

# UC San Diego

## UC San Diego Electronic Theses and Dissertations

### Title

Deciphering Transcriptional Control of Neuronal Identity and Diversity Using Direct Reprogramming

### Permalink

<https://escholarship.org/uc/item/59k3t2xt>

### Author

Tsunemoto, Rachel

### Publication Date

2016

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA, SAN DIEGO

Deciphering Transcriptional Control of Neuronal Identity and Diversity  
Using Direct Reprogramming

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Neurosciences

by

Rachel Kuen Tsunemoto

Committee in charge:

Professor Kristin Baldwin, Chair  
Professor Andrew Huberman, Co-Chair  
Professor Alysson Muotri  
Professor Samuel Pfaff  
Professor Nicholas Spitzer

2016



The Dissertation of Rachel Kuen Tsunemoto is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

---

---

---

---

Co-Chair

---

Chair

University of California, San Diego

2016

## DEDICATION

This dissertation is dedicated to my parents, Steve and Kit Mui, my sister, Hannah, and my partner in life, Yarin. Thank you for your unwavering support through it all.

## TABLE OF CONTENTS

Signature Page .....	iii
Dedication .....	iv
Table of Contents .....	v
List of Abbreviations .....	vii
List of Figures .....	x
List of Tables .....	xiii
Acknowledgements .....	xiv
Vita .....	xv
Abstract of the Dissertation .....	xvi
Chapter 1: Introduction and Background .....	1
1.1 Background .....	1
1.1.1 Master transcriptional regulators of cellular identity .....	1
1.1.2 Proneural genes and neuronal subtype diversity .....	2
1.1.3 Direct generation of neurons from other cell types and lineages using transcription factor-mediated transdifferentiation .....	3
1.1.4 Micro-RNA-mediated repression also converts fibroblasts into induced neurons .....	4
1.1.5 Biasing direct reprogramming to produce specific neuronal subtypes .....	5
Chapter 2: Direct reprogramming of mouse fibroblasts into functional neurons using <i>Neurogenin1</i> and <i>Neurogenin2</i> .....	15
2.1 Introduction .....	15
2.2 Methods: Testing candidate transcription factors to generate a specific neuronal subtype .....	16
2.3 Results .....	18
2.3.1 Transient overexpression of either <i>Ngn1</i> or <i>Ngn2</i> with <i>Brn2</i> and <i>Zic1</i> reprograms mouse fibroblasts to express pan-neuronal markers .....	18
2.3.2 Induced neurons generated by <i>Ngn1-Brn2-Zic1</i> and <i>Ngn2-Brn2-Zic1</i> exhibit physiological properties of functional neurons .....	23
2.3.3 Induced neurons generated using <i>Ngn1</i> and <i>Ngn2</i> differ in expression of cortical neuron markers compared to induced neurons reprogrammed using <i>Ascl1</i> .....	27
2.4 Discussion .....	31
Chapter 3: Identification of novel neuronal reprogramming factors using a large-scale screen .....	36
3.1 Introduction .....	36
3.2 Methods: Large-scale screening to identify transcription factor pairs capable of producing neurons directly from fibroblasts .....	36

3.3	Results .....	39
3.3.1	Screen reveals novel bHLH and POU transcription factor pairs capable of reprogramming fibroblasts into candidate induced neurons .....	39
3.3.2	Whole-transcriptome analysis of candidate iN populations .....	49
3.4	Discussion .....	64
Chapter 4: Functional and transcriptional diversity of induced neuron populations.....		68
4.1	Introduction .....	68
4.2	Methods.....	68
4.2.1	Functional analysis using whole-cell patch clamp technique.....	68
4.2.2	Weighted gene coexpression network analysis .....	69
4.3	Results.....	69
4.3.1	iN populations exhibit functional and diverse electrophysiological properties.....	69
4.3.2	Identification of genes selectively expressed in individual iN populations .....	74
4.3.3	Transcriptional patterning in iN populations.....	78
4.4	Discussion .....	83
Chapter 5: Conclusions.....		88
Appendix A: Supplementary Methods .....		93
A.1	Specific methods for Chapter 2 .....	93
A.1.1	Mouse fibroblast isolation.....	93
A.1.2	Molecular cloning, cell culture, and lentiviral transduction .....	93
A.1.3	Immunohistochemistry.....	93
A.1.4	Electrophysiology .....	93
A.1.5	Calcium imaging .....	94
A.2	Specific methods for Chapter 3 .....	94
A.2.1	Mouse fibroblast isolation.....	94
A.2.2	Molecular cloning, cell culture, and lentiviral transduction .....	94
A.2.3	Immunohistochemistry.....	94
A.2.4	FACS purification .....	94
A.2.5	RNA isolation .....	95
A.2.6	RNA-Seq library preparation and sequencing .....	95
A.2.7	RNA-Seq data analysis (DESeq2 and PCA).....	95
A.3	Specific methods for Chapter 4 .....	96
A.3.1	Electrophysiology .....	96
A.3.2	WGCNA.....	96

## LIST OF ABBREVIATIONS

A1	<i>Ascl1</i>
A2	<i>Ascl2</i>
A5	<i>Ascl5</i>
ALS	Amyotrophic Lateral Sclerosis
B	<i>Brn2</i>
B2	<i>Brn2</i>
B3a	<i>Brn3a</i>
B3b	<i>Brn3b</i>
B3c	<i>Brn3c</i>
B4	<i>Brn4</i>
BAM	Transcription factors <i>Brn2, Ascl1, Myt1l</i>
bHLH	Basic helix-loop-helix
CER	Cerebellum
CTX	Cortex
Dpi	Days post-induction
DRG	Dorsal root ganglion
ESCs	Embryonic stem cells
FACS	Fluorescence activated cell sorting
GO	Gene ontology
HCN	Hyperpolarization-activated and cyclic nucleotide-gated
HIP	Hippocampus
iDAs	Induced dopaminergic neurons
Ih	Hyperpolarization-activated mixed cation current
iMNs	Induced motor neurons
iNs	Induced neurons
iPSCs	Induced pluripotent stem cells

iSNs	Induced sensory neurons
ME	Module eigengene
MEFs	Mouse embryonic fibroblasts
mHb	Medial habenula
MHb-v	Ventral medial habenula
MHb-d	Dorsal medial habenula
MT	Mitral and tufted
N1	<i>Neurogenin1</i>
N2	<i>Neurogenin2</i>
N3	<i>Neurogenin3</i>
ND2	<i>NeuroD2</i>
Ngn1	Neurogenin1
Ngn2	Neurogenin2
Ngn3	Neurogenin3
NR	Nuclear receptor
O4	<i>Oct4</i>
OB	Olfactory bulb
OB-GC	Olfactory bulb – granule cells
OB-MT	Olfactory bulb – mitral and tufted cells
P1	<i>Pit1</i>
PC1	Principal component 1
PC2	Principal component 2
PC3	Principal component 3
PCA	Principal component analysis
POU	Pit-Oct-Unc
RIN	RNA integrity number
TH	Tyrosine hydroxylase

TTFs	Tail-tip fibroblasts
WGCNA	Weighted gene coexpression network analysis
<i>Z</i>	<i>Zic1</i>

## LIST OF FIGURES

Figure 2.1	Experimental design for reprogramming mouse fibroblasts into induced neurons. (A) Schematic of the breeder crosses to produce the mouse reporter line, <i>Pchd21:CRE x Ai9</i> . Representative images of TdTomato-labeling (red) of mitral and tufted cells in the olfactory bulb of <i>Pchd21:CRE x Ai9</i> mice, including the top view of the whole brain ....17
Figure 2.2	<i>Brn2</i> and <i>Zic1</i> with either <i>Ngn1</i> or <i>Ngn2</i> , but not <i>Ascl1</i> , produces Pcdh21-TdTomato and Tuj1 positive cells. (A) Transient expression of <i>Brn2</i> and <i>Zic1</i> with <i>Ascl1</i> , <i>Ngn1</i> , or <i>Ngn2</i> converted fibroblasts into cells expressing neuronal marker Tuj1 (green) with neuronal morphology. Only cells reprogrammed with <i>Ngn1</i> or <i>Ngn2</i> , and not <i>Ascl1</i> .....20
Figure 2.3	Pcdh21-TdTomato and Tuj1 positive cells express additional neuronal markers. (A) Pcdh21-TdTomato positive cells (red) induced with <i>Ngn1</i> -B-Z or <i>Ngn2</i> -B-Z expressed neuronal markers, PSA-NCAM, Map2 and NeuN (green). (B) Pcdh21-TdTomato positive cells (red) expressed synaptic marker, Synaptophysin, and glutamate receptor subunit...21
Figure 2.4	<i>Ngn1-Brn2-Zic1</i> and <i>Ngn2-Brn2-Zic1</i> can generate Pcdh21-TdTomato and Tuj1 positive cells from postnatal tail-tip fibroblasts. (A) Live imaging at day 11 post-induction of Pcdh21-TdTomato positive cells (red) with neuronal morphology induced from P5 tail-tip fibroblasts. (B) Representative images of cells expressing neuronal marker Tuj1 .....22
Figure 2.5	Pcdh21-TdTomato positive cells with neuronal morphology exhibit electrophysiological properties of functional neurons. (A) Representative images of a patch clamped Pcdh21-TdTomato positive cell (red) with neuronal morphology. Scale bar represents 100 $\mu$ m. (B) Quantification of the resting membrane potential of the patched cells.....25
Figure 2.6	Pcdh21-TdTomato positive cells are glutamate responsive. (A-C) Representative calcium traces of cells that responded to (A) L-glutamate and (B) buffer, and (C) for cells with no detected calcium responses. Calcium transients were measured in Pcdh21-TdTomato positive cells with neuronal morphology using Map2::GCaMP5.G. ....26
Figure 2.7	<i>Ngn1-Brn2-Zic1</i> - and <i>Ngn2-Brn2-Zic1</i> -derived induced neurons express upper layer cortical neuron markers. (A) Representative images of Tuj1-expressing cells (red) and upper layer cortical neuron marker, Satb2 (green). <i>Brn2</i> and <i>Zic1</i> , together with <i>Ngn1</i> or <i>Ngn2</i> , were able to generate significantly higher Satb2-expressing cells (green).....29
Figure 2.8	<i>Brn2</i> paired with <i>Ngn1</i> or <i>Ngn2</i> generates Satb2 and Tuj1 expressing cells. (A) Representative images of Tuj1-expressing cells (red) and upper layer cortical neuron marker, Satb2 (green). <i>Brn2</i> , without <i>Zic1</i> , when paired with <i>Ngn1</i> or <i>Ngn2</i> , were still able to generate Tuj1- (red) and Satb2-expressing (green) cells. Images are of cells.....30
Figure 3.1	Experimental design of a reprogramming screen using bHLH and POU transcription factors. (A) Schematic overview of the methodology of the screen. Transient co-expression of reprogramming factor pairs are induced in E13.5 mouse embryonic fibroblasts (MEFs) one day post plating (left image) for the following 8 days and .....38
Figure 3.2	bHLH and POU transcription factor pairs capable of reprogramming fibroblasts into Tuj1-positive cells. (A) Matrix of pairwise reprogramming factor combinations yielding Tuj1-positive cells from fibroblasts on day 14 post-induction. Numbers reflect the percentages of Tuj1-positive cells out of the total number of fibroblasts .....41

Figure 3.3	Non-normalized Tuj1 percentages resulting from bHLH and POU transcription factor pair overexpression (A) Non-normalized matrix of pairwise reprogramming factors yielding Tuj1-positive cells from fibroblasts on day 14 post-induction. Numbers reflect the percentages of Tuj1-positive cells out of the total number of fibroblasts plated.....42
Figure 3.4	Representative images of Tuj1 immunofluorescence labeling for the positive bHLH and POU pairs. (A) Tuj1 immunofluorescence labeling of 35 of the 76 positive combinations that were selected for whole-transcriptome analysis. (B) Tuj1 immunofluorescence labeling of the remaining 41 of the 76 positive combinations. (C) Tuj1.....43
Figure 3.5	Tuj1 positive cells generated from p75-depleted MEF populations. (A) Representative FACS gates of MEFs (~180,000 cells shown). MEFs were depleted of p75-positive neural crest cells, by first gating for DAPI-negative cells (not shown) and collecting only those that were p75-negative (~93% of the DAPI-negative population). (B).....46
Figure 3.6	Tuj1 positive cells derived from postnatal tail tip fibroblasts. (A) Tuj1 immunofluorescence labeling of tail tip fibroblasts (TTFs) derived from 3-day-old mice and transduced with select reprogramming combinations. Fixed and stained on day 16 post-induction. (B) Tuj1 immunofluorescence of TTFs treated with only rtTA.....47
Figure 3.7	TauEGFP, Tuj1 positive cells express markers of mature neurons (A) Immunofluorescence labeling of Tuj1 positive cells for TauEGFP and neuronal markers, Map2, and synapsin. Representative candidate induced neurons were generated with reprogramming pairs <i>Ngn3</i> and <i>Pit1</i> 12 to 16 days post-induction. Scale bars.....48
Figure 3.8	Workflow schematic of sample preparation for RNA sequencing (RNA-Seq). (A) Cell populations were purified using FACS and libraries were prepped using SMARTer amplification. RNA-Seq was conducted on duplicate populations of TauEGFP positive cells generated from 35 different transcription factor pairs. Additional sequenced.....52
Figure 3.9	Fluorescence-activated cell sorting and RNA-Seq library preparation of candidate iN populations. (A) Representative immunofluorescence labeling of TauEGFP positive candidate induced neuron population ( <i>Ascl2/Brn3b</i> ) on day 12 post-induction using neuronal antibodies, Tuj1 and Map2. Scale bars represent 100 $\mu$ m. (B).....53
Figure 3.10	Principal component analysis reveals similarities and differences between individual iN and endogenous neuron populations. (A) Principal component analysis (PCA) performed using DESeq2 vsd-normalized RNA-Seq data from TauEGFP positive candidate iN populations (green), MEFs (grey), endogenous neurons (purple), and whole brain .....61
Figure 3.11	Gene expression profiles of candidate iNs overlap with profiles of endogenous neurons. (A) Top 10 biological process (BP level 5) gene ontology (GO) terms overrepresented in the TauEGFP positive candidate iN populations (green) and MEFs (grey) as determined by DAVID. Input gene lists included significantly upregulated genes .....62
Figure 3.12	Candidate iNs express both neuronal and fibroblasts genes. (A) Heat map shows expression patterns of the top 200 significantly different genes (rows, p-adjusted < 0.05) equally split between MEFs and endogenous neuronal populations and whole brain, as determined by DESeq2. Expression levels are defined as DESeq2 vsd-normalized .....63
Figure 4.1	iN populations exhibit electrophysiological properties of endogenous neurons. (A) Representative membrane voltage responses from a TauEGFP-, synapsin-positive cell with neuronal morphology generated with <i>Ascl2/Brn3c</i> under whole-cell patch clamp conditions at max current injection (top) and current steps until the first induction .....72

- Figure 4.2 Different reprogramming pairs generate iNs with varying membrane input resistances and hyperpolarization-activated cation currents. (A) Representative membrane voltage responses to depolarizing current steps of TauEGFP-, synapsin-positive cells with neuronal morphology generated with *Ngn3/Pit1* (left), *Ascl2/Brn3c* (middle) ..... 73
- Figure 4.3 Appropriate reprogramming pairs are expressed in the sequenced iN populations. (A) Heat map shows expression of reprogramming factors in all replicate iN and MEF populations. Expression levels are defined as DESeq2 vsd-normalized RNA-Seq counts and scaled by row (gene). A1, *Ascl1*; A2, *Ascl2*; A5, *Ascl5*; N1, *Ngn1*; N2, *Ngn2*; ..... 76
- Figure 4.4 Select genes expressed in discrete iN populations. (A) Heat map shows expression of genes significantly upregulated (p-adjusted value < 0.05) in each iN population versus all other iN populations and MEFs as determined by DESeq2. Expression levels are defined as DESeq2 vsd-normalized RNA-Seq counts. Dendrogram represents hierarchical ..... 77
- Figure 4.5 Transcription factor specificity of WGCNA module expression. (A) Hierarchical clustering of merged modules M1-M27 based on correlation distance with iN populations fixed based on bHLH transcription factor identity. Modules were determined using individual replicates of all iN populations and MEFs, but for clarity, replicates were ..... 80
- Figure 4.6 Complex gene coexpression patterning across iN populations. (A) Bar plots of average module eigengene expression of two representative modules enriched in MEF and various iN populations. (B) Table of number of module genes, representative hub genes and associated GO Terms for modules shown in (A). (C) Bar plots of average ..... 82

## LIST OF TABLES

Table 3.1	Metadata of candidate iN populations selected for RNA-Seq. Table of experimental information related to cell sorting, RNA extraction, library preparation and sequencing of each candidate iN population selected for RNA-Seq analyses, with two biological replicates per population. dpi, days post-induction; RIN, RNA integrity .....55
Table 3.2	Metadata of endogenous neuron populations selected for RNA-Seq. Table of experimental information related to cell sorting, RNA extraction, library preparation and sequencing of each endogenous neuron population selected for RNA-Seq analyses, with 2-3 biological replicates per population. RIN, RNA integrity number .....59

## ACKNOWLEDGMENTS

First, thank you to Kristin Baldwin for her valuable scientific mentorship, guidance, and instruction. Thank you to present and former members of the Baldwin Lab, fellow collaborators and classmates for their daily support. In particular, I would like to thank Jennifer Hazen, Andrew Adler, Sohyon Lee, Pavel Chubukov, Joel Blanchard, Kevin Eade, and Attila Szűcs for their collaboration on the work presented and Melissa Lau, Kimberly Reinhold, Sarah Leinwand, and Sarah Smith for their productive discussions. I would also like to acknowledge Samuel Pfaff, Alysson Muotri, Nicholas Spitzer, Andrew Huberman and Ali Torkamani for their scientific advice and expertise. Finally, thank you to the members of The Scripps Research Institute Flow Cytometry and Next Generation Sequencing core facilities and Kathy Spencer and the vivarium staff of the Dorris Neuroscience Center for their technical support.

This work was supported in part by the National Science Foundation Graduate Research Fellowship Program and the California Institute of Regenerative Medicine Interdisciplinary Stem Cell Training Program Predoctoral Fellowship.

Chapter 1, in part, has been submitted for publication of the material as it may appear in Tsunemoto, R.K., Eade, K.T., Blanchard, J.W., and Baldwin, K.K., 2015, Forward Engineering Neuronal Diversity Using Direct Reprogramming, The EMBO Journal. The dissertation author was the primary investigator and author of this paper.

Chapter 2, includes representative images of a *Pchd21:CRE x Ai9* mouse brain as seen in Figure 2.1.A that was imaged and provided by Jennifer Hazen, PhD. Xiaofei Zhang, PhD, conducted the electrophysiological recordings as seen in Figure 2.5. Calcium recordings, as seen in Figure 2.6, were done in collaboration with Andrew Adler, PhD.

Chapter 3, 4 and 5, in part, is currently being prepared for submission for publication of the material. Tsunemoto, R.K., Lee, S., Szűcs, A., Chubukov, P., Blanchard, J.W., Eade, K.T., Torkamani, A., Sanna, P.P., and Baldwin, K.K. The dissertation author was the primary investigator and author of this paper.

## VITA

### Research Experience

- 2010-2016      Ph.D. student, The Scripps Research Institute  
                    Advisor: Professor Kristin Baldwin, Ph.D.
- 2008-2009      Research Assistant, Massachusetts Institute of Technology  
                    Advisor: Professor Li-Huei Tsai, Ph.D.
- Summer 2007    Research Associate, Merck Research Lab  
                    Neuroscience Drug Discovery Department
- 2006-2007      Research Assistant, Massachusetts Institute of Technology  
                    Advisor: Professor Elly Nedivi, Ph.D.

