

Integrative genomic analysis of Neurogenin2 reprogramming of human iPSCs



Muhammad Kaiser Bin Abdul Karim

Department of Clinical Neurosciences

University of Cambridge

This thesis is submitted for the degree of

Doctor of Philosophy

Declaration

This thesis is the result of my own work and includes nothing which is the outcome of work done in collaboration except as declared in the Acknowledgements section and specified in the text. It is not substantially the same as any that I have submitted, or, is being concurrently submitted for a degree or diploma or other qualification at the University of Cambridge or any other University or similar institution. I further state that no substantial part of my thesis has already been submitted, or, is being concurrently submitted for any such degree, diploma or other qualification at the University of Cambridge or any other University or similar institution. It does not exceed the prescribed word limit for the Faculties of Clinical and Veterinary Medicine Degree Committee.

Integrative genomic analysis of Neurogenin2 reprogramming of human iPSCs

Muhammad Kaiser Bin Abdul Karim

Abstract

Direct cell reprogramming is a rapidly growing field that has challenged traditional concepts of cellular identity. The expression of Neurogenin 2 (NGN2) results in rapid reprogramming of human pluripotent stem cells (PSCs) into functional excitatory neurons. My lab has previously demonstrated that gene targeting the components of a Tet-On system overexpressing NGN2 into two separate safe harbour sites overcome gene silencing and results in optimised transgene expression in hiPSCs (Opti-Ox), and consequently yields highly homogenous cultures of neurons within less than four days. The mechanisms that mediate this remarkable cellular metamorphosis however remain poorly understood. To explore this, I first sought to establish a protocol for long-term culture of electrophysiologically functional NGN2 induced neurons (iNs). This was achieved by co-culturing with primary rat-derived glial cells, enriched for astrocytes. iNs demonstrated functional activity around two weeks post-induction and by three weeks, formed networks of synchronous bursts that were mediated by glutamatergic AMPA-receptors. Based on these phenotypical hallmarks, I designed a time course-based genomic analysis that investigated the transcriptional and chromatin accessibility states of cells undergoing NGN2 reprogramming. Bulk RNA and ATAC-sequencing were performed at Day 0, 6h, 12h, Day 1, 36h, Day 2, Day 3, Day 4, Day 14 and Day 21 post-NGN2 induction. I also performed a scRNA-seq of the same time points, except for 6- and 36-hours post-induction, to investigate any heterogeneity in the time course and complement findings from the bulk RNA-seq. In order to differentiate direct from indirect NGN2 down-stream effectors, ChIP-seq of NGN2 binding acquired on day 1 after induction was subsequently overlaid with bulk-seq data. In addition, to study the genome-wide effects of astrocyte-enriched rat glia, I performed the same assays on neurons co-cultured with glia at Day 4, 14 and 21.

Together, they revealed rapid transcriptional and accessibility changes induced by NGN2 within 6 hours of reprogramming. The subsequent events show a stark similarity to the familiar stages of neurogenesis found in development or conventional differentiation protocols - shutting down of non-neuronal networks, in this case pluripotency, establishment of neuronal commitment in an NSC-like stage by Day 1 post-induction, cell cycle exit by Day 3 or Day 4 and subsequent onset of neuronal differentiation, followed by neuronal maturation. Up until now, this entire process was believed to be a highly homogenous occurrence, but findings from the scRNAseq analysis

showed that in addition to glutamatergic neurons, our NGN2 iNeurons are made up of two additional types of neurons: cholinergic neurons with a visceral motor phenotype and neurons with a hybrid profile of cholinergic and glutamatergic transcription. Comparison with neurons co-cultured with glia found an enrichment for synaptic genes and ontologies. Specifically, there was an enrichment for neuronal activity regulated genes in the post-synaptic compartment. Candidate transcription factors crucial to these processes and the states described above were identified from the transcriptional and epigenomic datasets. It is likely that some of these factors could potentially be crucial regulators of neuronal differentiation and function, not only in this protocol, but in development as well. Therefore, one major aim for the future would be to further explore the role these genes play in NGN2 reprogramming through gene knockdown or overexpression experiments along with investigation of their genomic binding sites. Ultimately, this could also lead to the discovery of new reprogramming strategies for producing neurons with enhanced maturity and functionality without the need of human or rodent glia. By uncovering broad, yet essential neuronal processes such as neuronal fate commitment and maturation, the dataset I have generated can be used as a useful resource for studying these processes and designing relevant disease models using this simple, yet robust and efficient protocol for generating functional excitatory human neurons.

Acknowledgements

First and foremost, I would like to thank my supervisor, Mark Kotter, for his unwavering support and mentorship throughout my PhD. Thanks to him and the collaborations he has fostered for this study, I had the wonderful opportunity to work with exceptional scientists and challenge myself to learn more than I could have ever imagined for my PhD.

This work would not have been possible without these collaborations. Firstly, I would like to thank Andrew Bassett at the Wellcome Sanger Institute for his overall help and guidance in carrying out the next-generation sequencing experiments. Specifically, his senior scientist, Erica Bello for her guidance in preparing samples for ChIP-sequencing, and Sara Cooper for her guidance in preparing samples for bulk RNA-seq, ATAC-seq and single-cell RNA-seq. Also affiliated to their group is Julie Jerber at the Wellcome Sanger Insititute, who I have to thank for her help in optimising my cell dissociation protocol for scRNAseq. Second, I would like to thank Emmanouil Metzakopian (University of Cambridge) and his brilliant bioinformatician, Nikolaos Patikas for his tremendous support in pre-processing the bulk RNA-seq and sc-RNAseq datasets, so that I can perform all the downstream analysis. Third group I would like to thank is the lab of John O'Neill (LMB), specifically David Tourigny, for our work together on analysing the electrophysiological properties of NGN2 iNeurons using their multi-electrode array system. Fourth group I would like to thank is Grant Belgard and this talented team at The Bioinformatics CRO (USA), for his advice on analysing the bulk RNA-seq, bulk ATAC-seq and ChIP-seq datasets and their tremendous direct contribution to analysing the datasets according to the questions and specifications I asked for. Next, I would like to thank the wonderful family of academics that I shared space with at the Anne McLaren Laboratory for Regenerative Medicine. Firstly, members of my own lab. Joana Tavares, for all the high-quality rat glial cultures I have used for this study. I am truly grateful for the never-ending hard work she has put in in making sure our lab had a supply of these cells for our co-culture experiments. Second, I would like to thank Koby Baranes for his diligence and guidance in helping with the gene targeting and molecular cloning steps. I would also like to thank the remaining members (and former members) of my lab, Elena Patili, Arko Paul, Aicha Masralli, Nataly Martynyuk, Mattias Pawlowski, Syed Yassir and Ana Amaral for their advice and support during this degree. Members from other labs that I am grateful for their advice and support are Daniel Ortmann and Rute Tomaz (Ludovic Vallier lab), William Bernard (Sanjay Sinha lab), Peter Kirwan (Florian Merkle lab) and Venkat Pisupati (Roger Barker lab).

Next, I would like to thank my degree funders Yayasan Daya Diri and Cambridge Trust for honouring me with a wonderful scholarship that has made my dream a reality.

Lastly, I would like to thank my family for their love and support. My siblings, for their love and words of encouragement when I needed it. My mother, who throughout my academic pursuits has shown the depth in love a parent can have for their child. My father, for inspiring me to be the scientist I am today and the ambition that will drive me for the rest of my life. Last but not least, my wonderful wife, Zahabiya, who has stood by patiently to support me at every difficult moment since we met during this degree. I can't imagine completing this degree without her support. I am forever grateful to her and everyone who has been with me on this incredible, life-changing journey.

Table of contents

Declaration.....	2
Abstract.....	3
Acknowledgements.....	5
Table of contents.....	7
List of figures.....	10
List of tables.....	13
1 Introduction.....	14
1.1 Cell reprogramming.....	14
1.2 Induced Neurons.....	15
1.3 OPTi-OX: Optimised inducible Overexpression System.....	18
1.4 NGN2 iNeurons.....	20
1.4.1 NGN2 – A Neurogenin bHLH transcription factor.....	20
1.5 Gaps in knowledge.....	26
1.6 Outlook.....	27
2 Methods.....	29
2.1 Cell culture.....	29
2.1.1 Maintenance of pluripotent stem cells.....	29
2.1.2 NGN2 iNeurons.....	29
2.1.3 Primary rat astrocyte culture.....	30
2.1.4 iN-rat astrocyte co-culture.....	30
2.1.5 iN culture on MEAs.....	31
2.2 Molecular cloning.....	31
2.3 Gene targeting.....	31
2.4 PCR screening of clones.....	32
2.5 Immunocytochemistry.....	32
2.6 Multi-electrode array (MEA) recordings.....	35
2.7 MEA data analysis.....	35

2.8 Karyotyping	35
2.9 Bulk RNA-sequencing	36
2.9.1 Bulk RNA-sequencing data analysis	36
2.9.2 Bulk RNA-seq data analysis of neurons co-cultured with glia against neurons cultured without glia	37
2.10 Dissociation of iPSCs and NGN2 iNs for single cell RNA-sequencing.....	38
2.11 Single cell RNA-sequencing.....	39
2.11.1 scRNA-seq data analysis	39
2.12 Bulk ATAC-sequencing	40
2.12.1 Bulk ATAC-seq data analysis.....	42
2.12.2 Bulk ATAC-seq data analysis of neurons co-cultured with glia against neurons cultured without glia	42
2.13 NGN2 ChIP-sequencing	43
2.13.1 NGN2 ChIP-seq data analysis	45
3 Characterisation of NGN2 iNeurons.....	46
3.1 Introduction.....	46
3.2 Aims.....	47
3.3 Results.....	48
3.3.1 Generation of an HA-tagged NGN2 OPTi-OX hiPSC line	48
3.3.2 Establishment of an optimised protocol for the long-term culture of iNeurons.	56
3.3.3 Electrophysiological properties of NGN2 iNs on MEAs	61
3.4 Discussion.....	72
4 Transcriptional profile of NGN2 reprogramming	75
4.1 Introduction.....	75
4.2 Aims.....	76
4.3 Results.....	78
4.3.1 Bulk RNA-seq data quality assessment and visualisation	78
4.3.2 Differential expression analysis of bulk RNAseq data	80
4.3.3 Weighted gene co-expression network analysis of bulk RNAseq data	83
4.3.4 scRNAseq analysis of NGN2 reprogramming.....	91
4.3.5 Effects of glia on transcriptome of NGN2 iNeurons	124

4.4 Discussion.....	134
5 Epigenomic profile of NGN2 reprogramming.....	139
5.1 Introduction.....	139
5.2 Aims and experimental design.....	141
5.3.1 PCA analysis of ATAC-seq data recapitulates bulk RNA-seq data	143
5.3.2 Rapid remodelling of the epigenome from the onset of NGN2 reprogramming.....	143
5.3.3 Nucleosome occupancy undergo dynamic changes globally over the course of reprogramming and maturation	148
5.3.4 Rapid establishment of nucleosome occupancy and positioning at NGN2-bound sites during reprogramming.	150
5.3.5 Identification of transcription factors driving changes in chromatin accessibility during NGN2 reprogramming and maturation.	154
5.4 Discussion.....	164
6 Conclusions and future directions.....	170
7 Appendix.....	174
7.1 List of antibodies	174
7.2 List of primers for qPCR	174
7.3 scRNAseq QC report	175
The QC filtering was done using Scanpy (Wolf, Angerer, & Theis, 2018) and involved removal of cells expressing less than 2000 genes, more than 100000 transcripts or more than 12% mitochondrial reads and genes that had less than 500 UMIs or expressed in less than 10 cells.	175
7.4 Full list of significant TFs from DASTk analysis.....	176
7.5 Differential accessible region analysis (cumulative version)	178
8 References.....	179

List of figures

Figure 1. 1: Optimised inducible overexpression system (OPTi-OX) for forward programming protocols.....	19
Figure 1. 2: Structure of bHLH transcription factors and NGN2.	22
Figure 2. 1: Mechanical dissociation method for increased dissociation of iNeurons into single cells.	39
Figure 3. 1: Overview of the cloning strategy for generating a Tet- HA-NGN2 AAVS1 targeting vector.	50
Figure 3. 2: Genotyping results for a clone targeted with HA-NGN2 OPTi-OX.....	51
Figure 3. 3: Verification of OPTi-OX system in targeted clones through qPCR analysis.	52
Figure 3. 4: Verification of HA-NGN2 expression.	54
Figure 3. 5: Karyotype report for the Bob HA-NGN2 OPTi-OX hiPSC line.	55
Figure 3. 6: Comparison of iNs cultured in the presence and absence of rat primary glia.....	58
Figure 3. 7: ICC of long-term cultures of iNeurons co-cultured with rat astrocytes.....	60
Figure 3. 8: Phase contrast time-lapse images of the first 3 days of iNeuron reprogramming....	63
Figure 3. 9: NGN2 iNs show spontaneous activity from day 11 onwards.	66
Figure 3. 10: Voltage-gated sodium channels mediate electrical activity in NGN2 iNs.....	67
Figure 3. 11: Synchronous burst firing emerge at day 20 post-induction.....	68
Figure 3. 12: Characteristics of synchronous burst firing.....	69
Figure 3. 13: Development of complex network activity from day 20 to day 27.....	71
Figure 4. 1: Time course used for bulk RNA-seq and scRNA-seq of NGN2 reprogramming.....	77
Figure 4. 2: Visualisation of NGN2 reprogramming bulk RNAseq data, processed using variance stabilising transformation.	79
Figure 4. 3: Top gene ontologies enriched for significantly upregulated and downregulated genes across linear time.	81
Figure 4. 4: Top gene ontologies enriched for significantly downregulated genes at specific points of NGN2 reprogramming.	82
Figure 4. 5: Top gene ontologies enriched for significantly upregulated genes at specific points of NGN2 reprogramming.....	84
Figure 4. 6: Gene clustering tree (dendrogram), module colours and labels, and a heatmap of individual gene association with each timepoint.	86

Figure 4. 7: Heatmap representation of time dependence of module eigengenes of main modules.....	87
Figure 4. 8: Boxplot of module eigengenes 3 & 7 vs. time.	88
Figure 4. 9: Boxplot of module eigengenes 2 & 5 vs. time.	90
Figure 4. 10: Boxplot of module eigengene 1 vs. time maturation module	92
Figure 4. 11: UMAP visualisation of NGN2 reprogramming time course.....	93
Figure 4. 12: UMAP visualisation of the early stage of NGN2 reprogramming and its Louvain clusters.	95
Figure 4. 13: Differential expression analysis of clusters from early stage of NGN2 reprogramming.....	96
Figure 4. 14: Rapid downregulation of pluripotency genes during early stage of NGN2 reprogramming.....	97
Figure 4. 15: Subpopulation of iPSCs expressed NPC genes.....	100
Figure 4. 16: Repressor of neurogenesis was efficiently downregulated within 24 hours of NGN2 reprogramming.....	101
Figure 4. 17: NGN2 reprogrammed cells strongly expressed CDKN1A at Day 2 before mass cell cycle exit.....	103
Figure 4. 18: Neuronal differentiation in the early stage of NGN2 reprogramming.	106
Figure 4. 19: Emergence of cholinergic fate in the early stage of NGN2 reprogramming.	107
Figure 4. 20: Pseudotime analysis of cholinergic fate at 2 days post-induction.....	110
Figure 4. 21: Pseudotime analysis of glutamatergic fate at 2 days post-induction.....	111
Figure 4. 22: Pseudotime analysis of glutamatergic and cholinergic fate at 3- and 4 days post-induction.	112
Figure 4. 23: UMAP visualisation of the late stage, NGN2 expression and its Louvain clusters.	115
Figure 4. 24: Differential expression analysis of clusters from late stage of NGN2 reprogramming.....	117
Figure 4. 25: Diversity in cholinergic markers at the late stage of NGN2 reprogramming.	118
Figure 4. 26: Diversity in glutamatergic markers at the late stage of NGN2 reprogramming. .	119
Figure 4. 27: NGN2 reprogramming produces three major types of neurons	120
Figure 4. 28: Immunocytochemistry of cholinergic and glutamatergic markers in Day 14 iNs.	122
Figure 4. 29: Heatmap of transcriptional expression of different markers from bulk RNAseq data of entire time course of NGN2 reprogramming.....	123

Figure 4. 30: UMAP visualisation of primary rat glia and differential expression analysis of its clusters	125
Figure 4. 31: Heterogeneity of rat glia.....	126
Figure 4. 32: Single-cell analysis of neuronal markers in rat glia.	127
Figure 4. 33: Addition of glia shows modest effect on transcriptome of cells at 4 days post NGN2 induction.....	129
Figure 4. 34: Enrichment of neuronal maturation ontologies in NGN2 neurons co-cultured with glia at Day 21	130
Figure 4. 35: Addition of glia shows no observable effect on lineage specification of NGN2 neurons.....	132
Figure 4. 36: Enrichment of synaptic genes in neurons co-cultured with glia at Day 14 and at Day 21.....	133
Figure 5. 1: Time course used for bulk ATAC-seq and NGN2 ChIP-seq of NGN2 reprogramming.....	142
Figure 5. 2: PCA of ATACseq data shows high similarity to PCA of bulk RNAseq	144
Figure 5. 3: Widespread gain of accessibility within six hours of NGN2 reprogramming.	145
Figure 5. 4: Loss of chromatin accessibility mainly occurs between 4 days and 14 days post-induction.	147
Figure 5. 5: Dynamic changes in nucleosome occupancy during NGN2 reprogramming	149
Figure 5. 6: Rapid establishment of nucleosomes at NGN2-bound sites during reprogramming.	151
Figure 5. 7: Nucleosome positioning and chromatin accessibility at NGN2 bound-sites for cells co-cultured with glia.	153
Figure 5. 8: ONECUT TFs show increased expression in the late maturation stage, as predicted by DASTk analysis.	158
Figure 5. 9: PBX1 and 3, not PBX2, show increased expression in the late maturation stage.	159
Figure 5. 10: Decrease in NEUROD1 expression in the late stage is likely independent of NGN2 expression.	160
Figure 5. 11: Expression of SIX1 in NGN2 reprogramming and in neurons with glia.	161
Figure 5. 12: Bulk RNA-seq expression profiles of a selection of TFs identified from DASTk and homologues for ONECUT and PBX.	163
Figure 6. 1: Schematic summary of the genomic analysis of NGN2 reprogramming of human iPSCs.....	173

Figure 7. 1:QC report for scRNAseq.	175
Figure 7. 2:Plot of cumulative counts (frequency) for sites gaining accessibility (opening loci) and sites losing accessibility (closing loci) throughout NGN2 reprogramming (samples without glia).	178

List of tables

Table 1. 1: A selection of studies that derived neurons through cellular reprogramming.	17
Table 2. 1: OPTi-OX genotyping primers and strategy.	34
Table 5. 1: Dominant transcription activity over the course of NGN2 reprogramming based on ATACseq data.	155
Table 5. 2: Dominant transcription activity between iNs cultured with glia and iNs cultured without glia, based on ATACseq data.	156
Table 7. 1: Dominant transcription activity over the course of NGN2 reprogramming based on ATACseq data.	176
Table 7. 2: Dominant transcription activity between iNs cultured with glia and iNs cultured without glia, based on ATACseq data.	177

1 Introduction

1.1 Cell reprogramming

There is a saying that *There are no shortcuts in life*; but *shortcuts* have become one of the greatest prospects of life science research in recent years. The term highlighted here refers to cellular reprogramming – the process of converting one cell type into another by overwriting the established starting cellular program.

The birth of cell programming began in 1962, when John Gurdon challenged the dogma that specialised cells are irreversibly committed to its fate. At that point, it was believed a cell loses the genomic information for making other cell types once it specialises. Gurdon tested the hypothesis that specialised cells do not lose this information by replacing the cell nucleus of a frog's egg cell with a nucleus from an intestinal cell of a tadpole. Surprisingly, the eggs developed into fully functional tadpoles and subsequently into adult frogs, proving that indeed the genomic information is not lost. It also showed that a specialised cell can be reprogrammed to an immature state capable of forming all cell types in the body, later defined as the pluripotent state (Gurdon, 1962). This ground-breaking discovery rewrote textbooks and eventually led to the landmark experiment of cloning Dolly the sheep. This method of cellular reprogramming is appropriately known as somatic cell nuclear transfer (SCNT) and is still widely applied in research, but scientists began to wonder if a cell could be reprogrammed to pluripotency without transferring the nucleus to a donor egg cell.

More than 40 years after Gurdon's discovery, Shinya Yamanaka answered this question by firstly identifying a list of genes that kept mouse embryonic stem cells (ESCs) in a pluripotent state. After testing different combinations of the genes, Yamanaka and his post-doc, Kazutoshi Takahashi, narrowed it down to four essential factors: OCT3/4, SOX2, C-MYC and KLF4, collectively known as Yamanaka factors (Takahashi & Yamanaka, 2006). Using retroviral transduction, they overexpressed these genes in mouse fibroblasts and successfully converted them into pluripotent stem cells (PSCs), which they called induced pluripotent stem cells (iPSCs). This breakthrough kickstarted the iPSC revolution which has since accelerated medical research.

However, the concept of ectopically overexpressing a gene to reprogram a cell into a different cell type came two decades before the discovery of iPSCs, when in 1987 it was shown that the transcription factor (TF) MyoD could reprogram fibroblasts to myoblasts (Davis, Weintraub, & Lassar, 1987). The discovery came from the observation that fibroblasts treated with 5-

azacytidine, an antagonist of DNA methyltransferase, converted them into myoblasts (Taylor & Jones, 1979). Davis and colleagues then identified MyoD as the key TF that is activated due to the epigenetic changes induced by 5-azacytidine. This finding opened up a new branch of cellular reprogramming known as direct reprogramming or transdifferentiation, where reprogramming occurs across different lineages. Since then, this approach has been used for the generation of numerous cell types such as hepatocytes, pancreas and neurons.

Whether it is reprogramming to pluripotency or direct reprogramming, these methods of *hacking* a cell's identity have not only shown us just how malleable the epigenome can be, but how just a few transcription factors can orchestrate this remarkable cellular process.

1.2 Induced Neurons

Neurons, more than most other cells in the body, are difficult to obtain and culture. Furthermore, post-mortem samples reflect the end-stage of a disease, making it difficult to uncover the pathogenic mechanisms of the disease. The advent of human ESCs (Thomson, 1998) and iPSCs (Takahashi & Yamanaka, 2006) led to the development of numerous protocols for making neurons (Chambers et al., 2009; Shi, Kirwan, & Livesey, 2012), applying an approach termed 'directed differentiation', where differentiation is based on recapitulation of developmental signalling cues in vitro. However, it soon became apparent that these protocols are laborious, requiring months of in vitro culture to produce functional neurons, and are often hard to reproduce.

An alternative approach was proposed by experiments in which induced neurons (iN) were derived through cellular reprogramming (Table 1). The first demonstration of reprogramming cells to neural lineages using transcription factors was conducted by Goetz and colleagues, who showed that forced-expression of PAX6 could induce neurogenesis in astrocytes isolated from Pax-6 mutant mice (Heins et al., 2002). Within 7 days, nearly half of astrocytes formed β III-tubulin- and NeuN-positive neurons. Later on, the same group showed that overexpressing the proneural basic-helix-loop-helix (bHLH) transcription factors, either ASCL1 or Ngn2 in mice postnatal astroglia also induces neurogenesis, but in both instances their neurons appeared to mature slowly and fail to generate functional presynaptic output, which they later attributed to silencing of their transgene (Berninger et al., 2007). This limitation was overcome by subcloning Ngn2 into a self-inactivating retroviral vector under the control of a chicken beta-actin promoter, which allowed stronger and more persistent expression of the transgene; thus, indicating the importance of the level of and persistence of expression of reprogramming factors towards successful reprogramming (Heinrich et al., 2010).

The use of astrocytes as a starting population was based on the notion that they would have an increased propensity for neurogenesis given that they share a common ancestral lineage. Furthermore, they were also motivated by the idea that astrocytes can be a great source for generating neurons *in vivo* through gene therapy and replace dying or damaged neurons in neurodegenerative diseases or traumatic brain injuries. While research into their utility for cell-based therapy continues, it wasn't clear if neurons could be generated from cells of non-neuronal lineages, particularly of mesodermal or endodermal origin. This highlights the significance of Vierbuchen and colleague's work, who successfully converted mouse fibroblasts into functional excitatory neurons using a combination of three factors: ASCL1, BRN2 and MYT1L (Vierbuchen et al., 2010). Although the use of ASCL1 alone was sufficient to induce immature neuronal features, albeit at a low yield, the additional expression of BRN2 and MYT1L was able to generate mature iNs with efficiencies of up to 19.5%. In addition to transcription factors, neuronal reprogramming has also been achieved using microRNAs. Specifically, the addition of miR-9/9* and miR-124 to NEUROD2, ASCL1 and MYT1L was shown to significantly improve the reprogramming efficiency of human fibroblasts into functional neurons (Yoo et al., 2011). The neurons generated by this approach were predominantly of a glutamatergic type, but reprogramming has also been used to generate other neuronal types.

GABAergic (γ -aminobutyric acid-secreting) interneurons (GINs) provide an inhibitory tone to the circuitry in the CNS. An imbalance between excitatory and inhibitory neuronal networks, especially in the cerebral cortex and hippocampus causes epilepsy and other neurological disorders (Colasante et al., 2015). In development, GINs arise from the lateral, medial and caudal ganglionic eminences in the ventral telencephalon, where ASCL1 is the only proneural bHLH TF that is expressed (Dennis, Han, & Schuurmans, 2018). Not surprisingly, all evidence of GINs generated through direct reprogramming so far require ASCL1 in combination with other factors (Colasante et al., 2015; Sun et al., 2016; Yang et al., 2017). In addition to ASCL1, these studies found variants of the Distal-less (DLX) genes - known for their role in GABAergic neuron generation and migration during development (Long, Cobos, Potter, & Rubenstein, 2009) - to be instrumental for reprogramming of cells into GABAergic fates. Although the first two demonstrations of induced GINs (iGINS) required other factors in combination with these two genes, Yang and colleagues, in an attempt to identify the best single factor in addition to either NGN2, or ASCL1, found that the combination of ASCL1 and DLX2 alone were most potent in generating functional iGINS from human PSCs (Yang et al., 2017). All three studies assessed their iGINS for specificity to any of the different subtypes of interneurons, such as somatostatin (SST), parvalbumin (PV), reelin, calbindin and neuropeptide Y. Interestingly, all

iN subtype	Factor(s)	Source	Species	Reference
Unspecified	Pax6	Cortical radial glia	Mouse	Heins et al., 2002
Glutamatergic	Ngn2/Mash1	Cortical astroglia	Mouse	Berninger et al., 2007, Heinrich et al., 2010
Glutamatergic	Ascl1, Brn2, Myt11	Embryonic fibroblasts	Mouse	Vierbuchen et al., 2010
Glutamatergic-GABAergic hybrid	miR-9/9*, miR-124, NEUROD2, ASCL1, MYT1	Fibroblasts	Human	Yoo et al., 2011
GABAergic	Ascl1, Dlx5, Lhx6, FoxG1, Sox2	Fibroblasts	Mouse, Human	Colasante et al., 2015
GABAergic	ASCL1, LHX6, DLX2, miR9/9*, miR-124	PSCs	Human	Sun et al., 2016
GABAergic	ASCL1, DLX2, MYTLL	PSCs	Human	Yang et al., 2017
Dopaminergic	ASCL1, BRN2, MYT1L, LMX1A, FOXA2	Fibroblasts	Human	Pfisterer et al., 2011
Dopaminergic	ASCL1, LMX1A, NURR1, NEUROD1, miR-218	Astrocytes	Human	Rivetti di Val Cervo et al., 2017
Motor neurons	ASCL1, BRN2, MYT1L, LHX3, HB9, ISL1, NGN2	Fibroblasts	Human	Son et al., 2011
Motor neurons	NGN2, SOX11, ISL1, LHX3	Fibroblasts	Human	Liu, Zang, & Zhang, 2016.
Motor neurons	Ngn2, Isl1, Lhx3	ESCs	Mouse	Velasco et al., 2017

Table 1. 1: A selection of studies that derived neurons through cellular reprogramming.

This technique has been used to derive various subtype of neurons, from both human and non-human sources and from both somatic and pluripotent cell types.

reported producing GINs expressing varying levels of the different marker genes, suggesting that the reprogramming produces a generic forebrain phenotype rather than any one subtype of interneurons. Yang and Sun's study both reported high expression of SST among the markers whereas, Colasante and colleagues reported PV as being the dominant subtype, which could be due to the latter using additional factors such as FoxG1 and Sox2.

Besides glutamatergic and GABAergic neurons, there have been considerable efforts to make other disease-relevant subtypes, such as dopaminergic and motor neurons, which are commonly implicated with Parkinson's disease and motor neuron disease, respectively (Table 1). Both these diseases are characterised by the degeneration of their associated neurons; thus, induced dopaminergic or motor neurons are great candidates for cell transplants. In addition, they are a good candidate for in vitro studies because the aetiological mechanisms underlying their pathology remains largely unknown. The speed that comes with direct reprogramming provides an unmatched advantage over cells sourced from conventional differentiations for both applications.

1.3 OPTi-OX: Optimised inducible Overexpression System

Typically, reprogramming protocols so far utilise a retroviral or lentiviral gene delivery system to deliver a reprogramming cassette that allows inducible overexpression of transgenes. Although these are powerful tools for the transduction of a broad range of mammalian cell types, limitations with regards to the transduction efficiency and timing of retroviral integration renders the starting population heterogeneous before the onset of the actual reprogramming process. In addition, random integration of the transgenes poses considerable limitations. Integrations into certain areas of the genome are susceptible to gene silencing, which is more pronounced in both proliferating and differentiating hPSCs (Qian et al., 2014). Delivery into other areas, such as coding sequences, may also pose impairments to cellular function or introduce a tumorigenic potential, which makes such protocols undesirable for cell-based therapies (Schambach, Zychlinski, Ehrnstroem, & Baum, 2013).

To circumvent these limitations, our lab developed a strategy of targeting the components of a doxycycline-dependent inducible transgene expression system into two separate genomic safe harbour sites (GSHs) in hPSCs (Pawlowski et al., 2017). These are areas of the genome that can accommodate predictable expression of newly integrated DNA without adverse effects on the host cell or organism and yield desired expression levels of the integrated transgene (Sadelain, Papapetrou, & Bushman, 2012). We have termed this approach OPTi-OX (optimised inducible overexpression system) (Fig. 1). It is comprised of two components of a Tet-On system: (1) a

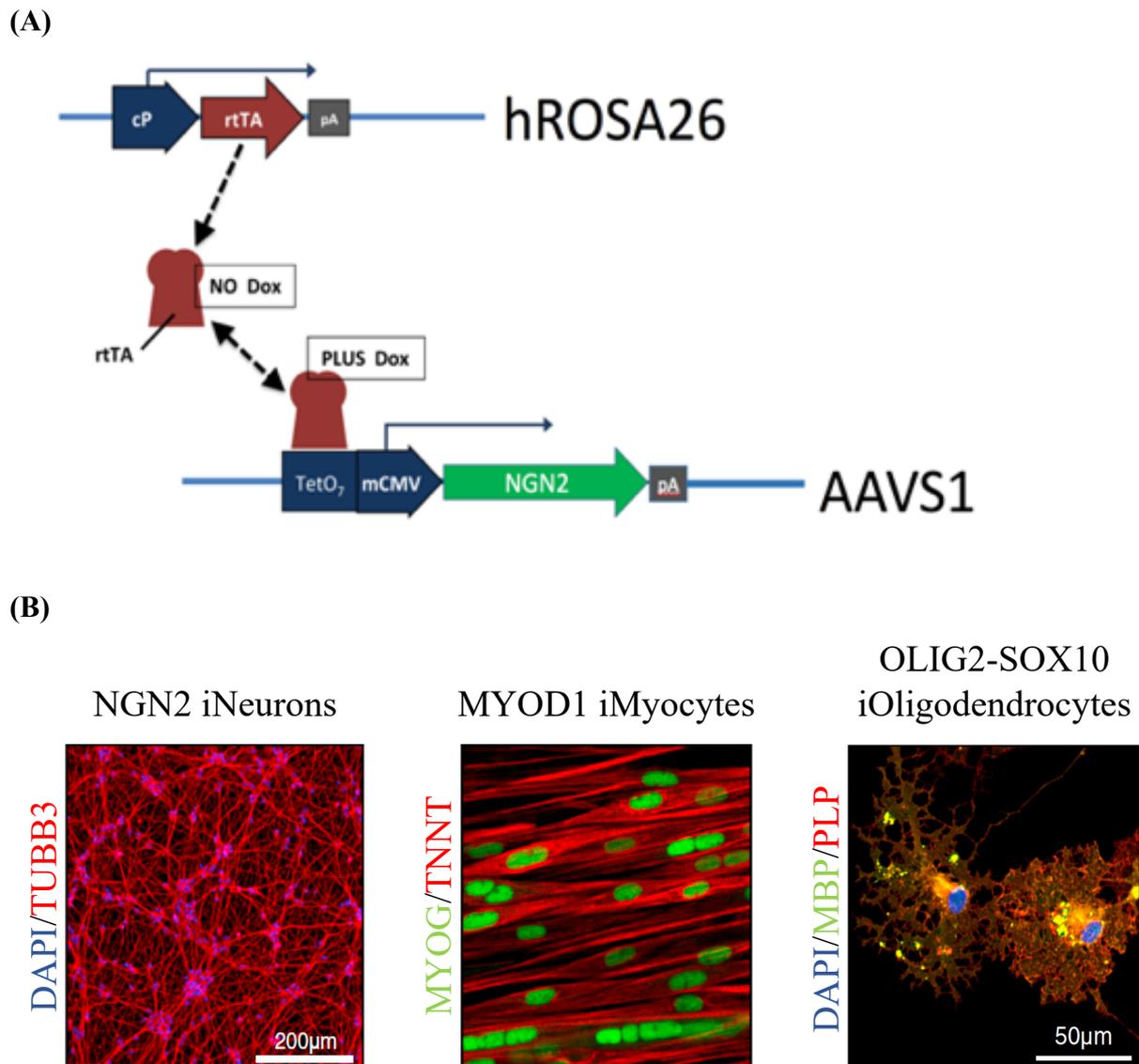


Figure 1. 1: Optimised inducible overexpression system (OPTi-OX) for forward programming protocols.

(A) The two components of the OPTi-OX system are targeted at their respective GSH to create transgenic lines that either are homozygous or heterozygous for each transgene. Clonal derivation of the cell lines ensures a highly consistent starting cell population. Embedded in the human Rosa26 locus, a constitutive promoter (cP) drives the expression of reverse-tetracycline Trans-Activator (rtTA), which in the presence of doxycycline, binds to a Tet-responsive element (TRE) placed in the AAVS1 locus, that then drives the overexpression of the transgene, in this case NGN2. (pA: Polyadenylation signal). (adapted from SWISS-MODEL Repository). (B) OPTi-OX has so far been used to produce three types of human cells: neurons derived overexpression of NGN2, skeletal myocytes through MYOD1, and finally oligodendrocytes through co-overexpression of OLIG2 and SOX10, shown here through immunocytochemistry stainings of important markers for the three cell types (adapted from Pawlowski et al., 2017).

constitutively expressed doxycycline (dox) -responsive transcriptional activator protein (reverse-tetracycline Trans- Activator, rtTA) embedded in the human Rosa26 locus , and (2) a responder cassette targeted into the AAVS1 locus, comprising a rtTA-sensitive inducible promoter (Tet Responsive Element, TRE) driving expression of the gene of interest. By using site specific genome targeting tools such as CRISPR/CAS9 and zinc-finger nucleases to place the two components within a specific site of their respective GSH, the system allows a robust overexpression of the transgene which can be maintained until maturity of the desired cell type.

Our lab has used the system successfully to reprogram human PSCs into functional skeletal myocytes through overexpression of MYOD1, oligodendrocyte precursor cells through overexpression of OLIG2 and SOX10 and, glutamatergic neurons through overexpression of NGN2 (Pawlowski et al., 2017).

1.4 NGN2 iNeurons

Application of our OPTi-OX strategy demonstrated that NGN2 reprogramming occurs highly efficiently, generating cultures made up entirely of neurons as early as 3- or 4-days post-induction, in every PSC cell line we tested. Furthermore, it was shown that only 4 days of dox-induced induction was needed for efficient reprogramming (Pawlowski et al., 2017). How a single transcription factor is able to orchestrate this rapid cellular process remains poorly understood. To answer this, it is crucial to have a solid underlying knowledge of NGN2, by firstly reviewing the work so far on NGN2-derived iNs. Francois Guillemot and Carol Schuurmans have contributed seminal work on proneural bHLHs, specifically ASCL1 and NGN2. Schuurmans and colleagues recently published an excellent review on bHLH transcription factors, including NGN2 (Dennis et al., 2018). The next few sections will cover some of the evidence on NGN2 discussed in their review supplemented further by findings from other sources.

1.4.1 NGN2 – A Neurogenin bHLH transcription factor

NGN2 or NEUROG2 stands for Neurogenin2 and together with NGN1 and NGN3, makes up the neurogenin family of transcription factors that, as the name suggests, play an important role in neurogenesis. They have widely been described as master regulators of neural fate and specification in the developing CNS, as evidenced by numerous loss and gain of function studies (Dixit et al., 2014; Kowalchuk, Maurer, Shoja-Taheri, & Brown, 2018; Lee, Lee, Ruiz, & Pfaff, 2005; Parras et al., 2002). Neurogenins are a part of a larger family of TFs called bHLH TFs, so-called because of their protein structure. These proteins are made up of two domains – a HLH domain comprised of a pair of alpha helices connected by a non-conserved loop that mediates

dimerization with other proteins, usually TFs; and a basic domain that binds E-box motifs in the DNA (**Fig. 2**) (Dennis et al., 2018). They are classified into two main groups, Class I which is comprised of ubiquitously expressed proteins such as Tcf4 and Tcf12, and Class II, comprised of bHLH proteins with tissue-specific expression. Neural-specific bHLH genes can be further be subdivided into proneural TFs and neural differentiation TFs. NGN2, along with bHLH TFs such as ATOH1 and ASCL1, are classified as proneural TFs because of their early expression in progenitor cells, which confers specification into a neural (neuronal and glial) or neuronal fate in the developing nervous system (Guillemot & Hassan, 2017). On the other hand, neuronal differentiation bHLH genes comprise members of the NeuroD family, which are first expressed in cells already committed to a neural fate and instead promote their differentiation.

1.4.2 Role of NGN2 in development

Like most proneural bHLH TFs, NGN2 driven neurogenesis occurs in various domains of the developing CNS. Perhaps the most well-known region associated with NGN2 is the neocortex which is part of the telencephalon. There, NGN2 is mainly expressed in cortical progenitors throughout the ventricular zone (VZ) in the dorsal telencephalon (E10.5-E17 in mouse), which then give rise to glutamatergic projection neurons (Dennis et al., 2018; Mattar et al., 2008). The earliest class of neurons to arise from NGN2⁺ progenitors here are early-born Cajal-Retzius cells, which were shown to populate layer I and then layer II/III neurons of the piriform cortex (Dixit et al., 2014). The next class of neurons to come from NGN2⁺ progenitors are subplate neurons, a transient neuronal population with important pioneering roles for guiding afferent and efferent axonal projections in cortical development (Mattar et al., 2004). NGN2 is excluded from the ventral telencephalon, where ASCL1 is the dominant proneural factor and induces the expression of genes involved in GABAergic neuron differentiation, such as DLX1 and 2 and LHX6 (Petryniak, Potter, Rowitch, & Rubenstein, 2007). Chromatin immunoprecipitation studies have shown that NGN2 blocks ASCL1-induced ventral-telencephalic identity through one of its direct targets, TBR2, which indirectly represses ASCL1 by inhibiting Ebf1, a positive regulator of ASCL1 (Kovach et al., 2013). In addition to regulating neuronal differentiation, NGN2 is also involved in cortical neuron migration. In newly formed mouse cortical neurons, it directly induces the expression of RND2, a GTP-binding protein that mediates neuronal migration through inhibition of RhoA activity (Heng et al., 2008).

Beyond the neocortex, in the developing hindbrain, NGN2 has been shown to regulate the generation of Purkinje neurons, which are a class of GABAergic neurons that make up the

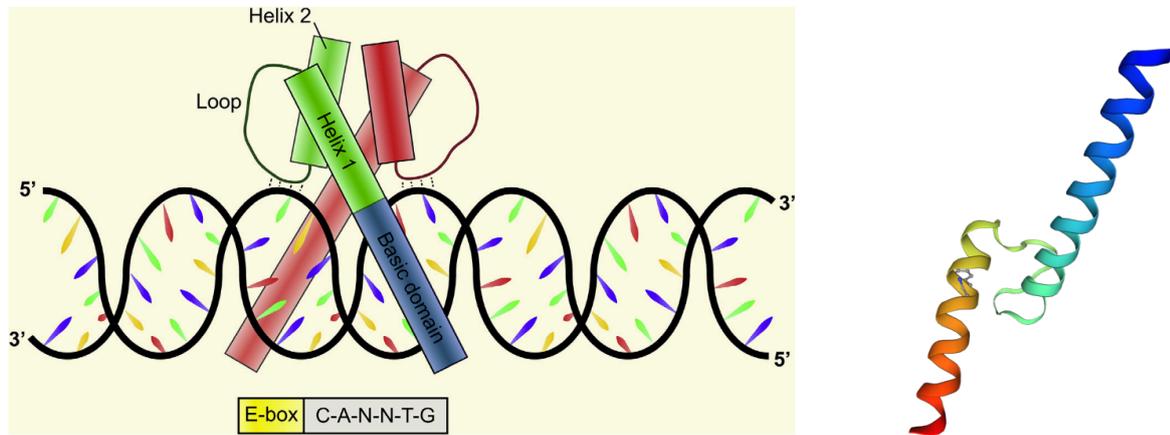


Figure 1. 2: Structure of bHLH transcription factors and NGN2.

Left: bHLH TFs are made up of two alpha helices connected by a non-conserved loop region that mediates dimerization with other protein, and a basic domain that binds E-box sequences in the DNA (adapted from Dennis *et al.*, 2019). Right: 3D-model of NGN2. Rainbow colour scheme depicts location of an amino acid relative to the N-terminus (blue) and the C-terminus (red) (adapted from SWISS-MODEL Repository).

principal neuron in cerebellar circuits (Florio et al., 2012). There, NGN2 is expressed in Purkinje cell progenitors poised to exit the cell cycle and also plays a key role in dendrite morphogenesis. This not only demonstrates NGN2's broad neurogenic properties in the developing CNS, but also how the ability to generate both glutamatergic and GABAergic phenotypes show its proneuronal functions are independent of its contribution to neuronal phenotype specification. Further evidence of this can be seen in the ventral spinal cord, where deletion of NGN2 was shown to lead to a reduction in motor neurons and ventral interneurons (Parras et al., 2002; Scardigli, Schuurmans, Gradwohl, & Guillemot, 2001). It was later demonstrated that high expression of NGN2 in motor neuron progenitors (pMN) competes with Olig2 for their shared DNA binding sites, which leads to a switch from Olig2-induced progenitor maintenance to NGN2-driven differentiation into post-mitotic motor neurons (Lee et al., 2005). Elsewhere, NGN2 has also been shown to play a role in the timing of neurogenesis of retinal ganglion cells (RGCs) in mice embryos (Hufnagel, 2010). In addition, it was recently reported that NGN2 is also required for postnatal retinal neurogenesis, specifically for the differentiation of cone and rod bipolar neurons and rod photoreceptors (Kowalchuk et al., 2018).

Collectively, these studies demonstrate NGN2's function as a master regulator of neurogenesis throughout CNS development, where it is engaged in various domains and generation of distinct phenotypes, namely glutamatergic, GABAergic, and cholinergic (motor) neurons. Realising this wide application of NGN2 would be integral for deciphering a complex genomic analysis of NGN2 reprogramming, especially a single cell transcriptomic analysis aimed at uncovering any heterogeneity of our NGN2 iNs.

1.4.3 NGN2 as a reprogramming factor

Given its important role as a proneuronal factor in development, it is no surprise that there have been numerous studies using NGN2 as a reprogramming factor to generate induced neurons. However, when compared to ASCL1, NGN2 seems less capable of trans-lineage reprogramming of fibroblast, unless it is overexpressed in combination with other factors. Therefore, reports of solely overexpressing NGN2 have either involved ESCs, iPSCs, NPCs or astrocytes as a starting cell population. A recent review by Colasante et al. (2019) provides a good account of the literature on NGN2 as a reprogramming factor, either on its own or in combination with other factors. The summary in Table 2 incorporates some of the publications reviewed there in addition to several other critical studies that were not covered. Furthermore, given the specific focus of my project on understanding NGN2 reprogramming, I also included information on genomic analyses that were carried out in the study, and also an account of the main cell culture components and

Factors	Source	iN subtype	Species	Main cell culture components	Genomic analysis	Reference
NGN2	iPSCs, ESCs	Glut	Human	NB, B27, FBS	sc RT-PCR	Zhang et al., 2013
Ngn2	ESCs	Glut	Mouse	Adv. DMEM/F-12, NB, N2, B27	RNAseq, scRNAseq, ATACseq, NGN2 ChIP, H3K27AC & H3K27ME3 ChIP	Aydin et al., 2019
Ngn2	ESCs	Glut	Mouse	Adv. DMEM/F-12, NB, N2, B27	Ngn2 ChIP	Velasco et al., 2017
NGN2	NPCs	Glut	Human	DMEM/F12, B27	RT-PCR	Ho et al., 2016
Ngn2	Cerebellar astrocytes	GABA & Glut	Mouse	DMEM/F12, B27		Chouchane et al., 2017
Ngn2	Astrocytes	Glut	Mouse	DMEM/F12, B27	RT-PCR	Heinrich et al., 2010
NGN1, NGN2	ESCs	Glut-chol. hybrid	Human	NB, B27	RNAseq, miRNA profiling	Busskamp et al., 2014
Ngn2, Bcl2	Astrocytes	Glut	Mouse	DMEM/F12, B27	RT-PCR	Gascón et al., 2016
NGN2, BRN3A	Fibroblasts	Sensory	Human	DMEM/F12, NB, N2, B27	RT-PCR	Blanchard et al., 2015
Ngn2, ASCL1	Astrocytes	Glut and GABA	Mouse	DMEM/F12, B27	Microarray, Micro-ChIP, qPCR	Masserdotti et al., 2015
NGN2, ASCL1	Fibroblasts	Glut	Human	DMEM/F12, NB, N2, B27	Whole-genome expression array	Ladewig et al., 2012
Ngn2, Isl1, Lhx3	ESCs	Motor	Mouse	Adv. DMEM/F-12, NB, N2, B27	RNAseq, ATACseq, NGN2 ChIP, H3K27AC & H3K27ME3 ChIP	Velasco et al., 2017
NGN2, SOX11, ISL1, LHX3, Small molecules*	Fibroblasts	Motor	Human	DMEM/F12, NB, N2, B27, <i>DM</i> , <i>FSK</i>	RT-PCR	Liu, Zang, & Zhang, 2016.
NGN2, Small molecules*	ESCs	Glut	Human	DMEM/F12, NB, N2, B27, <i>SB</i> , <i>LDN</i> , <i>XAV</i>	RNAseq, scRNAseq	Nehme et al., 2018
NGN2, SOX11, Small molecules*	Fibroblasts	Cholinergic	Human	DMEM/F12, NB, N2, B27, <i>DM</i> , <i>FSK</i>	RNAseq, ATACseq, NGN2 ChIP, H3K27AC & H3K27ME3 ChIP	Liu et al., 2013; Smith et al., 2016

Table 1.2: Summary of publications on NGN2 reprogramming.

Summary covers where NGN2 was used to directly reprogram somatic or pluripotent cells into specific iN subtypes, the main cell culture components they used, the species of the source cells and a list of any genomic analyses that were performed. NB=Neurobasal, DM = Dorsomorphin, FSK = Forskolin, SB = SB431542, LDN = LDN193189, XAV = XAV939, Glut = Glutamatergic. *Small molecules listed in italic in the “Main cell culture components” column.

the species of the cell source, since these factors can have a profound effect on reprogramming and the resulting cells.

When NGN2 is the sole reprogramming factor and the cells are cultured in typical, non-specifying neuronal cell culture conditions, such as Neurobasal, DMEM/F12, N2 supplement and B27 supplement, the neurons produced acquire a glutamatergic identity. Only when it is induced in murine astrocytes, specifically of cerebellar origin does NGN2 produce GABAergic neurons in addition to glutamatergic neurons (Chouchane et al., 2017). This is consistent with its role in the generation of Purkinje cells in the developing hindbrain, as described earlier (Florio et al., 2012).

NGN2 can also reprogram fibroblasts into glutamatergic neurons but only when it is co-expressed with ASCL1 (Ladewig et al., 2012). On the other hand, when co-expressed with Brn3a, NGN2 reprogrammed human and mice fibroblasts into induced sensory neurons which remarkably consisted all three of the functional subtypes of sensory neurons in mice. However, in all other instances covered in this review, co-expression of NGN2 produces neurons with a cholinergic phenotype, marked by expression of genes necessary for cholinergic neurotransmission such as CHAT and VACHT (SLC18A3). NGN2 specifically produced induced spinal motor neurons when overexpressed in combination with Isl1 and Lhx3, which are known for their role in motor neuron development (Guthrie, 2007; Liu, Zang, & Zhang, 2016; Velasco et al., 2017). In addition, the resulting motor neurons were able form functional neuromuscular junctions with skeletal muscles and model aspects of motor neuron degenerative diseases such as amyotrophic lateral sclerosis (ALS) (M.-L. Liu, Zang, & Zhang, 2016). NGN2 can also yield cholinergic neurons when combined with small molecules, specifically dorsomorphin, an inhibitor of AMP-activated protein kinase and bone morphogenetic protein type 1 receptors, and forskolin, a cyclic AMP (cAMP) synthesis activator (Liu et al., 2013). The addition of Sox11 to the reprogramming cassette as an additional factor was required for NGN2 to efficiently reprogram human postnatal and adult fibroblasts to highly pure cholinergic neurons under the defined conditions. Unlike the spinal motor neuron examples, these cholinergic neurons were not tested for lineage specificity, so it was not clear if they also included spinal motor neurons, or other subtypes such as forebrain cholinergic neurons or cranial motor neurons. Separately, overexpressing NGN2 in combination with NGN1 in human ESCs produced neurons with both a glutamatergic and cholinergic phenotype at 4 days post-induction, based solely on the co-expression of the glutamatergic marker VGLUT1 and cholinergic marker, ChAT (Busskamp et al., 2014). Despite making up 96% of the resulting culture, an extensive long-term study of their electrophysiological properties revealed that nearly all of their activity was mediated by AMPA glutamatergic receptors (Lam et al., 2017).

These reports of NGN2-associated cholinergic neurons are not surprising given that NGN2 itself is implicated in motor neuron development, as described in the previous section.

1.5 Gaps in knowledge

So far, we have seen how NGN2 has been used as a reprogramming factor in numerous ways, from reprogramming pluripotent cells, NPCs, astrocytes, to co-expression with other TFs or alongside perturbation with small molecules. In most cases, a rudimentary characterisation was carried out testing for electrophysiological function through patch-clamping and expression of common markers for neuronal maturity and lineage specificity, typically through RT-PCR and immunocytochemistry. A few studies have taken advantage of next-generation sequencing technologies, such as RNAseq, single cell RNAseq, ChIPseq and ATACseq to get insights into the transcriptional and epigenetic profile of their neurons and the reprogramming process itself (Aydin et al., 2019; Busskamp et al., 2014; Nehme et al., 2018; Smith et al., 2016; Velasco, Ibrahim, Kakumanu, et al., 2017; Yingsha Zhang et al., 2013). This typically involves a time course analysis where cells are sampled at key stages of the reprogramming.

In the earliest report of human ESC-derived NGN2 iNs, Zhang and colleagues quantitatively analysed expression of 73 genes in day 21 iNs at the single-cell level using Fluidigm-mediated RT-PCR (Yingsha Zhang et al., 2013). Although the gene list covered markers of different brain regions, types of glutamatergic receptors and markers of different neuronal types such as inhibitory neurons and serotonergic neurons, it did not test for the presence of cholinergic or motor neurons, which is more likely to be associated with NGN2 iNs based on the literature so far. Compared to the other studies reviewed here, the iNs described by Zhang and colleagues are more likely to be similar to our iNs, because it didn't involve any other TFs or small molecule perturbations, used human PSCs, and used culture conditions that is most similar to ours, except for the addition of fetal bovine serum in their co-cultures with rat glia. However, this study only analysed cells at a single time point. Busskamp and colleagues analysed the first four days of their human NGN1-NGN2 iNs using RNAseq and miRNA profiling, producing one of the most comprehensive transcriptional profile of an NGN2-associated reprogramming (Busskamp et al., 2014). They revealed significant ontologies for biological and cellular processes and transcription factors that are upregulated and downregulated during the reprogramming processes. Even though they derived their cells from human ESCs and used culture conditions identical to ours, the co-expression of NGN1 would probably make their iNs less comparable to ours. As for the remaining studies cited above, neither used parameters identical to our reprogramming.

On the other hand, the mechanisms of ASCL1 reprogramming from human fibroblasts into glutamatergic neurons has been extensively studied over the past few years by groups working in collaboration with the lab of Marius Wernig (Treutlein et al., 2016; Wapinski et al., 2017, 2013). Together, they have produced a rich transcriptional and epigenetic dataset to deconstruct ASCL1 reprogramming, demonstrating its capabilities as a pioneer factor to bind its cognate target sites in a relatively unfavourable donor environment, rapidly reorganise the chromatin structure and activate its target program. Furthermore, using recent advancements in single cell transcriptomics, they showed how the co-expression of BRN2 and MYT1L is necessary to silence ASCL1's off-target induction of a myogenic program, and therefore improve neuronal induction (Treutlein et al., 2016). It is also worth noting that their reprogramming was carried out in a non-instructive neuronal culture medium free of any small molecules; allowing any observations or conclusions made to be confidently attributed to the reprogramming factors.

Currently, a comprehensive body of work such as this hasn't been attempted for NGN2 reprogramming of human cells. As mentioned earlier, all previous studies have either co-expressed NGN2 with other TFs or in combination with small molecules. Furthermore, between the publications reporting a cholinergic phenotype, neither had parameters that matches our protocol. Therefore, it is not clear if overexpressing NGN2 alone in human PSCs in non-instructive culture conditions such as Neurobasal and B27, would also yield a cholinergic phenotype. Besides that, the use of our OPTi-OX model which results in homogenous expression of NGN2 across the entire cell population, as opposed to the variable expression that is commonly seen with the lentiviral expression systems, presents a unique opportunity to study the reprogramming process in a highly controlled setting. Moreover, the expression of a single transcription factor that is sufficient to reprogram cells into the desired target cell type provides an ideal reductionist model for studying the transcriptional and epigenetic changes that govern this process.

1.6 Outlook

The main aim of my PhD is to understand how NGN2 orchestrates the rapid and highly efficient conversion of human iPSCs into neurons. To achieve this, I will present a time course analysis of our NGN2 reprogramming protocol, integrating transcriptional- and epigenetic-based next-generation sequencing analyses, including an NGN2 ChIP-seq analysis to discern direct from indirect regulators. To dissect the process on a single cell level, I will also present findings from a single cell RNAseq analysis of the same time course. Designing a time course that best captures the crucial stages of this NGN2 reprogramming process requires a thorough characterisation of

its morphological and functional changes. Since only a rudimentary characterisation has been done on our NGN2 iNs, I will firstly address this need in Chapter 3, where I will present immunocytochemistry and electrophysiological data on our NGN2 iNs. Finally, I will discuss future directions which will include potential targets for future reprogramming protocols.

2 Methods

2.1 Cell culture

2.1.1 Maintenance of pluripotent stem cells

hPSCs were grown feeder-free in an in-house E8 media (LRM CCK), containing 25 ng/ml FGF2 and 2 ng/ml TGF- β . FGF2 was provided by Dr. Marko Hyvönen from the Department of Biochemistry (University of Cambridge). Cells were kept in a humidified incubator at 37°C and 5% CO₂. The medium was changed every day and cells were passaged every 4-6 days depending on confluency. For passaging, the medium was removed, washed once with Dulbecco's PBS (dPBS) and cells were incubated in dPBS-EDTA for 2-4 minutes, depending on the cell line. Subsequently, the solution was aspirated and cells were detached by forcefully releasing 1 ml of plain DMEM with a p1000. This was repeated until most of the cells were detached. With a 5ml stripette, cells were transferred to a 15 ml tube (Falcon) containing 5mls of DMEM. Cell clumps were allowed to settle for about 5 minutes and the supernatant was aspirated. The cell pellet was then re-suspended in fresh E8 medium and pipetted up and down twice with a p1000. Cells were then plated onto tissue-culture treated plates (Corning) that were coated with vitronectin (10 μ g/ml) (StemCell Tech) at room temperature for 1 hour. The splitting ratio was usually between 1:10 and 1:20.

2.1.2 NGN2 iNeurons

OPTi-NGN2 hPSCs were grown in colonies as described in 2.1.1. Then, they were dissociated into single cells using StemPro Accutase for 4 minutes and seeded (100 000 cells per well of a 12 well plate) onto Geltrex-coated plates. Cells were seeded in E8 media supplemented with 10 μ M of ROCK-Inhibitor. Differentiation was initiated 24 hours after seeding. For the first two days of differentiation, cells were switched to iN Induction media consisting of:

DMEM F12 (Gibco)

N2 supplement (100x, Gibco)

Glutamax (100x, Gibco)

Non-essential amino acids (100x, Gibco)

2-Mercaptoethanol (50 μ M, Gibco)

Penicillin-Streptomycin (100x)

Media was supplemented with dox (1 µg/ml) and changed daily. For subsequent differentiation and maintenance of neurons, cells were switched to iN Maintenance media consisting of:

Neurobasal (Gibco)

B27 supplement (100x, Gibco)

Glutamax (100x, Gibco)

2-Mercaptoethanol (50 µM, Gibco)

Penicillin-Streptomycin (100x)

Media was supplemented with dox (1 µg/ml), NT3 (10 ng/ml) and BDNF (5 ng/ml) and changed daily. Full media changes were performed until day 3 post-induction. There onwards, half-media changes were performed every other day. Dox was withdrawn from day 7 onwards.

2.1.3 Primary rat astrocyte culture

Primary mixed glial cultures were derived from P0-P2 neonatal Sprague Dawley rats and were generated along the previous guidelines (McCarthy & de Vellis, 1980), with minor modifications (Syed et al., 2008)(McCarthy & de Vellis, 1980). The pups were euthanized following Schedule 1 rules and regulations from the Home Office Animal Procedures Committee UK (APC). To maintain aseptic conditions, all procedures were performed in a laminar flow hood. A horizontal flow hood was used to perform dissections and a vertical flow hood for tissue culture.

In brief, the meninges, midbrain and olfactory bulbs were removed, and dissociated rat neonatal cortices cut thoroughly and incubated at 37°C for 30 min in Minimum Essential Medium Eagle (MEM) containing 4% Papain, 1% of 4 mg/mL DNase I Type IV and 1% of 24 mg/mL L-cysteine (Figure 1). After the digestion step the dissociated cells were plated into poly-D-lysine (PDL) coated cell culture flasks at a density of 2 brains/T75 flask. These mixed glia cultures were cultured for 10 days in Dulbecco's Modified Eagle's Medium (DMEM) supplemented with 10% FBS (Fetal Bovine Serum) and 1%Pen/Strep and kept under a humidified atmosphere at 37°C and 7% CO₂.

2.1.4 iN-rat astrocyte co-culture

Rat astrocytes, suspended in iN Maintenance media supplemented with dox, were added to iN cultures on day 3, at a 1:1 ratio. On day 5, medium was supplemented with 2 µM of Ara-C, to

inhibit astrocyte proliferation. Cultures were maintained in iN maintenance media and Dox was withdrawn from day 7 onwards.

2.1.5 iN culture on MEAs

For cultures on MEA, the MEA plates were coated with 100ug/ml of poly-D-lysine (PDL) for 1 hour at 37°C. After 3 washes with sterile water, they were air-dried in a Class II safety hood and sterilised under a UV light for 30 minutes. Then, a 20 µl drop of laminin (20 µg/ml; Sigma) was placed in the centre of the plate, so that it only covered the electrode surface. It was then left to incubate at 37°C for 30 minutes. Then, Day 3 iNs were dissociated and mixed with rat astrocytes at a ratio of 1:1 for a final total concentration of 4000 cells/µl. For seeding, cells were maintained in iN Maintenance media supplemented with 10 µM Rock-I and dox. The drop of laminin was then aspirated and replaced with 15 µl of the cell mixture. Cells were allowed to attach for one hour in an incubator at 37°C and 5% CO₂, before topping up with iN maintenance medium supplemented with dox. Half-media changes was performed from day 5 onwards, with 2 µM Ara-C being added to the medium on Day 5 to inhibit astrocyte proliferation. Dox was withdrawn from Day 7 onwards.

2.2 Molecular cloning

Flag-HA NGN2 OPTi-OX vector was generated using the NGN2 OPTi-OX vector (Pawlowski et al., 2017) as a backbone. The human NGN2 sequences was originally obtained from a plasmid that was received as a gift from Oliver Brüstle.

2.3 Gene targeting

The hROSA26 locus was targeted using a CRISPR/Cas9 gene editing system while the AAVS1 with a zinc finger gene editing system. For both loci, gene targeting was accomplished by nucleofection. At least 7 hours prior to nucleofection, cells were fed with maintenance media supplemented with Rock-inhibitor (10µM). On the day of nucleofection, cells were dissociated into single cells with StemPro Accutase for five minutes. Two million cells were used in a 100µl total volume and 12µg of DNA (two plasmids, each carrying guides or zinc fingers for the 5' or 3' end of the locus and the donor plasmid; 4µg of each plasmid). Nucleofections were performed using the Lonza P3 Primary Cell 4D-Nucleofector X Kit and the cycle CA-137 on a Lonza 4D-Nucleofector System, according to manufacturer's instructions. Subsequently, the cells were seeded onto 2 vitronectin-coated 10cm plates at a density of 1 million cells per plate, in cloning medium that consisted of P/S-free E8 media and CloneR (StemCell Technologies). Cells were

left in cloning medium for 48 hours, before a full medium change with cloning medium (E8 supplemented with P/S was used here onwards). The following day, the medium was exchanged cloning medium with 25% of the initial concentration of CloneR. The day after, use of CloneR in the medium was ceased and antibiotic selection was initiated using, either 100 µg/ml of G418 (Thermo Scientific) for hROSA26 targeting or 1µg/ml of puromycin for AAVS1 targeting. G418 selection proceeded for at least 5 days and puromycin selection was carried out for at least two days. Colonies were picked and expanded after around 10-14 days of culture. Clones were then screened for correct targeting by PCR.

2.4 PCR screening of clones

Genomic DNA was isolated using the Wizard Genomic DNA Isolation Kit (Promega) according to manufacturer's instructions. The PCR was performed with LongAmp Polymerase (NEB) according to the following protocol:

PCR grade H2O	4.1 µl
100% DMSO	0.2 µl
10 mM dNTPs	0.3 µl
5x buffer	2.0 µl
Primer F (10 µM)	0.5 µl
Primer R (10 µM)	0.5 µl
DNA (50 ng/µl)	2.0 µl
LongAmp Polymerase	0.4 µl

Through gel agarose electrophoresis, PCR products were then analysed for size against a 1 kb Hyperladder (BioLine).

The genotyping strategy carried using a list of primers described in Table 1.

2.5 Immunocytochemistry

Prior to immunocytochemistry, cells were washed in PBS to remove debris and medium and fixed in 4 % paraformaldehyde (in PBS) for 20 minutes at room temperature and subsequently washed again in PBS. Then, the cells were blocked and permeabilised in 0.02 % saponin with 10 % normal goat serum (in PBS) for 30 minutes. After that, the cells were incubated with the primary

antibodies in 2% goat serum for 2-3 hours at room temperature or at 4°C overnight. After three 5-minute washes with PBS, fluorescent-labelled secondary antibodies in 2% goat serum were added and allowed to incubate at room temperature for 1 hour. Then, cells were given two 5-minute washes with PBS, followed by a 10-minute incubation in 4',6-diamidino-2-phenylindole (DAPI). Lastly, they were washed once with PBS and kept in PBS for imaging. hiOPC-hiN co-cultures were imaged with Olympus IX71 microscope or a Zeiss LSM700 confocal microscope.

