression, which must be considered to validate that GRN inferred interactions occur between genes expressed in the same cell. Hence, with this study we provide for the first time a complete systematic description of the genes that orchestrate neural induction from human ESC/iPSCs. This provides valuable insight into mechanisms that may be used to control the fate specification of iPSC derived neuroectodermal stem cells and also cast light on the biological processes that might underwrite the initial stages of human neural development *in vivo*.

# REFERENCES


Abranches, E., Silva, M., Pradier, L., Schulz, H., Hummel, O., Henrique, D., & Bekman, E. (2009). Neural Differentiation of Embryonic Stem Cells In Vitro: A Road Map to Neurogenesis in the Embryo. *PLoS ONE*, *4*(7), e6286–14.

Adikusuma, F., Pederick, D., McAninch, D., Hughes, J., & Thomas, P. (2017). Functional Equivalence of the SOX2 and SOX3 Transcription Factors in the Developing Mouse Brain and Testes. *Genetics*, 117.202549–27.

Aigner, K., Dampier, B., Descovich, L., Mikula, M., Sultan, A., Schreiber, M., et al. (2007). The transcription factor ZEB1 (δEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene*, *26*(49), 6979–6988.

Andoniadou, C. L., & Martinez-Barbera, J. P. (2013). Developmental mechanisms directing early anterior forebrain specification in vertebrates. *Cellular and Molecular Life Sciences*, *70*(20), 3739–3752.

Antara De. (2011). Wnt/$Ca^{2+}$ signaling pathway: a brief overview. *Acta Biochimica et Biophysica Sinica*, 43(10), 745–756.

Antoni, D., Burckel, H., Josset, E., & Noel G. (2015). Three-Dimensional Cell Culture: A Breakthrough in Vivo. *Int. J. Mol. Sci*, 16, 5517-5527.

Antosova, B., Smolikova, J., Klimova, L., Lachova, J., Bendova, M., Kozmikova, I., et al. (2016). The Gene Regulatory Network of Lens Induction Is Wired through Meis-Dependent Shadow Enhancers of Pax6. *PLOS Genetics*, *12*(12), e1006441–24.

Archer, T. C., Jin, J., & Casey, E. S. (2011). Interaction of Sox1, Sox2, Sox3 and Oct4 during primary neurogenesis. *Developmental Biology*, *350*(2), 429–440.

Ardhanareeswaran, K., Mariani, J., Coppola, G., Abyzov, A., & Vaccarino, F. M. (2017). Human induced pluripotent stem cells for modelling neurodevelopmental disorders. *Nature Publishing Group*, *13*(5), 265–278.

Artus, J., & Chazaud, C. (2014). A close look at the mammalian blastocyst: epiblast and primitive endoderm formation. *Cellular and Molecular Life Sciences*, *71*(17), 3327–3338.

Aruga, J., & Mikoshiba, K. (2011). Role of BMP, FGF, Calcium Signaling, and Zic Proteins in Vertebrate Neuroectodermal Differentiation. *Neurochemical Research*, *36*(7), 1286–1292.

Barron, M., & Li, J. (2017). Identifying and removing the cell- cycle e ect from single-cell RNA- Sequencing data. *Scientific Reports*, 1–10.

Bartman, C. R., Hsu, S. C., Hsiung, C. C. S., Raj, A., & Blobel, G. A. (2016). Enhancer Regulation of Transcriptional Bursting Parameters Revealed by Forced Chromatin Looping. *Molecular Cell*, *62*(2), 237–247.

Bedzhov, I., Bialecka, M., Zielinska, A., Kosalka, J., Antonica, F., Thompson, A. J., et al. (2015). Development of the anterior-posterior axis is a self-organizing process in the absence of maternal cues in the mouse embryo. *Nature Publishing Group*, *25*(12), 1368–1371.

Begum, A. N., Guoynes, C., Cho, J., Hao, J., Lutfy, K., & Hong, Y. (2015). Rapid generation of sub-type, region-specific neurons and neural networks from human pluripotent stem cell-derived neurospheres. *Stem Cell Research*, *15*(3), 731–741.

Ben-Moshe S., Itzkovitz, S. (2016). Bursting through cell cycle. *eLife*, *5*, e14953–3.

Bergsland, M., Ramskold, D., Zaouter, C., Klum, S., Sandberg, R., & Muhr, J. (2011). Sequentially acting Sox transcription factors in neural lineage development. *Genes & Development*, *25*(23), 2453–2464.

Bergström, T., & Forsberg-Nilsson, K. (2012). Neural stem cells: Brain building blocks and beyond. *Upsala Journal of Medical Sciences*, *117*(2), 132–142.

Bertoli, C., Skotheim, J. M., & de Bruin, R. A. M. (2013). Control of cell cycle transcription during G1 and S phases. *Nature Reviews Molecular Cell Biology*, *14*(8), 518–528.

Biswas, S., & Acharyya, S. (2017). Neural model of gene regulatory network: a survey on supportive meta-heuristics. *Theory in Biosciences*, *135*(1), 1–19.

Blais, A., Dynlacht, B. (2005). Constructing transcriptional regulatory networks. *Genes & Development,* 19, 1499–1511.

Blake, J. A., & Ziman, M. R. (2014). Pax genes: regulators of lineage specification and progenitor cell maintenance. *Development*, *141*(4), 737–751.

Brix, J., Zhou, Y., & Luo, Y. (2015). The Epigenetic Reprogramming Roadmap in Generation of iPSCs from Somatic Cells. *Journal of Genetics and Genomics*, *42*(12), 661–670.

Broccoli, V., Colasante, G., Sessa, A., & Rubio, A. (2015). ScienceDirect Histone modifications controlling native and induced neural stem cell identity. *Current Opinion in Genetics & Development*, *34*(C), 95–101.

Brozzi, F., Arcuri, C., Giambanco, I., & Donato, R. (2009). S100B Protein Regulates Astrocyte Shape and Migration via Interaction with Src Kinase. *Journal of Biological Chemistry*, *284*(13), 8797–8811.

Bruce, A. W., & Zernicka-Goetz, M. (2010). Developmental control of the early mammalian embryo: competition among heterogeneous cells that biases cell fate. *Current Opinion in Genetics & Development*, *20*(5), 485–491.

Boroviak, T., & Rashbass, P. (2011). The Apical Polarity Determinant Crumbs 2 Is a Novel Regulator of ESC-Derived Neural Progenitors. *Stem Cells*, *29*(2), 193–205.

Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., et al. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnology*, *33*(2), 155–160.

Carey, L. B., van Dijk, D., Sloot, P. M. A., Kaandorp, J. A., & Segal, E. (2013). Promoter Sequence Determines the Relationship between Expression Level and Noise. *PLoS Biology*, *11*(4), e1001528–15.

Carletti, Carletti E., Motta A., Migliaresi C. (2011). Scaffolds for Tissue Engineering and 3D Cell Culture. In: Haycock J. (eds) 3D Cell Culture. Methods in Molecular Biology (Methods and Protocols), vol 695. Humana Press.

Caronna, E. A., Patterson, E. S., Hummert, P. M., & Kroll, K. L. (2013). Geminin Restrains Mesendodermal Fate Acquisition of Embryonic Stem Cells and is Associated with Antagonism of Wnt Signaling and Enhanced Polycomb-Mediated Repression. *Stem Cells*, *31*(8), 1477–1487.

Castro, D. S., Martynoga, B., Parras, C., Ramesh, V., Pacary, E., Johnston, C., et al. (2011). A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes & Development*, *25*(9), 930–945.

Chambers, S. M., Fasano, C. A., Papapetrou, E. P., Tomishima, M., Sadelain, M., & Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. *Nature Biotechnology*, *27*(3), 275–280.

Chanda, S., Ang, C. E., Davila, J., Pak, C., Mall, M., Lee, Q. Y., et al. (2014). Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1. *Stem Cell Reports*, *3*(2), 282–296.

Charney, R. M., Paraiso, K. D., Blitz, I. L., & Cho, K. W. Y. (2017). A gene regulatory program controlling early Xenopus mesendoderm formation: Network conservation and motifs. *Seminars in Cell and Developmental Biology*, *66*, 12–24.

Chazaud, C., & Yamanaka, Y. (2016). Lineage specification in the mouse preimplantation embryo. *Development*, *143*(7), 1063–1074.

Chen, B., Kim, E.-H., & Xu, P.-X. (2009). Initiation of olfactory placode development and neurogenesis is blocked in mice lacking both Six1 and Six4. *Developmental Biology*, *326*(1), 75–85.

Chi, L., Fan, B., Feng, D., Chen, Z., Liu, Z., Hui, Y., et al. (2016). The Dorsoventral Patterning of Human Forebrain Follows an Activation/Transformation Model. *Cerebral Cortex*, 152–14.

Choi, M., Shi, J., Jung, S., Chen, X., & Cho, K. (2012). Attractor Landscape Analysis Reveals Feedback Loops in the p53 Network That Control the Cellular Response to DNA Damage. *Computational Biology*, 251(5), 1–14.

Chng, Z., Teo, A., Pedersen, R. A., & Vallier, L. (2010). SIP1 Mediates Cell-Fate Decisions between Neuroectoderm and Mesendoderm in Human Pluripotent Stem Cells. *Stem Cell*, *6*(1), 59–70.

Chung, S., Shin, B. S., Hedlund, E., Pruszak, J., Ferree, A., Kang, U. J., et al. (2006). Genetic selection of sox1GFP-expressing neural precursors removes residual tumorigenic pluripotent stem cells and attenuates tumor formation after transplantation. *Journal of Neurochemistry*, *97*(5), 1467–1480.

Ciani, L., & Salinas, P. C. (2005). Signalling in neural development: WNTS in the vertebrate nervous system: from patterning to neuronal connectivity. *Nature Reviews Neuroscience*, *6*(5), 351–362.

Cohen, D., b, Cheng, C., Cheng, S., & Hui, C. (2000). Expression of two novel mouse Iroquois homeobox genes during neurogenesis. *Mechanisms of Development,* 91, 317-321.

Conidi, A., Cazzola, S., Beets, K., Coddens, K., Collart, C., Cornelis, F., et al. (2011). Few Smad proteins and many Smad-interacting proteins yield multiple functions and action modes in TGFβ/BMP signaling in vivo. *Cytokine and Growth Factor*

*Reviews*, *22*(5-6), 287–300.

Corrigan, A. M., Tunnacliffe, E., Cannon, D., & Chubb, J. R. (2016). A continuum model of transcriptional bursting. *eLife*, *5*, e13051–38.

Dar R., Razooky B., Singh A., Trimeloni T., McCollum J., Cox C., Simpson M., & Weinberger L. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc Natl Acad Sci*, 109, 17454–17459.

De Rosa, A., Pellegatta, S., Rossi, M., Tunici, P., Magnoni, L., Speranza, M. C., et al. (2012). A Radial Glia Gene Marker, Fatty Acid Binding Protein 7 (FABP7), Is Involved in Proliferation and Invasion of Glioblastoma Cells. *PLoS ONE*, *7*(12), e52113–16.

Deglincerti, A., Croft, G. F., Pietila, L. N., Zernicka-Goetz, M., Siggia, E. D., & Brivanlou, A. H. (2016). Self-organization of the in vitro attached human embryo. *Nature*, *533*(7602), 251–254.

Delaune, E. (2005). Neural induction in Xenopus requires early FGF signalling in addition to BMP inhibition. *Development*, *132*(2), 299–310.

Dey, S., Kester, L., Spanjaard, B., Bienko, M., & van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. *Nat. Biotechnol. 33*, 285–289.

Dhanesh, S. B., Subashini, C., & James, J. (2017). Hes1: the maestro in neurogenesis. *Cellular and Molecular Life Sciences*, *73*(21), 4019–4042.

Dincer, Z., Piao, J., Niu, L., Ganat, Y., Kriks, S., Zimmer, B., et al. (2013). Specification of Functional Cranial Placode Derivatives from Human Pluripotent Stem Cells. *Cell Reports*, *5*(5), 1387–1402.

Ding, Y., Ploper, D., Sosa, E. A., Colozza, G., Moriyama, Y., Benitez, M. D. J., et al. (2017). Spemann organizer transcriptome induction by early beta-catenin, Wnt,

Nodal, and Siamois signals in Xenopus laevis. *Proceedings of the National Academy of Sciences*, *114*(15), E3081–E3090.

Duan, L., Peng, C.-Y., Pan, L., & Kessler, J. A. (2015). Human Pluripotent Stem Cell-Derived Radial Glia Recapitulate Developmental Events and Provide Real-Time Access to Cortical Neurons and Astrocytes. *STEM CELLS Translational Medicine*, *4*(5), 437–447.

Ealy, M., Ellwanger, D. C., Kosaric, N., Stapper, A. P., & Heller, S. (2016). Single-cell analysis delineates a trajectory toward the human early otic lineage. *Proceedings of the National Academy of Sciences*, *113*(30), 8508–8513.