### Publications

Hazen J.L., Faust G, Rodriguez A.R., Ferguson W, Shumilina S, Clark R.A., Boland M.J., Martin G, Chubukov P, Tsunemoto R.K., Torkamani A, Kupriyanov S, Hall I.M., and Baldwin K.K. (2016) The complete genome sequences, unique mutational spectra and developmental potency of adult neurons revealed by cloning. *Neuron* 89, 1223-1236

Tsunemoto R.K., Eade K.T., Blanchard J.W., Baldwin K.K. (2015) Forward engineering neuronal diversity using direct reprogramming. *The EMBO Journal* 34, 1445-1455

Blanchard J.W., Eade K.T., Szucs A, Lo Sardo V, Tsunemoto R.K., Williams D, Sanna P.P., Baldwin K.K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nature Neuroscience* 18. 25-35

Kim D, Frank C.L., Dobbin M.M., Tsunemoto R.K., Tu W, Peng P.L., Guan J.S., Lee B.H., Moy L.Y., Giusti P, Broodie N, Mazitschek R, Delalle I, Haggarty S.J., Neve R.L., Lu Y, Tsai L.H. (2008). Deregulation of HDAC1 by p25/Cdk5 in neurotoxicity. *Neuron* 60, 803-817

### Education

- 2009-2016      Ph.D. Neurosciences, University of California, San Diego
- 2005-2009      B.S. Brain and Cognitive Sciences, minor in Biomedical Engineering,  
                    Massachusetts Institute of Technology

ABSTRACT OF THE DISSERTATION

Deciphering Transcriptional Control of Neuronal Identity and Diversity  
Using Direct Reprogramming

by

Rachel Kuen Tsunemoto

Doctor of Philosophy in Neurosciences

University of California, San Diego, 2016

Professor Kristin Baldwin, Chair

Professor Andrew Huberman, Co-Chair

The mammalian nervous system is comprised of an unknown, but recognizably large, number of diverse neuronal subtypes. Recently, direct reprogramming (also known as transdifferentiation) has become an established method to rapidly produce “induced” neurons of numerous different subtypes directly from fibroblasts by overexpressing specific combinations of transcription factors and/or microRNAs. This technique not only provides the means to study various neuronal subtype populations that are not easily accessible, particularly in humans, but it also serves as a tool to interrogate the transcriptional codes that regulate neuronal subtype identity and maintenance. Both *in vivo* studies and direct reprogramming protocols have demonstrated that basic helix-loop-helix (bHLH) and Pit-Oct-Unc

(POU) transcription factors can aid in the specification of distinct neuronal subtypes. Therefore, we set out to comprehensively and systematically address whether first, additional bHLH and POU factor pairings could reprogram fibroblasts into functional neurons and second, dissect out the discrete and synergistic roles of these factors in neuronal subtype specification. We discovered over 70 novel pairs of bHLH and POU (and non-POU) transcription factors sufficient to generate candidate induced neurons (iNs) from mouse embryonic fibroblasts. Transcriptomic analysis of 35 of these candidate iN populations revealed gene expression profiles similar to those of endogenous neuronal populations. Additionally, differences between iN populations were observed at both a transcriptional and functional level.

## **Chapter 1:**

### **Introduction and Background**

#### **1.1 Background**

##### **1.1.1 Master transcriptional regulators of cellular identity**

For centuries, uncovering how a single fertilized egg cell develops into the many different and distinct cell types that comprise a functional organism has occupied the minds of biologists. Of the numerous mechanisms discovered to regulate cell fate decisions, transcription factors have repeatedly been shown to be critical for the development and specification of cell lineages (for a review, see: Holmberg and Perlmann (2012)). Key transcription factors, termed “master regulators”, are of particular interest because of their unique characteristics. Master regulators are transcription factors that not only regulate multiple downstream genes resulting in the specification of a particular cell lineage, but are also capable of transforming a committed cell type, when misexpressed, into that particular cell lineage (for a review on master regulators, see: Chan and Kyba (2013)). The act of converting one cell type into another, also known as transdifferentiation or direct reprogramming, could be used as a tool in identifying master regulators and uncovering transcriptional networks that specify many different cell types.

An established example of a master regulator is the myogenic transcription factor, *MyoD*, which is capable of transdifferentiating fibroblast and adipoblast cells into myogenic cells (Davis et al., 1987; Lassar et al., 1986; Tapscott et al., 1988). *MyoD* was also shown to activate myogenesis in pigment, neural and liver cell lines (Weintraub et al., 1989). The hematopoietic transcription factor *SCL (Tal1)* is also a known master regulator as it both regulates hematopoietic differentiation (Elefanty et al., 1997; Porcher et al., 1996; Robb et al., 1996) and converts non-hematopoietic cells of the zebrafish embryo into hematopoietic progenitors (Gering et al., 1998; Gering et al., 2003). Similarly, ectopic expression of *GATA-1* rapidly reprograms myeloblasts into hematopoietic eosinophils, thromboplasts, and erythroblasts (Heyworth et al., 2002; Kulesa et al., 1995) and overexpression of *C/EBP $\alpha$*  and  $\beta$  in B cells efficiently reprograms them into macrophages (Xie et al., 2004). The discovery of these master regulators has led to a better understanding of what drives the differentiation programs of these diverse cell types (for a review on *MyoD* and myogenesis, see: Tapscott (2005)).

The concept of transcription factor-induced transdifferentiation also contributed to the seminal discovery by Shinya Yamanaka and colleagues in 2006 that transient expression of *Oct4*, *Klf4*, *Myc*, and *Sox2* converts fibroblasts into induced pluripotent stem cells (iPSCs) that greatly resemble embryonic stem cells (ESCs) (Takahashi and Yamanaka, 2006) (for a historical perspective, see: Graf (2011)). Following this discovery, multiple groups have demonstrated the method's robust reproducibility and that alternative combinations of factors can also produce iPSCs from fibroblasts, blood and a large number of other accessible cell types from both mouse and human (as reviewed in Robinton and Daley (2012)). Since it was possible to reprogram many different cell types to pluripotency, it also seemed plausible to convert cells of one germ layer into cells of a different germ layer. One such example would be to convert fibroblasts into neurons with the right set of transcription factors.

### 1.1.2 Proneural genes and neuronal subtype diversity

Uncovering the master regulators of neuronal specification may also reveal how neuronal diversity is established. The mammalian nervous system is comprised of an undefined, but recognizably large, number of diverse neuronal subtypes. These neuronal subtypes are classified based on a number of phenotypic differences, including morphology, spatial location, synaptic partners, firing patterns, and neurotransmitter identity. It is the assembly of these various neuronal subtypes into distinct neural circuits that underlie functions such as sensory perception, motor control, and cognition. Neurodevelopmental studies have identified transcription factors, termed "proneural genes" that are necessary and sufficient to specify neuronal identity, in the context of ectodermal tissue (for reviews, see: Bertrand et al. (2002) and Huang et al. (2014)).

First identified in *Drosophila melanogaster*, proneural genes include members of the *achaete-scute* complex (*asc*) (Alonso and Cabrera, 1988; Garcia-Bellido, 1979), as well as, *atonal* (*ato*) (Jarman et al., 1993), *amos* (Goulding et al., 2000b; Huang et al., 2000), and *cato* (Goulding et al., 2000a). In vertebrates, *ato* orthologues include the *Neurogenin* gene family (*Ngn1*, *Ngn2*, and *Ngn3*), the *Neurogenic differentiation* gene family (*NeuroD1*, *NeuroD2*, *NeuroD4*, *Atoh1*, *Atoh2*, *Atoh7*), and the *Olig* gene family (*Olig1*, *Olig2*, and *Olig3*) (Bertrand et al., 2002). Orthologues of *asc* genes include *Ascl1* and *Ascl2*

(Bertrand et al., 2002), as well as, *Ascl3*, *Ascl4* and *Ascl5*. However, of the vertebrate orthologues, only *Ascl1*, *Ngn1* (Ma et al., 1996), *Ngn2* (Gradwohl et al., 1996), and potentially *Atoh1* and *Atoh7* are currently considered true proneural genes (Huang et al., 2014). The other vertebrate orthologues have been found to function downstream during neural differentiation (Boutin et al., 2010).

Like *MyoD* and *SCL*, these proneural genes are within the basic helix-loop-helix (bHLH) family of transcription factors. Functionally, as defined by Huang et al., (2014), a proneural gene promotes neuronal differentiation, as a loss of activity results in a reduction in Notch signaling and depletion of specific neural progenitor and neuronal populations, while a gain-of-function promotes cell cycle exit and the production of neurons over glia (Bertrand et al., 2002). Importantly, proneural genes are also involved in neuronal subtype specification since their regionalized patterns of expression yield distinct neuronal subtype populations. *Ascl1* expression, for example, in ventral telencephalic progenitors is required and sufficient to generate GABAergic neurons (Casarosa et al., 1999; Horton et al., 1999; Parras et al., 2002), but expression in progenitors of the hindbrain results in the differentiation of noradrenergic neurons (Parras et al., 2002). *Ngn2* expression in dorsal telencephalic progenitors is required and sufficient to generate glutamatergic pyramidal neurons (Fode et al., 2000), while expression in the ventral mesencephalon is necessary for the differentiation of midbrain dopaminergic neurons (Kele et al., 2006). Clearly, these proneural genes are crucial for the proper development of many types of neurons, but a question that had yet to be addressed was whether these proneural genes were capable of converting a non-neural, somatic cell type into functional neurons.

### **1.1.3 Direct generation of neurons from other cell types and lineages using transcription factor-mediated transdifferentiation**

It had been previously shown, in 2002, that overexpressing *Pax6* in postnatal mouse cortical astrocytes could convert them into neurons (Heins et al., 2002). Additional studies have also demonstrated that expressing either proneural genes *Ascl1* or *Ngn2* in astrocytes can transdifferentiate them into neurons, showing that neuronal identity can be established via mechanisms not observed during normal development (Berninger et al., 2007). However, these conversions were accomplished between astrocytes and neurons,

which are both derived from the same ectodermal germ layer. It wasn't until 2010 that functional neurons were successfully generated from fibroblasts (Vierbuchen et al., 2010).

Using an approach similar to the Yamanaka group (Takahashi and Yamanaka, 2006), Marius Wernig and colleagues discovered that ectopic expression of three transcription factors, proneural bHLH *Ascl1*, with *Brn2*, and *Myt1l* (BAM factors) could produce induced neurons (iNs) from mouse embryonic fibroblasts (MEFs) (Vierbuchen et al., 2010). These iNs expressed neuron-specific genes, fired action potentials, and formed functional synapses *in vitro*. This method was also applicable to human cells, although it required the addition of *NeuroD1* to produce neurons with mature electrophysiological properties (Pang et al., 2011).

Subsequently, the Wernig group also demonstrated that the BAM factors could reprogram mouse hepatocytes into iNs, confirming that both endodermal and mesodermal lineages are capable of transdifferentiating into a neural, ectodermal lineage (Marro et al., 2011). Single cell and genome-wide expression analyses of iNs derived from fibroblasts with those derived from hepatocytes showed that direct reprogramming of each lineage involved the coordinated activation of neuronal transcriptional pathways and concomitant silencing of the transcriptional program of the source cell (Marro et al., 2011). Importantly, neither protocol involved a dividing precursor or detectable neural stem cell intermediate.

These studies raised the question of whether it would be possible to direct the differentiation of pluripotent cells into neurons using the same transcription factor combinations. Indeed, overexpressing the BAM factors in human ESCs resulted in the rapid and efficient generation of neurons after only a week of induction (Pang et al., 2011). Additionally, *Ascl1* (Chanda et al., 2014), *Ngn2* (Thoma et al., 2012; Zhang et al., 2013), and *NeuroD2* (Sugimoto et al., 2009) expressed alone efficiently and directly generated neurons from both mouse and human ESCs and iPSCs. Therefore, to date, transcription factors have been shown to induce neurons from fibroblasts (Pang et al., 2011; Vierbuchen et al., 2010), hepatocytes (Marro et al., 2011), ESCs (Pang et al., 2011), and astrocytes (Torper et al., 2013) of both mouse and human origin, underscoring the robustness of this method.

#### **1.1.4 Micro-RNA-mediated repression also converts fibroblasts into induced neurons**

In 2011, Gerald Crabtree and colleagues reported that neuronal microRNAs miR-9/9\* and miR-124 alone were sufficient to convert fibroblasts into neurons (Yoo et al., 2011). This could be considered surprising given that microRNAs act primarily to repress genes. However, during development, as neurons become specified from their precursors, these particular micro-RNAs are known to repress the SWI/SNF-like chromatin remodeling complex BAFa. This repression then induces the cells to assemble the BAFb complex, which facilitates cell cycle exit and helps to establish key features of neuronal identity. Therefore, this study revealed that repression of developmentally relevant non-neuronal programs results in the indirect activation of neuronal specific chromatin states and subsequent reprogramming into neurons. Additionally, one effect of expressing miR-124 is to repress the RNA binding polypyrimidine-tract-binding (PTB) protein. Accordingly, it was found that downregulating PTB with specific small hairpin RNA could also generate functional neurons from MEFs (Xue et al., 2013). Interestingly, numerous neuronal subtypes were generated through expression of mi9\*/124, including both excitatory and inhibitory neurons and cells positive for markers of multiple cortical layers. This showed that not only can mi9\*/124 induce neuronal identity, but also suggested that each individual cell can implement distinct downstream programs to further pattern subtype identity.

Because transcription factor activation and micro-RNA-mediated repression can both lead to neuronal fate conversion, it was of interest to know if these approaches would act synergistically. Neuronal reprogramming with miR-9/9\* and miR-124 can be enhanced by including *NeuroD2*, and/or *Ascl1* and *Myt1l* (Yoo et al., 2011). And, in a separate study, human fibroblasts were converted into functional neurons by co-expressing miR-124 with *Brn2* and *Myt1l* (Ambasudhan et al., 2011). These results showed that transcriptional and microRNA pathways are compatible and suggest that it may be useful to identify additional neuronal subtype-specific microRNAs and transcription factors in order to improve the quality or subtype identity of induced neurons.

### **1.1.5 Biasing direct reprogramming to produce specific neuronal subtypes**

As mentioned previously, the brain contains an undefined, but undoubtedly large number of distinct neuronal subtypes. Therefore, it is of interest to know whether the BAM factors produce any

recognizable subtypes, and if so, which ones and how many types. While this is a difficult question to answer due to currently imprecise definitions of neuronal subtype identity, gross scale characterization of mouse iNs generated by the BAM factors alone and in combination with microRNAs showed that they contained neurons with mixed characteristics. The largest group resembled excitatory glutamatergic neurons based on their expression of markers such as VGluT1, VGluT2, and Tbr1. However, some iNs produced using the same factors expressed markers of inhibitory GABAergic neurons such as GAD, GABA, GAD65, and GAD67 (Ambasudhan et al., 2011; Pang et al., 2011; Vierbuchen et al., 2010; Yoo et al., 2011). In human studies, where adding *NeuroD1* to the BAM factors generated iNs, small numbers of cells expressed either the dopaminergic neuronal marker tyrosine hydroxylase (TH), or peripherin, a marker of peripheral neurons (Pang et al., 2011). iNs produced from hepatocytes also displayed diversity; approximately 4% of neurons expressed markers of cholinergic neurons (choline acetyltransferase, ChAT), and a similar proportion expressed peripherin (Marro et al., 2011).

The diversity of neuronal subtypes produced using the same set of reprogramming factors is consistent with two models. In one model, the same set of factors can produce different neuronal subtypes depending on the state of each individual cell. This is either based on differences in the donor cell epigenome or intrinsic stochastic mechanisms. Alternatively, because these previous studies used combinations of factors delivered in independent viruses, subtype diversity could be dependent on differences in the ratio of factors or which factors actually get expressed in each cell. The identities of the reprogramming factors, as expected, play a critical role in the subtype identity of the resulting induced neuron and are described in the following sections.

### **Producing dopaminergic neurons**

The high demand for certain neuronal subtypes has led to rapid progress in refining direct reprogramming methods to enrich for neurons with desired characteristics. One set of studies aimed to produce dopaminergic neurons because of their importance for Parkinson's disease (PD). PD is marked by the specific loss of midbrain dopaminergic neurons and the subsequent loss of dopamine in the target striatum. If it were possible to produce autologous dopaminergic neurons for transplantation into patients,

symptoms of the disease might be alleviated. As such, production of dopaminergic neurons from pluripotent stem cells has been a major focus of stem cell research (Wakeman et al., 2011). Encouragingly, combining the BAM factors with two genes involved in midbrain and dopaminergic neuron specification (*Lmx1a* and *FoxA2*) produced human iNs that expressed tyrosine hydroxylase, TH (Pfisterer et al., 2011).

Additional studies demonstrated that TH-expressing dopaminergic neurons could also be generated with only three factors, *Ascl1*, *Lmx1a* and *Nurr1* (Caiazzo et al., 2011). In these studies, 85% of the iNs generated from MEFs were TH-positive, showing that it is possible to generate highly enriched or even pure populations of desired neuronal subtypes with direct reprogramming. Dopaminergic neurons, however, comprise a very diverse set of subtypes. In attempts to generate induced dopaminergic neurons (iDAs) that more closely resembled endogenous midbrain DA neurons and are functional when transplanted into the brain, Kim et al. (2011) screened factors and assayed for Pitx3 expression, a specific reporter for midbrain DA neurons. This study identified a cocktail of six transcription factors, *Ascl1*, *Pitx3*, *Lmx1a*, *Nurr1*, *Foxa2*, and *EN1*, and two exogenous factors, Shh and FGF8, that could generate Pitx3-expressing iNs.

Remarkably, when Pitx3-expressing iNs were transplanted into lesioned striatums, a mouse model of PD, these iNs could increase dopamine levels and alleviate motor impairments in the treated mice (Kim et al., 2011). Dell'Anno et al. (2014) also analyzed transplanted iDAs derived from *Ascl1*, *Nurr1*, *Lmx1a* and observed long-term survival, integration into the host circuitry, and improved motor function in PD mice. Furthermore, these transplanted cells were able to enhance the recovery of motor impairments in PD mice by pharmacologically manipulating activity of the engrafted iDAs *in vivo* using DREADD technology. Finally, microRNAs miR-181a/a\* and miR-125b overexpression also produced dopaminergic neurons from human long-term, self-renewing neuroepithelial-like stem cells (Stappert et al., 2013). These results support the possibility to specifically and directly generate neurons with a wide diversity of desired neuronal subtype properties. The data also provides an important proof of principle regarding the suitability of induced neurons for cell replacement therapy.

### **Generating spinal motor neurons**

Spinal cord motor neurons are another neuronal subtype of interest because of their relevance to diseases, such as amyotrophic lateral sclerosis (ALS), and spinal cord injury. By combining the BAM factors and *NeuroD1* with four transcription factors highly expressed in motor neurons, *Lhx3*, *Hb9*, *Isl1*, and *Ngn2*, Son et al. (2011) were able to generate neurons that expressed the key motor neuron marker Hb9 from mouse and human fibroblasts. These induced motor neurons (iMNs) expressed motor neuron-specific markers, were capable of forming cholinergic synapses with muscle, and migrated appropriately to the ventral horn and projected out to the ventral root after transplantation.

Subsequently, Liu et al. (2013) were able to convert human fetal lung fibroblasts into cholinergic neurons of greater than 90% purity. Similar to previously published iMNs, these cholinergic neurons expressed motor neuron markers and formed functional neuromuscular junctions. In this study, transdifferentiation to a motor neuron identity was accomplished by simply overexpressing one factor, *Ngn2*, in combination with exposure to two small molecules, forskolin and dorsomorphin. However, these factors were not sufficient to reprogram postnatal and adult skin fibroblasts, which required an additional factor, *Sox11*. These results suggest that specific aspects of the epigenetic or transcriptional state of the donor cell can be critical to the successful application of direct reprogramming methods. In fact, Mazzoni et al. (2013) generated mouse ESC lines that harbored the transcription factors *Ngn2*, *Isl1* and *Lhx3* driven by an inducible promoter. Upon induction, they observed highly efficient conversion; greater than 99% of the cells expressed the spinal motor neuron marker Hb9. Furthermore, they were able to shift the neuronal subtype identity by replacing *Lhx3* with *Phox2a*. In these experiments, 99% of the ESCs converted into induced neurons that expressed the cranial motor neuron marker *Phox2b*.

### **Specifying other subtypes including DRG sensory neurons**

Many broad classes of neurons, including midbrain dopaminergic and motor neurons, can be further classified into functionally distinct subpopulations that differ in their connectivity, expression profiles, and electrophysiological properties. Because iNs are generated *ex vivo*, in absence of their endogenous circuitry, it is difficult to characterize the precise subpopulations generated within each iN population. Blanchard et al. (2015), however, was able to show that two different transcription factor pairs

(*Ngn1* with *Brn3a*, or *Ngn2* with *Brn3a*) could produce iNs that exhibited characteristics of three distinct subclasses of sensory neurons found in the dorsal root ganglion (DRG). Similar to sensory neurons found *in vivo*, these induced sensory neurons (iSNs) diversified into three subpopulations distinguished by their selective expression of TrkA, TrkB, or TrkC and differences in soma size. Additionally, the iSNs, like endogenous sensory neurons, selectively responded to compounds that mimic pain, temperature and itch, as demonstrated through calcium imaging. Concurrently, Wainger et al. (2015) identified an alternative path to produce nociceptor sensory neurons that expressed the channel, TrpV1. By using the BAM factors plus *Isl2*, *Ngn1*, and *Klf7*, these induced TrpV1 positive neurons exhibited the appropriate electrophysiological properties and responded to capsaicin, a known ligand of TrpV1. Together these two studies demonstrated that different direct reprogramming approaches could produce similar functional nociceptive neurons.