Locus	PCR type	Primer	Primer location	Primer sequence	Amplicon ctr	Amplicon target	Amplicon plasmid
AAVS1	WT (locus)	FW	Genomic; 5' to 5' HAR	CTGTTTCCCCTCCAGGCGAG	1692 bp	Variable; missing if using CAG promoter (GC-rich)	no band
		REV	Genomic; 3' to 3' HAR	TGCAGGGGAACGGGGCTCA			
AAVS1	5'INT	FW	Genomic; 5' to 5' HAR	CTGTTTCCCCTCCAGGCGAG	no band	1103bp	no band
		REV	Puromycin	TCGTCGGGGGTGGCGAGGCGCACCG			
AAVS1	3'INT	FW	NGN2 cds 3'end	AGCTGCACCTTATCGCCCG	no band	1282 bp	no band
		REV	Genomic; 3' to 3' HAR	TGCAGGGGAACGGGGCTCA			
AAVS1	5' BB	FW	Backbone; 5' to HAR	ATGCTTCCGGCTCGTATGTT	no band	no band	1227 bp
		REV	Puromycin	TGAGGAAGAGTTCTTGACAGCTC			
AAVS1	3' BB	FW	NGN2 cds 3'end	AGCTGCACCTTATCGCCCG	no band	no band	2036 bp
		REV	Backbone; 3' to 3' HAR	ATGCACCACCGGGTAAAGTT			
ROSA	WT (locus)	FW	Genomic; 5' to 5' HAR	GAGAAAGAGGCTGTGCTTCGG	2186 bp	Variable; missing if using CAG promoter (GC-rich)	no band
		REV	Genomic; 3' to 3' HAR	ACAGTACAAGCCAGTAATGGAG			
ROSA	5'INT	FW	Genomic; 5' to 5' HAR	GAGAAAGAGGCTGTGCTTCGG	no band	1264 bp	no band
		REV	Splice acceptor	AAGACCCGGAAGAGTTGTCC			
ROSA	3'INT	FW	rTTA	GAAACTCGTCAAAAAGCTGGG	no band	1807 bp	no band
		REV	Genomic; 3' to 3' HAR	ACAGTACAAGCCAGTAATGGAG			
ROSA	5' BB	FW	Backbone; 5' to HAR	CGTTGTAAAACGACGGGCCAG	no band	no band	1148 bp
		REV	Neomycin	GTGCCCAGTCATAGCCGAAT			
ROSA	3' BB	FW	rTTA	CTGGCACGTGAAGAACAACAGC	no band	no band	1752 bp
		REV	Backbone; 3' to 3' HAR	TGACCATGATTACGCCAAGC			

Table 2. 1: OPTi-OX genotyping primers and strategy

2.6 Multi-electrode array (MEA) recordings

The electrical activity of cultured iNs was recorded using an MEA2100-System (Multichannel Systems, Reutlingen, Germany) with an integrated amplifier. Each MEA dish (60MEA100/10iR-Ti; Multichannel Systems) contained 64 electrodes (TiN, 30 μm diameter) arranged over an 8 \times 8 square grid. Recordings started 10min after the MEA plates were placed on the head stage, which was set to 37 $^{\circ}\text{C}$. All MEA recordings were performed in culture medium and each recording lasted 10 minutes. The electric signals were collected at 10 kHz using MCRack (Version 4.4.2; Multichannel Systems) and analysed offline. Spontaneous activity was recorded from Day 7 up to Day 27 post-induction, with recordings taking place at 10 am, on Monday, Wednesday and Friday of each week.

Treatment with tetrodotoxin (TTX, Sigma) was performed by first performing a routine recording as described above. Then, a full media change was carried out with regular iN media containing 50 μM of TTX. Measurements were started after a 10 minutes incubation. The culture was then washed with fresh medium and allowed to settle down in the incubator for 20 min before carrying out post-treatment recording.

2.7 MEA data analysis

The software MCRack was used for spike detection. Raw MEA data were first high-pass filtered at 200 Hz to remove low-frequency local field potentials. Spikes were detected using a threshold-based detector set to a downward excursion beyond 3.0 or 4.0 \times the standard deviation (calculated from 500 ms of filtered data that did not contain spike activity) above the peak-peak noise level. Bursts were detected using the “Burst detection” option in MCRack, where the min. spike interval was set to 10ms and the min. duration of burst set to 20ms. A synchronised burst firing (SBF) was defined as bursts that occur with an inter-burst interval of less than 200ms. From this, the *mean number of electrodes with SBF* was determined by averaging the number of electrodes with SBFs at five random time points during the 10-minute recording. The *mean burst rate (per minute)* was determined by averaging the total number of bursts detected for each electrode during the 10-minute recording, for all 64 electrodes. This value was then divided by ten.

2.8 Karyotyping

To prepare the targeted human PSCs for chromosome analysis, cells that had grown to 60-80% confluency in a 10cm dish were incubated in fresh culture media supplemented with Rho-kinase inhibitor Y-27632 (5 μM , Tocris) and KaryoMAX Colcemid (100ng/ml, Thermo Fisher) for 4h at

+37°C. Subsequently, cells were harvested as single cells, washed, and pelleted. To achieve nuclei swelling and spreading of the chromosomes, cells were treated with hypotonic 0.055M KCl-solution for 5-10 minutes. Finally, cells were preserved in fixative (mixture of absolute methanol and glacial acetic acid in the ratio 3:1), centrifuged and the pellet stored at -20°C until chromosome analysis. Chromosome analysis was done by the Medical Genetics Service at Addenbrooke's Hospital Cambridge.

2.9 Bulk RNA-sequencing

For neurons co-cultured with or without glia, total RNA was isolated in technical triplicates for each timepoint, where 3 independent 6 wells were derived from the same passage for each timepoint. For cell lysis and RNA extraction, cells were firstly treated with 5 minutes of StemPro Accutase to detach cells from culture surface and washed once to remove the accutase. Cell pellets were either flash frozen and stored at -80C, or immediately processed for RNA extraction using the Direct-zol RNA Miniprep kit (Zymo Research, R2071). Transcriptome libraries were generated at the Wellcome Sanger Institute with the Illumina TruSeq stranded RNAseq kit. All samples were sequenced using Illumina HiSeq with 40 million mapped reads on average for each sample.

2.9.1 Bulk RNA-sequencing data analysis

Adapters and basic read quality filter was done using utilities the from the biobambam2 package (Tischler & Leonard, 2014). Reads were aligned using the STAR aligner v 2.5.3a using the GRCh38 genome assembly (Dobin et al., 2013). Reads without a unique mapping location or non-canonical splicing were excluded from the analysis. To generate the gene expression vectors, featureCounts was used from the package, Subread-1.6.3, with the default parameters and the GRCh38.91 gene annotation file. Pre-processing involved removal of non- and low-expressed genes (retained only genes with at least 1 count per million aligned reads in at least 3 of the samples), followed by removal of outlier samples. Additionally, the data was normalized by variance stabilising transformation, using the R package DESeq2 (Love et al., 2014). DESeq2 was also used to perform differential expression (DE) analysis with a minimum $|\log_2(\text{fold change})| > 2$ and false discover rate (FDR) < 0.05 . Enrichment analysis of significantly DE genes was carried out using gene ontology collections from the Gene Ontology Consortium (Ashburner et al., 2000; The Gene Ontology Consortium, 2019), where, as background only 17844 protein-coding genes with valid Entrez ID were used.

For WGCNA (Langfelder & Horvath, 2008a) a Topological Overlap (TO) matrix (B. Zhang & Horvath, 2005) was constructed using a soft-thresholding power of 6 (the default for WGCNA) since the scale free topology fit index roughly saturates at around 6. WGCNA uses a topological overlap-based dissimilarity as input to average-linkage hierarchical clustering that results in a dendrogram. Modules were identified as branches in the dendrogram using Dynamic Tree Cut (Langfelder, Zhang, & Horvath, 2008). After module identification, enrichment analysis of genes in each module were evaluated using the following collections of literature sets:

1. Brain cell type and region markers, compiled from various literature sources by Jeremy A. Miller for the `userListEnrichment` function in WGCNA (Miller et al., 2011) and by PL for the `anRichment` R package;
2. BioSystems gene sets, including KEGG, Reactome, Lipid Pathways and BioCYC;
3. Gene Ontology (GO) (The Gene Ontology Consortium, 2019).
4. Molecular Signatures Database (MSigDB) version 6.2 (Subramanian et al., 2005);
5. Genomic position gene sets; each set contains genes in a 5 Mb window, with two adjacent windows overlapping by 2.5 Mb;
6. Enrichr 2016 ChEA library (E. Y. Chen et al., 2013);
7. Enrichr 2015 ENCODE histone modification library;
8. Enrichr 2015 ENCODE TF ChIP-seq library;
9. Enrichr 2017 mirTarBase library (Chou et al., 2018).

Again, as background for the enrichment calculations, only those 17844 genes that are protein-coding and have valid Entrez were used. This restriction avoids bias toward large terms (and strong p-values) from having less studied non-coding genes in the enrichment background. The calculations were carried out using R package `anRichment` (<https://horvath.genetics.ucla.edu/html/CoexpressionNetwork/GeneAnnotation/>) that implements standard Fisher exact test and a multiple-testing correction across all query and reference gene sets.

2.9.2 Bulk RNA-seq data analysis of neurons co-cultured with glia against neurons cultured without glia

Mixed-species RNA-seq reads from different timepoints (4, 14, and 21 days) were separated to species of origin with the Sargasso pipeline version 2.0.1 (Qiu et al. 2018). Sargasso was run with the “--conservative” option designed by the program authors.

The resulting aligned reads to the human genome from Sargasso were converted from *.bam files to *.fastq files with samtools version 1.7 (Li et al., 2009) and separated to paired-read files with

the unix command line tool grep. Paired-end fastq files were supplied to Salmon version 0.14.0, with options `-l A` and `--validateMappings` to determine the type of library automatically and to enable a more sensitive and specific mapping (Patro, Duggal, Love, Irizarry, & Kingsford, 2017). The Gencode human version 30 transcript was provided as a reference

(ftp://ftp.ebi.ac.uk/pub/databases/gencode/Gencode_human/release_30/gencode.v30.transcripts.fasta.gz).

Human-only reads provided as *.fastq files were aligned with Salmon with the same specifications to the same reference.

For gene differential expression, outputs from Salmon were imported into R with the `tximport` package version 1.10.1 (Soneson, Love, & Robinson, 2016) and tests of differential expression across each timepoint (4, 14 and 21 days) were conducted with R package `DESeq2` version 1.22.2 with an adjusted p - value of 0.05 (Love et al. 2014). The lowest mean adjusted read count present in genes determined to be differentially expressed was used as a filter to consider a gene expressed. Enrichment analysis of significantly DE genes was carried out as described as before, in section 2.9.1

2.10 Dissociation of iPSCs and NGN2 iNs for single cell RNA-sequencing

Each sample/timepoint was harvested from one well of a 6-well plate, except for Day 0, 12 hours and 1, which were harvested from 2 wells of 6 well plate. First, media was aspirated from each well and given a gentle one-time wash with dPBS (calcium and magnesium free). For Day 0 (iPSCs that have been seeded and kept in ROCK-I for 24 hours), 12 hours, Day 1, Day 2, Day 3 and Day 4 with and without rat glia co-culture, samples were treated with 1ml/well of plain Accutase for 6 minutes at 37°C. For Days 14 and 21 (with and without glia), samples were treated with 1ml/well of dissociation solution made up of papain (Worthington) resuspended in Accutase, for a final concentration of 20 U/ml. These late time points were then left in the incubator for 30 minutes at 37°C. A plain rat glia sample that had been cultured with the same culture protocol for iNeurons for up to 21 days, were also treated with this dissociation solution for 30 minutes at 37°C. At the end of their incubation period, 1 ml of dissociation buffer, made up of DMEM/F-12, ROCK-I (10µm) and DNase 1 (33µg/ml), was added to each well. Then, the cells were dissociated into a single-cell suspension by pipetting up and down with a p1000, at least 5 times, against the culture surface, as illustrated in Figure 2.1.

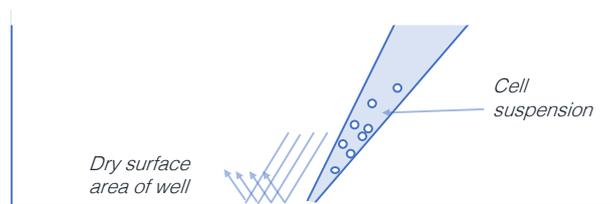


Figure 2. 1: Mechanical dissociation method for increased dissociation of iNeurons into single cells.

Cell suspensions were then transferred into a 15 ml falcon tube capped with a 40µm mini-cell strainer (Pluriselect), containing 1 ml of dissociation buffer (2ml for pooled samples). Wells were further washed with 1ml of dissociation buffer and transferred to the falcon tube. Samples were centrifuged at 300g for 3 mins at room temperature. Pellets were then resuspended in 2 ml of resuspension buffer made up of DMEM/F-12, 0.04% BSA and 10µM ROCK-I. Repeat centrifuge at 300g for 3 mins at RT. Finally, pellets were resuspended in 1ml of resuspension buffer and transferred to a 1.5ml Eppendorf tube capped with a 40µm mini-cell strainer, keeping the cells on ice for the rest of the protocol. Cell viability was determined using a Countess II (Thermofisher Scientific), with 4 counts obtained for each sample. The average viability and cell number was determined and recorded for each sample. From this main cell suspension, a 500 cells/ul cell suspension was prepared for each sample and kept on ice for further processing with the Chromium™ Single Cell 3' Reagent Kit (v2) by 10x Genomics.

2.11 Single cell RNA-sequencing

Each sample prepared in the previous section were loaded for a target recovery of 1000 cells per sample. GEM generation and barcoding steps of the Chromium protocol were carried out at the lab of Andrew Basett (Wellcome Sanger Institute). Samples were then handed over to the Wellcome Sanger Institute DNA Pipeline facility for further steps in the library construction and subsequent sequencing. All samples were sequenced using the Illumina HiSeq platform, with 50,000 reads on average per cell.

2.11.1 scRNA-seq data analysis

scRNAseq samples were processed using the 10x Genomics CellRanger pipeline v3.0.1 using the GRCh38 genome assembly to produce the gene expression matrices. For the glia co-culture experiments, CellRanger v3.0.2 was used using the GRCh38 and Rnor_6.0 genome assemblies. The filtered gene expression matrix was used without modifying any cell gating parameters for both cases. For the mapping of the rat genes to human, the following homology,

http://www.informatics.jax.org/faq/ORTH_dload.shtml was used. The QC filtering was done using Scanpy (Wolf, Angerer, & Theis, 2018) and involved removal of cells expressing less than 2000 genes, more than 100000 transcripts or more than 12% mitochondrial reads and genes that had less than 500 UMIs or expressed in less than 10 cells (Refer to . For the co-cultures, the cells were annotated as rat species, as human species or as multiplet using the following criteria: if more than 90% of the transcripts were mapping to one reference then the cell was annotated with the species of the dominant genome assembly, otherwise it was annotated as a multiplet. For dimensionality reduction, after normalizing with respect to sequencing depth of each cell and log-normalizing the top 50 principal components were calculated using only the highly variable genes which were found by using the `highly_variable_function` of scanpy with the default parameters. For visualization, UMAP embedding method was used after calculating the 15 nearest neighbors in the PCA space using the Euclidean metric (McInnes, Healy, Saul, & Großberger, 2018). The same UMAP embeddings were used for clustering of the cells using the Louvain method of scanpy by setting a more coarse resolution of 0.5 (Traag, Waltman, & van Eck, 2019). The Cell Cycle annotation of the cells was done using `score_genes_cell_cycle` of scanpy and the `regev_lab_cell_cycle_genes.txt` found here:

https://github.com/theislab/scanpy_usage/blob/master/180209_cell_cycle/data/regev_lab_cell_cycle_genes.txt.

The differential expression analysis was done using the `rank_gene_groups` of scanpy using the `t-test_overestim_var` method with multiple testing correction. Pseudotime analyses were performed with the monocle R package (version 3.0) (Trapnell et al., 2014a).

2.12 Bulk ATAC-sequencing

Each sample/timepoint was harvested from 2 wells of a 6 well-plate, with 2 biological replicates per timepoint. Media was removed and each well was washed once with dPBS (calcium and magnesium free). Then, they were treated with 1 ml per well of Accutase for 3 mins at 37 °C. Accutase was then aspirated off and cells lifted and dissociated in 3 ml of DMEM/F-12 for Days 0 to Day 2, and Neurobasal for the rest of the timepoints. Cell suspensions were transferred to a 15 ml falcon tube for centrifugation at 300g for 3 mins at RT. The pellet was then resuspended in 500 µl of ice-cold dPBS. From this main suspension, 500,000 viable cells per sample were transferred into a 1.5ml Eppendorf tube for centrifugation at 300g for 3 mins at RT. Each pellet was then resuspended in 500 µl of ice-cold sucrose buffer (nuclease free-water made up of 10mM Tris-Cl, pH7.5; 3mM CaCl₂; 2mM MgCl₂ and 0.32M sucrose) and kept on ice for 12 minutes.

Following that, cells were lysed by adding 25 μ l of freshly-prepared 10% Triton X-100 to each suspension and keeping on ice for a further 6 minutes. From each suspension, 100,000 nuclei were transferred to a 1.5 ml Eppendorf tube and centrifuged at 450g for 5 minutes at 4°C. The supernatant was then aspirated-off, making sure to remove as much as possible. Tagmentation and multiplexing was done using the Nextera DNA Sample Preparation kit (FC -121-10130) and Nextera XT Index kit (FC-131-1002), respectively, according to manufacturer’s protocols. Briefly, each nuclei pellet was resuspended in 50 μ l of master mix made up of 25 μ L of TD buffer, 20 μ L of water and 5 μ L of TDE1, and transferred to a 1.5 ml DNA LoBind Eppendorf tube. Each reaction was then incubated at 37°C for 30 mins. The reaction was stopped using 250 μ l of buffer PB (Qiagen PCR Clean-up Kit) and processed using the Qiagen PCR Clean-up Kit, according to manufacturer’s protocol, eluting the tagmented samples in 11.5 μ l of EB buffer. A PCR was then carried out using the following protocol:

Per reaction

Template from tagmentation	10uL
I7 primer	2.5ul
I5 primer	2.5ul
Nextera cocktail	2.5ul
Nextera PCR mastermix	7.5 μ l

Cycling as follows:

1	72 °C 3 minutes
2	98 °C 30 seconds
3	98 °C 10 seconds
4	63 °C 30 seconds

5	72 °C 3 minutes
6	Repeat step 3- 5 (x11).
7	Hold at 10 °C

Each PCR product was made-up to 45 μ l with EB buffer and mixed with 5x loading dye for the following steps. Primer removal and size selection was done using agarose gel electrophoresis. Bands between 120 and 1 kb were removed and processed using a Qiagen Gel Extraction Kit, eluting each sample in 20 μ l of EB buffer. Samples were then submitted to the Wellcome Sanger Institute DNA Pipeline facility sequencing using the Illumina HiSeq platform, with an average of 40 million reads per sample.

2.12.1 Bulk ATAC-seq data analysis

Raw sequence reads were trimmed with *cutadapt* (Martin, 2011) to remove sequencing adapters and low quality bases and subsequently aligned to the GRCh38 reference genome assembly using the Burrows-Wheeler Alignment (BWA) tool (H. Li & Durbin, 2009). Aligned reads were filtered to exclude unmapped reads and secondary alignments and written to bam format with *samtools* (Li et al., 2009). Reads mapping to chrM were removed with *awk* and duplicates were removed with *picard's MarkDuplicates*. Peaks of chromatin accessibility were predicted with the Model-based Analysis of ChIP-Seq (MACS2) (Zhang et al., 2008), combining replicate samples for peak prediction. Nucleosome occupancy profiles were determined using NucleoATAC (Schep et al., 2015). Differential analysis of ATAC-Seq peaks, from cell lines with and without glial support, were analyzed with DASTk (Tripodi, Allen, & Dowell, 2018) to assess changes in transcription factor activity. Putative binding sites or TF motifs corresponding to GRCh38 were defined by Homo sapiens Comprehensive Model Collection (HOCOMOCO) v11 (Kulakovskiy et al., 2018).

2.12.2 Bulk ATAC-seq data analysis of neurons co-cultured with glia against neurons cultured without glia

Mixed-species ATAC-seq reads from across timepoints were assigned to either human or rat reference genomes with the Sargasso pipeline version 2.0.1 (Qiu et al. 2018). Sargasso was run with the "--conservative" option created by the pipeline authors.

2.13 NGN2 ChIP-sequencing

ChIP-sequencing was performed on biological replicates at Day 1 post-induction. For chromatin immunoprecipitation, a two-step crosslinking protocol (Tian, Yang, & Brasier, 2012) using freshly-prepared 1mM DSG and 1% formaldehyde, was carried out on approximately 2.5 million cells per sample. These cross-linking reagents, ChIP lysis buffer and wash buffers were prepared as follows:

DSG, 125mM solution

9mg of DSG dissolved in 220µl of DMSO, prepared fresh just before cross-linking.

Formaldehyde, 1%

16 times dilution in PBS of Pierce™ 16% Formaldehyde (w/v), Methanol-free (Sigma). This was prepared fresh on the day.

Lysis Buffer 1

50 mM Hepes-KOH, pH 7.5; 140 mM NaCl; 1mM EDTA; 10% Glycerol; 0.5% Igepal (CA-630); 0.25% Triton X-100.

Lysis Buffer 2

10 mM Tris-HCL, pH8.0; 200 mM NaCl; 1 mM EDTA; 0.5 mM EGTA.

Lysis Buffer 3

10 mM Tris-HCl, pH 8; 100 mM NaCl; 1 mM EDTA; 0.5 mM EGTA; 0.1% Na-Deoxycholate; 0.5% N-lauroylsarcosine.

Elution buffer

1% SDS, 0.1M NaHCO₃

Wash buffer 1

0.1% SDS, 1% Triton X100, 2mM EDTA-NaOH pH 8.0, 150mM NaCl, 20mM Tris-Cl pH 8.0.

Wash buffer 2

0.1% SDS, 1% Triton X100, 2mM EDTA-NaOH pH 8.0, 500mM NaCl, 20mM Tris-Cl pH 8.0.

Wash buffer 3

0.25 M LiCl, 1% NP-40, 1% Na-Deoxycholate (Sodium deoxycholate), 1mM EDTA-NaOH pH 8.0, 10mM Tris-HCl pH 8.0.

Cells were first dissociated using Accutase and resuspended in 5 ml of PBS. 80 μ l of 125mM DSG was then added to each 5 ml suspension (final DSG concentration = 2mM) and incubated at RT for 45 minutes. Cells were then washed with PBS three times before being resuspended in 1% formaldehyde, for 10 minutes at RT. Then, cells were treated with 0.125mM glycine for 5 minutes at RT, before being mixed with an equal volume of ice-cold PBS containing protease inhibitors. Cells were then pelleted at 2000g for 10 minutes at 4°C. Next, samples were washed by resuspending in ice-cold PBS containing 10% FBS and then centrifuged again at 2000g for 5 minutes at 4°C. Cells were then resuspended in 100 μ l Lysis buffer 1 containing protease inhibitors and left on ice for 5 minutes, mixing gently every now and then, followed by centrifugation at 2000g at 4°C for 5 minutes. This was then repeated for 100 μ l of Lysis buffer 2 containing protease inhibitors. Next, cells were resuspended in 100 μ l of Lysis buffer 3 and each sample then transferred to a microTUBE AFA Fiber Snap-Cap (PN 520045) for sonication using the Covaris E220 Sonicator with the following settings: treatment time=430sec, peak incident power=175, duty factor 10%, cycles per burst= 200. 10 μ l of 10% Triton X-100 was added to each sonicated lysate and left on ice for 5 minutes before using centrifuged at 16,000 g for 10 minutes at 4°C. The supernatant was then transferred to a fresh Eppendorf tube and pre-cleared with 10 μ l of Pierce™ Control Agarose Resin (pre-washed twice) beads for 1h at 4°C on a rotating wheel. Beads were then removed by centrifugation at 2000 g for 30 s at 4C. Then, 10% of lysate was kept aside at -20°C, as the Input sample. For antibody mix, a 20 μ l 1:1 slurry of α -HA resin (Roche 3F10) (no need prior wash) was mixed with a 20 μ l 1:1 slurry of Pierce™ Control Agarose Resin (pre-washed). This 40 μ l antibody mix was then added to each remaining lysate and incubated overnight at 4°C on a rotator. The next day, each sample was washed with 1ml of Wash Buffer 1, centrifuging at 2000g for 30s at 4°C. Each sample was then resuspended in 1ml of Wash Buffer 2 and transferred to a new tube before centrifugation as before. This was followed by washing with Wash Buffer 3 and two-time wash with TE buffer, keeping samples in the same tube. Samples were then centrifuged again at 2000g for 30s at 4°C and resuspended in 110 μ l of Elution buffer, and incubated, first statically at 65°C for 5 minutes and subsequently for 30 minutes at RT with vortexing at 800 rpm. Beads were then removed by centrifugation at 10,000g for 10s. 100 μ l of the eluant was transferred to a new tube. Before proceeding with the following steps, input samples were thawed and topped-up to 110 μ l with elution buffer. NaCl was then

added to all eluates and input samples to a final concentration of 200 mM for reverse cross-linking at 65°C, overnight. The next day, 100 µl of TE buffer was added to each sample to dilute the SDS. Then, 4 µl of 1mg/ml of RNaseA (Ambion) was added to each sample, mixed and incubated for 30 minutes at 37°C. This was followed by incubation with 2 µl of 20mg/ml of Proteinase K (Invitrogen) for 1 hour at 55°C. Finally, DNA was purified using Zymo DNA Concentrator kit according to manufacturer's protocol, eluting samples in a 60 µl volume into 1.5 ml DNA LoBind Eppendorf tubes. Library preparation was performed using the MicroPlex Library Preparation kit v2 by Diagenode, according to manufacturer's instructions. Sequencing was performed using the Illumina HiSeq platform at the Wellcome Sanger Institute, with an average of 40 million reads per sample.

2.13.1 NGN2 ChIP-seq data analysis

Raw ChIP sequence reads were trimmed with Trim_galore (version 0.6.1) that performed trimming of the low-quality sequences and adapter dimers (--paired -q 25). The trimmed reads were aligned against the GRch38(hg38) genome using bwa-mem aligner (BWA) with the default setting. Only uniquely mapped reads were used for downstream analysis (e.g peak calling and motif analysis). Peak-calling was performed using MACS2 (version 2.1.1). The uniquely mapped reads were used to predict potential Ngn2 binding sites against input control (default setting 0.05). A stringent cutoff ($Q < 0.01$) was used.

3 Characterisation of NGN2 iNeurons

3.1 Introduction

The generation of NGN2 iNeurons using our OPTi-OX technology demonstrated the utility of this technology for generating mature human cell types from PSCs in a rapid and efficient manner. Pawlowski and colleagues carried out a rudimentary characterisation on the iNs (Pawlowski et al., 2017). This involved an immunocytochemistry analysis of important neuronal markers and qPCR analysis of both pluripotency and neuronal genes at days 0, 7 and 14 of NGN2 reprogramming. The results indicated that a complete downregulation of pluripotency genes had occurred at day 4 post-induction, with a concomitant upregulation of neuronal genes such as MAP2, TUBB3, BRN2, VGLUT2 and SYP. At day 7 onwards, iNeurons strongly expressed forebrain markers BRN2 and FOXG1, and glutamatergic neuronal genes GRIA4, VGLUT1 and VGLUT2, indicative of an excitatory cortical neuronal identity, consistent with previous reports (Zhang et al., 2013; Busskamp et al., 2014). In addition, the study showed that transient administration of doxycycline (dox) treatment for 4 days was sufficient for complete conversion and that the cells did not rely on continuous overexpression of NGN2. Finally, the study also demonstrated that near 100% of the resulting cultures at day 7 stained positive for the pan-neuronal marker TUBB3.

While this study provided a timeline and preliminary understanding of the important morphological and transcriptional changes of NGN2 reprogramming, an in-depth analysis of the transcriptional and electrophysiological properties of our NGN2 iNs was not investigated. Based on other reports of iNs, action potentials can be detected as early as 7 days post-induction, with the resulting spontaneous activity being glutamatergic in nature and largely mediated by AMPA-receptors (Busskamp et al., 2014; Lam et al., 2017; Yingsha Zhang et al., 2013). When co-cultured with rat glia, NGN1/NGN2 iNs could be maintained in culture for extended periods, showing many hallmark properties for neuronal maturation, including robust electrical and synaptic activity (Lam et al., 2017).

The majority of electrophysiological studies on iNs have been conducted using patch-clamp approaches. Alternatively, recent years have seen an increase in the use of multi electrode arrays (MEAs) to assess electrophysiological properties of neurons. This approach is complementary to the patch clamp approach and allows further investigation: MEAs can be used for non-invasive recording of extracellular signals simultaneously from multiple electrodes at a time. This allows for studying network development and plasticity over time. Patch clamp approaches, on the other

hand, allow recording from only few cells at a time and the approach is invasive to the cell. So far only three studies have used this approach on iNs (Frega et al., 2017; Ho et al., 2016; Nehme et al., 2018). The most comprehensive MEA study was carried out on NGN2 iNs derived from human ESCs using the protocol from the Wernig lab (Zhang et al., 2013), which most closely resembles the Pawlowski protocol (Frega et al., 2017). There, it was shown that NGN2 iNs develop functionally active neuronal networks, showing spontaneous electrophysiological activity, two weeks after induction. The level of network activity increases in the subsequent weeks, exhibiting synchronous network bursts in the fourth week post-induction. In a subsequent study, Frega and colleagues successfully used this approach to model Kleefstra syndrome, a neurodevelopmental disorder with the clinical features of both Intellectual Disability and Autism Spectrum Disorder, showing that network dysfunction in this model is mediated by NMDAR signalling (Frega et al., 2019). Therefore, studying the network development of our NGN2 iNs is not only an important question in the context of this PhD, but also to establish whether it too can be used to study neuronal networks in both diseased and healthy models.

To study and characterise its functional and mature properties, it was crucial to have a protocol that would allow culturing iNeurons for extended periods. The existing protocol detailed in Pawlowski et al. (2017) resulted in cells aggregating their cell bodies forming clumps within two weeks of culture. This hinders any studies of their electrophysiological properties on MEAs, as the approach depends on cultures having an even spread of cells over its electrodes. Besides this, after three weeks of culture NGN2 iN cultures easily detach from the bottom of the wells, making it impossible to conduct long-term studies of iN development. Therefore, developing a culture protocol that permits long-term culture is paramount.

3.2 Aims

The aim of my PhD project was to uncover the mechanisms underlying NGN2 reprogramming in our OPTi-OX based model, through an integrative time course genomic analysis of this process. To design a study that best captures important hallmarks of this process, it is important the study is based on a solid characterisation of the iNs. Additionally, crucial to the planned ChIP-seq analysis of NGN2 binding, is the quality of the antibody for detecting NGN2. Currently, there aren't any verified ChIP grade antibodies for NGN2 in the market. Alternatively, a popular approach is to engineer a recombinant version of the protein that expresses Human influenza hemagglutinin (HA), which is a surface glycoprotein required for the infectivity of the human virus. The HA tag has been extensively used as a general epitope tag in expression vectors as it

does not appear to interfere with the bioactivity or the biodistribution of the recombinant protein. This tag facilitates the detection, isolation, and purification of the protein.

Taken together, this chapter addresses the following aims:

1) Generation of an HA-tagged NGN2 OPTi-OX hiPSC line

I will present the strategy used for generating this line that will be used for the planned NGS experiments, along with steps taken to ensure a functional line is generated that is homozygously targeted for both components of the OPTi-OX system, with no off-target integration of the backbone.

2) Establishment of an optimised protocol for the long-term culture of iNeurons

To meet this aim, I have tested different substrates commonly used for supporting human- and rodent-derived neurons. In addition, previous electrophysiological studies of NGN2 iNs have indicated that the addition of rodent glia is crucial for the survival and maturation of iNs (Frega et al., 2017; Yingsha Zhang et al., 2013). Therefore, I will also present my attempt of incorporating this strategy to overcome the limitations associated with the current culture protocol.

3) Investigate the electrophysiological properties of NGN2 iNs on MEAs

To characterise the electrophysiological properties of the iNs, I cultured our NGN2 iNs on MEAs to investigate the following questions:

- 1) Do iNeurons show spontaneous electrical activity?
- 2) Do iNeurons demonstrate synchronous network activity over time?

3.3 Results

3.3.1 Generation of an HA-tagged NGN2 OPTi-OX hiPSC line

For this objective, I first had to modify the AAVS1-targeting vector that we have previously used to generate our NGN2 OPTi-OX line, to incorporate the HA epitope tag. Towards this aim, I used the existing vector as a template to generate a HA-NGN2 fragment using a 5' PCR primer incorporating the HA sequence and a portion of the 5' end of the NGN2 coding sequence (CDS) (Fig. 3.1). This was then cloned into the backbone of the existing vector using restriction digest, replacing the NGN2 CDS with a HA-tagged NGN2 CDS, under the control of a Tet-responsive element (TRE). Then, this vector was used to target the OPTi-OX system into a well-established

hiPSC line referred to as 'Bob' that is routinely used in our facility as a control (Yusa et al., 2011). Cells were then nucleofected on two separate occasions; targeting the hROSA26 locus with the rtTA component and then the AAVS1 locus with the TRE-HA-NGN2 transgene. Several clones for each round of targeting were chosen and genotyped as described in Chapter 2. The goal was to select a clone that was homozygously targeted for both components of the system, with no detectable off-target integration of the backbone vector and no karyotypic abnormalities.

Genetic targeting using CRISPR-Cas or zinc finger approaches have a risk of unwanted off-target insertion of transgenes (X.-H. Zhang, Tee, Wang, Huang, & Yang, 2015) and may result in chromosomal abnormalities (Brunet et al., 2009). To determine whether my clones were correctly targeted with the OPTi-OX system, I performed PCR-based genotyping (**Fig. 3.2**). Briefly, for each locus, I assessed if a clone had homozygous integration of the transgene. Then, I determined if the entire length of the transgene had integrated by sequencing the 5' and 3' end of the construct. Lastly, I determined if there were any backbone vector integration by using primers for the 5' and 3' portions of the plasmid backbone. From this I selected two homozygously targeted clones, clones 5 and 6.

I then sought to determine whether the OPTi-OX system is functional in these two clones, i.e. that upon addition of dox, they express high levels of the transgene and can reprogram the iPSCs into neurons. For this, I carried out a qPCR analysis of cells at days 7 and 14 post-induction, checking for the expression of the transgene NGN2 and two pan-neuronal neuronal genes MAP2 and SYP (**Fig. 3.3**). The expression dynamics of NGN2 was as expected, showing high levels of expression at day 7 and significant drop at day 14, due to dox-withdrawal. There were no significant differences between the two clones, for both timepoints. As for the neuronal genes, again, both clones showed similar expression levels for each gene at each time point. As expected, expression for both genes increased from day 7 to day 14, with MAP2 showing the most noticeable increase. Since there were no significant difference between the two clones, I picked clone 5 for subsequent tests.

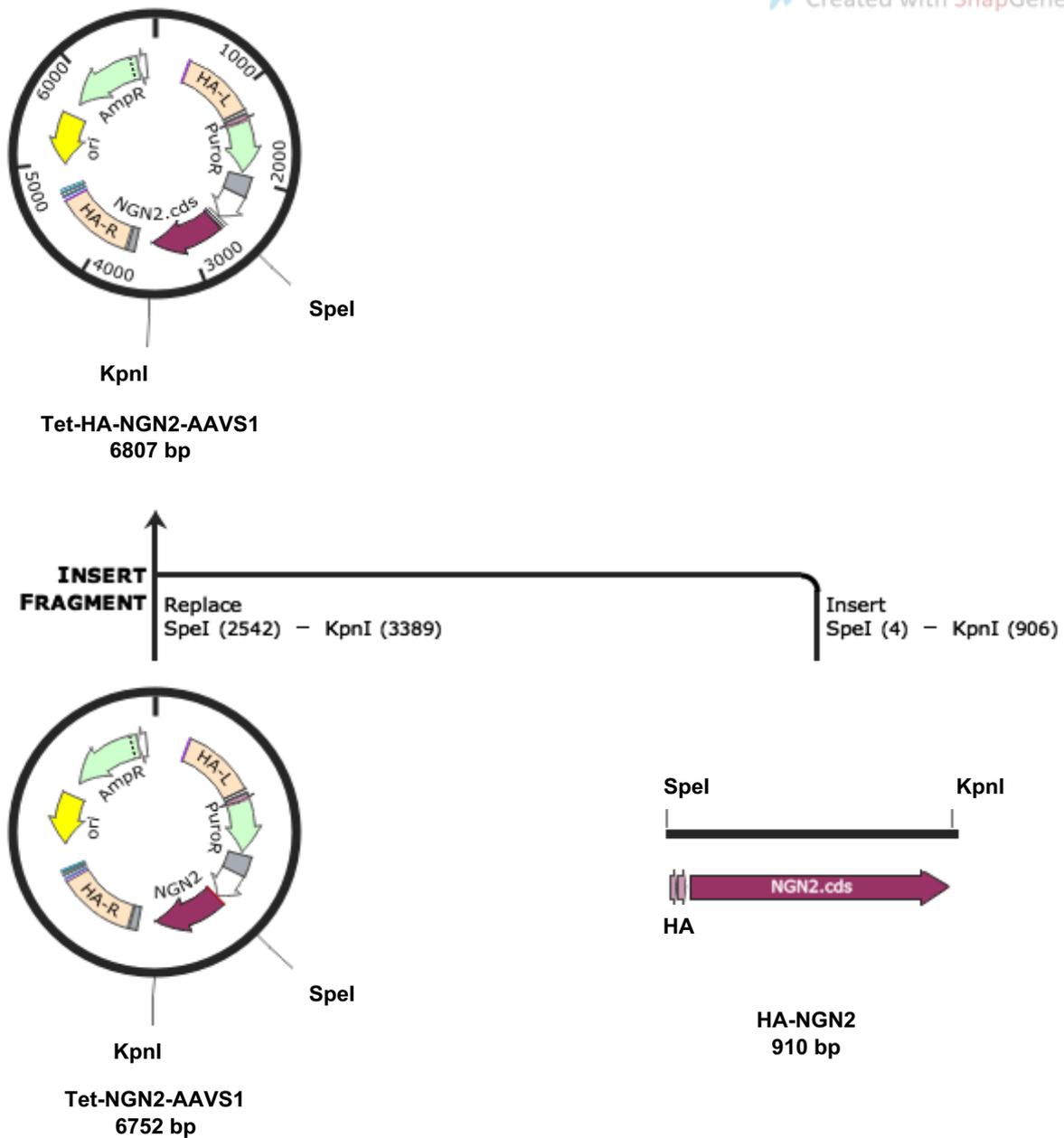


Figure 3. 1: Overview of the cloning strategy for generating a Tet- HA-NGN2 AAVS1 targeting vector.

Using a forward primer incorporating HA sequence, I used the existing Tet-NGN2-AAVS1 plasmid as a template to generate HA-NGN2 fragment. This was then cloned into the Tet-NGN2-AAVS1 backbone using restriction digest.

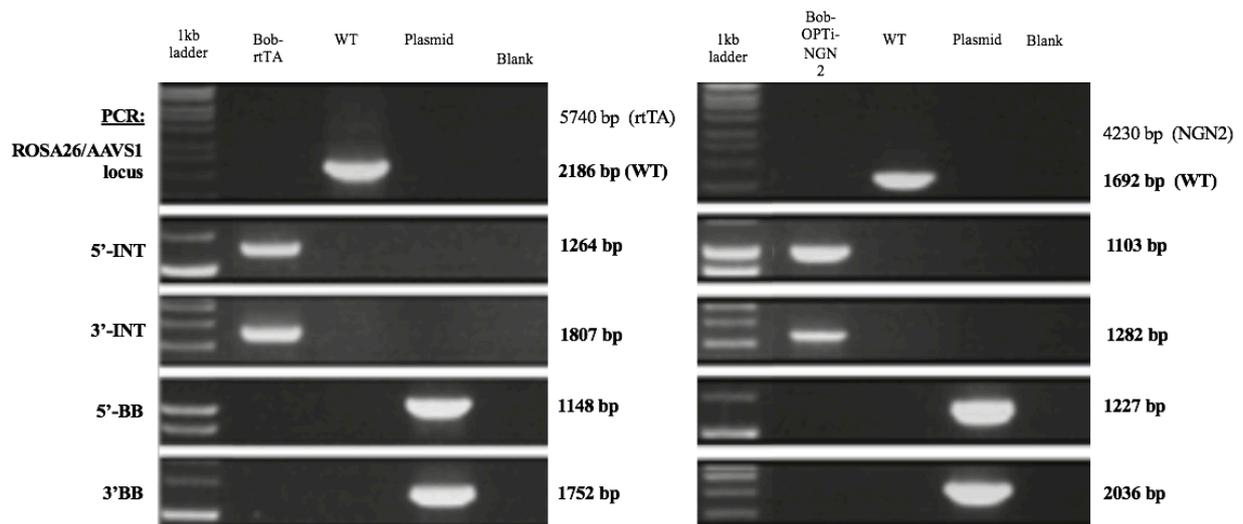


Figure 3. 2: Genotyping results for a clone targeted with HA-NGN2 OPTi-OX.

The gel image on the left shows results for targeting hROSA26 locus of the Bob iPSC line with the rtTA construct, while the gel image on the right for targeting the AAVS1 locus with the responder construct. These results show that the clone was homozygously targeted for both constructs and possess no backbone integration.

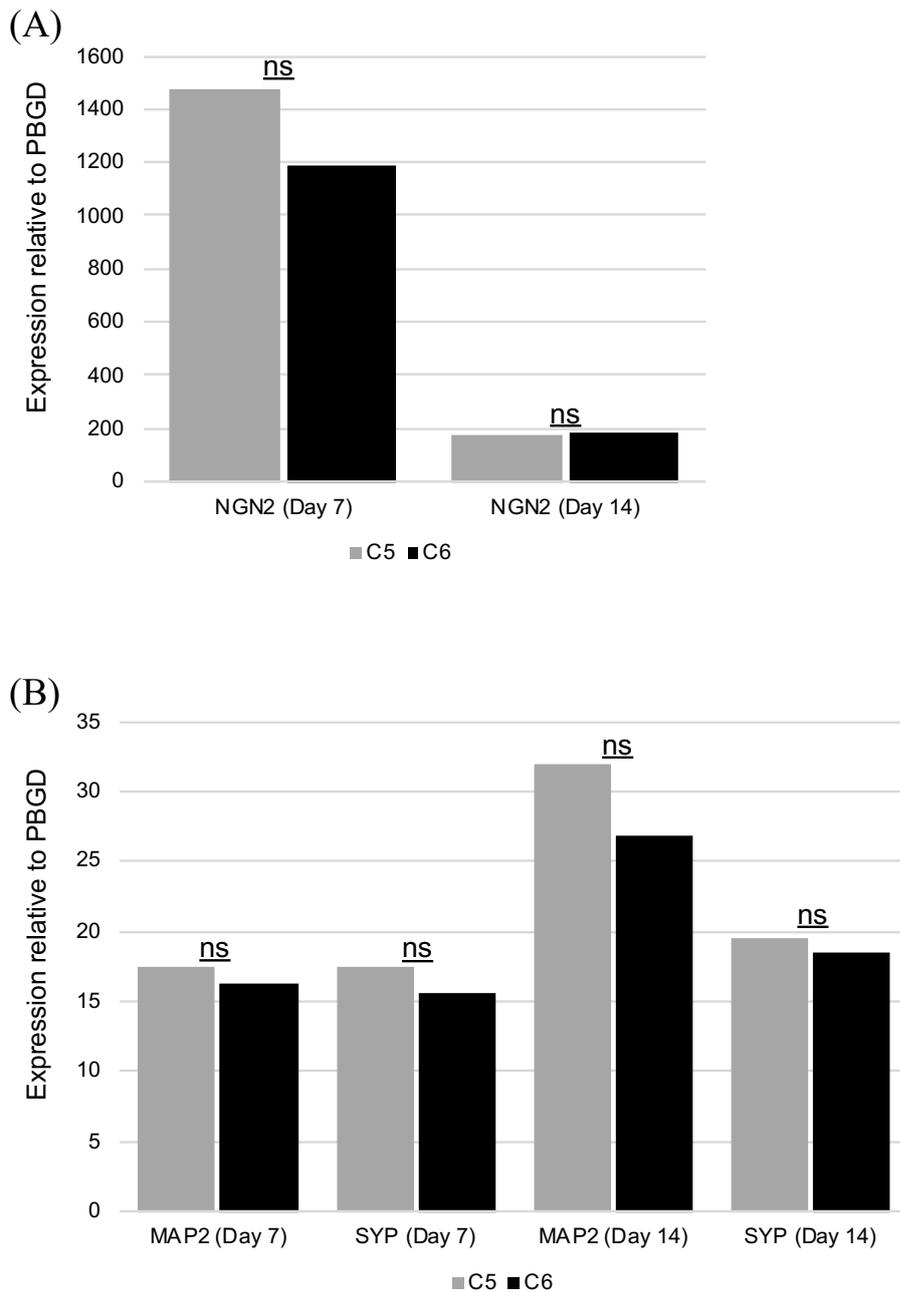


Figure 3. 3: Verification of OPTi-OX system in targeted clones through qPCR analysis.

(A) Expression levels of NGN2 transgene relative to PBGD, in clones 5 and 6 at days 7 and 14 post-induction, showing no significant differences between the two clones. (B) Expression levels of the pan-neuronal genes MAP2 and SYP, for both clones at days 7 and 14. Again, with no significant differences between both clones.

Then, I determined whether the cells expressed HA-tagged NGN2 when induced with doxycycline. For this, I carried out an immunocytochemistry test on cells after 2 days of dox induction, using the same anti-HA antibody that will be used for the NGN2 ChIPseq experiment. Compared to a non-tagged NGN2 line, the clone 5 HA-NGN2 Bob line showed high expression of HA that is localised in the cell nucleus, as expected given NGN2's role as a TF (**Fig. 3.4**). Next, a karyotype test on this line revealed that it was free of any karyotypic abnormalities (**Fig. 3.5**). Together, these results verify that I had successfully generated a functional HA-tagged NGN2 OPTi-OX hiPSC line that is fit for the planned ChIP-Seq experiments.

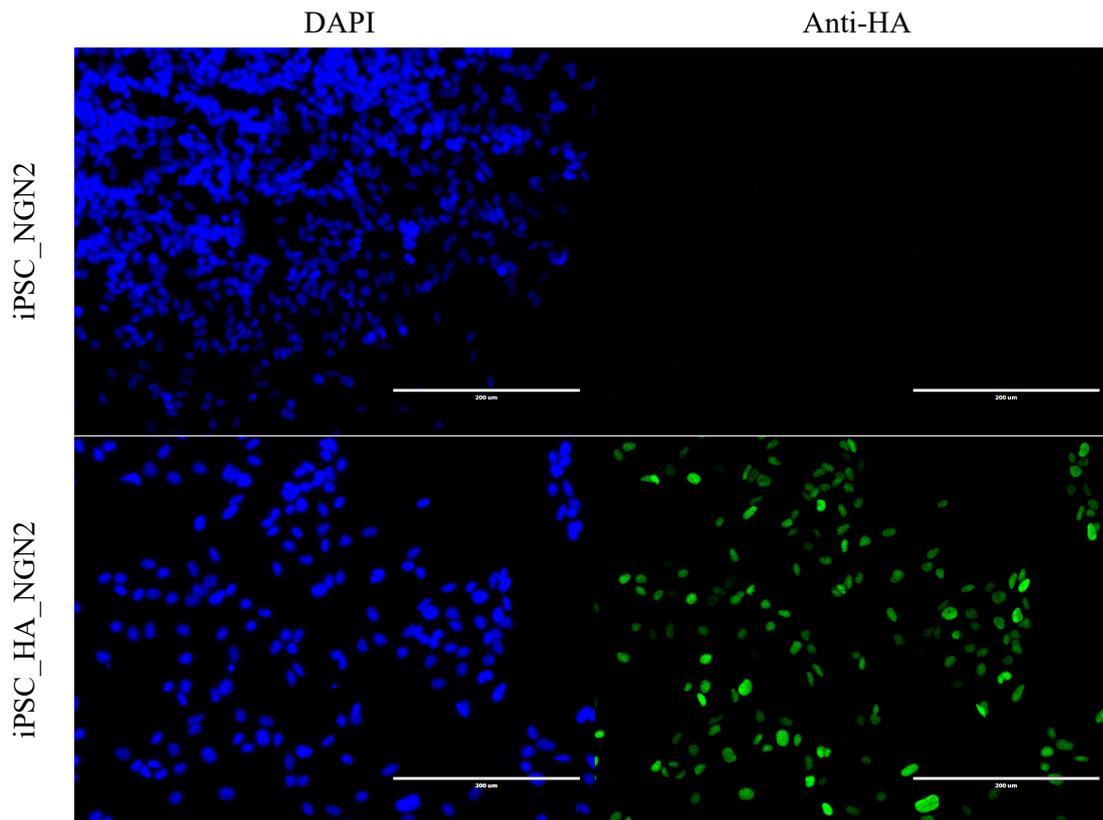


Figure 3. 4: Verification of HA-NGN2 expression.

An untagged NGN2 OPTi-OX iPSC line (top panel) and HA-tagged NGN2 OPTi-OX hiPSC line (bottom panel) were stained with the nucleus counter stain, DAPI (left panel, Blue) and antibodies against the HA epitope tag (right panel, Green), after 2 days of dox-induced transgene overexpression. The HA-tagged line shows clear expression of HA that is localised to the nucleus, consistent with the spatial expression of NGN2. (*Scale bars: 200 μm*).

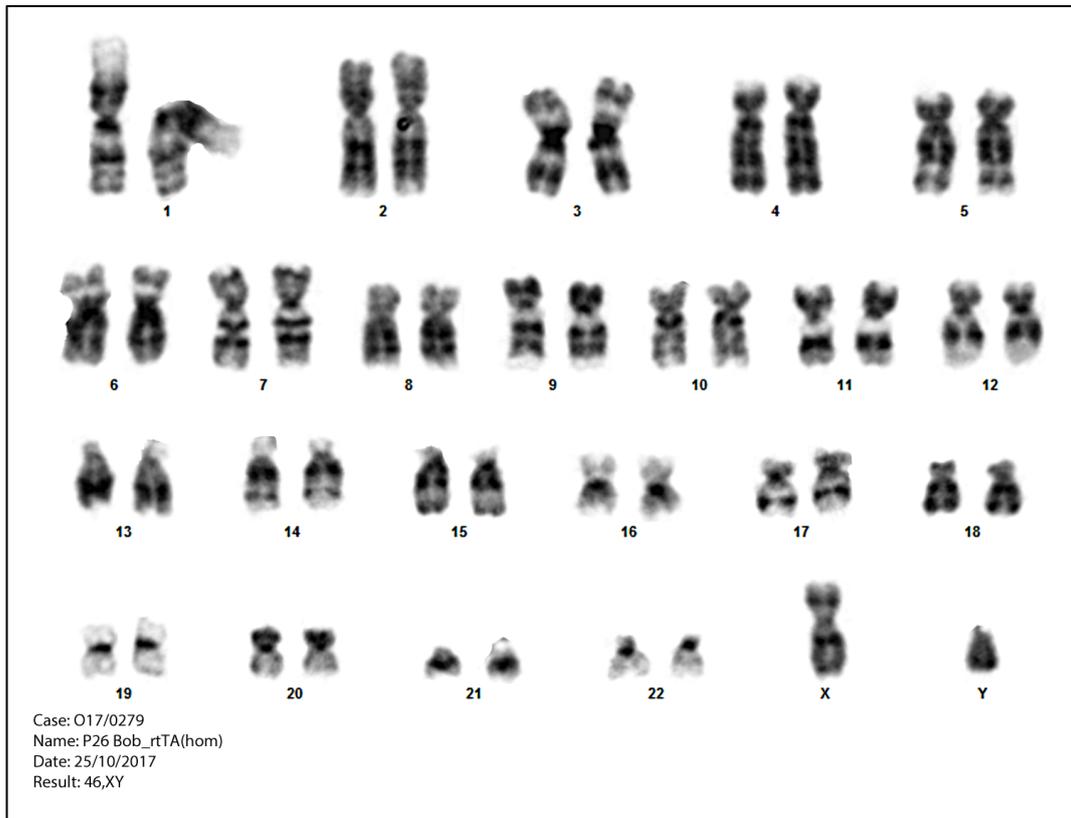


Figure 3. 5: Karyotype report for the Bob HA-NGN2 OPTi-OX hiPSC line.

The report shows a normal male karyotype (46-XY) of the Bob HA-NGN2 OPTi-OX hiPSC line at passage 26, following OPTi-OX targeting.

3.3.2 Establishment of an optimised protocol for the long-term culture of iNeurons.

The existing protocol described in Pawlowski et al. (2017) results in dense clusters of neuronal bodies as early one week after induction and eventually in entire cultures detaching off the surface after two to three weeks of culture.

I hypothesised that optimising the coating matrix may provide a means of overcoming this problem. I tested four substrates commonly used for supporting human- and rodent-derived neurons: Matrigel (Corning), Geltrex (Gibco), poly-D-lysine plus laminin and polyethyleneimine plus laminin. However, all four coatings also resulted in detachment of the cultures. Since most reports of NGN2 iNs reported the use of primary rodent glia (mainly astrocytes) to aid neuronal maturation and function (Busskamp et al., 2014; Lam et al., 2017; Yingsha Zhang et al., 2013), I chose to compare iNeurons cultured in the presence and absence of rat glia and determine if it reduced neuronal clustering and if it could facilitate its long-term culture. The protocol I adopted for the co-cultures was based on a synthesis of methods used in three previous reports (Busskamp et al., 2014; Lam et al., 2017; Zhang et al., 2013) and is described in the Chapter 2 of this thesis.

I compared iNeurons cultured in the presence or absence of rat glia, cultured on PDL-laminin coated culture plates, at three different time points: day 7, 14 and 21 post-induction (**Fig. 3.6**). When co-cultured with glia, iNs showed no signs of significant clustering after two weeks of culture, compared to monocultures of iNs. After three weeks, co-cultured iNeurons showed signs of clustering compared to the two-week mark. However, this was noticeably less compared to cells cultured in the absence of glia. In addition, many of the monocultures peeled-off from the culture surface; whereas, all co-cultures remained attached (data not shown). Therefore, the presence of rat glia seemed to prevent clustering of neuronal cell bodies and promotes surface attachment for long-term culture.

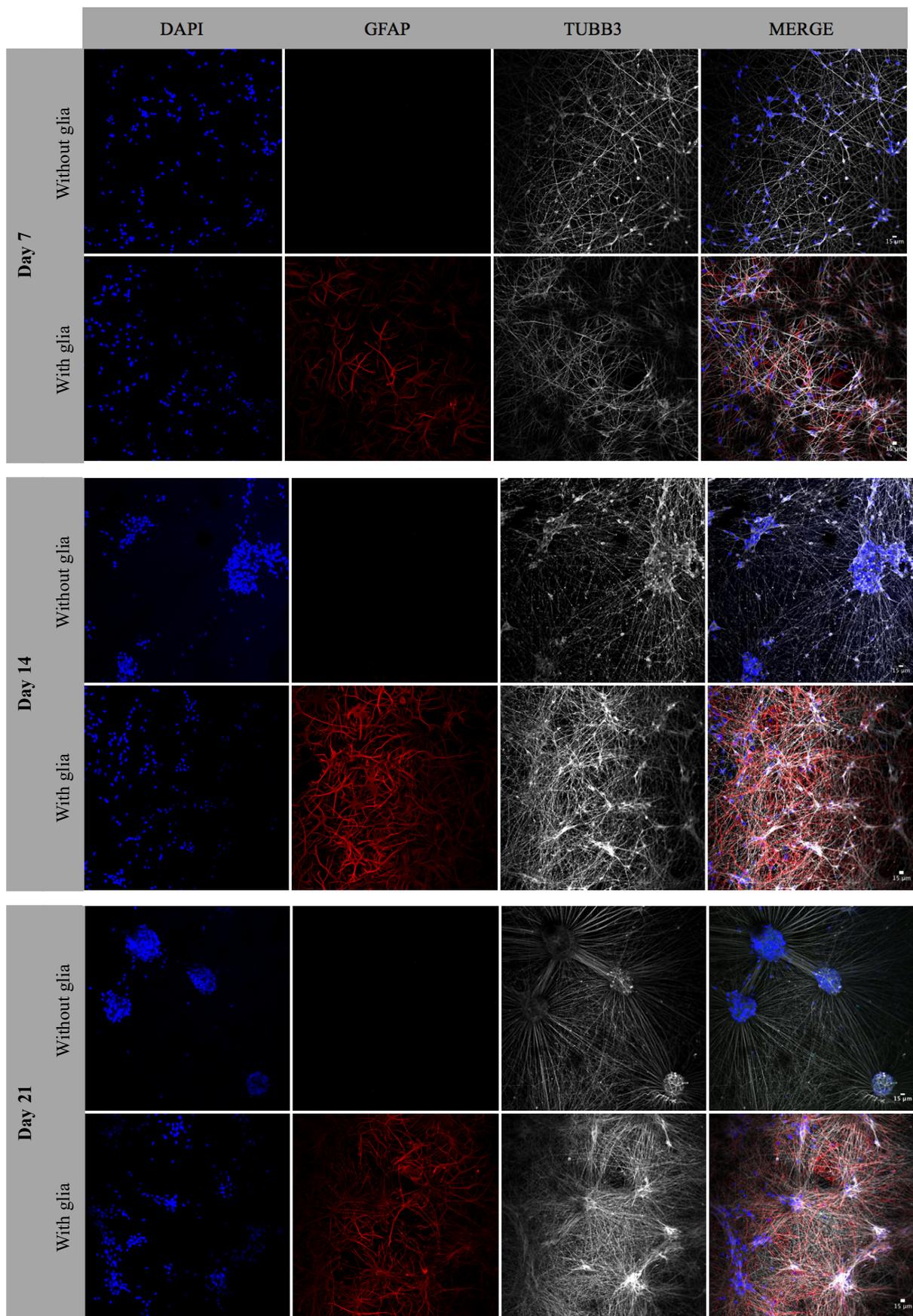


Figure 3. 6: Comparison of iNs cultured in the presence and absence of rat primary glia.

Immunocytochemistry of iNs cultured in the absence (top panel) and presence (bottom panel) of rat-derived glia (mainly astrocytes), at Day 7, Day 14 and Day 21 post-induction. Cells were stained for the astrocyte marker, GFAP (red) and the pan-neuronal marker, TUBB3 (gray). Nuclei were visualised with DAPI (blue). Images show decreased neuronal clumping in the presence of rat astrocytes (Scale bars: 15 μ m).

To investigate the mature and functional properties of iNs over time, it was important to develop a protocol that would facilitate long-term culture. Previous reports indicated that the use of mouse glia (Yingsha Zhang et al., 2013) or rat (Frega et al., 2017; Lam et al., 2017) enriched for astrocytes may facilitate long-term culture of iNs. For my project, I chose to use rat glia because these are readily available in our lab. To determine whether our rat glia (astrocyte enriched) preparations are also capable of facilitating long-term culture of human NGN2 OPTi-OX iNs, cells were maintained in co-culture up to 75 days post-induction (**Fig. 3.7**). The cultures remained stable over the observed time-period and rat astrocytes prevented dislocation from the culture dish. To further characterise the cultures, they were fixed and stained at day 50 and day 75 post-induction. At both 50- and 75-days post-induction, hiNeurons appeared healthy and had formed complex networks with their processes. GFAP staining demonstrated a persisting presence of astrocytes, which seemed to be dispersed throughout the cultures but predominantly located next to the perikarya of neurons. Synapsin-1 puncta overlapped with MAP2 immunofluorescence, indicating the presence of functional synapses in the cultures.

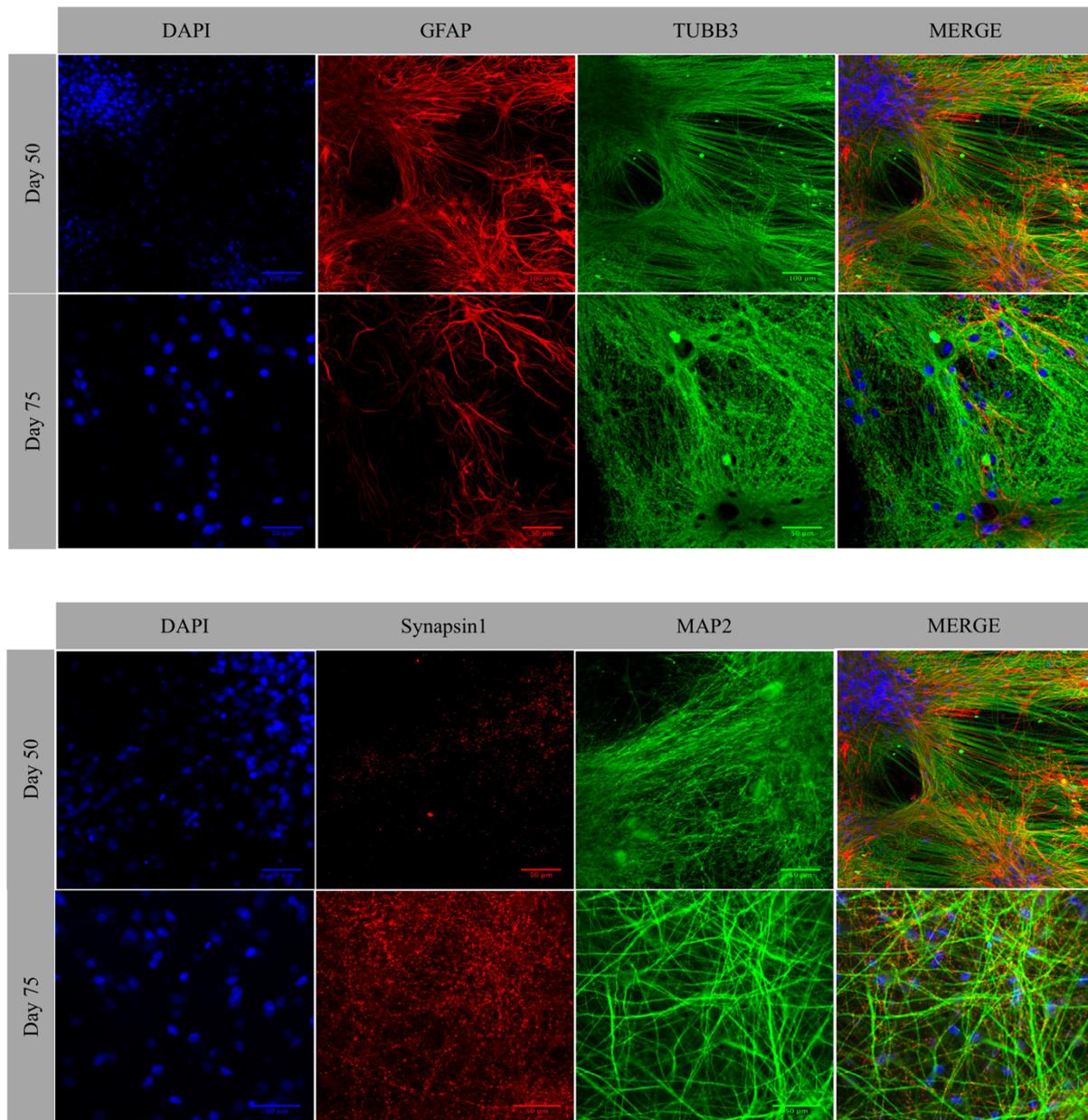


Figure 3. 7: ICC of long-term cultures of iNeurons co-cultured with rat astrocytes

At day 50 and 75 (A), co-cultures were fixed and stained for glial fibrillary acidic protein (GFAP; red), the hallmark intermediate filament protein in astrocytes (Hol & Pekny, 2015) and the pan-neuronal marker, beta-III tubulin (TUBB3; green), a major constituent of microtubules in neurons (Cicchillitti et al., 2008). At the same time points (B), cells were fixed and stained for Synapsin-1 (red), a pre-synaptic marker that is indicative of histological presence of synapses, and the pan-neuronal marker microtubule-associated protein 2 (MAP2; green), a protein that serves to cross-link microtubules with intermediate filaments and other microtubules and has been implicated in the formation of dendritic spines (Shirao & González-Billault, 2013). Nuclei were visualised with DAPI (blue) (Scale bars: A, 100 μ m; B, 50 μ m).

3.3.3 Electrophysiological properties of NGN2 iNs on MEAs

Having established a culture protocol that reduced neuronal clustering and allowed long-term culture, I next sought to study the electrical activity of iNs. Previous reports indicated that NGN2-derived iNs display electrophysiological activity within two weeks of culture. Specifically, patch clamping established that spontaneous excitatory activity appears as early as 7 days post-induction (Frega et al., 2017; Lam et al., 2017; Yingsha Zhang et al., 2013). MEA recordings after three weeks demonstrated that spontaneous activity by hiN and synchronous burst firing (Frega et al., 2017). To examine whether OPTi-NGN2 iNs display similar patterns of activity, several modifications had to be made to the culture protocol described in the previous chapter.