Edmonson, R., Broglie, J., Adcock, A., & Yang, L. (2014). Three-Dimensional Cell Culture Systems and Their Applications in Drug Discovery and Cell-Based Biosensors. *ASSAY and Drug Development Technologies*, 12(4): 207-218.

Elkabetz, Y., Panagiotakos, G., Shamy, Al, G., Socci, N. D., Tabar, V., & Studer, L. (2008). Human ES cell-derived neural rosettes reveal a functionally distinct early neural stem cell stage. *Genes & Development*, *22*(2), 152–165.

Elowitz, M., Levine, A., Siggia, E., Swain, P. (2002). Stochastic Gene Expression in a Single Cell. Science, 297 (5584), 1183–1186.

Enríquez, P. (2016). CRISPR-Mediated Epigenome Editing *Yale Journal of Biology and Medicine*, 86, 471-486.

Evseenko, D., Zhua, Y., Schenke-Laylandb, K., Kuoa, J., Latoura, B., Gea, S., Scholesa, J., Dravida,G., et al (2010). Mapping the first stages of mesoderm commitment during differentiation of human embryonic stem cells. *PNAS*, 107 (31), 13742–13747.

Fagnocchi, L., Cherubini, A., Hatsuda, H., Fasciani, A., Mazzoleni, S., Poli, V., et al. (2016). A Myc-driven self-reinforcing regulatory network maintains mouse embryonic stem cell identity. *Nature Communications*, *7*, 1–17.

Fagnocchi, L., & Zippo, A. (2017). Multiple Roles of MYC in Integrating Regulatory Networks of Pluripotent Stem Cells. *Frontiers in Cell and Developmental Biology*, *5*, 981–19.

Faial, T., Bernardo, A. S., Mendjan, S., Diamanti, E., Ortmann, D., Gentsch, G. E., et al. (2015). Brachyury and SMAD signalling collaboratively orchestrate distinct mesoderm and endoderm gene regulatory networks in differentiating human embryonic stem cells. *Development*, *142*(12), 2121–2135.

Faissner, A., & Reinhard, J. (2015). The extracellular matrix compartment of neural stem and glial progenitor cells. *Glia*, *63*(8), 1330–1349.

Fujita, K., Iwaki, M., & Yanagida, T. (2017). Transcriptional bursting is intrinsically caused by interplay between RNA polymerases on DNA. *Nature Communications*, *7*, 1–10.

Fukaya, T., Lim, B., & Levine, M. (2016). Enhancer Control of Transcriptional Bursting. *Cell*, *166*(2), 358–368.

Fukuoka, M., Uehara, A., Niki, K., Goto, S., Kato, D., Utsugi, T., et al. (2013). Identification of preferentially reactivated genes during early G1 phase using nascent mRNA as an index of transcriptional activity. *Biochemical and Biophysical Research Communications*, *430*(3), 1005–1010.

Fyfe, J. C. (2006). An 140-kb deletion associated with feline spinal muscular atrophy implies an essential LIX1 function for motor neuron survival. *Genome Research*, *16*(9), 1084–1090.

Gaur, S., Mandelbaum, M., Herold, M., Majumdar, H. D., Neilson, K. M., Maynard, T. M., et al. (2016). Neural transcription factors bias cleavage stage blastomeres to give rise to neural ectoderm. *Genesis*, *54*(6), 334–349.

Gao, S., Hou, X., Jiang, Y., Xu, Z., Cai, T., Chen, J., & Chang, G. (2017). Integrated analysis of hematopoietic differentiation outcomes and molecular characterization

reveals unbiased differentiation capacity and minor transcriptional memory in HPC/HSC-iPSCs. *Stem Cell Research & Therapy*, 1–12.

Ghaffarizadeh, A., Podgorski, G. J., & Flann, N. S. (2017). Applying attractor dynamics to infer gene regulatory interactions involved in cellular differentiation. *Biosystems*, *155*, 29–41.

Gomez-Skarmeta, J. L., & Modolell, J. (2002). Iroquois genes: genomic organization and function in vertebrate neural development. *Current Opinion in Genetics & Development*, *12*(4), 403–408.

Gomez, I. Mojtahedi1, M., Wu1, W., & Cao, W. (2013). Beyond Cancer Stem Cells: Understanding Cancer Heterogeneity Through Gene Regulatory Networks and Single Cell Analysis. *Journal of Stem Cell Research & Therapy*, *01*(S1), 1–4.

Gonzales, K. A. U., Liang, H., Lim, Y.-S., Chan, Y.-S., Yeo, J.-C., Tan, C.-P., et al. (2015). Deterministic Restriction on Pluripotent State Dissolution by Cell-Cycle Pathways. *Cell*, *162*(3), 564–579.

Goolam, M., Scialdone, A., Graham, S. J. L., Macaulay, I. C., Jedrusik, A., Hupalowska, A., et al. (2016). Heterogeneity in Oct4 and Sox2 Targets Biases Cell Fate in 4-Cell Mouse Embryos. *Cell*, *165*(1), 61–74.

Graham, V., Khudyakov, J., Ellis, P., & Pevny, L. (2003). SOX2 Functions to Maintain Neural Progenitor Identity. *Neuron*, *39*(5), 749–765.

Hachem, S., Laurenson, A.-S., Hugnot, J.-P., & Legraverend, C. (2007). Expression of S100B during embryonic development of the mouse cerebellum. *BMC Developmental Biology*, *7*(1), 17–15.

Harding, M. J., McGraw, H. F., & Nechiporuk, A. (2014). The roles and regulation of multicellular rosette structures during morphogenesis. *Development*, *141*(13), 2549–2558.

Harrison-Uy, S. J., & Pleasure, S. J. (2012). Wnt Signaling and Forebrain Development. *Cold Spring Harbor Perspectives in Biology*, *4*(7), a008094–a008094.

Heeg-Truesdell, E., & LaBonne, C. (2006). Neural induction in Xenopus requires inhibition of Wnt-β-catenin signaling. *Developmental Biology*, *298*(1), 71–86.

Hecker, M., Lambeck, S., Toepfer, S., van Someren, E., & Guthke, R. (2009). Gene regulatory network inference: Data integration in dynamic models—A review. *Biosystems*, *96*(1), 86–103.

Hendy, O., Campbell, J., Weissmana, J., Larsonb, D., & Singera, S. (2017). Differential context-speci c impact of individual core promoter elements on transcriptional dynamics. *Molecular Biology of the Cell*, 17(06), 3360-3369.

Hollister, S. (2012). Porous scaffold design for tissue engineering. *Nature materials*, 4, 518-525.

Holmberg, J., Hansson, E., Malewicz, M., Sandberg, M., Perlmann, T., Lendahl, U., & Muhr, J. (2008). SoxB1 transcription factors and Notch signaling use distinct mechanisms to regulate proneural gene function and neural progenitor differentiation. *Development*, *135*(10), 1843–1851.

Hsiung, C. C. S., Bartman, C. R., Huang, P., Ginart, P., Stonestrom, A. J., Keller, C. A., et al. (2016). A hyperactive transcriptional state marks genome reactivation at the mitosis–G1 transition. *Genes & Development*, *30*(12), 1423–1439.

Hu, B. Y., Weick, J. P., Yu, J., Ma, L. X., Zhang, X. Q., Thomson, J. A., & Zhang, S. C. (2010). Neural differentiation of human induced pluripotent stem cells follows developmental principles but with variable potency. *Proceedings of the National Academy of Sciences*, *107*(9), 4335–4340.

Hu, W., Lu, H., SWang, S., Yin, W., Liu, X., Dong, L., Chiu, R., Shen, L., Lu, W., Lan, F. (2016). Suppression of Nestin reveals a critical role for p38-EGFR pathway in neural progenitor cell proliferation. *Oncotarget*, 7(52), 87052-87063.

Huang, S.-M. A., Mishina, Y. M., Liu, S., Cheung, A., Stegmeier, F., Michaud, G. A., et al. (2009a). Tankyrase inhibition stabilizes axin and antagonizes Wnt signalling. *Nature*, *461*(7264), 614–620.

Huang, S., Ernberg, I., & Kauffman, S. (2009b). Cancer attractors: A systems view of tumors from a gene network dynamics and developmental perspective. *Seminars in Cell and Developmental Biology*, *20*(7), 869–876.

Huang, C. T.-L., Tao, Y., Lu, J., Jones, J. R., Fowler, L., Weick, J. P., & Zhang, S.-C. (2016). Time-Course Gene Expression Profiling Reveals a Novel Role of Non-Canonical WNT Signaling During Neural Induction. *Scientific Reports*, *6*(1), 13–30.

Inoue, T., Ota, M., Ogawa, M., Mikoshiba, K., & Aruga, J. (2007). Zic1 and Zic3 Regulate Medial Forebrain Development through Expansion of Neuronal Progenitors. *Journal of Neuroscience*, *27*(20), 5461–5473.

Islam, S., Kjallquist, U., Moliner, A., Zajac, P., Fan, J. B., Lonnerberg, P., & Linnarsson, S. (2011). Characterization of the single-cell transcriptional landscape by highly multiplex RNA-seq. *Genome Research*, *21*(7), 1160–1167.

Janesick, A., Shiotsugu, J., Taketani, M., & Blumberg, B. (2012). RIPPLY3 is a retinoic acid-inducible repressor required for setting the borders of the pre-placodal ectoderm. *Development*, *139*(6), 1213–1224.

Kadowaki, M., Nakamura, S., Machon, O., Krauss, S., Radice, G. L., & Takeichi, M. (2007). N-cadherin mediates cortical organization in the mouse brain. *Developmental Biology*, *304*(1), 22–33.

Kageyama, R., Ohtsuka, T., Hatakeyama, J., & Ohsawa, R. (2005). Roles of bHLH genes in neural stem cell differentiation. *Experimental Cell Research*, *306*(2), 343–348.

Kageyama, R., Shimojo, H., & Imayoshi, I. (2014). Dynamic expression and roles of Hes factors in neural development. *Cell and Tissue Research*, *359*(1), 125–133.

Kamiya, D., Banno, S., Sasai, N., Ohgushi, M., Inomata, H., Watanabe, K., et al. (2011). Intrinsic transition of embryonic stem-cell differentiation into neural progenitors. *Nature*, *470*(7335), 503–509.

Kar, G., Kim, J. K., Kolodziejczyk, A. A., Natarajan, K. N., Triglia, E. T., Mifsud, B., et al. (2017). Flipping between Polycomb repressed and active transcriptional states introduces noise in gene expression. *Nature Communications*, 1–13. http://doi.org/10.1038/s41467-017-00052-2

Karlebach, G., & Shamir, R. (2008). Modelling and analysis of gene regulatory networks. *Nature Reviews Molecular Cell Biology*, *9*(10), 770–780.

Kathuria, A ., Nowosiad, P., Jagasia, R., Aigner, S., Taylor, RD., Andrea, LC., Gatford, NJF., Lucchesi, W., Srivastava, DP., & Price, J. (2017). Stem cell-derived neurons from autistic individuals with SHANK3 mutation show morphogenetic abnormalities during early development. Molecular Psychiatry, 00, 1–12.

Kedziora, K. M., Leyton-Puig, D., Argenzio, E., Boumeester, A. J., van Butselaar, B., Yin, T., et al. (2016). Rapid Remodeling of Invadosomes by G i-coupled Receptors. *Journal of Biological Chemistry*, *291*(9), 4323–4333.

Kharchenko, P. V., Silberstein, L., & Scadden, D. T. (2014). Bayesian approach to single-cell differential expression analysis. *Nature Methods*, *11*(7), 740–742.

Kim, Y. S., Yi, B.R., Kim, N. H., & Choi, K. C. (2014). Role of the epithelial & mesenchymal transition and its effects on embryonic stem cells, *46*(8), e108–6.

Klein, S. L., & Moody, S. A. (2015). Early neural ectodermal genes are activated by siamois and twin during blastula stages. *Genesis*, *53*(5), 308–320.

Kærn, M., Elston, T. C., Blake, W. J., & Collins, J. J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nature Reviews Genetics*, *6*(6), 451–464.

Koch, P., Opitz, T., Steinbeck, J., Ladewig, Brutstle, J. (2009). A rosette-type, self-renewing human ES cell-derived neural stem cell with potential for in vitro instruction and synaptic integration. *PNAS*, 106(9), 3225–3230.

Kohn, A., & Moon, R. (2005). Wnt and calcium signaling: β-Catenin-independent pathways. Cell Calcium. 38, 439–446.

Komia Y., & Habas R. (2008). Wnt signal transduction pathways. *Organogenesis*, 4(2), 68-75.

Kordmahalleh, M., Sefidmazgi, M., Harrison, S., Homaifar, A. (2017). Identifying time-delayed gene regulatory networks via an evolvable hierarchical recurrent neural network, *BioData Mining*, 10(29), 1–25.

La Manno, G., Gyllborg, D., Codeluppi, S., Nishimura, K., Salto, C., Zeisel, A., et al. (2016). Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell*, *167*(2), 566–580.e19.

Lamouille, S., Xu, J., & Derynck, R. (2014). Molecular mechanisms of epithelial–mesenchymal transition. *Nature Reviews Molecular Cell Biology*, *15*(3), 178–196.