In addition to the induced neuronal subtypes described, other groups have reported that direct reprogramming methods can also produce medium spiny neurons (Victor et al., 2014), GABAergic neurons (Colasante et al., 2015), and serotonergic neurons (Vadodaria et al., 2015). Collectively, these studies highlight that direct reprogramming is capable of generating many different neuronal subtypes, each expressing their distinct characteristics. Conversely, these studies also demonstrate that many different factor combinations and methods are capable of generating the same neuronal subtypes with shared patterns of gene expression, such as expression of TH, Hb9, or Trp channels.

A question that remains is whether these distinct reprogramming methods converge on a limited number of feed forward transcriptional cascades or rather, are each set of induced neurons quite different from the other? Additionally, can we generate all the different types of neurons found in the brain using direct reprogramming or is there a limit to what we can recapitulate in a dish using solely intrinsic cues? In the chapters that follow, I will describe our attempts at elucidating the answers to these questions by performing a transcription factor screen that identifies more than seventy new reprogramming factor combinations sufficient to produce iNs directly from fibroblasts and analyzing the transcriptional and functional properties of the resulting iN populations.

Chapter 1, in part, has been submitted for publication of the material as it may appear in Tsunemoto, R.K., Eade, K.T., Blanchard, J.W., and Baldwin, K.K., 2015, Forward Engineering Neuronal

Diversity Using Direct Reprogramming, *The EMBO Journal*. The dissertation author was the primary investigator and author of this paper.

### References:

Alonso, M.C., and Cabrera, C.V. (1988). The achaete-scute gene complex of *Drosophila melanogaster* comprises four homologous genes. *The EMBO journal* 7, 2585-2591.

Ambasudhan, R., Talantova, M., Coleman, R., Yuan, X., Zhu, S., Lipton, S.A., and Ding, S. (2011). Direct reprogramming of adult human fibroblasts to functional neurons under defined conditions. *Cell Stem Cell* 9, 113-118.

Berninger, B., Costa, M.R., Koch, U., Schroeder, T., Sutor, B., Grothe, B., and Gotz, M. (2007). Functional properties of neurons derived from in vitro reprogrammed postnatal astroglia. *J Neurosci* 27, 8654-8664.

Bertrand, N., Castro, D.S., and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. *Nature reviews Neuroscience* 3, 517-530.

Blanchard, J.W., Eade, K.T., Szucs, A., Lo Sardo, V., Tsunemoto, R.K., Williams, D., Sanna, P.P., and Baldwin, K.K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nature neuroscience* 18, 25-35.

Boutin, C., Hardt, O., de Chevigny, A., Core, N., Goebbels, S., Seidenfaden, R., Bosio, A., and Cremer, H. (2010). NeuroD1 induces terminal neuronal differentiation in olfactory neurogenesis. *Proc Natl Acad Sci U S A* 107, 1201-1206.

Caiazzo, M., Dell'Anno, M.T., Dvoretzkova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T.D., Menegon, A., Roncaglia, P., Colciago, G., Russo, G., Carninci, P., Pezzoli, G., Gainetdinov, R.R., Gustinich, S., Dityatev, A., and Broccoli, V. (2011). Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* 476, 224-227.

Casarosa, S., Fode, C., and Guillemot, F. (1999). Mash1 regulates neurogenesis in the ventral telencephalon. *Development* 126, 525-534.

Chan, S.S., and Kyba, M. (2013). What is a Master Regulator? *J Stem Cell Res Ther* 3.

Chanda, S., Ang, C.E., Davila, J., Pak, C., Mall, M., Lee, Q.Y., Ahlenius, H., Jung, S.W., Sudhof, T.C., and Wernig, M. (2014). Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1. *Stem cell reports* 3, 282-296.

Colasante, G., Lignani, G., Rubio, A., Medrihan, L., Yekhleif, L., Sessa, A., Massimino, L., Giannelli, S.G., Sacchetti, S., Caiazzo, M., Leo, D., Alexopoulou, D., Dell'Anno, M.T., Ciabatti, E., Orlando, M., Studer, M., Dahl, A., Gainetdinov, R.R., Taverna, S., Benfenati, F., and Broccoli, V. (2015). Rapid Conversion of Fibroblasts into Functional Forebrain GABAergic Interneurons by Direct Genetic Reprogramming. *Cell Stem Cell*.

Davis, R.L., Weintraub, H., and Lassar, A.B. (1987). Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* 51, 987-1000.

Dell'Anno, M.T., Caiazzo, M., Leo, D., Dvoretzkova, E., Medrihan, L., Colasante, G., Giannelli, S., Theka, I., Russo, G., Mus, L., Pezzoli, G., Gainetdinov, R.R., Benfenati, F., Taverna, S., Dityatev, A., and

- Broccoli, V. (2014). Remote control of induced dopaminergic neurons in parkinsonian rats. *The Journal of clinical investigation* *124*, 3215-3229.
- Elefanty, A.G., Robb, L., Birner, R., and Begley, C.G. (1997). Hematopoietic-specific genes are not induced during in vitro differentiation of scl-null embryonic stem cells. *Blood* *90*, 1435-1447.
- Fode, C., Ma, Q., Casarosa, S., Ang, S.L., Anderson, D.J., and Guillemot, F. (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes & Development* *14*, 67-80.
- Garcia-Bellido, A. (1979). Genetic Analysis of the Achaete-Scute System of DROSOPHILA MELANOGASTER. *Genetics* *91*, 491-520.
- Gering, M., Rodaway, A.R., Göttgens, B., Patient, R.K., and Green, A.R. (1998). The SCL gene specifies haemangioblast development from early mesoderm. *The EMBO journal* *17*, 4029-4045.
- Gering, M., Yamada, Y., Rabbitts, T.H., and Patient, R.K. (2003). Lmo2 and Scf/Tal1 convert non-axial mesoderm into haemangioblasts which differentiate into endothelial cells in the absence of Gata1. *Development* *130*, 6187-6199.
- Goulding, S.E., White, N.M., and Jarman, A.P. (2000a). cato encodes a basic helix-loop-helix transcription factor implicated in the correct differentiation of Drosophila sense organs. *Developmental biology* *221*, 120-131.
- Goulding, S.E., zur Lage, P., and Jarman, A.P. (2000b). amos, a proneural gene for Drosophila olfactory sense organs that is regulated by lozenge. *Neuron* *25*, 69-78.
- Gradwohl, G., Fode, C., and Guillemot, F. (1996). Restricted expression of a novel murine atonal-related bHLH protein in undifferentiated neural precursors. *Developmental biology* *180*, 227-241.
- Graf, T. (2011). Historical origins of transdifferentiation and reprogramming. *Cell Stem Cell* *9*, 504-516.
- Heins, N., Malatesta, P., Cecconi, F., Nakafuku, M., Tucker, K.L., Hack, M.A., Chapouton, P., Barde, Y.A., and Gotz, M. (2002). Glial cells generate neurons: the role of the transcription factor Pax6. *Nat Neurosci* *5*, 308-315.
- Heyworth, C., Pearson, S., May, G., and Enver, T. (2002). Transcription factor-mediated lineage switching reveals plasticity in primary committed progenitor cells. *The EMBO journal* *21*, 3770-3781.
- Holmberg, J., and Perlmann, T. (2012). Maintaining differentiated cellular identity. *Nat Rev Genet* *13*, 429-439.
- Horton, S., Meredith, A., Richardson, J.A., and Johnson, J.E. (1999). Correct coordination of neuronal differentiation events in ventral forebrain requires the bHLH factor MASH1. *Molecular and cellular neurosciences* *14*, 355-369.
- Huang, C., Chan, J.A., and Schuurmans, C. (2014). Proneural bHLH genes in development and disease. *Current topics in developmental biology* *110*, 75-127.
- Huang, M.L., Hsu, C.H., and Chien, C.T. (2000). The proneural gene amos promotes multiple dendritic neuron formation in the Drosophila peripheral nervous system. *Neuron* *25*, 57-67.
- Jarman, A.P., Grau, Y., Jan, L.Y., and Jan, Y.N. (1993). atonal is a proneural gene that directs chordotonal organ formation in the Drosophila peripheral nervous system. *Cell* *73*, 1307-1321.

Kele, J., Simplicio, N., Ferri, A.L., Mira, H., Guillemot, F., Arenas, E., and Ang, S.L. (2006). Neurogenin 2 is required for the development of ventral midbrain dopaminergic neurons. *Development* *133*, 495-505.

Kim, J., Su, S.C., Wang, H., Cheng, A.W., Cassady, J.P., Lodato, M.A., Lengner, C.J., Chung, C.Y., Dawlaty, M.M., Tsai, L.H., and Jaenisch, R. (2011). Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. *Cell Stem Cell* *9*, 413-419.

Kulesa, H., Frampton, J., and Graf, T. (1995). GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboplasts, and erythroblasts. *Genes & Development* *9*, 1250-1262.

Lassar, A.B., Paterson, B.M., and Weintraub, H. (1986). Transfection of a DNA locus that mediates the conversion of 10T1/2 fibroblasts to myoblasts. *Cell* *47*, 649-656.

Liu, M.L., Zang, T., Zou, Y., Chang, J.C., Gibson, J.R., Huber, K.M., and Zhang, C.L. (2013). Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons. *Nature communications* *4*, 2183.

Ma, Q., Kintner, C., and Anderson, D.J. (1996). Identification of neurogenin, a vertebrate neuronal determination gene. *Cell* *87*, 43-52.

Marro, S., Pang, Z.P., Yang, N., Tsai, M.C., Qu, K., Chang, H.Y., Sudhof, T.C., and Wernig, M. (2011). Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* *9*, 374-382.

Mazzoni, E.O., Mahony, S., Closser, M., Morrison, C.A., Nedelec, S., Williams, D.J., An, D., Gifford, D.K., and Wichterle, H. (2013). Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat Neurosci*.

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Sudhof, T.C., and Wernig, M. (2011). Induction of human neuronal cells by defined transcription factors. *Nature* *476*, 220-223.

Parras, C.M., Schuurmans, C., Scardigli, R., Kim, J., Anderson, D.J., and Guillemot, F. (2002). Divergent functions of the proneural genes *Mash1* and *Ngn2* in the specification of neuronal subtype identity. *Genes & Development* *16*, 324-338.

Pfisterer, U., Kirkeby, A., Torper, O., Wood, J., Nelander, J., Dufour, A., Bjorklund, A., Lindvall, O., Jakobsson, J., and Parmar, M. (2011). Direct conversion of human fibroblasts to dopaminergic neurons. *Proc Natl Acad Sci U S A* *108*, 10343-10348.

Porcher, C., Swat, W., Rockwell, K., Fujiwara, Y., Alt, F.W., and Orkin, S.H. (1996). The T cell leukemia oncoprotein SCL/tal-1 is essential for development of all hematopoietic lineages. *Cell* *86*, 47-57.

Robb, L., Elwood, N.J., Elefanty, A.G., Kontgen, F., Li, R., Barnett, L.D., and Begley, C.G. (1996). The *scl* gene product is required for the generation of all hematopoietic lineages in the adult mouse. *The EMBO journal* *15*, 4123-4129.

Robinton, D.A., and Daley, G.Q. (2012). The promise of induced pluripotent stem cells in research and therapy. *Nature* *481*, 295-305.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* *9*, 205-218.

Stappert, L., Borghese, L., Roese-Koerner, B., Weinhold, S., Koch, P., Terstegge, S., Uhrberg, M., Wernet, P., and Brustle, O. (2013). MicroRNA-Based Promotion of Human Neuronal Differentiation and Subtype Specification. *PLoS One* 8, e59011.

Sugimoto, Y., Furuno, T., and Nakanishi, M. (2009). Effect of NeuroD2 expression on neuronal differentiation in mouse embryonic stem cells. *Cell biology international* 33, 174-179.

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

Tapscott, S.J. (2005). The circuitry of a master switch: MyoD and the regulation of skeletal muscle gene transcription. *Development* 132, 2685-2695.

Tapscott, S.J., Davis, R.L., Thayer, M.J., Cheng, P.F., Weintraub, H., and Lassar, A.B. (1988). MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. *Science* 242, 405-411.

Thoma, E.C., Wischmeyer, E., Offen, N., Maurus, K., Siren, A.L., Scharl, M., and Wagner, T.U. (2012). Ectopic expression of neurogenin 2 alone is sufficient to induce differentiation of embryonic stem cells into mature neurons. *PLoS One* 7, e38651.

Torper, O., Pfisterer, U., Wolf, D.A., Pereira, M., Lau, S., Jakobsson, J., Bjorklund, A., Grealish, S., and Parmar, M. (2013). Generation of induced neurons via direct conversion in vivo. *Proc Natl Acad Sci U S A*.  
Vadodaria, K.C., Mertens, J., Paquola, A., Bardy, C., Li, X., Jappelli, R., Fung, L., Marchetto, M.C., Hamm, M., Gorris, M., Koch, P., and Gage, F.H. (2015). Generation of functional human serotonergic neurons from fibroblasts. *Molecular Psychiatry*.

Victor, Matheus B., Richner, M., Hermansteyne, Tracey O., Ransdell, Joseph L., Sobieski, C., Deng, P.-Y., Klyachko, Vitaly A., Nerbonne, Jeanne M., and Yoo, Andrew S. (2014). Generation of Human Striatal Neurons by MicroRNA-Dependent Direct Conversion of Fibroblasts. *Neuron* 84, 311-323.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035-1041.

Wainger, B.J., Buttermore, E.D., Oliveira, J.T., Mellin, C., Lee, S., Saber, W.A., Wang, A.J., Ichida, J.K., Chiu, I.M., Barrett, L., Huebner, E.A., Bilgin, C., Tsujimoto, N., Brenneis, C., Kapur, K., Rubin, L.L., Eggan, K., and Woolf, C.J. (2015). Modeling pain in vitro using nociceptor neurons reprogrammed from fibroblasts. *Nature neuroscience* 18, 17-24.

Wakeman, D.R., Dodiya, H.B., and Kordower, J.H. (2011). Cell transplantation and gene therapy in Parkinson's disease. *Mount Sinai Journal of Medicine: A Journal of Translational and Personalized Medicine* 78, 126-158.

Weintraub, H., Tapscott, S.J., Davis, R.L., Thayer, M.J., Adam, M.A., Lassar, A.B., and Miller, A.D. (1989). Activation of muscle-specific genes in pigment, nerve, fat, liver, and fibroblast cell lines by forced expression of MyoD. *Proc Natl Acad Sci U S A* 86, 5434-5438.

Xie, H., Ye, M., Feng, R., and Graf, T. (2004). Stepwise Reprogramming of B Cells into Macrophages. *Cell* 117, 663-676.

Xue, Y., Ouyang, K., Huang, J., Zhou, Y., Ouyang, H., Li, H., Wang, G., Wu, Q., Wei, C., Bi, Y., Jiang, L., Cai, Z., Sun, H., Zhang, K., Zhang, Y., Chen, J., and Fu, X.D. (2013). Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits. *Cell* 152, 82-96.

Yoo, A.S., Sun, A.X., Li, L., Shcheglovitov, A., Portmann, T., Li, Y., Lee-Messer, C., Dolmetsch, R.E., Tsien, R.W., and Crabtree, G.R. (2011). MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* 476, 228-231.

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., Xu, W., Yang, N., Danko, T., Chen, L., Wernig, M., and Sudhof, T.C. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78, 785-798.

## Chapter 2:

### Direct reprogramming of mouse fibroblasts into functional neurons using *Neurogenin1* and *Neurogenin2*

#### 2.1 Introduction

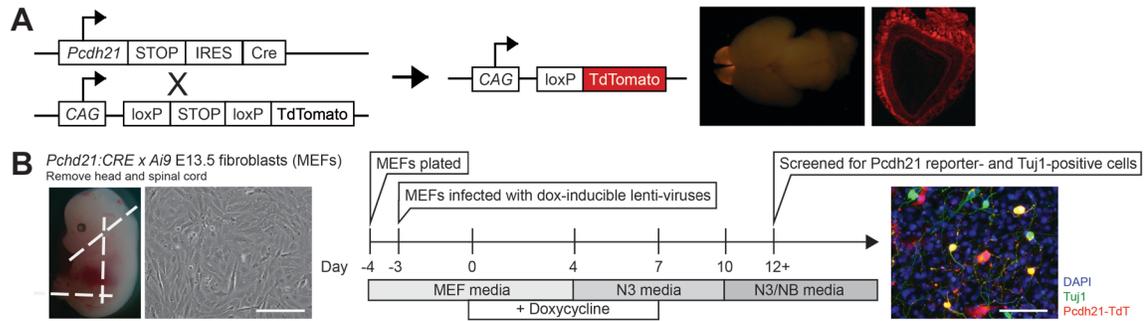
Immediately following the discovery that *Ascl1*, *Brn2*, and *Myt1l* could together convert fibroblasts into functional induced neurons (Vierbuchen et al., 2010), other groups also utilized *Ascl1* in their combinations to generate specific neuronal subtypes (Caiazzo et al., 2011; Kim et al., 2011b; Pfisterer et al., 2011; Son et al., 2011). Mechanistically, *Ascl1* is described as an “on-target pioneering factor” that can immediately bind neuronal targets across the fibroblast genome and recruit *Brn2* to those sites (Wapinski et al., 2013). Even alone, *Ascl1* expression is sufficient to induce immature neuronal features from fibroblasts (Chanda et al., 2014; Vierbuchen et al., 2010).

Interestingly, *Ascl1* expression in the developing nervous system is limited to spatially distinct populations of neural progenitors that give rise to both neurons and oligodendrocytes. *Ascl1* progenitors produce many, but importantly, not all of the neuronal subtypes and oligodendrocytes found in the brain (Kim et al., 2008). Therefore, following the publications utilizing *Ascl1* as a primary reprogramming factor, there were two obvious questions that had not been addressed. First, was *Ascl1* required for reprogramming from fibroblast to neurons? Second, does the use of *Ascl1* as a reprogramming factor limit the types of neurons capable of being produced?

To address these questions, we initially aimed to generate mitral and tufted cells, the principal excitatory projection neurons of the olfactory bulb. Mitral and tufted (MT) cells were a neuronal subtype not previously generated using reprogramming methods. More critically, MT cells are derived from progenitors expressing the pro-neural bHLH transcription factors *Neurogenin1* (*Ngn1*) and *Neurogenin2* (*Ngn2*), not *Ascl1* (Kim et al., 2008; Kim et al., 2011a; Winpenny et al., 2011). In fact, *Ascl1* and *Neurogenins* are expressed in roughly mutually exclusive domains of the developing mouse brain (Fode et al., 2000; Ma et al., 1997; Osorio et al., 2010). Therefore, we tested whether replacing *Ascl1* with either *Ngn1* or *Ngn2* could convert mouse fibroblasts into functional neurons and if such a swap would alter the neuronal subtype identity of the resulting iNs and confer a MT cell identity.

## 2.2 Methods: Testing candidate transcription factors to generate a specific neuronal subtype

To test whether the candidate transcription factors, *Ngn1* and *Ngn2*, were capable of generating functional iNs directly from mouse fibroblasts, we modified the reprogramming protocol published by the Wernig group (Vierbuchen et al., 2010). In short, we prepared either E13.5 mouse embryonic fibroblasts (MEFs) or P3 tail-tip fibroblasts (TTFs) from a mitral and tufted cell reporter line, *Pchd21:CRE x Ai9* (Figure 2.1A). *Ngn1* and *Ngn2*, along with *Brn2* and *Zic1*, were transiently expressed in the fibroblasts by transducing the cells with doxycycline-inducible lentiviruses containing the cDNA of each factor and exposing the cells to doxycycline for 8 days. Cells were first maintained in MEF media, then subsequently in neural maintenance medias, N3 and N3/NB media. After 12 days post-induction, cells were assessed for neuronal properties and neuronal subtype-specific features using immunostaining, electrophysiological recordings and calcium imaging (Figure 2.1B). Detailed methods are provided in Appendix A.1.



**Figure 2.1 Experimental design for reprogramming mouse fibroblasts into induced neurons.** (A) Schematic of the breeder crosses to produce the mouse reporter line, *Pcdh21:CRE x Ai9*. Representative images of TdTomato-labeling (red) of mitral and tufted cells in the olfactory bulb of *Pcdh21:CRE x Ai9* mice, including the top view of the whole brain (left) and coronal section of the olfactory bulb (right) at P21. (B) Schematic of the reprogramming process, which includes representative images of E13.5 MEFs one day post plating (left) and *Pcdh21*-TdTomato (red) and Tuj1 (green) positive cells at day 14 post-induction (right). Scale bars represent 100  $\mu$ m.