As a first step, the right seeding density had to be established as several studies have shown that this can strongly influence the functional properties of developing neuronal networks (Biffi et al., 2013). Cortical neurons seeded at a high density are known to mature faster and record faster firing rates (Wagenaar, Pine, & Potter, 2006). Published protocols using rat (Wagenaar et al., 2006) and human iPS-derived cortical neurons (Heikkilä et al., 2009) have plated between 30,000 and 50,000 cells within the active electrode area of a 64 electrode MEA. In addition, the volume of the cell suspension has to be taken into account, as a higher volume will cover a larger surface area. The objective was to keep neurons within the field of the electrodes. A too large volume therefore could lead to cells outside the area covered by the array and affect the network activity recorded. Taken together, these effects could introduce variability between cultures.

Another important factor with regards to optimising the culture conditions for MEA recordings from iNeurons was to decide whether to plate PSCs onto the MEAs and then initiate induction or to plate cells when they already were induced to neurons. To determine the optimal time point for seeding, I analysed time-lapse series of iPSCs as they convert to iN (**Fig. 3.8**). These demonstrated that cells continue to divide during the first 24 hours of induction. As a consequence, controlling the final number of neurons may pose difficulties. Moreover, iPSCs tend to form clusters as they grow in colonies, and unless iN migrate out and form an even distribution this may compromise future MEA recordings. On the other hand, cells at day 3 appeared to have reached a neuron lineage on the basis of their morphology but yet have to fully extend axonal and dendritic processes. The notion that cells are committed to a neuronal lineage is corroborated by the finding that discontinuing doxycycline at day 3 has minimal effects on subsequent iN cultures (Pawlowski et al., 2017). Therefore, cells were dissociated and plated onto MEA after three days of induction.

To promote maturation and maintain iNeurons in culture over prolonged periods, these had to be cultured in the presence rat glia (astrocyte-enriched). I adopted this approach for MEA recordings using 64-electrode MEAs (Multichannel Systems). A recent report showed that a 1:1 ratio of neurons to glia yields an optimal ratio for recording network activity (Frega et al., 2017). Therefore, at three days post-induction, iNs were dissociated and mixed with rat glia at this ratio and seeded onto the active electrode area of each MEA. The seeding volume and density was adjusted to give a total of 30,000 iNs over the recording area. Cells were allowed to attach for one hour in an incubator at 37°C and 5% CO₂, before topping up with culture medium.

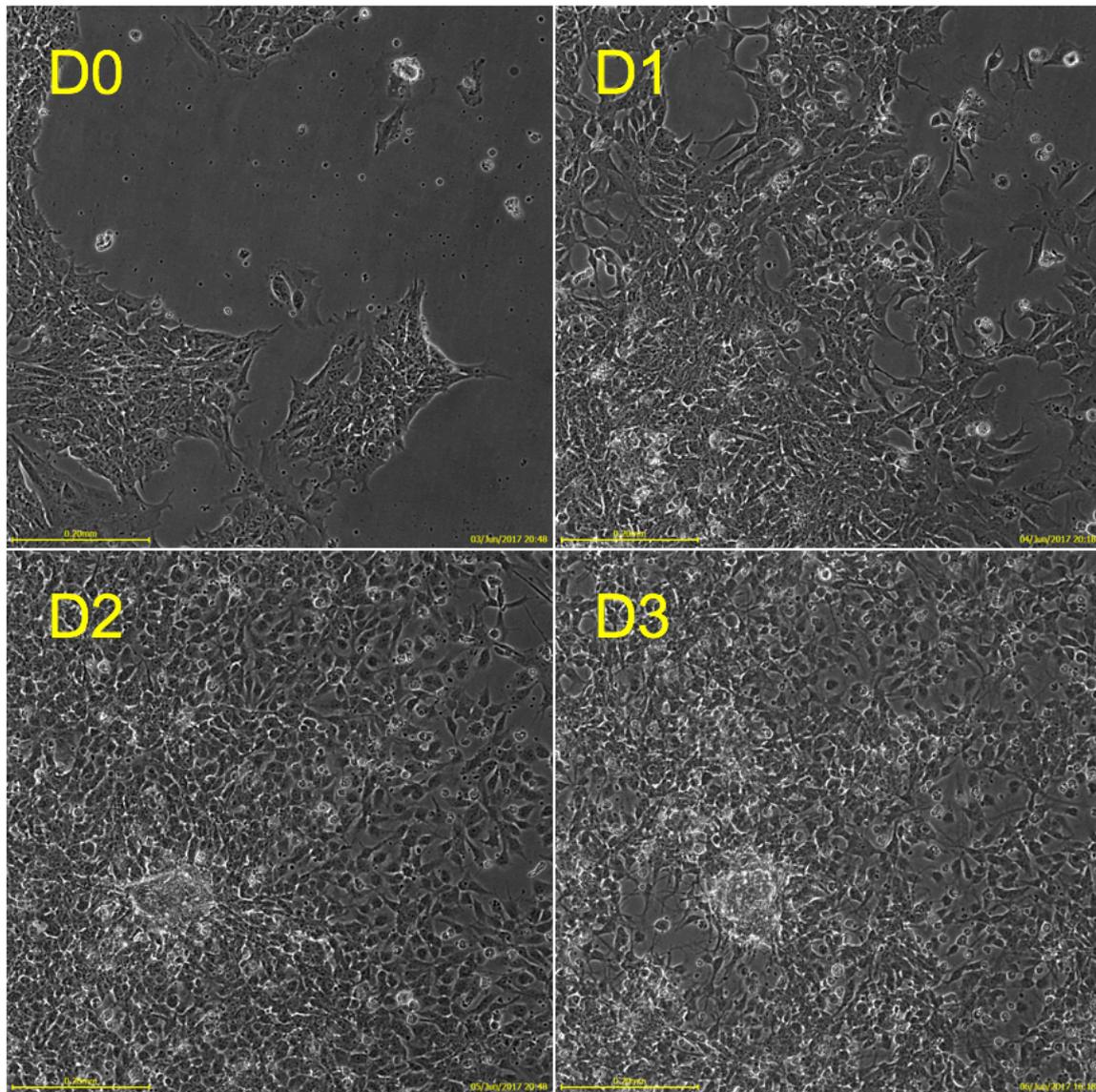


Figure 3. 8: Phase contrast time-lapse images of the first 3 days of iNeuron reprogramming. At the start of an induction (D0), OPTi-NGN2 hiPSCs had formed a colony of cells. 24 hours after induction (D1), colony integrity was disrupted as cells appeared to have dispersed. It is likely that cells continue to divide up to this point, due to the apparent increase in cells. At 48 hours (D2), majority of the cells have acquired a dense, regular-shaped, soma-like cell body. After 72 hours (D3), cells appear more neuronal-like, with occasional outgrowth of short neuronal processes.

Having arrived at an optimised culture system for MEA studies, I proceeded to record from the cultures. The electrical activity of cultured iNs was recorded using an MEA2100-System (Multichannel Systems, Reutlingen, Germany) with an integrated amplifier. Each MEA dish (60MEA100/10iR-Ti; Multichannel Systems) contained 64 electrodes (TiN, 30 μ m diameter) arranged over an 8 \times 8 square grid. Recordings started 10min after the MEA plates were placed on the head stage, which was set to 37 $^{\circ}$ C. All MEA recordings were performed in culture medium and each recording lasted 10 minutes. The electric signals were collected at 10 kHz using MCRack (Version 4.4.2; Multichannel Systems) and analysed offline. At the time point of the present report, I generated MEA recordings of hiN/rat astrocyte co-cultures starting at day 7 until day 27 post-induction. Based on these recordings, the first instance of spontaneous activity appeared on day 11 post-induction and was detected across all electrodes (**Fig. 3.9**). On day 16 the response to tetrodotoxin (TTX), a voltage-gated sodium channel blocker was tested in one of the cultures. Prior to treatment, a routine recording was performed. Then, a full media change was carried out with regular iN media containing 50 μ M of TTX. Activity was recorded after a 10 minutes incubation. The culture was then washed with fresh medium and allowed to settle down in the incubator for 20 min before carrying out post-treatment recording. Raster plots of threshold-based activity showed spontaneous activity in several of the electrodes prior to treatment, which was blocked during treatment (**Fig. 3.10**). Upon wash-out, activity was restored but with a noticeable increase in the number of electrodes with spontaneous firing.

At day 20, the morphology of spontaneous firing changed from random spiking activity to synchronised bursting patterns (**Fig. 3.11**). Bursts of varying amplitudes were recorded across several electrodes, with no detectable activity in the remaining electrodes. Based on the spatial nature of the synchronised bursts, they appear to be spread across the MEA grid, suggesting that the functional networks formed between the neurons are not localised to cells close to another, but between neurons spread across the culture. To confirm that synchronous bursts are dependent on excitatory glutamatergic signalling we subjected our cultures to treatment with 40 μ M cyanquixaline (CNQX) following the developmental time course and used a custom-built synchronous burst detection algorithm to analyse bursts from raw electrode data. CNQX is a specific, competitive inhibitor of excitatory AMPA/kainite receptors (Blake, Brown, & Collingridge, 1988). 10-minute incubation of cultures in the presence of this drug resulted in total inhibition of synchronous bursts without affecting spontaneous firing of action potentials (**Fig. 3.12**). Conversely, administration of 40 μ M bicuculline, a competitive antagonist of the primary inhibitory GABA receptor, had no detectable effect on synchronous bursts (not shown). After day 20, our iN cultures showed an increase in neuronal activity, with increasing electrodes showing

synchronous bursting activity (**Fig. 3.13**). Within a week, the bursting activity evolved dramatically from clear bursting phenotypes in several electrodes to more complex, MEA-wide spontaneous activity.

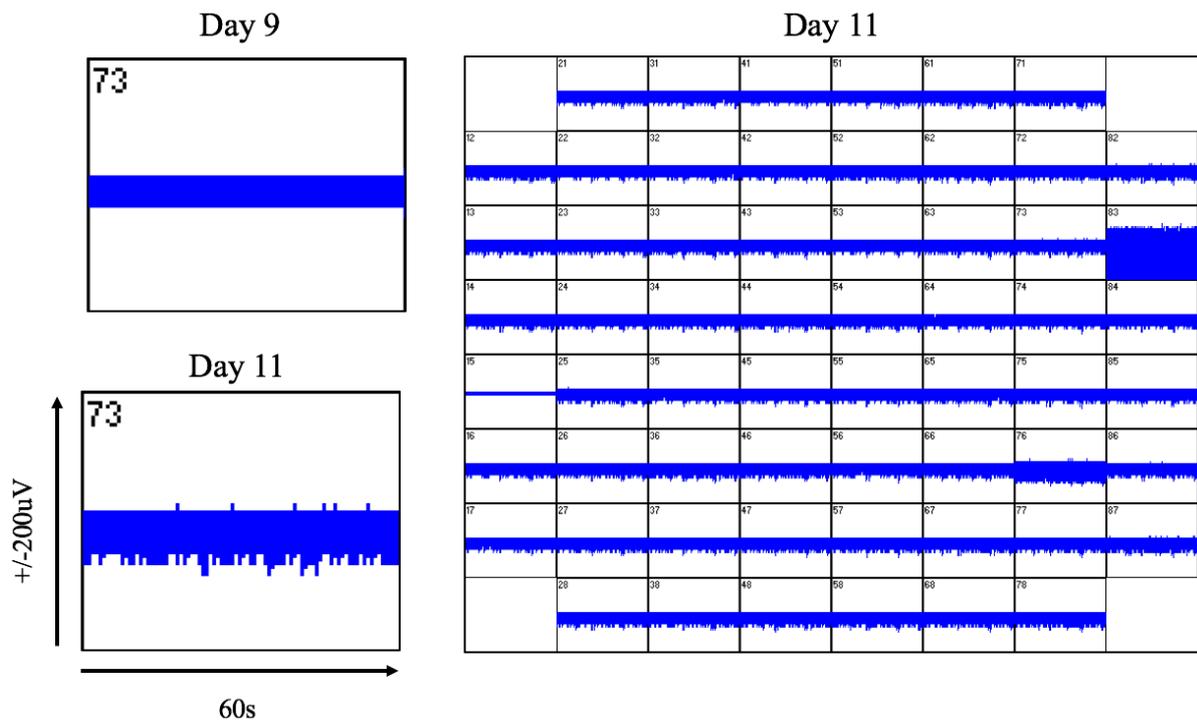


Figure 3. 9: NGN2 iNs show spontaneous activity from day 11 onwards.

In this snapshot of a recording on Day 11 (right), each box represents recording from one electrode during a 60s period. Here, recording for electrode 73 is enlarged (bottom left) and compared to a recording taken for the same electrode on Day 9 (top left). It shows that spontaneous firing above the noise level first appears on Day 11, where majority of the electrodes showed activity.

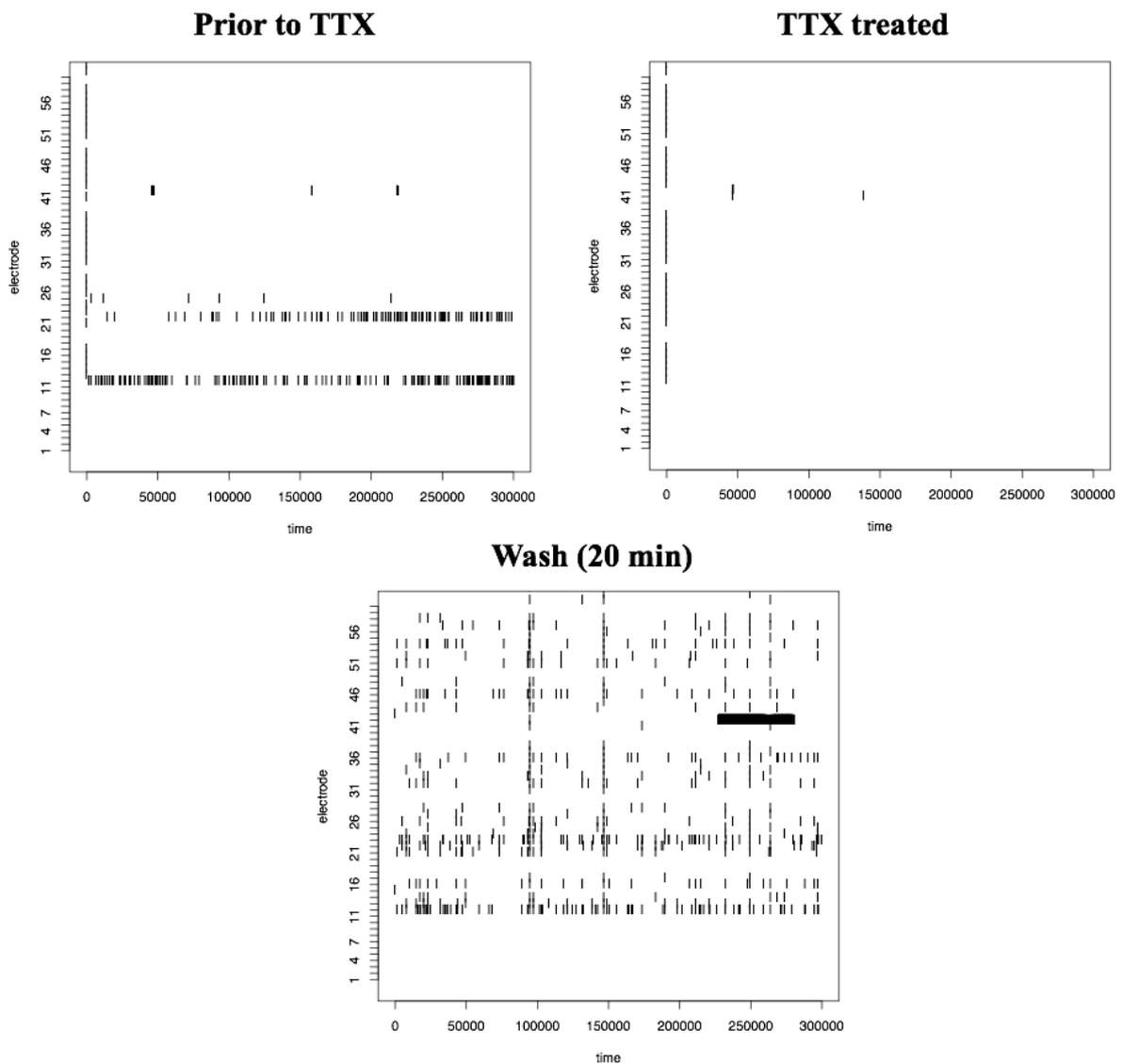


Figure 3. 10: Voltage-gated sodium channels mediate electrical activity in NGN2 iNs

Threshold-based spike detection (3 standard deviations of the noise level) was used to identify and collect timestamps for each electrode. Spike raster plots for each electrode against time (ms), show spontaneous firing in several of the electrodes at 16 days' post-induction (top left). Each thin black line is representative of a spike/action potential. This activity was blocked during TTX treatment (top right). 20 minutes upon wash-out (bottom), spontaneous firing was detected in most electrodes in the culture.

Day 20

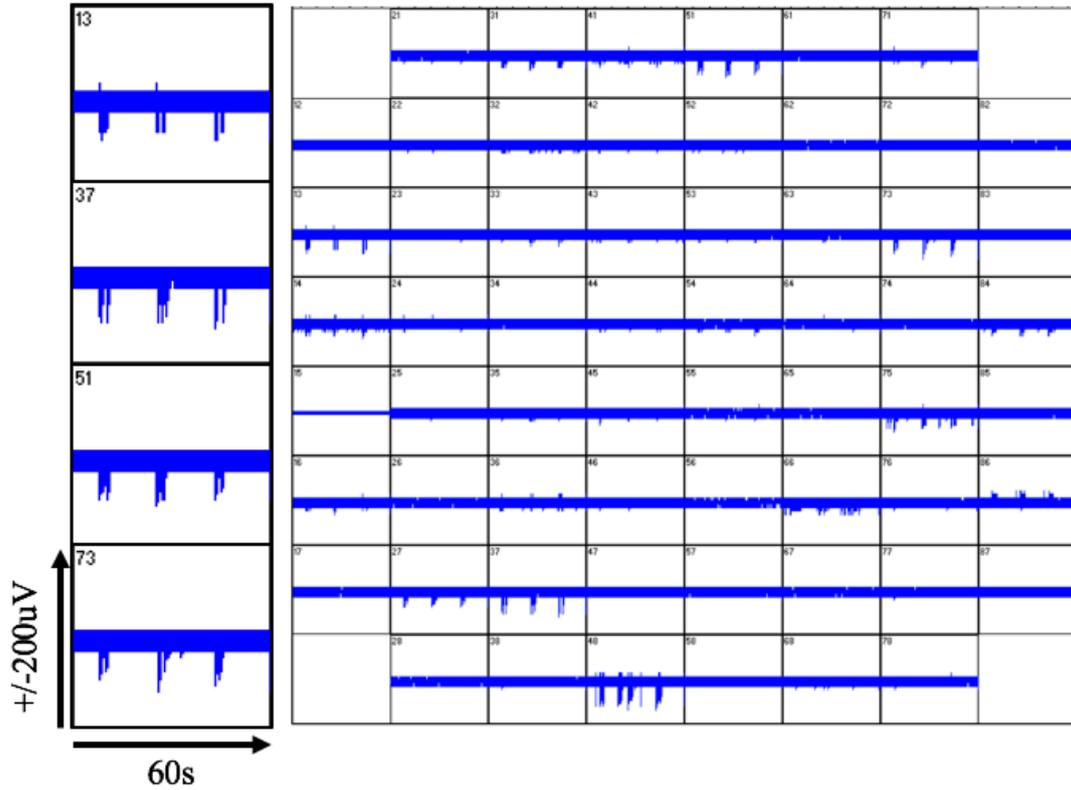
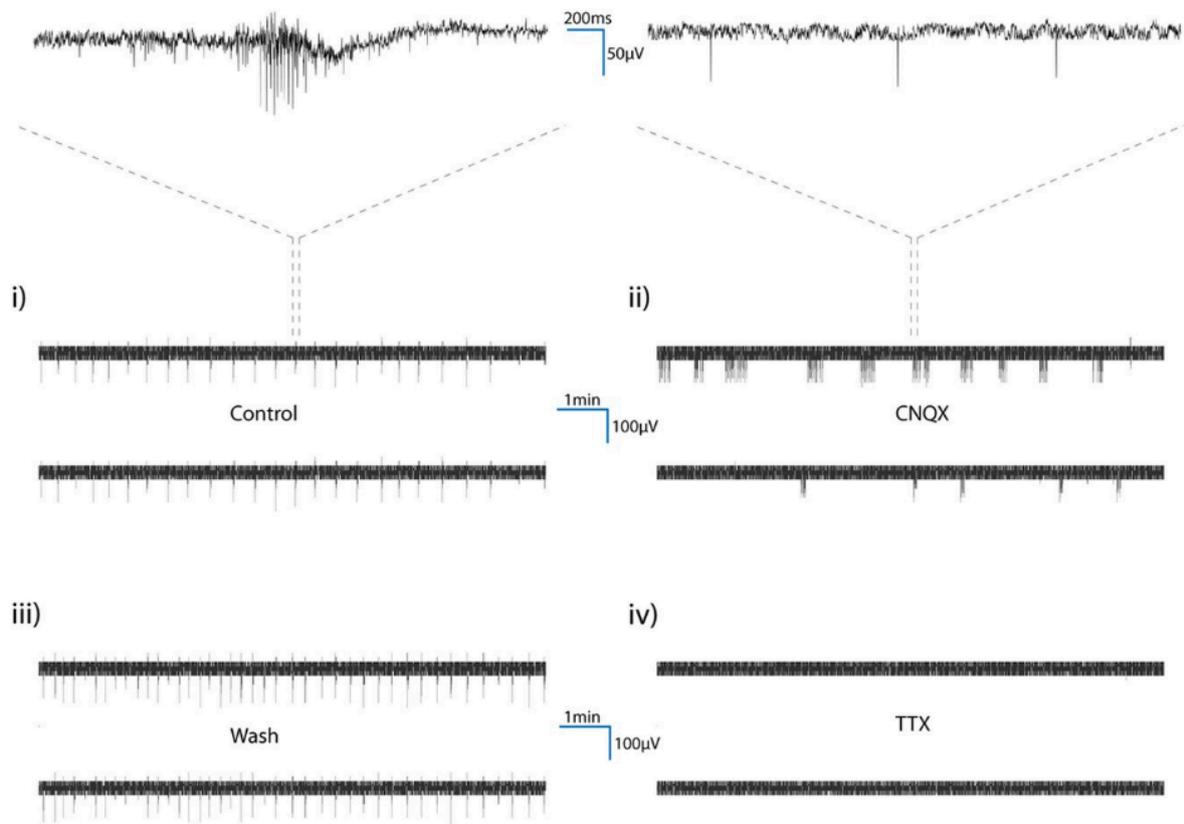


Figure 3. 11: Synchronous burst firing emerge at day 20 post-induction

In this snapshot of spontaneous activity recorded on day 20 post-induction (right), burst firing was recorded in multiple electrodes. Recordings from four of these electrodes (electrodes 13, 37, 51 and 73) are enlarged for visual comparison.

(A)



(B)

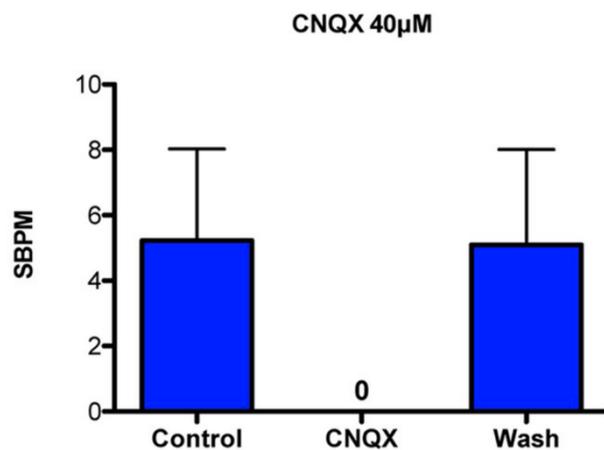


Figure 3.12: Characteristics of synchronous burst firing

(A) Synchronized bursts consisting of trains of action potentials are clearly visible in raw electrode data. Representative data from two neighbouring electrodes showing temporal correlation of bursts occurring in control conditions (i). Synchronized bursts are obliterated (although spontaneous action potentials persist) by the presence of the AMPA/kainite receptor

antagonist CNQX (40 μ M) (ii), but this effect is immediately reversible following a wash-off (iii). There is total inhibition of electrical activity upon treatment with TTX (iv). For illustrative purposes, upper panels displaying fast time scale are smoothed using a 2ms Gaussian window. (B) Zero synchronous bursts per minute (SBPM) were observed in the presence 40 μ M CNQX treatment, further indicating the dependence of synchronous bursting on excitatory glutamatergic signalling.

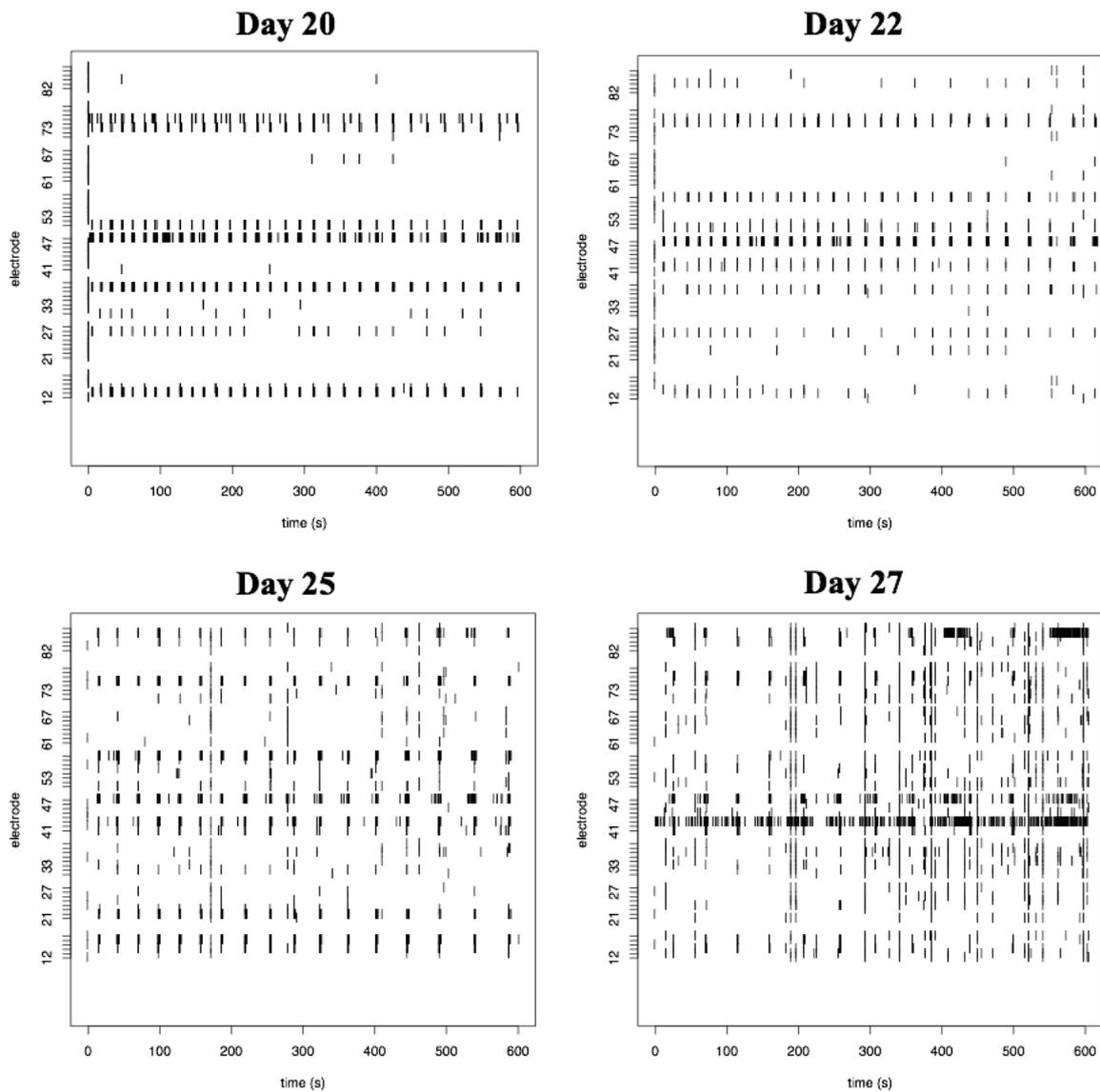


Figure 3. 13: Development of complex network activity from day 20 to day 27

Spike raster plots for each electrode against time (s) for days 20, 22, 25 and 27 post-induction. Each thin black line is representative of a spike/action potential, while bursts of electrical activity represented by thicker blocks. Based on this, there is an overall increase in neuronal activity from day 20 to day 27, with different, more complex bursting morphologies emerging at days 25 and 27.

3.4 Discussion

The experiments outlined in this section firstly demonstrate the successful generation of an HA-tagged NGN2 OPTi-OX hiPSC line with a low passage and intact karyotype for the proposed genomic time course analysis of NGN2 reprogramming. In addition, I have developed an optimised protocol for their long-term *in vitro* culture. Use of other coating substrates to overcome issues with cell clumping and culture attachment were not successful; instead, co-culturing iNs with rat glia proved to be the only method for culturing mature and functional neurons for extended periods.

Previous reports indicated that the presence of mouse astrocyte-enriched glia facilitates the long-term survival and maturation of both primary and stem cell derived-human neurons (Frega et al., 2017; Yingsha Zhang et al., 2013). The work presented so far validated the concept that the presence of rat astrocytes is also able to maintain NGN2 iNs for prolonged culture periods, to prevent cells from clustering and from peeling off of the surface. The fact that rat astrocytes have comparable effects to the previously used mouse-derived cells alludes to the fundamentally important role of astrocytes across species. In a separate study, clustering of neurons derived through differentiation of human iPSCs was overcome by culturing cells on micropatterned substrates (Burbulla, Beaumont, Mrksich, & Krainc, 2016). Therefore, astrocytes could simply contribute by providing “micropatterns” and thus act as physical support for the neurons. Alternative ways of mediating conducive effects of neurons include the paracrine secretion of relevant neurotrophic factors such as neurotrophin-3 (NT-3) and brain-derived neurotrophic factor (BDNF) (Verkhatsky, Matteoli, Parpura, Mothet, & Zorec, 2016), or direct cell-to-cell contact interactions, as been described in the context of the tri-partate synapse, in which astrocytes are known to play an important role (Amaral, Meisingset, Kotter, & Sonnewald, 2013; Eroglu & Barres, 2010).

Using the protocol, I demonstrated that cultures can be maintained up to 75 days *in vitro*. As they remained healthy throughout the entire culturing period, it is to be expected that the time frame can easily be extended. ICC analysis at day 50 and 75 demonstrated that astrocytes were in abundance and in close contact with the iNeurons. Close physical contact is known to play an important role for synaptic maturation and function (Odawara, Saitoh, Alhebshi, Gotoh, & Suzuki, 2014). This was corroborated in hiNeuron cultures by staining of the pre-synaptic marker Synapsin-1 at day 50 and 75, which showed numerous punctate pre-synapses distributed along neuronal processes across the culture. To the best of our knowledge, no other publication has demonstrated such long-term culture for NGN2 neurons, but punctate expression has been shown

to occur at earlier timepoints (Fernandopulle et al., 2018; Frega et al., 2017). Punctate patterns are consistent with the formation of mature synaptic boutons, evolving from diffuse axonal staining patterns in immature neurons (Fletcher, Cameron, De Camilli, & Banker, 1991). Expression of Synapsin-1 appears higher at day 75 in these samples, suggesting increased synaptic connectivity over time.

This protocol was then successfully adapted for MEA cultures, which requires a dense distribution of firmly attached neurons over its electrodes for effective recording. Using this protocol, the first sign of activity was observed as early as day 11 post-induction, in support of other reports which also detected activity before two weeks *in vitro* through patch-clamping (Lam et al., 2017; Yingsha Zhang et al., 2013). This activity was confirmed to be of neuronal origin, demonstrated by treatment with TTX. TTX dampens sodium channels that are exclusively expressed in neurons (Bane, Lehane, Dikshit, O’Riordan, & Furey, 2014). Moreover, contaminations by rat-derived neurons is highly unlikely: astrocytes were generated from primary mixed glial cultures via a process that is known to break axons and as a consequence induce cell death in neurons (Cole & de Vellis, 2001). In addition, glial preparations were cultured for a minimum of one passage before co-cultures, which further eliminates potential contaminating neurons, as they are highly susceptible to cell death during passaging (Lam et al., 2017). A single-cell sequencing analysis of the glial cultures, outlined in the next chapter, further corroborates this finding.

Synchronized spike bursts are critical for information transfer within the cortex and are indicators of functional maturation (Odawara, Katoh, Matsuda, & Suzuki, 2016). Previous studies with MEA analysis of human NGN2 iNs have also reported the emergence of synchronous bursting at around three weeks post-induction (Frega et al., 2017; Nehme et al., 2018). Furthermore, this activity was shown to be entirely mediated by AMPA-receptors, with no contribution from GABA receptor-mediated synaptic transmission, consistent with previous reports (Frega et al., 2017; Lam et al., 2017; Nehme et al., 2018; Yingsha Zhang et al., 2013). The network activity displayed by our iNs between day 20 and 27 is comparable to NGN2 iNs derived through a combination of NGN2 overexpression and small molecule neuronal patterning, a protocol which was reported to increase neuralising efficiency and production of mature functional neurons that express transcriptional states found in superficial levels of the cortex (Nehme et al., 2018). The evolution of largely bursting activity into a more heterogenous, complex activity also draws a similar comparison to network patterns seen with differentiated human cortical neurons between days 58 and 77 *in vitro* (Kirwan et al., 2016). Given that our cultures are devoid of GABAergic neurons, it is intriguing how a completely excitatory network can exhibit attenuation of spontaneous activity. Whether this attenuation is driven by cell-intrinsic mechanisms or modulated by the supporting glial cells,

will be an interesting subject of future studies. Nonetheless, these findings show that our protocol produces neurons with biologically relevant electrophysiological features seen in both reprogramming protocols and conventional human PSC-derived cortical differentiation.

The present characterisation of our NGN2 iNs has built substantially on our previous study. Together they highlight several important hallmarks in their development. The first four days of reprogramming is characterised by rapid proliferation and change in morphology that begin within hours of induction, producing neuronal-like cells by day four. The cells continue to differentiate, becoming electrophysiologically active near the two-week mark when co-cultured with rat glia. By three-weeks, the iNs have begun establishing functional synaptic connections with one another which result in synchronous burst firing. These distinct stages in NGN2 iN development serve as suitable timepoints for a time course genomic analysis of NGN2 reprogramming enabling the capture of the two main stages of NGN2 iN development: an early ‘reprogramming’ stage, Days 0 to 4 post-NGN2 induction, that encompasses the NGN2-driven reprogramming of PSCs into neurons; and a late ‘maturation’ stage, Days 14 and 21 post-NGN2 induction, that identifies two separate important electrophysiological events.

Investigating the ‘maturation’ stage would also provide the opportunity to study how the addition of rat-glia (primarily astrocytes) modulates the transcriptome and epigenome, which results in the functional differences seen at the two late stage time points. MEA data of iNs cultured without glia were not available for this thesis mainly due to the poor attachment of iN monocultures to an MEA surface, making it problematic to record any activity. However, numerous studies have reported that *in vitro* cultures of neurons, particularly neurons differentiated *in vitro*, show little, or no activity in the absence of glia. In fact, Nehme and colleagues’ who successfully recorded activity from their NGN2 iN monocultures showed that although there may be some activity in the absence of glia, it is generally delayed, scarce and lower in intensity compared to their NGN2 iN-glia co-cultures (Nehme et al., 2018). Additionally, synchronous burst firing is absent even after 42 DIV. Therefore, comparing iNs cultured with glia to iNs cultured without at the two late stage time points would very likely provide novel insights into the mechanisms by which glia mediate the electrophysiological maturation of neurons.

4 Transcriptional profile of NGN2 reprogramming

4.1 Introduction

The rapid transition of proliferative pluripotent stem cells into functional post-mitotic neurons induced by OPTI-OX-mediated NGN2 reprogramming, is likely governed by an equally remarkable change in their transcriptome. However, the mechanisms and specifically the gene regulatory networks by which NGN2 induces cell fate changes in iPSCs remain poorly understood. As discussed in Chapter 1, a number of studies have reported transcriptional states induced by NGN2; yet the data remain limited.

Cell type transition likely involves not only primary interactions of the respective reprogramming factors with their transcriptional network, but also second and higher order interactions (Dunn, Martello, Yordanov, Emmott, & Smith, 2014; Guo, Morris, Wang, & Esteban, 2017). As a consequence, it is insufficient to limit the study to only one time point. An ideal study design would enable to assess transcriptional changes over the entire reprogramming process. For NGN2, it was shown that a minimum of 4 days of dox-induced expression in PSCs is required for efficient reprogramming to neurons (Pawlowski et al., 2017), suggesting that persistent activation of direct targets of NGN2 is necessary up until day 4 to firmly establish a neuronal fate in all cells. As shown in the previous chapter, this also marks the point where most cells have a neuronal-like morphology. In addition, it was also observed that rapid changes occur within hours of reprogramming. Therefore, a time course that captures these early changes would provide valuable information on the early transcriptional networks triggered by NGN2 and subsequent networks that lead to efficient neuronal reprogramming of human iPSCs.

Similar to conventional neuronal differentiation protocols, neuronal reprogramming can also produce a heterogeneous population of cells. The most notable example of this is with *Ascl1* reprogramming of MEFs, which has been shown to generate a sub-population of myocyte-like cells alongside glutamatergic neurons, independently of any transcriptional heterogeneity in the starting population (Vierbuchen et al., 2010, Treutlin et al., 2016). The rapid growth of single cell transcriptomics in recent years has made it possible to easily study heterogeneity in such protocols and understand how they arise. They have not only allowed us to devise strategies for improving efficiencies, but have also provided new insights into the developmental function and mechanisms of reprogramming factors (Schiebinger et al., 2019; Stone et al., 2019; Treutlein et al., 2016). So far, only two studies have employed this method to investigate NGN2 reprogramming; one where *Ngn2*-reprogrammed cells from mouse ESCs were compared to *Ascl1*-reprogrammed cells

(Aydin et al., 2019), and another which investigated the reprogramming of human iPSCs through a combination of NGN2 reprogramming and small molecule patterning (Nehme et al., 2018). However, none have used single cell transcriptomics to solely investigate NGN2 reprogramming of human PSCs.

Therefore, a rich transcriptional account of the reprogramming trajectory of NGN2 in human iPSCs and the terminal population will provide a new type of dataset that, coupled with the ever growing range of bioinformatic tools, will not only offer new insights into the functions of NGN2 as both a proneural and reprogramming factor, but also generalised principles that govern cell type transitions.

4.2 Aims

To uncover the transcriptional states underlying the reprogramming process of NGN2 in our human OPTi-OX model, I carried out bulk RNA sequencing of NGN2 reprogramming using a time course built on data from the previous chapter. Essentially, this time course comprises 10 time points, that make up the early stage of NGN2 reprogramming (Day 0 to 4) and late maturation stage (Day 14 and 21) (Fig 4.1). Given the rapid morphological transformation that occurs within hours of reprogramming, the first 48 hours of this time course sampled cells at 6 and the 12-hour intervals. From this rich and granular dataset, genes significant to the reprogramming process were extracted in two ways: first, by carrying out a differential expression analysis for associations across the time course and for pairwise associations between specific timepoints. Second, by identifying networks of co-expressed genes across the time course using weighted gene co-expression network analysis (WGCNA). Then, these lists of genes were subjected to an enrichment analysis for different collections of literature sets, such as Gene Ontology and Molecular Signature Database (MSigDB). Using this workflow, I investigated the following:

- 1) When do the cells transition from a pluripotent transcriptional network into a neuronal network?
- 2) Do the cells transition through a neuronal progenitor stage?

Then, I supplemented the findings from this bulk dataset with a single cell RNAseq analysis of the same timepoints, except for 6- and 36-hours post-induction, to investigate the same questions aforementioned. In addition, I investigated the heterogeneity of the reprogramming process; in particular, to determine if there are other types of neurons in the terminal population apart from glutamatergic neurons.

Finally, I used both bulk and single cell approaches to investigate the transcriptional changes underlying the functional features seen in NGN2 iNs when co-cultured with rat glia, by comparing it to iNs cultured without glia, at Days 4, 14 and 21 post-induction.

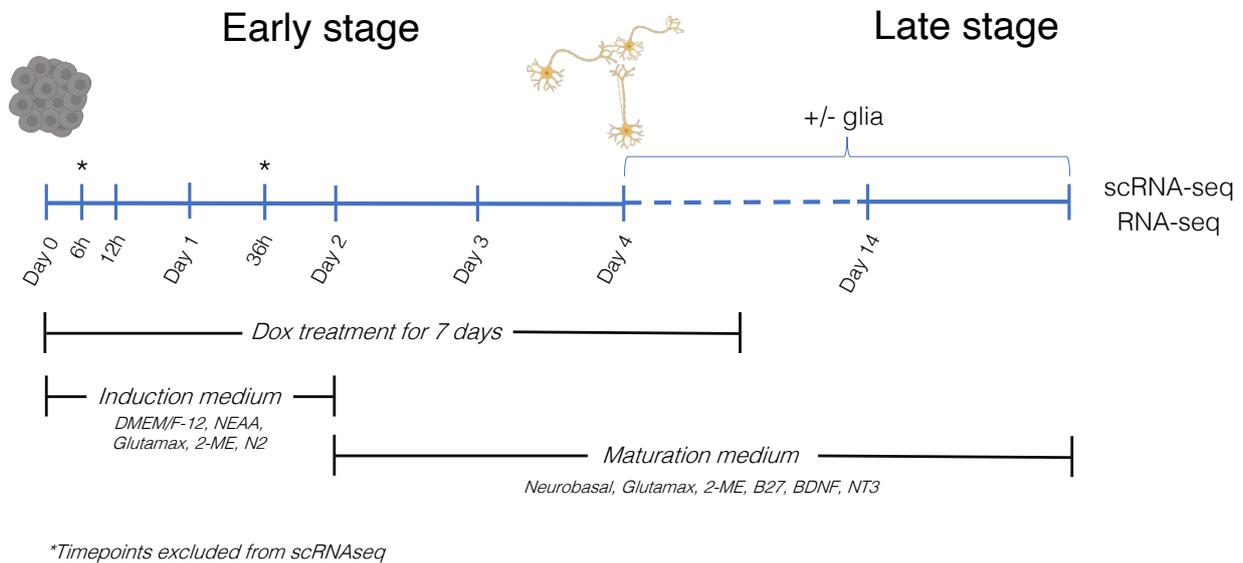


Figure 4. 1: Time course used for bulk RNA-seq and scRNA-seq of NGN2 reprogramming.

For bulk RNA-seq analysis, ten time points were grouped in two stages. For the early (reprogramming) stage, biological triplicates were sampled between Day 0, made up of iPSCs, to Day 4 post-NGN2 induction, at which most cells have formed neuronal-like features. For the late (maturation) stage, cells were sampled at Day 14 and 21. To investigate the effects of rat-derived glia, cells co-cultured with glia were sampled at Day 4 (24 hours after co-culture with glia), Day 14 and Day 21. scRNA-seq was performed on the same time points, except for 6- and 36h post-induction. Cells were kept in induction medium for the first 2 days of reprogramming and then switched to maturation medium for the remainder of the protocol. Dox was added to the culture medium for the first 7 days of induction.

4.3 Results

4.3.1 Bulk RNA-seq data quality assessment and visualisation

Two main steps were carried out in pre-processing the bulk RNAseq data for sample clustering, differential expression analysis and WGCNA. In brief, this involved removal of non- and low-expressed genes, followed by removal of outlier samples. WGCNA works best with data whose variance is (at least approximately) independent of the mean. Therefore, an additional pre-processing step involved normalizing the data using variance stabilising transformation, which was carried out using the R package DESeq2 (Love et al., 2014). This transformed dataset was also used to quantify the sample-to-sample distances for hierarchical clustering of the samples.

A first glance at a heatmap combining these two sets of information, reassuringly reveals two main clusters in the NGN2 reprogramming time course: an early stage comprising Day 0 to Day 4 post-induction, and a late stage comprising Day 14 and Day 21 (**Fig. 4.2A**). The early, reprogramming stage is further clustered into three main groups: the first 12 hours, Day 1 to Day 2 and lastly, Day 3 to Day 4. This initial visualisation of the data provides an overview of the main transcriptional phases that occurred during the reprogramming, especially at the early stage, and will prove useful in guiding and interpreting downstream analysis. Furthermore, it is also worth noting that there are large similarities between the replicates for each time point, indicating the high quality of the data. This is also evident in a principal component analysis (PCA) plot of the time course, where apart for Day 3 and Day 21, replicates within a timepoint show high correlation to each other (**Fig. 4.2B**). In addition, the plot also shows a defined developmental trajectory, especially from the pluripotency stage to Day 4, with an ordering of samples in chronological order. Unsurprisingly, an enrichment analysis for the top 300 genes in each of the first two components revealed neuronal terms, with the first component largely explained by neuronal development and processes associated with action potentials. Doxycycline was added to the medium until Day 7 of the reprogramming; given that Day 14 and Day 21 are negatively correlated with the first component, it is likely that the genes in the first component were mostly highly expressed genes in response to NGN2 overexpression. Interestingly, the second component also showed an enrichment for cholinergic and neuromuscular terms, that are positively correlated with Day 14 and Day 21. This observation suggests the emergence of a motor neuron subpopulation in the later stages of reprogramming and will therefore be further investigated in the scRNAseq dataset analysis.

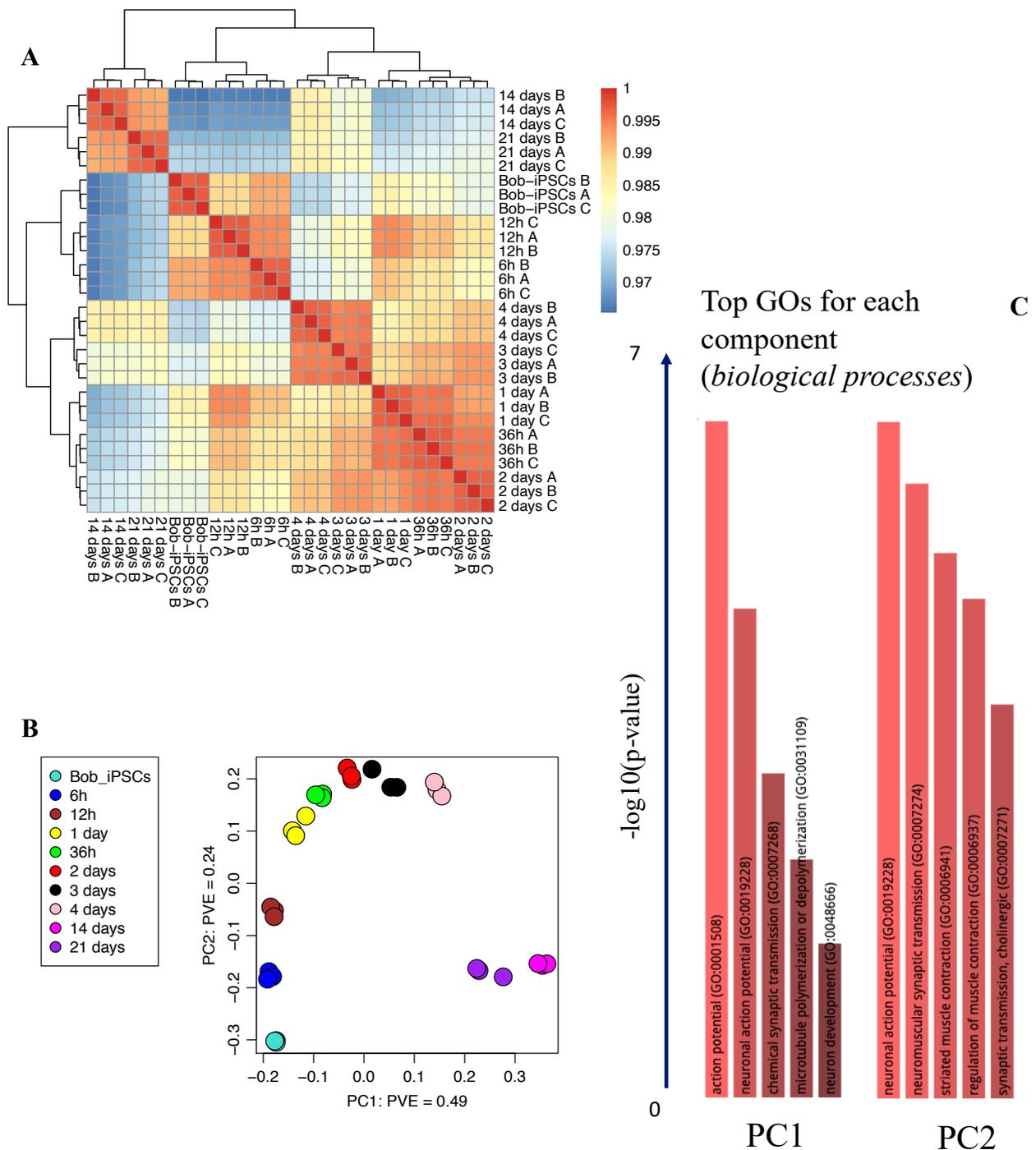


Figure 4. 2: Visualisation of NGN2 reprogramming bulk RNAseq data, processed using variance stabilising transformation.

(A) Heatmap showing the sample-to-sample distances for each timepoint (3 biological replicates for each) from the NGN2 reprogramming time course, grouped using hierarchical clustering. (B) PCA plot of the time course showing the top two principal components and the proportion of variance explained (PVE) for each component. (C) The top 5 gene ontologies (biological processes) for genes identified in PC1 and PC2. PVE: principal variance explained, PC: Principal component.

4.3.2 Differential expression analysis of bulk RNAseq data

Confident with the quality of the pre-processed bulk RNAseq data, I next carried out differential expression (DE) analysis on the bulk RNAseq data using DeSeq2 (Love, Huber, & Anders, 2014), with default arguments except for disabling outlier replacement (since weights have already been applied to downweigh potential outliers) and independent filtering (since genes have already been pre-filtered based on expression levels). The dataset was tested in two ways: 1) for association with **linear time** using a general linear model with time in days as the predictor, and 2) for association with **pairwise time contrasts** (e.g., 6 hours vs. Bob-iPCs, 3 days vs. 2 days, etc.) for a total of 45 different pairwise comparisons. For each test, genes that were significantly upregulated or downregulated below a false discovery rate (FDR) of 0.05, were retrieved.

To gain a deeper understanding of the events occurring at each stage of the reprogramming, I carried out a gene ontology enrichment analysis of significantly DE genes, both up- and down-regulated, for all 46 associations (linear time and 45 pairwise time contrasts). Given the relatively large size of information this generates, many with overlapping enriched terms, I will present here the most important and relevant results from this analysis. As a start, analysis for linear time provides a broad overview of the whole time course, showing enrichment for neuronal terms in upregulated genes such as *nervous system development* and *synaptic signaling* (**Fig. 4.3**). As for downregulated genes, the biological significance of the terms enriched are less obvious – with terms such as *RNA binding*, *nuclear lumen*, and *RNA processing*. Given that the reprogramming takes cells from pluripotency to non-proliferative neurons, it is likely that these terms are associated with maintaining pluripotency or in a cycling state. Indeed, some of the top DE genes listed in these terms are associated with pluripotency such as ZFP42 (REX1) and THOC5.

Next, to gain an idea of the most significant changes occurring after each timepoint, I looked at terms enriched between a timepoint against the one immediately before it (e.g., 6 hours vs Bob iPSCs, 12 hours vs 6 hours and so on). First glance at the downregulated terms for each pairwise comparison reveals largely similar terms to linear time, with broad terms associated with the *nuclear compartment*, *RNA binding* and *intracellular organelle* dominating downregulated terms throughout the time course – but there are a few stages with unique enrichments. Interestingly, these stages also correlate with the start of the main subpopulations identified in Fig. 4.1A. Within the first 6 hours of overexpression, top downregulated terms are associated with transcriptional activity (**Fig. 4.4A**). This then progress to the dominating terms listed above, up until Day 3 vs Day 2 post-induction, which sees an enrichment for cell cycle terms, suggesting this stage of the reprogramming marks cell cycle exit and the point where the neurons become post-mitotic (**Fig.**

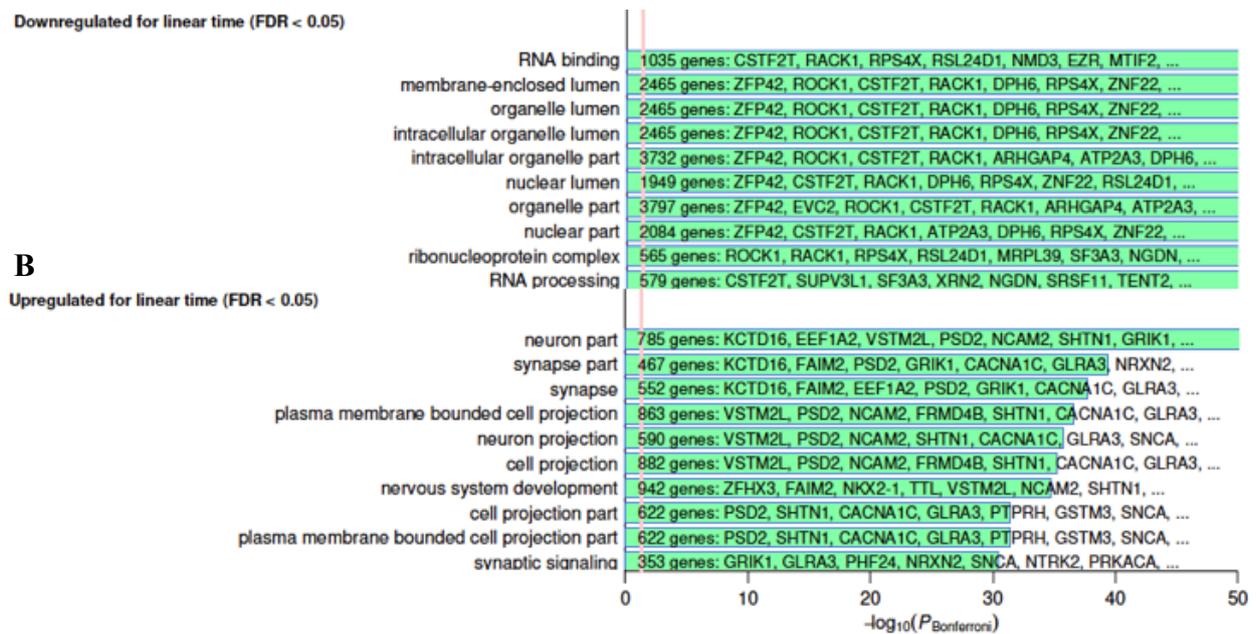


Figure 4. 3: Top gene ontologies enriched for significantly upregulated and downregulated genes across linear time.

Top gene ontologies enriched for significantly (A) downregulated and (B) upregulated genes across linear time. For each set of DE genes, the plot shows the 10 highest-enriched GO terms. Bars represent $-\log_{10}$ of the Bonferroni-corrected enrichment p-value. The red line represents the threshold $P_{\text{Bonferroni}} = 0.05$.

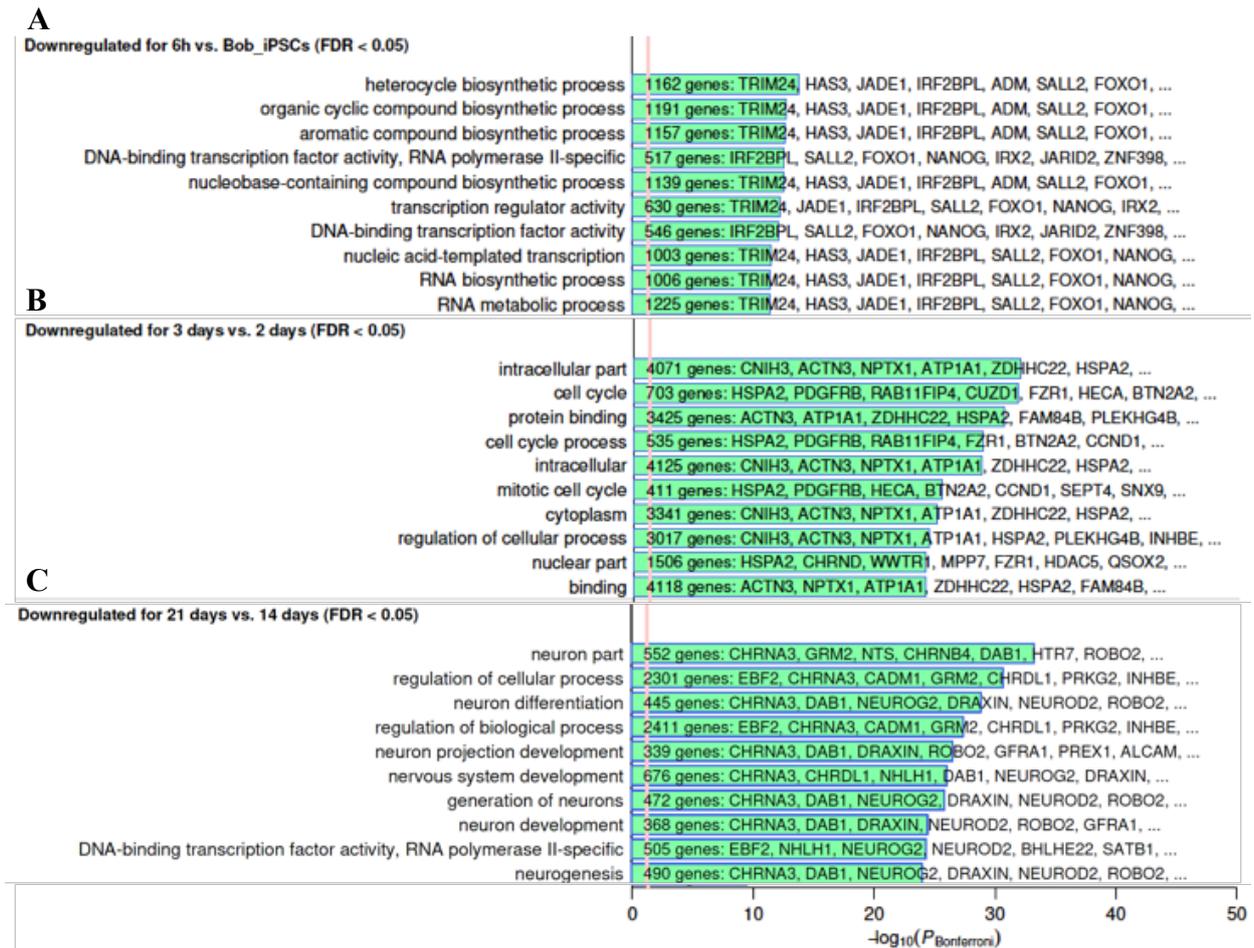


Figure 4. 4: Top gene ontologies enriched for significantly downregulated genes at specific points of NGN2 reprogramming.

Top gene ontologies enriched for significantly downregulated genes between (A) 6 hours post-induction vs iPSC stage, (B) 3 days vs 2 days post-induction and (C) 21 days vs 14 days post-induction. For each set of DE genes, the plot shows the 10 highest-enriched GO terms. Bars represent -log₁₀ of the Bonferroni-corrected enrichment p-value. The red line represents the threshold $P_{\text{Bonferroni}} = 0.05$.

4.4B). From this point onwards, again the same dominant terms are enriched in following pairwise comparisons, up until D21 vs 14, where intriguingly there is a downregulation of neuronal terms such as *synapse*, *neurogenesis* and *nervous system development*, with NGN2 among the top genes associated with these terms (**Fig. 4.4C**). One explanation for this could be that the NGN2 transgene is still being overexpressed up until Day 14. Addition of doxycycline was stopped from Day 7 onwards; however, because half-medium changes were performed from Day 5 onwards, it is likely that the effects of doxycycline were not fully diluted out by Day 14, but by Day 21 instead. Hence, most of the downregulated genes at this point would come from downregulation of the NGN2 transgene and its downstream effectors. A look at NGN2 expression in the single cell dataset later (**Fig. 4.23B**), finds that this was highly the case, as NGN2 was homogeneously expressed across all cells at Day 14, in a pattern that is consistent with expression driven by transgene overexpression.

As for upregulated terms in the pairwise comparisons, a similar pattern is observed, with neuronal development terms dominating most stages of the time course except for the same stages highlighted before. Here, the first 6 hours of reprogramming was enriched for terms associated with protein synthesis and localization to the endoplasmic reticulum (**Fig. 4.5A**). By 12 hours, neuronal ontologies begin to appear in the top 10 terms but not until Day 1 does it dominate the top list of ontologies, suggesting key neuronal genes become accessible and thus highly expressed by the 24-hour mark (**Fig. 4.5B**). This continues until Day 3 vs Day 2, where as before, there is an upregulation for a different set of terms, mainly associated with mitochondria and intracellular transport (**Fig. 4.5C**). Given cell cycle exit likely begins at this stage, these terms highlight a crucial role for mitochondrial function and intracellular transport in neurons becoming post-mitotic. At Day 21 vs Day 14, again we see the likely effect of NGN2 downregulation, interestingly with terms that were downregulated throughout the time course being upregulated in Day 21 neurons (**Fig. 4.5D**).

Collectively, this analysis provided insights into the dominant processes that occurred during NGN2 reprogramming - the shutting down of a pluripotency network and a concomitant activation of neurogenesis. In addition, it highlighted stages of the reprogramming where key events could be occurring such as the increase in neuronal terms at Day 1, cell cycle exit at Day 3 and downregulation of NGN2 transgene at D21 (see genes from **Fig. 4.17**).

4.3.3 Weighted gene co-expression network analysis of bulk RNAseq data

WGCNA identifies clusters (termed ‘modules’) of highly correlated genes (nodes) which are then correlated with categorical or quantitative traits in the experiment (Langfelder & Horvath, 2008b).

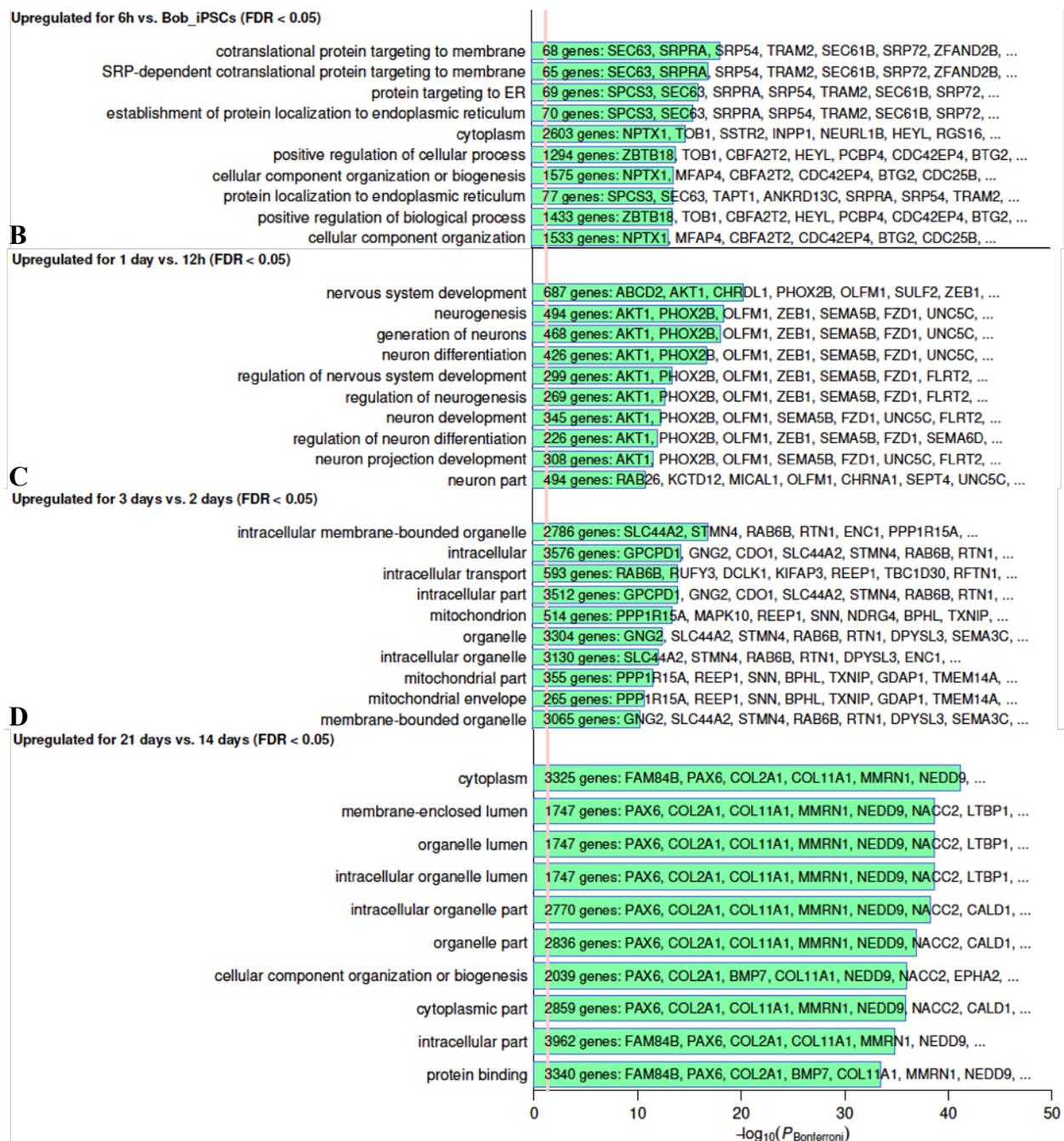


Figure 4. 5: Top gene ontologies enriched for significantly upregulated genes at specific points of NGN2 reprogramming.