Larsen, K. B., Lutterodt, M. C., Møllgård, K., & Møller, M. (2009). Expression of the Homeobox Genes OTX2and OTX1in the Early Developing Human Brain. *Journal of Histochemistry & Cytochemistry*, *58*(7), 669–678.

Lanctot, A. A., Guo, Y., Le, Y., Edens, B. M., Nowakowski, R. S., & Feng, Y. (2017). Loss of Brap Results in Premature G1/S Phase Transition and Impeded Neural Progenitor Differentiation. *CellReports*, *20*(5), 1148–1160.

Lanner, F. (2014). Lineage specification in the early mouse embryo. *Experimental Cell Research*, *321*(1), 32–39.

Lee, H.-K., Lee, H.-S., & Moody, S. A. (2014). Neural Transcription Factors: from Embryos to Neural Stem Cells. *Molecules and Cells*, *37*(10), 705–712.

Lee, H.-K., Lee, H.-S., & Moody, S. A. (2014). Neural Transcription Factors: from Embryos to Neural Stem Cells. *Molecules and Cells*, *37*(10), 705–712.

Leung, A. W., Morest, D. K., & Li, J. Y. H. (2013). Differential BMP signaling controls formation and differentiation of multipotent preplacodal ectoderm progenitors from human embryonic stem cells. *Developmental Biology*, *379*(2), 208–220.

Leung, A. W., Murdoch, B., Salem, A. F., Prasad, M. S., Gomez, G. A., & Garcia-Castro, M. I. (2016). WNT/catenin signaling mediates human neural crest induction via a pre-neural border intermediate. *Development*, *143*(3), 398–410.

Li, E., & Davidson, E. H. (2009). Building developmental gene regulatory networks. *Birth Defects Research Part C: Embryo Today: Reviews*, *87*(2), 123–130.

Li, L., Bennett, S. A. L., & Wang, L. (2014). Role of E-cadherin and other cell adhesion molecules in survival and differentiation of human pluripotent stem cells. *Cell Adhesion & Migration*, *6*(1), 59–73.

Li, M., Zou, Y., Lu, Q., Tang, N., Heng, A., Islam, I., et al. (2017). Efficient derivation of dopaminergic neurons from SOX1- floor plate cells under defined culture conditions. *Journal of Biomedical Science*, 1–13.

Liu, Y., Pelham-Webb, B., Di Giammartino, D. C., Li, J., Kim, D., Kita, K., et al. (2017). Widespread Mitotic Bookmarking by Histone Marks and Transcription Factors in Pluripotent Stem Cells. *CellReports*, *19*(7), 1283–1293.

Lim, J. W., Hummert, P., Mills, J. C., & Kroll, K. L. (2010). Geminin cooperates with Polycomb to restrain multi-lineage commitment in the early embryo. *Development*, *138*(1), 33–44.

Löhle, M., Hermann, A., Glaß, H., Kempe, A., Schwarz, S. C., Kim, J. B., et al. (2012). Differentiation Efficiency of Induced Pluripotent Stem Cells Depends on the Number of Reprogramming Factors. *Stem Cells*, *30*(3), 570–579.

Lopes, F. M., de Oliveira, E. A., & Cesar, R. M., Jr. (2011). Inference of gene regulatory networks from time series by Tsallis entropy. *BMC Systems Biology*, *5*(1), 61.

Lukovic, D., Diez Lloret, A., Stojkovic, P., Rodríguez-Martínez, D., Perez Arago, M. A., Rodriguez-Jimenez, F. J., et al. (2017). Highly Efficient Neural Conversion of Human Pluripotent Stem Cells in Adherent and Animal-Free Conditions. *STEM CELLS Translational Medicine*, *6*(4), 1217–1226.

Luo, Z., Gao, X., Lin, C., Smith, E. R., Marshall, S. A., Swanson, S. K., et al. (2015). Zic2 Is an Enhancer-Binding Factor Required for Embryonic Stem Cell Specification. *Molecular Cell*, *57*(4), 685–694.

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

MacDonald, B. T., Tamai, K., & He, X. (2009). Wnt/β-Catenin Signaling: Components, Mechanisms, and Diseases. *Developmental Cell*, *17*(1), 9–26.

MacNeil, L. T., & Walhout, A. J. M. (2011). Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression. *Genome Research*, *21*(5), 645–657.

McKey, J., Martire, D., de Santa Barbara, P., & Faure, S. (2016). LIX1 regulates YAP1 activity and controls the proliferation and differentiation of stomach mesenchymal progenitors. *BMC Biology*, 1–16.

Maden, M. (2007). Retinoic acid in the development, regeneration and maintenance of the nervous system. *Nature Reviews Neuroscience*, *8*(10), 755–765.

Mahler, J., & Driever, W. (2007). Expression of the zebrafish intermediate neurofilament Nestin in the developing nervous system and in neural proliferation zones at postembryonic stages. *BMC Developmental Biology*, *7*(1), 89–11.

Malchenko, S., Xie, J., de Fatima Bonaldo, M., Vanin, E. F., Bhattacharyya, B. J., Belmadani, A., et al. (2014). Onset of rosette formation during spontaneous neural differentiation of hESC and hiPSC colonies. *Gene*, *534*(2), 400–407.

Malysheva, V., Mendoza-Parra, M. A., Saleem, M.-A. M., & Gronemeyer, H. (2018). Reconstruction of gene regulatory networks reveals chromatin remodelers and key transcription factors in tumorigenesis. *Genome Medicine*, 1–16.

Marchal, L., Luxardi, G., Thome, V., & Kodjabachian, L. (2009). BMP inhibition initiates neural induction via FGF signaling and Zic genes. *PNAS*. 106(41), 17437-17442.

Manuel, M., Martynoga, B., Yu, T., West, J. D., Mason, J. O., & Price, D. J. (2010). The transcription factor Foxg1 regulates the competence of telencephalic cells to adopt subpallial fates in mice. *Development*, *137*(3), 487–497.

Mathieu, J., Zhou, W., Xing, Y., Sperber, H., Ferreccio, A., Agoston, Z., et al. (2014). Hypoxia-Inducible Factors Have Distinct and Stage-Specific Roles during Reprogramming of Human Cells to Pluripotency. *Stem Cell*, *14*(5), 592–605.

Matson, J. P., & Cook, J. G. (2016). Cell cycle proliferation decisions: the impact of single cell analyses. *The FEBS Journal*, *284*(3), 362–375.

Matsuoka, A. J., Morrissey, Z. D., Zhang, C., Homma, K., Belmadani, A., Miller, C. A., et al. (2016). Directed Differentiation of Human Embryonic Stem Cells Toward Placode-Derived Spiral Ganglion-Like Sensory Neurons. *STEM CELLS Translational Medicine*, *6*(3), 923–936.

Menchero, S., Rayon, T., Andreu, M. J., & Manzanares, M. (2016). Signaling pathways in mammalian preimplantation development: Linking cellular phenotypes to lineage decisions. *Developmental Dynamics*, 1–17.

Moignard, V., & Göttgens, B. (2014). Transcriptional mechanisms of cell fate decisions revealed by single cell expression profiling. *BioEssays*, *36*(4), 419–426.

Moody, S., Klein, S., Kapinski, B., Maynard, T., & LaMantia, A. (2013). *Review Article:* On becoming neural: what the embryo can tell us about differentiating neural stem cells. *Am J Stem Cells,* 2(2), 74-94.

Moody, S. A., & LaMantia, A.-S. (2015). Transcriptional Regulation of Cranial Sensory Placode Development. In *Neural Crest and Placodes* , 111, 301–350.

Morris, S. A. (2017). Human embryos cultured in vitro to 14 days. *Open Biology*, *7*(1), 170003–5.

Munski, B., Neuert, G., Oudenaarde, A. (2012). Using Gene Expression Noise to Understand Gene Regulation. Science, 336 (6078), 183–187.

Muñoz-Descalzo, S., Hadjantonakis, A.-K., & Arias, A. M. (2015). Wnt/ß-catenin signalling and the dynamics of fate decisions in early mouse embryos and embryonic stem (ES) cells. *Seminars in Cell and Developmental Biology*, 1–9.

Muñoz-Sanjuán, I., & Brivanlou, A. H. (2002). Neural induction, the default model and embryonic stem cells. *Nature Reviews Neuroscience*, *3*(4), 271–280.

Muramoto, T., MUller, I., Thomas, G., Melvin, A., & Chubb, J. R. (2010). Methylation of H3K4 Is Required for Inheritance of Active Transcriptional States. *Current Biology*, *20*(5), 397–406.

Muratore, C. R., Srikanth, P., Callahan, D. G., & Young-Pearse, T. L. (2014). Comparison and Optimization of hiPSC Forebrain Cortical Differentiation Protocols. *PLoS ONE*, *9*(8), e105807–18.

Nakanishi, M & Otsu, M. (2012). Development of Sendai Virus Vectors and their Potential Applications in Gene Therapy and Regenerative Medicine. *Current Gene Therapy*, *12,* 410-416**.**

Nandadasa, S., Tao, Q., Menon, N. R., Heasman, J., & Wylie, C. (2009). N- and E-cadherins in Xenopus are specifically required in the neural and non-neural ectoderm, respectively, for F-actin assembly and morphogenetic movements. *Development*, *136*(8), 1327–1338.

Nat, R., Nilbratt, M., Narkilahti, S., Winblad, B., Hovatta, O., & Nordberg, A. (2007). Neurogenic neuroepithelial and radial glial cells generated from six human embryonic stem cell lines in serum-free suspension and adherent cultures. *Glia*, *55*(4), 385–399.

Neely, M. D., Litt, M. J., Tidball, A. M., Li, G. G., Aboud, A. A., Hopkins, C. R., et al. (2012). DMH1, a Highly Selective Small Molecule BMP Inhibitor Promotes Neurogenesis of hiPSCs: Comparison of PAX6 and SOX1 Expression during Neural Induction. *ACS Chemical Neuroscience*, *3*(6), 482–491.

Neriec, N., & Desplan, C. (2014). Different ways to make neurons: parallel evolution in the SoxB family. *Genome Biology*, *15*(5), 116–3.

Newman, A. M., & Cooper, J. B. (2010). Lab-Specific Gene Expression Signatures in Pluripotent Stem Cells. *Stem Cell*, *7*(2), 258–262.

Niakan, K. K., & Eggan, K. (2013). Analysis of human embryos from zygote to blastocyst reveals distinct gene expression patterns relative to the mouse. *Developmental Biology*, *375*(1), 54–64.

Nicolas, D., Phillips, N. E., & Naef, F. (2017). What shapes eukaryotic transcriptional bursting? *Molecular BioSystems*, *13*, 1280–1290.

Nishimura, N., Kamimura, Y., Ishida, Y., Takemoto, T., Kondoh, H., & Uchikawa, M. (2012). A Systematic Survey and Characterization of Enhancers that Regulate Sox3 in Neuro-Sensory Development in Comparison with Sox2 Enhancers. *Biology*, *1*(3), 714–735.

Niu, W., Zang, T., Smith, D. K., Vue, T. Y., Zou, Y., Bachoo, R., et al. (2015). SOX2 Reprograms Resident Astrocytes into Neural Progenitors in the Adult Brain. *Stem Cell Reports*, *4*(5), 780–794.

Nuse, R. (2012). WNT signaling. *Cold Spring Harb Perspect Biol,* a011163.

Ohkubo, N., Matsubara, E., Yamanouchi, J., Akazawa, R., Aoto, M., Suzuki, Y., et al. (2014). Abnormal Behaviors and Developmental Disorder of Hippocampus in Zinc Finger Protein 521 (ZFP521) Mutant Mice. *PLoS ONE*, *9*(3), e92848–13.

Ohnishi, Y., Huber, W., Tsumura, A., Kang, M., Xenopoulos, P., Kurimoto, K., *et al*. (2013). Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nature Cell Biology*, *16*(1), 27–37.

Okano, H., & Temple, S. (2009). Cell types to order: temporal specification of CNS stem cells. *Current Opinion in Neurobiology*, *19*(2), 112–119.

Olsen, C., Fleming, K., Prendergast, N., Rubio, R., Emmert-Streib, F., Bontempi, G., et al. (2014). Inference and validation of predictive gene networks from biomedical literature and gene expression data. *Genomics*, *103*(5-6), 329–336.

Olsen, C., Fleming, K., Prendergast, N., Rubio, R., Emmert-Streib, F., Bontempi, G., et al. (2015). Using shRNA experiments to validate gene regulatory networks. *Gdata*, *4*(C), 123–126.

O'Malley, J., Skylaki, S., Iwabuchi, K. A., Chantzoura, E., Ruetz, T., Johnsson, A., et al. (2013). High-resolution analysis with novel cell-surface markers identifies routes to iPS cells. *Nature*, *498*(7456), 88–91.

Overton, K. W., Spencer, S. L., Noderer, W. L., Meyer, T., & Wang, C. L. (2014). Basal p21 controls population heterogeneity in cycling and quiescent cell cycle states. *Proceedings of the National Academy of Sciences*, *111*(41), E4386–E4393.