## 2.3 Results

### 2.3.1 Transient overexpression of either *Ngn1* or *Ngn2* with *Brn2* and *Zic1* reprograms mouse fibroblasts to express pan-neuronal markers

Previously published reports utilizing transcription factors to directly reprogram fibroblasts to neurons, at the time of this study, had all used *Ascl1* in their reprogramming factor combinations. However, *in vivo* neurodevelopmental studies have shown that *Ascl1*-expressing neural progenitors cells give rise to only a limited number of neuronal subtype populations in the adult mouse brain, not including MT cells of the olfactory bulb (Kim et al., 2008). Rather, MT cells derive from *Ngn1*- and *Ngn2*-expressing neural progenitors (Kim et al., 2011a; Winpenny et al., 2011). Thus, we tested whether replacing *Ascl1* with *Ngn1* and *Ngn2* in a variation of the Wernig factor combination, *Ascl1-Brn2-Zic1*, could generate induced neurons with MT cell properties\*.

We first established that *Ngn1* and *Ngn2*, like *Ascl1*, when combined with *Brn2* and *Zic1*, could transform MEFs into cells that exhibit neuronal morphology and express the pan-neuronal marker, *Tuj1*. Reprogramming efficiencies, or the percent of *Tuj1*-positive cells out of the total number of MEFs plated, between combinations containing either *Ascl1*, *Ngn1*, or *Ngn2* were not significantly different from each other (when normalized to *Ascl1-Brn2-Zic1* for each experiment; raw percentages were between 1-5%) (Figure 2.2).

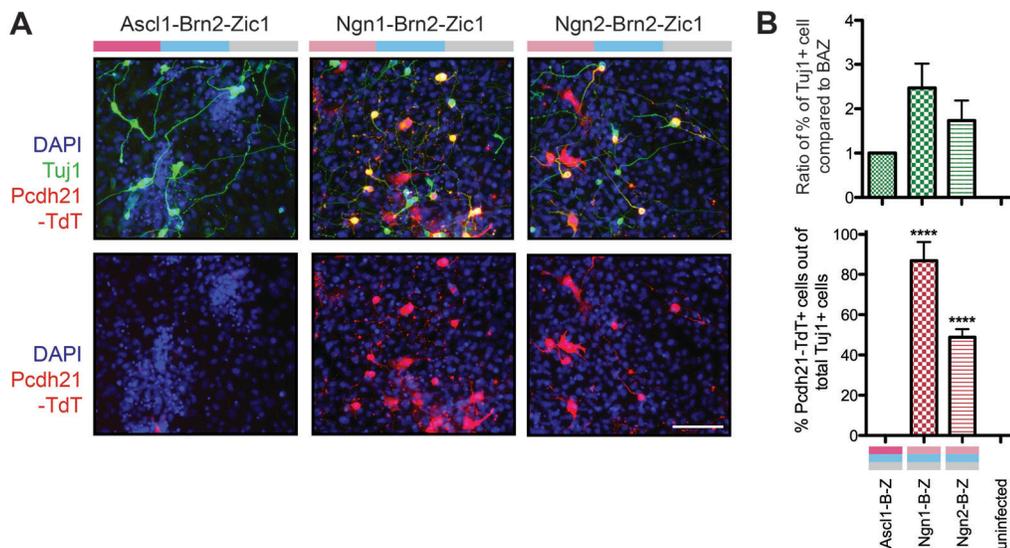
To determine if these *Tuj1*-expressing cells also express the MT cell marker, *Pcdh21*, we utilized MEFs from the MT mouse reporter line, *Pcdh21:CRE x Ai9*, that specifically and permanently labels MT cells with the fluorescent protein TdTomato in the developing and adult mouse brain (Figure 2.1A). The majority of *Tuj1*-expressing cells derived from *Ngn1-Brn2-Zic1* and *Ngn2-Brn2-Zic1* were co-labeled with the *Pcdh21*-TdTomato reporter (*Ngn1*-B-Z, 86.8% ± 9.3% and *Ngn2*-B-Z, 48.9% ± 4.0%). No *Tuj1*-positive cells from the *Ascl1-Brn2-Zic1* condition expressed *Pcdh21*-TdTomato (Figure 2.2). Together this demonstrates that *Ngn1* and *Ngn2*, when combined with *Brn2* and *Zic1*, can generate putative induced neurons and these *Tuj1*-expressing cells, unlike *Ascl1*-derived iNs, express the MT marker, *Pcdh21*.

---

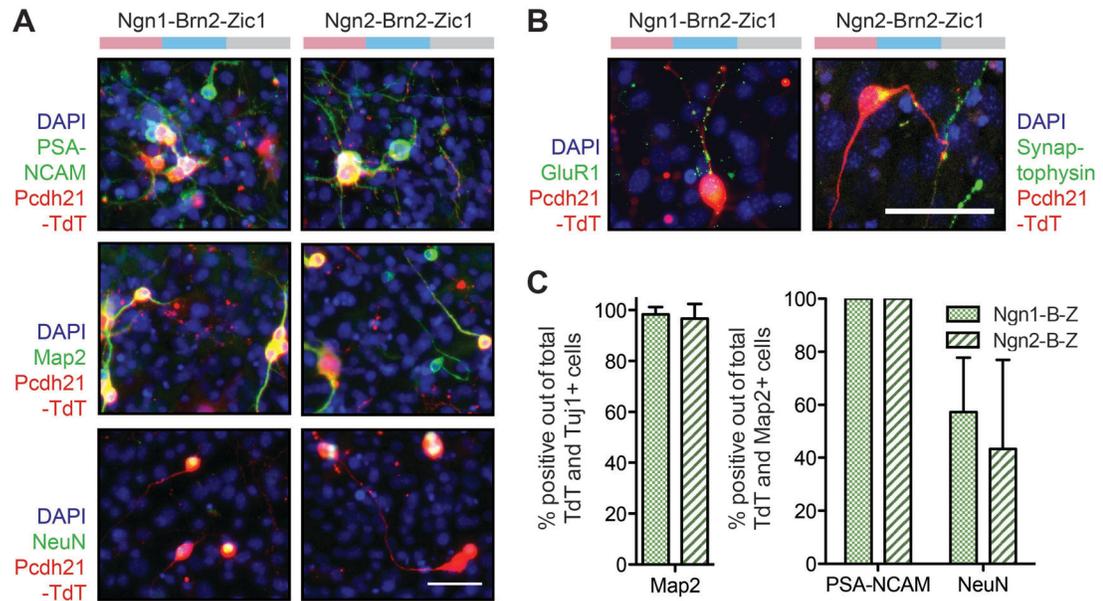
\* We used *Ascl1-Brn2-Zic1* instead of the more commonly used *Ascl1-Brn2-Myt1l* (BAM) combination because at the time we were developing polycistronic constructs in which the shorter gene *Zic1* was preferable to use than the longer *Myt1l*. According to Vierbuchen et al. (2010), both combinations were efficient at generating iNs.

To further establish that the *Ngn1*- and *Ngn2*-derived Pcdh21-TdTomato and Tuj1 double positive cell are indeed putative induced neurons, we confirmed the expression of additional neuronal markers, including PSA-NCAM, Map2 and NeuN, synaptic marker, Synaptophysin, and glutamate receptor subunit, GluR1. Nearly all Pcdh21-TdTomato and Tuj1 double positive cells expressed neuronal migration and synaptogenesis marker, PSA-NCAM (*Ngn1*-B-Z, 100%  $\pm$  0% and *Ngn2*-B-Z, 100%  $\pm$  0%), and mature neuronal marker, Map2 (*Ngn1*-B-Z, 98.3%  $\pm$  2.9% and *Ngn2*-B-Z, 96.7%  $\pm$  5.8%). NeuN was expressed in a subset of Pcdh21-TdTomato and Map2 double positive cells (*Ngn1*-B-Z, 57.3%  $\pm$  20.5% and *Ngn2*-B-Z, 43.3%  $\pm$  33.6%) (Figure 2.3). Interestingly, NeuN is not expressed in mitral and tufted cells *in vivo* (Mullen et al., 1992), suggesting at least a subpopulation of these Pcdh21-TdTomato positive cells do not recapitulate MT cell gene expression patterns.

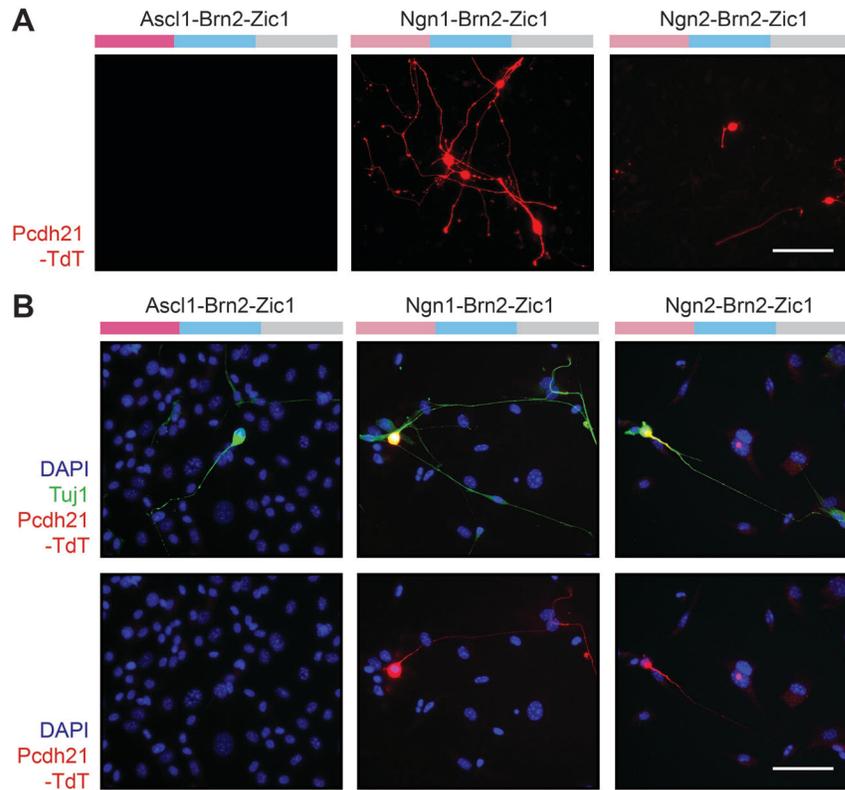
So far, we have demonstrated that *Ngn1-Brn2-Zic1* and *Ngn2-Brn2-Zic1* can transform MEFs to express a panel of neuronal markers and establish neuronal morphology. However, to exclude the possibilities that successful reprogramming with *Ngn1* and *Ngn2* is limited to using an embryonic starting cell population or *Ngn1* and *Ngn2* are reprogramming potentially contaminating neural precursor cells within the MEF population, we also generated Tuj1-expressing cells from P3 tail tip fibroblasts (Figure 2.4B). Like in the MEF experiments, *Ngn1* and *Ngn2*, combined with *Brn2* and *Zic1*, generated Tuj1 positive cells that also expressed the Pcdh21-TdTomato reporter (Figure 2.4).



**Figure 2.2 *Brn2* and *Zic1* with either *Ngn1* or *Ngn2*, but not *Ascl1*, produces *Pcdh21*-*TdTomato* and *Tuj1* positive cells.** (A) Transient expression of *Brn2* and *Zic1* with *Ascl1*, *Ngn1*, or *Ngn2* converted fibroblasts into cells expressing neuronal marker *Tuj1* (green) with neuronal morphology. Only cells reprogrammed with *Ngn1* or *Ngn2*, and not *Ascl1*, together with *Brn2* and *Zic1* were positive for *Pcdh21*-*TdTomato* (red). Images are of cells immunostained at day 14 post-induction. Scale bars represent 100  $\mu$ m. (B) Quantification of the percent of *Tuj1* positive cells out of total number of cells plated and normalized to BAZ. (C) Quantification of the percent of *Pcdh21*-*TdTomato* positive cells out of the total number of *Tuj1* positive cells. Data for (B) and (C) is presented as the mean  $\pm$  SD from three independent experiment. \*\*\*\* represents a p-value  $< 0.0001$  (Bonferroni's Multiple Comparison Test).



**Figure 2.3 Pcdh21-TdTomato and Tuj1 positive cells express additional neuronal markers.** (A) Pcdh21-TdTomato positive cells (red) induced with *Ngn1*-B-Z or *Ngn2*-B-Z expressed neuronal markers, PSA-NCAM, Map2 and NeuN (green). (B) Pcdh21-TdTomato positive cells (red) express synaptic marker, Synaptophysin, and glutamate receptor subunit, GluR1 (green). Images from (A) and (B) are of cells immunostained at day 14 post-induction. Scale bars represent 50  $\mu$ m. (C) Quantification of the percent of Pcdh21-TdTomato and Tuj1 double positive cells that also express Map2 (right). Quantification of the percent of Pcdh21-TdTomato and Map2 double positive cells that also express PSA-NCAM and NeuN (left). Data are presented as the mean  $\pm$  SD from one experiment, with counts from three fields ( $n > 50$  cells per combination).



**Figure 2.4** *Ngn1-Brn2-Zic1* and *Ngn2-Brn2-Zic1* can generate *Pcdh21-TdTomato* and *Tuj1* positive cells from postnatal tail-tip fibroblasts. (A) Live imaging at day 11 post-induction of *Pcdh21-TdTomato* positive cells (red) with neuronal morphology induced from P5 tail-tip fibroblasts. (B) Representative images of cells expressing neuronal marker *Tuj1* (green). Only *Brn2* and *Zic1*, together with *Ngn1* or *Ngn2*, were able to generate *Tuj1*-expressing cells (green) that were also *Pcdh21-TdTomato* positive (red). Images are of cells immunostained at day 14 post-induction. Scale bars represent 100  $\mu\text{m}$ .

### 2.3.2 Induced neurons generated by *Ngn1-Brn2-Zic1* and *Ngn2-Brn2-Zic1* exhibit physiological properties of functional neurons

To assess the electrophysiological properties of the putative induced neurons and confirm they are indeed functional induced neurons, we performed whole cell patch clamp on Pcdh21-TdTomato positive cells with neuronal morphology between days 12-21 post induction in MEFs (Figure 2.5A). The resting membrane potentials for *Ngn1-Brn2-Zic1*- and *Ngn2-Brn2-Zic1*-derived putative iNs were  $-32.6 \pm 8.4$  mV ( $n = 12$ ) and  $-47.2 \pm 10.2$  mV ( $n = 6$ ), respectively (Figure 2.5B). This is within the range of resting membrane potentials of other induced neuron populations previously reported (Blanchard et al., 2015; Caiazzo et al., 2011; Kim et al., 2011b; Son et al., 2011; Vierbuchen et al., 2010). The majority of patched cells, 10 out of 16 cells (62.5%) derived from *Ngn1-Brn2-Zic1* and 5 out of 6 cells (83.3%) for *Ngn2-Brn2-Zic1*, exhibited action potentials in response to depolarizing current steps (Figure 2.5C-D). Additionally, we observed spontaneous action potentials in 2 out of 16 *Ngn1-Brn2-Zic1* cells (12.5%) and 1 out of 6 *Ngn2-Brn2-Zic1* cells (16.6%) (Figure 2.5C, E). This is consistent with the expression of functioning voltage-dependent sodium and potassium channels in the cells and was also observed as fast inward and subsequent slow outward currents, respectively, with depolarizing voltage steps (Figure 2.5F). At least one cell exhibited ligand-gated ion channel activity, as seen as fast inward currents, with the application of 100  $\mu$ M AMPA and 200  $\mu$ M GABA (Figure 2.5G). Observing a response to AMPA is consistent with positive immunostaining for the glutamatergic AMPA receptor subunit, GluR1.

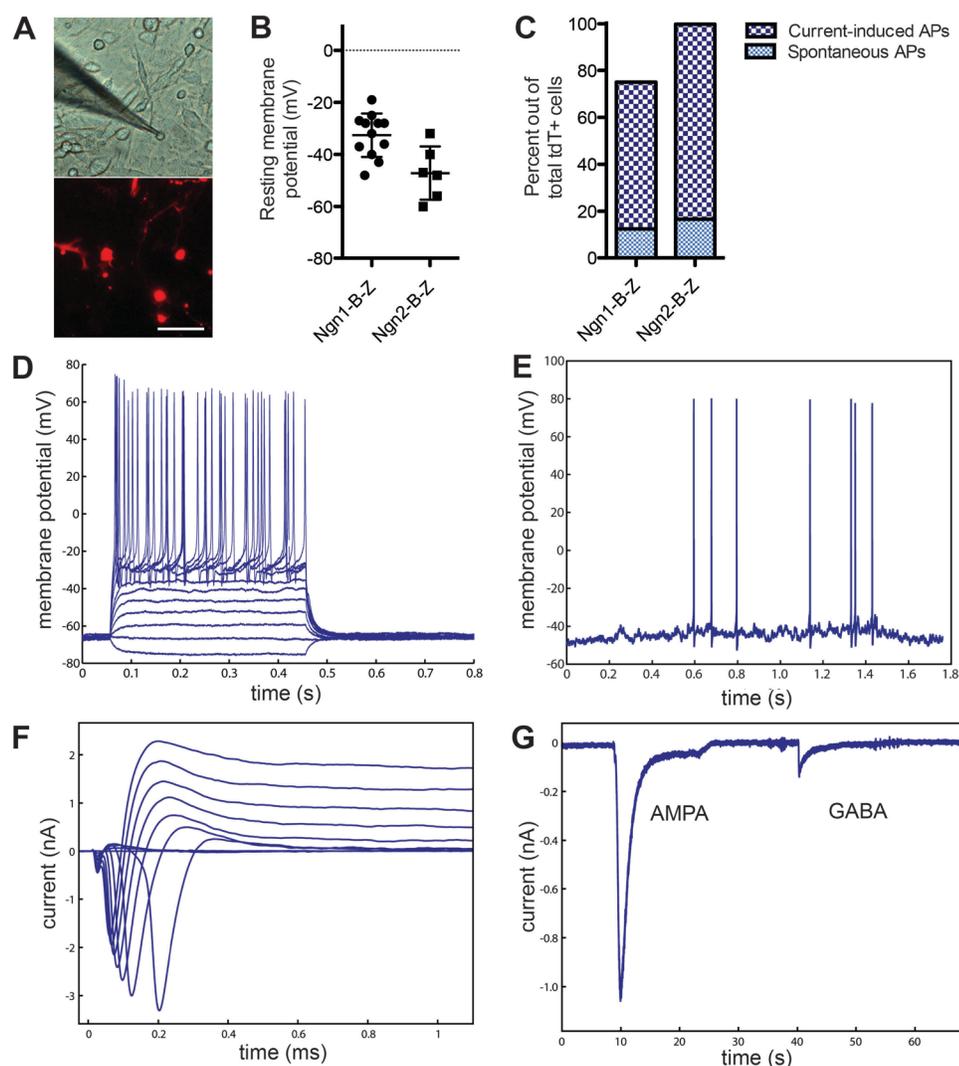
We further investigated ligand responses of the Pcdh21-TdTomato positive cells using calcium imaging by first transducing the cells with a lentivirus encoding the fluorescent calcium indicator GCaMP5.G under the control of a *Map2* promoter (*Map2::GCaMP5.G*) (Addis et al., 2011)<sup>†</sup>. Cells were sequentially exposed to NRS buffer, L-glutamate (1 mM) and KCl (25 mM) using direct application (Figure 2.6A-C). Buffer was added first to assess whether the cells were mechanosensitive, while KCl was applied last to confirm neuronal identity and functional *Map2::GCaMP5.G* expression. Only cells that did not respond to buffer, but did respond to KCl were included in these analyses (Figure 2.6A). 13 out of 16 *Ngn1-Brn2-Zic1* cells (83%) and 1 out of 4 *Ngn2-Brn2-Zic1* cells (25%) exhibited sharp calcium

---

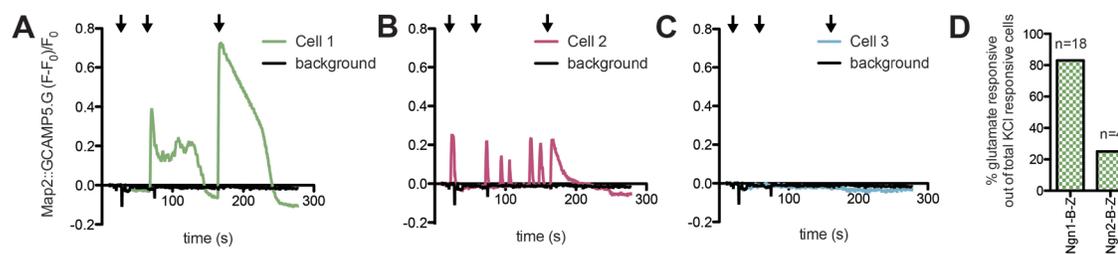
<sup>†</sup> *Map2::GCaMP5.G* is similar to the *Map2::GCaMP3* plasmid published in Addis et al. (2011), but with the *GCaMP3* replaced with *GCaMP5.G*.

fluctuations in response to L-glutamate. This suggests that the majority of Pcdh21-TdTomato and Map2-expressing cells are responsive to glutamate, a feature common to many endogenous neuron populations. However, this experiment needs to be repeated to confirm these findings.