Top gene ontologies enriched for significantly upregulated genes between (A) 6 hours post-induction vs iPSC stage, (B) 1day vs 12 hours, (C) 3 days vs 2 days and (D) 21 days vs 14 days post-induction. For each set of DE genes, the plot shows the 10 highest-enriched GO terms. Bars represent $-\log_{10}$ of the Bonferroni-corrected enrichment p-value. The red line represents the threshold $P_{\text{Bonferroni}} = 0.05$.

Modules can then be analysed using enrichment analysis to provide insights into their biological relevance. In addition, WGCNA can be used to identify highly connected genes that are centrally located in the module. The value of a gene's membership to a module is calculated, with highly connected genes given a higher membership score. Genes with the highest module membership are called hub genes. Some genes may have high continuous module membership in two or more modules and may, in this sense, be considered members of (or intermediate between) several modules.

WGCNA uses a topological overlap-based dissimilarity as input to average-linkage hierarchical clustering that results in a dendrogram. Modules are identified as branches in the dendrogram using Dynamic Tree Cut. The gene dendrogram and final module labels are shown in **Fig. 4.6**. A total of 21 different main modules were identified and the following results will only consider these modules.

After module identification, an enrichment analysis was performed on genes in each module in the same collections that were used for enrichment calculations of the DE analysis. The enrichment labels are used in the following module heatmap and boxplot. Each labelled module was then given a single representative expression profile called a module eigengene, making it easier to analyse their association to individual timepoints in the NGN2 reprogramming time course. This analysis is visualised here in two ways: a heatmap, for each eigengene and each time point, where the values of the three replicates for each timepoint are averaged (**Fig. 4.7**); and boxplots of module eigengenes against time, showing the expression value of each replicate for a single timepoint, enriched terms and top hubs in the module (**Fig. 4.8, 4.9 and 4.10**).

With 21 different modules identified, several showed no enrichment for any biological terms, while the relevance of some that were identified were not very clear. Therefore, I have chosen to focus on the most relevant modules that will prove insightful towards the aims of this study. The first of this is module 3 (M.3), which as shown in **Fig. 4.8A**, represents a network of genes that was gradually downregulated from the iPSC stage to below zero at Day 1 of reprogramming. Interestingly, the most significant term enriched for this module is *Down-regulated in Huez2.3line NSCs vs ESC*, a term obtained from the Drug SIGNature Database (DSigDB). This suggests the network of genes suppressed by NGN2 overexpression was not just a pluripotency network, but specifically a network of genes downregulated as embryonic stem cells (ESCs) differentiate into neural stem cells (NSCs). Inverse to module 3, is module 7, which is marked by the term *Up-regulated in SIVF017 line NSCs vs ESCs* (DSigDB), showing upregulation beginning at Day 1, but only up to Day 4, after which it was noticeably downregulated (**Fig. 4.8B**). The enrichment

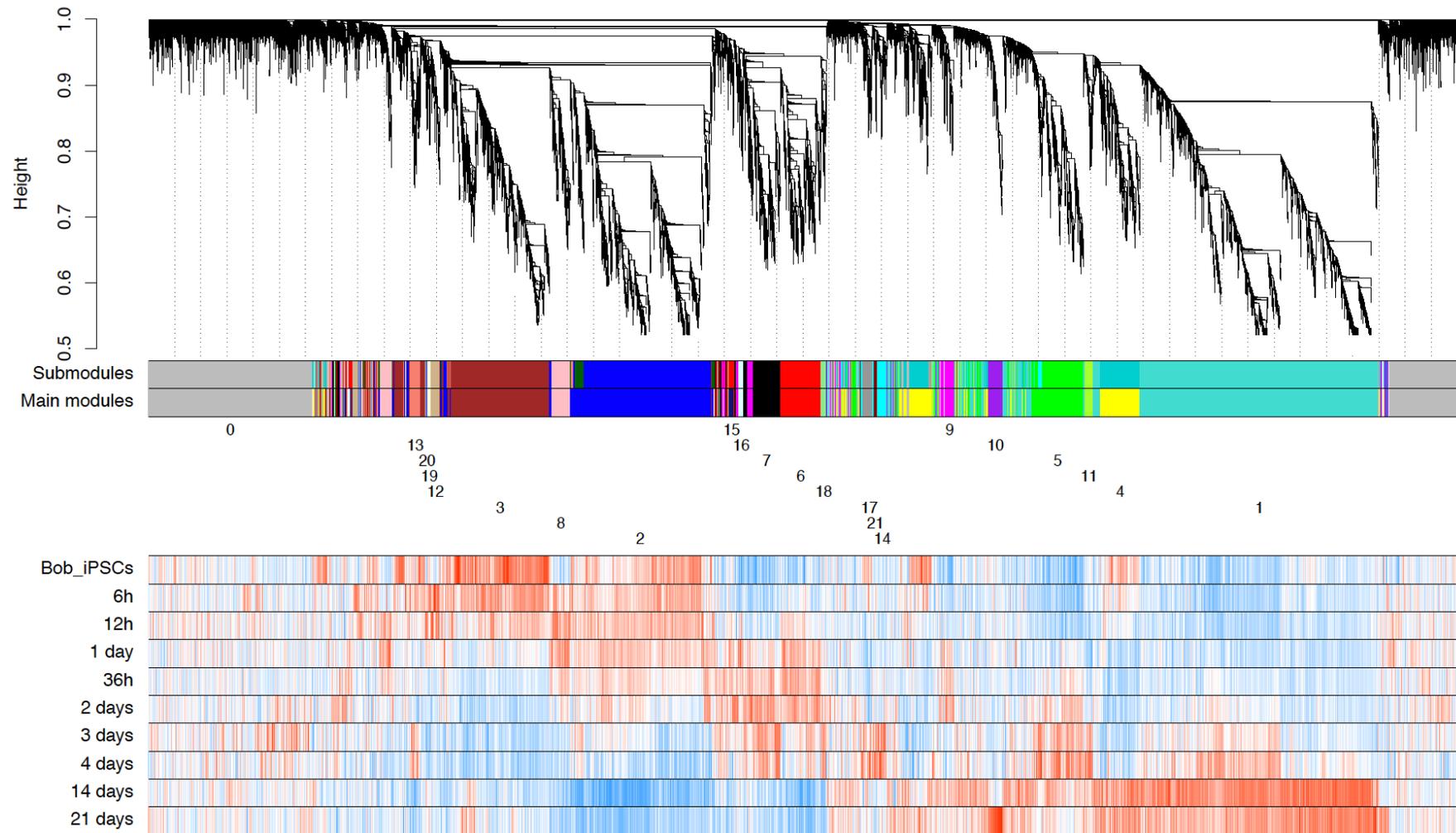


Figure 4. 6: Gene clustering tree (dendrogram), module colours and labels, and a heatmap of individual gene association with each timepoint. Numeric module labels are only shown for main modules. Heatmap below the clustering tree indicates mean expression for each gene in the 3 samples corresponding to each time point. Blue (red) means below (above) the average.

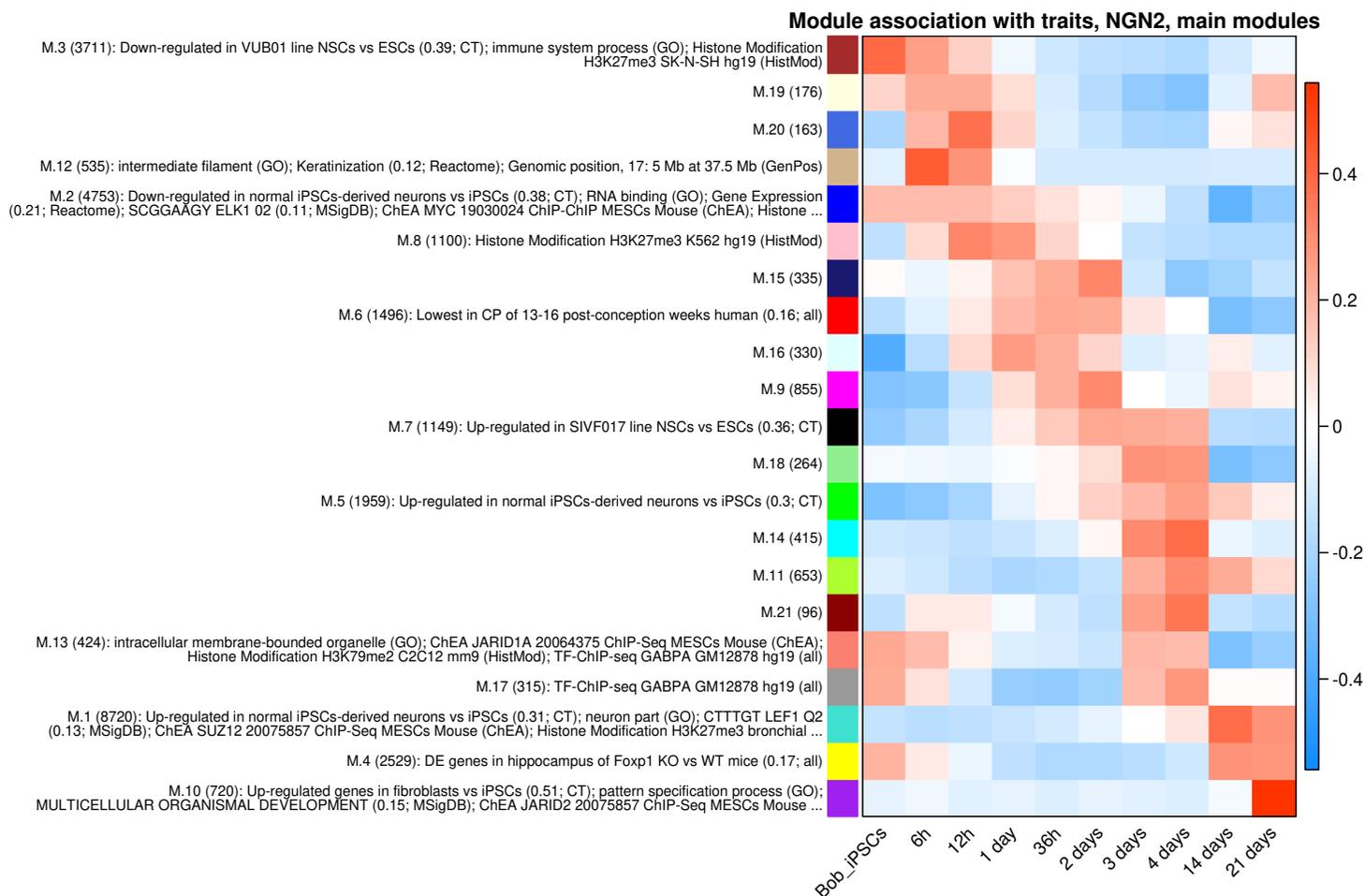


Figure 4. 7: Heatmap representation of time dependence of module eigengenes of main modules.

Each row corresponds to a module. Row labels indicate the numeric module label, module size and top enrichment terms. Each column corresponds to a time point. Heatmap shows the mean value of the scaled eigengenes at each time point (blue and red represent under- and over-expression relative to the mean of each eigengene). The eigengenes are ordered by hierarchical clustering of their distances, which places eigengenes with similar profiles in close proximity.

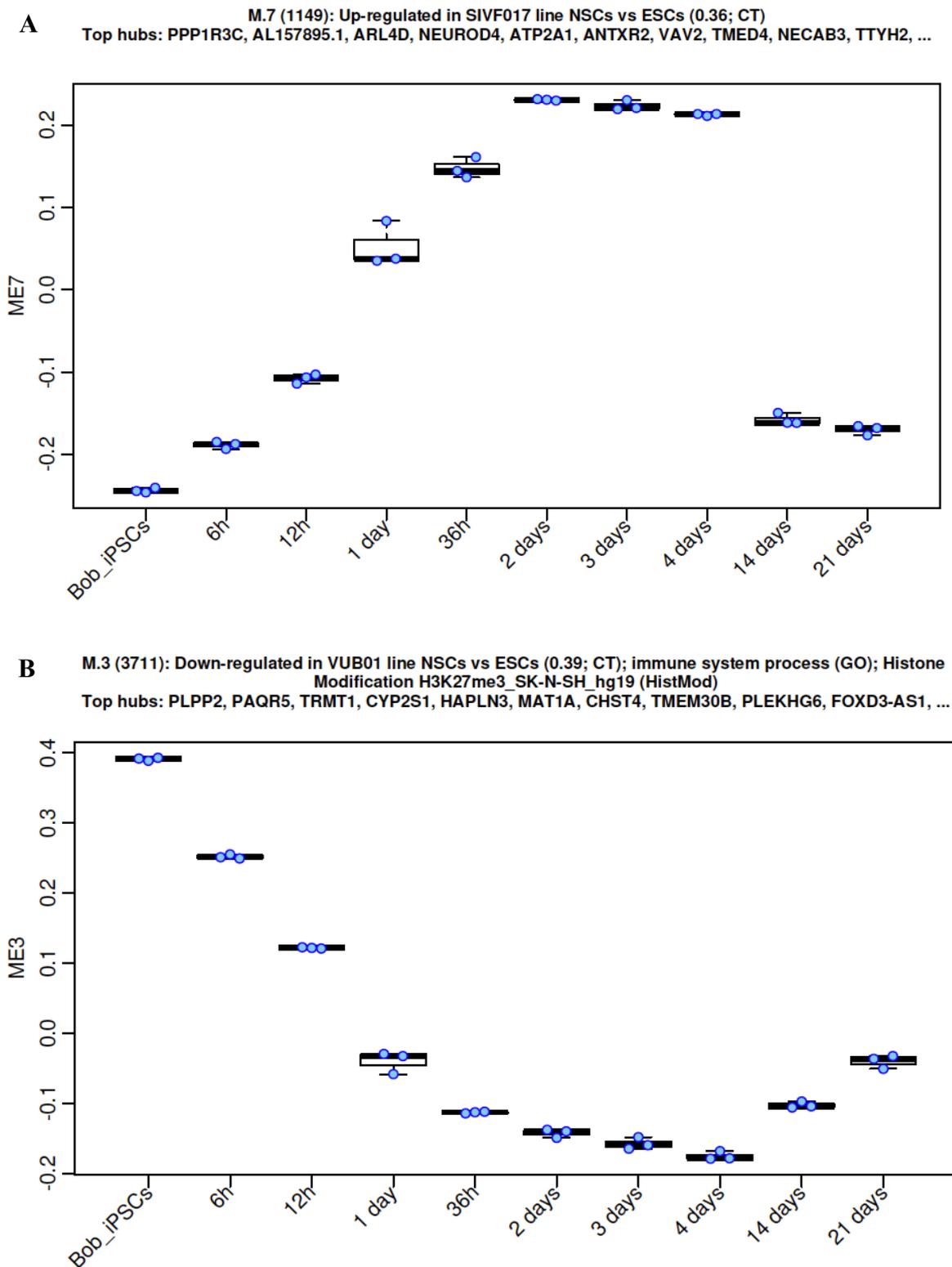
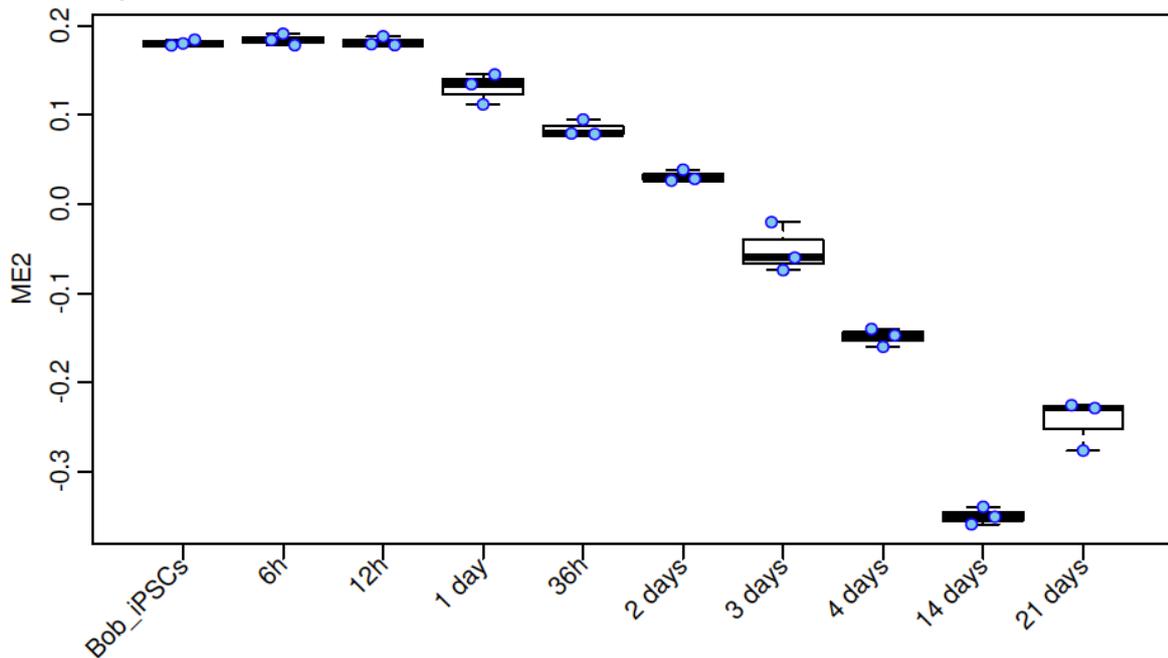


Figure 4. 8: Boxplot of module eigengenes 3 & 7 vs. time.

(A) Module 3 and (B) module 7 identify a network of genes associated with NSCs. Each box represents the inter-quartile range and the thick line in the box represents the median. Blue points show the actual expression values (one dot per sample). The title shows the module label and

number of genes as well as the trait, top enrichment terms and the top hub genes (ordered by decreasing kME).

A M.2 (4753): Down-regulated in normal iPSCs-derived neurons vs iPSCs (0.38; CT); RNA binding (GO); Gene Expression (0.21; Reactome); SCGGAAGY_ELK1_02 (0.11; MSigDB); ChEA MYC_19030024_ChIP-ChIP_MESCs_Mouse (ChEA); Histone Modification H3K79me2_C2C12_mm9 (HistMod); mirTarBase hsa-miR-193b-3p (mirTar)
 Top hubs: NPM1, CSE1L, USP39, SLC39A1, FBL, BRX1, GNL3, RFWD3, DESI2, HSPD1, MIS18A, LAPTM4B, ...



B M.5 (1959): Up-regulated in normal iPSCs-derived neurons vs iPSCs (0.3; CT)
 Top hubs: ADARB2, UNCX, XKR7, CALCOCO1, RGS10, SGCB, CREBRF, RBM18, TTC14, DNAJC27, PKD1, ...

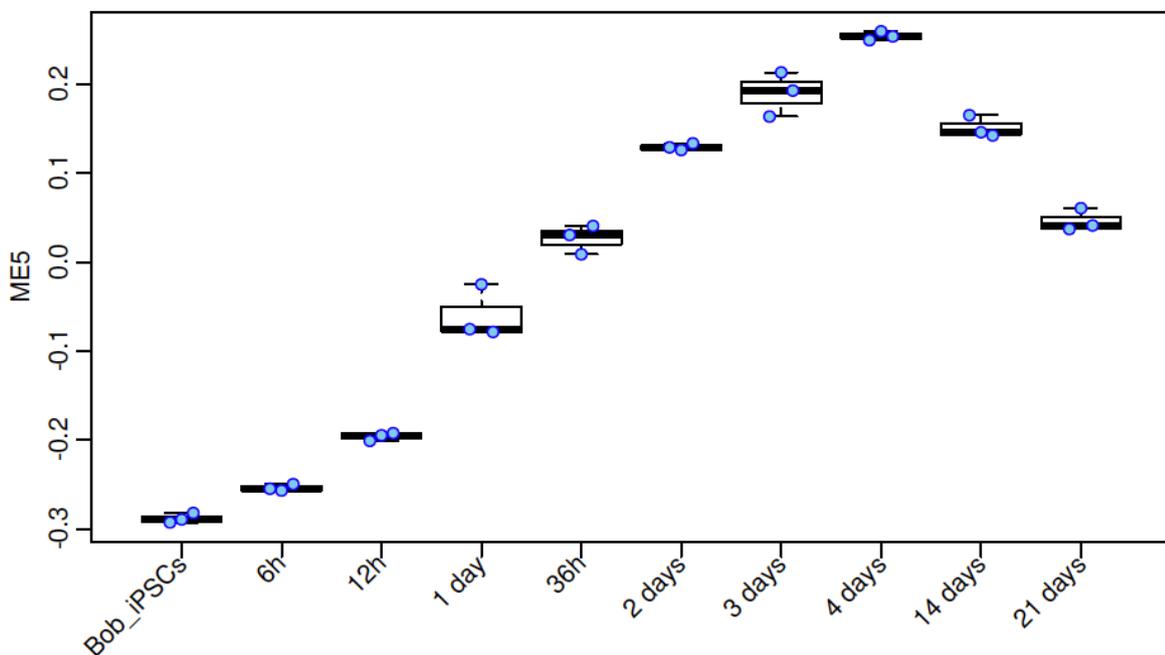


Figure 4.9: Boxplot of module eigengenes 2 & 5 vs. time.

(A) Module 2 and (B) module 5 identify a network of genes associated with iPSC-derived neurons. Each box represents the inter-quartile range and the thick line in the box represents the median. Blue points show the actual expression values (one dot per sample). The title shows the module label and number of genes as well as the trait, top enrichment terms and the top hub genes (ordered by decreasing kME).

for these two modules suggests that cells undergoing NGN2 reprogramming from iPSCs progressed through a transient progenitor-like stage, after 24 hours of reprogramming. One of the top hubs identified from this module is NEUROD4, a well-known target of NGN2. This result may explain why samples from Day 1 to Day 2 clustered more closely together in **Fig. 4.2A** and neuronal terms were highly enriched beginning at Day 1 (**Fig. 4.5.B**).

Similarly, module 2 is topped by the term *Down-regulated in iPSC-derived neurons vs iPSCs* (DSigDB), which reaches near zero around Day 2, coinciding with the appearance of neuronal morphologies during the reprogramming (**Fig. 4.9A**); thus, suggesting that shutting down the network of genes in this module is potentially critical for the establishment of neuronal fate. Conversely, module 5 identifies a network of genes that showed highest association with the term *Up-regulated in normal iPSC-derived neurons vs iPSCs* (DSigDB) and begins upregulation around 36 hours to 48 hours post-induction, coinciding with the downregulation of module 2 (**Fig. 4.9B**). Two genes in module 5, NEUROD1 and POU3F2 (BRN2) are TFs known to be involved in neuronal differentiation. Another interesting network of genes is module 1, which upregulation goes above zero at Day 3 onwards before being highly expressed at Day 14 and Day 21 (**Fig. 4.10**). Just like module 5, this network showed highest association with the term *Up-regulated in normal iPSC-derived neurons vs iPSCs* (DSigDB); however, unlike module 5, this cluster contained notable markers of neuronal maturity such as all 4 glutamate AMPA-receptor subunit genes, GRIA1 to GRIA4, indicating that this module could be a network important for establishing neuronal maturity and the functionality observed in the electrophysiological studies.

4.3.4 scRNAseq analysis of NGN2 reprogramming

Analysis of the scRNAseq dataset was done using two different bioinformatic analytical tools for scRNAseq: Scanpy (Wolf et al., 2018) and Monocle 3 (Trapnell et al., 2014b). Scanpy was used for UMAP-based visualisation of the data and clustering, whereas, Monocle 3 was used to construct lineage trajectories and pseudotime analysis. Differential expression analysis was carried out in both instances, to identify genes DE in a specific population of cells.

As a first step, I considered the entire time course for visualisation using UMAP-based dimensionality reduction, to get an overview of the time course (**Fig. 4.11**). Although the time points are in chronological order, it also placed the different stages of the programming, such as the iPSC stage, the following 48 hours, Day 3 and Day 4, and the late stage of the reprogramming, far from each other, which would make it difficult to infer any lineage and pseudotime trajectories

M.1 (8720): Up-regulated in normal iPSCs-derived neurons vs iPSCs (0.31; CT); neuron part (GO); CTTTGT_LEF1_Q2 (0.13; MSigDB); ChEA SUZ12_20075857_ChIP-Seq_MESCs_Mouse (ChEA); Histone Modification H3K27me3_bronchial epithelial cell_hg19 (HistMod); Up-regulated in normal human brain vs other fetal tissues (0.34; all)
 Top hubs: CPEB3, KCNB1, ZCCHC18, ZNF540, MYO5A, ZNF713, CELF4, ANKRD36C, NA, AC096733.2, ...

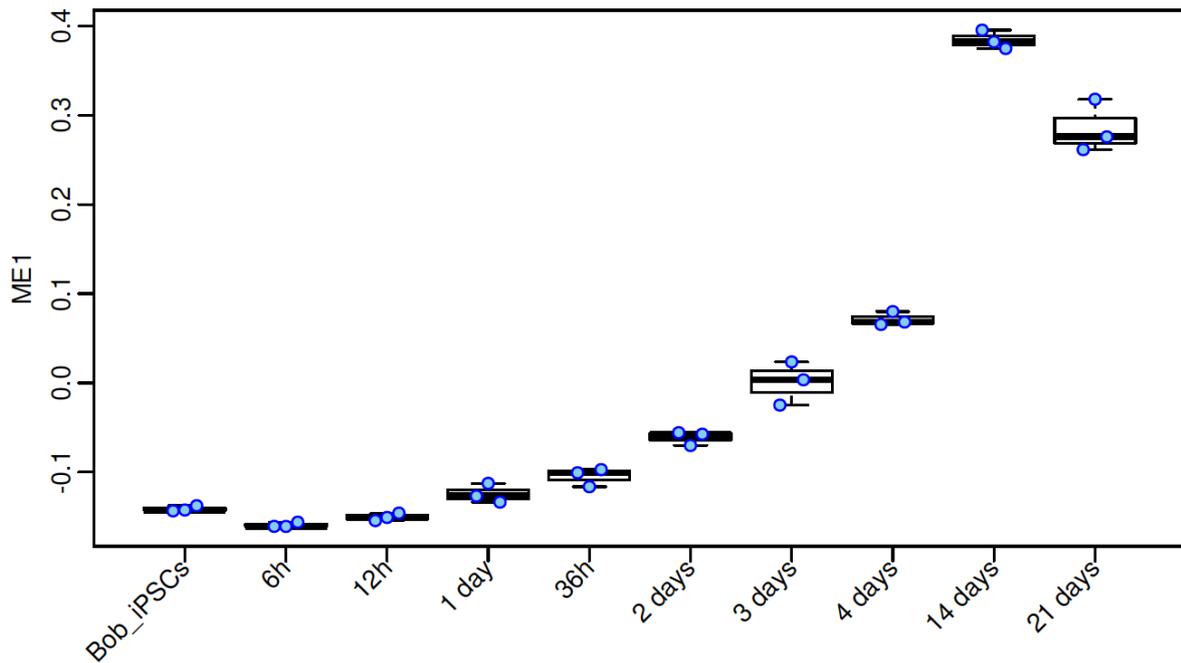


Figure 4. 10: Boxplot of module eigengene 1 vs. time maturation module

Module 1 identifies a network of genes associated with iPSC-derived neurons. Each box represents the inter-quartile range and the thick line in the box represents the median. Blue points show the actual expression values (one dot per sample). The title shows the module label and number of genes as well as the trait, top enrichment terms and the top hub genes (ordered by decreasing kME).

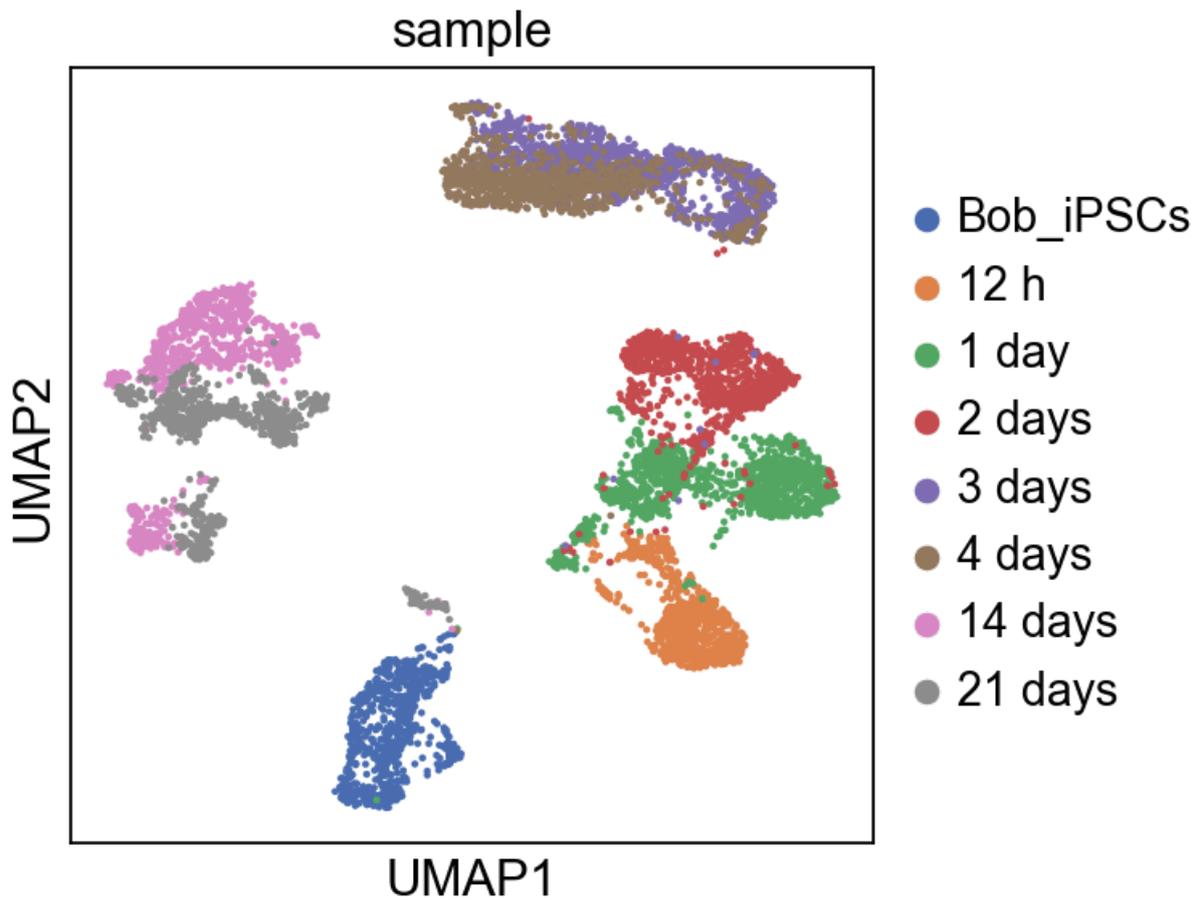


Figure 4. 11: UMAP visualisation of NGN2 reprogramming time course.

UMAP plot showing visualisation of single cells collected from 8 time points of the NGN2 reprogramming time course. The arrangement of each timepoint follows a chronological order from the iPSC stage, with little overlap between them, except for 3 days and 4 days post-induction. In addition, the data is also arranged in 4 distinct partitions: the iPSC stage, 12 hours to 2 days, 3 and 4 days, and finally 14 and 21 days.

in downstream analysis. This is likely because there is a large transcriptional difference between the early and late stage of reprogramming; thus, I decided to analyse these two stages separately.

Visualising just the early stage samples, their spatial arrangement followed a chronological order, up to Day 3 and Day 4, where there appears to be little differences between the two timepoints (**Fig. 4.12A**). A visualisation of NGN2 showed that the gene was strongly expressed in all cells after the iPSC stage (Day 0), with a noticeable increase in expression in the Day 2 population, which may be due to upregulation of the endogenous NGN2 (**Fig. 4.12B**). To identify clusters within the UMAP plot, I used the Louvain graph-clustering method, where community detection is based on optimising modularity (Levine et al., 2015) (**Fig. 4.12C**). This identified 8 different clusters, most of which largely correlating with a specific time point, except for clusters 2 and 4, which identified two separate populations at Day 1, and clusters 0, 6 and 7 which identified two different populations containing both Day 3 and Day 4 cells. In addition, cluster 2 identified a subpopulation of cells at 12 hours post-induction.

Next, as a first step towards annotating each cluster and understanding the biological processes underlying them, I carried out a DE analysis to identify the top genes in each cluster using Scanpy's Wilcoxon rank-sum (Mann-Whitney-U) test (**Fig. 4.13**). This revealed several genes that correlate with the main stages detected from the bulk RNAseq dataset, such as the NPC gene vimentin enriched at 24 hours post-induction (cluster 2) (**Fig. 4.15**), cell cycle genes such as UBE2C and HMGB2 in a subpopulation of Day 3 and Day 4 (cluster 6) (**Fig. 4.17**), and genes associated with neuronal development such as NEUROD1, DCX and TUBA1A enriched in the largest subpopulation of Day 3 and Day 4 samples (cluster 0) (**Fig. 4.18**). Interestingly, the DE analysis also identified a small sub-population of Day 3 and Day 4 cells that expressed ISL1 (cluster 7), which is recognised for its role in development of motor neuron (**Fig. 4.19**).

I next sought to further explore these different observations from the early stage of reprogramming, by visualising known genes associated with them. Firstly, I looked at pluripotency genes such as NANOG, POU5F1 (OCT4) and SOX2 to determine when the cells exit pluripotency and how heterogenous this process is. As seen with the bulk datasets, downregulation of pluripotency genes was rapid, with NANOG and MYC being entirely suppressed after 12 hours (**Fig. 4.14 A & B**). However, visualisation of SOX2 and POU5F1 found that they were heterogeneously expressed at the iPSC stage and subsequently at 12 hours and Day 1 (**Fig. 4.14 C**). I hypothesised that this could be due to a population of iPSCs that were drifting towards a neuroepithelium fate. A check for an early marker of this, PAX6, found that indeed the SOX2⁺/POU5F1⁻ population (SOX2⁺ here onwards) were also positive for this gene, along with

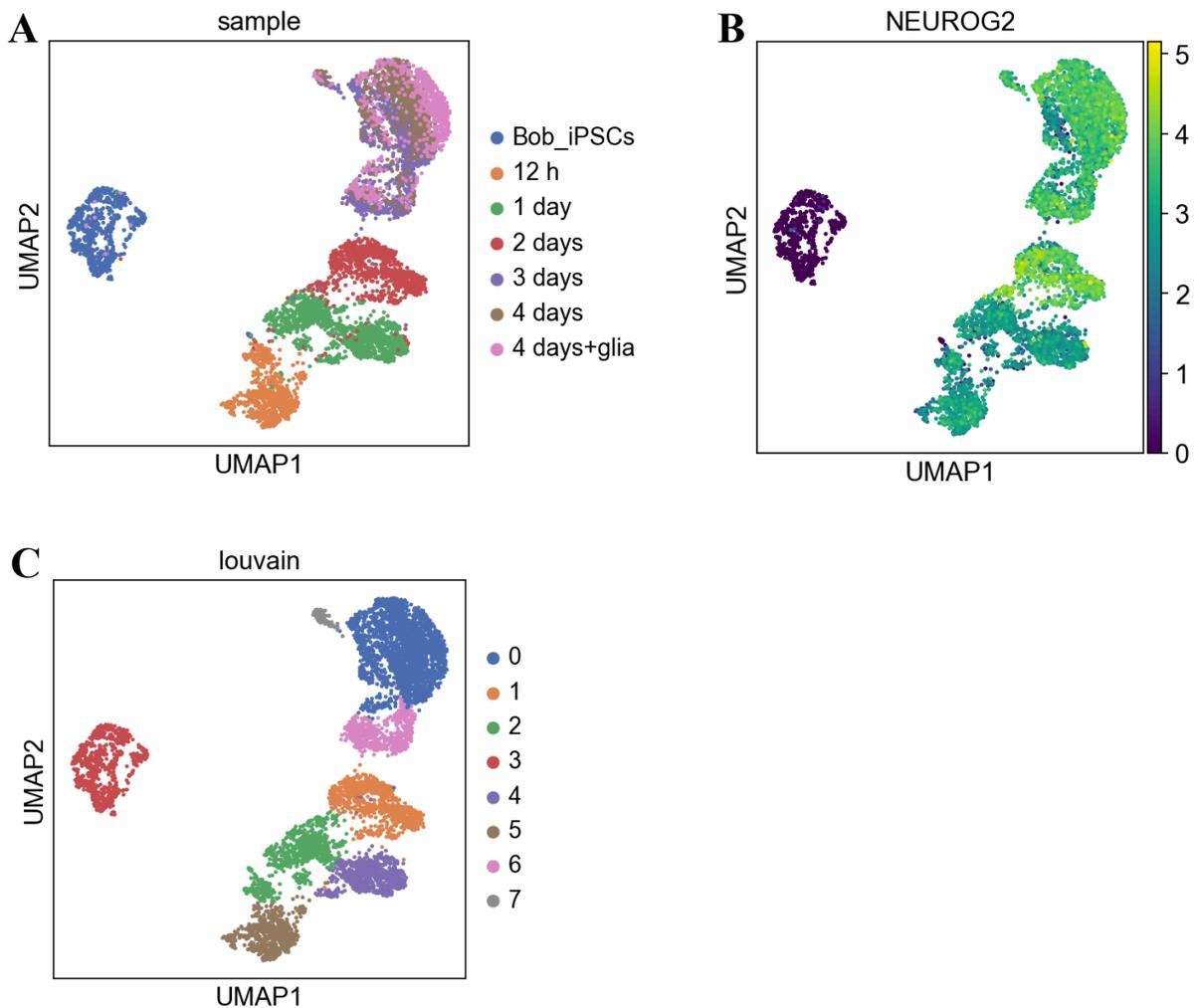


Figure 4. 12: UMAP visualisation of the early stage of NGN2 reprogramming and its Louvain clusters.

(A) The early stage of reprogramming is visualised by considering only the first four days. (B) UMAP visualisation for the expression levels of NGN2 across the population. (C) Louvain clustering of these time points identified seven different clusters, with only the iPSC stage forming an individual cluster, while 12 hours to day 2 being part of two different clusters and days 3 and 4 being part of three different clusters.

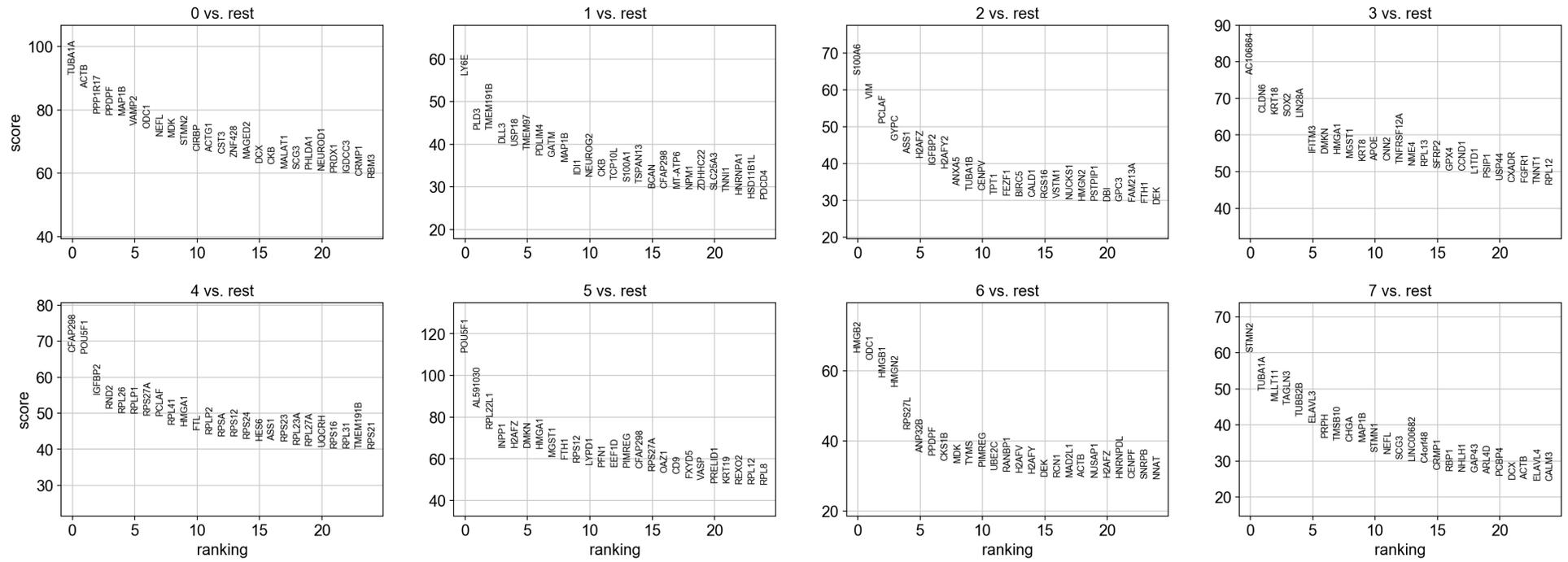


Figure 4.13: Differential expression analysis of clusters from early stage of NGN2 reprogramming.

The top 25 DE genes are shown for each of the 7 clusters from the early stage of reprogramming, with each gene ranked (x-axis) in order of decreasing test score (y-axis).

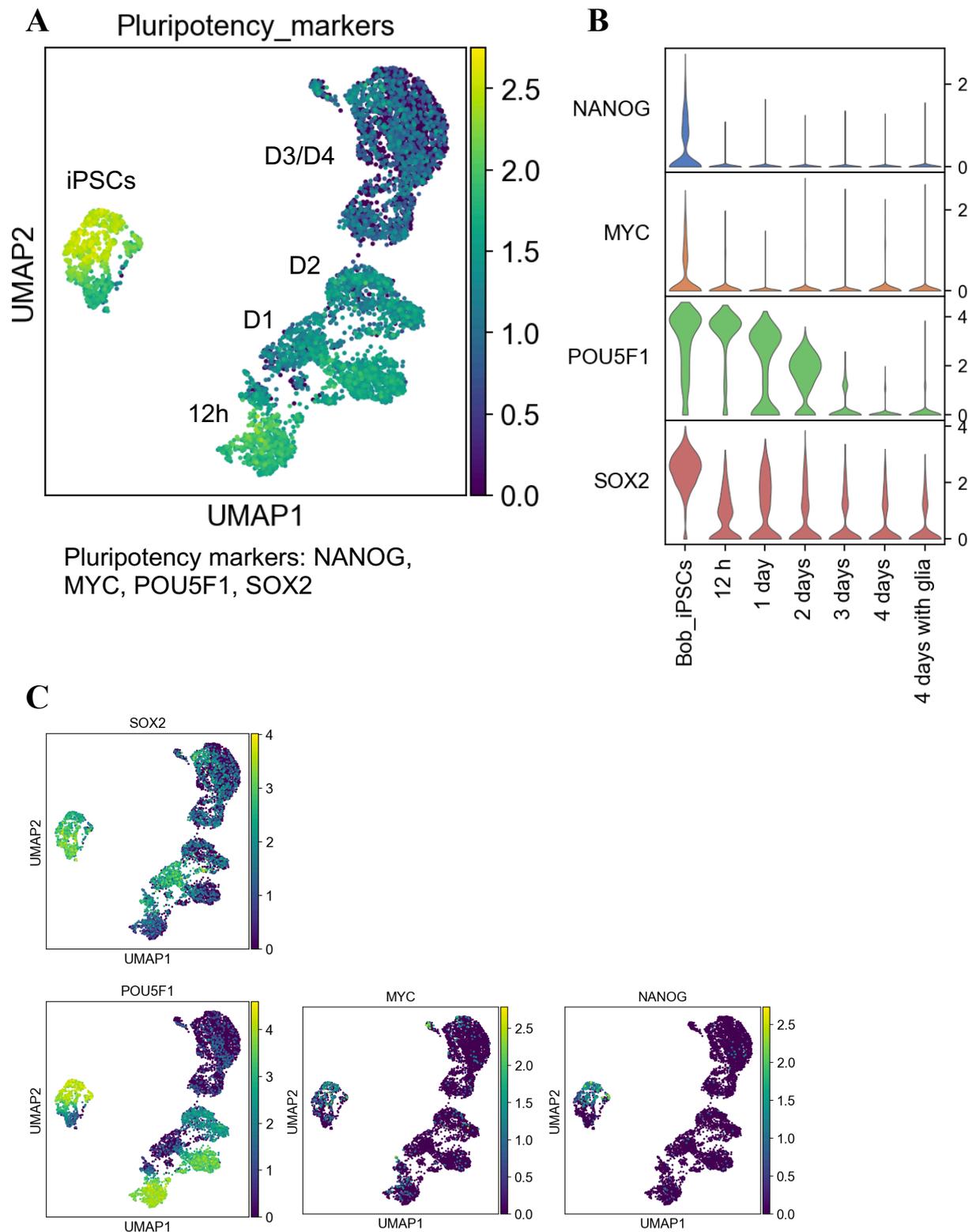


Figure 4. 14: Rapid downregulation of pluripotency genes during early stage of NGN2 reprogramming.

(A) UMAP visualisation showing an overlay of pluripotency markers over the first four days of reprogramming. Time points are placed next to their corresponding cluster of cells, and the genes used for the overlay are listed beneath the plot. (B) Stacked violin plots for expression levels of

four pluripotency genes at each time point. (C) Individual UMAP plots for each of the four pluripotent genes, showing a subpopulation of the iPSCs not expressing POU5F1, MYC and NANOG. This population also has a noticeably higher expression of SOX2.

NSC-associated genes such Vimentin, FABP7, FEZF1 and HES1 (**Fig. 4.15B**). At 12 hours and Day 1, we see a similar pattern of expression for some of these genes, strongly suggesting that the heterogeneity of the starting iPSC population was driving the clustering observed at 12 hours (cluster 2 vs cluster 5) and Day 1 (cluster 2 vs cluster 4), where cluster 2 is distinguished by SOX2 expression and the NPC associated genes found in the iPSC subpopulation, while cluster 4 and 5 distinguished by POU5F1 expression. The POU5F1⁺ cells at 12h and Day 1 have noticeably downregulated SOX2 expression, while the SOX2⁺ cells have downregulated POU5F1 expression. One way of determining if this heterogeneity originated from their associated populations at the iPSC stage is to plot a trajectory path between the samples; however, as mentioned earlier, it is visually evident that there are large transcriptional differences between some of the timepoints, mostly likely due to the rapid transcriptional changes induced by NGN2 as eluded to in the bulk RNAseq data earlier. This makes it difficult for the tools used to infer a trajectory path, between the iPSCs and the 12h sample in this case. Therefore, the following interpretation of the data will take into account the heterogeneity described and work on the high likelihood that it does indeed originate from the heterogenous iPSC population.

Despite this heterogeneity, several genes were identified that were not affected by this, and with that, provide some interesting insights about the dominant events that occur during the first 24 hours of NGN2 reprogramming. Firstly, the overexpression of the transgene was largely homogeneously expressed after 12 hours (**Fig 4.12B**); a possible demonstration of how the OPTi-OX system allows homogenous overexpression of a transgene. In addition, I decided to look at a known repressor of neurogenesis, REST (**Fig 4.16**), and a target of its repression, NEUROD4 (Drouin-Ouellet et al., 2017; Wang, Fong, & Huang, 2015). UMAP visualisation of REST expression shows that even at 12 hours, its expression had noticeably reduced, followed by significant downregulation at Day 1. This coincided with an increase in NEUROD4 expression that was expressed across Day 1 onwards. Compared to NEUROD1 and 2, NEUROD4 was the earliest of the NEUROD genes to be upregulated in this time course. The expression dynamics of REST and NEUROD4 does not appear to be affected by the difference in SOX2 and POU5F1 expression at this early stage, providing further evidence that this crucial repressor of neurogenesis is strongly and homogeneously downregulated in the first 24 hours of NGN2 reprogramming, regardless of whether the cells had a more pluripotent (POU5F1⁺) or NPC-like (SOX2⁺) transcriptional profile.

At Day 2 of the reprogramming, the expression of SOX2 was noticeably downregulated across the timepoint, with POU5F1 homogeneously expressed at a low level. This most likely explains why the whole timepoint was identified by a single cluster (cluster 1). As mentioned earlier, this

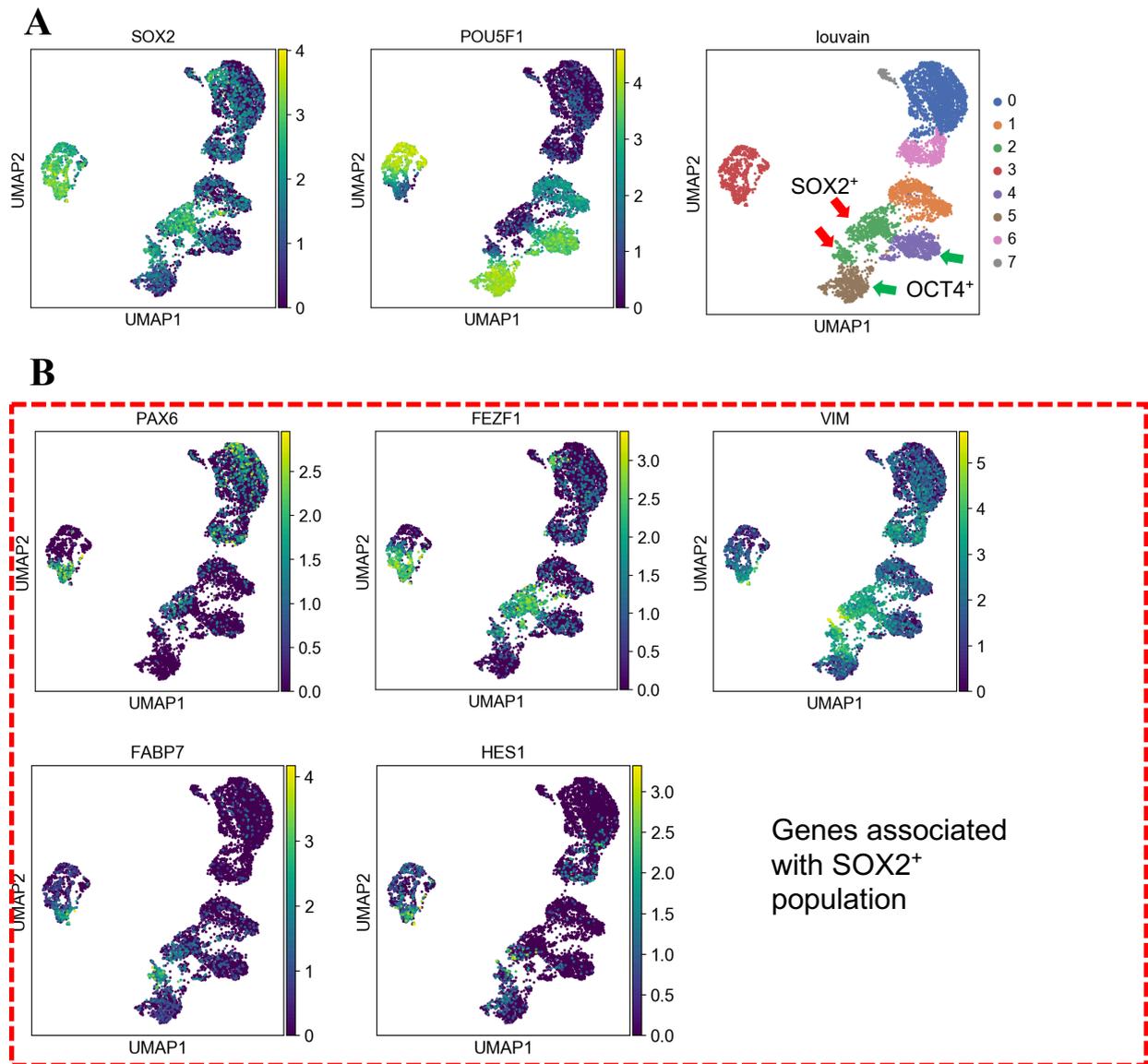


Figure 4. 15: Subpopulation of iPSCs expressed NPC genes

(A) UMAP of SOX2, POU5F1 and Louvain clusters. The heterogeneous expression of these two genes correlates with the clustering at 12 hours and Day 1 timepoint. Red arrows show clusters associated with SOX2 expressing cells (SOX2⁺), green arrows show clusters associated with POU5F1 expressing cells (POU5F1⁺). **(B)** Individual UMAP plots of NPC genes that are expressed in iPSCs and in SOX2⁺ clusters at 12h and Day 1.

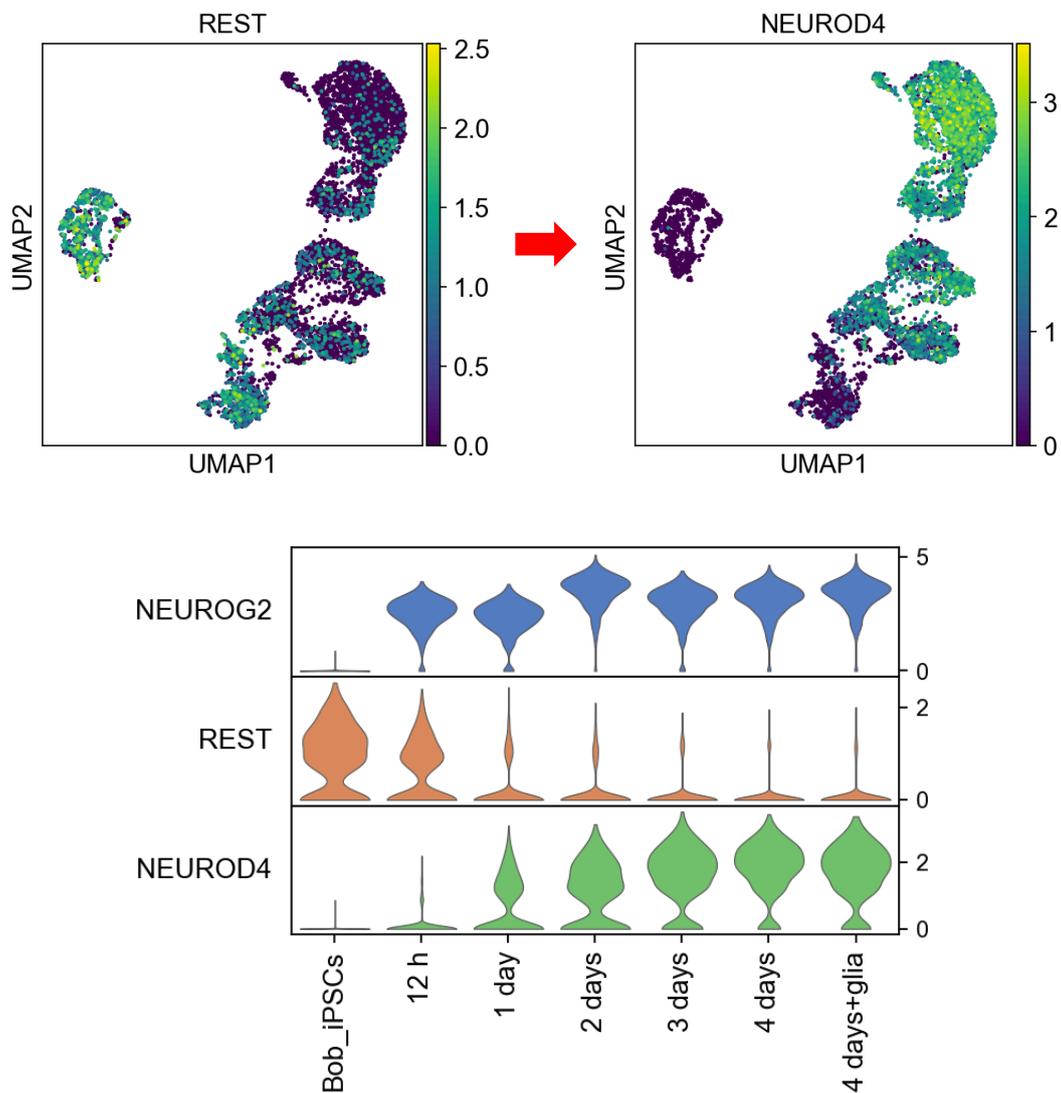


Figure 4. 16: Repressor of neurogenesis was efficiently downregulated within 24 hours of NGN2 reprogramming.

Top panel: UMAP plot showing the single cell expression of REST (left) and NEUROD4 (right) for the first four days of reprogramming. Bottom panel: Violin plots of the same genes, with the addition of NEUROG2 (NGN2). Together, they show that the neuronal repressor REST was rapidly downregulated following NGN2 overexpression and that this was followed by a concomitant increase in the pro-neuronal factor, NEUROD4

is also the point where there was a peak in NGN2 expression, which could have originated either from the transgene itself or endogenous NGN2 expression.

I next sought to investigate how the cell cycle was regulated in the early stage of reprogramming. An overlay of a collection of cell cycle related genes showed that cells undergoing NGN2 reprogramming became post-mitotic *en masse* at Day 3 or Day 4 post-induction, but this occurred as early as Day 1 of the reprogramming which is why a sizeable proportion of Day 2 cells had already exited the cell cycle (**Fig 4.17A**). A look at the proliferative marker MKI67 shows that its expression gradually increased from the iPSC stage until Day 2, after which it was drastically downregulated at Day 3 onwards. This large halt to cell cycle from Day 3 onwards, prompted me to look at cell cycle inhibitors, in particular CDKN1A, B and C. Between the three genes, CDKN1A had a significant increase at Day 2 of the reprogramming compared to previous time points (**Fig 4.17B**). Interestingly, this was widely expressed across the Day 2 population including the cells that had already exited the cell cycle, suggesting that it could be a gene that is closely regulated by NGN2 (**Fig 4.17C**). This increase in CDKN1A at Day 2 is most likely why a majority of the population became post-mitotic at Day 3 onwards.

Since some of the cells exited the cell cycle as early as Day 1, we see a similar pattern for markers of early differentiation such as SOX4 and NEUROD1, where predominantly they were upregulated just after cell cycle exit at Day 3 and 4, but also expressed earlier in the reprogramming (**Fig 4.18B**). There is also a noticeable enrichment for these genes at the Day 1 cells that were SOX2⁺, suggesting that their initial NPC-like state may have accelerated neuronal differentiation. DCX, a gene involved in guiding differentiating neurons to their appropriate anatomical positions in the early brain (Gleeson et al., 1999), was also among the genes immediately upregulated upon cell cycle exit (**Fig 4.18A**). As mentioned earlier, NEUROD genes play a key role in the formation of glutamatergic neurons in the cortex; therefore, these cells most likely identify the emerging glutamatergic neurons in the population. The upregulation of early differentiation genes was followed by pan-neuronal genes such as TUBB3 and SYT1, that are commonly expressed in immature neurons (**Fig 4.18C**). MAP2 and MAPT, which is associated with more mature neurons, was mostly expressed in some cells after the upregulation of the pan-neuronal genes

Next, I investigated the emergence of the ISL1⁺ cells in the population. Driven by the hypothesis that this could be identifying a sub-population of cholinergic neurons, I tested each Louvain cluster for the expression of other transcription factors with a known role in cholinergic development (**Fig 4.19B**). This involved PHOX2B, known for specifying cranial visceral motor

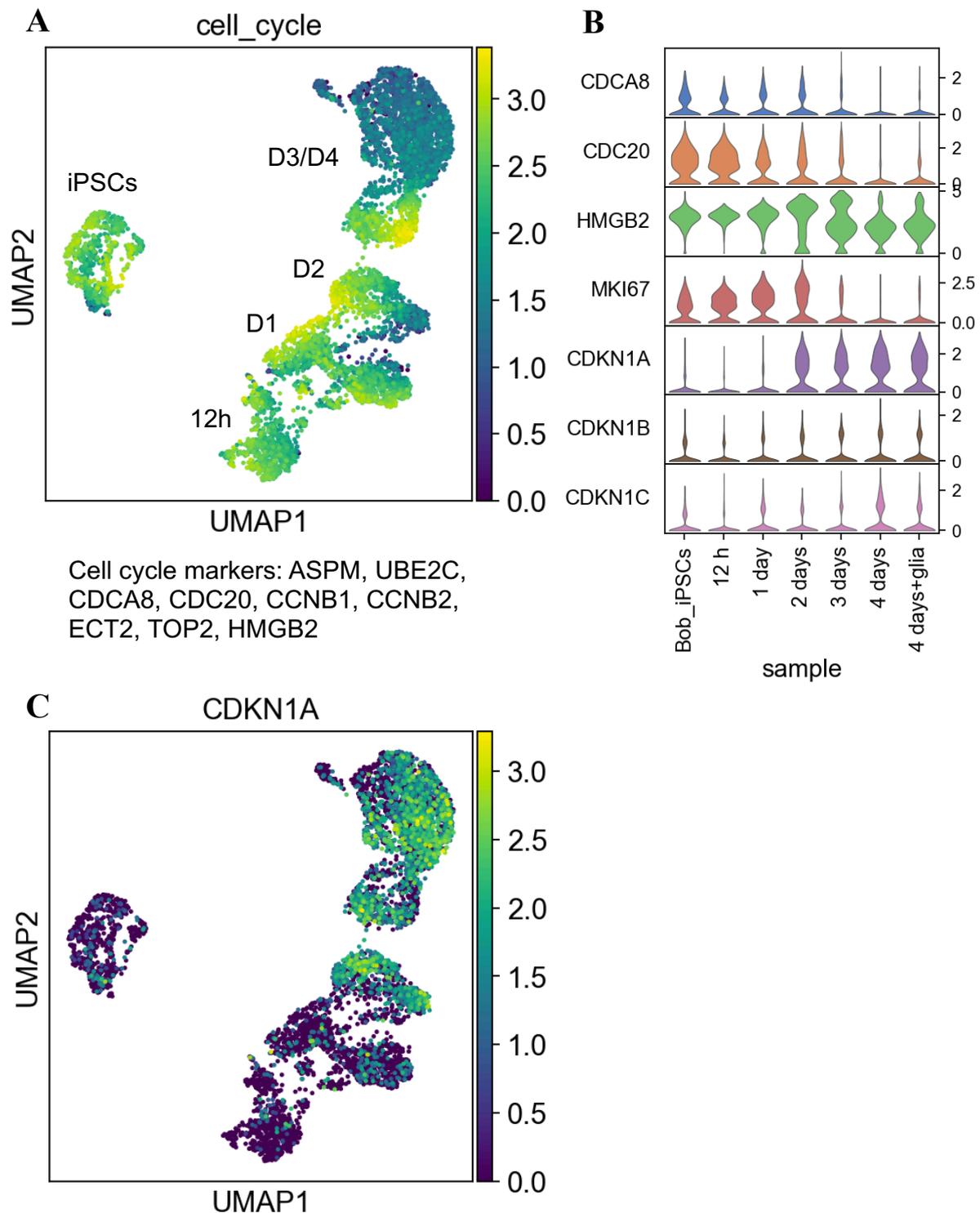


Figure 4. 17: NGN2 reprogrammed cells strongly expressed CDKN1A at Day 2 before mass cell cycle exit

(A) UMAP visualisation showing an overlay of cell cycle markers over the first four days of reprogramming. Time points are placed next to their corresponding cluster of cells, and the genes used for the overlay are listed beneath the plot. (B) Stacked violin plots for expression levels of cell cycle genes (top four) and inhibitor of cell cycle genes (bottom three) for each timepoint. (C)

UMAP visualisation for the expression of CDKN1A, showing homogenous expression across Day 2, including sub-population that exited cell cycle.