Ozair, M. Z., Kintner, C., & Brivanlou, A. H. (2013). Neural induction and early patterning in vertebrates. *Wiley Interdisciplinary Reviews: Developmental Biology*, *2*(4), 479–498.

Ozair, M. Z., Noggle, S., Warmflash, A., Krzyspiak, J. E., & Brivanlou, A. H. (2012). SMAD7 Directly Converts Human Embryonic Stem Cells to Telencephalic Fate by a Default Mechanism. *Stem Cells*, *31*(1), 35–47.

Park, D., Xiang, A. P., Mao, F. F., Zhang, L., Di, C.-G., Liu, X.-M., et al. (2010). Nestin Is Required for the Proper Self-Renewal of Neural Stem Cells. *Stem Cells*, *28*(12), 2162–2171.

Pauklin, S., & Vallier, L. (2013). The Cell-Cycle State of Stem Cells Determines Cell Fate Propensity. *Cell*, *155*(1), 135–147.

Padovan-Merhar, O., Nair, G. P., Biaesch, A. G., Mayer, A., Scarfone, S., Foley, S. W., et al. (2015). Single Mammalian Cells Compensate for Differences in Cellular Volume and DNA Copy Number through Independent Global Transcriptional Mechanisms. *Molecular Cell*, *58*(2), 339–352.

Pelton, T. A., Sharma, S., Schulz, T. C., Rathjen, J., & Rathjen, P. D. (2001). Transient pluripotent cell populations during primitive ectoderm formation: correlation of in vivo and in vitro pluripotent cell development. *Journal of Cell Science*, *115*(2), 329–339.

Peng, Y., Zhang, X., Feng, X., Fan, X., & Jin, Z. (2017). The crosstalk between microRNAs and the Wnt/β-catenin signaling pathway in cancer**.** *Oncotarget*, 8(8), 14089-14106.

Peter, I. S., & Davidson, E. H. (2017). Assessing regulatory information in developmental gene regulatory networks. *Proceedings of the National Academy of Sciences*, *114*(23), 5862–5869.

Phillips, B. T., Kwon, H.-J., Melton, C., Houghtaling, P., Fritz, A., & Riley, B. B. (2006). Zebrafish msxB, msxC and msxE function together to refine the neural–nonneural

border and regulate cranial placodes and neural crest development. *Developmental Biology*, *294*(2), 376–390.

Picelli, S., Faridani, O. R., rklund, A. S. K. B. O., Winberg, G. O. S., Sagasser, S., & Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. *Nature Protocols*, *9*(1), 171–181.

Pieters, T., & Van Roy, F. (2014). Role of cell-cell adhesion complexes in embryonic stem cell biology. *Journal of Cell Science*, *127*(12), 2603–2613.

Prasanth, S. G., Méndez, J., Prasanth, K. V., & Stillman, B. (2004). Dynamics of pre-replication complex proteins during the cell division cycle. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *359*(1441), 7–16.

Premarathne, S., Murtaza, M., Matigian, N., Jolly, L. A., & Wood, S. A. (2017). Loss of Usp9x disrupts cell adhesion, and components of the Wnt and Notch signaling pathways in neural progenitors. *Nature Publishing Group*, 1–18.

Provenzano, C., Pascucci, B., Lupari, E., & Civitareale, D. (2010). Large scale analysis of transcription factor TTF-1/NKX2.1 target genes in GnRH secreting cell line GT1-7. *Molecular and Cellular Endocrinology*, *323*(2), 215–223.

Poliandri, A., Miller, D., Howard, S., Nobles, M., Ruiz-Babot, G., Harmer, S., et al. (2017). Generation of kisspeptin-responsive GnRH neurons from human pluripotent stem cells. *Molecular and Cellular Endocrinology*, *447*, 12–22.

Pollen, A. A., Nowakowski, T. J., Chen, J., Retallack, H., Sandoval-Espinosa, C., Nicholas, C. R., et al. (2015). Molecular Identity of Human Outer Radial Glia during Cortical Development. *Cell*, *163*(1), 55–67.

Pombo, M. A., Zheng, Y., Fei, Z., Martin, G. B., & Rosli, H. G. (2018). Use of RNA-seq data to identify and validate RT-qPCR reference genes for studying the tomato-Pseudomonas pathosystem. *Nature Publishing Group*, 1–11.

Poulin, J.-F., Tasic, B., Hjerling-Leffler, J., Trimarchi, J. M., & Awatramani, R. (2016). Disentangling neural cell diversity using single-cell transcriptomics. *Nature Neuroscience*, *19*(9), 1131–1141.

Raposo, A. A. S. F., Vasconcelos, F. F., Drechsel, D., Marie, C., Johnston, C., Dolle, D., et al. (2015). Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis. *CellReports*, *10*(9), 1544–1556.

Qiao, Y., Yang, X., & Jing, N. (2016). Epigenetic regulation of early neural fate commitment. *Cellular and Molecular Life Sciences*, *73*(7), 1399–1411.

Qin, S., Ma, F., & Chen, L. (2015). Gene regulatory networks by transcription factors and microRNAs in breast cancer. *Bioinformatics*. 31(1), 76–83.

Raj, A., Peskin, C. S., Tranchina, D., Vargas, D. Y., & Tyagi, S. (2006). Stochastic mRNA Synthesis in Mammalian Cells. *PLoS Biology*, *4*(10), e309–13.

Range, R. C., Angerer, R. C., & Angerer, L. M. (2013). Integration of Canonical and Noncanonical Wnt Signaling Pathways Patterns the Neuroectoderm Along the Anterior–Posterior Axis of Sea Urchin Embryos. *PLoS Biology*, *11*(1), e1001467–18.

Rodríguez-Seguel, E., Alarcón, P., & Gómez-Skarmeta, J. L. (2009). The Xenopus Irx genes are essential for neural patterning and define the border between prethalamus and thalamus through mutual antagonism with the anterior repressors Fezf and Arx. *Developmental Biology*, *329*(2), 258–268.

Rogers, C. D., Moody, S. A., & Casey, E. S. (2009). Neural induction and factors that stabilize a neural fate. *Birth Defects Research Part C: Embryo Today: Reviews*, *87*(3), 249–262.

Roode, M., Blair, K., Snell, P., Elder, K., Marchant, S., Smith, A., & Nichols, J. (2012). Human hypoblast formation is not dependent on FGF signalling. *Developmental Biology*, *361*(2), 358–363.

Ruiz, S., Panopoulos, A. D., Herrerías, A., Bissig, K.-D., Lutz, M., Berggren, W. T., et al. (2011). A High Proliferation Rate Is Required for Cell Reprogramming and Maintenance of Human Embryonic Stem Cell Identity. *Current Biology*, *21*(1), 45–52.

Saint-Jeannet, J.-P., & Moody, S. A. (2014). Establishing the pre-placodal region and breaking it into placodes with distinct identities. *Developmental Biology*, *389*(1), 13–27.

Saito, S., Lin, Y.-C., Tsai, M.-H., Lin, C.-S., Murayama, Y., Sato, R., & Yokoyama, K. K. (9999). Emerging roles of hypoxia-inducible factors and reactive oxygen species in cancer and pluripotent stem cells. *Kaohsiung Journal of Medical Sciences*, *31*(6), 279–286.

Sakaue-Sawano, A., Kurokawa, H., Morimura, T., Hanyu, A., Hama, H., Osawa, H., et al. (2008). Visualizing Spatiotemporal Dynamics of Multicellular Cell-Cycle Progression. *Cell*, *132*(3), 487–498.

Sankar, S., Yellajoshyula, D., Zhang, B., Teets, B., Rockweiler, N., & Kroll, K. L. (2016). Gene regulatory networks in neural cell fate acquisition from genome- wide chromatin association of Geminin and Zic1. *Scientific Reports*, 1–16.

Sato, T. (2005). Neural crest determination by co-activation of Pax3 and Zic1 genes in Xenopus ectoderm. *Development*, *132*(10), 2355–2363.

Schrode, N., Xenopoulos, P., Piliszek, A., Frankenberg, S., Plusa, B., & Hadjantonakis, A. K. (2013). Anatomy of a blastocyst: Cell behaviors driving cell fate choice and morphogenesis in the early mouse embryo. *Genesis*, *51*(4), 219–233.

Schrode, N., Saiz, N. S., Talia, S. D., & Hadjantonakis, A.-K. (2014). GATA6 Levels Modulate Primitive Endoderm Cell Fate Choice and Timing in the Mouse Blastocyst. *Developmental Cell*, *29*(4), 454–467.

Schoech A. & N. Zabet. (2014). Facilitated diffusion buffers noise in gene expression, *Phys. Rev. E*, 90(3), 032701.

Schroter, C., Rue, P., Mackenzie, J. P., & Martinez Arias, A. (2015). FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells. *Development*, *142*(24), 4205–4216.

Shahbazi, M. N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V., et al. (2016). Self-organization of the human embryo in the absence of maternal tissues. *Nature Cell Biology*, *18*(6), 700–708.

Shahbazi, E., Moradi, S., Nemati, S., Satarian, L., Basiri, M., Gourabi, H., et al. (2016). Conversion of Human Fibroblasts to Stably Self-Renewing Neural Stem Cells with a Single Zinc-Finger Transcription Factor. *Stem Cell Reports*, *6*(4), 539–551.

Shalek, A. K., Satija, R., Adiconis, X., Gertner, R. S., Gaublomme, J. T., Raychowdhury, R., et al. (2013). Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature*, *498*(7453), 236–240.

Shen, S., Pu, J., Lang, B., & McCaig, C. (2011). A zinc nger protein Zfp521 directs neural diferentiation and beyond. *Stem Cell Research & Therapy*, 2:20.

Sherman, J. H., Karpinski, B. A., Fralish, M. S., Cappuzzo, J. M., Dhindsa, D. S., Thal, A. G., et al. (2017). Foxd4 is essential for establishing neural cell fate and for neuronal differentiation. *Genesis*, *55*(6), e23031–13.

Shi, X., Richard, J., Zirbes, K. M., Gong, W., Lin, G., Kyba, M., et al. (2014). Cooperative interaction of Etv2 and Gata2 regulates the development of endothelial and hematopoietic lineages. *Developmental Biology*, *389*(2), 208–218.

Shi, Y., Inoue, H., Wu, J. C., & Yamanaka, S. (2016). Induced pluripotent stem cell technology: a decade of progress. *Nature Publishing Group*, *16*(2), 115–130.

Shigetani, Y., Wakamatsu, Y., Tachibana, T., & Okabe, M. (2016). Conversion of neural plate explants to pre-placodal ectoderm-like tissue in vitro. *Biochemical and Biophysical Research Communications*, *477*(4), 807–813.

Shimojo, H., Ohtsuka, T., & Kageyama, R. (2008). Oscillations in Notch Signaling Regulate Maintenance of Neural Progenitors. *Neuron*, *58*(1), 52–64.

Shin, S., Sun, Y., Liu, Y., Khaner, H., Svant, S., Cai, J., et al. (2007). Whole Genome Analysis of Human Neural Stem Cells Derived from Embryonic Stem Cells and Stem and Progenitor Cells Isolated from Fetal Tissue. *Stem Cells*, *25*(5), 1298–1306.

Skinner, S. O., Xu, H., Nagarkar-Jaiswal, S., Freire, P. R., Zwaka, T. P., & Golding, I. (2016). Single-cell analysis of transcription kinetics across the cell cycle. *eLife*, *5*, e59928–3.

Silver, S. J., Davies, E. L., Doyon, L., & Rebay, I. (2003). Functional Dissection of Eyes absent Reveals New Modes of Regulation within the Retinal Determination Gene Network. *Molecular and Cellular Biology*, *23*(17), 5989–5999.

Soltani, M., & Singh, A. (2016). Effects of cell-cycle-dependent expression on random fluctuations in protein levels. *Royal Society Open Science*, *3*(12), 160578–12.

Soufi, A., & Dalton, S. (2016). Cycling through developmental decisions: how cell cycle dynamics control pluripotency, differentiation and reprogramming. *Development*, *143*(23), 4301–4311.

Stanganello, E., & Scholpp, S. (2016). Role of cytonemes in Wnt transport. *Journal of Cell Science*, *129*(4), 665–672.

Stern, C. D. (2005). Neural induction: old problem, new findings, yet more questions. *Development*, *132*(9), 2007–2021.

Steventon, B., Mayor, R., & Streit, A. (2012). Mutual repression between Gbx2 and Otx2 in sensory placodes reveals a general mechanism for ectodermal patterning. *Developmental Biology*, *367*(1), 55–65.

Su, H., Wang, L., Huang, W., Qin, D., Cai, J., Yao, X., et al. (2013). Immediate expression of Cdh2 is essential for efficient neural differentiation of mouse induced pluripotent stem cells. *Stem Cell Research*, *10*(3), 338–348.

Suh, W. (2017). A new era of disease modeling and drug discovery using induced pluripotent stem cells. *Archives of Pharmacal Research*, *40*(1), 1–12.