Taken together, it is evident that the Pcdh21-TdTomato, Tuj1 double positive expressing cells are indeed functional induced neurons. Therefore, *Ascl1* is in fact not required for reprogramming fibroblasts into induced neurons, but can be replaced by *Ngn1* and *Ngn2*, when combined with *Brn2* and *Zic1*.



**Figure 2.5 Pcdh21-TdTomato positive cells with neuronal morphology exhibit electrophysiological properties of functional neurons.** (A) Representative images of a patch clamped Pcdh21-TdTomato positive cell (red) with neuronal morphology. Scale bar represents 100  $\mu\text{m}$ . (B) Quantification of the resting membrane potential of the patched cells ( $-32.6 \pm 8.4$  mV and  $-47.2 \pm 10.2$  mV for *Ngn1-B-Z* ( $n = 12$ ) and *Ngn2-B-Z* ( $n = 6$ ), respectively). (C) Majority of patched cells displayed either current-evoked action potentials (*Ngn1-B-Z* = 62.5% and *Ngn2-B-Z* = 83.3%) or spontaneous action potentials (*Ngn1-B-Z* = 12.5%,  $n = 16$ , and *Ngn2-B-Z* = 16.6%,  $n = 6$ ). Data are presented as the mean  $\pm$  SD. (D-G) Representative traces of Pcdh21-TdTomato positive cells showing (D) action potentials induced by depolarizing current ( $-100$  to  $400$  pA current clamp,  $50$  pA steps), (E) voltage-gated  $\text{Na}^+$  and  $\text{K}^+$  currents (voltage clamp from  $-60$  mV to  $+30$  mV,  $10$  mV steps), (F) spontaneous action potentials, and (G) ligand-gated ion channel activity from  $100$   $\mu\text{M}$  AMPA and  $200$   $\mu\text{M}$  GABA application (voltage clamped at  $-60$  mV).



**Figure 2.6 Pcdh21-TdTomato positive cells are glutamate responsive.** (A-C) Representative calcium traces of cells that responded to (A) L-glutamate and (B) buffer, and (C) for cells with no detected calcium responses. Calcium transients were measured in Pcdh21-TdTomato positive cells with neuronal morphology using Map2::GCaMP5.G. Calcium responses were calculated as the change in fluorescence intensity ( $\Delta F$ ) over the initial fluorescence intensity ( $F_0$ ) and normalized to background. Arrows (from left to right) correspond to application of NRS buffer, L-glutamate (1 mM) and KCl (25 mM), respectively. KCl was applied last to confirm neuronal identity and functional Map2::GCaMP5.G expression. (D) Quantification of the percent of L-glutamate responsive cells out of the total number of Pcdh21-TdTomato positive cells that responded to KCl, but not buffer alone. Data are presented from one experiment (*Ngn1-B-Z*,  $n = 18$ , *Ngn2-B-Z*,  $n = 4$ ).

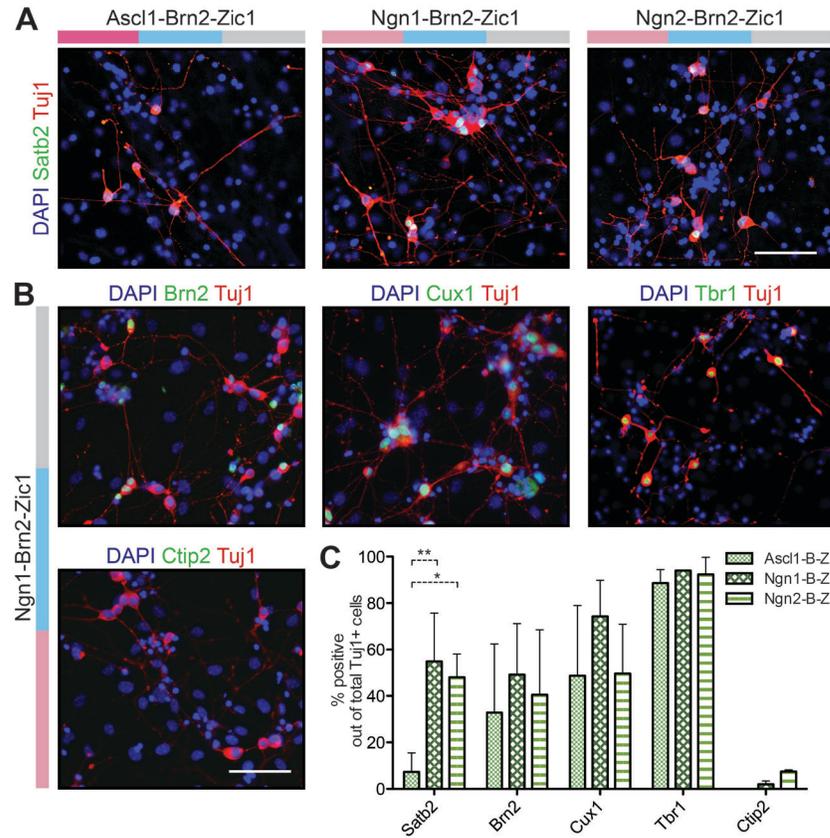
### 2.3.3 Induced neurons generated using *Ngn1* and *Ngn2* differ in expression of cortical neuron markers compared to induced neurons reprogrammed using *Ascl1*

To determine whether the iNs expressing the MT cell reporter, Pcdh21-TdTomato, were similar to endogenous MT cells, we immunostained the iNs for MT cell markers Tbr2 and Tbx21. We also immunostained for Pcdh21 to confirm the continued expression of the protein since we were using an irreversible Cre reporter line to label the cells. Surprisingly, we did not detect the expression of Tbr2, Tbx21, or Pcdh21 in the Pcdh21-TdTomato positive iNs (data not shown). Tbr2 and Pcdh21 were only detected if doxycycline was kept in the media throughout the reprogramming process (data not shown). Therefore, this suggests that the *transient* expression of *Ngn1-Brn2-Zic1* or *Ngn2-Brn2-Zic1* results in iNs that are not similar to endogenous MT cells.

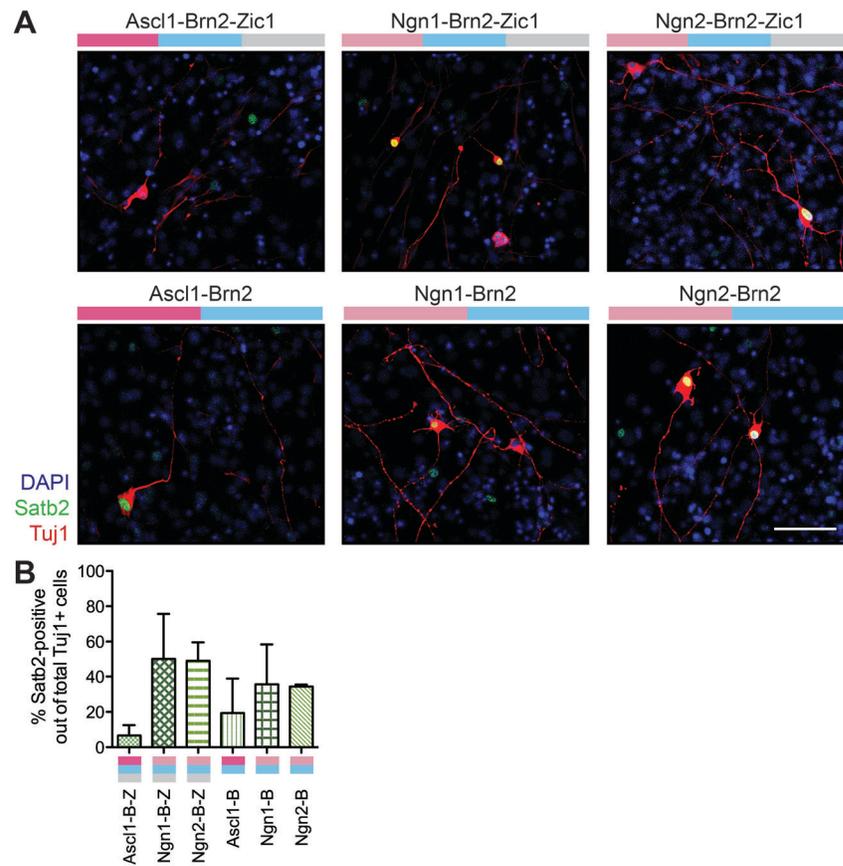
Since *Ngn1* and *Ngn2* are known to be involved in the specification of projection neurons of the cortex *in vivo* (Fode et al., 2000; Schuurmans et al., 2004), we next tested whether the iNs derived from *Ngn1* and *Ngn2* expressed cortical neuron markers. Indeed, iNs generated using *Ngn1-Brn2-Zic1* and *Ngn2-Brn2-Zic1* were positive for cortical neuron markers, *Satb2*, *Brn2*, *Cux1* and *Tbr1* (Figure 2.7). Interestingly, *Satb2*, *Brn2* and *Cux1* expression *in vivo* is limited to the upper-layers (layers II – IV) of the cortex. *Satb2*, a post-mitotic specifier of upper layer identity and marker of a subpopulation of upper layer cortical neurons (Britanova et al., 2008), was expressed in significantly greater percentages of iNs derived from *Ngn1* and *Ngn2* ( $54.8\% \pm 20.9\%$  and  $48.0\% \pm 10.0\%$ , respectively) compared to *Ascl1* ( $7.3\% \pm 8.1\%$ ) (Figure 2.7C). *Tbr1*, on the other hand, was detected in the majority of iNs derived from *Ngn1*, *Ngn2*, and *Ascl1*. Although *Tbr1* is expressed most predominantly in layer VI of the cortex, expression has been observed throughout the layers of the rostral cortex (Bedogni et al., 2010). *Ctip2*, a marker for deep-layer (layer V) cortical neurons, was only detected in less than 10% of the iNs (Figure 2.7C). Taken together, we have preliminary evidence that the iNs generated using *Ngn1-Brn2-Zic1* and *Ngn2-Brn2-Zic1* resemble *Satb2*-positive upper layer cortical neurons.

Intriguingly, *Brn2 in vivo* is critical for the proper production of upper layer cortical neurons (Doerks et al., 2002), while *Zic1* has been observed to inhibit neuronal differentiation during forebrain development (Inoue et al., 2007). Therefore, we subsequently tested whether *Ngn1* and *Ngn2* paired with

*Brn2* alone would be sufficient to generate iNs that express upper layer cortical neuron markers. *Ngn1-Brn2* and *Ngn2-Brn2* were indeed sufficient to reprogram MEFs into Tuj1- and Satb2-expressing cells (Figure 2.8).



**Figure 2.7** *Ngn1-Brn2-Zic1*- and *Ngn2-Brn2-Zic1*-derived induced neurons express upper layer cortical neuron markers. (A) Representative images of Tuj1-expressing cells (red) and upper layer cortical neuron marker, Satb2 (green). *Brn2* and *Zic1*, together with *Ngn1* or *Ngn2*, were able to generate significantly higher Satb2-expressing cells (green) compared to *Brn2* and *Zic1* combined with *Ascl1*. (B) Representative images of Tuj1-expressing cells (red) derived from *Ngn1-Brn2-Zic1* and counterstained with cortical neurons markers *Brn2*, *Cux1*, *Tbr1* and lower layer neuron marker, *Ctip2* (green). Images from (A) and (B) are of cells immunostained at day 14 post-induction. Scale bars represent 100  $\mu$ m. (C) Quantification of the percent of Tuj1-positive cells that express the cortical neuron marker. Data are presented as the mean  $\pm$  SD from six independent experiments for Satb2, *Brn2*, *Cux1* and *Tbr1* quantification and two independent experiments for *Ctip2* quantification. \*\*, \* represent p-values < 0.01, <0.05, respectively (Bonferroni's Multiple Comparison Test).



**Figure 2.8 *Brn2* paired with *Ngn1* or *Ngn2* generates *Satb2* and *Tuj1* expressing cells.** (A) Representative images of *Tuj1*-expressing cells (red) and upper layer cortical neuron marker, *Satb2* (green). *Brn2*, without *Zic1*, when paired with *Ngn1* or *Ngn2*, were still able to generate *Tuj1*- (red) and *Satb2*-expressing (green) cells. Images are of cells immunostained at day 18 post-induction. Scale bars represent 100  $\mu$ m. (B) Quantification of the percent of *Tuj1*-positive cells that express *Satb2*. Data are presented as the mean  $\pm$  SD from three independent experiments.

## 2.4 Discussion

These experiments show that *Ascl1* is not required to reprogram fibroblasts into neurons and establish that replacing *Ascl1* with either *Ngn1* or *Ngn2* can induce fibroblasts to attain morphological and electrophysiological characteristics of functional iNs. In support of these findings, our group and others have recently reported that *Ngn1* and *Ngn2* without *Ascl1* can generate functional iNs from fibroblasts (Blanchard et al., 2015; Liu et al., 2013). *Ngn2* has also been shown to directly generate induced neurons from pluripotent stem cells (Mazzoni et al., 2013; Thoma et al., 2012; Zhang et al., 2013) and astrocytes (Berninger et al., 2007; Heinrich et al., 2010). Even reprogramming of astrocytes to neurons using only a cocktail of small molecules involves the transcriptional activation of *Ngn2* and *NeuroD1* (Zhang et al., 2015). Collectively, these studies demonstrate that several members of the bHLH family, independent of *Ascl1*, are capable of generating induced neurons from many cell sources, from both mouse and human.

To test whether changing the identity of the bHLH gene would alter the cellular subtype identity of the resulting iNs, we tested iNs generated with different transcription factors for expression of markers of MT neurons. Although we observed the activation of one MT cell-reporter in the iNs derived from the *Ngn1*- and *Ngn2*- but not the *Ascl1*-reprogramming combinations, we did not detect the expression of other canonical MT cell-specific markers. This highlights the importance of using multiple independent and robust assays to test for the characteristics of the desired neuronal subtype. This is especially the case when using a Cre-reporter line that labels cells non-reversibly in contrast to a line with a real-time reporter. In fact, we only detected the expression of Pcdh21 protein when doxycycline exposure was maintained throughout the reprogramming experiment (data not shown). This suggests that either the induced neurons cannot maintain a MT cell identity without expression of the reprogramming factors or the factors exclusively activate Pcdh21 and no other MT identity pathways. We did observe expression of Tbr2, a marker of mature MT cells, in the Pcdh21-TdTomato positive iNs with constant doxycycline exposure. However, Tbr2 is also a marker of intermediate cortical progenitors (Englund et al., 2005), which would also imply that continued reprogramming factor expression could keep iNs in an immature state.

If we were to further pursue the direct generation of MT cells, we could continue testing additional transcription factors known to be involved in MT cell development *in vivo*. However, this strategy relies

heavily on already knowing potential regulators that specify the chosen neuronal subtype. Therefore, neuronal subtype populations successfully generated using direct reprogramming methods are limited to those that have been well studied in the field. Since only a handful of published studies have specifically focused on MT cell development at a transcriptional level (Campbell et al., 2011), it is possible that the transcription factors critical to generate induced MT cells are currently not cited as such. Another possibility, although harder to prove, is that MT cells cannot be generated using solely intrinsic cues.

We did find that *Ngn1-Brn2-Zic1* and *Ngn2-Brn2-Zic1* generated iNs that expressed markers of mature cortical neurons, similar to what has been reported on iNs generated using *Ascl1-Brn2-Myt1l* (Vierbuchen et al., 2010). This is not surprising since *Ngn1* and *Ngn2* are known to specify cortical and glutamatergic identity *in vivo* (Fode et al., 2000; Schuurmans et al., 2004). Intriguingly though, we found that using *Ngn1* and *Ngn2* versus *Ascl1* as a reprogramming factor led to an upregulation of the upper-layer cortical neuron marker, *Satb2*. This is unexpected since *in vivo*, upper-layer cortical neurons are specified in an *Ngn*-independent manner (Schuurmans et al., 2004). However, *Brn2* is a known specifier of upper-layer cortical identity (Doerks et al., 2002; Dominguez et al., 2013) and therefore, it may be the synergistic pairing of *Ngns* with *Brn2* that produce *Satb2*-positive iNs.

When we paired *Ascl1*, *Ngn1*, or *Ngn2* with *Brn2* alone, we observed a similar upregulation of *Satb2* in the *Ngn1* and *Ngn2* conditions compared to the *Ascl1* condition (although not statistically significant). This suggests that it is the pairing of the different bHLH transcription factors with *Brn2*, and not *Zic1*, that accounts for this difference. This is particularly interesting since concurrent to these findings, fellow colleagues found that transient co-expression of *Ngn1* or *Ngn2* with *Brn3a* was sufficient to produce iNs resembling peripheral sensory neurons (Blanchard et al., 2015). *Brn3a* (*Pou4f1*), like *Brn2* (*Pou3f2*), is in the POU family of transcription factors expressed in the murine nervous system. Together, these studies imply that different bHLH transcription factors paired with different POU transcription factors result in iNs that exhibit distinctly different patterns of gene expression and functionality.

Chapter 2, includes representative images of a *Pchd21:CRE x Ai9* mouse brain as seen in Figure 2.1.A that was imaged and provided by Jennifer Hazen, PhD. Xiaofei Zhang, PhD, conducted the

electrophysiological recordings as seen in Figure 2.5. Calcium recordings, as seen in Figure 2.6, were done in collaboration with Andrew Adler, PhD.

### References:

Addis, R.C., Hsu, F.C., Wright, R.L., Dichter, M.A., Coulter, D.A., and Gearhart, J.D. (2011). Efficient conversion of astrocytes to functional midbrain dopaminergic neurons using a single polycistronic vector. *PLoS One* 6, e28719.

Bedogni, F., Hodge, R.D., Elsen, G.E., Nelson, B.R., Daza, R.A., Beyer, R.P., Bammler, T.K., Rubenstein, J.L., and Hevner, R.F. (2010). *Tbr1* regulates regional and laminar identity of postmitotic neurons in developing neocortex. *Proc Natl Acad Sci U S A* 107, 13129-13134.

Berninger, B., Costa, M.R., Koch, U., Schroeder, T., Sutor, B., Grothe, B., and Gotz, 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, 8654-8664.

Blanchard, J.W., Eade, K.T., Szucs, A., Lo Sardo, V., Tsunemoto, R.K., Williams, D., Sanna, P.P., and Baldwin, K.K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nature neuroscience* 18, 25-35.

Britanova, O., de Juan Romero, C., Cheung, A., Kwan, K.Y., Schwark, M., Gyorgy, A., Vogel, T., Akopov, S., Mitkovski, M., Agoston, D., Sestan, N., Molnar, Z., and Tarabykin, V. (2008). *Satb2* is a postmitotic determinant for upper-layer neuron specification in the neocortex. *Neuron* 57, 378-392.

Caiazzo, M., Dell'Anno, M.T., Dvoretzkova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T.D., Menegon, A., Roncaglia, P., Colciago, G., Russo, G., Carninci, P., Pezzoli, G., Gainetdinov, R.R., Gustinich, S., Dityatev, A., and Broccoli, V. (2011). Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* 476, 224-227.

Campbell, G.R., Baudhuin, A., Vranizan, K., and Ngai, J. (2011). Transcription factors expressed in olfactory bulb local progenitor cells revealed by genome-wide transcriptome profiling. *Molecular and cellular neurosciences* 46, 548-561.

Chanda, S., Ang, C.E., Davila, J., Pak, C., Mall, M., Lee, Q.Y., Ahlenius, H., Jung, S.W., Sudhof, T.C., and Wernig, M. (2014). Generation of Induced Neuronal Cells by the Single Reprogramming Factor ASCL1. *Stem cell reports* 3, 282-296.

Doerks, T., Copley, R.R., Schultz, J., Ponting, C.P., and Bork, P. (2002). Systematic identification of novel protein domain families associated with nuclear functions. *Genome research* 12, 47-56.

Dominguez, M.H., Ayoub, A.E., and Rakic, P. (2013). POU-III transcription factors (*Brn1*, *Brn2*, and *Oct6*) influence neurogenesis, molecular identity, and migratory destination of upper-layer cells of the cerebral cortex. *Cerebral cortex (New York, NY : 1991)* 23, 2632-2643.