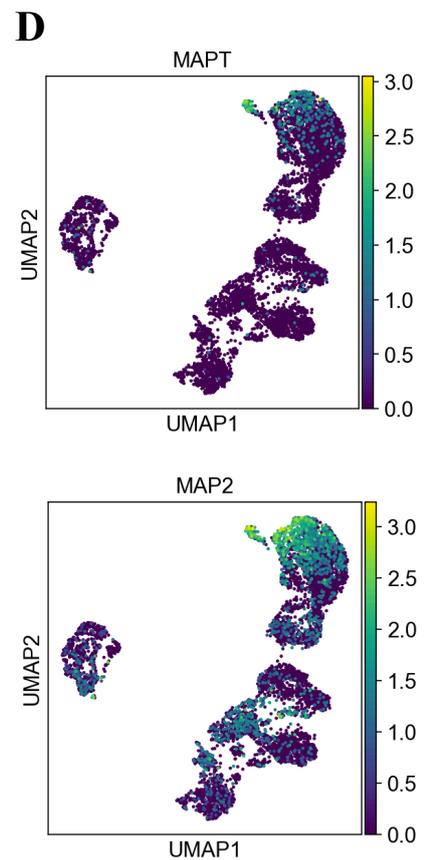
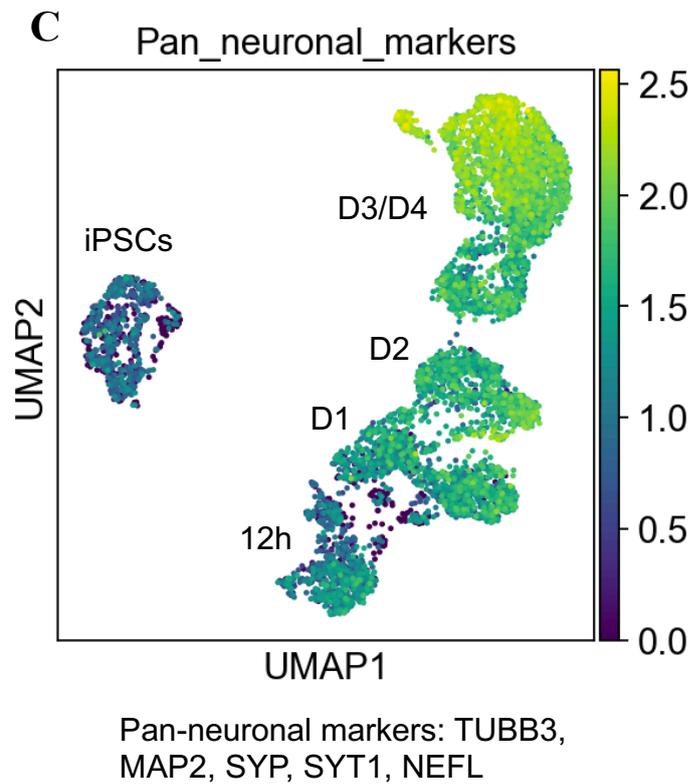
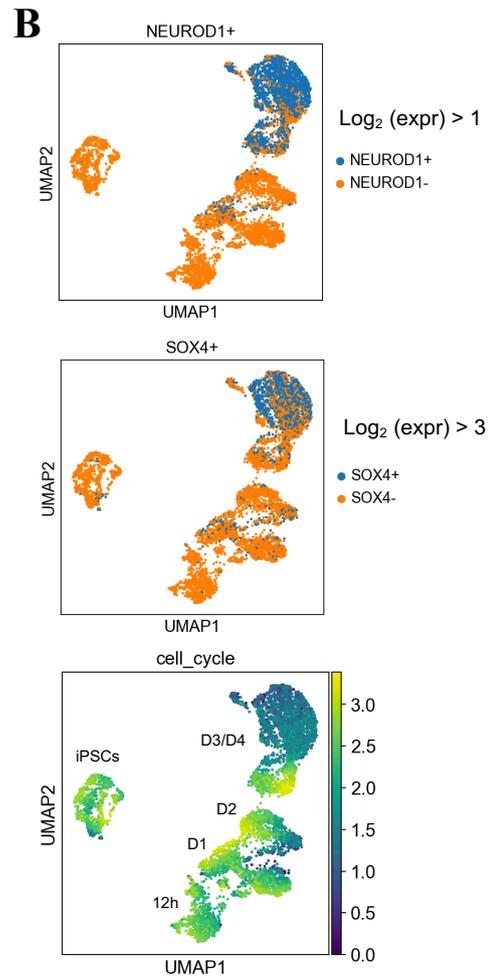
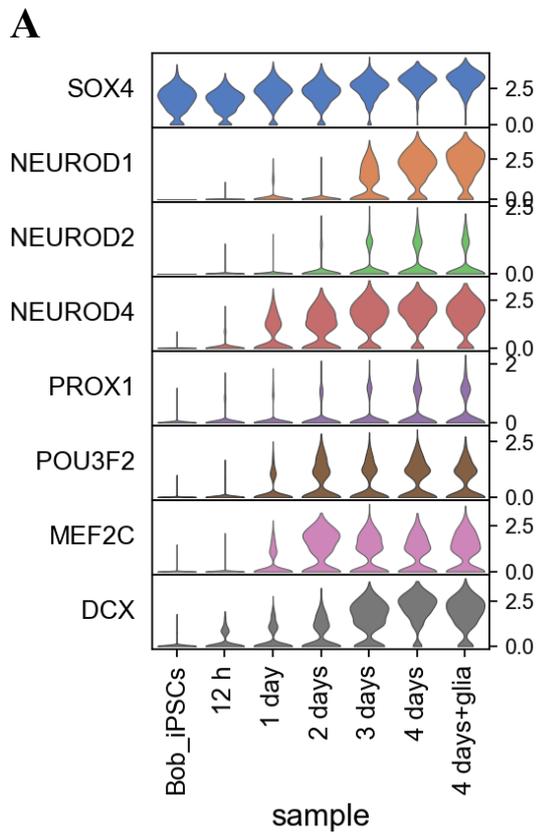


Figure 4. 18: Neuronal differentiation in the early stage of NGN2 reprogramming.

(A) Stacked violin plot of expression levels of neuronal differentiation genes for each time point. (B) UMAP visualisation of cells expressing NEUROD1 or SOX4, above a \log_2 expression cut-off of 1 and 3, respectively. UMAP plot of an overlay of cell cycle markers is included to illustrate that expression of neuronal differentiation genes occurred largely at Day 3 and Day 4 after mass cell cycle exit, but also at Day 1 and 2. (C) UMAP visualisation for the expression levels of pan-neuronal genes, TUBB3 and SYT1. (D) UMAP visualisation for the expression levels of mature neuronal genes, MAP2 and MAPT, which are largely expressed after the pan-neuronal genes.

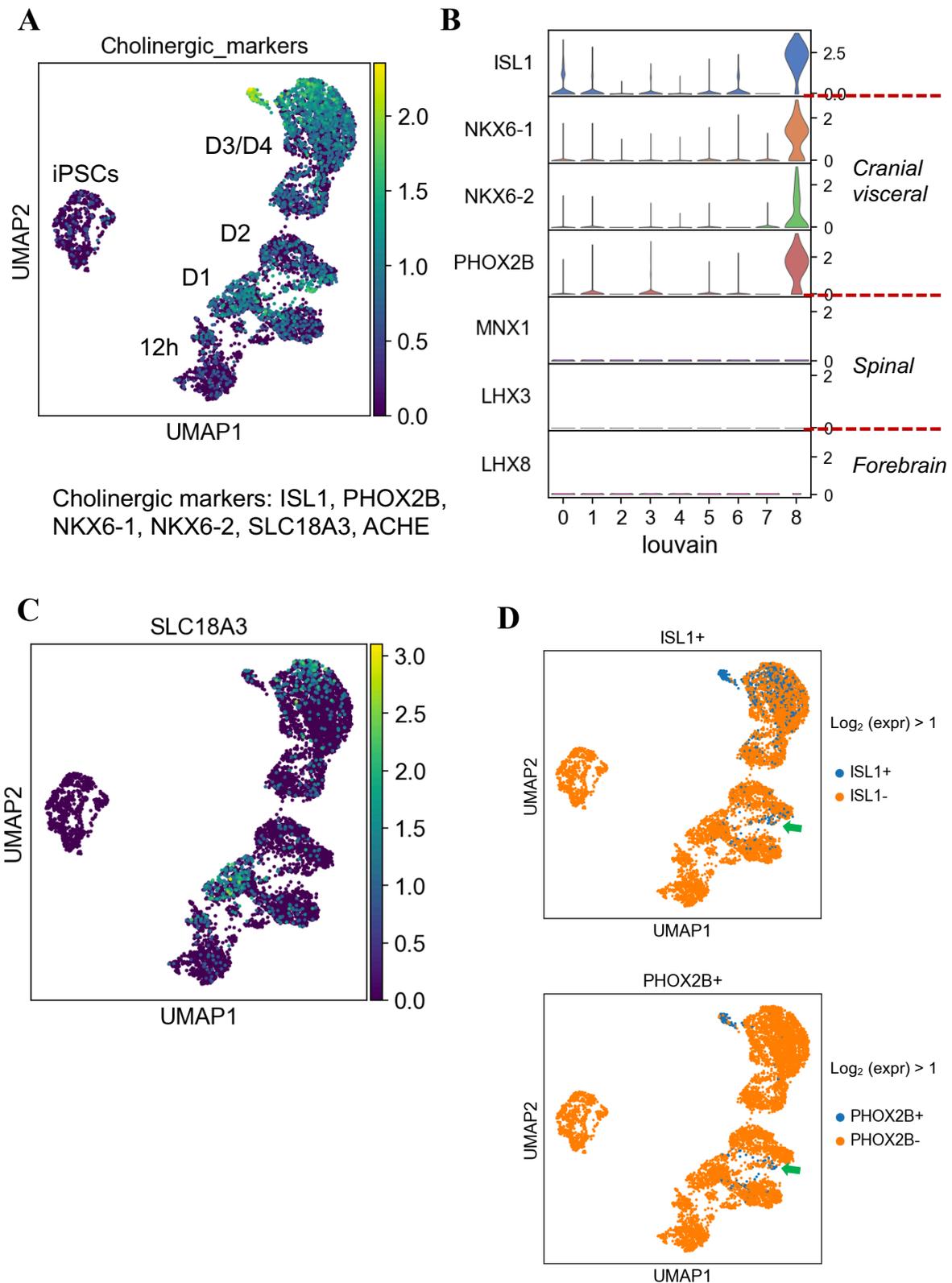


Figure 4. 19: Emergence of cholinergic fate in the early stage of NGN2 reprogramming.

(A) UMAP visualisation showing an overlay of cholinergic markers over the first four days of reprogramming. Time points are placed next to their corresponding cluster of cells, and the genes used for the overlay are listed beneath the plot. Red arrow indicates the earliest emergence of the

cholinergic fate. (B) Stacked violin plots for expression levels of ISL1 and genes for three different cholinergic lineages, at each time point. (C) UMAP visualisation for the expression level of SLC18A3, a gene integral to cholinergic neurotransmission. (D) UMAP visualisation of cells expressing ISL1 or PHOX2B, above a \log_2 expression cut-off of 1, showing earliest enrichment for both genes occurring at Day 2 (green arrow).

neurons; MNX1 and LHX3, which are involved in specifying spinal motor neurons and finally LHX8, which specifies forebrain motor neurons (Guthrie, 2007; Velasco, Ibrahim, Kakumanu, et al., 2017). In addition, I also tested for NKX6-1 and NKX6-2, two TFs expressed in progenitor domains that give rise to motor neurons in the developing brain (Guthrie, 2007). This showed a noticeable enrichment for ISL1, PHOX2B, NKX6-1 and NKX6-2 in cluster 7, which is made up of a subpopulation of cells from Day 3 and Day 4 of the reprogramming (**Fig 4.19B**). Individual inspection of the genes in a UMAP plot showed that both ISL1 and PHOX2B were also enriched in a small population of cells at Day 2 of the reprogramming, which interestingly, are also part of the Day 2 population that have exited the cell cycle (**Fig 4.19D**). On the other hand, NKX6-1 and NKX6-2 were not enriched in this early population of ISL1+/PHOX2B+ cells. Given that this population is made up of only a small number of cells, they were not identified as a separate cluster in the Louvain clustering step. I next investigated the expression of two genes involved in cholinergic neurotransmission, SLC18A3 (VACht) and ACHE (**Fig 4.19C**). Another important gene in this process is SLC5A7, a choline transporter at cholinergic synapses; however, it was one of the drop-out genes from this experiment. In cluster 7, there is a noticeable upregulation for ACHE, but a lower expression for SLC18A3.

To further demonstrate the emergence of this population at Day 2 of the reprogramming, specifically, emerging just after cell cycle exit, I conducted a pseudotime analysis of the Day 2 population using Monocle 3 (**Fig 4.20**). Further clustering of this specific timepoint, revealed 6 different clusters. For the pseudotime analysis, I chose cells with a higher expression of the cell cycle associated gene, HMGB2 as the root node (starting point) for the pseudotime, which involved a mix of cells in cluster 16 and cluster 14. Then, I analysed the pseudotime expression of HMGB2, the cholinergic factors determined from before, the pan-neuronal marker MAP2 and a gene associated with mature neurons, MAPT. In addition, I also investigated the expression of ONECUT2 and EBF2, two TFs that have been shown to be involved in generation of motor neurons from mouse ESCs through Ngn2, Isl1 and Lhx3 reprogramming (Velasco, Ibrahim, Kakumanu, et al., 2017). This revealed a coordinated regulation, where the downregulation of HMGB2 is followed by the simultaneous upregulation of ONECUT2, EBF2 and ISL1 (**Fig 4.20C**). This is then followed by PHOX2B *later* down the path in cluster 13. Although NKX 6-1 was upregulated in a few cells in this cluster, the effect appears to be non-significant. MAP2 was also upregulated late in this path; whereas, MAPT was not significantly upregulated. Next, I carried out the same analysis on the Day 3 and Day 4 population, to determine how these genes change in respect to pseudotime (**Fig 4.22**). As before, the population with cells that were still cycling were set as the starting point for pseudotime, in this case, corresponding to cells in cluster

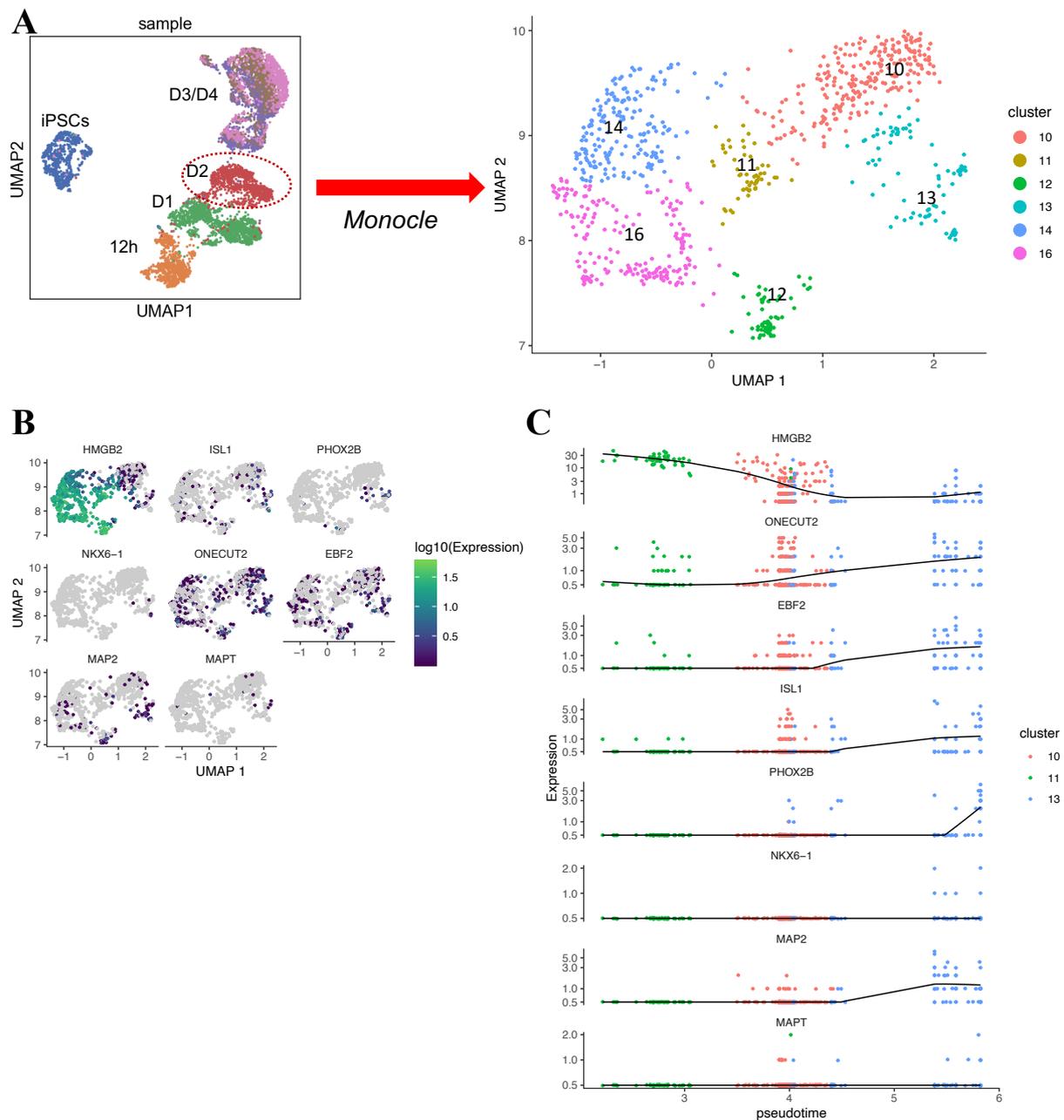


Figure 4.20: Pseudotime analysis of cholinergic fate at 2 days post-induction.

(A) Cells from Day 2 of reprogramming (left) were chosen for trajectory and pseudotime analysis using Monocle 3. Diagram shows UMAP visualisation of Day 2 cells only and 6 different clusters determined by Monocle. (B) Individual UMAP plots showing expression levels for the cell cycle marker, HMGB2, genes associated with cholinergic fate, ISL1, PHOX2B, NKX6-1, ONECUT2 and EBF2, and two pan-neuronal genes, MAP2 and MAPT. (C) Pseudotime analysis of the same genes for clusters 10, 11 and 13, showing the expression levels for each gene in individual cells, arranged in order of increasing pseudotime. Cells are coloured according to their cluster of origin.

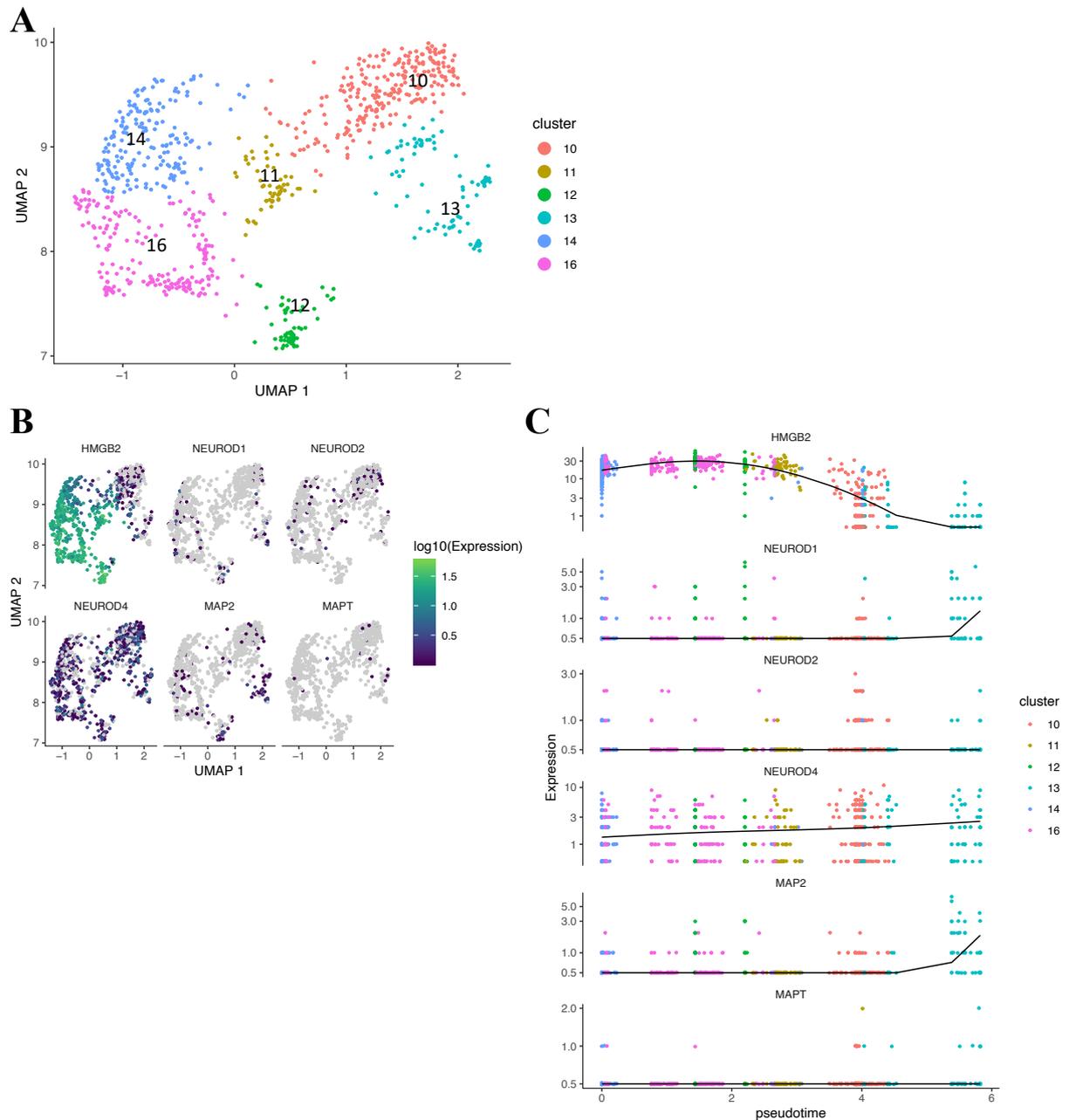


Figure 4.21: Pseudotime analysis of glutamatergic fate at 2 days post-induction.

(A) UMAP visualisation of Day 2 cells only, showing 6 different clusters identified by Monocle. (B) Individual UMAP plots showing expression levels for the cell cycle marker, HMGB2, genes associated with glutamatergic fate, NEUROD1, 2, and 4, and two pan-neuronal genes, MAP2 and MAPT. (C) Pseudotime analysis of the same genes for all Day 2 clusters, showing the expression levels for each gene in individual cells, arranged in order of increasing pseudotime. Cells are coloured according to their cluster of origin.

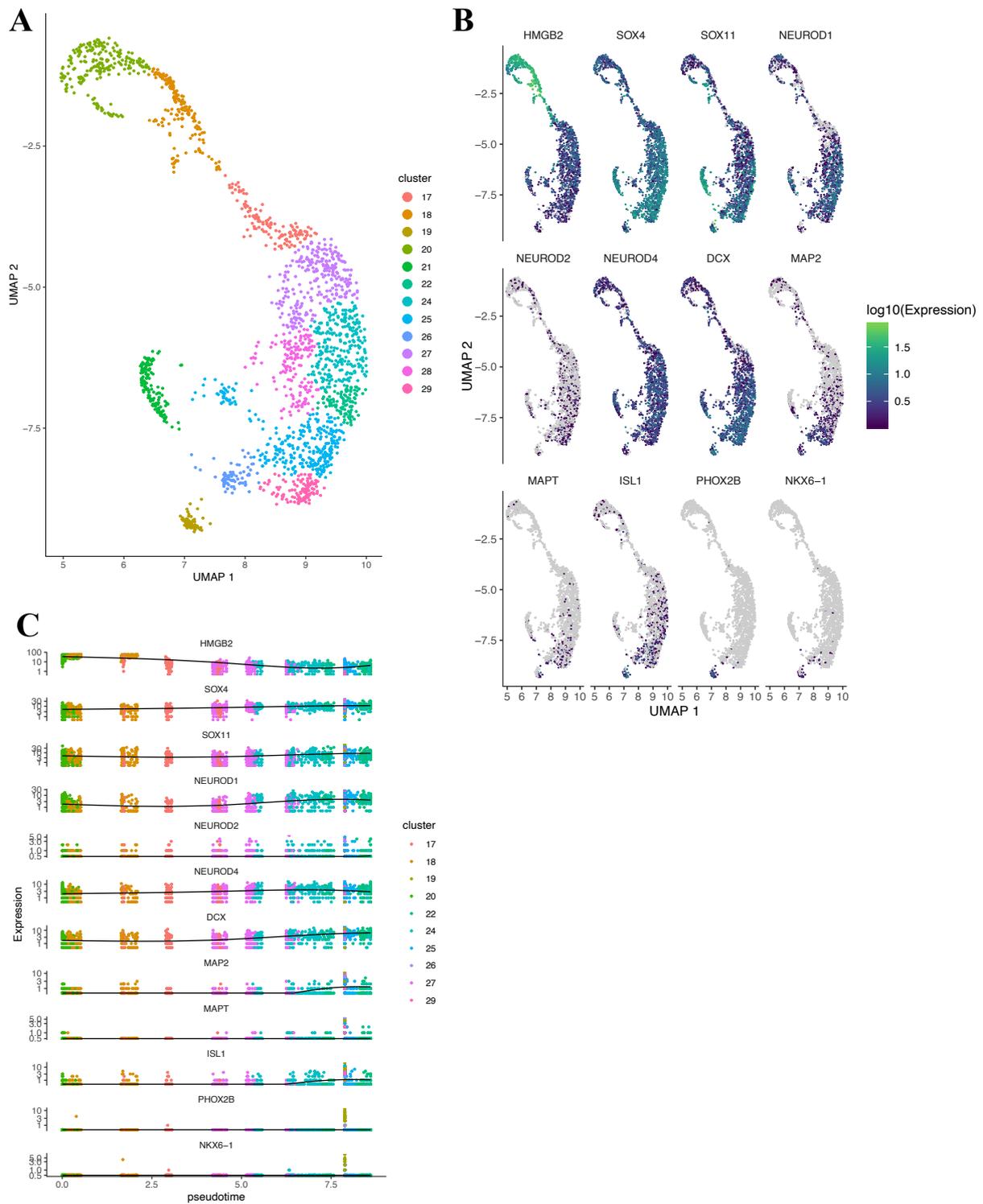


Figure 4. 22: Pseudotime analysis of glutamatergic and cholinergic fate at 3- and 4 days post-induction.

(A) UMAP visualisation of Day 3 and 4 cells only, showing 12 different clusters identified by Monocle. (B) Individual UMAP plots showing expression levels for the cell cycle marker, HMGB2, genes associated with neuronal differentiation (SOX4, SOX11, DCX), glutamatergic fate (NEUROD1, NEUROD2 and NEUROD 4), cholinergic fate (ISL1, and two pan-neuronal

genes, MAP2 and MAPT. (C) Pseudotime analysis of the same genes for all Day 2 clusters, showing the expression levels for each gene in individual cells, arranged in order of increasing pseudotime. Cells are coloured according to their cluster of origin.

20, of the 12 clusters determined using Monocle 3. Here, the population of ISL1+ and PHOX2B+ cells are in cluster 19, which corresponds to cluster 7 from the Scanpy analysis. From the pseudotime analysis, the upregulation of the cholinergic genes is not as obvious as before. Although the analysis shows a slight increase in ISL1 upon downregulation of the cell cycle gene HMGB2, the cells in cluster 19 are squeezed into a single point on the pseudotime path, making the change insignificant to the rest of the population. However, it is visually evident that this cluster of cells highly expressed ISL1, PHOX2B, NKX6-1 and even the marker for mature neurons, MAPT.

I then investigated this early stage of NGN2 reprogramming for glutamatergic neurons. As mentioned earlier, NEUROD genes identified these cells in the population, emerging as early as Day 2 and being enriched in the post-mitotic cells. Therefore, as done before, I carried out a pseudotime analysis of cells at Day 2 (**Fig 4.21**). This showed an increase in cells expressing NEUROD1 and 4 along the pseudotime path, which followed a decrease in cells expressing HMGB2 (**Fig 4.21 C**). As with the analysis of the cholinergic population, a significant number of cells expressed MAP2 later down the pseudotime path, but no significant upregulation of MAPT. Pseudotime analysis of these genes in the Day 3 and Day 4 population, showed a similar trend, with a gradual increase in NEUROD1 and NEUROD4 followed by upregulation of MAP2. I then searched for expression of two genes involved in glutamatergic neurotransmission SLC17A6 (vGLUT2) and SLC17A7 (vGLUT1). The former showed no detectable expression across the population, except for cluster 7, which was also enriched for the cholinergic markers investigated earlier. On the other hand, SLC17A7 was found to be upregulated in all cells except the iPSC population (cluster 3), with the highest expression, again being in cluster 7.

Analysing the late stage, days 14 and 21, followed the same workflow used for the early stage, but also builds on the findings from the early stage. To begin with, a UMAP visualisation of the two timepoints showed a clear separation between them, suggesting significant transcriptional differences (**Fig 4.23A**). However, as seen with the bulk RNAseq dataset, this distinction was mostly driven by the difference in NGN2 overexpression between them, where most likely residual doxycycline in the medium was still driving the overexpression of the NGN2 transgene at Day 14, before being completely depleted at Day 21 (**Fig 4.23B**). In fact, the level of NGN2 expression at Day 14 is similar to the expression levels of the first 24 hours of the reprogramming stage, and only slightly lower than expression levels for Day 2 to Day 4 (**Fig 4.23C**). Despite this discrepancy, Louvain clustering of these timepoints appears to be not entirely affected by this experimental artefact (**Fig 4.23D**). Three of the clusters; clusters 0, 2 and 4 are comprised of both

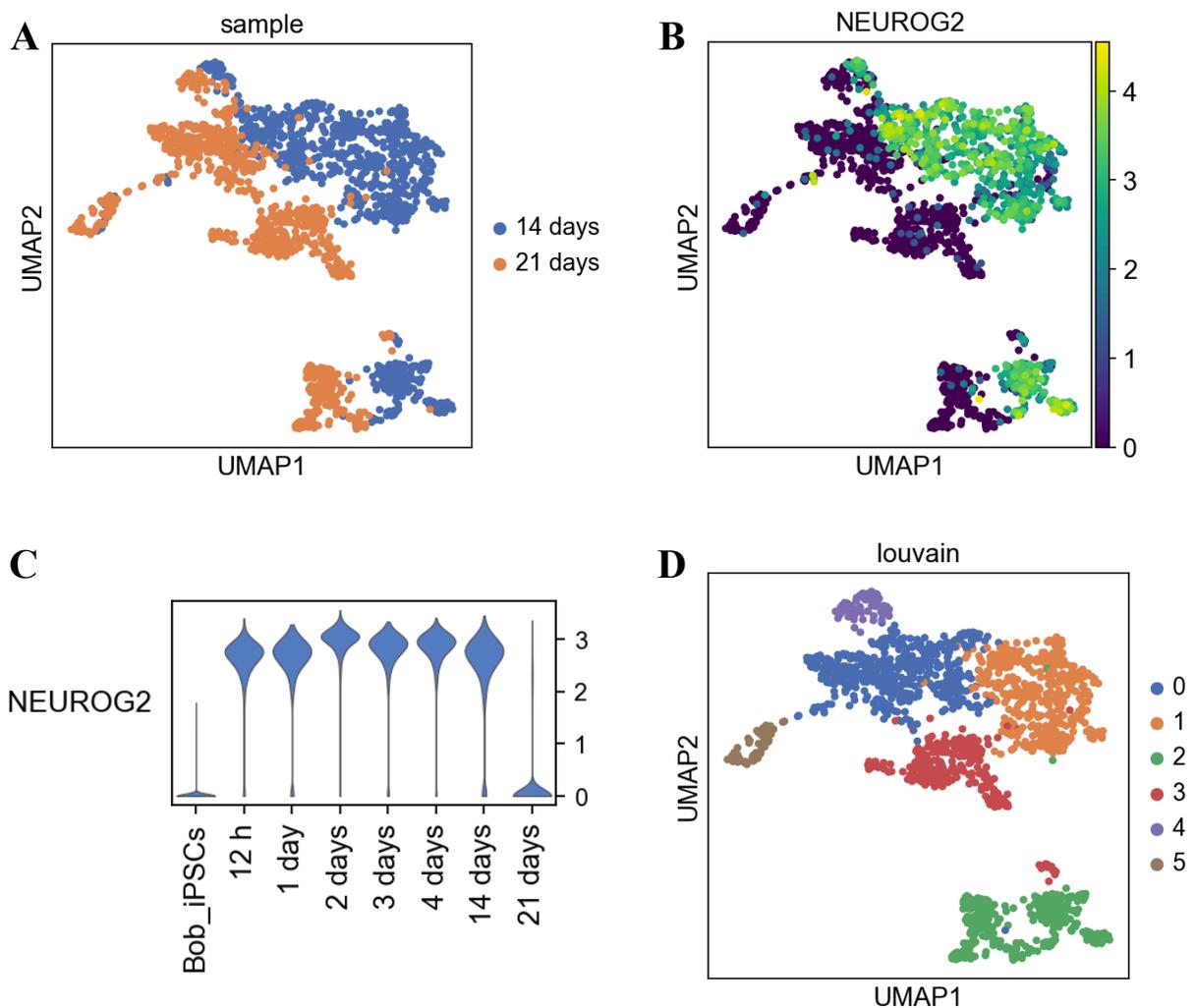


Figure 4. 23: UMAP visualisation of the late stage, NGN2 expression and its Louvain clusters.

(A) The late stage of reprogramming is visualised by considering only 14- and 21 days post-induction. (B) UMAP visualisation of NEUROG2 (NGN2) expression across the late of stage of reprogramming. Expression of NGN2 is homogenously downregulated at 21 days. (C) Violin plot of NGN2 (NEUROG2) expression (from scRNAseq) across all time points (D) Louvain clustering of these time points identified six different clusters.

timepoints, suggesting that their transcriptional profile was not mainly driven by NGN2 overexpression. Next, as a step towards annotating these clusters, I performed a DE analysis (Scanpy Wilcoxon test) to determine the top DE genes in each cluster. This produced familiar genes identified in the early stage analysis, such as ISL1, PHOX2B and NKX6-1 in cluster 2 (**Fig 4.24**).

As seen at the early stage of NGN2 reprogramming, ISL1 and PHOX2B identifies a separate cluster of cells in cluster 2 (**Fig 4.25**). However, in addition to these two genes, this cluster was also enriched for genes associated with cholinergic neurotransmission such as SLC18A3, SLC5A7 and SV2C, suggesting that the neurons have acquired functional properties between day 4 and day 14 of the reprogramming. Intriguingly, ISL1 expression is not unique to this cluster, as it was also expressed in cluster 3 and part of cluster 1, albeit at a lower level. Unlike cluster 2 however, these cells did not express PHOX2B and SLC5A7, but did express SLC18A3 and SV2C, As with ISL1, these genes were expressed at a noticeably lower level compared to cluster 2 (**Fig 4.25C**). Together these group of genes identified two separate population of cells with cholinergic transcription, but with varying levels of each gene.

As for glutamatergic neurons, an overlay of glutamatergic genes demonstrated that they are expressed at varying levels across the population, except for cluster 5 (**Fig 4.26A**). However, cluster 0 and part of cluster 1 had a higher enrichment for glutamatergic genes, especially the AMPA-mediated glutamate receptor GRIA2 and the NMDA-mediated glutamate receptor GRIN2A. NEUROD genes 1 and 2 were also enriched in these group of cells, but mainly in cluster 1 which corresponds to Day 14 neurons (**Fig 4.26C**). Interestingly, the cholinergic population expressing both ISL1 and PHOX2B (cluster 2), had a much lower expression of glutamatergic genes, being almost devoid of some genes such as SV2B and SLC17A7. In contrast, the ISL1 positive but PHOX2B negative cells in cluster 3 and part of cluster 1, had a slightly higher expression of glutamatergic genes like SV2A and SLC17A6, both key components of glutamatergic neurotransmission.

Together, this analysis identified three major neuronal populations (**Fig 4.27**). The first, possibly a *bona fide* cholinergic population, specifically, cranial visceral motor neurons based on the strong expression of PHOX2B, NKX6-1 and NKX6-2. Enrichment analysis of the top 300 DE genes in this population identifies the terms *motor neuron* and *spinal cord* as the top terms, followed by *midbrain* and *sensory neuron*. The second population are glutamatergic neurons based on having higher expression of glutamatergic genes compared to the rest of the population. In addition, enrichment analysis for this population produced the terms *prefrontal cortex* and *cerebral cortex*

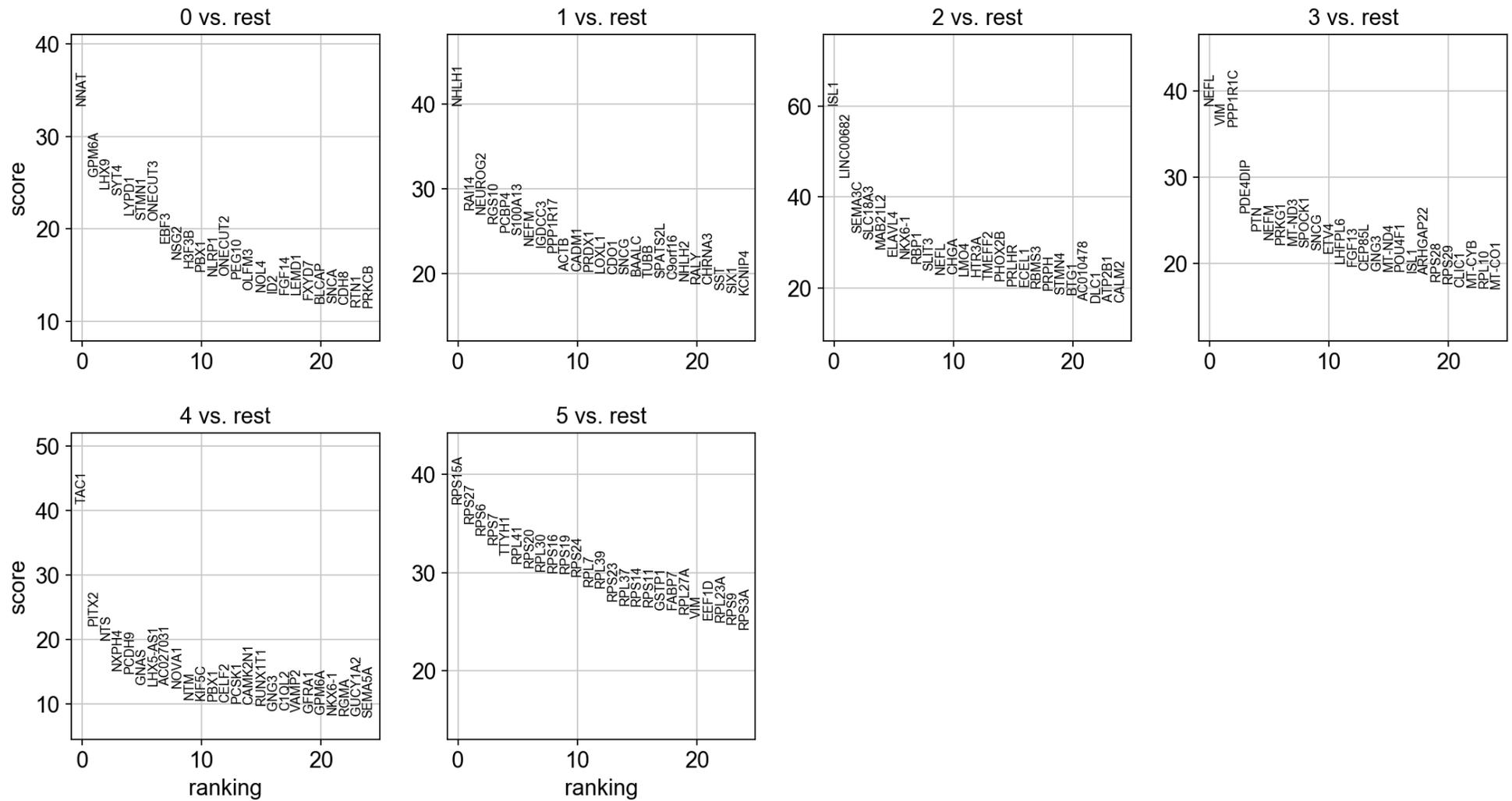


Figure 4. 24: Differential expression analysis of clusters from late stage of NGN2 reprogramming.

The top 25 DE genes are shown for each of the 7 clusters from the early stage of reprogramming, with each gene ranked (x-axis) in order of decreasing test score (y-axis).

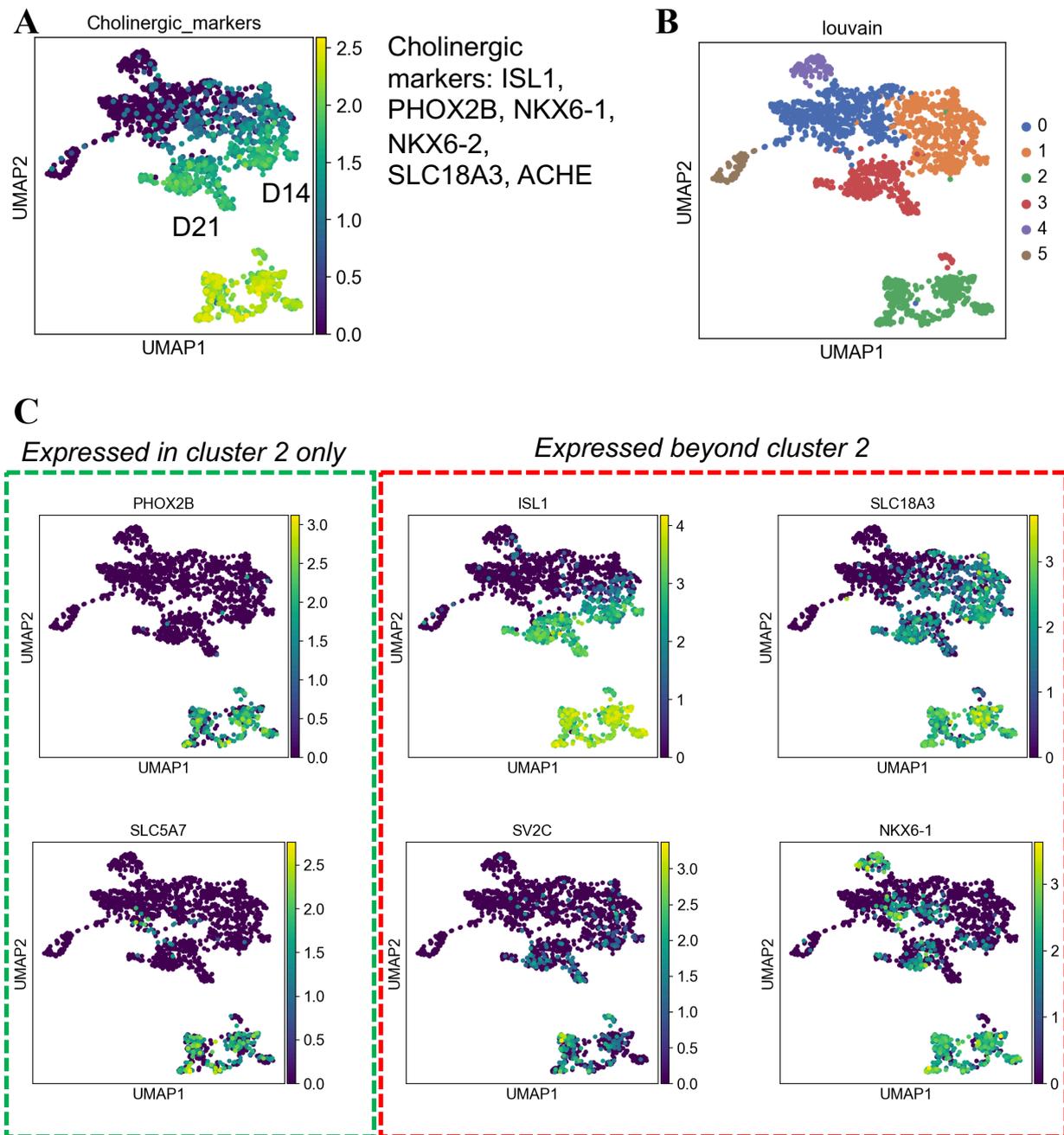


Figure 4.25: Diversity in cholinergic markers at the late stage of NGN2 reprogramming.

(A) UMAP visualisation showing an overlay of cholinergic markers over the late stage of reprogramming. Time points are placed next to their corresponding cluster of cells, and the genes used for the overlay are listed on the right. (B) UMAP visualisation of the Louvain clustering for late stage. (C) UMAP visualisation for the expression levels of cholinergic genes showing the diversity in their expression across the population, with some exclusive to cluster 2 and some that are expressed beyond cluster 2.

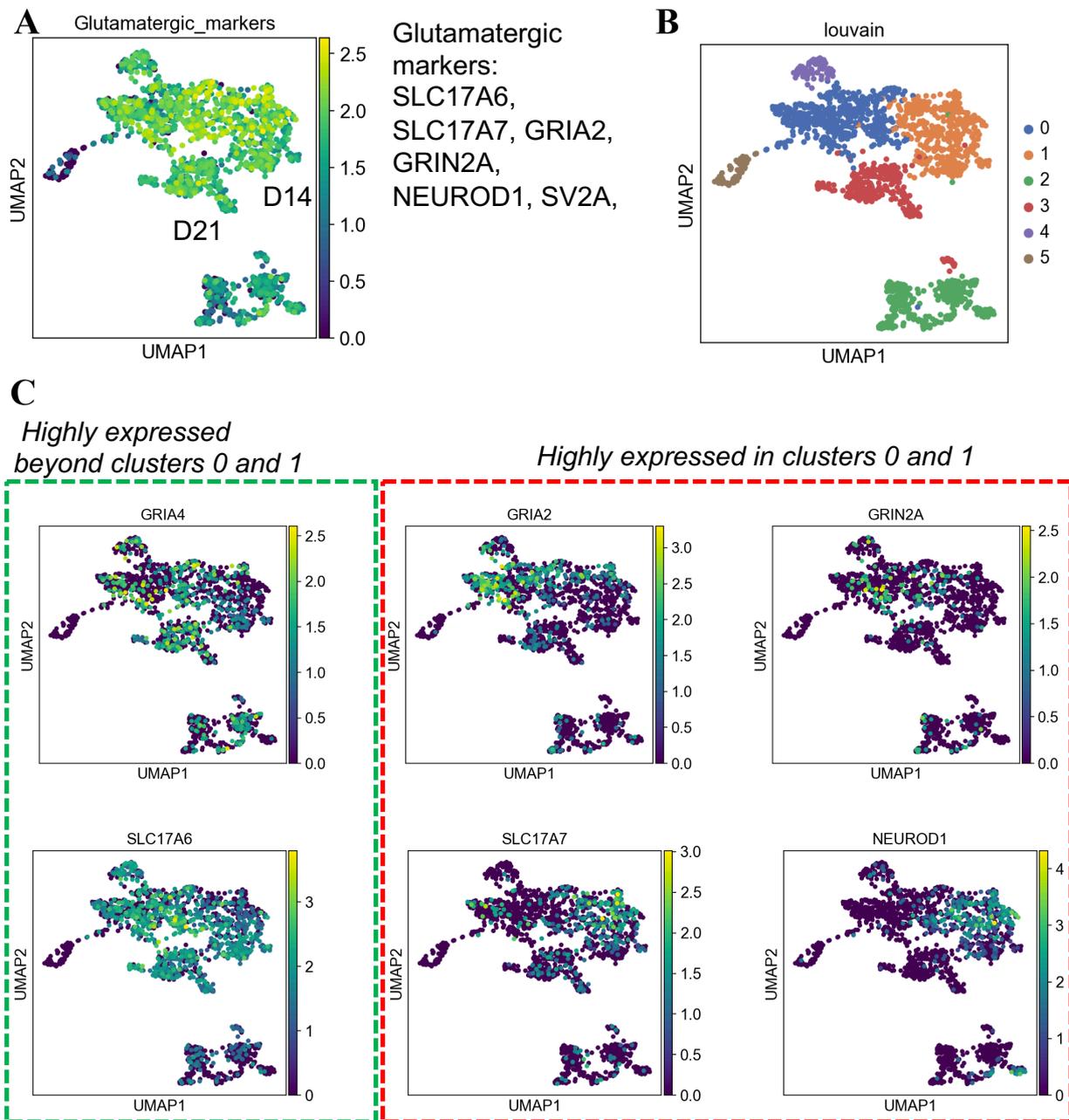


Figure 4. 26: Diversity in glutamatergic markers at the late stage of NGN2 reprogramming. (A) UMAP visualisation showing an overlay of glutamatergic markers over the late stage of reprogramming. Time points are placed next to their corresponding cluster of cells, and the genes used for the overlay are listed on the right. (B) UMAP visualisation of the Louvain clustering for late stage. (C) UMAP visualisation for the expression levels of glutamatergic genes showing the diversity in their expression across the population, with some highly expressed in clusters 0 and 1 and some that are expressed beyond cluster 0 and 1.

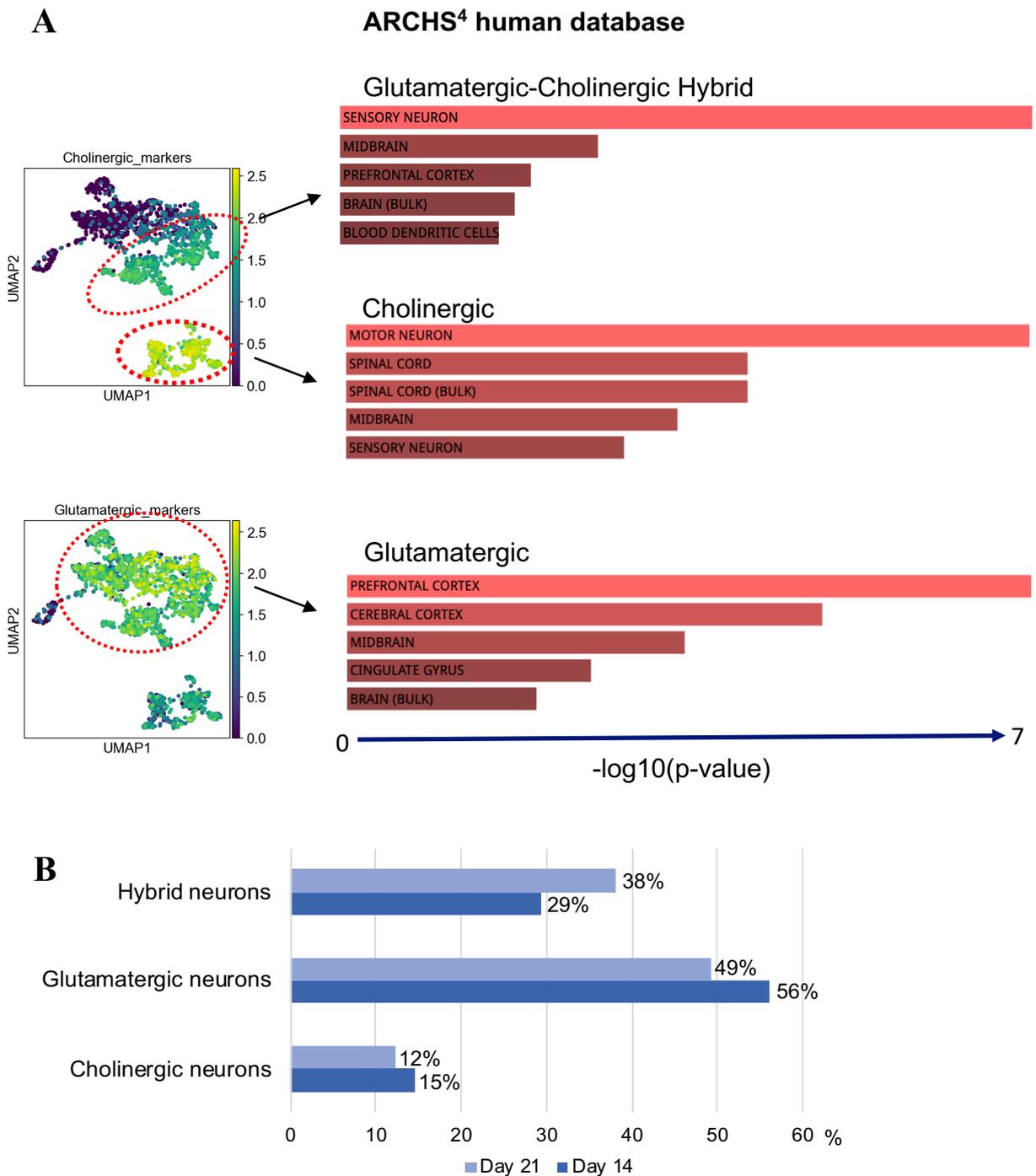


Figure 4. 27: NGN2 reprogramming produces three major types of neurons

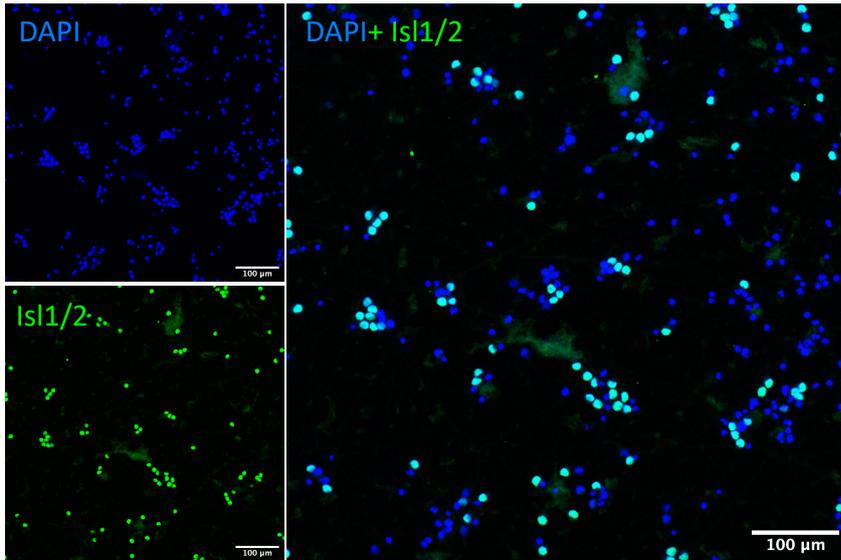
(A) Top 5 terms enriched from the ARCHS4 human database (right) for the top 300 DE genes in the corresponding populations (in red circle) from the UMAP plots (left). The inferred cell type is above each set of terms. (B) Percentage of the three populations at Day 14 and Day 21 of NGN2 reprogramming.

as the top terms, which are usually associated with glutamatergic neurons, followed by the terms *midbrain* and *cingulate gyrus*. The third population, appears to be a hybrid of the first two, expressing varying levels of glutamatergic and cholinergic genes, but with no detectable expression of a key component of cholinergic neurotransmission, SLC5A7. This population also had a noticeably lower expression of ISL1 compared to the *bona fide* cholinergic population. Fittingly, enrichment analysis of the top 300 DE genes in this population identified a mix of terms from the other two populations, including *motor neuron*, which has a lower p-value but is still among the top ten terms.

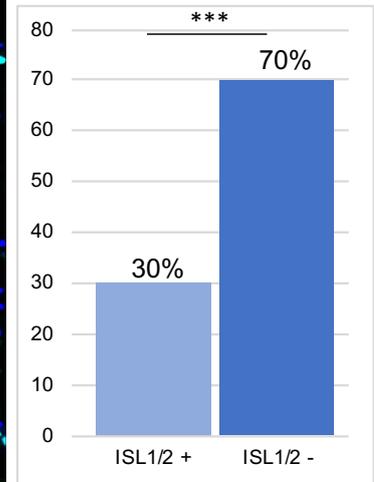
I next sought to validate these findings through immunocytochemistry. At the time, only two cholinergic antibodies were available at hand, ISL1/2 and VACHT (SLC18A3); nevertheless, their use together with an antibody for the glutamatergic marker, VGLUT1, provided some important preliminary results. Firstly, stainings of Day 14 neurons for ISL1/2 and the nuclear counterstain DAPI showed that not all cells were positive for ISL1/2, in agreement with observations from the scRNAseq dataset (**Fig 4.28**). However, their proportion in the population, which has a value of 30%, is lower compared to the proportion determined from the Day 14 scRNAseq sample, which is at 44%. This difference could be due to post-transcriptional regulation of ISL1, which can result in lower levels of the final protein. In addition to ISL1, stainings of the same timepoint with VACHT and VGLUT1, markers of cholinergic and glutamatergic neurotransmission respectively, also revealed positive outcomes, with both markers detected in the cultures. It was not possible to determine the proportion of cells expressing VACHT because unlike VGLUT1, the localisation of VACHT stainings were not consistent, either staining the axons in some cells, or the soma in others. However, the images do indicate colocalization of both proteins in some cells, supporting the finding from the scRNAseq data that there could be a population of glutamatergic-cholinergic hybrid neurons in cultures of NGN2 iNeurons.

Although the analysis largely explains the generation of three different types of neurons, I also investigated other neuronal types, in case they occur in small, rare populations in the dataset. For this, I used known genes for four other major types of neurons: GABAergic, noradrenergic, serotonergic and dopaminergic. However, many of the genes tested were drop-out genes from this batch of samples, making it difficult to draw any conclusions on the presence of these cell types in the culture. As an alternative, I checked for their expression in the bulk RNAseq dataset. As expected, the genes tested were hardly expressed in comparison to cholinergic and glutamatergic genes (**Fig 4.29**).

A I



II



B

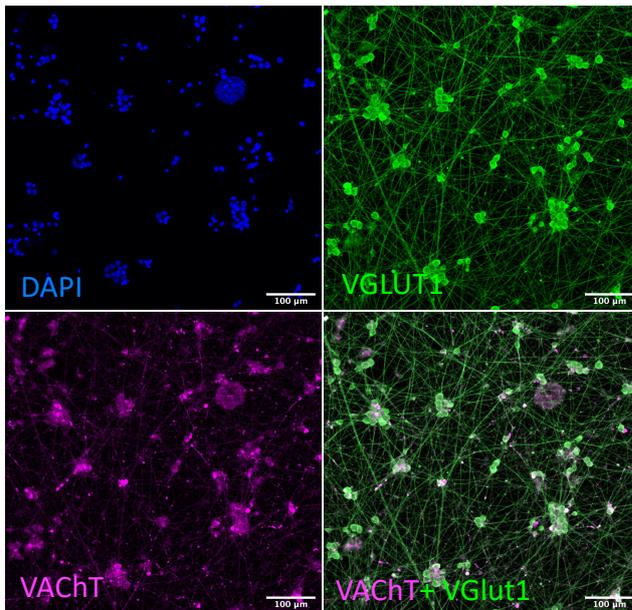


Figure 4. 28: Immunocytochemistry of cholinergic and glutamatergic markers in Day 14 iNs.

(A) (I) Immunostainings for ISL1/2 (green) of iNs at Day 14. (II) Percentage of cells that were either positive or negative for labelling of ISL1/2 ($p < 0.001$). (B) Immunostainings for VGLUT1 (green) and VACHT (magenta) of iNs at Day 14. In both panels, nuclei were counterstained with DAPI (blue) (Scale bars: 100 μ M).

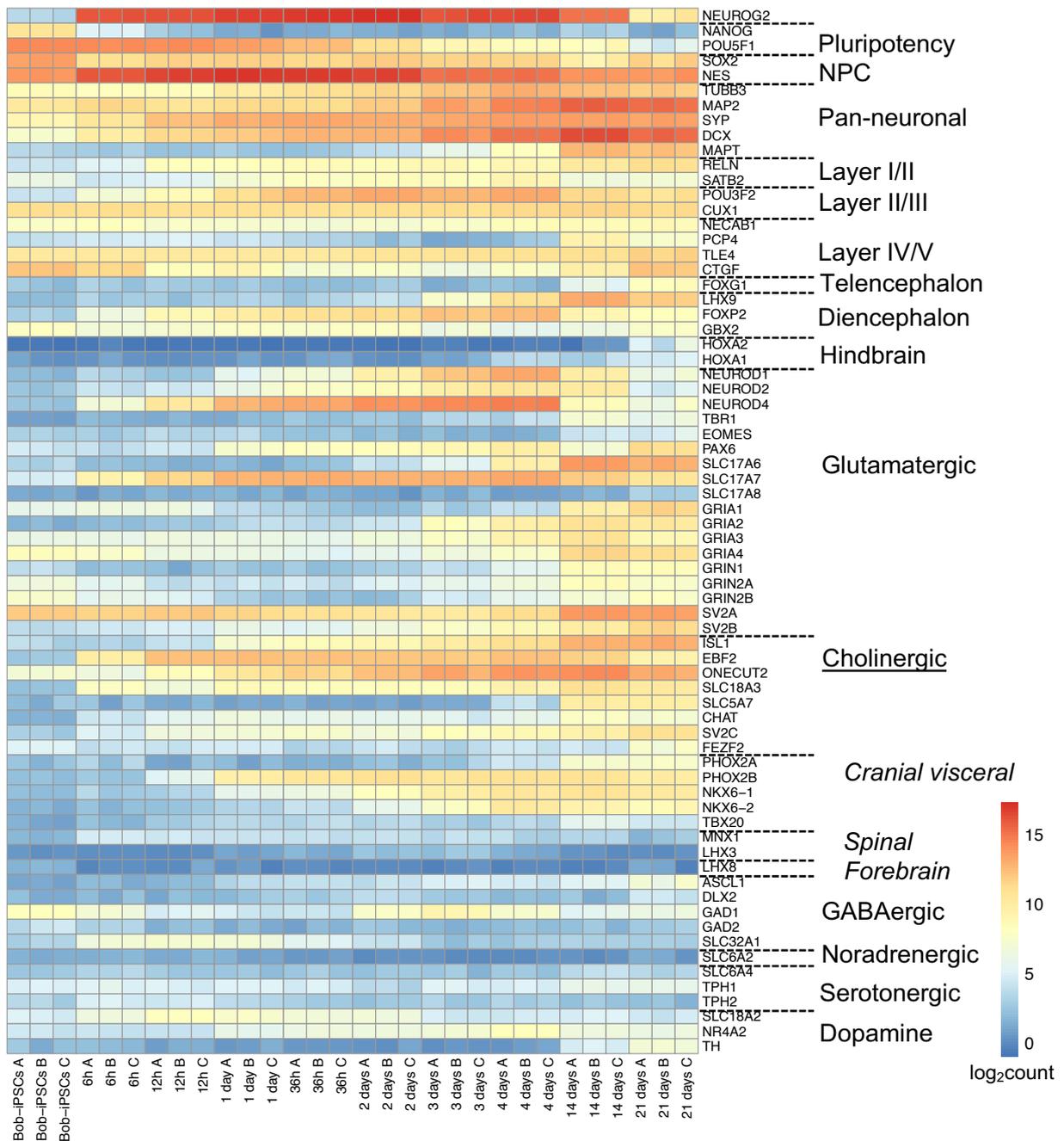


Figure 4. 29: Heatmap of transcriptional expression of different markers from bulk RNAseq data of entire time course of NGN2 reprogramming.

Heatmap of log₂ expression for selected genes, arranged in groups, with their corresponding group label on the right side of the plot. Specific subtypes of cholinergic neurons are labelled in italic. Data is shown for all ten time points from the time course, with 3 biological replicates (A, B and C) for each timepoint.

4.3.5 Effects of glia on transcriptome of NGN2 iNeurons

To get insights into the changes induced by the addition of primary rat glia on the transcriptome of NGN2 iNeurons, I carried out a two-pronged approach, using bulk RNA sequencing and single cell RNA sequencing to compare neurons cultured with glia to neurons cultured without glia. This involved sampling cells at three timepoints from the time course used in the previous section: 4 days post-induction (24 hours after addition of glia), 14 days post-induction and finally 21 days post-induction.