Sun, J., Rockowitz, S., Xie, Q., Ashery-Padan, R., Zheng, D., & Cvekl, A. (2015). Identification of in vivoDNA-binding mechanisms of Pax6 and reconstruction of Pax6-dependent gene regulatory networks during forebrain and lens development. *Nucleic Acids Research*, *43*(14), 6827–6846.

Suter, D. M., Tirefort, D., Julien, S., & Krause, K.-H. (2009). A Sox1 to Pax6 Switch Drives Neuroectoderm to Radial Glia Progression During Differentiation of Mouse Embryonic Stem Cells. *Stem Cells*, *27*(1), 49–58.

Takahashi, K., & Yamanaka, S. (2006). Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, *126*(4), 663–676.

Tang, F., Barbacioru, C., Nordman, E., Bin Li, Xu, N., Bashkirov, V. I., et al. (2010). RNA-Seq analysis to capture the transcriptome landscape of a single cell. *Nature Protocols*, *5*(3), 516–535.

Tang, K., Peng, G., Qiao, Y., Song, L., & Jing, N. (2015). Intrinsic regulations in neural fate commitment. *Development, Growth & Differentiation*, *57*(2), 109–120.

Taube, J., Herschkowitzb, J., Komurovc, K., Zhou, A., Guptaf, S., Yang, J., et al., (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. *Proceedings of the National Academy of Sciences*, *107*(44), 19132–19132.

Tesarova, L., Simara, P., Stejskal, S., & Koutna, I. (2016). The Aberrant DNA Methylation Profile of Human Induced Pluripotent Stem Cells Is Connected to the

Reprogramming Process and Is Normalized During In Vitro Culture. *PLoS ONE*, *11*(6), e0157974–16.

Thakurela, S., Tiwari, N., Schick, S., Garding, A., Ivanek, R., Berninger, B., & Tiwari, V. K. (2018). Mapping gene regulatory circuitry of Pax6 during neurogenesis. *Nature Publishing Group*, 1–22.

Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., et al. (2014). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*, *32*(4), 381–386.

Ulloa, F., Briscoe, J. (2007). Morphogens and the Control of Cell Proliferation and Patterning in the Spinal Cord. *Cell Cycle*, 6(21), 2640-2649.

Vallejos, C. A., Risso, D., Scialdone, A., Dudoit, S., & Marioni, J. C. (2017). Normalizing single-cell RNA sequencing data: challenges and opportunities. *Nature Methods*, *14*(6), 565–571.

Vallier, L. (2015). Cell Cycle Rules Pluripotency. *Stem Cell*, *17*(2), 131–132.

van Caam, A., Madej, W., de Vinuesa, A. G., Goumans, M.-J., Dijke, ten, P., Davidson, E. B., & van der Kraan, P. (2017). TGFβ1-induced SMAD2/3 and SMAD1/5 phosphorylation are both ALK5-kinase- dependent in primary chondrocytes and mediated by TAK1 kinase activity, 1–13.

van de Leemput, J., Boles, N. C., Kiehl, T. R., Corneo, B., Lederman, P., Menon, V., et al. (2014). CORTECON: A Temporal Transcriptome Analysis of In Vitro Human Cerebral Cortex Development from Human Embryonic Stem Cells. *Neuron*, *83*(1), 51–68.

Vashishtha, S., Broderick, G., Craddock, T. J. A., Fletcher, M. A., & Klimas, N. G. (2015). Inferring Broad Regulatory Biology from Time Course Data: Have We Reached an Upper Bound under Constraints Typical of In Vivo Studies? *PLoS ONE*, *10*(5), e0127364–27.

Wang, Y. (2009). Wnt/Planar cell polarity signaling: A new paradigm for cancer therapy. *Mol Cancer Ther*, 8(8), 2103-2109.

Wang, F., & Higgins, J. M. G. (2013a). Histone modifications and mitosis: countermarks, landmarks, and bookmarks. *Trends in Cell Biology*, *23*(4), 175–184.

Wang, H., Wang, X., Xu, X., Zwaka, T. P., & Cooney, A. J. (2013b). Epigenetic Reprogramming of the Germ Cell Nuclear Factor Gene Is Required for Proper Differentiation of Induced Pluripotent Cells. *Stem Cells*, *31*(12), 2659–2666.

Wang, P., Qin, J., Qin, Y., Zhu, Y., Wang, L. Y., Li, M. J., et al. (2015). ChIP-Array 2: integrating multiple omics data to construct gene regulatory networks. *Nucleic Acids Research*, *43*(W1), W264–W269.

Wegner, M. (2011). SOX after SOX: SOXession regulates neurogenesis. *Genes & Development*, *25*(23), 2423–2428.

Wiese, C., Rolletschek, A., Kania, G., Blyszczuk, P., Tarasov, K. V., Tarasova, Y., et al. (2004). Nestin expression? a property of multi-lineage progenitor cells? *Cellular and Molecular Life Sciences*, *61*(19-20), 2510–2522.

Wills, A. E., Choi, V. M., Bennett, M. J., Khokha, M. K., & Harland, R. M. (2010). BMP antagonists and FGF signaling contribute to different domains of the neural plate in Xenopus. *Developmental Biology*, *337*(2), 335–350.

Wilson, P & Stice, S. (2006). Development and Differentiation of Neural Rosettes Derived From Human Embryonic Stem Cells. *Stem Cell Reviews*, 2, 67-77.

Wray, S. (2010). From Nose to Brain: Development of Gonadotrophin-Releasing Hormone -1 Neurones. *Journal of Neuroendocrinology*, *22*(7), 743–753.

Xia, X., Zhang, Y., Zieth, C & Zhang, S. (2012). Transgenes Delivered by Lentiviral Vector Are Suppressed in Human Embryonic Stem Cells in a Promoter-Dependent Manner. *Stem Cells Development*, 16(1), 167–176.

Xu, R., Wunsch, D. C., & Frank, R. L. (2007). Inference of Genetic Regulatory Networks with Recurrent Neural Network Models Using Particle Swarm Optimization. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, *4*(4), 681–692.

Yao, Z., Mich, J. K., Ku, S., Menon, V., Krostag, A.-R., Martinez, R. A., et al. (2016). A Single-Cell Roadmap of Lineage Bifurcation in Human ESC Models of Embryonic Brain Development. *Stem Cell*, 1–42.

Yellayoshyula, D., Patterson, E., Ellt, M., & Kroll, K. (2010). Geminin promotes neural fate acquisition of embryonic stem cells by maintaining chromatin in an accessible and hyperacetylated state. PNAS, 108(8), 3294-3299.

Yoshida, Y., Takahashi, K., Okita, K., Ichisaka, T., & Yamanaka, S. (2009). Hypoxia Enhances the Generation of Induced Pluripotent Stem Cells. *Stem Cell*, *5*(3), 237–241.

Young, D. W., Hassan, M. Q., Pratap, J., Galindo, M., Zaidi, S. K., Lee, S.-H., et al. (2007). Mitotic occupancy and lineage-specific transcriptional control of rRNA genes by Runx2. *Nature*, *445*(7126), 442–446.

Yu, D., Lim, J., Wang, X., Liang, X., & Xiao, G. (2017). Enhanced Construction of Gene Regulatory Networks using Hub Gene Information, *BMC Bioinformatics,* 186(18), 1–20.

Yuan, F., Fang, K.-H., Cao, S.-Y., Qu, Z.-Y., Li, Q., Krencik, R., et al. (2015). Efficient generation of region-specific forebrain neurons from human pluripotent stem cells under highly defined condition. *Scientific Reports*, 1–11.

Zaret, K. S., & Carroll, J. S. (2011). Pioneer transcription factors: establishing competence for gene expression. *Genes & Development*, *25*(21), 2227–2241.

Zhang, X., Huang, C. T., Chen, J., Pankratz, M. T., Xi, J., Li, J., et al. (2010). Pax6 Is a Human Neuroectoderm Cell Fate Determinant. *Cell Stem Cell*, *7*(1), 90–100.

Zhang, P., Sun, Y., & Ma, L. (2015). ZEB1: At the crossroads of epithelial-mesenchymal transition, metastasis and therapy resistance. *Cell Cycle*, *14*(4), 481–487.

Zirra, A., Wiethoff, S., & Patani, R. (2016). Neural Conversion and Patterning of Human Pluripotent Stem Cells: A Developmental Perspective. *Stem Cells International*, 1–14.

Ziv, O., Zaritsky, A., Yaffe, Y., Mutukula, N., Edri, R., & Elkabetz, Y. (2015). Quantitative Live Imaging of Human Embryonic Stem Cell Derived Neural Rosettes Reveals Structure-Function Dynamics Coupled to Cortical Development. *PLoS Computational Biology*, *11*(10), e1004453–21.

# APPENDICES

**Appendix. 3.1. Description of the statistical analyses results from the Q-PCR assays (CD-ROM).** The table includes the differences between means, the significance and the adjusted p-value of the two-way ANOVA test preformed to determine the significance of the difference between time points.

**Appendix. 3.2. Description of the statistical results of the differentially up-regulated transcription factors of cells induced with 2i media (CD-ROM).** The table includes the details of the statistical analyses preformed to identify differentially expressed genes. The number of genes in the list is higher that the number established of total TF differentially expressed since some TF are up-regulated between more than one time point. The data is organized in chronological order. The last comparison in the table represents the analyses between time point d0 and d8.

**Appendix. 3.3. Description of the statistical results of the differentially down-regulated transcription factors of cells induced with 2i media (CD-ROM).** The table includes the details of the statistical analyses preformed to identify differentially expressed genes. The number of genes in the list is higher that the number established of total TF differentially expressed since some TF are down-regulated between more than one time point. The data is organized in chronological order. The last comparison in the table represents the analyses between time point d0 and d8.

**Appendix. 3.4. Description of the statistical results of the differentially up-regulated transcription factors of cells induced with 2i-WNT media (CD-ROM).** The table includes the details of the statistical analyses preformed to identify differentially expressed genes. The number of genes in the list is higher that the number established of total TF differentially expressed since some TF are up-regulated between more than one time point. The data is organized in chronological order. The last comparison in the table represents the analyses between time point d0 and d8.

**Appendix. 3.5. Description of the statistical results of the differentially down-regulated transcription factors of cells induced with 2i-WNT media (CD-ROM).** Details of the statistical analyses preformed to identify differentially expressed genes. The number of genes in the list is higher that the number established of total TF differentially expressed since some TF are down-regulated between more than one time point. The data is organized in chronological order. The last comparison in the table represents the analyses between time point d0 and d8.