Englund, C., Fink, A., Lau, C., Pham, D., Daza, R.A., Bulfone, A., Kowalczyk, T., and Hevner, R.F. (2005). *Pax6*, *Tbr2*, and *Tbr1* are expressed sequentially by radial glia, intermediate progenitor cells, and postmitotic neurons in developing neocortex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* 25, 247-251.

Fode, C., Ma, Q., Casarosa, S., Ang, S.L., Anderson, D.J., and Guillemot, F. (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes & Development* *14*, 67-80.

Heinrich, C., Blum, R., Gascon, S., Masserdotti, G., Tripathi, P., Sanchez, R., Tiedt, S., Schroeder, T., Gotz, M., and Berninger, B. (2010). Directing astroglia from the cerebral cortex into subtype specific functional neurons. *PLoS biology* *8*, e1000373.

Inoue, T., Ota, M., Ogawa, M., Mikoshiba, K., and Aruga, J. (2007). *Zic1* and *Zic3* regulate medial forebrain development through expansion of neuronal progenitors. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *27*, 5461-5473.

Kim, E.J., Battiste, J., Nakagawa, Y., and Johnson, J.E. (2008). *Ascl1* (*Mash1*) lineage cells contribute to discrete cell populations in CNS architecture. *Molecular and cellular neurosciences* *38*, 595-606.

Kim, E.J., Hori, K., Wyckoff, A., Dickel, L.K., Koundakjian, E.J., Goodrich, L.V., and Johnson, J.E. (2011a). Spatiotemporal fate map of neurogenin1 (*Neurog1*) lineages in the mouse central nervous system. *The Journal of comparative neurology* *519*, 1355-1370.

Kim, J., Su, S.C., Wang, H., Cheng, A.W., Cassady, J.P., Lodato, M.A., Lengner, C.J., Chung, C.Y., Dawlaty, M.M., Tsai, L.H., and Jaenisch, R. (2011b). Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. *Cell Stem Cell* *9*, 413-419.

Liu, M.-L., Zang, T., Zou, Y., Chang, J.C., Gibson, J.R., Huber, K.M., and Zhang, C.-L. (2013). Small molecules enable neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons. *Nat Commun* *4*.

Ma, Q., Sommer, L., Cserjesi, P., and Anderson, D.J. (1997). *Mash1* and neurogenin1 expression patterns define complementary domains of neuroepithelium in the developing CNS and are correlated with regions expressing notch ligands. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *17*, 3644-3652.

Mazzoni, E.O., Mahony, S., Closser, M., Morrison, C.A., Nedelec, S., Williams, D.J., An, D., Gifford, D.K., and Wichterle, H. (2013). Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nature neuroscience* *16*, 1219-1227.

Mullen, R.J., Buck, C.R., and Smith, A.M. (1992). NeuN, a neuronal specific nuclear protein in vertebrates. *Development* *116*, 201-211.

Osorio, J., Mueller, T., Retaux, S., Vernier, P., and Wullimann, M.F. (2010). Phylotypic expression of the bHLH genes *Neurogenin2*, *Neurod*, and *Mash1* in the mouse embryonic forebrain. *The Journal of comparative neurology* *518*, 851-871.

Pfisterer, U., Kirkeby, A., Torper, O., Wood, J., Nelander, J., Dufour, A., Bjorklund, A., Lindvall, O., Jakobsson, J., and Parmar, M. (2011). Direct conversion of human fibroblasts to dopaminergic neurons. *Proc Natl Acad Sci U S A* *108*, 10343-10348.

Schuurmans, C., Armant, O., Nieto, M., Stenman, J.M., Britz, O., Klenin, N., Brown, C., Langevin, L.-M., Seibt, J., Tang, H., Cunningham, J.M., Dyck, R., Walsh, C., Campbell, K., Polleux, F., and Guillemot, F. (2004). Sequential phases of cortical specification involve Neurogenin-dependent and -independent pathways. *The EMBO journal* *23*, 2892-2902.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* 9, 205-218.

Thoma, E.C., Wischmeyer, E., Offen, N., Maurus, K., Siren, A.L., Scharl, M., and Wagner, T.U. (2012). Ectopic expression of neurogenin 2 alone is sufficient to induce differentiation of embryonic stem cells into mature neurons. *PLoS One* 7, e38651.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035-1041.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., Drechsel, D., Martynoga, B., Castro, D.S., Webb, A.E., Sudhof, T.C., Brunet, A., Guillemot, F., Chang, H.Y., and Wernig, M. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* 155, 621-635.

Winpenny, E., Lebel-Potter, M., Fernandez, M.E., Brill, M.S., Gotz, M., Guillemot, F., and Raineteau, O. (2011). Sequential generation of olfactory bulb glutamatergic neurons by Neurog2-expressing precursor cells. *Neural development* 6, 12.

Zhang, L., Yin, J.C., Yeh, H., Ma, N.X., Lee, G., Chen, X.A., Wang, Y., Lin, L., Chen, L., Jin, P., Wu, G.Y., and Chen, G. (2015). Small Molecules Efficiently Reprogram Human Astroglial Cells into Functional Neurons. *Cell Stem Cell*.

Zhang, Y., Pak, C., Han, Y., Ahlenius, H., Zhang, Z., Chanda, S., Marro, S., Patzke, C., Acuna, C., Covy, J., Xu, W., Yang, N., Danko, T., Chen, L., Wernig, M., and Sudhof, T.C. (2013). Rapid single-step induction of functional neurons from human pluripotent stem cells. *Neuron* 78, 785-798.

## **Chapter 3:**

### **Identification of novel neuronal reprogramming factors using a large-scale screen**

#### **3.1 Introduction**

Colleagues have demonstrated that two transcription factors, a bHLH factor paired with a POU factor, are sufficient to reprogram fibroblasts into functional induced sensory neurons (Blanchard et al., 2015). We have also provided evidence that swapping the bHLH and POU factors with corresponding family members results in iNs that exhibit distinctly different patterns of gene expression and functionality (Chapter 2 and Blanchard et al. (2015)). This is not surprising since, as previously described, different proneural bHLH transcription factors coordinate neurogenesis in defined areas of the nervous system (Fode et al., 2000; Ma et al., 1997; Osorio et al., 2010). The same is true for POU transcription factors; although *Brn1* and *Brn2* are the most widely expressed in the brain, other POU factors such as *Brn3a*, *Brn3b*, *Brn3c* and *Pit1* are expressed in more restricted patterns throughout the nervous system (He et al., 1989). Therefore, the overlap between the distinct patterns of expression between each bHLH and POU transcription factor during neurodevelopment could provide a combinatorial code that helps specify the many neuronal phenotypes of the nervous system.

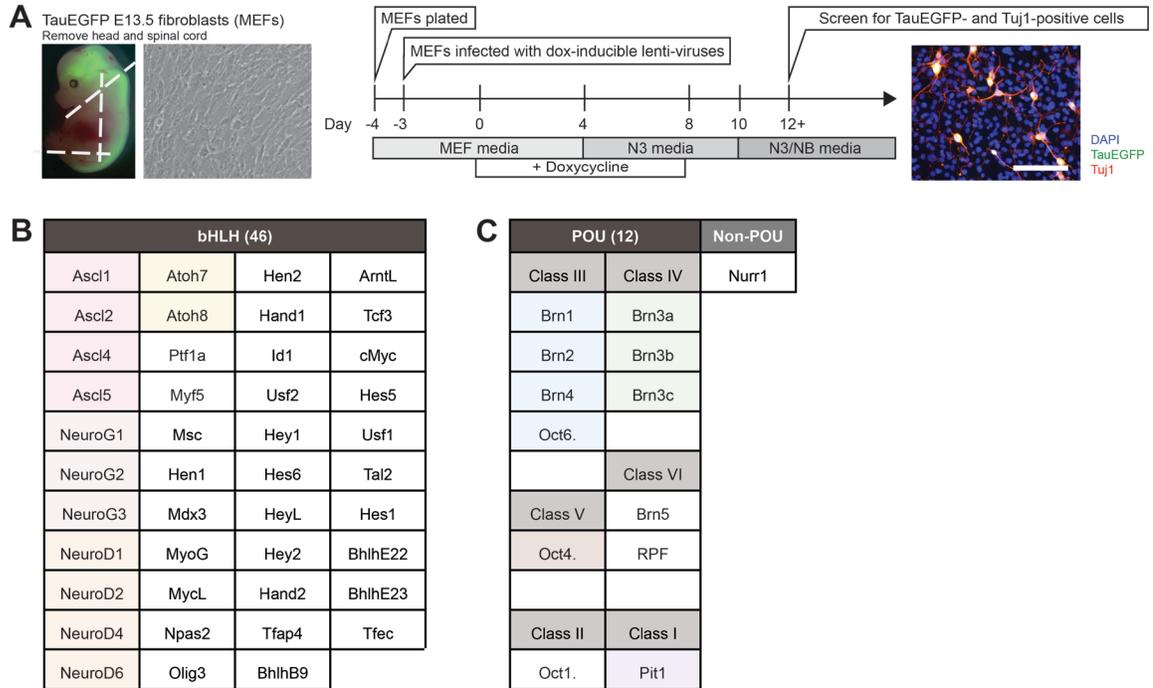
Mechanistically, the ability of bHLH factors to specifically target promoters of neuron-related genes has been attributed to their functional interactions with cofactors (Lee and Pfaff, 2003), including members of the POU family. For example, cooperative binding of *Ascl1* with POU proteins, *Brn1* (*Pou3f3*) and *Brn2* (*Pou3f2*) activates genes related to neurogenesis, including, mouse *Delta1* (Castro et al., 2006; Josephson et al., 1998). During reprogramming from fibroblast to neuron, *Ascl1* is known to recruit *Brn2* to those neuron-related gene targets (Soufi et al., 2015; Wapinski et al., 2013). Taken together, we hypothesized that additional bHLH and POU factor pairings could reprogram fibroblasts into functional neurons and that different pairings would result in markedly diverse iN populations.

#### **3.2 Methods: Large-scale screening to identify transcription factor pairs capable of producing neurons directly from fibroblasts**

To identify additional pairs of bHLH and POU transcription factors capable of producing neurons directly from fibroblasts, we conducted a large-scale screen. First, we identified bHLH and POU transcription factors reported to be expressed in the mouse nervous system at any point during development or adulthood<sup>‡</sup>. We cloned cDNAs for as many of the identified genes into doxycycline (dox)-inducible lentiviral vectors for a total of 46 bHLH (Figure 3.1B), 12 POU transcription factors and one nuclear receptor (NR) transcription factor for comparison (Figure 3.1C). Next, we transduced MEFs with each bHLH factor paired with either one POU factor or NR factor, thus screening 598 distinct combinations, and followed the same reprogramming protocol as described in Chapter 2.2 (Figure 3.1A). MEFs were cultured from E13.5 pups of the neuronal reporter knock-in mouse strain, TauEGFP, or wild-type mice. Factor pairs that produced Tuj1 positive cells were further characterized using immunostaining and whole transcriptome analyses. Detailed methods are provided in Appendix A.2.

---

<sup>‡</sup> Although the POU transcription factor *Oct4* is not expressed in the nervous system, we included it in our screen because of its unique expression in ESCs and its role in iPSC reprogramming.



**Figure 3.1 Experimental design of a reprogramming screen using bHLH and POU transcription factors.** (A) Schematic overview of the methodology of the screen. Transient co-expression of reprogramming factor pairs are induced in E13.5 MEFs one day post plating (left image) for the following 8 days and screened for TauEGFP (green) and Tuj1 (red) positive cells at day 14 post-induction (right image). Scale bars represent 100  $\mu$ m. (B) List of bHLH and (C) POU and NR transcription factors cloned and used in the screen. Factors are color-coded based on subclass assignment.

### 3.3 Results

#### 3.3.1 Screen reveals novel bHLH and POU transcription factor pairs capable of reprogramming fibroblasts into candidate induced neurons

Fascinatingly, in 76 (12.7%) of the 598 conditions tested, we observed Tuj1-positive cells exhibiting stereotypical neuronal morphology at greater than 0.01% efficiency (percent of Tuj1-positive cells out of total number of cell plated) after 14 days of reprogramming (Figure 3.2, Figure 3.3, Figure 3.4). The 76 positive factor combinations were comprised of distinct pairings between 16 bHLH, 9 POU, and 1 NR factors. Of the 76 pairs, 72 were novel combinations not previously reported to generate neuronal-like cells directly from fibroblasts. Comparisons of efficiency between combinations, however, should take into account the variability in viral titer between the factors tested. Additionally, we cannot exclude the possibility of false negatives based on this same reasoning. In control experiments using individual bHLH, POU, and NR transcription factors, we only detected Tuj1-positive cells in conditions containing *Ascl1*, *Ascl2*, *Neurog1* (*Ngn1*), and *Neurog3* (*Ngn3*), but at lower numbers than in the reported 76 positive conditions (Figure 3.3, Figure 3.4C). No Tuj1-positive cells were found in untreated MEFs (Figure 3.4D).

To address the possibility that the observed Tuj1-positive cells arose from contaminating neural crest cells in the source MEF cultures, despite removing neural tissue, we used fluorescence activated cell sorting (FACS) to reduce the number of cells expressing p75, a marker of neural crest cells, prior to reprogramming (1.6% down to 0.9%,  $n = 3$ , Figure 3.5A-B). After reprogramming, the percentages of Tuj1-positive cells generated from the source and p75-depleted populations were not significantly different from each other (Figure 3.5C). Additionally, to determine whether embryonic cells were required for reprogramming, we infected tail-tip fibroblasts (TTFs) from 3-day-old pups with a subset of the positive factor combinations. Following the same MEF-reprogramming protocol, Tuj1-positive neuronal-like cells were observed in all 16 combinations tested on TTFs (Figure 3.6). Although these results do not eliminate the possibility that some of the Tuj1-positive cells arose from precursor cells, taken together, this suggests that the newly discovered transcription factor pairs do not require embryonic precursor cells to induce Tuj1 expression and neuronal morphology.

In addition to expressing neuronal marker Tuj1 and exhibiting stereotypical neuronal morphology, these same cells expressed EGFP when factors were transduced in MEFs from TauEGFP knock-in mice that specifically express EGFP in neurons (Figure 3.7). Tuj1 positive cells also expressed markers of mature neurons, including Map2 (85-99% across 5 populations) and the presynaptic protein, Synapsin (86-98%) (Figure 3.7). Collectively, this suggests that the observed Tuj1-positive cells have similar expression profiles as endogenous neurons. Therefore, we will refer to the Tuj1-positive cells as candidate induced neurons.

**A**

POU		Ascl				Neurogenin			NeuroD				Atoh1					
		Ascl1	Ascl2	Ascl4	Ascl5	Ngn1	Ngn2	Ngn3	ND1	ND2	ND4	ND6	Atoh1	Atoh7	Atoh8	Myf5	Ptf1a	None
Class IV	Brn3c	5.7	3.0	0.02	0.95	0.42	0.29	2.0	0.16	0.31		0.02	0.30	0.04		0.05		
	Brn3a	3.2	1.8	0.02	0.29	0.33	0.32	0.82	0.03	0.63	0.06	0.02	0.26	0.11	0.01			
	Brn3b	2.8	2.7			0.21	0.28	0.97	0.20	0.33			0.05			0.04	0.01	
Class III	Brn2	1.6	1.4			0.76	0.56	1.8	0.07	0.02			0.01					
	Brn4	0.20	0.15			0.36	0.13	3.0	0.04									
	Brn1					0.66	0.09	0.15	0.02									
	Oct6.	0.03				0.13	0.02	0.11	0.01									
I	Pit1	0.13	0.25			0.02	0.03	0.34	0.06									
V	Oct4.	1.1	0.55			0.06	0.05	0.84	0.07	0.01			0.02					
II	Oct1.																	
VI	RPF																	
	Brn5																	
	None																	
	Nurr1	0.05	0.03															

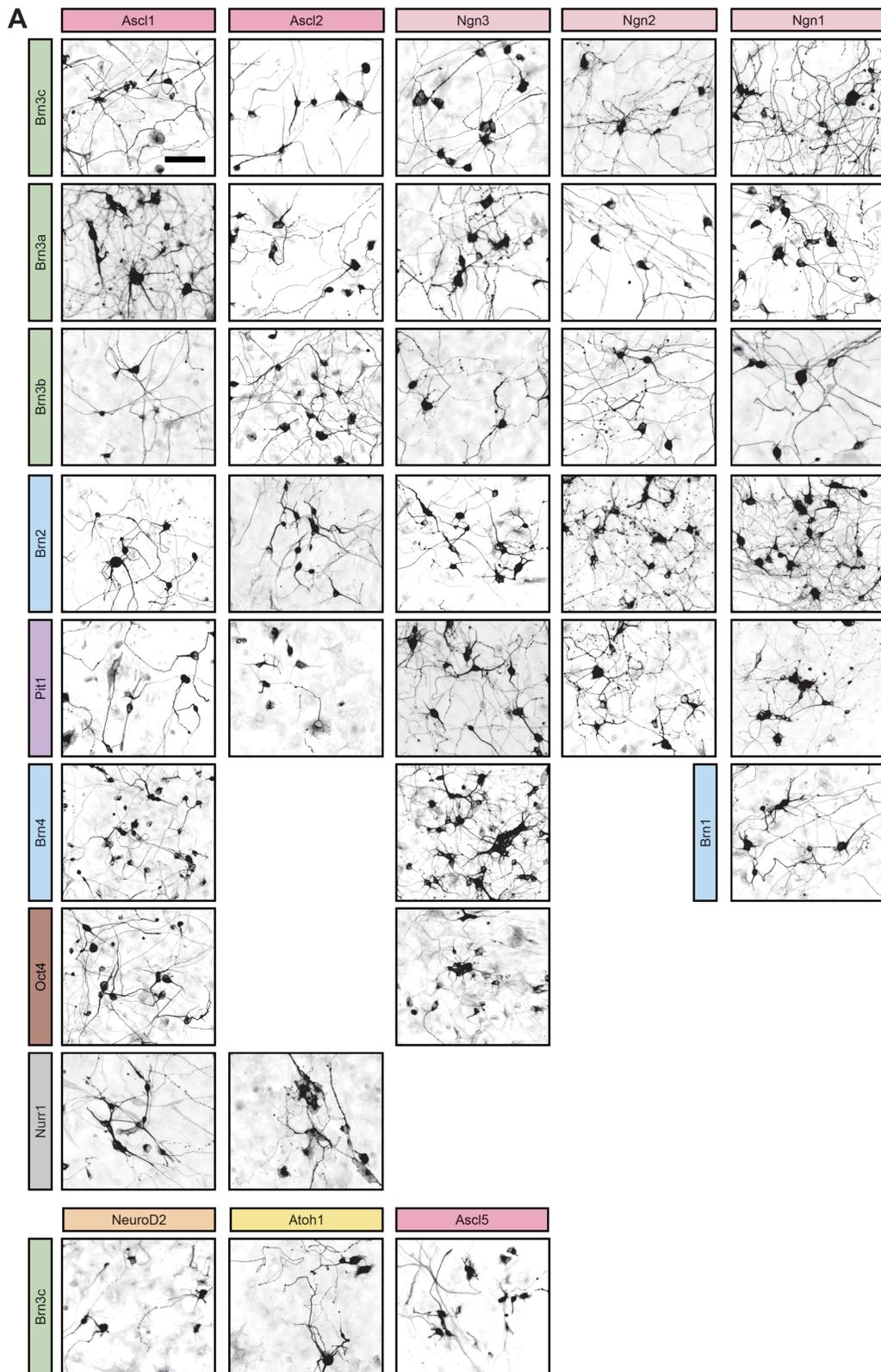
**Figure 3.2 bHLH and POU transcription factor pairs capable of reprogramming fibroblasts into Tuj1-positive cells.** (A) Matrix of pairwise reprogramming factor combinations yielding Tuj1-positive cells from fibroblasts on day 14 post-induction. Numbers reflect the percentages of Tuj1-positive cells out of the total number of fibroblasts plated. Normalized percentages from combinations that included bHLH factors *Ascl1*, *Ascl2*, *Ngn1* and *Ngn3* were calculated by subtracting the percentage of Tuj1-positive cells generated from the bHLH factors alone (this may result in an underestimation of total efficiency). Green cells highlight combinations that yielded Tuj1-positive cells greater than 0.01% efficiency after normalization, with darker green cells marking combinations with higher efficiency. Grey cells represent negative percentages after normalization. Blank cells resulted in no observable Tuj1-positive cells after normalization. Pairwise combinations not shown in matrix also resulted in no observable Tuj1-positive cells.