Before considering the effects the glia had on the neurons, it is imperative to know what are the different cell types that made-up the glia used for the co-cultures. Although the glial preparation used for our co-cultures are processed for removal of neurons and non-astrocyte glial cells, it is still possible that these cells persist in the final preparation used for the co-culture. Therefore, I sampled some glial cells from the same batch used for the bulk and single cell experiments and carried out a single cell RNAseq analysis of the glial cells. In the first instance, a UMAP visualisation of this sample shows that there was indeed heterogeneity in the glial preparation, with four different sub-populations detected by Louvain clustering (**Fig 4.30A & B**). As before, I carried out a DE analysis (Scanpy Wilcoxon test) to determine the top DE genes in each cluster as a first step towards uncovering the identity of each cluster (**Fig 4.30C**). Using this list of genes, I was able to determine the cell types in each cluster (**Fig 4.31**). The biggest cluster (0), were mostly astrocytes, based on the expression of *Pea15* and *Gfap*. Cluster 1 identified microglia in the population, whereas cluster 3 identified pericytes. Cluster 2 appears to be made up of two separate sub-populations, and one of these identified ependymal cells based on the expression of *Ccdc153* and *Tmem212*. The other sub-population most likely represents proliferative ependymal cells based on the expression of cell cycle markers and it's clustering with ependymal cells. It is worth pointing out that the ependymal cells also expressed several classic astrocyte markers, albeit at lower levels, such as *Pea15*, *Gfap* and *S100b* (not shown). Since it is likely that some oligodendrocytes could also be in the glial preparation, I tested for two of their characteristic markers, *Olig2* and *Mbp*. This found a small population of cells, just adjacent to the cluster of astrocytes, but not a uniquely identified cluster on its own, probably because of its small size relative to the rest of population. To rule out that there could be any rat neurons in the sample, I tested for 4 different neuronal markers, and although there appears to be strong expression of each gene in some cells, none of them appear to overlap in a single cluster of cells as seen with the other cell markers. Therefore, it is likely that they were random misexpression in the population (**Fig 4.32**). Together, these results indicate that although astrocytes make up the majority of the glial preparation used for the co-culture experiments and there were no contaminating neurons,

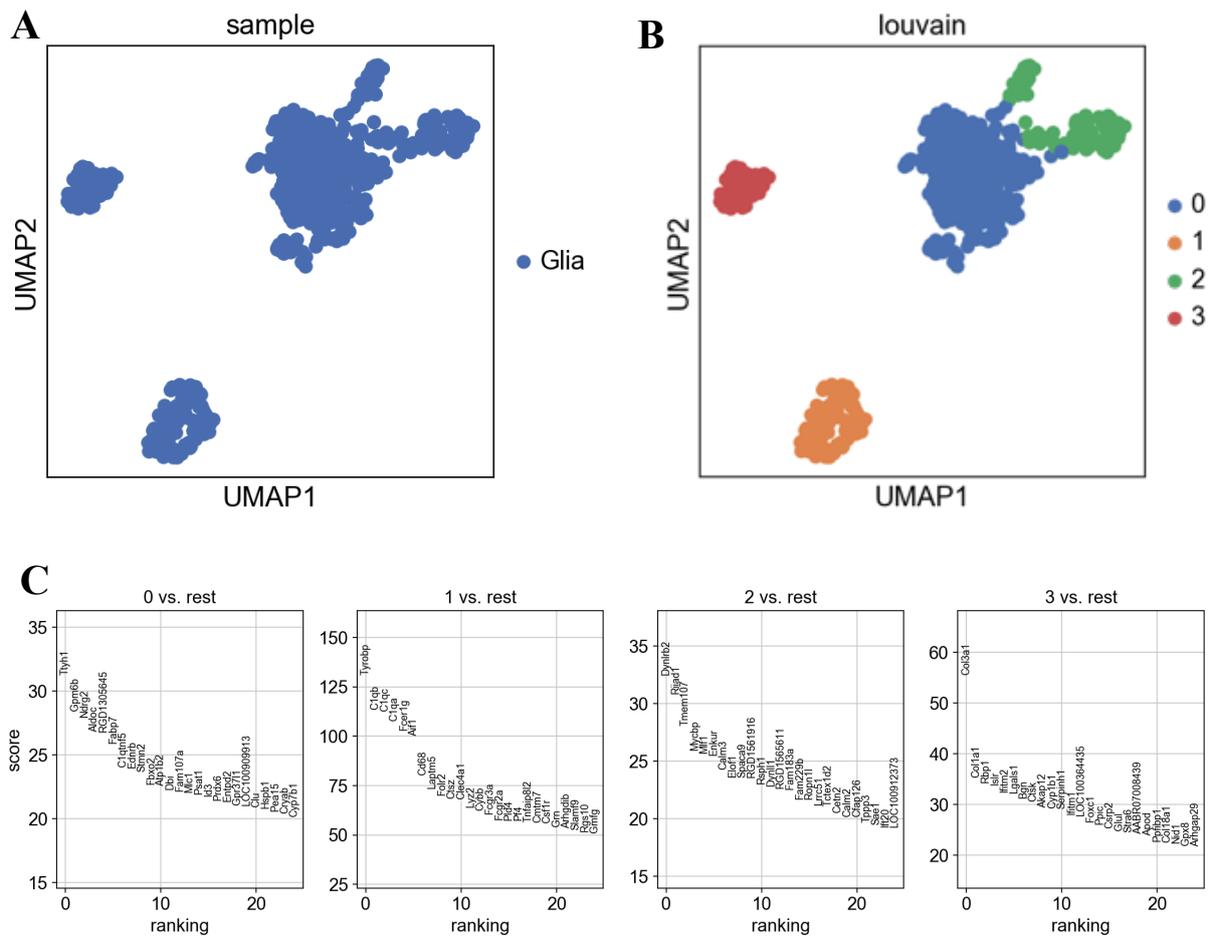


Figure 4.30: UMAP visualisation of primary rat glia and differential expression analysis of its clusters

(A) UMAP visualisation of primary rat glia, showing three main sub-populations. (B) Louvain clustering of rat glia identified four different clusters. (C) The top 25 DE genes are shown for each of the 4 clusters from the rat glia sample, with each gene ranked (x-axis) in order of decreasing test score (y-axis).

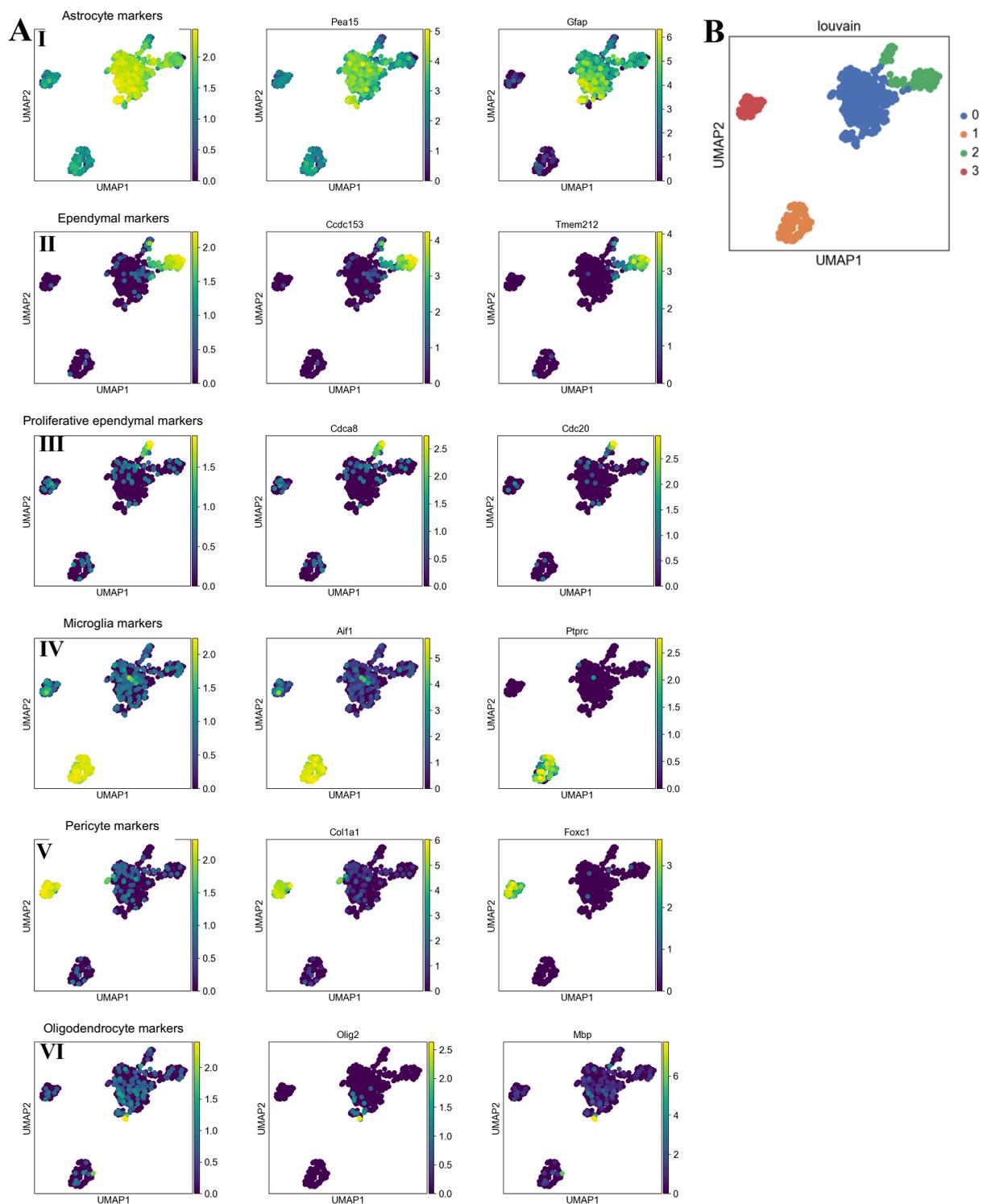


Figure 4.31: Heterogeneity of rat glia

(A) UMAP plots for 6 different population of cells in the sample of rat glia (I – VI). Each row corresponds to a single cell type, with combined overlay of gene expression shown in the left column, and the expression levels for the two genes used for the overlay, in the middle and right columns. (B) Louvain clustering of rat glia.

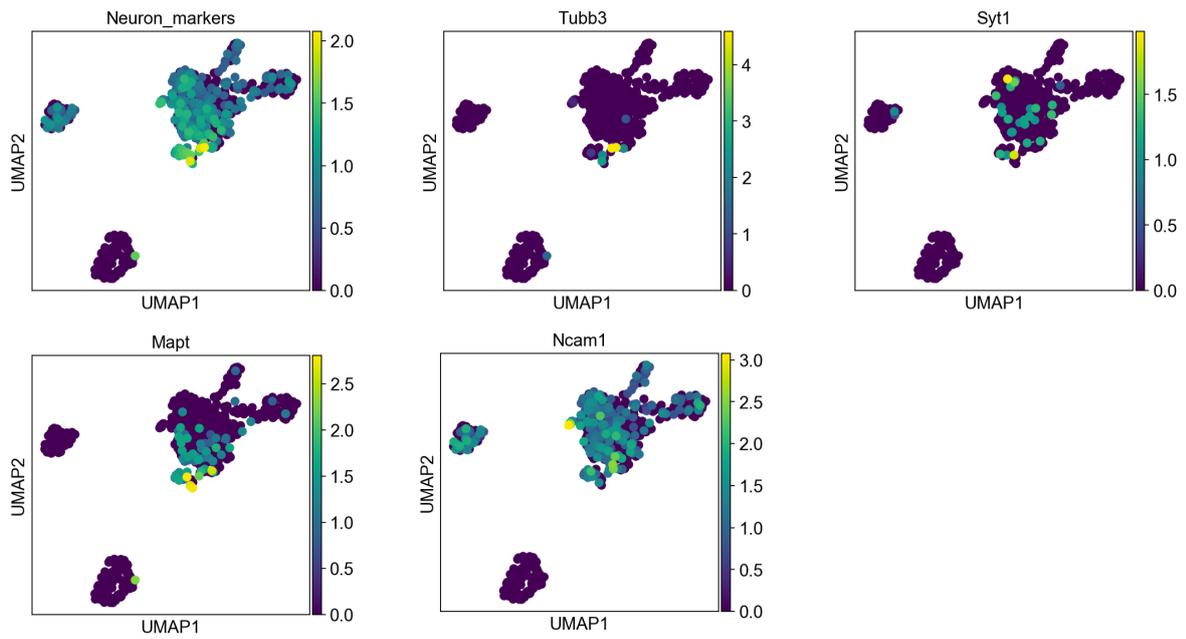


Figure 4. 32: Single-cell analysis of neuronal markers in rat glia.

UMAP visualisation of the rat glia sample showing a combined overlay and their individual expression of four neuronal markers, Tubb3, Syt1, Mapt and Ncam1.

there were other glial cell types in the population; therefore, the following results and discussion will be in the context of rat glia, not just astrocytes.

Bulk RNA sequencing analysis of the three co-culture timepoints utilised the Sargasso bioinformatic tool to disambiguate human cell reads from rat glial reads (Qiu et al., 2018). Then, this allowed the transcriptome of the human cells undergoing NGN2 reprogramming that were co-cultured with rat glia to be compared with their matching timepoint of human cells cultured without glia. This was followed by a DE analysis of the three pair of timepoints, and then an enrichment analysis of the top significant DE genes.

Between the Day 4 neurons with glia (4 days+glia) and Day 4 neurons without glia (4 days), the top terms upregulated in the Day 4 neurons with glia mostly explain broad ontologies such as *nuclear lumen*, *nucleus* and *chromosome* (**Fig 4.33A**). As for Day 4 neurons without glia, the top terms seem to be related to protein and ribosome metabolism. Interestingly, many of the terms for Day 4 neurons without glia were also enriched in 6 hour and 12-hour post-induction samples. Based on this result for the Day 4 comparison, it is challenging to conclude what the differences are between the two conditions; however, it is worth pointing out that there also no differences related to neuronal development between cells at this timepoint. This is also evident in a UMAP analysis of the single cells between the two conditions, where there is large overlap between Day 4 neurons co-cultured with glia and Day 4 neurons cultured without glia (**Fig 4.33B**). It shows that both conditions contained cells progressing out of the cell cycle and into a neuronal development trajectory. Furthermore, there is also a sub-population of the early cholinergic cells in the 4 days+glia sample, demonstrating that this early side population was preserved in the Day 4 co-cultures.

Moving on to Day 14, we begin to see differences associated with neuronal maturation. Although the top terms in the Day 14 neurons co-cultured with glia sample are mostly broad terms, NEUROD1 was ranked as the top DE gene in many of the terms (**Fig 4.34A**). A closer look at other genes enriched in these terms identified other known neuronal development genes such as NEUROD2, NEUROD4, even NEUROG1. At 21 days post-induction, there is a clear enrichment for neuronal maturation terms in neurons co-cultured with glia, with terms associated with axonal growth and postsynaptic function (**Fig 4.34B**). Interestingly, some of the top ranked genes in these terms are also linked to glutamatergic neurotransmission such as GRIN2A, an NMDA-mediated glutamatergic receptor, and CACNG5; the latter being associated with AMPA receptor transport to the postsynaptic compartment (L. Chen et al., 2000). In addition to maturation-associated terms, again, there is also an enrichment for developmental-related terms such as *neuronal differentiation*, *neurogenesis* and *neuronal development*, the latter being a term also enriched at

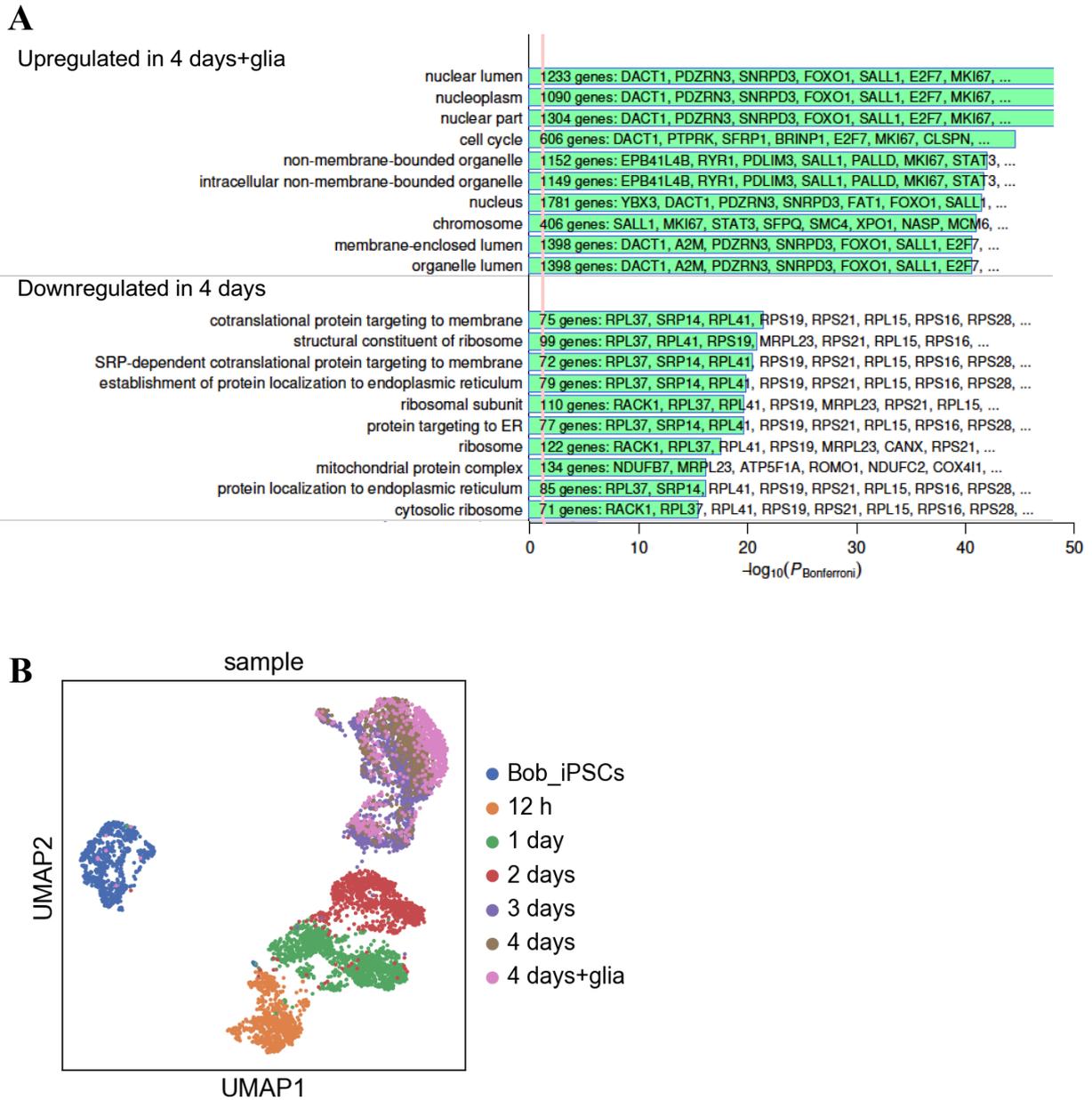


Figure 4. 33: Addition of glia shows modest effect on transcriptome of cells at 4 days post NGN2 induction.

(A) Bulk RNAseq data: Top gene ontologies enriched for significantly upregulated genes in cells cultured with rat glia and the downregulated genes in cells cultured without glia at 4 days post-induction. For each set of DE genes, the plot shows the 10 highest-enriched GO terms. Bars represent $-\log_{10}$ of the Bonferroni-corrected enrichment p-value. The red line represents the threshold $P_{\text{Bonferroni}} = 0.05$. (B) scRNAseq data: UMAP visualisation of the early stage of reprogramming, including Day 4 cells cocultured with glia (4 days+glia).

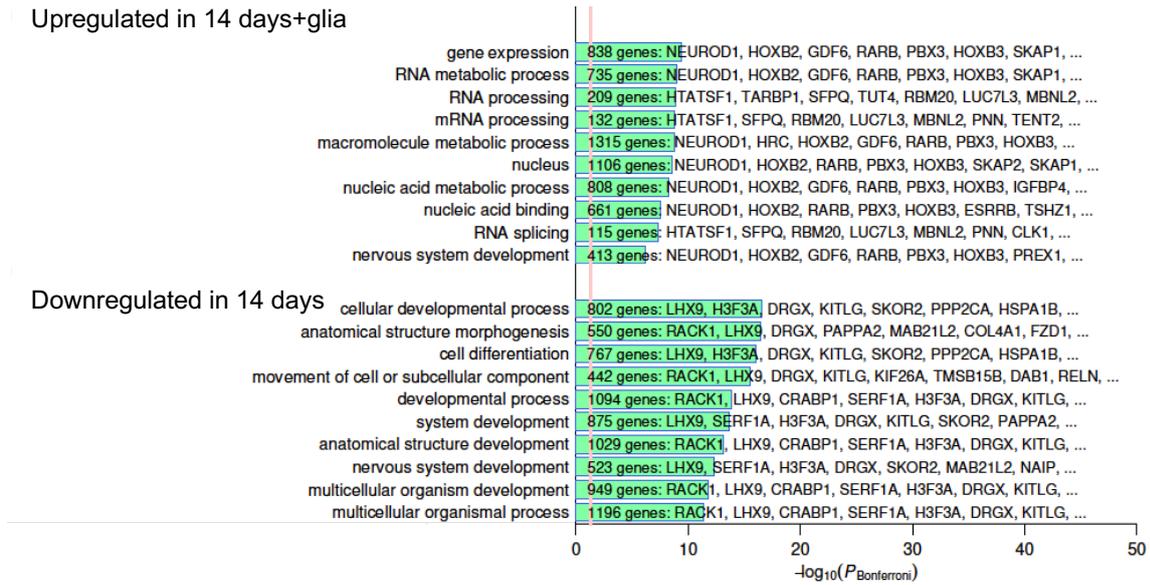
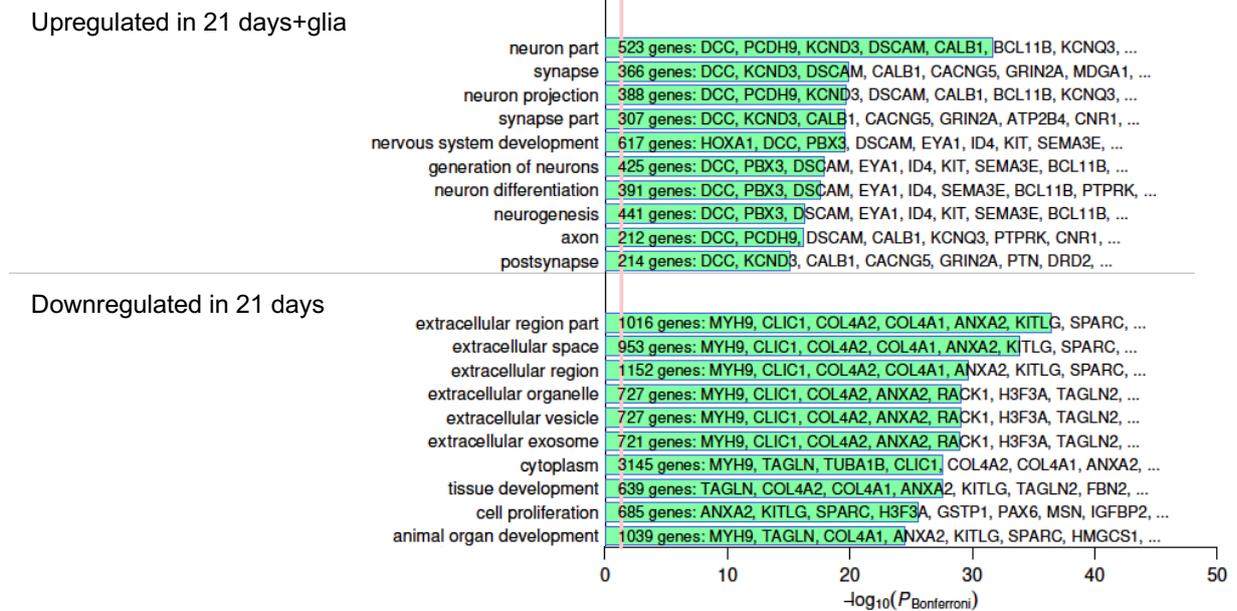
A**B**

Figure 4. 34: Enrichment of neuronal maturation ontologies in NGN2 neurons co-cultured with glia at Day 21

Top gene ontologies enriched for significantly upregulated genes in cells cultured with rat glia and the downregulated genes in cells cultured without glia at (A) 14 days post-induction and (B) 21 days post-induction. For each set of DE genes, the plot shows the 10 highest-enriched GO terms. Bars represent $-\log_{10}$ of the Bonferroni-corrected enrichment p-value. The red line represents the threshold $P_{\text{Bonferroni}} = 0.05$. Data derived from bulk RNA sequencing.

Day 14 neurons co-cultured with glia. Together with the enrichment for genes such as NEUROD1 and NEUROD2 in Day 14 co-cultured neurons, they suggest that co-culturing with glia potentiates neuronal differentiation and development.

I next sought to investigate differences between the two conditions for Day 14 and Day 21 at the single cell level. As a start, a UMAP visualisation of the cells cultured with or without glia at these two timepoints show a near-mirror image of each other (**Fig 4.35A**). A look at the three main cell types identified from before, shows that neurons co-cultured with glia were also comprised of a glutamatergic population, a cranial visceral cholinergic population and a glutamatergic-cholinergic hybrid (**Fig 4.35C & D**). I then performed a Wilcoxon DE test between the neurons cultured with and without glia at the late stage timepoints (**Fig 4.36A**). This revealed a similar picture to the bulk RNAseq comparison. Among the top DE genes here are genes associated with synaptic function such as HOMER1, CAMK2N1, TNFRSF12A and NPTX2 (**Fig 4.36A, B & C**). The latter three genes have been described as neural activity-regulated genes crucial to promoting synaptic function (Yap & Greenberg, 2018). Intrigued by the strong expression of these genes, I interrogated the expression of another synaptic gene that is regulated by neural activity- ARC. Although not amongst the top 30 most DE genes, ARC is highly expressed in neurons co-cultured with glia at both timepoints (**Fig 4.36B & C**). Furthermore, with the exception of NPTX2, this enrichment for synaptic genes in co-cultured neurons occurred in all three different cell type populations – the glutamatergic, cholinergic and the hybrid population. In addition, the activity-dependent transcription factor, EGR1 was also among the highly expressed genes in co-cultured neurons (**Fig 4.36A & B**). A look at pre-synaptic genes such as SYN1, SYP and SNAP25, showed similar levels between both conditions at Day 14 and 21. Besides synaptic genes, the bulk RNAseq comparison also found NEUROD1 as being highly expressed in Day 14 neurons co-cultured with glia compared to Day 14 neurons without glia. This is also recapitulated in the single cell dataset. At Day 14, although both conditions had similar levels of NGN2 expression, NEUROD1, was expressed at a higher level in Day 14 neurons co-cultured with glia, suggesting that this enrichment is independent of transgene overexpression (**Fig 4.36C**).

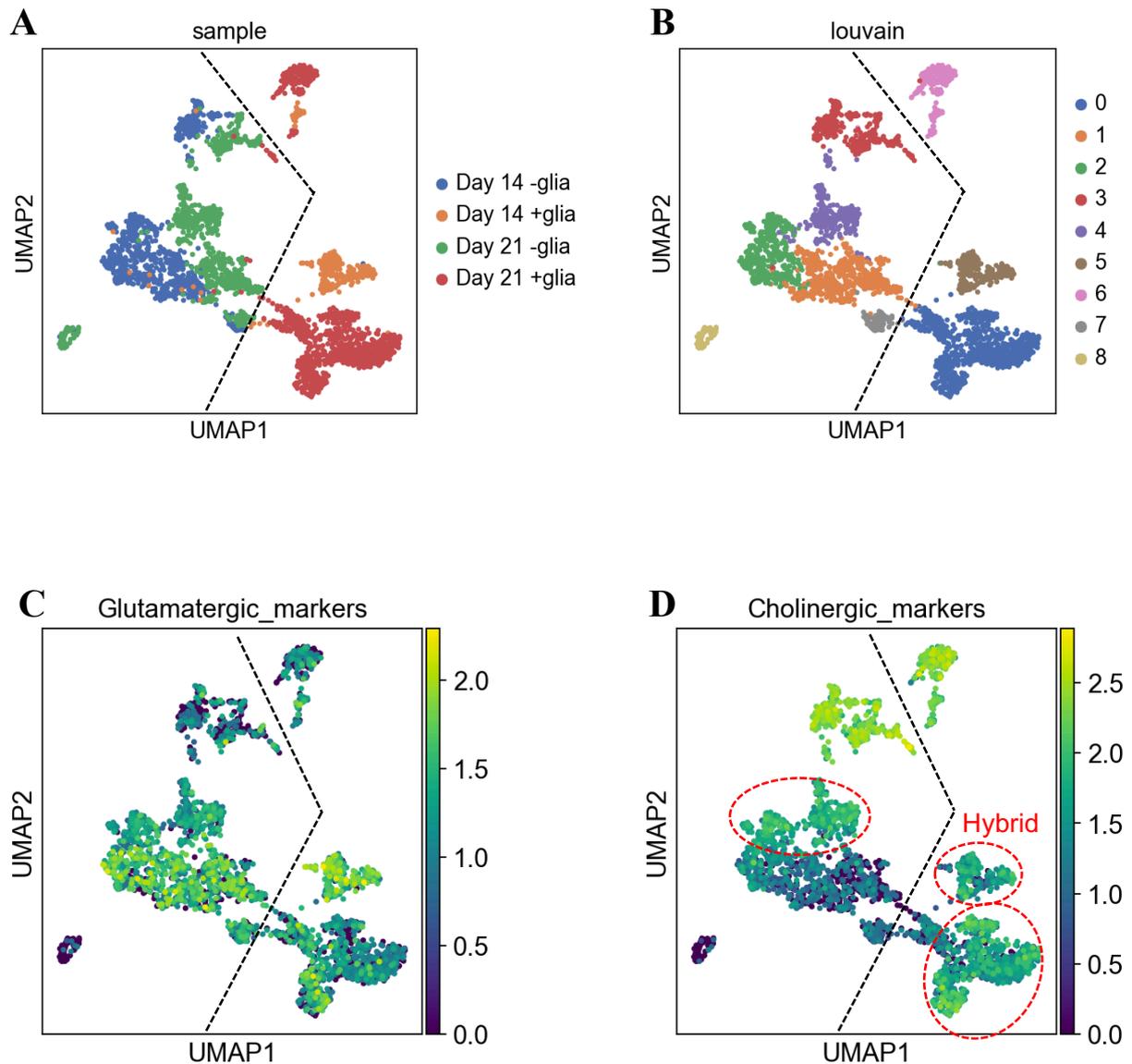


Figure 4. 35: Addition of glia shows no observable effect on lineage specification of NGN2 neurons.

(A) UMAP visualisation of the late stage of reprogramming, including Day 14 and Day 21 cells cocultured with glia (+glia). (B) Louvain clustering of the same samples. UMAP plots showing overlay of (C) glutamatergic genes and (D) cholinergic genes, with population in red circle representing glutamatergic-cholinergic hybrid. Black dotted line in each plot partitions the population into cells cocultured with glia (right) and cell cultured without (left).

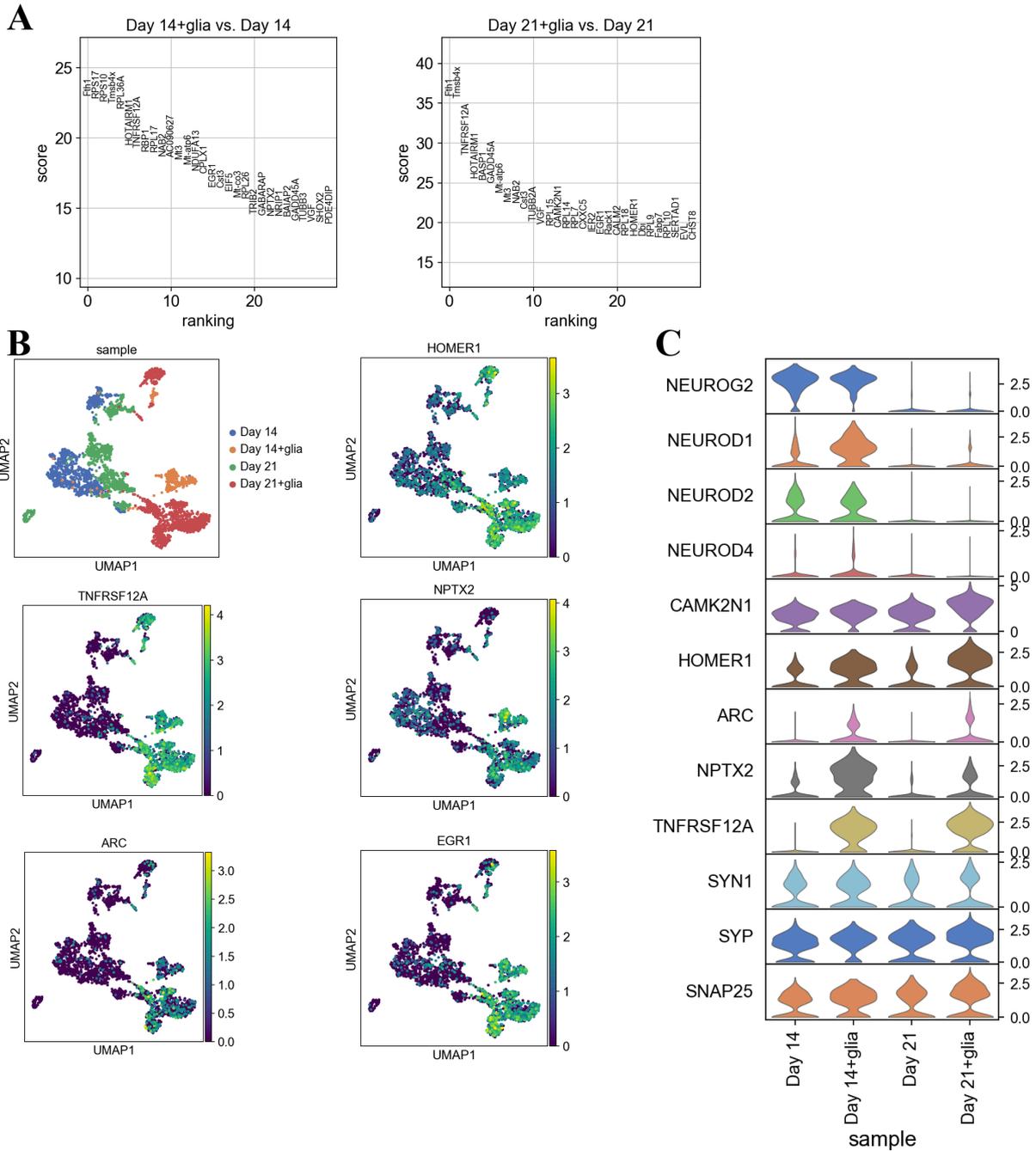


Figure 4.36: Enrichment of synaptic genes in neurons co-cultured with glia at Day 14 and at Day 21.

(A) Differential expression analysis of neurons co-cultured with glia vs neurons cultured without, for Days 14 and 21. The top 30 DE genes are shown, with each gene ranked (x-axis) in order of decreasing test score (y-axis). (B) UMAP plots showing the visualisation for cells co-cultured with glia (+glia) and without glia, and the expression of synaptic genes HOMER1, TNFRSF12A, NPTX2 and ARC, and the activity dependent TF, EGR1. Together, they show enrichment for these genes in the neurons co-cultured with glia. (C) Stacked violin plots for expression levels of genes associated with neuronal differentiation (NEUROG2, NEUROD1, 2 and 4) and synaptic function (remaining genes).

4.4 Discussion

Here, I have established a rich dataset of the transcriptional events that govern the transition of human iPSCs into mature neurons following NGN2 overexpression. By using both bulk and single cell RNA sequencing, I provide insights into how this transcription factor is able to achieve this feat with such speed and high efficiency. Essentially, this analysis uncovered 4 major stages in the time course: 1) the first 12 hours post-induction that is characterised by rapid downregulation of the pluripotency network and repressors of neurogenesis, 2) Day 1 to Day 2 where neuronal network of genes are first upregulated 3) Day 3 to Day 4 where most of the cells become post-mitotic and initiate differentiation into neurons, and finally 4) Day 14 and D21, that sees a downregulation of the pro-differentiation network, and a concomitant increase in maturation genes. Up until now, this entire process was believed to be a highly homogenous occurrence, but findings from the scRNAseq analysis in this chapter, showed that in addition to glutamatergic neurons, our NGN2 iNeurons are made up of two additional types of neurons, visceral motor neurons and neurons with a hybrid profile of cholinergic and glutamatergic transcription.

This reprogramming protocol begins with a dramatic transcriptional change that is marked by rapid downregulation of pluripotency genes within 12 hours, just before the onset of a pro-neural transcriptional network. However, it is important to consider that this effect could also be mediated by the change to a pro-neural medium containing N2 supplement, a well-known formulation for inducing neurogenesis *in vitro* that has been widely used for the past three decades (Bottenstein & Sato, 1979). In addition, it could also be due to the removal of growth factors for PSC maintenance, in particular TGF β and FGF, which are the main pluripotency growth factors in the E8 medium used for our iPS cell culture. The former is a well-known inducer of SMAD-signalling, the inhibition of which is a crucial early step for neural induction in hESCs (Chambers et al., 2009). While these effects are certainly contributing factors, NGN2 is likely the dominant driver of this rapid early transition. Based on the scRNAseq analysis, one of the many genes that is dramatically downregulated by 24 hours is the transcriptional repressor REST, which is known for repressing neurogenesis. In fact, REST inhibition has even been used in combination with neuronal reprogramming cocktails to improve the conversion efficiency of fibroblasts into neurons (Drouin-Ouellet et al., 2017). Recently, it was demonstrated that NGN2 competes with REST for the promoter region of NEUROD4, and therefore initiates neurogenesis when REST is displaced (Masserdotti et al., 2015). In this study, we get a visualisation of this process at the single cell level, showing the upregulation of NEUROD4 in cells at 24 hours post-induction that have also substantially suppressed REST expression. There was a homogenous response despite

heterogeneity in the starting iPSC population. This interplay between NGN2 and REST may be one of the main reasons why NGN2 can successfully reprogram cells at such high efficiency.

As evidenced from both the bulk and single cell RNAseq dataset, NEUROD4 is one of the earliest NEUROD genes expressed during NGN2 reprogramming. This is expected as it is a downstream target of NGN2 in primary neurogenesis and its expression timing and pattern is consistent with a role in driving the transition between proliferation and differentiation (Hardwick & Philpott, 2015). In this chapter, we see a likely mechanism of how this occurs - the increase in NEUROD4 expression at Day 1 was followed by an upregulation in the cell cycle inhibitor CDKN1A at Day 2, which was then followed by mass cell cycle exit, at Day 3 and Day 4 post-induction. Although, a closer look at the single cell transcriptional level revealed that cell cycle exit can occur as early as Day 1 in some cells. These events were then followed by a pro-differentiation program evidenced by the expression of NEUROD1, SOX4 and pan-neuronal genes such as TUBB3 and SYT1. This transition from a cycling state to the onset of neuronal differentiation is analogous to the transition of NSCs to differentiating neurons during development or in conventional differentiation protocols. In fact, NEUROD4 was one of the top hubs identified from the NSC-like module (M.5) in the WGCNA analysis. Based on these findings, I propose a possible window in NGN2 reprogramming for studying early neuronal development, which involves the transition from a proliferative, progenitor-like state, at Day 1 or 2, to early differentiation into immature neurons at Day 3 and Day 4. This brief transition through a progenitor-like stage was also observed in *Ascl1*-mediated reprogramming of fibroblasts into neurons (Treutlein et al., 2016). Recently, hESC-derived NGN2 iNeurons were successfully used to model neurodevelopment neurotoxicity caused by prenatal exposure to valproic acid (Chanda et al., 2019). This landmark study was the first to use NGN2 iNeurons to model a neurodevelopmental disorder, however, their analysis was performed on cells from Day 4 onwards, which would have missed early neuronal commitment and formation of immature neurons, which were documented in this chapter.

Besides capturing these events in neuronal development, my analysis also showed that NGN2 reprogramming of hiPSCs in basic neuronal culture conditions, also generated cholinergic neurons and cholinergic-glutamatergic hybrids. The side-occurrence of a cholinergic fate is not surprising given that NGN2 is known to cooperate with other TFs like ISL1 and LHX3 in production of motor neurons during development (Lee et al., 2005). In fact, both cholinergic and hybrid types have been associated with NGN2 reprogramming before, but in very different ways. The addition of dorsomorphin and forskolin proved to be instrumental in specifying cholinergic neurons from human skin fibroblasts, however, it wasn't clear which subtype of motor neurons they belong to (M.-L. Liu et al., 2013; Smith et al., 2016). In a more similar cell culture condition

but with co-expression of NGN1, Busskamp and colleagues showed that almost 100% of their iNeurons tested positive for VGLUT1 immunostaining while 98% tested positive for the cholinergic marker CHAT, concluding that their protocol produces entirely hybrid neurons, although it wasn't mentioned at what timepoint the analysis was carried out (Busskamp et al., 2014). On the other hand, my analysis found the occurrence of hybrid neurons to be around 30% at Day 14 based on the transcriptional expression of ISL1 alone. CHAT was not detected in the scRNAseq analysis but based on the expression of other important cholinergic markers such as VACHT (SLC18A3) and SLC5A7, it is unlikely that our cultures are made up of such a high proportion of cholinergic hybrids. In addition to this, my analysis identified a specific propensity for visceral motor neurons over other subtypes based on the expression of PHOX2B and absence of other subtype markers. Recently, a comparison of *Ascl1* and *Ngn2* at 48 hours post induction in mouse ESCs, found *Phox2b* to be enriched in *Ascl1* reprogrammed cells, with only a few in the *Ngn2* population (Aydin et al., 2019). Although this was carried out in mouse cells, based on the results from this chapter, 48 hours may be too early to estimate their final proportion in a fully differentiated population.

An important feature of this dataset is that it captured the point at which the cholinergic and glutamatergic fates emerged during the early stage of NGN2 reprogramming. However, trajectory and pseudotime reconstruction of their paths were not successful in delineating the bifurcation of the resulting cell types. Pseudotime analysis of cholinergic and glutamatergic TFs such as ISL1 and NEUROD1, respectively, identified cells expressing factors as early as Day 2, just after the cells exited cell cycle. However, there wasn't a clear transcriptional profile of the trajectory leading up to this lineage bifurcation. Even at Day 3 and Day 4, where there was a distinct subpopulation of cholinergic cells identified by Louvain clustering, there were no cells bridging the main population to this cluster, making it difficult to infer a trajectory. There can be several reasons for this limitation in the dataset. One explanation could be that the emergence of the cholinergic population is not entirely driven by transcriptional regulation, rather by epigenetic mechanisms instead. If so, the bulk ATAC sequencing detailed in the following chapter may not be able to capture this brief and rare process. Alternatively, it is also possible that the single cell dataset was not rich enough to distinguish between the emerging population due to a lack of sequencing depth or dropout events, which can occur due to technical limitations of the procedure itself. Besides this, it could also be due to a bioinformatic limitation. Although both Scanpy and Monocle are currently widely used for single cell transcriptomics, both utilise a similar approach for determining transcriptional similarity between cells, which is by Euclidean distance (Trapnell et al., 2014a; Wolf et al., 2018). Additionally, improved methods for batch-correction and data

integration may also solve this problem. There is great ongoing effort to improve methods for analysing such datasets, which in future may allow us to get a better understanding of how these different cell fates arise in NGN2 reprogramming.

Co-culturing NGN2 iNeurons with glia has been commonly adopted as a way for promoting its maturation and electrophysiological function, as demonstrated with our cultures in the previous chapter. Here, I have both bulk and single-cell transcriptional evidence that provide insights into how glia supports these crucial processes. At both time points, genes associated with synaptic function are highly enriched in neurons co-cultured with glia. Pre-synaptic genes such as SYN1, SYP and SNAP25 tend to be similarly expressed between both conditions. Instead, there seems to be a specific enrichment for activity-dependent genes in the postsynaptic compartment, with most of them showing an increasing trend from Day 14 to Day 21. Amongst this is HOMER1, a well-known activity-dependent scaffold protein which forms part of the postsynaptic density in excitatory neurons (Brakeman et al., 1997; Xiao et al., 1998). Other genes in this list - CAMKIIN, ARC, and NPTX2 - have been discovered over the past two decades (Chang et al., 2010; Chowdhury et al., 2006), but TNFRSF12A, has only recently been described (Cheadle et al., 2018). Therefore, this dataset not only provides an avenue for validating new and long-standing findings of genes crucial to synaptic activity but could also potentially uncover novel genes involved in this process. In accordance with these activity-regulated genes, another highly expressed gene in co-cultured neuros is EGR1, an immediate early gene (IEG) TF that has been shown to be an integral part of processes underlying neuronal activity such as neurotransmission, synaptic plasticity and higher order processes such as learning and memory (Duclot & Kabbaj, 2017). In fact, EGR1 binds to the promoter of ARC *in vivo* following synaptic activation, to trigger its transcription (Li et al., 2005).

The enrichment of these activity-regulated genes are likely in response to the enhanced neuronal activity when NGN2 iNs are co-cultured with glia; however, it is difficult to discern if they are also triggered by glial-derived (mainly astrocyte) cues, which are a combination of contact-mediated and secreted cues (Allen & Eroglu, 2017). The enrichment of these genes provide transcriptional evidence for the network of genes necessary for maintaining functional activity, but they don't necessarily constitute a likely mechanism by which the addition of glia promotes enhanced neuronal activity. One likely mechanism suggested by this data is that there is a potentiation of neuronal development in neurons co-cultured with glia. Ontologies such as *neuronal differentiation* and *neurogenesis* are enhanced in co-cultured neurons, particularly at Day 21. In fact, NEUROD1 was found to be higher in Day 14 co-cultured neurons, and this seemed to be independent of NGN2 expression levels. Classically viewed as a promoter of

neuronal fate specification, differentiation and migration in the developing neocortex (Hodge, Kahoud, & Hevner, 2012), NEUROD1 has also been shown to play a role in differentiated, functional neurons (Aprea, Nonaka-Kinoshita, & Calegari, 2014; Boutin et al., 2010), but the evidence remains limited. The findings in the present study seem to support a role in this late stage of differentiation, but a further investigation is needed. One important question to explore is if sustained expression of NEUROD1 in this stage of the iNeuron protocol could promote some of the functional properties observed when iNs are co-cultured with glia. Towards this, the collection of synaptic genes uncovered in this study can serve as a guide not only for investigating NEUROD1, but other candidate factors that could promote electrophysiological function. In the following chapter, such candidate factors are identified from an ATAC-seq analysis of the same time points.

5 Epigenomic profile of NGN2 reprogramming

5.1 Introduction

Cellular identity is associated with the transcriptional profile of individual cells. In addition, epigenetic signatures have been linked to cellular identity. A number of studies demonstrated that cells undergo significant remodelling of the epigenetic landscape during cellular reprogramming (Aydin et al., 2019; Chronis et al., 2017; Wapinski et al., 2017). An important aspect of my project therefore was to study epigenetic changes that are associated with NGN2 reprogramming of iPSC to human iNeurons.

Numerous genomic assays have been used to investigate chromatin dynamics and organisation such as DNase-seq, MNase-seq, and ChIP of histone modifications. A recent method called ATAC-seq (Assay for transposase-accessible chromatin) is rapidly gaining popularity due to its high sensitivity, requiring less starting material, and enabling mapping of chromatin even in single cells (Buenrostro et al., 2013; Buenrostro et al., 2015; Chen, Miragaia, Natarajan, & Teichmann, 2018). ATAC-seq uses a hyperactive Tn5 transposase, loaded with adaptors that can simultaneously fragment and tag a genome with the sequencing adaptors, a process called tagmentation. Since transposons integrate into active regulatory elements, this method allows to interrogate regions of accessible chromatin.

Apart from identifying regions that gain or lose accessibility, ATAC-seq also allows to investigate nucleosome organisation, specifically, nucleosome positioning and occupancy. Nucleosome positioning is defined as the probability of a nucleosome reference point (typically, a dyad) being at a specific genomic coordinate relative to surrounding coordinates; whereas, nucleosome occupancy is defined as the probability of nucleosomes being present over a specific genomic region within a population of cells (Lai & Pugh, 2017). Dynamics of nucleosome positioning and occupancy play a key role in transcription. Transcriptionally active genes possess a nucleosome-free region (NFR) at the transcriptional start site (TSS), allowing the site to be accessible to chromatin regulators, as well as transcription and replication machineries (Jiang & Pugh, 2009). Surrounding this region, are the +1 and -1 nucleosomes, which are highly regulated and well-positioned nucleosomes, residing at a canonical distance downstream (in the direction of transcription) and upstream of the TSS, respectively (Lai & Pugh, 2017). Nucleosome positioning and occupancy can be extrapolated from ATAC-seq data on the basis of fragments of sequences that are associated with nucleosomes showing periodicity of approximately 150-200bp, with single nucleosomes generating approximately 200bp fragments followed by 400- and 600-bp

fragments generated by di- and trinucleosomes, respectively (Buenrostro et al., 2013; Schep et al., 2015). Smaller than 150-bp reads account for nucleosomal-free DNA (NFR).

In addition, when considered in the context of known TF sequence motifs, ATAC-seq can also be used to infer TF binding profiles. Two recent bioinformatic tools that have demonstrated this utility of ATACseq datasets are Differential ATAC-seq toolkit (DASTk) and Bivariate Genomic Footprinting (BaGFoot) (Baek, Goldstein, & Hager, 2017; Tripodi et al., 2018). Although both techniques can assess changes in TF activity induced by a perturbation, the algorithm used by DASTk corrects for localised sequence bias observed at promoters and enhancers, thus providing results with higher confidence and fewer false positives. Taken together, the various analyses that an ATACseq dataset offers makes it a useful approach for this study.

NGN2's ability to efficiently reprogram human PSCs may be associated with an ability to function as a pioneer transcription factor. Pioneer factors can engage silent, unmarked chromatin and initiate the recruitment of other factors to activate genes specific to a new fate (Iwafuchi-Doi & Zaret, 2016). However, there is some debate as to whether NGN2 can be classed as a pioneer factor. Reports of successful reprogramming with NGN2 alone have come either from PSCs or neural lineages such as NSCs and cortical astrocytes (refer to Table 1). In contrast, several studies have indicated that NGN2 cannot reprogram human and mouse fibroblasts efficiently, as for example compared to *Ascl1* (Chanda et al., 2014; M.-L. Liu et al., 2013). This poor efficiency in reprogramming fibroblasts can be improved dramatically when NGN2 is expressed in the presence of other TFs and/or small molecules, although they do not result in neurons with glutamatergic function (Blanchard et al., 2015; M.-L. Liu et al., 2013; X. Liu et al., 2012; Smith et al., 2016; Son et al., 2011). One of these additional factors are the small molecules forskolin and dorsomorphin (FD), whose addition mediate a cholinergic neuronal phenotype. ATACseq analysis investigating NGN2 and NGN2 plus FD reprogramming revealed that the addition of FD noticeably improves accessibility of NGN2 to neurogenesis-related factors (Smith et al., 2016). Although the intensity of the accessibility signal for NGN2 at these sites is much lower in the absence of FD, there is still an increase over the control (without NGN2 or FD). Therefore, the authors concluded that NGN2 is a pioneer factor, but requires the addition of FD to improve accessibility at pro-neural sites. Regardless of whether or not NGN2 should be classified as a pioneer factor, these findings imply that NGN2's ability to reprogram PSCs, NSCs or cortical astrocytes may be due to its target sites being more accessible as compared to non-neural somatic cells such as fibroblasts.

So far, few studies have investigated the epigenetic changes associated with NGN2 reprogramming (Aydin et al., 2019; Smith et al., 2016; van der Raadt et al., 2019; Velasco et al., 2017); however, none have studied this process in human PSCs and in the absence of small molecules. More importantly, this has never been investigated in a single time course that encompasses the early, rapid onset of changes that lead to neurogenesis and subsequently maturation and gain of function.

5.2 Aims and experimental design

Where the previous chapter elucidated the major transcriptional changes during NGN2 reprogramming, the main aim of this chapter is to generate a more wholistic understanding of this process by incorporating the epigenomic landscape. To address this question, I performed an ATAC-seq analysis of NGN2 reprogramming using the same time course designed for transcriptional profiling in Chapter 4 (Figure 5.1). In addition, I performed ChIP-sequencing of NGN2 binding at Day 1 post-induction. Using this approach, I sought to address the following questions:

- 1) Which sites gain and lose accessibility during NGN2 reprogramming and when do these changes occur? Are these direct or indirect targets of NGN2?
- 2) The global dynamics in nucleosome occupancy and positioning.
- 3) Which are the transcription factor binding sites that open up/close down following the time course of NGN2 programming? Which of these are direct versus indirect targets of NGN2?
- 4) How does co-culture of glia affect chromatin remodelling at the late, maturation stage, including transcription factor binding sites? Which of those are direct or indirect targets of NGN2?

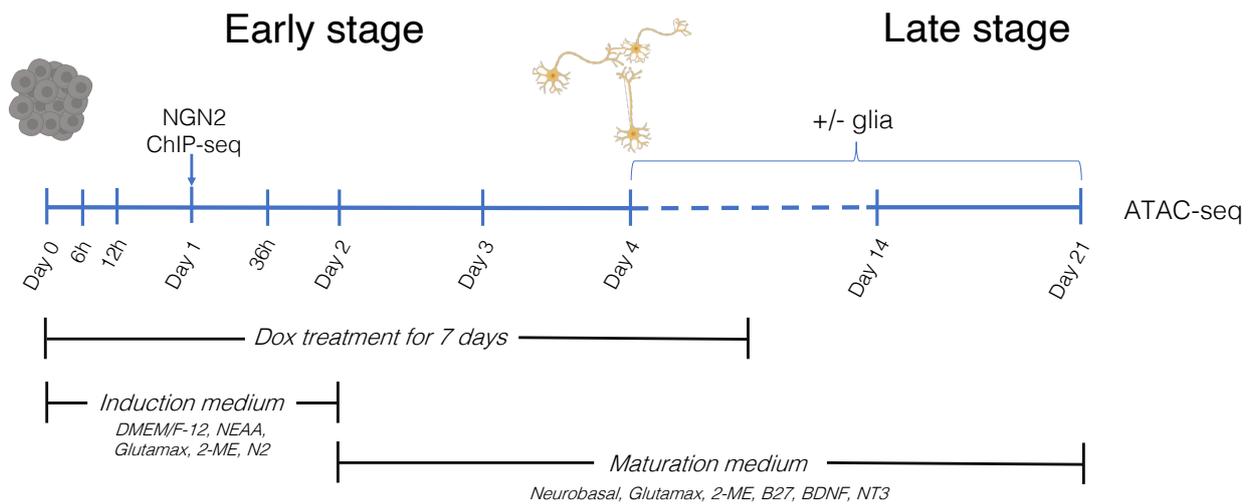


Figure 5. 1: Time course used for bulk ATAC-seq and NGN2 ChIP-seq of NGN2 reprogramming.

For bulk ATAC-seq analysis, the same time course for transcriptional profiling was used, where there were ten time points grouped in two stages. For the early (reprogramming) stage, biological duplicates were sampled between Day 0, made up of iPSCs, to Day 4 post-NGN2 induction, at which most cells have formed neuronal-like features. For the late (maturation) stage, cells were sampled at Day 14 and 21. To investigate the effects of rat-derived glia, cells co-cultured with glia were sampled at Day 4 (24 hours after co-culture with glia), Day 14 and Day 21. Biological duplicates of Day 1 cells were also sampled for NGN2 ChIP-seq. Cells were kept in induction medium for the first 2 days of reprogramming and then switched to maturation medium for the remainder of the protocol. Dox was added to the culture medium for the first 7 days of induction.

5.3 Results

5.3.1 PCA analysis of ATAC-seq data recapitulates bulk RNA-seq data

To visualise the bulk ATAC-Seq data generated and to assess their quality, I carried out a PCA. Over the time course, the PCA of the ATAC-seq data was highly reminiscent to the bulk RNAseq data PCA presented in Chapter 4 (**Fig 5.2**). The spatial arrangements of the time points resemble the RNAseq data and the proportions of the two principal components were also similar. In addition, PC1 mainly separates the early (iPSCs to 4 days) and the late (days 14 and 21) stage of the time course, while PC2 separates day 1 to day 4, from the rest of the time course. The changes in the chromatin accessibility over the time course reflect two expected principle biological processes – cell type specification driven by NGN2 reprogramming and maturation of the resulting cells.

5.3.2 Rapid remodelling of the epigenome from the onset of NGN2 reprogramming

To get insights into the chromatin dynamics over the course of NGN2 reprogramming, I carried out a differentially accessible region (DAR) analysis of the ATACseq data. This involved comparing each consecutive pair of timepoints from the time course. The number of regions that gained accessibility (opening loci) or lost accessibility (closing loci) were recorded for each pairwise comparison. The parameters used for the DAR analysis were a log fold change above 1.5 log counts per million and an FDR score below 0.05. This revealed several interesting changes over the time course. First, the highest gain in accessibility occurred after 6 hours post-induction, with low gain in accessibility in comparison for the following time points up until 4 days (**Fig 5.3A**). In contrast, there were hardly any sites losing accessibility during the first four days of reprogramming. Genes associated with opening loci at 6h, were associated with GO terms involved in early and late neuronal development (**Fig 5.3B**). The Notch signalling genes *DLL3*, *DLL4* and *DLL1*, which are associated with NSC maintenance and early neuronal differentiation, were among the genes with the highest significant change. Indeed, *positive regulation of Notch signalling pathway* were among the most highly ranked terms for regions that gained accessibility within 6 hours of NGN2 induction (**Fig 5.3C**). In addition, GO terms associated with neuronal differentiation were also significantly enriched.

Between 4 days and 14 days post-reprogramming, there is another increase in chromatin accessibility; this is accompanied by a significant loss of accessibility (**Fig 5.4 A**). As expected, some of the ontologies enriched for the opening loci during this time window were associated with neuronal maturation such as *positive regulation of neuron maturation* and *dendritic spine*

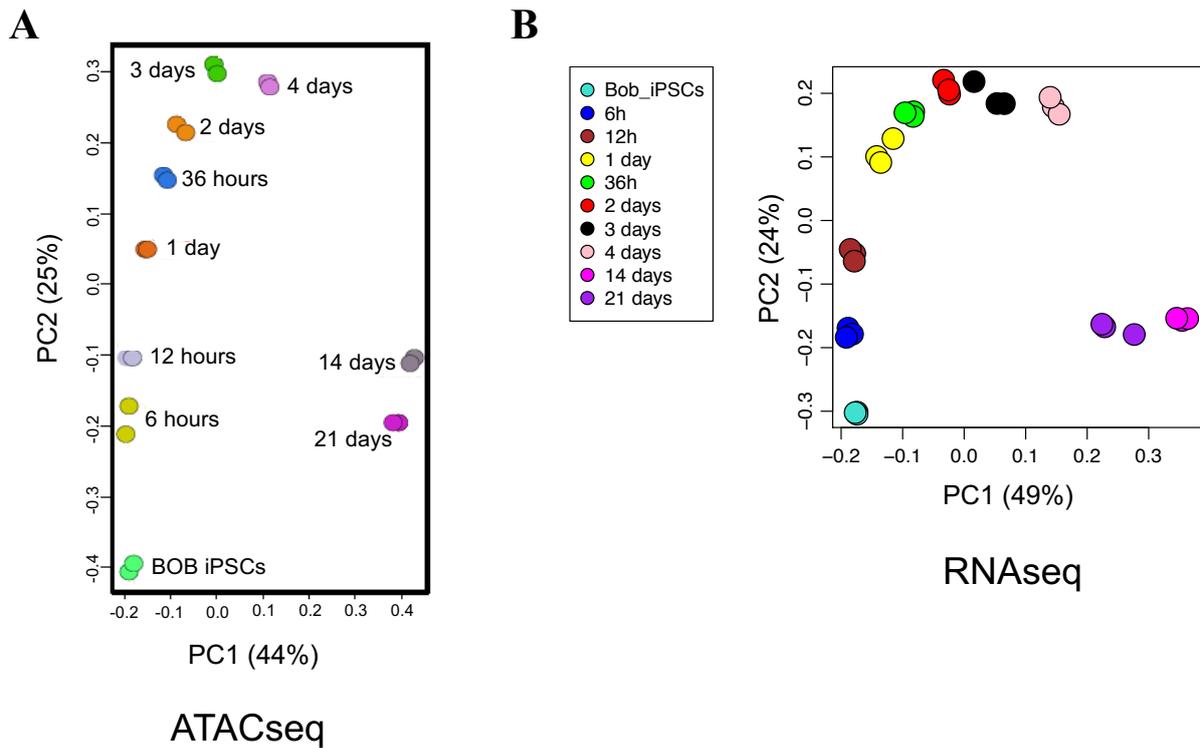


Figure 5. 2: PCA of ATACseq data shows high similarity to PCA of bulk RNAseq

PCA plot of the top two principal components for (A) bulk ATACseq and (B) bulk RNAseq analysis of the NGN2 reprogramming time course. The proportion of each component is shown in brackets. Both plots show a similar trend for the time course, and similar proportions of the top two components.

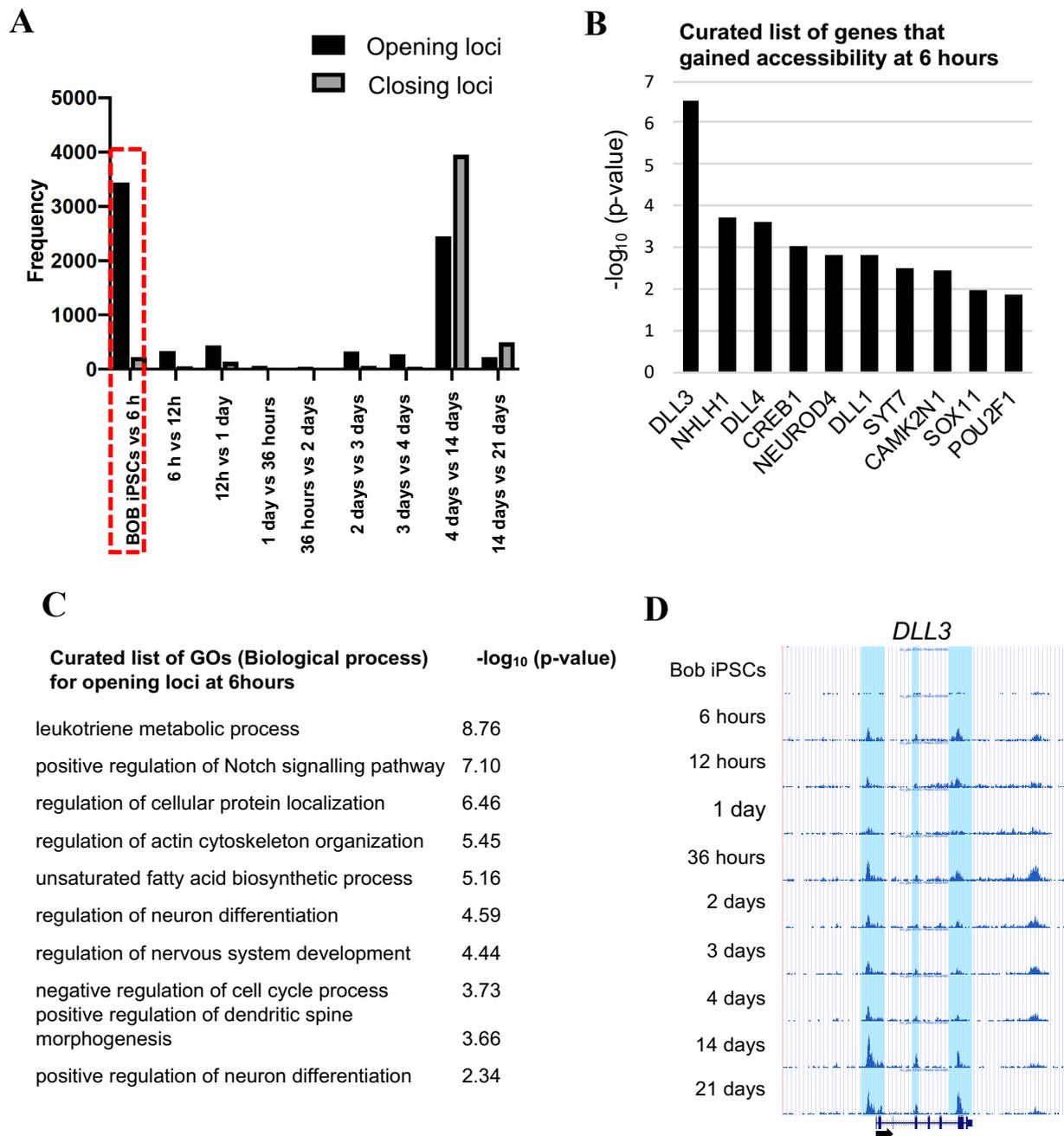
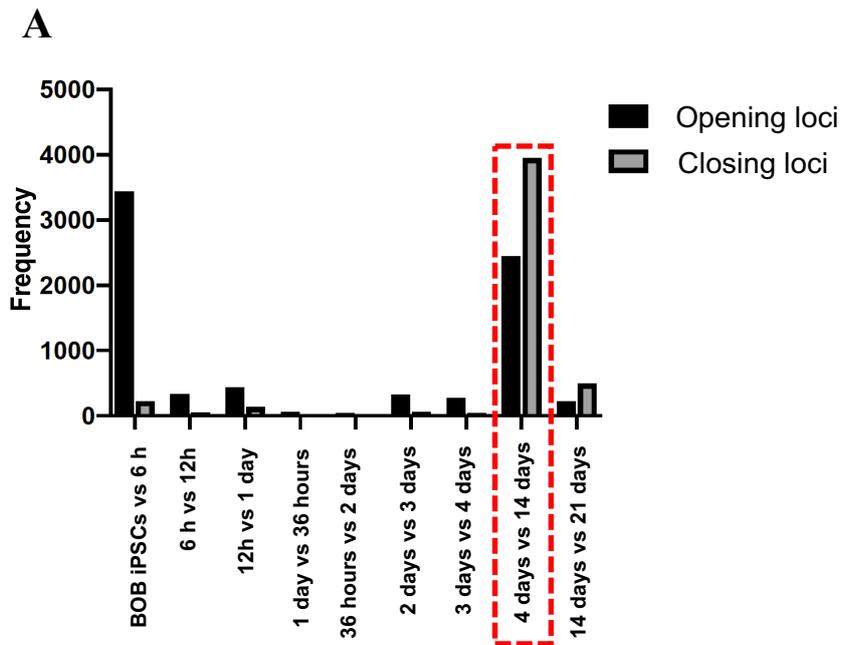


Figure 5. 3: Widespread gain of accessibility within six hours of NGN2 reprogramming.