| GO Term | Count | P | Fold | Bonferroni | FDR |
|---|---|---|---|---|---|
| Regulation of ion transport | 73 | 1.27E-05 | 1.671 | 4.08E-02 | 2.32E-02 |
| Hindbrain development | 29 | 1.08E-05 | 2.456 | 3.48E-02 | 1.97E-02 |
| Spinal cord development | 22 | 5.46E-06 | 3.022 | 1.77E-02 | 9.93E-03 |
| Positive regulation of phosphorus metabolic | 121 | 5.23E-06 | 1.491 | 1.69E-02 | 9.50E-03 |
| Epithelial tube morphogenesis | 51 | 3.18E-06 | 1.978 | 1.04E-02 | 5.79E-03 |
| Embryonic camera-type eye development | 14 | 2.86E-06 | 4.590 | 9.30E-03 | 5.19E-03 |
| Neuron recognition | 14 | 2.86E-06 | 4.590 | 9.30E-03 | 5.19E-03 |
| Positive regulation of neurogenesis | 62 | 2.84E-06 | 1.843 | 9.25E-03 | 5.17E-03 |
| Tube morphogenesis | 55 | 2.83E-06 | 1.928 | 9.20E-03 | 5.14E-03 |
| Negative regulation of neurogenesis | 46 | 8.36E-07 | 2.165 | 2.73E-03 | 1.52E-03 |
| Cellular response to BMP stimulus | 33 | 6.18E-07 | 2.600 | 2.02E-03 | 1.12E-03 |
| Neurons | 20 | 2.88E-07 | 3.835 | 9.41E-04 | 5.24E-04 |
| Tube development | 83 | 2.81E-07 | 1.769 | 9.17E-04 | 5.10E-04 |
| Regulation of synapse organization | 29 | 2.13E-07 | 2.947 | 6.96E-04 | 3.87E-04 |
| Regulation of cell motility | 100 | 2.04E-07 | 1.677 | 6.68E-04 | 3.71E-04 |
| Synaptic signaling | 85 | 1.59E-07 | 1.778 | 5.19E-04 | 2.89E-04 |
| Cell projection morphogenesis | 115 | 8.84E-08 | 1.635 | 2.89E-04 | 1.61E-04 |
| Blood vessel morphogenesis | 74 | 7.78E-08 | 1.899 | 2.54E-04 | 1.41E-04 |
| Regulation of cell migration | 97 | 4.34E-08 | 1.748 | 1.42E-04 | 7.89E-05 |
| Telencephalon development | 48 | 9.91E-09 | 2.439 | 3.24E-05 | 1.80E-05 |
| Vasculature development | 91 | 6.31E-09 | 1.857 | 2.06E-05 | 1.15E-05 |
| Response to growth factor | 99 | 5.02E-09 | 1.810 | 1.64E-05 | 9.12E-06 |
| Synapse assembly | 35 | 2.99E-09 | 3.067 | 9.78E-06 | 5.44E-06 |
| Cell-cell adhesion via plasma-membrane | 48 | 2.91E-09 | 2.528 | 9.53E-06 | 5.30E-06 |
| Diencephalon development | 25 | 1.93E-09 | 4.033 | 6.31E-06 | 3.51E-06 |
| Regulation of neurogenesis | 108 | 6.37E-10 | 1.820 | 2.08E-06 | 1.16E-06 |
| Positive regulation of nervous system | 80 | 5.36E-10 | 2.053 | 1.75E-06 | 9.74E-07 |
| G-protein coupled receptor signaling pathway | 96 | 4.51E-10 | 1.913 | 1.47E-06 | 8.20E-07 |
| Neuron projection guidance | 51 | 8.72E-11 | 2.686 | 2.85E-07 | 1.59E-07 |
| Cell migration | 156 | 5.15E-11 | 1.671 | 1.68E-07 | 9.36E-08 |
| Cell motility | 170 | 4.84E-11 | 1.628 | 1.58E-07 | 8.80E-08 |
| Neuron projection development | 129 | 3.48E-11 | 1.789 | 1.14E-07 | 6.33E-08 |
| Epithelium development | 147 | 4.11E-12 | 1.760 | 1.35E-08 | 7.48E-09 |
| Cell morphogenesis differentiation | 122 | 3.73E-12 | 1.884 | 1.22E-08 | 6.78E-09 |
| Regulation of nervous system development | 126 | 1.64E-12 | 1.883 | 5.36E-09 | 2.98E-09 |
| Cell-cell signaling | 193 | 8.00E-13 | 1.642 | 2.61E-09 | 1.45E-09 |
| Cell surface receptor signaling pathway | 301 | 2.68E-13 | 1.464 | 8.75E-10 | 4.87E-10 |
| Eye development | 71 | 6.52E-14 | 2.596 | 2.13E-10 | 1.18E-10 |
| Brain development | 118 | 3.40E-14 | 2.040 | 1.11E-10 | 6.18E-11 |
| Neuron development | 153 | 3.14E-14 | 1.840 | 1.02E-10 | 5.69E-11 |
| Forebrain development | 77 | 1.69E-14 | 2.541 | 5.52E-11 | 3.07E-11 |
| Sensory perception | 93 | 2.69E-16 | 2.455 | 7.26E-13 | 4.00E-13 |
| Pattern specification process | 85 | 6.45E-17 | 2.618 | 3.63E-13 | 2.00E-13 |
| Central nervous system development | 157 | 9.28E-20 | 2.075 | 3.03E-16 | 1.69E-16 |
| Sensory organ development | 106 | 2.12E-20 | 2.577 | 6.92E-17 | 3.85E-17 |
| Nervous system development | 334 | 2.13E-31 | 1.811 | 6.96E-28 | 3.87E-28 |

**Appendix. 3.6. Description of the significant GO results of the differentially up-regulated genes of cells induced with 2i media.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Positive regulation of synapse assembly | 18 | 9.29E-06 | 3.410 | 4.30E-02 | 1.76E-02 |
| Cellular response to growth factor stimulus | 84 | 6.61E-06 | 1.634 | 3.07E-02 | 1.25E-02 |
| Central nervous system neuron development | 20 | 6.40E-06 | 3.223 | 2.98E-02 | 1.21E-02 |
| Synaptic signaling | 79 | 5.57E-06 | 1.673 | 2.60E-02 | 1.05E-02 |
| Trans-synaptic signaling | 79 | 5.57E-06 | 1.673 | 2.60E-02 | 1.05E-02 |
| Hindbrain development | 30 | 3.64E-06 | 2.551 | 1.71E-02 | 6.89E-03 |
| Stem cell differentiation | 22 | 3.34E-06 | 3.126 | 1.57E-02 | 6.32E-03 |
| Epithelial cell differentiation | 73 | 1.64E-06 | 1.771 | 7.70E-03 | 3.10E-03 |
| Enzyme linked receptor protein signaling | 121 | 1.33E-06 | 1.534 | 6.28E-03 | 2.52E-03 |
| Regulation of synapse assembly | 22 | 1.27E-06 | 3.299 | 6.00E-03 | 2.41E-03 |
| Tube morphogenesis | 55 | 9.48E-07 | 2.000 | 4.47E-03 | 1.79E-03 |
| Embryonic eye morphogenesis | 14 | 9.13E-07 | 5.039 | 4.30E-03 | 1.73E-03 |
| Regulation of cell motility | 96 | 6.19E-07 | 1.661 | 2.92E-03 | 1.17E-03 |
| Epithelial tube morphogenesis | 52 | 3.61E-07 | 2.111 | 1.70E-03 | 6.82E-04 |
| Neuron migration | 30 | 2.74E-07 | 2.867 | 1.29E-03 | 5.18E-04 |
| Eye morphogenesis | 32 | 2.03E-07 | 2.787 | 9.59E-04 | 3.84E-04 |
| Regulation of signal transduction | 287 | 1.82E-07 | 1.311 | 8.61E-04 | 3.45E-04 |
| Regulation of cell migration | 93 | 1.43E-07 | 1.731 | 6.75E-04 | 2.70E-04 |
| Telencephalon development | 46 | 6.32E-08 | 2.365 | 2.99E-04 | 1.20E-04 |
| Forebrain generation of neurons | 21 | 5.53E-08 | 4.049 | 2.62E-04 | 1.05E-04 |
| Camera-type eye development | 52 | 4.05E-08 | 2.255 | 1.91E-04 | 7.66E-05 |
| Diencephalon development | 23 | 3.72E-08 | 3.821 | 1.76E-04 | 7.04E-05 |
| Forebrain neuron differentiation | 20 | 3.71E-08 | 4.319 | 1.75E-04 | 7.02E-05 |
| Cell projection morphogenesis | 114 | 2.51E-08 | 1.679 | 1.18E-04 | 4.74E-05 |
| G-protein coupled receptor signaling | 93 | 1.90E-08 | 1.806 | 8.97E-05 | 3.59E-05 |
| Regulation of neurogenesis | 101 | 1.58E-08 | 1.762 | 7.46E-05 | 2.99E-05 |
| Positive regulation of nervous system | 75 | 1.14E-08 | 1.980 | 5.39E-05 | 2.16E-05 |
| Vasculature development | 90 | 6.84E-09 | 1.865 | 3.23E-05 | 1.29E-05 |
| Eye development | 60 | 5.21E-09 | 2.226 | 2.46E-05 | 9.85E-06 |
| Axon development | 75 | 4.74E-09 | 2.020 | 2.24E-05 | 8.98E-06 |
| Blood vessel development | 87 | 3.63E-09 | 1.913 | 1.72E-05 | 6.87E-06 |
| Neural crest cell development | 24 | 2.38E-09 | 4.180 | 1.12E-05 | 4.50E-06 |
| Regulation of neuron differentiation | 91 | 1.94E-09 | 1.904 | 9.15E-06 | 3.66E-06 |
| Synapse assembly | 35 | 1.65E-09 | 3.150 | 7.82E-06 | 3.13E-06 |
| Neuron projection morphogenesis | 89 | 9.06E-10 | 1.949 | 4.28E-06 | 1.71E-06 |
| Central nervous system neuron | 40 | 3.66E-10 | 3.021 | 1.73E-06 | 6.93E-07 |
| Cell morphogenesis involved in neuron | 85 | 2.06E-10 | 2.044 | 9.74E-07 | 3.90E-07 |
| Neuron projection guidance | 50 | 1.35E-10 | 2.700 | 6.36E-07 | 2.55E-07 |
| Axon guidance | 50 | 1.11E-10 | 2.713 | 5.26E-07 | 2.11E-07 |
| Regulation of nervous system development | 120 | 1.66E-11 | 1.859 | 7.84E-08 | 3.14E-08 |
| Cell surface receptor signaling pathway | 293 | 6.34E-13 | 1.463 | 2.99E-09 | 1.20E-09 |
| Neuron development | 146 | 5.40E-13 | 1.812 | 2.55E-09 | 1.02E-09 |
| Forebrain development | 75 | 8.54E-14 | 2.515 | 4.03E-10 | 1.62E-10 |
| Sensory organ morphogenesis | 59 | 1.01E-14 | 3.034 | 4.77E-11 | 1.91E-11 |
| Brain development | 118 | 6.16E-15 | 2.092 | 2.89E-11 | 1.15E-11 |
| Central nervous system development | 149 | 1.80E-17 | 2.014 | 8.51E-14 | 3.41E-14 |
| Sensory organ development | 100 | 7.02E-18 | 2.477 | 3.32E-14 | 1.33E-14 |
| Neurogenesis | 221 | 1.81E-21 | 1.860 | 8.58E-18 | 3.43E-18 |
| Neuron differentiation | 200 | 1.36E-22 | 1.980 | 6.45E-19 | 2.58E-19 |
| Nervous system development | 320 | 6.13E-29 | 1.789 | 2.90E-25 | 1.16E-25 |

**Appendix. 3.7. Description of the significant GO results of the differentially up-regulated genes of cells induced with 2i-WNT media**. Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enrich | Bonferroni | FDR |
|---|---|---|---|---|---|
| Leukocyte activation | 103 | 1.43E-05 | 1.508 | 4.91E-02 | 2.62E- |
| Inflammatory response | 85 | 1.41E-05 | 1.583 | 4.85E-02 | 2.59E- |
| Regulation of intracellular signal | 246 | 9.96E-06 | 1.286 | 3.44E-02 | 1.83E- |
| Response to decreased oxygen levels | 61 | 8.24E-06 | 1.773 | 2.86E-02 | 1.51E- |
| Chemotaxis | 87 | 5.77E-06 | 1.608 | 2.01E-02 | 1.06E- |
| NADH regeneration | 13 | 5.59E-06 | 4.413 | 1.95E-02 | 1.02E- |
| Calcium ion transport into cytosol | 31 | 5.53E-06 | 2.387 | 1.93E-02 | 1.01E- |
| ATP generation from ADP | 22 | 3.55E-06 | 2.987 | 1.24E-02 | 6.50E- |
| Positive regulation of ion transport | 46 | 3.06E-06 | 2.033 | 1.07E-02 | 5.60E- |
| Intracellular signal transduction | 370 | 2.90E-06 | 1.230 | 1.02E-02 | 5.32E- |
| Sensory perception | 85 | 2.69E-06 | 1.649 | 9.40E-03 | 4.92E- |
| Signal transduction by protein | 143 | 9.26E-07 | 1.475 | 3.25E-03 | 1.70E- |
| Regulation of secretion | 110 | 5.37E-07 | 1.589 | 1.89E-03 | 9.84E- |
| Regulation of angiogenesis | 49 | 4.73E-07 | 2.103 | 1.66E-03 | 8.68E- |
| Regulation of ion homeostasis | 45 | 4.00E-07 | 2.197 | 1.41E-03 | 7.33E- |
| Blood circulation | 92 | 1.66E-07 | 1.713 | 5.82E-04 | 3.03E- |
| Regulation of vasculature development | 54 | 1.43E-07 | 2.090 | 5.02E-04 | 2.62E- |
| Synaptic signaling | 107 | 1.35E-07 | 1.644 | 4.74E-04 | 2.47E- |
| Ion transmembrane transport | 155 | 1.22E-07 | 1.496 | 4.30E-04 | 2.24E- |
| Positive regulation of phosphorus metabolic | 164 | 8.41E-08 | 1.485 | 2.96E-04 | 1.54E- |
| MAPK cascade | 142 | 8.28E-08 | 1.537 | 2.91E-04 | 1.52E- |
| Extracellular matrix organization | 68 | 8.27E-08 | 1.931 | 2.91E-04 | 1.52E- |
| Regulation of response to external stimulus | 51 | 5.54E-08 | 2.202 | 1.95E-04 | 1.02E- |
| Angiogenesis | 81 | 4.92E-08 | 1.833 | 1.73E-04 | 9.01E- |
| Negative regulation of cell proliferation | 115 | 4.69E-08 | 1.642 | 1.65E-04 | 8.59E- |
| Transmembrane transport | 204 | 1.29E-08 | 1.448 | 4.55E-05 | 2.37E- |
| Regulation of phosphorus metabolic process | 245 | 1.20E-08 | 1.395 | 4.21E-05 | 2.19E- |
| Muscle contraction | 72 | 1.16E-08 | 1.977 | 4.08E-05 | 2.13E- |
| Regulation of cell migration | 125 | 6.81E-09 | 1.655 | 2.40E-05 | 1.25E- |
| Oxoacid metabolic process | 144 | 5.52E-09 | 1.596 | 1.94E-05 | 1.01E- |
| Response to other organism | 116 | 5.51E-09 | 1.699 | 1.94E-05 | 1.01E- |
| Positive regulation of cell migration | 85 | 7.57E-10 | 1.959 | 2.66E-06 | 1.39E- |
| Vasculature development | 117 | 6.18E-10 | 1.755 | 2.17E-06 | 1.13E- |
| Regulation of cell motility | 137 | 2.87E-10 | 1.688 | 1.01E-06 | 5.25E- |
| Cell-cell signaling | 237 | 6.87E-11 | 1.481 | 2.42E-07 | 1.26E- |
| Cell motility | 216 | 4.87E-11 | 1.520 | 1.71E-07 | 8.92E- |
| Positive regulation of cell motility | 90 | 4.16E-11 | 2.019 | 1.46E-07 | 7.63E- |
| Regulation of ion transport | 111 | 3.55E-11 | 1.867 | 1.25E-07 | 6.51E- |
| Small molecule biosynthetic process | 101 | 3.03E-11 | 1.939 | 1.07E-07 | 5.56E- |
| G-protein coupled receptor signaling | 124 | 1.64E-11 | 1.816 | 5.76E-08 | 3.00E- |
| Cell migration | 203 | 1.92E-12 | 1.598 | 6.75E-09 | 3.51E- |
| Ion transport | 254 | 5.98E-18 | 1.658 | 2.10E-14 | 1.10E- |