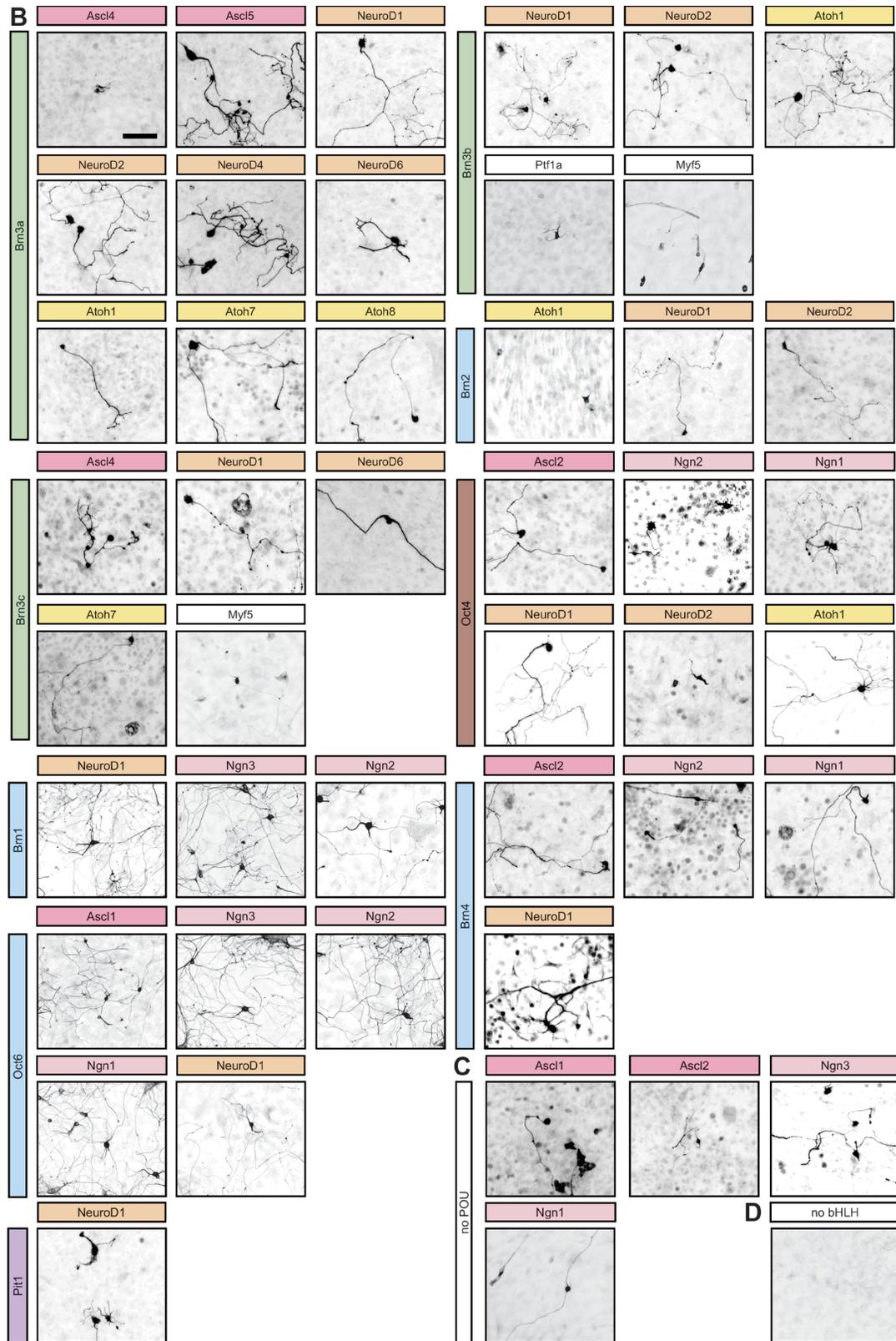
**A**

POU		Ascl				Neurogenin			NeuroD				Atoh1			Myf5	Ptf1a	None
		Ascl1	Ascl2	Ascl4	Ascl5	Ngn1	Ngn2	Ngn3	ND1	ND2	ND4	ND6	Atoh1	Atoh7	Atoh8			
Class IV	Brn3c	6.06	3.1	0.02	0.95	0.43	0.29	2.14	0.16	0.31	0	0.02	0.3	0.04	0	0	0.05	0
	Brn3a	3.59	1.88	0.02	0.29	0.34	0.32	1.02	0.03	0.63	0.06	0.02	0.26	0.11	0.01	0	0	0
	Brn3b	3.23	2.73	0	0	0.21	0.28	1.16	0.2	0.33	0	0	0.05	0	0	0.01	0.04	0
III	Brn2	2.02	1.49	0	0	0.77	0.56	1.95	0.07	0.02	0	0	0.01	0	0	0	0	0
	Brn4	0.59	0.22	0	0	0.37	0.13	3.18	0.04	0	0	0	0	0	0	0	0	0
V	Oct4.	1.52	0.62	0	0	0.07	0.05	1.03	0.07	0.01	0	0	0.02	0	0	0	0	0
I	Pit1	0.52	0.32	0	0	0.03	0.03	0.53	0.06	0	0	0	0	0	0	0	0	0
VI	Brn5	0.35	0.06	0	0	0	0	0.17	0	0	0	0	0	0	0	0	0	0
	None	0.39	0.07	0	0	0.01	0	0.19	0	0	0	0	0	0	0	0	0	0
																		0
III	Brn1	0	0	0	0	0.66	0.09	0.15	0.02	0	0	0	0	0	0	0	0	0
	Oct6.	0.07	0	0	0	0.13	0.02	0.11	0.01	0	0	0	0	0	0	0	0	0
V	RPF	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
II	Oct1.	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	None	0.04	0	0	0	0	0	0.01	0	0	0	0	0	0	0	0	0	0
	Nurr1	0.09	0.03	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

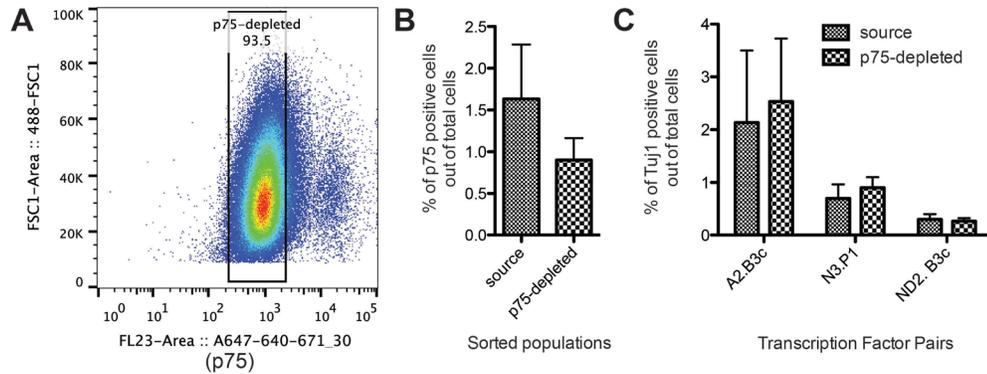
**Figure 3.3 Non-normalized Tuj1 percentages resulting from bHLH and POU transcription factor pair overexpression** (A) Non-normalized matrix of pairwise reprogramming factors yielding Tuj1-positive cells from fibroblasts on day 14 post-induction. Numbers reflect the percentages of Tuj1-positive cells out of the total number of fibroblasts plated from two separate screen experiments.

**Figure 3.4 Representative images of Tuj1 immunofluorescence labeling for the positive bHLH and POU pairs.** (A) Tuj1 immunofluorescence labeling of 35 of the 76 positive combinations that were selected for whole-transcriptome analysis. (B) Tuj1 immunofluorescence labeling of the remaining 41 of the 76 positive combinations. (C) Tuj1 immunofluorescence labeling of conditions with individual bHLH factors *Ascl1*, *Ascl2*, *Ngf1* and *Ngf3*. (D) Tuj1 immunofluorescence labeling of MEFs treated with only rtTA, without reprogramming factors. Scale bars represent 100  $\mu\text{m}$ .

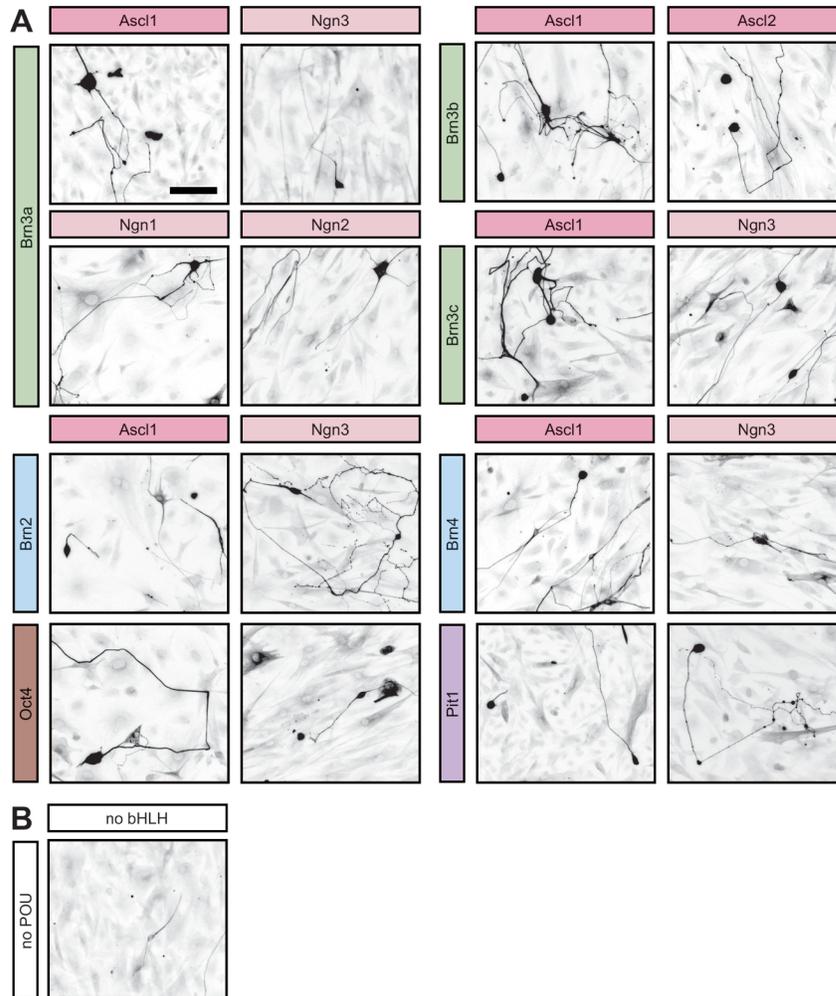




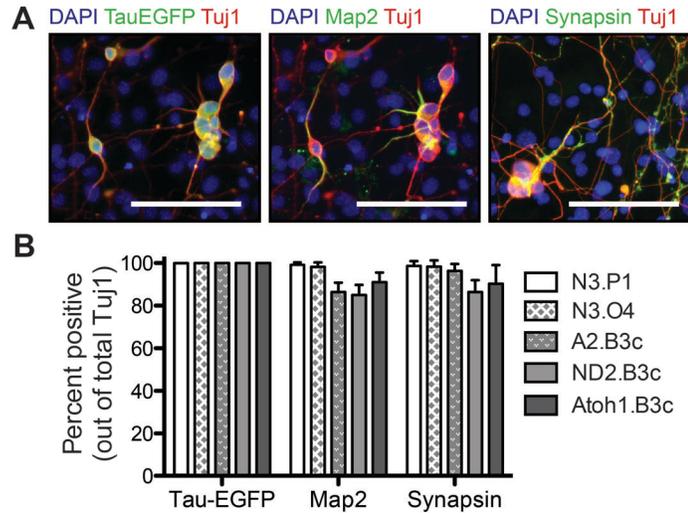
**Figure 3.4** Representative images of Tuj1 immunofluorescence labeling for the positive bHLH and POU pairs, Continued.



**Figure 3.5 TuJ1 positive cells generated from p75-depleted MEF populations.** (A) Representative FACS gates of MEFs (~180,000 cells shown). MEFs were depleted of p75-positive neural crest cells, by first gating for DAPI-negative cells (not shown) and collecting only those that were p75-negative (~93% of the DAPI-negative population). (B) Quantification of immunostaining for p75 positive cells in source and p75-depleted MEF populations after expansion for 4 days post-FACS, the day of transduction for reprogramming. Data are presented as the mean  $\pm$  SD,  $n = 3$ . (C) Quantification of the percent of TuJ1 positive cells derived from source and p75-depleted MEF populations 16 days post-induction. Data are presented as the mean  $\pm$  SD,  $n = 3$ . Percentages between source and p75-depleted are not significantly different ( $p$ -value  $> 0.05$ , Bonferroni multiple comparisons).



**Figure 3.6 TuJ1 positive cells derived from postnatal tail tip fibroblasts.** (A) TuJ1 immunofluorescence labeling of TTFs derived from 3-day-old mice and transduced with select reprogramming combinations. Fixed and stained on day 16 post-induction. (B) TuJ1 immunofluorescence of TTFs treated with only rtTA, without reprogramming factors. Scale bars represent 100  $\mu$ m.



**Figure 3.7 TauEGFP, Tuj1 positive cells express markers of mature neurons** (A) Immunofluorescence labeling of Tuj1 positive cells for TauEGFP and neuronal markers, Map2, and synapsin. Representative candidate induced neurons were generated with reprogramming pairs *Ngn3* and *Pit1* 12 to 16 days post-induction. Scale bars represent 100  $\mu$ m. (B) Quantification of immunofluorescence. Data are presented as the mean  $\pm$  SD from at least three independent experiments.

### 3.3.2 Whole-transcriptome analysis of candidate iN populations

To determine the extent to which the candidate iN populations resembled endogenous neuron populations, we systematically performed whole-transcriptome analyses on select iN, MEF, and endogenous neuronal subtype populations (Figure 3.8A). For the candidate iN populations, we selected 35 of the 76 positive combinations based on efficiency and those with unique synergistic properties. To enrich for candidate iNs prior to RNA-seq library preparation, we first transduced the selected combinations in TauEGFP MEFs. TauEGFP served as a good marker for the candidate iNs since greater than 98% of Tuj1-positive cells were TauEGFP positive and 100% of TauEGFP positive cells were Tuj1-positive, as observed by co-immunofluorescence staining (Figure 3.9A-B). At day 16 of reprogramming, we used FACS to isolate the TauEGFP positive cells from MEFs (~93% purity) (Figures 3.9C-E, Appendix A.2). RNA-seq libraries were then generated from two biological replicates for each of the purified candidate iN populations, as well as, MEFs transduced with rtTA only and maintained in parallel with iN reprogramming. We obtained high quality total RNA from all purified populations with RNA integrity numbers (RIN) ranging from 6 to 10 (median = 8.7). Approximately 1,500 to 2,000 cells were required to obtain 10 ng of total RNA, which served as input for SMARTer Ultra Low Input RNA Kit for Sequencing – v3 and subsequently prepped for Illumina TruSeq RNA-seq. A few replicate libraries were prepped from 2 or 5 ng of input total RNA (Table 3.1). These lower input libraries were comparable to libraries prepped from 10 ng of RNA since correlation coefficients were greater than 0.98 between libraries prepped from 1, 5, and 10 ng of the same total RNA (Figure 3.9F-H). Libraries were sequenced to a mean of ~37.5 million uniquely mapped 75 base pair single-end reads per replicate (Table 3.1) and aligned and assembled as detailed in Appendix A.2.

Similar methods were also used to isolate cells and prepare RNA-Seq libraries of endogenous neuronal subtype populations (Table 3.2). In efforts to represent the diversity of subtypes found in the nervous system, we selected endogenous neuronal subtypes from both the central and peripheral nervous system and of excitatory and inhibitory identity. Neuronal populations from the olfactory bulb (OB), cortex (CTX), hippocampus (HIP), medial habenula (mHb), cerebellum (CER), and DRG were isolated

using appropriate fluorescence reporter mouse lines and FACS. Whole brain RNA provided by Clontech was also prepared for sequencing.

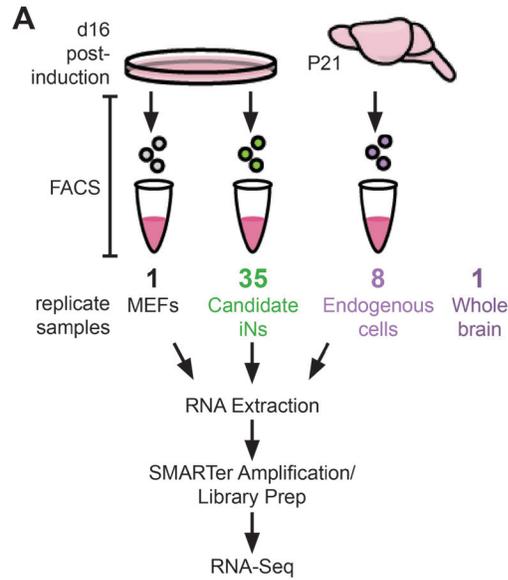
To explore the relationship of the candidate iN populations (TauEGFP positive cells) with fibroblasts and the representative endogenous neuronal and whole brain populations, we conducted principal component analysis (PCA) on all groups (Figure 3.10). The first principal component (PC1) accounted for 19.1% of the variability in this dataset. This component appeared to place the candidate iNs in between the MEF starting population and the endogenous neurons, suggesting that the candidate iN populations exhibit similar expression profiles to endogenous neurons while retaining some expression of fibroblast genes. Intriguingly, principal component 2 (PC2) and 3 (PC3), which accounted for 16.7% and 8.7% of the variability respectively, appeared to reflect similarities and differences between individual populations within the iN and endogenous neuron groups (Figure 3.10).

To further investigate whether the candidate iNs share similar expression profiles with endogenous neurons, fibroblasts, or both populations, we next used DESeq2 (Love et al., 2014) to identify significant differentially expressed genes between the 35 TauEGFP positive, candidate iN populations and source MEFs ( $p$ -adjusted  $< 0.05$ ). Gene ontology (GO) terms overrepresented in the 3,860 significantly upregulated genes of the candidate iNs compared to MEFs revealed terms of biological processes (BP level 5) highly associated with neuronal development and function according to DAVID (Dennis et al., 2003) (Figure 3.11A). Of the top ten significant GO terms ranked by fold enrichment, six were related to synaptic function, suggesting that the candidate iNs share features with synaptically mature neurons. In contrast, GO terms overrepresented in the 3,467 significantly downregulated genes of the candidate iNs were associated with immune function and cell division (Figure 3.11A). Since endogenous neurons are post-mitotic and MEFs are proliferative, the downregulation of cell cycle-related genes in the candidate iNs further supports their successful conversion from fibroblasts to neuronal-like cells.