(A) Plot of counts (frequency) for sites gaining accessibility (opening loci) and sites losing accessibility (closing loci) throughout NGN2 reprogramming. (B) A selection of genes associated with loci that gained accessibility at 6 hours post-induction (p -value <0.05). (C) Biological process gene ontologies associated with loci that gained accessibility at 6 hours post-induction (p -value <0.05) (D) Genome tracks showing ATACseq data for the entire time course at *DLL3* locus. Arrow indicates direction of transcription, while the height of each track is set at 127 pixels. Regions that gain accessibility are highlighted in light blue. Note the immediate gain in accessibility at 6 hours.



B Curated list of GOs for opening and closing loci at 14 days

Opening loci		Closing loci	
GO (Biological process)	$-\log_{10}$ (p-value)	GO (Biological process)	$-\log_{10}$ (p-value)
positive regulation of neuron maturation	3.30	regulation of neurogenesis	5.37
dendritic spine organisation	2.35	epithelial to mesenchymal transition involved in cardiac fibroblast development	5.18
axonal transport of mitochondrion	2.26	muscle cell differentiation	4.38
positive regulation of exit from mitosis	1.73	neuron migration	3.55
visceral motor neuron differentiation	1.43	regulation of Notch signalling pathway	3.36

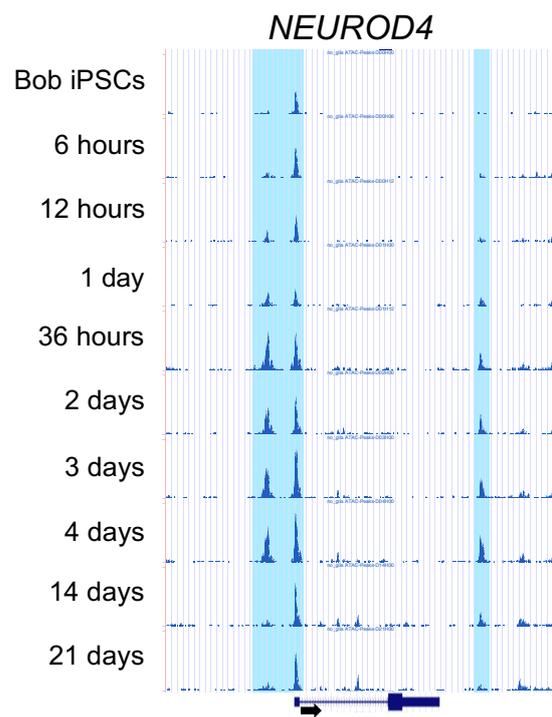
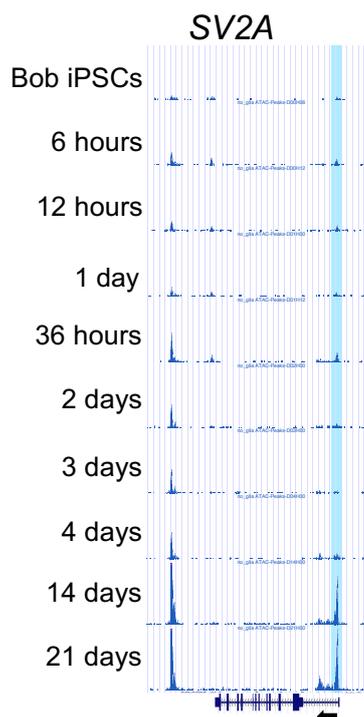


Figure 5. 4: Loss of chromatin accessibility mainly occurs between 4 days and 14 days post-induction.

(A) Plot of counts (frequency) for sites gaining accessibility (opening loci) and sites losing accessibility (closing loci), throughout the course of NGN2 reprogramming. (B, Top panel) A selection of significant Gene Ontologies (biological processes) for opening and closing loci at 14 days post-induction, compared to 4 days post-induction (p-value<0.05). (B, Bottom panel) Genome tracks showing ATACseq data of representative genes for opening and closing loci between day 4 and day 14. Arrow indicates direction of transcription, while the height of each track is set at 127 pixels. Regions that gain or lose accessibility are highlighted in light blue.

organisation (Fig 5.4B). One such gene associated with this is the synaptic vesicle gene SV2A, which showed a noticeable gain in accessibility around its promoter region at Day 14 onwards. In the previous chapter, scRNAseq analysis of the reprogramming uncovered a subpopulation of cells with a transcriptional profile which in part resembled motor neurons. Here, one of the GOs identified in the opening loci between Day 4 and 14 supports this (*visceral motor neuron differentiation*), however, the genes enriched for this GO only include ISL1 and TBX20. TBX20 was not detected in the scRNAseq dataset. Closing chromatin loci included terms related to neuronal commitment and early development such as *regulation of neurogenesis* and *neuron migration*. One of the genes that showed significant loss of accessibility at Day 14 was NEUROD4 (Fig 5.4B). The NEUROD4 locus displays a gradual gain in accessibility around its promoter region from 6 hours onwards with peak accessibility between 36 hours and Day 4, after which there is loss in accessibility. This correlates with both bulk and single cell transcriptional expression data of NEUROD4 in the previous chapter (Fig 4.16 & 4.29).

Together, these results indicate there is rapid remodelling of the epigenome immediately after NGN2 reprogramming. During the first 4 days of reprogramming, chromatin regions mostly gain accessibility. After reaching neuronal identity at day 4, the subsequent maturation phase sees another gain of accessibility in regions associated with neuronal maturation, but this time accompanied by a major loss of chromatin accessibility in regions associated with neurogenesis and early differentiation.

5.3.3 Nucleosome occupancy undergo dynamic changes globally over the course of reprogramming and maturation

I next interrogated the global changes in nucleosome occupancy and positioning. Nucleosome occupancy can be inferred from the ATACseq data using NucleoATAC, as described in section 2.12.1 of the Chapter 2. Once the global occupancy profiles for each time point were established, I sought to determine how they vary from one time point to another. This is represented in the correlation heatmap of nucleosome occupancy profiles **Figure 5.5**. Most notably, the occupancy profile for each time point is distinct: most correlation coefficients range between 0.3 and 0.4, with the lowest correlation 0.24 between Day 14 and iPSCs and the highest 0.47 between Day 1 and 36 hours. This suggests that there are dynamic changes in nucleosome occupancy throughout the time course. Hierarchical clustering of the time points showed similarities to the clustering of transcriptional profiles, with analogous pairs such as 6 hours and 12 hours, and 3 days and 4 days, clustering together.

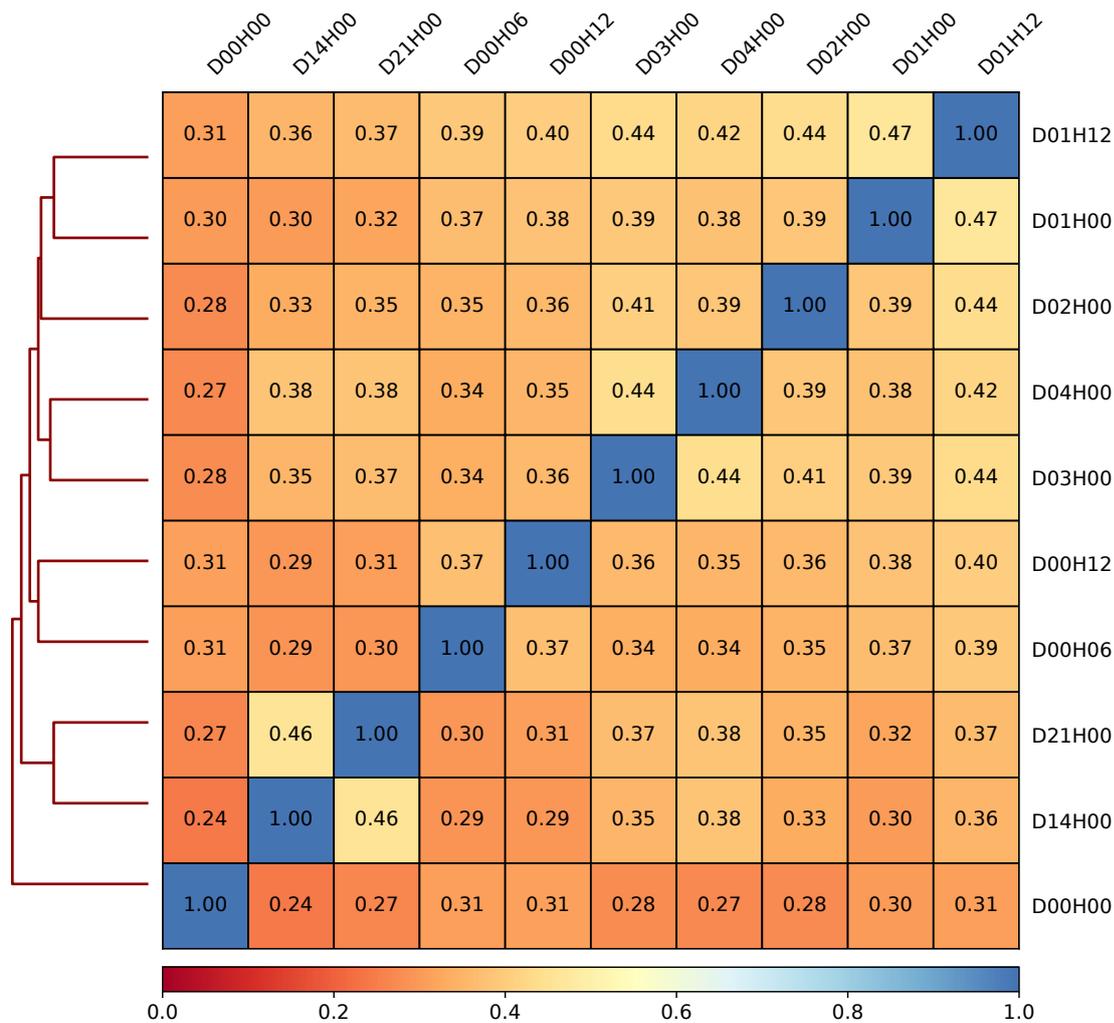


Figure 5. 5: Dynamic changes in nucleosome occupancy during NGN2 reprogramming

Correlation heatmap of nucleosome occupancy over the course of NGN2 reprogramming. Overall, it shows that the nucleosome occupancy profile of each time point is vastly different from one another, with the highest correlation only being at 0.47 (D01H00 vs D01H12).

5.3.4 Rapid establishment of nucleosome occupancy and positioning at NGN2-bound sites during reprogramming.

To investigate the dynamics of the chromatin landscape and nucleosome positioning at NGN2-bound sites, V-plots were generated in which mapping of ATACseq fragments between 36 and 500bp are centred around a putative NGN2 motif (**Fig 5.6**) (Henikoff, Belsky, Krassovsky, MacAlpine, & Henikoff, 2011; Wapinski et al., 2017). The motif was determined separately from the ChIPseq analysis of NGN2 at 24 hours post-induction. From the plots, short (<150-bp) fragments indicative of nucleosome-free DNA accumulate at the NGN2 motif midpoint, and the accumulation of ~200-bp fragments, indicative of +1 and -1 nucleosomes, are also centred around the NGN2 motif midpoint. This is followed by the accumulation of ~400-bp fragments indicative of dinucleosomes, further away from NGN2 motif midpoint. Together, they recapitulate earlier observations of rapid gain in accessibility within 6 hours of reprogramming followed by a loss of accessibility between 4 days and 14 days post-induction. Single and even double nucleosomes (dinucleosomes) are observed around NGN2 motifs starting at 6 hours, with a gradual increase up to 4 days post-induction. The loss of dinucleosomes from 14 days onwards and the notable decline in single nucleosomes and nucleosome-free regions suggest that NGN2 targets play an important role during the early, reprogramming stage and are then silenced during the late, maturation stage of our NGN2 iN protocol.

For example, the nucleosome occupancy profile of the NGN2-target NEUROD1 demonstrates stabilisation of +1 and -1 nucleosomes around its promoter at 36 hours, followed by the placement of a single nucleosome at the start of its 3' UTR at day 3 (**Fig 5.6**). This coincides with a high expression between 36h and Day 3. At Day 14, where NEUROD1 is still expressed, the single nucleosome at the 3' end is lost, but single nucleosomes around the promoter region are still maintained. At Day 21 when NEUROD1 is mostly downregulated, a noticeable decline in nucleosome occupancy at the promoter region was found, followed by an occupancy profile at the 3' UTR that is similar to samples up until Day 2.

To investigate how the addition of glia might have affected nucleosome positioning around NGN2-bound sites, a similar approach was applied. The resulting V-plots show a similar set of features between the two conditions at each time point, albeit with fewer reads, suggesting that the addition of glia did not have a significant impact on the chromatin landscape at NGN2-bound sites (**Fig 5.7**).

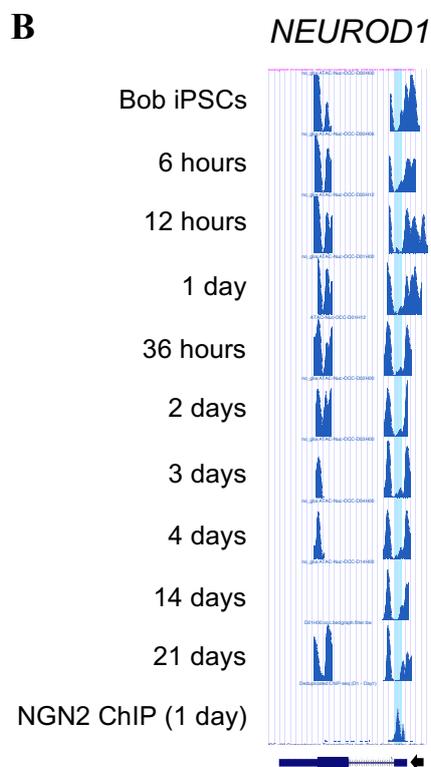
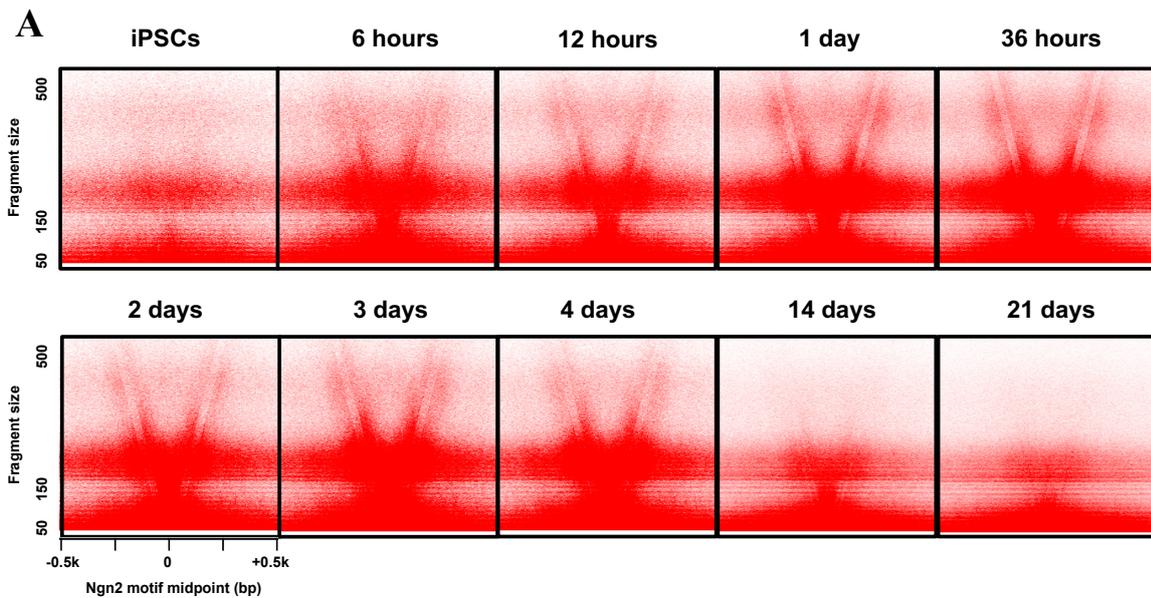


Figure 5. 6: Rapid establishment of nucleosomes at NGN2-bound sites during reprogramming.

(A) V-plot of number of ATAC-seq reads as function of fragment size and distance from NGN2 motif at the centre of NGN2 binding sites. All data are centred on NGN2 motifs across the genome. As early as 6 hours, short (<150-bp) fragments indicative of nucleosome-free DNA accumulate at the NGN2 motif midpoint, along with the accumulation of ~200-bp fragments indicative of +1 and -1 nucleosomes and ~400-bp fragments indicative of dinucleosomes, further away from NGN2 motif midpoint. Note that there is a noticeable gain in all three features during the first 4 days of reprogramming, after which, there is a loss of dinucleosomes and a decline in

single nucleosomes and nucleosome-free loci at NGN2 motifs. (B) Genome tracks showing nucleosome occupancy (entire time course) and NGN2 ChIPseq data (at 1 day post-induction) for NEUROD1. It shows that is a direct target of NGN2 and that changes in nucleosome occupancy occur around NGN2's binding site from 36 hours onwards. Note the decline in occupancy at 14 and 21 days post-induction. Loci corresponding to NGN2 binding is highlighted in blue. Height of nucleosome tracks are between 0 and 1, and represent probability of nucleosome occupancy, with 1 being the highest probability of nucleosome occupancy. Height of ChIPseq track is set at 127 pixels. Arrow indicates the direction of transcription. sizeable

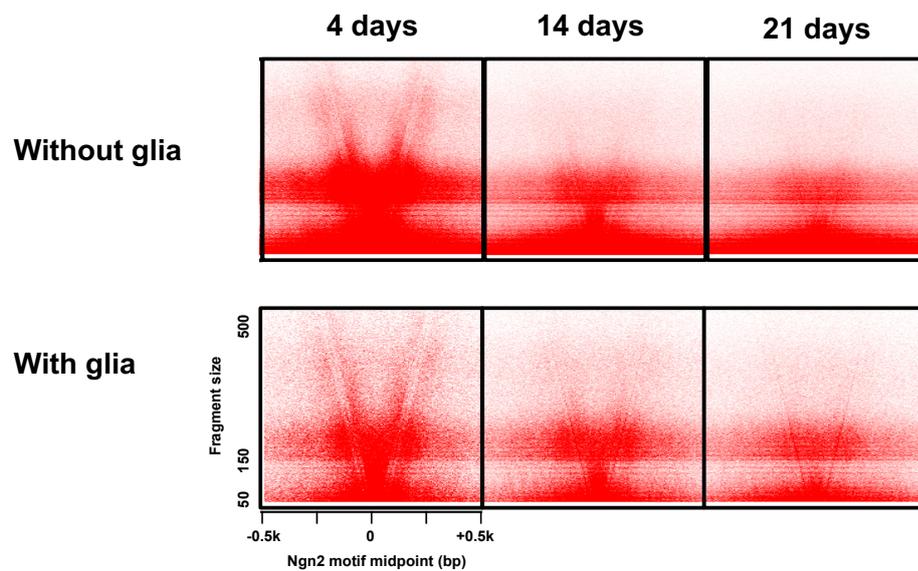


Figure 5. 7: Nucleosome positioning and chromatin accessibility at NGN2 bound-sites for cells co-cultured with glia.

V-plot of number of ATAC-seq reads as function of fragment size and distance from NGN2 motif at the centre of NGN2 binding sites, for time points where cells were either cultured without glia (top panel) or with glia (bottom panel). All data are centred on NGN2 motifs across the genome. At 4 days, for both conditions, there is evident presence of features for nucleosome-free loci at NGN2 bound sites (<150bp fragments), single nucleosomes (~200bp fragments) and dinucleosomes (~400bp fragments). Then, both conditions show loss of dinucleosomes and evident decrease in single nucleosomes and nucleosome-free loci.

5.3.5 Identification of transcription factors driving changes in chromatin accessibility during NGN2 reprogramming and maturation.

To infer significant TF activity from the changes in chromatin accessibility over this time course, I used the recently described DASTk package (Tripodi, Allen, & Dowell, 2018), which utilises 680 position weight matrices (PWMs) of human motifs in the HOCOMOCO database (Kulakovskiy et al., 2018). This tool applies a metric called a motif-displacement score (MD-score), which reflects the enrichment of a TF sequence motif within the midpoint of detected ATAC-seq peaks. A positive MD score reflects TFs that show increased activity at a particular timepoint compared to the previous timepoint, while negative MD scores reflect decreased activity. In addition, the higher the absolute value of the score, the greater the activity, i.e; the greater the frequency of its target loci either gaining or losing accessibility. Because the HOCOMOCO database depends on published, experimentally determined motifs, the confidence in a motif in turn depends greatly on the quality of the experiment, where the lowest rated motifs can only provide a rough description of a binding pattern and should therefore be used primarily in exploration studies. This quality rating is recorded for each motif in the collection and will therefore be considered in the present analysis.

This approach generated a sizeable list of TFs (Appendix 7.3), and for many of them, would require further investigation beyond the time frame of this degree. Therefore, I chose to focus on TFs which were identified in the previous chapter and are backed by sufficient evidence from existing literature (**Table 5.1 and 5.2**, p-value <0.05).

By considering only TFs with highly significant activity (p-value < 10^{-7}) and large MD scores (absolute values of above 0.100), we get a distribution that coincides with the major opening and closing events described in Figure 5.2 and 5.3, suggesting that these could be the main TFs driving these changes. The first of these is NEUROD1, coinciding with the large gain in accessibility at 6 hours. Being an established target of NGN2, this result provides additional evidence to the nucleosome positioning dynamics in Figure 5.5, that NGN2 targets are the main sites gaining accessibility during the first 6 hours of reprogramming.

The second set of TFs with highly significant activity coincides with the large gain in accessibility between Day 4 and Day 14. Between the two, ONECUT1 contributed the highest activity, followed by PBX2 (Table 5.1). ONECUT1 showed a modest decrease in activity from Day 14 to Day 21, suggesting its target sites are highly regulated in the late stage of this time course. I thus sought to explore the transcriptional profile of ONECUT1, -2 and -3 along the observed time course. In support of the DASTk analysis, single-cell transcriptional data of these TFs displayed

Timepoint	TF	+ MD Score	TF	- MD Score
6 hours	NEUROD1	0.219	-	-
12 hours	CUX1	0.042	POU5F1 NANOG KLF3	-0.087 -0.084 -0.020
1 day	-	-	REST NANOG	-0.038 -0.052
36 hours	-	-	-	-
2 days	CUX1 NEUROD1	0.087 0.049		
3 days	-	-	-	-
4 days	PAX6 ONECUT1	0.059 0.046	-	-
14 days	ONECUT1 PBX2 POU3F2	0.219 0.131	NEUROD1 PAX6	-0.201 -0.056
21 days	BACH2	0.073	CUX1 NEUROD1 ONECUT1	-0.059 -0.083 -0.079

Table 5. 1: Dominant transcription activity over the course of NGN2 reprogramming based on ATACseq data.

Table listing significant TFs predicted using DASTK analysis (p-value <0.05), for every timepoint from the NGN2 time course. For each timepoint, the +MD score reflects TFs that show increased activity at that particular timepoint, compared to the previous timepoint., while -MD scores reflect decreased activity. TFs with high activity (absolute MD score of above 0.100 AND p-value < 10⁻⁷) are highlighted in grey. CHIP-validated targets of NGN2 are highlighted in bold.

Timepoint	TF	+ MD Score	TF	- MD Score
4 days with glia	-	-	ONECUT1 POU3F2	-0.045 -0.057
14 days with glia	-	-	POU3F2	-0.075
21 days with glia	CUX1 SIX2 SIX1	0.037 0.064 0.058	-	-

Table 5. 2: Dominant transcription activity between iNs cultured with glia and iNs cultured without glia, based on ATACseq data

Table listing significant TFs predicted using DASTK analysis, for time points where iNs were co-cultured with glia. For each timepoint, the +MD score reflects TFs that show increased activity at that particular timepoint, compared to iNs cultured without glia at the same timepoint., while - MD scores reflect decreased activity. TFs with high activity (absolute MD score of above 0.100 AND p-value < 10⁻⁷) are highlighted in grey.

an increase in expression from Day 4 to Day 14, implying that these class of TFs may play a role for cell maturation (**Fig 5.8 & Fig 5.12**). Similar to ONECUT TFs, PBX2 and its homologues, PBX1 and PBX3 share similar binding motifs to PBX2 on the database and therefore could also be a group of TFs playing an important role in the late stage. Although PBX2 expression remains mostly unchanged across the time course, PBX1 and PBX3 show increased expression at Days 14 and 21, with PBX3 showing specific enrichment in cells co-cultured with glia (**Fig 5.9 & Fig 5.12**).

The third TF with highly significant activity coincides with the large loss of accessibility between day 4 and 14. Here, NEUROD1 motifs were highly ranked. In addition, it was also detected, albeit with a much lower score, for sites closing between Day 14 and Day 21. As shown in Figures 4.35 of chapter 4 and **Figure 5.12**, both bulk and single-cell RNAseq analysis show only a modest decrease in NGN2 expression from Day 4 to Day 14, likely due to residual dox in the medium; but the decrease in NEUROD1 expression is more noticeable, suggesting that this decrease in NEUROD1 activity is independent of the dox-induced NGN2 expression. Nonetheless, these results further evidence to a decreasing role for NGN2 and its immediate downstream targets during the maturation stage.

Although of lower significance in p-value, this analysis also highlighted several other important TFs. Most notably are the pluripotency TFs, POU5F1, NANOG and KLF, which showed decreased activity at 12 hours, followed by the neuronal repressor, REST at Day 1 (**Table 5.1**). This coincides with their transcriptional downregulation in the time course, as demonstrated in the previous chapter. Together, this result indicates that these non-neuronal TFs and their targets are rapidly shut down within the first 24 hours. On the other hand, there is a subsequent increase in the time course for PAX6, CUX1 and POU3F2, which are TFs commonly associated with glutamatergic differentiation in the dorsal cortex (**Table 5.1**). PAX6 in particular, showed an increase at Day 4, followed by a decline in activity at Day 14, suggesting that PAX6 plays a role in the early stage of differentiation in NGN2 reprogramming. Another interesting TF from the analysis is BACH2, an NGN2 target which showed a small gain in activity at Day 21 (**Table 5.1**). A look at its transcriptional profile in the time course revealed that its expression increases in a small portion of post-mitotic cells at Day 3 onwards (**Fig 5.11 & 5.12**). Its expression increases in a greater number of cells at Day 14 and Day 21, with expression being higher in co-cultured cells, especially Day 14 neurons co-cultured with glia (**Fig 5.11**). Together, the accessibility and transcriptional profile suggest that BACH2 could be an important regulator of the late, maturation stage.

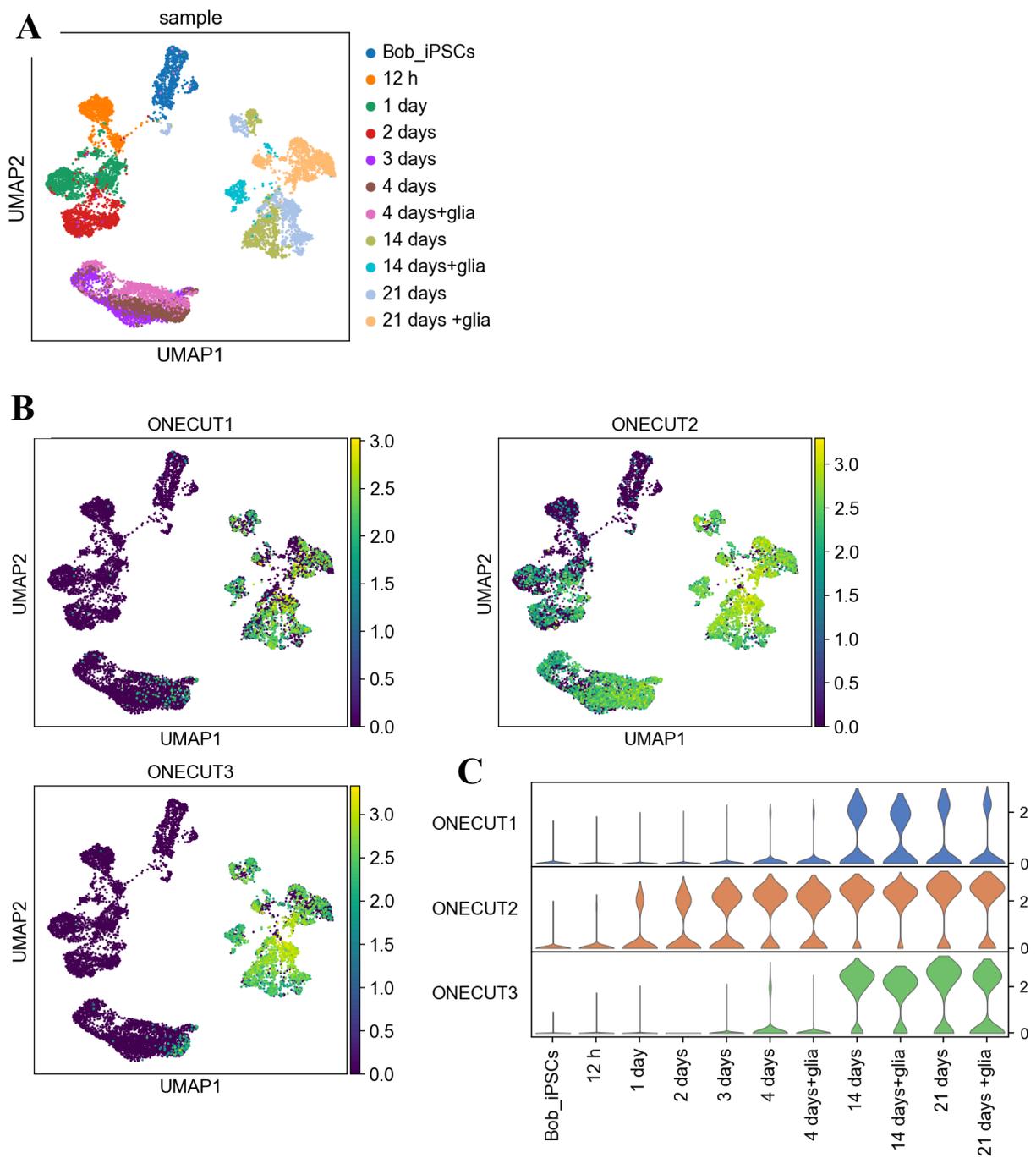


Figure 5. 8: ONECUT TFs show increased expression in the late maturation stage, as predicted by DASTk analysis.

(A) UMAP visualisation of all time points used in scRNAseq of NGN2 reprogramming time course, including iNs cocultured with glia. (B) UMAP visualisation of ONECUT1, -2 and -3 across the same time points. (C) Stacked violin plots showing the overall single-cell expression levels across the time course.

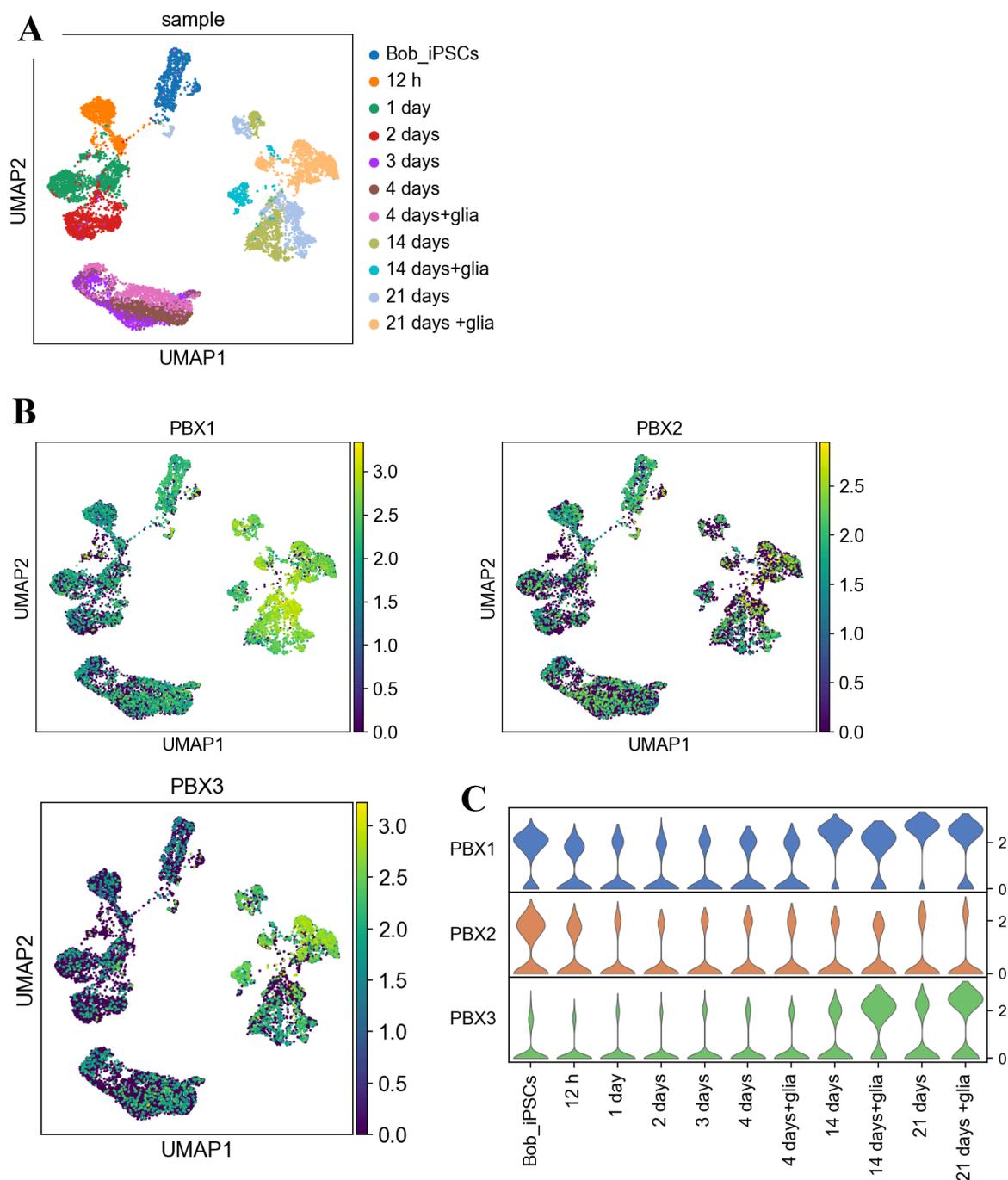


Figure 5. 9: PBX1 and 3, not PBX2, show increased expression in the late maturation stage.

(A) UMAP visualisation of all time points used in scRNAseq of NGN2 reprogramming time course, including iNs cocultured with glia. (B) UMAP visualisation of PBX1, -2 and -3 across the same time points. (C) Stacked violin plots showing the overall expression levels for each TF across the time course.

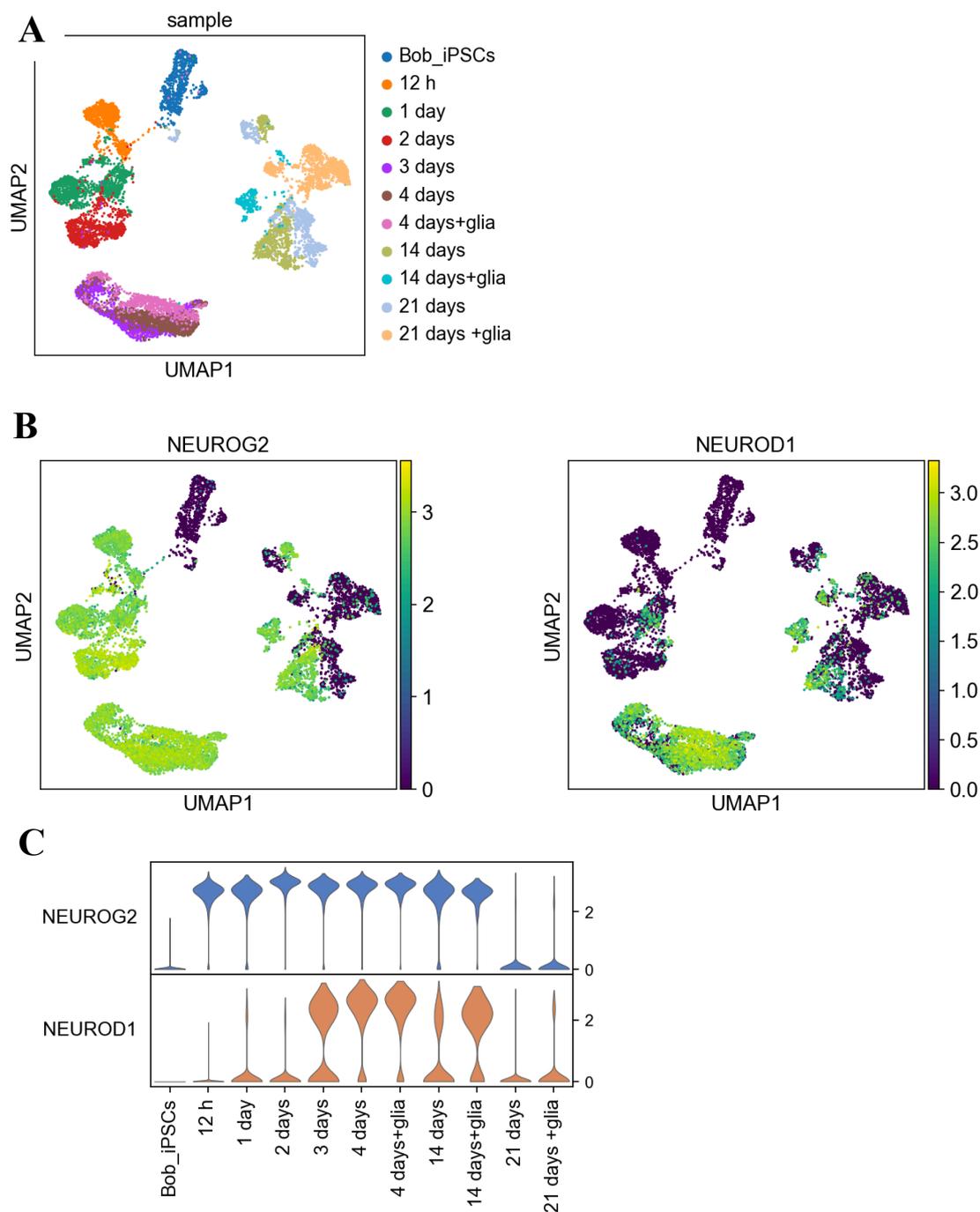


Figure 5. 10: Decrease in NEUROD1 expression in the late stage is likely independent of NGN2 expression.

(A) UMAP visualisation of all time points used in scRNAseq of NGN2 reprogramming time course, including iNs cocultured with glia. (B) UMAP visualisation of NGN2 (NEUROG2) and NEUROD1. (C) Stacked violin plots showing the overall single-cell expression levels across the time course.

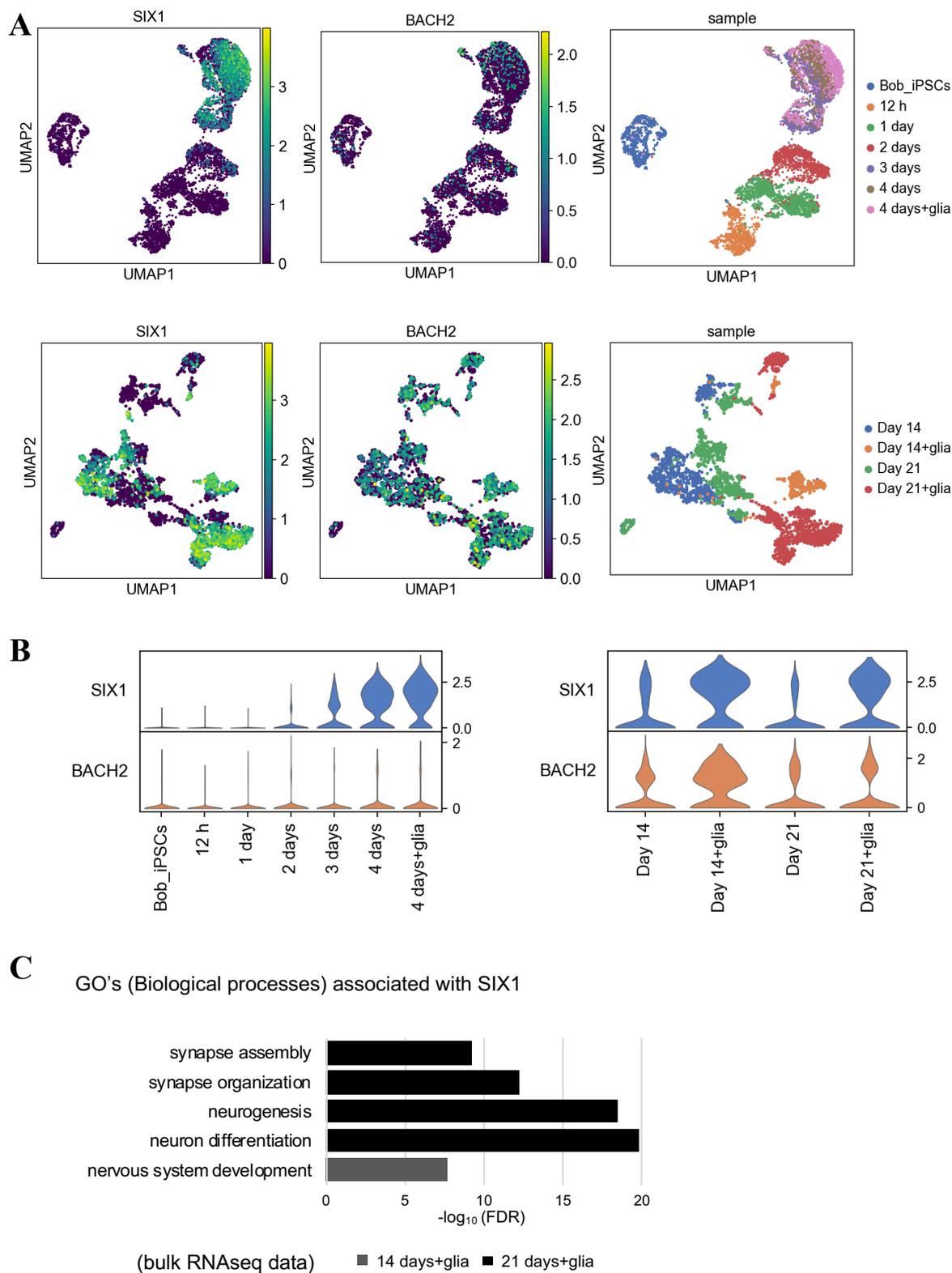


Figure 5. 11: Expression of SIX1 in NGN2 reprogramming and in neurons with glia.

(A) UMAP plots of scRNAseq data showing the expression levels of SIX1 across the population during the early stage of NGN2 reprogramming (top left) and late stage reprogramming (bottom left). The sample arrangement for each stage is shown to the right, for reference. (B) Stacked violin plots showing the overall expression levels for each timepoint in the early (left) and late

stage (right) of reprogramming. There is a slight increase in SIX1 expression at 2 days, before significant expression from 3 days onwards. SIX1 expression is higher in all timepoints with glia compared to those without. (C) A selection of gene ontologies (biological processes) that are associated with SIX1, based on bulk RNAseq data. These ontologies had higher expression in iNs cocultured with glia at days 14 and 21.

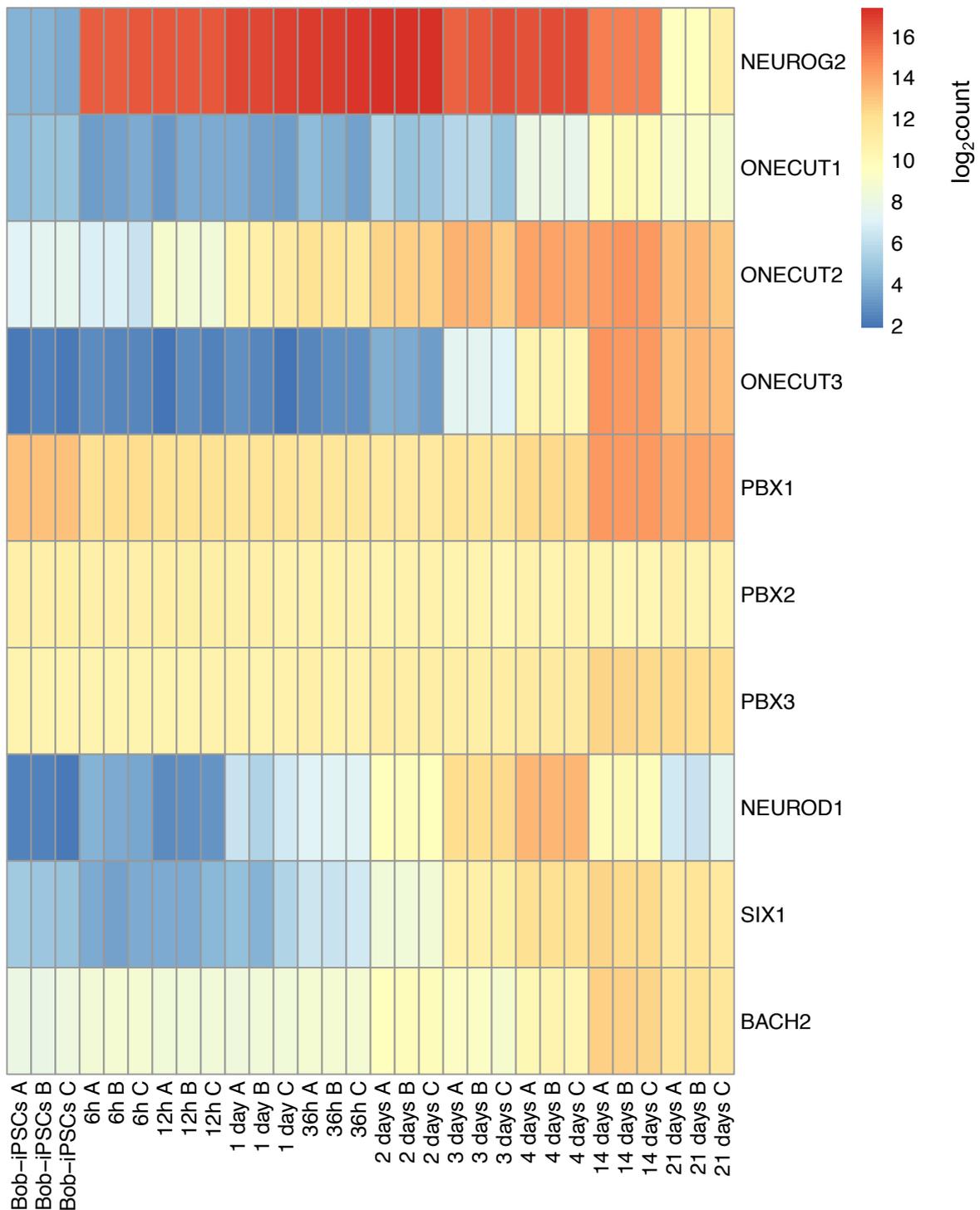


Figure 5. 12: Bulk RNA-seq expression profiles of a selection of TFs identified from DASTk and homologues for ONECUT and PBX.

Heatmap of log₂ expression for NGN2 (NEUROG2), ONECUT1, -2, -3, PBX1, -2, -3, NEUROD1, SIX1 and BACH2 for all ten time points from the time course (without glia), with 3 biological replicates (A, B and C) for each timepoint.

I also used this approach to identify TFs regulated between neurons cultured with and without glia (**Table 5.2**). Some of the TFs from this list were also identified in the analysis of the whole time course, such as ONECUT1, POU3F2 and CUX1. Also amongst the active genes, is the transcription factor SIX1, a gene that has not yet been identified in the context of NGN2 reprogramming, which showed increased activity for Day 21 neurons cultured with glia. The expression for SIX1 in the scRNA-seq dataset homogeneously increased from Day 2 to Day 4 of reprogramming (**Fig 5.11**). Heterogeneous expression was observed at Day 14 and 21 in both neurons cultured with and without glia. Cells cultured with glia, overall had a higher expression of SIX1 and a greater proportion with high expression of the TF (**Fig 5.11**). To obtain insights into its putative function, gene ontologies associated with SIX1 were searched at the same timepoints in the functional enrichment analysis of the bulk RNA-seq data (**Fig 5.11**). This indicated associations with neurogenesis, neuronal differentiation and synaptogenesis. Together, these results point to a neurogenic role for SIX1 in the generation of these neurons and that the addition of glia potentiates its expression.

5.4 Discussion

This chapter provided insights into the epigenomic landscape of NGN2 reprogramming of human iPSCs. It highlighted the main changes and the genes that drive them; thus, building on the transcriptional evidence examined in the previous chapter. PCA analysis of samples revealed similarities to the transcriptional changes across the time course. NGN2 induction leads to significant reorganisation of chromatin. This begins with a rapid gain in accessibility, largely centred around NGN2 binding sites, within 6 hours of reprogramming. Subsequent time points in the first four days see little gain in accessibility. There is then a large gain in accessibility from Day 4 to Day 14. In contrast, the vast majority of closing events take place between Day 4 and Day 21.

The quick unravelling of the chromatin around NGN2 binding sites draws a similar comparison to the chromatin dynamics in Ascl1 reprogramming of MEFs (Wapinski et al., 2017), but the speed in which it occurs may explain why NGN2 reprogramming of PSCs is so efficient and successful. In a comparable time course study, Wapinski and colleagues showed that a single major “concerted” switch occurs between 2 days and 5 days post-Ascl1 induction of mouse embryonic fibroblasts, where there is both a large opening and closing of the chromatin within this period, which are also centred around the binding sites of the proneuronal TF. In contrast, my results showed that large-scale opening and closing of the chromatin in NGN2 reprogramming occurs during distinct phases of reprogramming and maturation. Even as majority of the cells have

exited the cell cycle and committed to a neuronal fate between Day 3 and Day 4, large-scale closing of non-neuronal sites only occurs later on.

Rapid chromatin opening from the onset of induction may be driven by pioneering activity of NGN2, binding its targets in closed, compacted regions. Indeed, NGN2 binding sites appeared to be mostly located in closed, inaccessible regions of the iPSC genome, before being drastically opened-up within the first 6 hours. In NGN2 reprogramming of fibroblasts, the addition of small molecules activate signalling cascades that promote NEUROG2 and CREB1 co-transcription, induce SOX4 expression, enhance H3K27 acetylation, and promote SOX4-dependent chromatin remodelling (Smith et al., 2016). It is therefore conceivable that the success of NGN2 reprogramming in the present study depends on a combination of its innate ability to access closed chromatin in PSCs and the availability of co-operating factors, like CREB1. This suggests limitations to NGN2's pioneer factor activity compared to ASCL1. Nevertheless, the results in this study indicate that NGN2's *pioneering* abilities may have been underestimated.

It was anticipated that NGN2 itself would be the dominant TF identified by DASTk, driving the gain in accessibility at 6 hours; but the analysis identified NEUROD1 instead. NEUROD1 and NEUROG2 share highly similar binding motifs, both in the HOCOMOCO database used for DASTk, and the widely-used JASPAR database (Fornes et al., 2020). In addition, the quality of the experiment used for NGN2 in the HOCOMOCO database, is of a lower quality compared to NEUROD1 (Kulakovskiy et al., 2018). Given that NEUROD1 is mainly expressed from 36 hours post-induction, both in the bulk and single cell RNA-seq datasets, it is very likely that NGN2 is the intended TF for this particular change in accessibility.

Chromatin compaction largely occurred between Day 4 and Day 14, i.e. between the early, reprogramming stage and the late, maturation stage, where cell type is successfully established. Motif enrichment analysis of these regions and of sites centred around the NGN2 motif showed that they are mainly regions targeted by NGN2 and NEUROD1. This closure of chromatin could have occurred as a direct result of decreased NGN2 overexpression. However, the expression level of NGN2 in Day 14 neurons was only slightly lower compared to Day 4 neurons; whereas, the decrease in NEUROD1 expression in Day 14 neurons was considerable. This result suggests two things. First, that the decrease in NEUROD1 activity determined by DASTk is not a result of any decrease in NGN2 overexpression. Second, that sustained NGN2 overexpression has little effect on the overall transcriptome of the neurons in the late stage of the protocol, as many of its targets are being packed into closed chromatin. In agreement with this, some protocols for making NGN2

iNs maintain dox-induced expression of NGN2 throughout their protocol (Nehme et al., 2018; Yingsha Zhang et al., 2013).

Besides NGN2 dominant changes in the epigenome, this chapter also suggested a role for ONECUT TFs in triggering large opening of the chromatin. This occurred beyond the early stage of reprogramming and is likely implicated with the switch from an early reprogramming stage where neuronal identity is established to one that establishes neuronal maturation. This is in agreement with a previous study that found that motifs for all three subtypes (ONECUT1, 2 and 3) are the most enriched TFs in Day 21 NGN2-induced hiPSC iNeurons, when compared to accessibility data of human iPSCs and fibroblasts (van der Raadt et al., 2019). They subsequently showed when iPSCs are reprogrammed with ONECUT TFs, the ontologies enriched for highly expressed genes are associated with synaptic function, membrane permeability, axon development, and neuronal differentiation, suggesting that these TFs could be particularly important for neuronal maturation, rather than neuronal fate establishment and early differentiation. In the present study, I not only provide further evidence to their involvement in NGN2 reprogramming, but additionally inform on the temporal nature of their regulation. Their role in neuronal development has been demonstrated before, primarily in the spinal cord, involving diversification of motor neurons (Roy et al., 2012) and the differentiation and distribution of dorsal interneurons (Kabayiza et al., 2017). In addition, they have also been shown to cooperate with *Isl1* and *Lhx3* in reprogramming mouse ESCs into spinal motor neurons (Velasco, Ibrahim, Ohler, et al., 2017). However, apart from the recent work by van der Raadt and colleagues, no other study have indicated a role for ONECUT TFs in the generation of predominantly glutamatergic neurons. Their analysis for motif enrichment used a separate motif database to HOCOMOCO; this therefore provides strong confidence in the discovery of ONECUT TFs in driving this late, maturation stage of NGN2 iNs. Interestingly, ONECUT TFs were not one of the TFs driving maturation of *Ascl1*-reprogrammed MEFS (Wapinski et al., 2017), suggesting that ONECUT-driven neuronal maturation could be unique to NGN2 reprogramming.

In addition to ONECUT TFs, PBX2 also emerged as a TF with high activity in the late stage; although, its transcriptional profile was largely unchanged across the time course. This prompted the investigation of three other PBX genes, PBX1 and PBX3, of which PBX1 and -3 did show increased expression in the late stage of reprogramming. All four TFs have been suggested to play a role in forebrain development (Long et al., 2009; Toresson, Parmar, & Campbell, 2000). More recently, PBX1 and -2 were elegantly shown to function in progenitors and postmitotic neurons of the developing mouse neocortex to regulate patterning of the cerebral cortex, in part by

repressing genes that promote a dorsocaudal fate (Golonzhka et al., 2015). Interestingly, in van der Raadt's 2019 study that was mentioned earlier, motifs for all four PBX's were enriched in regions that gained accessibility following ONECUT1 and ONECUT2 overexpression, suggesting that the ONECUT TFs could be acting upstream of PBX in regulating the late stage of NGN2 reprogramming. Together, these provide compelling evidence that both class of TFs could be significant players in driving the late stage of NGN2 iN differentiation. While van der Raadt and colleagues have demonstrated this directly by overexpressing ONECUT1, -2 and -3 in human fibroblasts, the supporting evidence in this chapter warrants a similar test for PBX TFs.

This chapter also elucidated the involvement of TFs more familiar to the generation of glutamatergic neurons in the developing cortex, that is PAX6, POU3F2, and CUX1. The activity of PAX6 and POU3F2 in this time course seems to coincide with their temporal expression in development. PAX6, believed to be crucial for the initiation of differentiation in NPCs (Götz, Stoykova, & Gruss, 1998; Hodge et al., 2012), was found to be active in Day 4 of the reprogramming, where most of the cells have just exited the cell cycle. POU3F2 on the hand, believed to act downstream of PAX6 in cortical development (Hodge et al., 2012), was found to be active in between Day 4 and 14 of the reprogramming. CUX1 activity however was detected in more than one point in the time course – active at 12 hours and Day 2, and then decreased activity at Day 21. Interestingly, Day 21 neurons cocultured with glia showed higher CUX1 activity in the ATAC-seq data compared to Day 21 monocultures. In murine cortical development, CUX1 activity is also found in both mitotically active and post-mitotic populations, with expression in progenitors of the VZ and later on in distinct subpopulations in layer II/III of the upper cortex (Nieto et al., 2004; Rodríguez-Tornos et al., 2016). The increased expression of CUX1 in neurons co-cultured with glia could point to a key transcriptional network that mediates the increased functionality and maturity seen in these neurons over plain monocultures.

Transcriptional analysis of neurons co-cultured with glia in Chapter 4 revealed a network of maturation and synaptic genes being enriched in co-cultured neurons at Day 14 and Day 21. From the DASTk analysis in the present chapter, two TFs were identified that could be regulators of these important processes, SIX1 and BACH2. The DASTk activity of the latter wasn't specifically detected in co-cultured neurons, but along with SIX1, transcriptional evidence showed that both TFs began to be highly expressed in post-mitotic neurons, at around Day 3 and 4, and showed higher expression in neurons co-cultured with glia at the later stage. Neither TFs have been described in the context of NGN2 reprogramming. SIX1 has been shown to be crucial for neurogenesis in the inner ear, acting upstream of NEUROG1 and NEUROD1 (Ahmed, Xu, & Xu, 2012) and in the development of sensory neurons in the peripheral nervous system (Sato et

al., 2015; Yajima et al., 2014). Evidence on BACH2 is even more limited, but very promising. A recent pioneering study that produced a single-cell transcriptomic atlas of human neocortical development during mid-gestation found BACH2 as a crucial TF in the gene regulatory network specific to maturing excitatory neurons. Further investigation of these factors could uncover novel roles in promoting neuronal maturation and function. In addition, they could serve as potential targets for promoting these crucial properties in the absence of glia co-culture.

The DASTk motif enrichment analysis also detected many several other new and unexpected TFs, which will require additional interrogation beyond the time frame of this degree. The most important of these are the AP-1 associated genes, JUN, JUND and FOSL2. AP-1 and its components are immediate early genes that are expressed rapidly in the CNS in response to various stimuli, including synaptic transmission (Tuvikene, Pruunsild, Orav, Esvald, & Timmusk, 2016; Yap & Greenberg, 2018). In *in vitro* cultures, it can also be expressed in response to dissociation reagents such as trypsin, and collagenase, but especially in in dissociations for single cell sequencing experiments (Wu et al., 2017; Van Den Brink et al., 2017). Even though, cells for the ATACseq experiment were only exposed to a short treatment of StemPro Accutase, a mild, gentler form of trypsin, it is likely that this could have been enough to induce the significant activity observed in the late stage monocultures of our reprogramming protocol. In the previous chapter, there was an enrichment for activity regulated genes in the post-synaptic compartment such as ARC, TNFRSF12A and EGR1. Thus, it is also possible that the AP-1 TFs identified in this chapter are in response to electrophysiological activity.

In summary, investigating the chromatin accessibility of NGN2 reprogramming complemented the findings from the transcriptional dataset, providing a wholistic view of this remarkable process. While the early stage sees cells transitioning through several states – moving out of pluripotency into an NSC-like state and then exiting the cell cycle before committing to a neuronal fate - major change to chromatin occurs in sites open up by NGN2. Major compaction of chromatin only occurs beyond this early stage, possibly helping to maintain neuronal identity and pave the way for a maturation program. Within this same period, sites reserved for maturation are accessed. This is triggered mainly by ONECUT1 and possibly its homologues. It is worth noting that their discovery is limited to the use a single database and the 680 human motifs in it. Further analysis can be performed using other databases such as HOMER and JASPAR, that could verify the findings of this chapter and potentially build on it. Hence, as with the previous chapter, this dataset can serve as a valuable resource for designing future studies and for uncovering novel functions for numerous genes and TFs, which will continue to add to our understanding of NGN2

and its ability to efficiently reprogram human iPSCs into functional and mature excitatory neurons.

6 Conclusions and future directions

In conclusion, the integrative genomic analysis detailed in this thesis provides a rich account of the transcriptional and epigenetic changes that occurs in NGN2 reprogramming of human iPSCs. Using a time course designed based on morphological and electrophysiological changes proved to be successful in capturing the main events in this process. Surprisingly, these events show a stark similarity to the familiar stages of neurogenesis found in development or conventional differentiation protocols - shutting down of non-neuronal networks, in this case pluripotency, establishment of neuronal commitment in an NSC-like stage, cell cycle exit and subsequent onset of neuronal differentiation, followed by neuronal maturation. The granularity of this time course showed how these changes occur in a rapid and controlled manner during the early stages of reprogramming, providing an unparalleled view of how NGN2 is able to successfully reprogram human iPSCs into neurons with high efficiency. In addition, this study also revealed the transcriptional and epigenetic changes underlying the gain in neuronal function when the neurons are co-cultured with rat-derived glia that consisted mainly of astrocytes. Taking into account all the information gathered from this analysis, I have identified the dominant genes, especially transcription factors, associated with each stage of the reprogramming (Figure 6.1). Some of these genes have only recently been described in the context of NGN2 reprogramming, such as REST, NEUROD4 and ONECUTs, but several others have yet to be explored, such as the PBX class of TFs, BACH2 and SIX1. It is likely that some of these factors could potentially be crucial regulators of neuronal differentiation and function, not only in this protocol, but in development as well. Therefore, one major aim for the future would be to explore the role these genes play in NGN2 reprogramming through gene knockdown or overexpression experiments along with investigation of their genomic binding sites. Ultimately, this could also lead to the discovery of new reprogramming strategies for producing neurons with enhanced maturity and functionality without the need of human or rodent glia.

The use of our OPTi-OX model presented a unique opportunity to study the NGN2 reprogramming process in a highly controlled setting, where there is homogenous expression of NGN2 across nearly the entire cell population, as opposed to the variable expression that is commonly seen with lentiviral expression systems. Using scRNAseq, this allowed us to appreciate, how a seemingly homogenous process that produces functional glutamatergic neurons, also produces a side-population of cholinergic neurons. While it has been eluded to in a few studies before, here, I provide conclusive evidence for its emergence in NGN2 reprogramming of human iPSCs and that it forms cholinergic neurons with a specific cranial visceral phenotype.