**Appendix. 3.8. Description of the significant GO results of the differentially down-regulated genes of cells induced with 2i media**. Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enrich | Bonferroni | FDR |
|---|---|---|---|---|---|
| Mesenchyme development | 49 | 1.37E-05 | 1.884 | 4.78E-02 | 2.52E-02 |
| Coenzyme metabolic process | 65 | 1.20E-05 | 1.718 | 4.19E-02 | 2.20E-02 |
| T cell activation | 71 | 1.10E-05 | 1.677 | 3.84E-02 | 2.02E-02 |
| Regulation of signal transduction | 380 | 1.01E-05 | 1.211 | 3.53E-02 | 1.85E-02 |
| Regulation of sequestering of calcium ion | 28 | 1.00E-05 | 2.453 | 3.52E-02 | 1.84E-02 |
| Positive regulation of cell adhesion | 69 | 6.62E-06 | 1.716 | 2.33E-02 | 1.21E-02 |
| Positive regulation of ion transport | 46 | 6.60E-06 | 1.981 | 2.33E-02 | 1.21E-02 |
| Epithelium development | 162 | 5.31E-06 | 1.395 | 1.87E-02 | 9.74E-03 |
| Response to cAMP | 27 | 4.28E-06 | 2.608 | 1.51E-02 | 7.85E-03 |
| ATP generation from ADP | 23 | 2.86E-06 | 2.937 | 1.02E-02 | 5.26E-03 |
| Leukocyte activation | 109 | 2.59E-06 | 1.544 | 9.18E-03 | 4.75E-03 |
| Regulation of programmed cell death | 216 | 1.66E-06 | 1.346 | 5.91E-03 | 3.05E-03 |
| Negative regulation of cell motility | 51 | 1.64E-06 | 1.991 | 5.83E-03 | 3.01E-03 |
| Leukocyte cell-cell adhesion | 58 | 1.49E-06 | 1.900 | 5.28E-03 | 2.73E-03 |
| Response to cytokine | 124 | 1.18E-06 | 1.519 | 4.18E-03 | 2.16E-03 |
| Regulation of angiogenesis | 50 | 8.73E-07 | 2.047 | 3.11E-03 | 1.60E-03 |
| Regulation of cell-cell adhesion | 70 | 5.43E-07 | 1.825 | 1.94E-03 | 9.97E-04 |
| Transmembrane transport | 203 | 3.16E-07 | 1.393 | 1.13E-03 | 5.80E-04 |
| Ion transmembrane transport | 158 | 2.92E-07 | 1.470 | 1.04E-03 | 5.35E-04 |
| NADH regeneration | 15 | 2.65E-07 | 4.709 | 9.43E-04 | 4.85E-04 |
| Regulation of vasculature development | 55 | 2.55E-07 | 2.041 | 9.09E-04 | 4.68E-04 |
| Calcium ion transport into cytosol | 35 | 1.97E-07 | 2.560 | 7.01E-04 | 3.61E-04 |
| Leukocyte cell-cell adhesion | 83 | 1.44E-07 | 1.777 | 5.12E-04 | 2.64E-04 |
| Regulation of tube size | 37 | 1.43E-07 | 2.512 | 5.11E-04 | 2.63E-04 |
| Cellular homeostasis | 137 | 1.39E-07 | 1.538 | 4.97E-04 | 2.56E-04 |
| Inflammatory response | 96 | 1.08E-07 | 1.706 | 3.87E-04 | 1.99E-04 |
| MAPK cascade | 145 | 4.69E-08 | 1.543 | 1.67E-04 | 8.61E-05 |
| Cardiovascular system development | 164 | 3.26E-08 | 1.505 | 1.16E-04 | 5.98E-05 |
| Regulation of blood circulation | 67 | 2.79E-08 | 1.995 | 9.94E-05 | 5.12E-05 |
| Regulation of ion homeostasis | 49 | 2.67E-08 | 2.293 | 9.52E-05 | 4.90E-05 |
| Angiogenesis | 85 | 1.90E-08 | 1.840 | 6.76E-05 | 3.48E-05 |
| Positive regulation of cell migration | 86 | 8.99E-10 | 1.946 | 3.20E-06 | 1.65E-06 |
| Regulation of phosphorus metabolic | 256 | 6.07E-10 | 1.426 | 2.16E-06 | 1.11E-06 |
| Single-organism transport | 470 | 5.43E-10 | 1.272 | 1.94E-06 | 9.97E-07 |
| Chemotaxis | 104 | 2.66E-10 | 1.852 | 9.47E-07 | 4.87E-07 |
| Positive regulation of cell motility | 91 | 7.04E-11 | 1.993 | 2.51E-07 | 1.29E-07 |
| Protein coupled receptor signaling | 133 | 5.00E-12 | 1.802 | 1.78E-08 | 9.17E-09 |
| Regulation of cell motility | 145 | 8.56E-05 | 1.751 | 1.79E-08 | 9.22E-09 |
| Extracellular matrix organization | 80 | 2.60E-12 | 2.232 | 9.28E-09 | 4.78E-09 |
| Cell-cell signaling | 255 | 6.45E-14 | 1.549 | 2.30E-10 | 1.18E-10 |
| Cell migration | 215 | 2.18E-14 | 1.641 | 7.76E-11 | 3.99E-11 |
| Ion transport | 256 | 6.20E-16 | 1.607 | 2.37E-12 | 1.22E-12 |

**Appendix. 3.9. Description of the significant GO results of the differentially down-regulated genes of cells induced with 2i-WNT media.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Cell activation | 88 | 2.21E-05 | 1.573 | 2.54E-02 | 3.56E-02 |
| Single organism cell adhesion | 78 | 2.16E-05 | 1.629 | 2.48E-02 | 3.47E-02 |
| Single organismal cell-cell adhesion | 74 | 1.91E-05 | 1.660 | 2.20E-02 | 3.08E-02 |
| Positive regulation of cell adhesion | 47 | 1.13E-05 | 1.986 | 1.30E-02 | 1.82E-02 |
| Cell-cell adhesion | 109 | 8.48E-06 | 1.523 | 9.81E-03 | 1.37E-02 |
| Transmembrane transport | 121 | 7.42E-06 | 1.489 | 8.59E-03 | 1.20E-02 |
| Defense response | 136 | 5.40E-06 | 1.457 | 6.25E-03 | 8.69E-03 |
| Sulfur compound metabolic process | 47 | 4.56E-06 | 2.055 | 5.28E-03 | 7.34E-03 |
| Regulation of cell proliferation | 139 | 4.15E-06 | 1.457 | 4.81E-03 | 6.68E-03 |
| Transport | 351 | 4.07E-06 | 1.224 | 4.72E-03 | 6.56E-03 |
| Extracellular structure organization | 46 | 6.01E-07 | 2.234 | 6.99E-04 | 9.68E-04 |
| Single-organism transport | 281 | 3.27E-08 | 1.338 | 3.80E-05 | 5.27E-05 |
| Angiogenesis | 45 | 2.56E-04 | 1.774 | 5.09E-01 | 4.56E-01 |

**Appendix. 3.10. Description of the GO results of the genes assigned to the profile 31 of cells induced with 2i media.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Organophosphate metabolic process | 111 | 3.81E-05 | 1.464 | 4.42E-02 | 6.15E-02 |
| Positive regulation of cell adhesion | 49 | 3.79E-05 | 1.860 | 4.39E-02 | 6.12E-02 |
| Defence response | 144 | 3.76E-05 | 1.386 | 4.35E-02 | 6.06E-02 |
| Regulation of signal transduction | 237 | 3.71E-05 | 1.269 | 4.30E-02 | 5.99E-02 |
| Angiogenesis | 52 | 2.86E-05 | 1.840 | 3.34E-02 | 4.62E-02 |
| Phosphorus metabolic process | 265 | 2.81E-05 | 1.253 | 3.28E-02 | 4.54E-02 |
| Response to hypoxia | 40 | 2.79E-05 | 2.044 | 3.26E-02 | 4.51E-02 |
| Epithelial cell proliferation | 46 | 2.52E-05 | 1.937 | 2.94E-02 | 4.07E-02 |
| Regulation of cell communication | 262 | 1.36E-05 | 1.268 | 1.60E-02 | 2.20E-02 |
| Cellular response to chemical stimulus | 234 | 1.34E-05 | 1.291 | 1.58E-02 | 2.17E-02 |
| Regulation of body fluid levels | 62 | 8.59E-06 | 1.801 | 1.01E-02 | 1.39E-02 |
| Organic acid metabolic process | 98 | 6.01E-06 | 1.579 | 7.10E-03 | 9.71E-03 |
| Response to hormone | 94 | 4.51E-06 | 1.608 | 5.33E-03 | 7.28E-03 |
| Cell-cell adhesion | 122 | 1.68E-06 | 1.531 | 1.99E-03 | 2.71E-03 |
| Cell communication | 506 | 1.07E-06 | 1.178 | 1.27E-03 | 1.73E-03 |
| Single-organism biosynthetic process | 139 | 5.02E-07 | 1.516 | 5.95E-04 | 8.11E-04 |
| Transmembrane transport | 139 | 2.29E-07 | 1.536 | 2.71E-04 | 3.70E-04 |
| Response to organic substance | 260 | 1.12E-07 | 1.346 | 1.32E-04 | 1.80E-04 |
| Organonitrogen compound metabolic process | 215 | 4.86E-08 | 1.414 | 5.76E-05 | 7.85E-05 |
| Small molecule metabolic process | 195 | 2.31E-08 | 1.460 | 2.74E-05 | 3.73E-05 |
| Regulation of cell proliferation | 164 | 9.69E-09 | 1.544 | 1.15E-05 | 1.56E-05 |
| Transport | 409 | 1.50E-09 | 1.282 | 1.78E-06 | 2.42E-06 |
| Single-organism transport | 326 | 1.26E-11 | 1.395 | 1.49E-08 | 2.03E-08 |

**Appendix. 3.11. Description of the GO results genes assigned to the profile 31 of cells induced with 2i-WNT media.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Positive regulation of synapse assembly | 11 | 3.92E-05 | 5.251 | 0.03750178 | 0.061720 |
| Embryonic morphogenesis | 39 | 3.92E-05 | 2.054 | 0.03749994 | 0.061717 |
| Neural precursor cell proliferation | 16 | 2.47E-05 | 3.730 | 0.02381975 | 0.038933 |
| Positive regulation of developmental | 63 | 2.10E-05 | 1.740 | 0.02027063 | 0.033073 |
| Positive regulation of cell differentiation | 53 | 1.14E-05 | 1.895 | 0.01107415 | 0.017985 |
| Telencephalon development | 23 | 9.45E-06 | 2.994 | 0.00917583 | 0.014888 |
| Inner ear development | 20 | 7.52E-06 | 3.379 | 0.00730600 | 0.011843 |
| Organ morphogenesis | 61 | 4.70E-06 | 1.849 | 0.00457674 | 0.007409 |
| Cell development | 102 | 4.49E-06 | 1.558 | 0.00436383 | 0.007063 |
| Animal organ development | 152 | 3.95E-07 | 1.461 | 3.85E-04 | 6.22E-04 |
| Sensory organ development | 42 | 1.40E-07 | 2.491 | 1.36E-04 | 2.20E-04 |
| Head development | 53 | 1.15E-07 | 2.214 | 1.12E-04 | 1.81E-04 |
| Synapse organization | 27 | 5.61E-08 | 3.515 | 5.47E-05 | 8.83E-05 |
| Forebrain development | 35 | 4.74E-08 | 2.916 | 4.62E-05 | 7.46E-05 |
| Brain development | 53 | 2.17E-08 | 2.330 | 2.12E-05 | 3.42E-05 |
| Cellular developmental process | 191 | 1.33E-08 | 1.439 | 1.30E-05 | 2.10E-05 |
| Anatomical structure morphogenesis | 143 | 8.43E-10 | 1.634 | 8.22E-07 | 1.33E-06 |
| Pattern specification process | 43 | 5.99E-10 | 2.966 | 5.84E-07 | 9.43E-07 |
| Cell projection organization | 87 | 4.41E-10 | 2.011 | 4.30E-07 | 6.95E-07 |
| Cellular component morphogenesis | 89 | 1.79E-10 | 2.026 | 1.74E-07 | 2.82E-07 |
| System development | 213 | 2.58E-11 | 1.493 | 2.51E-08 | 4.06E-08 |
| Multicellular organism development | 240 | 9.55E-13 | 1.480 | 9.31E-10 | 1.50E-09 |