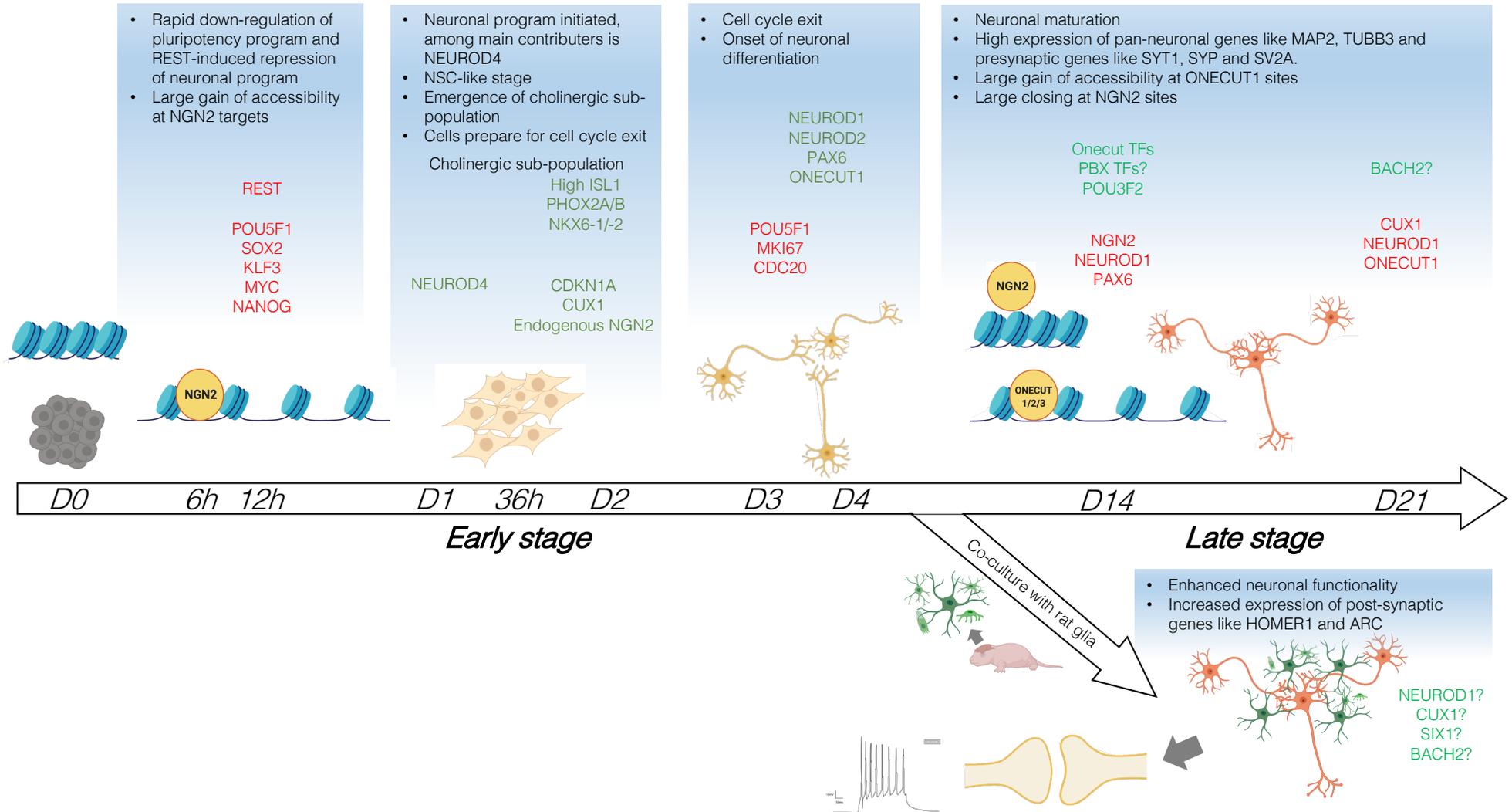
Future studies will seek to determine how they arise in this protocol and how it can be mitigated to produce a homogenous population of glutamatergic neurons.

By uncovering broad, yet essential neuronal processes in this time course, I propose that this system can be used to model aspects of neuronal development, such as neuronal fate commitment and establishment, early differentiation from a progenitor-like stage and late differentiation leading to maturation and gain of function. Towards this, the dataset I have generated can be used as a useful resource for studying these processes and designing relevant disease models.

Due to time constraints, the understanding of NGN2 reprogramming was built by manually merging the information from the two datasets. Although this proved to be very informative, a more systematic approach envisaged for the future is to build a network model of NGN2 reprogramming that integrates the epigenomic and transcriptomic changes uncovered in this study. There are various ways for achieving this, but one simple method would be to design a custom script that uses the motif enrichment analysis and bulk RNA-seq data to filter TF-gene interactions that resulted both in a change in the target gene's chromatin accessibility and differential expression, as demonstrated by Wapinski and colleagues for *Ascl1* reprogramming of fibroblasts (Wapinski et al., 2017). The TF-gene interactions for each timepoint can then be organised as a plot where each node is a TF and each edge represents a relationship between two genes.

The ongoing advancements in assays for next-generation sequencing will see a growth in data from various human tissues. With this, comes an improved curation of databases of human genes and their function. The big data age then is about the sophistication of computational tools to analyse such complex data and extract valuable and meaningful information. Therefore, it is hoped that this rich dataset will continue to reveal new, valuable information on the remarkable cellular metamorphosis of NGN2 reprogramming of human iPSCs into functional excitatory neurons.

NGN2 reprogramming of human iPSCs



(refer to next page for figure description; figure created with BioRender.com)

Figure 6. 1: Schematic summary of the genomic analysis of NGN2 reprogramming of human iPSCs.

The schematic highlights the key events identified across the time course. Upregulated activity of TFs and genes are shown in green; whilst downregulated activity are shown in red, with each gene ordered according to their temporal activity in the time course. TFs accompanied with a question mark are candidate TFs for roles associated with maturation and gain of function.

Upon dox-induced overexpression of NGN2 in human iPSCs, there is large remodelling of the chromatin within the first 6 hours, characterised by pioneer-like binding of NGN2 to previously closed chromatin to consequently open them. In the next 6 hours, there is rapid downregulation of the pluripotency program, marked by downregulation in the activity of pluripotency TFs like POU5F1, SOX2 and KLF3. In parallel, there is also a significant shut-down of REST and its repression of neuronal genes. Together, this leads to an activation of a neuronal program at 24 hours following NGN2 induction, with an NSC-like transcriptome driven by NEUROD4. This period also sees the earliest emergence of a sub-population of cholinergic neurons with a visceral motor identity, based on the co-expression of PHOX2A/B and NKX6-1/2. The expression of CDKN1A at the end of this stage on Day 2 marks the transitioning out of this cycling, progenitor-like stage. Mass exit from the cell-cycle occurs between Day 3 and Day 4, with downregulation in genes like CDC20 and MKI67. The pluripotency gene POU5F1 is completely downregulated at this period. Following this, cells initiate differentiation, most-notably through the downstream targets of NGN2 - NEUROD1 and NEUROD2, and probably through PAX6 as well. By the end of this early stage of NGN2 reprogramming, most cells have attained a neuronal-like morphology. The transition to the late stage sees two more major chromatin remodelling events - large closing of NGN2 targets and a concurrent gain of accessibility largely at ONECUT1 sites (and most likely ONECUT2 and -3), followed by PBX2 and the forebrain TF, POU3F2. This gain in accessibility is associated with neuronal maturation where pan-neuronal genes like MAP2 and TUBB3 and indicators of synaptogenesis such as SYT1 and SV2A see a significant increase in expression from Day 4. At this stage, neurons in the culture have visibly formed a network of connections with one another. BACH2, a TF recently associated with maturation of human forebrain excitatory neurons, could also be playing a similar role in this stage of the reprogramming. When co-cultured with rat-derived glia that mainly consists of astrocytes, neurons at the late stage see enhanced neuronal functionality, evident by an increase in post-synaptic genes like HOMER1 and ARC. Among the candidate TFs that could be playing a role in this increased functionality are NEUROD1, CUX1, SIX1 and BACH2. (Figure created with BioRender.com).

7 Appendix

7.1 List of antibodies

Antibody	Species	Type	Clonality	Company	Catalog number	Dilution
HA peptide	Rat	IgG	Monoclonal	Sigma/Roche	11867423001	1:1000
TUBB3	Mouse	IgG	Monoclonal	Biologend	801213	1:1000
GFAP	Rabbit	IgG	Polyclonal	Agilent	Z033429-2	1:1000
Synapsin1	Rabbit	IgG	Polyclonal	Abcam	ab64581	1:200
MAP2	Chicken	IgY	Polyclonal	Abcam	ab5392	1:2000

7.2 List of primers for qPCR

Gene	Direction	Sequence
NGN2	F	TGTTTCGTCAAATCCGAGACCT
	R	CGATCCGAGCAGCACTAACA
MAP2	F	AGACTGCAGCTCTGCCTTTAG
	R	AGACTGCAGCTCTGCCTTTAG
Synaptophysin (SYP)	F	ACCTCGGGACTCAACACCTCGG
	R	GAACCACAGGTTGCCGACCCAG

7.3 scRNAseq QC report

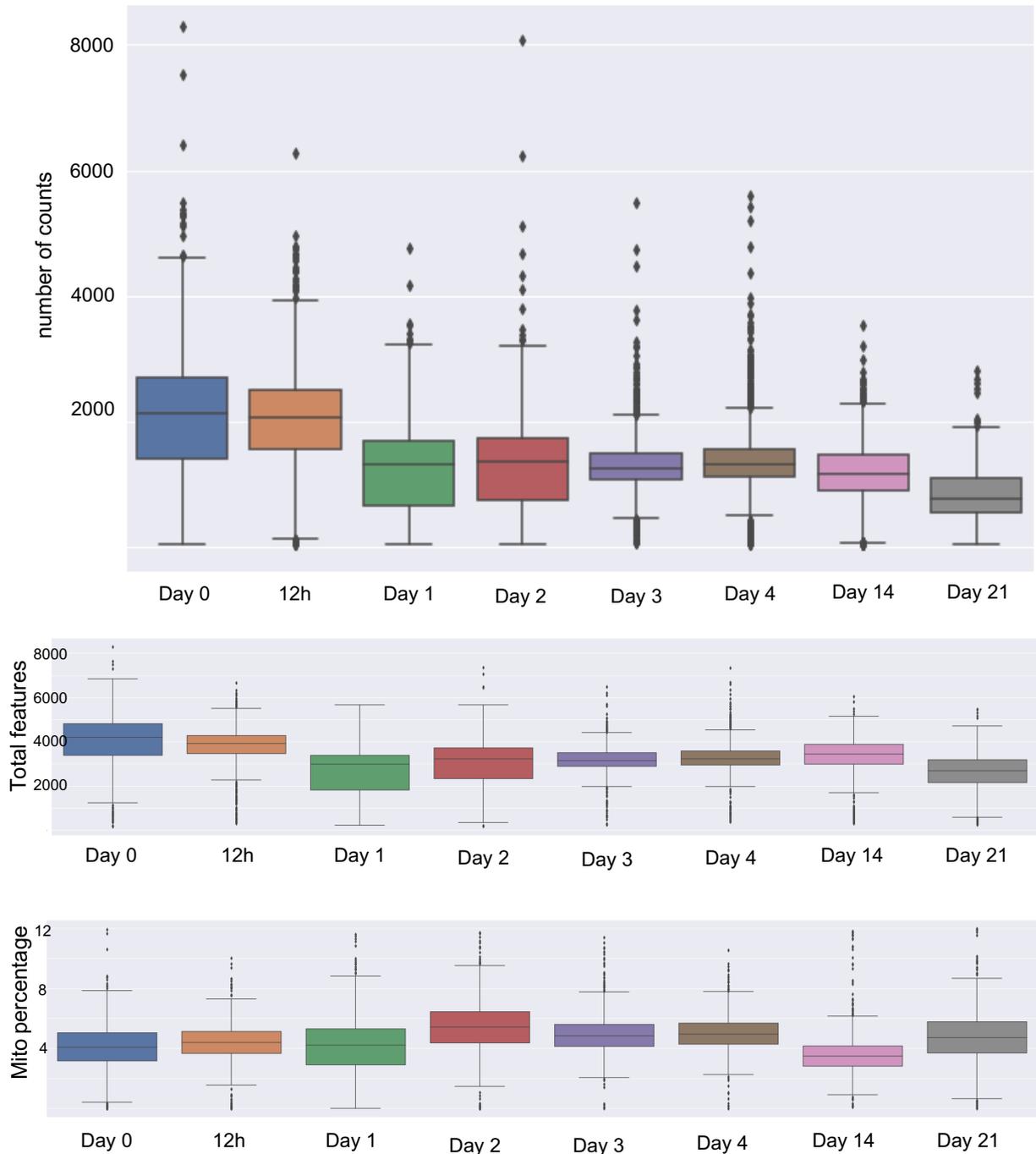


Figure 7. 2: QC report for scRNAseq analysis of NGN2 time course

The QC filtering was done using Scanpy (Wolf, Angerer, & Theis, 2018) and involved removal of cells expressing less than 2000 genes, more than 100000 transcripts or more than 12% mitochondrial reads and genes that had less than 500 UMIs or expressed in less than 10 cells.

7.4 Full list of significant TFs from DASTk analysis

Timepoint	TF	+ MD Score	TF	- MD Score
6 hours	NEUROD1	0.219	ZNF384	-0.05
	OLIG2	0.076	HINFP	-0.107
12 hours	CUX1	0.042	POU5F1	-0.087
	-	-	NANOG	-0.084
	-	-	KLF3	-0.02
1 day	HXA10	0.064	DUX4	-0.04
	EBF1	0.034	REST	-0.038
	-	-	NANOG	-0.052
36 hours	THAP1	0.016	-	-
	LHX2	0.043	-	-
2 days	CUX1	0.087	-	-
	NRF1	0.042	-	-
	CTCF	0.032	-	-
	NEUROD1	0.049	-	-
3 days	PIT1	0.052	ZNF384	-0.079
	PATZ1	0.009	CTCF	-0.047
	SP2	0.017	FOSL2	-0.062
4 days	HXB8	0.065	FOXC1	-0.085
	PAX6	0.059	DBP	-0.056
	ONECUT1	0.046	SP2	-0.009
14 days	ONECUT1	0.219	FOXC1	-0.288
	PBX2	0.131	NEUROD1	-0.201
	CTCF	0.034	MYOD1	-0.072
	POU3F2	0.033	ASCL1	-0.059
	-	-	PAX6	-0.056
	-	-	-	-
21 days	FOXC1	0.19	CUX1	-0.059
	FOSL2	0.186	NEUROD1	-0.083
	JUN	0.201	ONECUT1	-0.079
	BATF	0.074	-	-
	BACH2	0.073	-	-

Table 7. 1: Dominant transcription activity over the course of NGN2 reprogramming based on ATACseq data.

Table listing significant TFs predicted using DASTk analysis (p-value <0.05), for every timepoint from the NGN2 time course. For each timepoint, the +MD score reflects TFs that show increased activity at that particular timepoint, compared to the previous timepoint., while -MD scores reflect

decreased activity. TFs with high activity (absolute MD score of above 0.100 AND p-value < 10⁻⁷) are highlighted in grey. ChIP-validated targets of NGN2 are highlighted in bold.

Timepoint	TF	+ MD Score	TF	- MD Score
4 days with glia	NR1H4	0.058	FOXC1	-0.365
	ZNF384	0.06	ONECUT1	-0.045
	MEF2C	0.054	POU3F2	-0.057
14 days with glia	ZNF384	0.081	CTCF	-0.067
	HNF1A	0.03	POU3F2	-0.075
	-	-	CREB1	-0.093
21 days with glia	ZN384	0.132	FOXC1	-0.237
	CUX1	0.037	JUN	-0.203
	SIX2	0.064	JUND	-0.166
	SIX1	0.058	FOSL2	-0.155

Table 7. 2: Dominant transcription activity between iNs cultured with glia and iNs cultured without glia, based on ATACseq data

Table listing significant TFs predicted using DASTK analysis, for time points where iNs were co-cultured with glia. For each timepoint, the +MD score reflects TFs that show increased activity at that particular timepoint, compared to iNs cultured without glia at the same timepoint., while - MD scores reflect decreased activity. TFs with high activity (absolute MD score of above 0.100 AND p-value < 10⁻⁷) are highlighted in grey.

7.5 Differential accessible region analysis (cumulative version)

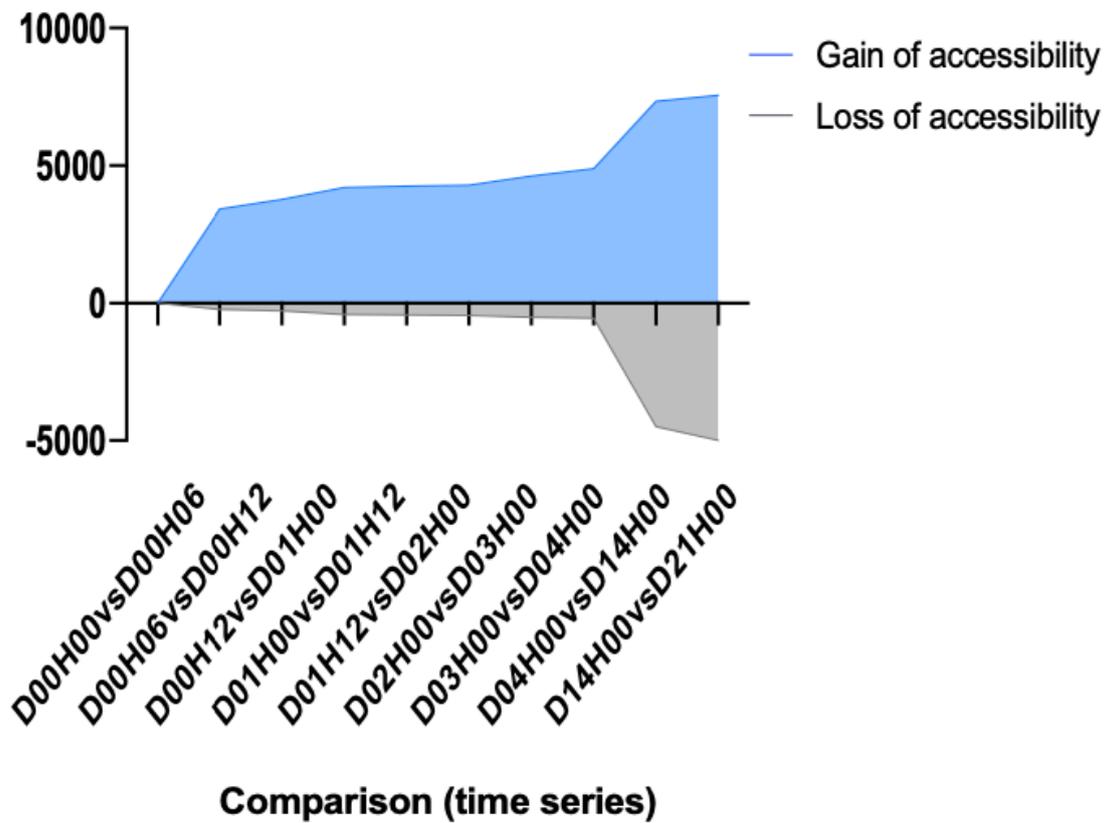


Figure 7. 2:Plot of cumulative counts (frequency) for sites gaining accessibility (opening loci) and sites losing accessibility (closing loci) throughout NGN2 reprogramming (samples without glia).

A cumulative version to Figures 5.3 and 5.4, this plot shows that there is a gradual increase in opening sites between 6 hours and Day 4, and in closing sites from the Day 0 to Day 4. This allows the change between these timepoints to be viewed differently - as a steady opening of sites essential to reprogramming after a rapid and large gain of accessibility triggered by NGN2 by 6 hours. This rate of opening possibly continues at a similar pace between Day 4 and Day 14. Conversely, sites are slowly closed from the onset reprogramming, but a major compaction event occurs between Day 4 and 14.

8 References

- Ahmed, M., Xu, J., & Xu, P. X. (2012). EYA1 and SIX1 drive the neuronal developmental program in cooperation with the SWI/SNF chromatin-remodeling complex and SOX2 in the mammalian inner ear. *Development*, *139*(11), 1965–1977. <https://doi.org/10.1242/dev.071670>
- Allen, N. J., & Eroglu, C. (2017). Cell Biology of Astrocyte-Synapse Interactions. *Neuron*. <https://doi.org/10.1016/j.neuron.2017.09.056>
- Amaral, A. I., Meisingset, T. W., Kotter, M. R., & Sonnewald, U. (2013). Metabolic aspects of Neuron-Oligodendrocyte-Astrocyte interactions. *Frontiers in Endocrinology*. <https://doi.org/10.3389/fendo.2013.00054>
- Aprèa, J., Nonaka-Kinoshita, M., & Calegari, F. (2014). Generation and characterization of Neurod1-CreERT2 mouse lines for the study of embryonic and adult neurogenesis. *Genesis*. <https://doi.org/10.1002/dvg.22797>
- Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., ... Sherlock, G. (2000, May). Gene ontology: Tool for the unification of biology. *Nature Genetics*. NIH Public Access. <https://doi.org/10.1038/75556>
- Aydin, B., Kakumanu, A., Rossillo, M., Moreno-Estellés, M., Garipler, G., Ringstad, N., ... Mazzoni, E. O. (2019). Proneural factors *Ascl1* and *Neurog2* contribute to neuronal subtype identities by establishing distinct chromatin landscapes. *Nature Neuroscience*, *1*. <https://doi.org/10.1038/s41593-019-0399-y>
- Baek, S., Goldstein, I., & Hager, G. L. (2017). Bivariate Genomic Footprinting Detects Changes in Transcription Factor Activity. *Cell Reports*, *19*(8), 1710–1722. <https://doi.org/10.1016/j.celrep.2017.05.003>
- Bane, V., Lehane, M., Dikshit, M., O’Riordan, A., & Furey, A. (2014). Tetrodotoxin: Chemistry, Toxicity, Source, Distribution and Detection. *Toxins*, *6*(2), 693–755. <https://doi.org/10.3390/toxins6020693>
- Berninger, B., Costa, M. R., Koch, U., Schroeder, T., Sutor, B., Grothe, B., & Götz, M. (2007). Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, *27*(32),

- Biffi, E., Regalia, G., Menegon, A., Ferrigno, G., & Pedrocchi, A. (2013). The influence of neuronal density and maturation on network activity of hippocampal cell cultures: a methodological study. *PloS One*, 8(12), e83899. <https://doi.org/10.1371/journal.pone.0083899>
- Blake, J. F., Brown, M. W., & Collingridge, G. L. (1988). CNQX blocks acidic amino acid induced depolarizations and synaptic components mediated by non-NMDA receptors in rat hippocampal slices. *Neuroscience Letters*, 89(2), 182–186. [https://doi.org/10.1016/0304-3940\(88\)90378-3](https://doi.org/10.1016/0304-3940(88)90378-3)
- Blanchard, J. W., Eade, K. T., Szűcs, A., Lo Sardo, V., Tsunemoto, R. K., Williams, D., ... Baldwin, K. K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nature Neuroscience*, 18(1), 25–35. <https://doi.org/10.1038/nn.3887>
- Bottenstein, J. E., & Sato, G. H. (1979). Growth of a rat neuroblastoma cell line in serum-free supplemented medium. *Proceedings of the National Academy of Sciences of the United States of America*, 76(1), 514–517. <https://doi.org/10.1073/pnas.76.1.514>
- Boutin, C., Hardt, O., De Chevigny, A., Coré, N., Goebbels, S., Seidenfaden, R., ... Cremer, H. (2010). NeuroD1 induces terminal neuronal differentiation in olfactory neurogenesis. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0909015107>
- Brakeman, P. R., Lanahan, A. A., O'Brien, R., Roche, K., Barnes, C. A., Huganir, R. L., & Worley, P. F. (1997). Homer: A protein that selectively binds metabotropic glutamate receptors. *Nature*. <https://doi.org/10.1038/386284a0>
- Brunet, E., Simsek, D., Tomishima, M., DeKolver, R., Choi, V. M., Gregory, P., ... Jasin, M. (2009). Chromosomal translocations induced at specified loci in human stem cells. *Proceedings of the National Academy of Sciences of the United States of America*, 106(26), 10620–10625. <https://doi.org/10.1073/pnas.0902076106>
- Buenrostro, J. D., Giresi, P. G., Zaba, L. C., Chang, H. Y., & Greenleaf, W. J. (2013). transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, dnA-binding proteins and nucleosome position. <https://doi.org/10.1038/nmeth.2688>
- Buenrostro, J. D., Wu, B., Litzenburger, U. M., Ruff, D., Gonzales, M. L., Snyder, M. P., ...

- Greenleaf, W. J. (2015). Single-cell chromatin accessibility reveals principles of regulatory variation. *Nature*, *523*(7561), 486–490. <https://doi.org/10.1038/nature14590>
- Burbulla, L. F., Beaumont, K. G., Mrksich, M., & Krainc, D. (2016). Micropatterning Facilitates the Long-Term Growth and Analysis of iPSC-Derived Individual Human Neurons and Neuronal Networks. *Advanced Healthcare Materials*, *5*(15), 1894–1903. <https://doi.org/10.1002/adhm.201500900>
- Busskamp, V., Lewis, N. E., Guye, P., Ng, A. H., Shipman, S. L., Byrne, S. M., ... Church, G. M. (2014). Rapid neurogenesis through transcriptional activation in human stem cells. *Molecular Systems Biology*, *10*(11), 760–760. <https://doi.org/10.15252/msb.20145508>
- 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. <https://doi.org/10.1038/nbt.1529>
- Chanda, S., Ang, C. E., Davila, J., Pak, C., Mall, M., Lee, Q. Y., ... Wernig, M. (2014). Generation of induced neuronal cells by the single reprogramming factor ASCL1. *Stem Cell Reports*, *3*(2), 282–296. <https://doi.org/10.1016/j.stemcr.2014.05.020>
- Chanda, S., Ang, C. E., Lee, Q. Y., Ghebrial, M., Haag, D., Shibuya, Y., ... Südhof, T. C. (2019). Direct Reprogramming of Human Neurons Identifies MARCKSL1 as a Pathogenic Mediator of Valproic Acid-Induced Teratogenicity. *Cell Stem Cell*. <https://doi.org/10.1016/J.STEM.2019.04.021>
- Chang, M. C., Park, J. M., Pelkey, K. A., Grabenstatter, H. L., Xu, D., Linden, D. J., ... Worley, P. F. (2010). Narp regulates homeostatic scaling of excitatory synapses on parvalbumin-expressing interneurons. *Nature Neuroscience*. <https://doi.org/10.1038/nn.2621>
- Cheadle, L., Tzeng, C. P., Kalish, B. T., Harmin, D. A., Rivera, S., Ling, E., ... Greenberg, M. E. (2018). Visual Experience-Dependent Expression of Fn14 Is Required for Retinogeniculate Refinement. *Neuron*, *99*(3), 525-539.e10. <https://doi.org/10.1016/j.neuron.2018.06.036>
- Chen, E. Y., Tan, C. M., Kou, Y., Duan, Q., Wang, Z., Meirelles, G. V., ... Ma'ayan, A. (2013). Enrichr: Interactive and collaborative HTML5 gene list enrichment analysis tool. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-14-128>
- Chen, L., Chetkovich, D. M., Petralia, R. S., Sweeney, N. T., Kawasaki, Y., Wenthold, R. J., ... Nicoll, R. A. (2000). Stargazin regulates synaptic targeting of AMPA receptors by two

- distinct mechanisms. *Nature*, 408(6815), 936–943. <https://doi.org/10.1038/35050030>
- Chen, X., Miragaia, R. J., Natarajan, K. N., & Teichmann, S. A. (2018). A rapid and robust method for single cell chromatin accessibility profiling. *Nature Communications*. <https://doi.org/10.1038/s41467-018-07771-0>
- Chou, C. H., Shrestha, S., Yang, C. D., Chang, N. W., Lin, Y. L., Liao, K. W., ... Huang, H. Da. (2018). MiRTarBase update 2018: A resource for experimentally validated microRNA-target interactions. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkx1067>
- Chouchane, M., Melo de Farias, A. R., Moura, D. M. de S., Hilscher, M. M., Schroeder, T., Leão, R. N., & Costa, M. R. (2017). Lineage Reprogramming of Astroglial Cells from Different Origins into Distinct Neuronal Subtypes. *Stem Cell Reports*, 9(1), 162–176. <https://doi.org/10.1016/j.stemcr.2017.05.009>
- Chowdhury, S., Shepherd, J. D., Okuno, H., Lyford, G., Petralia, R. S., Plath, N., ... Worley, P. F. (2006). Arc/Arg3.1 Interacts with the Endocytic Machinery to Regulate AMPA Receptor Trafficking. *Neuron*. <https://doi.org/10.1016/j.neuron.2006.08.033>
- Chronis, C., Fiziev, P., Papp, B., Butz, S., Bonora, G., Sabri, S., ... Plath, K. (2017). Cooperative Binding of Transcription Factors Orchestrates Reprogramming. *Cell*. <https://doi.org/10.1016/j.cell.2016.12.016>
- Cicchillitti, L., Penci, R., Di Michele, M., Filippetti, F., Rotilio, D., Donati, M. B., ... Ferlini, C. (2008). Proteomic characterization of cytoskeletal and mitochondrial class III beta-tubulin. *Molecular Cancer Therapeutics*, 7(7), 2070–2079. <https://doi.org/10.1158/1535-7163.MCT-07-2370>
- Colasante, G., Lignani, G., Rubio, A., Medrihan, L., Yekhle, L., Sessa, A., ... Broccoli, V. (2015). Rapid Conversion of Fibroblasts into Functional Forebrain GABAergic Interneurons by Direct Genetic Reprogramming. *Cell Stem Cell*, 17(6), 719–734. <https://doi.org/10.1016/j.stem.2015.09.002>
- Colasante, G., Rubio, A., Massimino, L., & Broccoli, V. (2019). Direct Neuronal Reprogramming Reveals Unknown Functions for Known Transcription Factors. *Frontiers in Neuroscience*, 13, 283. <https://doi.org/10.3389/fnins.2019.00283>
- Cole, R., & de Vellis, J. (2001). Preparation of Astrocyte, Oligodendrocyte, and Microglia Cultures from Primary Rat Cerebral Cultures. In *Protocols for Neural Cell Culture* (pp. 117–

- 127). New Jersey: Humana Press. <https://doi.org/10.1385/1-59259-207-4:117>
- Davis, R. L., Weintraub, H., & Lassar, A. B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell*, *51*(6), 987–1000. [https://doi.org/10.1016/0092-8674\(87\)90585-X](https://doi.org/10.1016/0092-8674(87)90585-X)
- Dennis, D. J., Han, S., & Schuurmans, C. (2018). bHLH transcription factors in neural development, disease, and reprogramming. *Brain Research*. <https://doi.org/10.1016/J.BRAINRES.2018.03.013>
- Dixit, R., Wilkinson, G., Cancino, G. I., Shaker, T., Adnani, L., Li, S., ... Schuurmans, C. (2014). Neurog1 and Neurog2 Control Two Waves of Neuronal Differentiation in the Piriform Cortex. *Journal of Neuroscience*, *34*(2), 539–553. <https://doi.org/10.1523/JNEUROSCI.0614-13.2014>
- Dobin, A., Davis, C. A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., ... Gingeras, T. R. (2013). STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/bts635>
- Drouin-Ouellet, J., Lau, S., Brattås, L., Ottosson, D. R., Pircs, K., Grassi, D. A., ... Parmar, M. (2017). REST suppression mediates neural conversion of adult human fibroblasts via microRNA-dependent and -independent pathways. *EMBO Mol Med*, *9*, 1117–1131. <https://doi.org/10.15252/emmm>
- Duclot, F., & Kabbaj, M. (2017). The role of early growth response 1 (EGR1) in brain plasticity and neuropsychiatric disorders. *Frontiers in Behavioral Neuroscience*. <https://doi.org/10.3389/fnbeh.2017.00035>
- Dunn, S. J., Martello, G., Yordanov, B., Emmott, S., & Smith, A. G. (2014). Defining an essential transcription factor program for naïve pluripotency. *Science*, *344*(6188), 1156–1160. <https://doi.org/10.1126/science.1248882>
- Emily Wu, Y., Pan, L., Zuo, Y., Li, X., & Hong, W. (2017). Detecting Activated Cell Populations Using Single-Cell RNA-Seq. *Neuron*, *96*, 313-328.e7. <https://doi.org/10.1016/j.neuron.2017.09.026>
- Eroglu, C., & Barres, B. A. (2010). Regulation of synaptic connectivity by glia. *Nature*, *468*(7321), 223–231. <https://doi.org/10.1038/nature09612>

- Fernandopulle, M. S., Prestil, R., Grunseich, C., Wang, C., Gan, L., & Ward, M. E. (2018). Transcription Factor–Mediated Differentiation of Human iPSCs into Neurons. *Current Protocols in Cell Biology*, 79(1). <https://doi.org/10.1002/cpcb.51>
- Fletcher, T. L., Cameron, P., De Camilli, P., & Banker, G. (1991). The distribution of synapsin I and synaptophysin in hippocampal neurons developing in culture. *The Journal of Neuroscience*, 11(6), 1617–1626. <https://doi.org/10.1103/PhysRevLett.114.173002>
- Florio, M., Leto, K., Muzio, L., Tinterri, A., Badaloni, A., Croci, L., ... Consalez, G. G. (2012). Neurogenin 2 regulates progenitor cell-cycle progression and Purkinje cell dendritogenesis in cerebellar development. *Development (Cambridge, England)*, 139(13), 2308–2320. <https://doi.org/10.1242/dev.075861>
- Fornes, O., Castro-Mondragon, J. A., Khan, A., van der Lee, R., Zhang, X., Richmond, P. A., ... Mathelier, A. (2020). JASPAR 2020: update of the open-access database of transcription factor binding profiles. *Nucleic Acids Research*. <https://doi.org/10.1093/nar/gkz1001>
- Frega, M., Linda, K., Keller, J. M., Gümüş-Akay, G., Mossink, B., Rhijn, J.-R. van, ... Kasri, N. N. (2019). Neuronal network dysfunction in a human model for Kleefstra syndrome mediated by enhanced NMDAR signaling. *BioRxiv*, 585596. <https://doi.org/10.1101/585596>
- Frega, M., Van Gestel, S. H. C., Linda, K., Van Der Raadt, J., Keller, J., Van Rhijn, J. R., ... Kasri, N. N. (2017). Rapid neuronal differentiation of induced pluripotent stem cells for measuring network activity on micro-electrode arrays. *Journal of Visualized Experiments*, (119), e54900–e54900. <https://doi.org/10.3791/54900>
- Gascón, S., Murenu, E., Masserdotti, G., Ortega, F., Russo, G. L., Petrik, D., ... Götz, M. (2016). Identification and Successful Negotiation of a Metabolic Checkpoint in Direct Neuronal Reprogramming. *Cell Stem Cell*, 18(3), 396–409. <https://doi.org/10.1016/j.stem.2015.12.003>
- Gleeson, J. G., Peter T, L., Flanagan, L. A., & Walsh, C. A. (1999). Doublecortin is a microtubule-associated protein and is expressed widely by migrating neurons. *Neuron*. [https://doi.org/10.1016/S0896-6273\(00\)80778-3](https://doi.org/10.1016/S0896-6273(00)80778-3)
- Golonzhka, O., Nord, A., Tang, P. L. F., Lindtner, S., Ypsilanti, A. R., Ferretti, E., ... Rubenstein, J. L. R. (2015). Pbx Regulates Patterning of the Cerebral Cortex in Progenitors and Postmitotic Neurons. *Neuron*, 88(6), 1192–1207.

<https://doi.org/10.1016/j.neuron.2015.10.045>

- Götz, M., Stoykova, A., & Gruss, P. (1998). Pax6 controls radial glia differentiation in the cerebral cortex. *Neuron*, *21*(5), 1031–1044. [https://doi.org/10.1016/S0896-6273\(00\)80621-2](https://doi.org/10.1016/S0896-6273(00)80621-2)
- Guillemot, F., & Hassan, B. A. (2017). Beyond proneural: emerging functions and regulations of proneural proteins. *Current Opinion in Neurobiology*, *42*, 93–101. <https://doi.org/10.1016/J.CONB.2016.11.011>
- Guo, C., Morris, S. A., Wang, J., & Esteban, M. (2017). Engineering cell identity: establishing new gene regulatory and chromatin landscapes This review comes from a themed issue on Cell reprogramming. *Current Opinion in Genetics & Development*, *46*, 50–57. <https://doi.org/10.1016/j.gde.2017.06.011>
- Gurdon, J. B. (1962). The developmental capacity of nuclei taken from intestinal epithelium cells of feeding tadpoles. *Journal of Embryology and Experimental Morphology*, *10*, 622–640. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/13951335>
- Guthrie, S. (2007). Patterning and axon guidance of cranial motor neurons. *Nature Reviews Neuroscience*. <https://doi.org/10.1038/nrn2254>
- H, L., B, H., A, W., T, F., J, R., N, H., ... R, D. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16), 2078–2079.
- Hardwick, L. J. A., & Philpott, A. (2015). Multi-site phosphorylation regulates NeuroD4 activity during primary neurogenesis: a conserved mechanism amongst proneural proteins. *Neural Development*, *10*(1), 15. <https://doi.org/10.1186/s13064-015-0044-8>
- Heikkilä, T. J., Ylä-Outinen, L., Tanskanen, J. M. A., Lappalainen, R. S., Skottman, H., Suuronen, R., ... Narkilahti, S. (2009). Human embryonic stem cell-derived neuronal cells form spontaneously active neuronal networks in vitro. *Experimental Neurology*, *218*(1), 109–116. <https://doi.org/10.1016/j.expneurol.2009.04.011>
- Heinrich, C., Blum, R., Gascón, S., Masserdotti, G., Tripathi, P., Sánchez, R., ... Berninger, B. (2010). Directing Astroglia from the Cerebral Cortex into Subtype Specific Functional Neurons. *PLoS Biology*, *8*(5), e1000373. <https://doi.org/10.1371/journal.pbio.1000373>
- Heins, N., Malatesta, P., Ceconi, F., Nakafuku, M., Tucker, K. L., Hack, M. A., ... Götz, M. (2002). Glial cells generate neurons: the role of the transcription factor Pax6. *Nature*

- Neuroscience*, 5(4), 308–315. <https://doi.org/10.1038/nn828>
- Heng, J. I.-T., Nguyen, L., Castro, D. S., Zimmer, C., Wildner, H., Armant, O., ... Guillemot, F. (2008). Neurogenin 2 controls cortical neuron migration through regulation of Rnd2. *Nature*, 455(7209), 114–118. <https://doi.org/10.1038/nature07198>
- Henikoff, J. G., Belsky, J. A., Krassovsky, K., MacAlpine, D. M., & Henikoff, S. (2011). Epigenome characterization at single base-pair resolution. *Proceedings of the National Academy of Sciences of the United States of America*, 108(45), 18318–18323. <https://doi.org/10.1073/pnas.1110731108>
- Ho, S.-M., Hartley, B. J., TCW, J., Beaumont, M., Stafford, K., Slesinger, P. A., & Brennand, K. J. (2016). Rapid Ngn2-induction of excitatory neurons from hiPSC-derived neural progenitor cells. *Methods*, 101, 113–124. <https://doi.org/10.1016/j.ymeth.2015.11.019>
- Hodge, R. D., Kahoud, R. J., & Hevner, R. F. (2012). Transcriptional control of glutamatergic differentiation during adult neurogenesis. *Cellular and Molecular Life Sciences: CMLS*, 69(13), 2125–2134. <https://doi.org/10.1007/s00018-011-0916-y>
- Hol, E. M., & Pekny, M. (2015). Glial fibrillary acidic protein (GFAP) and the astrocyte intermediate filament system in diseases of the central nervous system. *Current Opinion in Cell Biology*, 32, 121–130. <https://doi.org/10.1016/J.CEB.2015.02.004>
- Iwafuchi-Doi, M., & Zaret, K. S. (2016). Cell fate control by pioneer transcription factors. *Development (Cambridge)*. <https://doi.org/10.1242/dev.133900>
- Jiang, C., & Pugh, B. F. (2009, March). Nucleosome positioning and gene regulation: Advances through genomics. *Nature Reviews Genetics*. <https://doi.org/10.1038/nrg2522>
- Kabayiza, K. U., Masgutova, G., Harris, A., Rucchin, V., Jacob, B., & Clotman, F. (2017). The Onecut Transcription Factors Regulate Differentiation and Distribution of Dorsal Interneurons during Spinal Cord Development. *Frontiers in Molecular Neuroscience*, 10, 157. <https://doi.org/10.3389/fnmol.2017.00157>
- Kirwan, P., Turner-Bridger, B., Peter, M., Momoh, A., Arambepola, D., Robinson, H. P. C., & Livesey, F. J. (2016). Development and function of human cerebral cortex neural networks from pluripotent stem cells in vitro. <https://doi.org/10.1242/dev.123851>
- Kovach, C., Dixit, R., Li, S., Mattar, P., Wilkinson, G., Elsen, G. E., ... Schuurmans, C. (2013).

- Neurog2 Simultaneously Activates and Represses Alternative Gene Expression Programs in the Developing Neocortex. *Cerebral Cortex*, 23(8), 1884–1900. <https://doi.org/10.1093/cercor/bhs176>
- Kowalchuk, A. M., Maurer, K. A., Shoja-Taheri, F., & Brown, N. L. (2018). Requirements for Neurogenin2 during mouse postnatal retinal neurogenesis. *Developmental Biology*, 442(2), 220–235. <https://doi.org/10.1016/J.YDBIO.2018.07.020>
- Kulakovskiy, I. V., Vorontsov, I. E., Yevshin, I. S., Sharipov, R. N., Fedorova, A. D., Rumynskiy, E. I., ... Makeev, V. J. (2018). HOCOMOCO: Towards a complete collection of transcription factor binding models for human and mouse via large-scale ChIP-Seq analysis. *Nucleic Acids Research*, 46(D1), D252–D259. <https://doi.org/10.1093/nar/gkx1106>
- Ladewig, J., Mertens, J., Kesavan, J., Doerr, J., Poppe, D., Glaue, F., ... Brüstle, O. (2012). Small molecules enable highly efficient neuronal conversion of human fibroblasts. *Nature Methods*, 9(6), 575–578. <https://doi.org/10.1038/nmeth.1972>
- Lai, W. K. M., & Pugh, B. F. (2017). Understanding nucleosome dynamics and their links to gene expression and DNA replication. *Nature Reviews Molecular Cell Biology*, 18(9), 548–562. <https://doi.org/10.1038/nrm.2017.47>
- Lam, R. S., Töpfer, F. M., Wood, P. G., Busskamp, V., Bamberg, E., & Wernig, M. (2017). Functional Maturation of Human Stem Cell-Derived Neurons in Long-Term Cultures. *PLOS ONE*, 12(1), e0169506. <https://doi.org/10.1371/journal.pone.0169506>
- Langfelder, P., & Horvath, S. (2008a). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9(1), 559. <https://doi.org/10.1186/1471-2105-9-559>
- Langfelder, P., & Horvath, S. (2008b). WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-9-559>
- Langfelder, P., Zhang, B., & Horvath, S. (2008). Defining clusters from a hierarchical cluster tree: The Dynamic Tree Cut package for R. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btm563>
- Lee, S.-K., Lee, B., Ruiz, E. C., & Pfaff, S. L. (2005). Olig2 and Ngn2 function in opposition to modulate gene expression in motor neuron progenitor cells. *Genes & Development*, 19(2), 282. <https://doi.org/10.1101/GAD.1257105>

- Levine, J. H., Simonds, E. F., Bendall, S. C., Davis, K. L., Amir, E. D., Tadmor, M. D., ... Nolan, G. P. (2015). Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell*, *162*(1), 184–197. <https://doi.org/10.1016/J.CELL.2015.05.047>
- Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. <https://doi.org/10.1093/bioinformatics/btp324>
- Li, L., Carter, J., Gao, X., Whitehead, J., & Tourtellotte, W. G. (2005). The Neuroplasticity-Associated Arc Gene Is a Direct Transcriptional Target of Early Growth Response (Egr) Transcription Factors. *Molecular and Cellular Biology*. <https://doi.org/10.1128/mcb.25.23.10286-10300.2005>
- Liu, M.-L., Zang, T., & Zhang, C.-L. (n.d.). Direct lineage reprogramming reveals disease-specific phenotypes of motor neurons from human ALS patients. <https://doi.org/10.1016/j.celrep.2015.12.018>
- Liu, M.-L., Zang, T., & Zhang, C.-L. (2016). Direct Lineage Reprogramming Reveals Disease-Specific Phenotypes of Motor Neurons from Human ALS Patients. *Cell Reports*, *14*(1), 115–128. <https://doi.org/10.1016/j.celrep.2015.12.018>
- Liu, M.-L., Zang, T., Zou, Y., Chang, J. C., Gibson, J. R., Huber, K. M., & Zhang, C.-L. (2013). ARTICLE Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons. <https://doi.org/10.1038/ncomms3183>
- Liu, M. L., Zang, T., & Zhang, C. L. (2016). Direct Lineage Reprogramming Reveals Disease-Specific Phenotypes of Motor Neurons from Human ALS Patients. *Cell Reports*, *14*(1), 115–128. <https://doi.org/10.1016/j.celrep.2015.12.018>
- Liu, X., Li, F., Stubblefield, E. A., Blanchard, B., Richards, T. L., Larson, G. A., ... Li, C. Y. (2012). Direct reprogramming of human fibroblasts into dopaminergic neuron-like cells. *Cell Research*, *22*(2), 321–332. <https://doi.org/10.1038/cr.2011.181>
- Long, J. E., Cobos, I., Potter, G. B., & Rubenstein, J. L. R. (2009). Dlx1&2 and Mash1 transcription factors control MGE and CGE patterning and differentiation through parallel and overlapping pathways. *Cerebral Cortex (New York, N.Y. : 1991)*, *19 Suppl 1*(Suppl 1), i96-106. <https://doi.org/10.1093/cercor/bhp045>
- Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion

- for RNA-seq data with DESeq2. *Genome Biology*. <https://doi.org/10.1186/s13059-014-0550-8>
- Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.Journal*. <https://doi.org/10.14806/ej.17.1.200>
- Masserdotti, G., Gillotin, S., Sutor, B., Drechsel, D., Irmeler, M., Jørgensen, H. F., ... Götz, M. (2015). Transcriptional Mechanisms of Proneural Factors and REST in Regulating Neuronal Reprogramming of Astrocytes. *Cell Stem Cell*, 17(1), 74–88. <https://doi.org/10.1016/j.stem.2015.05.014>
- Mattar, P., Britz, O., Johannes, C., Nieto, M., Ma, L., Rebeyka, A., ... Schuurmans, C. (2004). A screen for downstream effectors of Neurogenin2 in the embryonic neocortex. *Developmental Biology*, 273(2), 373–389. <https://doi.org/10.1016/J.YDBIO.2004.06.013>
- Mattar, P., Langevin, L. M., Markham, K., Klenin, N., Shivji, S., Zinyk, D., & Schuurmans, C. (2008). Basic helix-loop-helix transcription factors cooperate to specify a cortical projection neuron identity. *Molecular and Cellular Biology*, 28(5), 1456–1469. <https://doi.org/10.1128/MCB.01510-07>
- McCarthy, K. D., & de Vellis, J. (1980). Preparation of separate astroglial and oligodendroglial cell cultures from rat cerebral tissue. *The Journal of Cell Biology*, 85(3), 890–902. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/6248568>
- McInnes, L., Healy, J., Saul, N., & Großberger, L. (2018). UMAP: Uniform Manifold Approximation and Projection. *Journal of Open Source Software*. <https://doi.org/10.21105/joss.00861>
- Miller, J. A., Cai, C., Langfelder, P., Geschwind, D. H., Kurian, S. M., Salomon, D. R., & Horvath, S. (2011). Strategies for aggregating gene expression data: The collapseRows R function. *BMC Bioinformatics*. <https://doi.org/10.1186/1471-2105-12-322>
- Nehme, R., Zuccaro, E., Dia Ghosh, S., Fu, Z., Feng, G., & Egan Correspondence, K. (2018). Combining NGN2 Programming with Developmental Patterning Generates Human Excitatory Neurons with NMDAR-Mediated Synaptic Transmission Data and Software Availability GSE112732. <https://doi.org/10.1016/j.celrep.2018.04.066>
- Nieto, M., Monuki, E. S., Tang, H., Imitola, J., Haubst, N., Khoury, S. J., ... Walsh, C. A. (2004). Expression of Cux-1 and Cux-2 in the subventricular zone and upper layers II-IV of the

- cerebral cortex. *The Journal of Comparative Neurology*, 479(2), 168–180. <https://doi.org/10.1002/cne.20322>
- Odawara, A., Katoh, H., Matsuda, N., & Suzuki, I. (2016). Physiological maturation and drug responses of human induced pluripotent stem cell-derived cortical neuronal networks in long-term culture. *Nature Publishing Group*. <https://doi.org/10.1038/srep26181>
- Odawara, A., Saitoh, Y., Alhebshi, A. H., Gotoh, M., & Suzuki, I. (2014). Long-term electrophysiological activity and pharmacological response of a human induced pluripotent stem cell-derived neuron and astrocyte co-culture. <https://doi.org/10.1016/j.bbrc.2013.12.142>
- Parras, C. M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D. J., & Guillemot, F. (2002). Divergent functions of the proneural genes *Mash1* and *Ngn2* in the specification of neuronal subtype identity. *Genes & Development*, 16(3), 324–338. <https://doi.org/10.1101/gad.940902>
- Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*. <https://doi.org/10.1038/nmeth.4197>
- Pawlowski, M., Ortmann, D., Bertero, A., Tavares, J. M., Pedersen, R. A., Vallier, L., ... al., et. (2017). Inducible and Deterministic Forward Programming of Human Pluripotent Stem Cells into Neurons, Skeletal Myocytes, and Oligodendrocytes. *Stem Cell Reports*, 455(0), 627–632. <https://doi.org/10.1016/j.stemcr.2017.02.016>
- Petryniak, M. A., Potter, G. B., Rowitch, D. H., & Rubenstein, J. L. R. (2007). *Dlx1* and *Dlx2* control neuronal versus oligodendroglial cell fate acquisition in the developing forebrain. *Neuron*, 55(3), 417–433. <https://doi.org/10.1016/j.neuron.2007.06.036>
- Pfisterer, U., Kirkeby, A., Torper, O., Wood, J., Nelander, J., Dufour, A., ... Parmar, M. (2011). Direct conversion of human fibroblasts to dopaminergic neurons. *Proceedings of the National Academy of Sciences of the United States of America*, 108(25), 10343–10348. <https://doi.org/10.1073/pnas.1105135108>
- Qiu, J., Dando, O., Baxter, P. S., Hasel, P., Heron, S., Simpson, T. I., & Hardingham, G. E. (2018). Mixed-species RNA-seq for elucidation of non-cell-autonomous control of gene transcription. *Nature Protocols*, 13(10), 2176–2199. <https://doi.org/10.1038/s41596-018->

- Rivetti di Val Cervo, P., Romanov, R. A., Spigolon, G., Masini, D., Martín-Montañez, E., Toledo, E. M., ... Arenas, E. (2017). Induction of functional dopamine neurons from human astrocytes in vitro and mouse astrocytes in a Parkinson's disease model. *Nature Biotechnology*, *35*(5), 444–452. <https://doi.org/10.1038/nbt.3835>
- Rodríguez-Tornos, F. M., Briz, C. G., Weiss, L. A., Sebastián-Serrano, A., Ares, S., Navarrete, M., ... Nieto, M. (2016). Cux1 Enables Interhemispheric Connections of Layer II/III Neurons by Regulating Kv1-Dependent Firing. *Neuron*, *89*(3), 494–506. <https://doi.org/10.1016/j.neuron.2015.12.020>
- Roy, A., Francius, C., Rousso, D. L., Seuntjens, E., Debruyne, J., Luxenhofer, G., ... Clotman, F. (2012). Onecut transcription factors act upstream of Isl1 to regulate spinal motoneuron diversification. *Development (Cambridge, England)*, *139*(17), 3109–3119. <https://doi.org/10.1242/dev.078501>
- Sato, S., Yajima, H., Furuta, Y., Ikeda, K., & Kawakami, K. (2015). Activation of Six1 Expression in Vertebrate Sensory Neurons. <https://doi.org/10.1371/journal.pone.0136666>
- Scardigli, R., Schuurmans, C., Gradwohl, G., & Guillemot, F. (2001). Crossregulation between Neurogenin2 and Pathways Specifying Neuronal Identity in the Spinal Cord. *Neuron*, *31*(2), 203–217. [https://doi.org/10.1016/S0896-6273\(01\)00358-0](https://doi.org/10.1016/S0896-6273(01)00358-0)
- Schambach, A., Zychlinski, D., Ehrnstroem, B., & Baum, C. (2013). Biosafety Features of Lentiviral Vectors. *Human Gene Therapy*, *24*(2), 132–142. <https://doi.org/10.1089/hum.2012.229>
- Schep, A. N., Buenrostro, J. D., Denny, S. K., Schwartz, K., Sherlock, G., & Greenleaf, W. J. (2015). Structured nucleosome fingerprints enable high-resolution mapping of chromatin architecture within regulatory regions. *Genome Research*, *25*(11), 1757–1770. <https://doi.org/10.1101/gr.192294.115>
- Schiebinger, G., Shu, J., Tabaka, M., Cleary, B., Subramanian, V., Solomon, A., ... Lander, E. S. (2019). Optimal-Transport Analysis of Single-Cell Gene Expression Identifies Developmental Trajectories in Reprogramming. *Cell*, *176*(4), 928–943.e22. <https://doi.org/10.1016/j.cell.2019.01.006>
- Shi, Y., Kirwan, P., & Livesey, F. J. (2012). Directed differentiation of human pluripotent stem

- cells to cerebral cortex neurons and neural networks. *Nature Protocols*, 7(10), 1836–1846. <https://doi.org/10.1038/nprot.2012.116>
- Shirao, T., & González-Billault, C. (2013). Actin filaments and microtubules in dendritic spines. *Journal of Neurochemistry*. <https://doi.org/10.1111/jnc.12313>
- Smith, D. K., Yang, J., Liu, M.-L., & Zhang, C.-L. (2016). Small Molecules Modulate Chromatin Accessibility to Promote NEUROG2-Mediated Fibroblast-to-Neuron Reprogramming. *Stem Cell Reports*, 7, 955–969. <https://doi.org/10.1016/j.stemcr.2016.09.013>
- Son, E. Y., Ichida, J. K., Wainger, B. J., Toma, J. S., Rafuse, V. F., Woolf, C. J., & Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell*, 9(3), 205–218. <https://doi.org/10.1016/j.stem.2011.07.014>
- Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: Transcript-level estimates improve gene-level inferences [version 2; referees: 2 approved]. *F1000Research*. <https://doi.org/10.12688/F1000RESEARCH.7563.2>
- Stone, N. R., Gifford, C. A., Thomas, R., Pratt, K. J. B., Samse-Knapp, K., Mohamed, T. M. A., ... Srivastava, D. (2019). Context-Specific Transcription Factor Functions Regulate Epigenomic and Transcriptional Dynamics during Cardiac Reprogramming. *Cell Stem Cell*, 25(1), 87-102.e9. <https://doi.org/10.1016/J.STEM.2019.06.012>
- Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... Mesirov, J. P. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*. <https://doi.org/10.1073/pnas.0506580102>
- Sun, A. X., Yuan, Q., Tan, S., Xiao, Y., Wang, D., Khoo, A. T. T., ... Je, H. S. (2016). Direct Induction and Functional Maturation of Forebrain GABAergic Neurons from Human Pluripotent Stem Cells. *Cell Reports*, 16(7), 1942–1953. <https://doi.org/10.1016/j.celrep.2016.07.035>
- Syed, Y. A., Baer, A. S., Lubec, G., Hoeger, H., Widhalm, G., & Kotter, M. R. (2008). Inhibition of oligodendrocyte precursor cell differentiation by myelin-associated proteins. *Neurosurgical Focus*, 24(3–4), E5. <https://doi.org/10.3171/FOC/2008/24/3-4/E4>
- 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.

<https://doi.org/10.1016/j.cell.2006.07.024>

- Taylor, S. M., & Jones, P. A. (1979). Multiple new phenotypes induced in 10T1/2 and 3T3 cells treated with 5-azacytidine. *Cell*, *17*(4), 771–779. [https://doi.org/10.1016/0092-8674\(79\)90317-9](https://doi.org/10.1016/0092-8674(79)90317-9)
- The Gene Ontology Consortium. (2019). The Gene Ontology Resource: 20 years and still GOing strong. *Nucleic Acids Research*, *47*(D1), D330–D338. <https://doi.org/10.1093/nar/gky1055>
- Thomson, J. A. (1998). Embryonic stem cell lines derived from human blastocysts. *Science*. <https://doi.org/10.1126/science.282.5391.1145>
- Tian, B., Yang, J., & Brasier, A. R. (2012). Two-step cross-linking for analysis of protein-chromatin interactions. *Methods in Molecular Biology*, *809*, 105–120. https://doi.org/10.1007/978-1-61779-376-9_7
- Tischler, G., & Leonard, S. (2014). Biobambam: Tools for read pair collation based algorithms on BAM files. *Source Code for Biology and Medicine*. <https://doi.org/10.1186/1751-0473-9-13>
- Toresson, H., Parmar, M., & Campbell, K. (2000). Expression of Meis and Pbx genes and their protein products in the developing telencephalon: Implications for regional differentiation. *Mechanisms of Development*, *94*(1–2), 183–187. [https://doi.org/10.1016/S0925-4773\(00\)00324-5](https://doi.org/10.1016/S0925-4773(00)00324-5)
- Traag, V. A., Waltman, L., & van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*. <https://doi.org/10.1038/s41598-019-41695-z>
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., ... Rinn, J. L. (2014a). Pseudo-temporal ordering of individual cells reveals dynamics and regulators of cell fate decisions. *Nature Biotechnology*, *32*(4), 381. <https://doi.org/10.1038/NBT.2859>
- Trapnell, C., Cacchiarelli, D., Grimsby, J., Pokharel, P., Li, S., Morse, M., ... Rinn, J. L. (2014b). The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nature Biotechnology*. <https://doi.org/10.1038/nbt.2859>
- Treutlein, B., Lee, Q. Y., Camp, J. G., Mall, M., Koh, W., Ali, S., ... Quake, S. R. (2016). Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. <https://doi.org/10.1038/nature18323>

- Tripodi, I. J., Allen, M. A., & Dowell, R. D. (2018). Detecting Differential Transcription Factor Activity from ATAC-Seq Data. *Molecules : A Journal of Synthetic Chemistry and Natural Product Chemistry*, 23(5). <https://doi.org/10.3390/MOLECULES23051136>
- Tuvikene, J., Pruunsild, P., Orav, E., Esvald, E. E., & Timmusk, T. (2016). AP-1 transcription factors mediate BDNF-positive feedback loop in cortical neurons. *Journal of Neuroscience*, 36(4), 1290–1305. <https://doi.org/10.1523/JNEUROSCI.3360-15.2016>
- van der Raadt, J., van Gestel, S. H. C., Nadif Kasri, N., & Albers, C. A. (2019). ONECUT transcription factors induce neuronal characteristics and remodel chromatin accessibility. *Nucleic Acids Research*, 47(11), 5587–5602. <https://doi.org/10.1093/nar/gkz273>
- Van Den Brink, S. C., Sage, F., Vértesy, Á., Spanjaard, B., Peterson-Maduro, J., Baron, C. S., ... Van Oudenaarden, A. (2017, October 1). Single-cell sequencing reveals dissociation-induced gene expression in tissue subpopulations. *Nature Methods*. Nature Publishing Group. <https://doi.org/10.1038/nmeth.4437>
- Velasco, S., Ibrahim, M. M., Kakumanu, A., Garipler, G., Aydin, B., Al-Sayegh, M. A., ... Mazzoni, E. O. (2017). A Multi-step Transcriptional and Chromatin State Cascade Underlies Motor Neuron Programming from Embryonic Stem Cells. *Cell Stem Cell*, 20(2), 205-217.e8. <https://doi.org/10.1016/j.stem.2016.11.006>
- Velasco, S., Ibrahim, M. M., Ohler, U., Mahony, S., Correspondence, E. O. M., Kakumanu, A., ... Mazzoni, E. O. (2017). A Multi-step Transcriptional and Chromatin State Cascade Underlies Motor Neuron Programming from Embryonic Stem Cells. *Cell Stem Cell*, 20, 205–217. <https://doi.org/10.1016/j.stem.2016.11.006>
- Verkhatsky, A., Matteoli, M., Parpura, V., Mothet, J.-P., & Zorec, R. (2016). Astrocytes as secretory cells of the central nervous system: idiosyncrasies of vesicular secretion. *The EMBO Journal*, 35(3), 239–257. <https://doi.org/10.15252/emboj.201592705>
- Vierbuchen, T., Ostermeier, A., Pang, Z. P., Kokubu, Y., Südhof, T. C., & Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*, 463(7284), 1035–1041. <https://doi.org/10.1038/nature08797>
- Wagenaar, D. A., Pine, J., & Potter, S. M. (2006). An extremely rich repertoire of bursting patterns during the development of cortical cultures. *BMC Neuroscience*, 7, 11. <https://doi.org/10.1186/1471-2202-7-11>

- Wang, C., Fong, H., & Huang, Y. (2015). Direct Reprogramming of RESTing Astrocytes. *Cell Stem Cell*, *17*(1), 1–3. <https://doi.org/10.1016/J.STEM.2015.06.011>
- Wapinski, O. L., Lee, Q. Y., Chen, A. C., Li, R., Corces, M. R., Ang, C. E., ... Chang, H. Y. (2017). Rapid Chromatin Switch in the Direct Reprogramming of Fibroblasts to Neurons. *Cell Reports*, *20*(13), 3236–3247. <https://doi.org/10.1016/j.celrep.2017.09.011>
- Wapinski, O. L., Vierbuchen, T., Qu, K., Lee, Q. Y., Chanda, S., Fuentes, D. R., ... Wernig, M. (2013). Hierarchical Mechanisms for Direct Reprogramming of Fibroblasts to Neurons. *Cell*, *155*, 621–635. <https://doi.org/10.1016/j.cell.2013.09.028>
- Wolf, F. A., Angerer, P., & Theis, F. J. (2018). SCANPY: large-scale single-cell gene expression data analysis. *Genome Biology*, *19*(1), 15. <https://doi.org/10.1186/s13059-017-1382-0>
- Xiao, B., Tu, J. C., Petralia, R. S., Yuan, J. P., Doan, A., Breder, C. D., ... Worley, P. F. (1998). Homer regulates the association of group 1 metabotropic glutamate receptors with multivalent complexes of Homer-related, synaptic proteins. *Neuron*. [https://doi.org/10.1016/S0896-6273\(00\)80588-7](https://doi.org/10.1016/S0896-6273(00)80588-7)
- Yajima, H., Suzuki, M., Ochi, H., Ikeda, K., Sato, S., Yamamura, K. ichi, ... Kawakami, K. (2014). Six1 is a key regulator of the developmental and evolutionary architecture of sensory neurons in craniates. *BMC Biology*, *12*(1), 40. <https://doi.org/10.1186/1741-7007-12-40>
- Yang, N., Chanda, S., Marro, S., Ng, Y. H., Janas, J. A., Haag, D., ... Wernig, M. (2017). Generation of pure GABAergic neurons by transcription factor programming. *Nature Methods*. <https://doi.org/10.1038/nmeth.4291>
- Yap, E.-L., & Greenberg, M. E. (2018). Activity-Regulated Transcription: Bridging the Gap between Neural Activity and Behavior. *Neuron*, *100*(2), 330–348. <https://doi.org/10.1016/j.neuron.2018.10.013>
- Yoo, A. S., Sun, A. X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., ... Crabtree, G. R. (2011, August 11). MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature*. <https://doi.org/10.1038/nature10323>
- Yusa, K., Rashid, S. T., Strick-Marchand, H., Varela, I., Liu, P. Q., Paschon, D. E., ... Vallier, L. (2011). Targeted gene correction of α 1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, *478*(7369), 391–394. <https://doi.org/10.1038/nature10424>

- Zhang, B., & Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical Applications in Genetics and Molecular Biology*. <https://doi.org/10.2202/1544-6115.1128>
- Zhang, X.-H., Tee, L. Y., Wang, X.-G., Huang, Q.-S., & Yang, S.-H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Molecular Therapy. Nucleic Acids*, 4, e264. <https://doi.org/10.1038/mtna.2015.37>
- Zhang, Yingsha, Pak, C. H., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., ... Südhof, T. C. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron*, 78(5), 785–798. <https://doi.org/10.1016/j.neuron.2013.05.029>
- Zhang, Yong, Liu, T., Meyer, C. A., Eeckhoute, J., Johnson, D. S., Bernstein, B. E., ... Shirley, X. S. (2008). Model-based analysis of ChIP-Seq (MACS). *Genome Biology*. <https://doi.org/10.1186/gb-2008-9-9-r137>