**Appendix. 3.12. Description of the significant GO results of the genes assigned to the profile 65 of cells induced with 2i media.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Inner ear development | 19 | 2.92E-05 | 3.193 | 0.02676116 | 0.045689 |
| Cellular component assembly involved in | 27 | 6.40E-06 | 2.747 | 0.00592534 | 0.010011 |
| Forebrain development | 31 | 4.45E-06 | 2.568 | 0.00412107 | 0.006957 |
| Head development | 49 | 4.04E-06 | 2.035 | 0.00374273 | 0.006317 |
| Synapse organization | 24 | 3.09E-06 | 3.108 | 0.0028653 | 0.004834 |
| Anatomical structure morphogenesis | 130 | 2.30E-06 | 1.478 | 0.00213067 | 0.003593 |
| Brain development | 48 | 2.23E-06 | 2.099 | 0.00207310 | 0.003496 |
| Cellular component morphogenesis | 87 | 1.22E-09 | 1.970 | 1.14E-06 | 1.91E-06 |
| Cell projection organization | 91 | 1.94E-11 | 2.092 | 1.80E-08 | 3.03E-08 |
| Neural precursor cell proliferation | 14 | 3.79E-04 | 3.246 | 0.29669337 | 0.591215 |
| Telencephalon development | 20 | 2.85E-04 | 2.590 | 0.23276794 | 0.445407 |

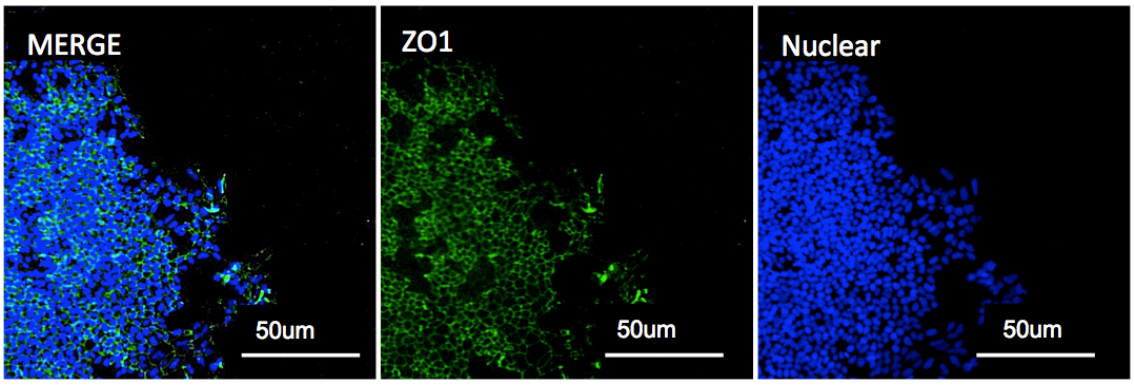
**Appendix. 3.13. Description of the significant GO results genes assigned to the profile 65 of cells induced with 2i-WNT media.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

**Appendix. 3.14. Description of the statistical results of genes differentially up-regulated in cells induced with 2i media compare with cells induced with 2i-WNT media (CD-ROM).** Gene expression statistical analyses between cells induced with 2i and 2i-WNT media. The data has being organized in chronological order of the time points compared and includes d4 and d8, which were the days with significant differential expression between treatments.

**Appendix. 3.15. Description of the statistical results of genes differentially down-regulated in cells induced with 2i-WNT compare with cells induced with 2i- media (CD-ROM).** Gene expression statistical analyses between cells induced with 2i-WNT and 2i media. The data has being organized in chronological order of the time points compared and includes d4 and d8, which were the days with significant differential expression between treatments.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Anatomical morphogenesis morphogenesis | 31 | 4.20E-05 | 2.111 | 0.02733981 | 0.062821 |
| Tissue development | 24 | 4.07E-05 | 2.501 | 0.02648550 | 0.060831 |
| Organ development | 35 | 1.98E-05 | 2.034 | 0.01297799 | 0.029608 |
| System development | 47 | 2.76E-07 | 1.981 | 1.82E-04 | 4.14E-04 |

**Appendix. 3.16. Description of the GO results of the differentially up-regulated genes of cells induced with 2i media compared with cells induced with 2i-WNT media at d4 and d8.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.




**Appendix. 3.17. ZO1 immunofluorescence.** The immunofluorescence analyses of the rosette marker ZO1 in cell induced with 2i after time point d8. The nuclei are stained with Hoescht 33342.

**Appendix. 4.1. Gene regulatory network (CD-ROM).** Gene expression statistical analyses between cells induced with 2i-WNT and 2i media. The data has been organized in chronological order of the time points compared and includes d4 and d8, which were the days with significant differential expression between treatments.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Developmental process | 97 | 1.40E-19 | 2.142 | 3.22E-18 | 1.11E-16 |
| Response to stimulus | 108 | 3.91E-13 | 1.612 | 8.99E-12 | 3.10E-10 |
| Locomotion | 40 | 2.91E-11 | 3.251 | 6.69E-10 | 2.31E-08 |
| Reproductive process | 32 | 8.68E-08 | 2.898 | 2.00E-06 | 6.88E-05 |
| Rhythmic process | 15 | 2.02E-07 | 5.980 | 4.64E-06 | 1.60E-04 |
| Immune system process | 42 | 1.96E-06 | 2.119 | 4.50E-05 | 0.001550 |
| Biological adhesion | 32 | 8.98E-06 | 2.334 | 2.07E-04 | 0.007119 |
| Localization | 69 | 1.20E-04 | 1.461 | 0.002767 | 0.095469 |
| Cellular component organization or biogenesis | 72 | 1.50E-04 | 1.429 | 0.003451 | 0.119067 |

**Appendix. 4.2. Description of the GO results of the genes in the regulatory network.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Cell proliferation | 15 | 1.15E-05 | 3.702 | 2.56E-02 | 2.01E-02 |
| Intrinsic apoptotic signaling pathway | 8 | 3.81E-06 | 11.55 | 8.53E-03 | 6.64E-03 |
| Programmed cell death | 16 | 3.61E-06 | 3.767 | 8.09E-03 | 6.29E-03 |
| Cell death | 17 | 1.26E-06 | 3.784 | 2.84E-03 | 2.20E-03 |
| Regulation of response to stress | 15 | 1.41E-07 | 5.334 | 3.16E-04 | 2.45E-04 |
| Regulation of cell proliferation | 17 | 2.26E-08 | 5.049 | 5.09E-05 | 3.95E-05 |

**Appendix. 4.3. Description of the GO results of the MYC module.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| RNA biosynthetic process | 8 | 7.98E-05 | 3.847 | 4.43E-02 | 1.17E-01 |
| Nervous system development | 7 | 6.49E-05 | 5.895 | 3.62E-02 | 9.51E-02 |
| Regulation of RNA metabolic process | 8 | 5.91E-05 | 4.015 | 3.30E-02 | 8.65E-02 |
| Neural tube development | 4 | 5.67E-05 | 41.80 | 3.17E-02 | 8.30E-02 |
| Forebrain regionalization | 3 | 4.98E-05 | 236.7 | 2.79E-02 | 7.30E-02 |
| Regulation of RNA biosynthetic process | 8 | 4.64E-05 | 4.156 | 2.60E-02 | 6.81E-02 |
| Transcription, DNA-templated | 8 | 4.56E-05 | 4.166 | 2.56E-02 | 6.68E-02 |
| Regulation of transcription, DNA-templated | 8 | 4.29E-05 | 4.203 | 2.41E-02 | 6.29E-02 |
| Central nervous system development | 6 | 1.57E-05 | 12.31 | 8.87E-03 | 2.30E-02 |
| Forebrain development | 5 | 1.15E-05 | 25.62 | 6.49E-03 | 1.68E-02 |
| Dorsal/ventral pattern formation | 4 | 9.07E-06 | 76.98 | 5.14E-03 | 1.33E-02 |
| Sensory organ morphogenesis | 5 | 2.06E-06 | 39.45 | 1.17E-03 | 3.02E-03 |
| Positive regulation of gene expression | 8 | 1.98E-07 | 9.050 | 1.13E-04 | 2.91E-04 |
| Sensory organ development | 8 | 4.15E-11 | 30.20 | 2.36E-08 | 6.08E-08 |

**Appendix. 4.4. Description of the GO results of the PAX6 module.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Blood vessel development | 7 | 4.62E-05 | 9.402 | 4.05E-02 | 7.20E-02 |
| RNA biosynthetic process | 15 | 1.84E-05 | 2.885 | 1.63E-02 | 2.87E-02 |
| Regulation of RNA metabolic process | 15 | 1.08E-05 | 3.011 | 9.59E-03 | 1.68E-02 |
| Regulation of RNA biosynthetic process | 15 | 7.01E-06 | 3.117 | 6.25E-03 | 1.09E-02 |
| Transcription, DNA-templated | 15 | 6.78E-06 | 3.125 | 6.04E-03 | 1.06E-02 |
| Regulation of transcription, DNA-templated | 15 | 6.08E-06 | 3.152 | 5.42E-03 | 9.47E-03 |
| Negative regulation of RNA metabolic process | 10 | 5.82E-06 | 6.238 | 5.19E-03 | 9.07E-03 |
| Negative regulation of RNA biosynthetic process | 10 | 4.14E-06 | 6.501 | 3.69E-03 | 6.44E-03 |
| Negative regulation of transcription, DNA- | 10 | 2.84E-06 | 6.802 | 2.53E-03 | 4.42E-03 |
| Regulation of gene expression | 17 | 8.34E-07 | 2.949 | 7.46E-04 | 1.30E-03 |
| Negative regulation of gene expression | 12 | 2.75E-07 | 6.124 | 2.46E-04 | 4.29E-04 |
| Positive regulation of RNA biosynthetic process | 12 | 1.08E-07 | 6.704 | 9.69E-05 | 1.69E-04 |
| Positive regulation of transcription, DNA- | 12 | 9.01E-08 | 6.824 | 8.06E-05 | 1.40E-04 |
| Positive regulation of RNA metabolic process | 13 | 1.05E-08 | 7.002 | 9.42E-06 | 1.64E-05 |
| Positive regulation of gene expression | 14 | 4.96E-09 | 6.335 | 4.44E-06 | 7.73E-06 |

**Appendix. 4.5. Description of the GO results of the SMAD2 module.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.

| GO Term | Count | P Value | Fold Enric | Bonferroni | FDR |
|---|---|---|---|---|---|
| Regulation of cell cycle G1/S phase transition | 5 | 6.64E-05 | 21.02 | 3.65E-02 | 9.71E-02 |
| DNA damage checkpoint | 5 | 5.94E-05 | 21.62 | 3.27E-02 | 8.69E-02 |
| Nucleic acid metabolic process | 17 | 4.80E-05 | 2.317 | 2.65E-02 | 7.02E-02 |
| DNA recombination | 6 | 2.05E-05 | 16.17 | 1.14E-02 | 3.00E-02 |
| Cell cycle checkpoint | 6 | 1.68E-05 | 16.85 | 9.37E-03 | 2.46E-02 |
| Regulation of mitotic cell cycle | 11 | 4.67E-10 | 14.47 | 2.62E-07 | 6.84E-07 |
| DNA metabolic process | 14 | 3.73E-11 | 9.641 | 2.09E-08 | 5.45E-08 |
| DNA replication initiation | 7 | 6.10E-12 | 131.5 | 3.41E-09 | 8.92E-09 |
| DNA replication | 12 | 4.37E-14 | 27.02 | 2.45E-11 | 6.40E-11 |
| G1/S transition of mitotic cell cycle | 12 | 2.86E-15 | 34.52 | 1.62E-12 | 4.22E-12 |
| Mitotic cell cycle process | 17 | 5.13E-16 | 12.11 | 3.11E-13 | 8.10E-13 |
| Cell cycle phase transition | 16 | 1.54E-17 | 18.64 | 8.62E-15 | 2.25E-14 |
| Mitotic cell cycle phase transition | 16 | 6.88E-18 | 19.67 | 3.85E-15 | 1.01E-14 |

**Appendix. 4.6. Description of the GO results of the CDC6 module.** Details of the GO analyses including the terms, the number of genes per term (count), the p-value, the adjusted p-value (Bonferroni) and the false discovery rate.