Next, we determined that the upregulated genes in the candidate iN populations compared to MEFs were indeed neuronal genes found in endogenous neuronal populations. This is visually depicted in a volcano plot of  $\log_2$  fold change versus  $-\log_{10}$   $p$ -adjusted value for each gene when comparing MEFs to the sequenced endogenous neuron and whole brain populations using DESeq2 (Figure 3.11B). Of the

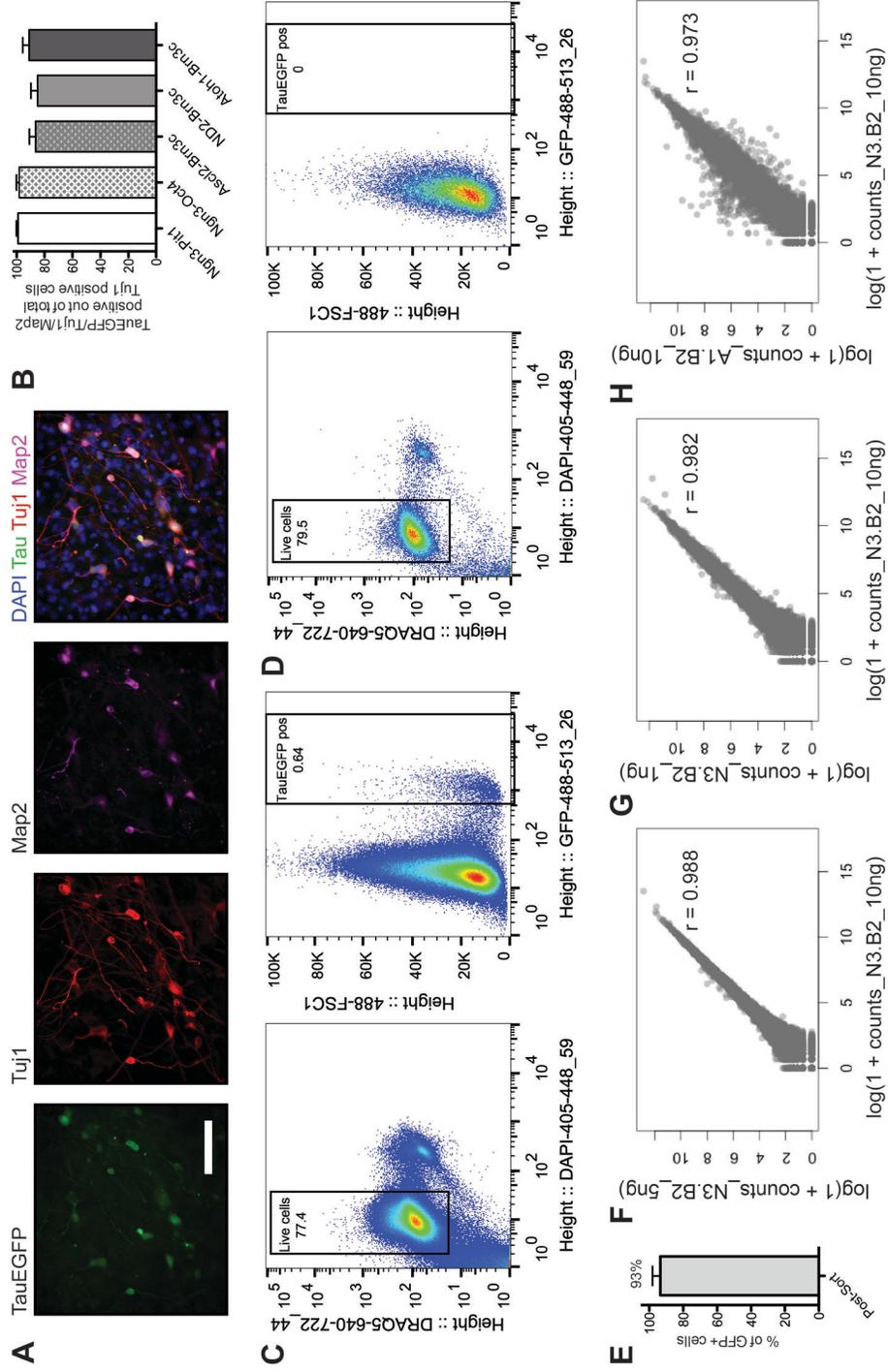
2,965 significantly upregulated genes in the endogenous neurons and whole brain populations compared to MEFs, 75.5% (2,239) of the genes were also upregulated in the candidate iNs (orange dots). Only 24.5% of genes were unique to the endogenous neuron and whole brain population (purple dots). The high percentage of shared genes between the candidate iNs and endogenous neuronal populations is comparable to the percentages observed when comparing the upregulated genes of each individual endogenous neuron population to the upregulated genes of the endogenous neuronal populations as a whole ( $78\% \pm 7\%$ ). Indeed, the expression of the top 100 significantly upregulated genes in the endogenous neuronal and whole brain populations compared to MEFs were also high in the candidate iN populations (Figure 3.12A, top half). Together, this suggests that the candidate iNs populations collectively share similar gene expression profiles to endogenous neurons.

Of the remaining upregulated genes in the candidate iNs not shared with the endogenous neuronal populations (42%, green dots, Figure 3.10B), the majority fell below the significance line (0.05 p-adjusted value). Only 19 genes (< 1%) were significantly upregulated in the fibroblast population, which included nine Hox family genes. Hox genes are known patterning genes expressed both in neuronal and fibroblast populations, but may only be significantly upregulated in the fibroblast populations because the selected endogenous neuronal populations do not encompass all neuronal populations along the anterior-posterior axis (Philippidou and Dasen, 2013). Additionally, the candidate iN populations had low expression levels of the top 100 significantly upregulated genes in MEFs. However, expression levels were not as reduced as those seen in the endogenous neuronal and whole brain samples (Figure 3.12A, bottom half). This implies that there is residual, low level expression of MEF genes in the candidate iN populations. This could be explained by the presence of fibroblasts and/or cells in the intermediate stages of transitioning from fibroblast to neuron in the sorted iN populations. Alternatively, this could also reflect the persistence of fibroblast gene expression in the candidate iNs, which might be reduced by culturing the iNs for a longer period of time, co-culturing them with other neural cell types such as glia, or further refining reprogramming methods.



**Figure 3.8 Workflow schematic of sample preparation for RNA sequencing (RNA-Seq).** (A) Cell populations were purified using FACS and libraries were prepped using SMARTer amplification. RNA-Seq was conducted on duplicate populations of TauEGFP positive cells generated from 35 different transcription factor pairs. Additional sequenced populations included MEFs, representative endogenous neuronal populations, and whole brain samples.

**Figure 3.9 Fluorescence-activated cell sorting and RNA-Seq library preparation of candidate iN populations.** (A) Representative immunofluorescence labeling of TauEGFP positive candidate induced neuron population (*Ascl2/Brn3b*) on day 12 post-induction using neuronal antibodies, Tuj1 and Map2. Scale bars represent 100  $\mu\text{m}$ . (B) Quantification of co-labeling between the TauEGFP reporter, Tuj1 and Map2 on day 12 post-induction calculated from various reprogramming pairs. Data are presented as the mean  $\pm$  SD from at least three independent experiments. (C-D) Representative FACS gates of (C) an *Ascl2/Brn3b* induced neuron population (500,000 cells shown) and (D) a negative rtTA-only control (40,000 cells shown) sorted on day 16 post-induction. Live, TauEGFP-positive cells were enriched by first gating DRAQ5-positive, DAPI-negative cells, then collecting only those that were GFP-positive. (E) Percent of TauEGFP positive cells out of total number of cells collected post-FACS ( $n = 4$  sorts,  $>100$  cells/sort). (F-H) Correlation plots between aligned counts from sequenced libraries of an *Ngn3/Brn2* induced neuron population generated from 10 ng vs. 5 ng (F) and 10 ng vs. 1 ng (G) of input RNA. *Ngn3/Brn2* (10 ng) counts were also plotted against *Ascl1/Brn2* (10 ng) counts.



**Table 3.1 Metadata of candidate iN populations selected for RNA-Seq.** Table of experimental information related to cell sorting, RNA extraction, library preparation and sequencing of each candidate iN population selected for RNA-Seq analyses, with two biological replicates per population. dpi, days post-induction; RIN, RNA integrity number.

sample name	biological replicate	factor 1	factor 2	strain	cell source	final cell type	sorting date	purification method	input RNA (ng)	RIN (RNA)	PI/cell (RNA)	amplification method	average cDNA fragment length (bp)	sequencing platform	sequencing parameters	uniquely mapped reads
1 A1-B2-3	1	Asc11 (A1)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9.2	11.6	Clontech SMARTER Ultra Low-v3	409	Illumina NextSeq	Single-read 1x75	32808353
2 A1-B2-4	2	Asc11 (A1)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9.2	4.1	Clontech SMARTER Ultra Low-v3	360	Illumina NextSeq	Single-read 1x75	39920737
3 A1-B3a-4	1	Asc11 (A1)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9.4	4.2	Clontech SMARTER Ultra Low-v3	348	Illumina NextSeq	Single-read 1x75	42970282
4 A1-B3a-6	2	Asc11 (A1)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9	11.3	Clontech SMARTER Ultra Low-v3	345	Illumina NextSeq	Single-read 1x75	37848170
5 A1-B3b-5-2	1	Asc11 (A1)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.5	5.2	Clontech SMARTER Ultra Low-v3	352	Illumina NextSeq	Single-read 1x75	27765162
6 A1-B3b-5	2	Asc11 (A1)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.3	7.8	Clontech SMARTER Ultra Low-v3	369	Illumina NextSeq	Single-read 1x75	38244688
7 A1-B3c-2	1	Asc11 (A1)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.4	13.2	Clontech SMARTER Ultra Low-v3	377	Illumina NextSeq	Single-read 1x75	3324709
8 A1-B3c-3	2	Asc11 (A1)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9.1	11.0	Clontech SMARTER Ultra Low-v3	384	Illumina NextSeq	Single-read 1x75	33174349
9 A1-B4-2	1	Asc11 (A1)	Pou3F4 (Bm4, B4)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.5	5.8	Clontech SMARTER Ultra Low-v3	380	Illumina NextSeq	Single-read 1x75	28991805
10 A1-B4-4	2	Asc11 (A1)	Pou3F4 (Bm4, B4)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	7.7	5.0	Clontech SMARTER Ultra Low-v3	347	Illumina NextSeq	Single-read 1x75	41883701
11 A1-Nurr1-6-2	1	Asc11 (A1)	Nr4a2 (Nurr1)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.6	8.0	Clontech SMARTER Ultra Low-v3	356	Illumina NextSeq	Single-read 1x75	40819525
12 A1-Nurr1-6	2	Asc11 (A1)	Nr4a2 (Nurr1)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	7.9	8.8	Clontech SMARTER Ultra Low-v3	348	Illumina NextSeq	Single-read 1x75	37969587
13 A1-O4-2	1	Asc11 (A1)	Pou5F1 (Oct4)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.3	5.4	Clontech SMARTER Ultra Low-v3	379	Illumina NextSeq	Single-read 1x75	34031541
14 A1-O4-4	2	Asc11 (A1)	Pou5F1 (Oct4)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.9	3.6	Clontech SMARTER Ultra Low-v3	356	Illumina NextSeq	Single-read 1x75	37721632
15 A1-P1-2	1	Asc11 (A1)	Pou1F1 (Pt1)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.9	6.5	Clontech SMARTER Ultra Low-v3	376	Illumina NextSeq	Single-read 1x75	32191663
16 A1-P1-4	2	Asc11 (A1)	Pou1F1 (Pt1)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.4	4.8	Clontech SMARTER Ultra Low-v3	351	Illumina NextSeq	Single-read 1x75	36877002
17 A2-B2-5-2	1	Asc12 (A2)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9	11.9	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	38311839
18 A2-B2-5	2	Asc12 (A2)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9	13.6	Clontech SMARTER Ultra Low-v3	381	Illumina NextSeq	Single-read 1x75	46490508
19 A2-B3a-4-2	1	Asc12 (A2)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.8	5.4	Clontech SMARTER Ultra Low-v3	351	Illumina NextSeq	Single-read 1x75	40177732
20 A2-B3a-4	2	Asc12 (A2)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.3	5.4	Clontech SMARTER Ultra Low-v3	345	Illumina NextSeq	Single-read 1x75	42928906
21 A2-B3b-2	1	Asc12 (A2)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.4	9.3	Clontech SMARTER Ultra Low-v3	368	Illumina NextSeq	Single-read 1x75	31511333
22 A2-B3b-3	2	Asc12 (A2)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9.3	5.1	Clontech SMARTER Ultra Low-v3	392	Illumina NextSeq	Single-read 1x75	35841419
23 A2-B3c-5	1	Asc12 (A2)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.5	7.1	Clontech SMARTER Ultra Low-v3	377	Illumina NextSeq	Single-read 1x75	36434070
24 A2-B3c-6	2	Asc12 (A2)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.9	12.3	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	36298467
25 A2-Nurr1-5	1	Asc12 (A2)	Nr4a2 (Nurr1)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	7.7	6.5	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	46208833
26 A2-Nurr1-6	2	Asc12 (A2)	Nr4a2 (Nurr1)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	5	7.9	1.2	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	34997649
27 A2-P1-6-2	1	Asc12 (A2)	Pou1F1 (Pt1)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.1	7.8	Clontech SMARTER Ultra Low-v3	356	Illumina NextSeq	Single-read 1x75	28759704
28 A2-P1-6	2	Asc12 (A2)	Pou1F1 (Pt1)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8	6.7	Clontech SMARTER Ultra Low-v3	367	Illumina NextSeq	Single-read 1x75	40037897
29 A5-B3c-5	1	Asc15 (A5)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.3	15.2	Clontech SMARTER Ultra Low-v3	354	Illumina NextSeq	Single-read 1x75	36617187
30 A5-B3c-6	2	Asc15 (A5)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	8.7	4.7	Clontech SMARTER Ultra Low-v3	353	Illumina NextSeq	Single-read 1x75	34431959
31 Atoh1-B3c-6-2	1	Atoh1	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9	14.8	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	34363835
32 Atoh1-B3c-6	2	Atoh1	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive INs	16 dpi	FACS	10	9.7	10.2	Clontech SMARTER Ultra Low-v3	347	Illumina NextSeq	Single-read 1x75	36055651

Table 3.1 Metadata of candidate iN populations selected for RNA-Seq, Continued.

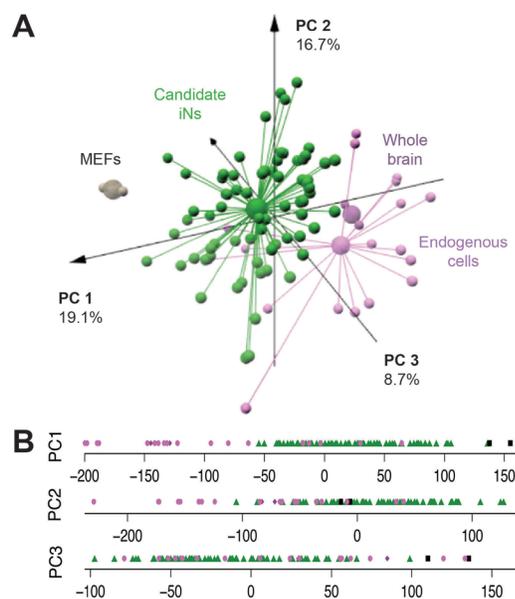
sample name	condition	biological replicate	factor1	factor2	strain	cell source	final cell type	sorting date	purification method	input RNA (ng)	RIN (RNA)	pg/cell (RNA)	amplification_method	average cDNA fragment length (bp)	sequencing platform	sequencing parameters	uniquely mapped reads
33 N1-B1-5	N1-B1	1	Neurog1 (N1)	Pou3F3 (Bm1)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.6	3.6	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	38154782
34 N1-B1-6	N1-B1	2	Neurog1 (N1)	Pou3F3 (Bm1)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	5	8.7	2.7	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	27168695
35 N1-B2-5-2	N1-B2	1	Neurog1 (N1)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	7.7	7.4	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	37766152
36 N1-B2-5	N1-B2	2	Neurog1 (N1)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.7	4.4	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	42932053
37 N1-B3a-2	N1-B3a	1	Neurog1 (N1)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.3	12.1	Clontech SMARTER Ultra Low-v3	381	Illumina NextSeq	Single-read 1x75	28327651
38 N1-B3a-3	N1-B3a	2	Neurog1 (N1)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.1	N/A	Clontech SMARTER Ultra Low-v3	401	Illumina NextSeq	Single-read 1x75	32186269
39 N1-B3b-5-2	N1-B3b	1	Neurog1 (N1)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9	14.6	Clontech SMARTER Ultra Low-v3	387	Illumina NextSeq	Single-read 1x75	39913841
40 N1-B3b-5	N1-B3b	2	Neurog1 (N1)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.6	15.3	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	40747713
41 N1-B3c-6-2	N1-B3c	1	Neurog1 (N1)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.4	10.8	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	45602575
42 N1-B3c-6	N1-B3c	2	Neurog1 (N1)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.2	12.8	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	35973747
43 N1-P1-5-2	N1-P1	1	Neurog1 (N1)	Pou1F1 (P1)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	6.9	9.7	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	31754253
44 N1-P1-5	N1-P1	2	Neurog1 (N1)	Pou1F1 (P1)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.6	7.4	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	41808539
45 N2-B2-5-2	N2-B2	1	Neurog2 (N2)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	6	12.1	Clontech SMARTER Ultra Low-v3	360	Illumina NextSeq	Single-read 1x75	33519276
46 N2-B2-5	N2-B2	2	Neurog2 (N2)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.1	8.2	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	33403294
47 N2-B3a-2	N2-B3a	1	Neurog2 (N2)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9	10.9	Clontech SMARTER Ultra Low-v3	385	Illumina NextSeq	Single-read 1x75	32721372
48 N2-B3a-3	N2-B3a	2	Neurog2 (N2)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	N/A	N/A	Clontech SMARTER Ultra Low-v3	397	Illumina NextSeq	Single-read 1x75	36905579
49 N2-B3b-5	N2-B3b	1	Neurog2 (N2)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.7	12.6	Clontech SMARTER Ultra Low-v3	376	Illumina NextSeq	Single-read 1x75	35786689
50 N2-B3b-6	N2-B3b	2	Neurog2 (N2)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	5	8.7	2.0	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	36557772
51 N2-B3c-6-2	N2-B3c	1	Neurog2 (N2)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.2	15.0	Clontech SMARTER Ultra Low-v3	360	Illumina NextSeq	Single-read 1x75	31238999
52 N2-B3c-6	N2-B3c	2	Neurog2 (N2)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.6	10.1	Clontech SMARTER Ultra Low-v3	349	Illumina NextSeq	Single-read 1x75	34734490
53 N2-P1-5-2	N2-P1	1	Neurog2 (N2)	Pou1F1 (P1)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.5	6.8	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	47338698
54 N2-P1-5	N2-P1	2	Neurog2 (N2)	Pou1F1 (P1)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.4	11.6	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	33475970
55 N3-B2-3	N3-B2	1	Neurog3 (N3)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.5	15.3	Clontech SMARTER Ultra Low-v3	406	Illumina NextSeq	Single-read 1x75	37984077
56 N3-B2-4	N3-B2	2	Neurog3 (N3)	Pou3F2 (Bm2, B2)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.6	9.2	Clontech SMARTER Ultra Low-v3	348	Illumina NextSeq	Single-read 1x75	38150092
57 N3-B3a-2	N3-B3a	1	Neurog3 (N3)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.4	4.9	Clontech SMARTER Ultra Low-v3	385	Illumina NextSeq	Single-read 1x75	24504005
58 N3-B3a-6	N3-B3a	2	Neurog3 (N3)	Pou4F1 (Bm3a, B3a)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9	15.4	Clontech SMARTER Ultra Low-v3	353	Illumina NextSeq	Single-read 1x75	38192085
59 N3-B3b-5-2	N3-B3b	1	Neurog3 (N3)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9	9.7	Clontech SMARTER Ultra Low-v3	N/A	Illumina NextSeq	Single-read 1x75	47138267
60 N3-B3b-5	N3-B3b	2	Neurog3 (N3)	Pou4F2 (Bm3b, B3b)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.8	14.3	Clontech SMARTER Ultra Low-v3	384	Illumina NextSeq	Single-read 1x75	37222476
61 N3-B3c-2	N3-B3c	1	Neurog3 (N3)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.6	8.7	Clontech SMARTER Ultra Low-v3	393	Illumina NextSeq	Single-read 1x75	31430699
62 N3-B3c-3	N3-B3c	2	Neurog3 (N3)	Pou4F3 (Bm3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.8	6.0	Clontech SMARTER Ultra Low-v3	402	Illumina NextSeq	Single-read 1x75	36112521
63 N3-B4-2	N3-B4	1	Neurog3 (N3)	Pou3F4 (Bm4, B4)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9	3.4	Clontech SMARTER Ultra Low-v3	397	Illumina NextSeq	Single-read 1x75	31363295

Table 3.1 Metadata of candidate iN populations selected for RNA-Seq, Continued.

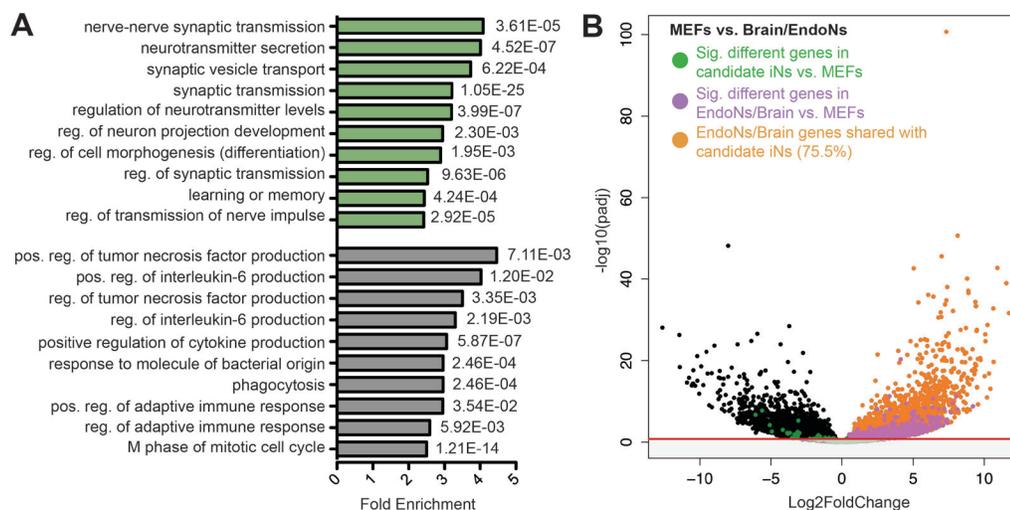
sample name	condition	biological replicate	factor 1	factor 2	strain	cell source	final cell type	sorting date	purification method	input RNA (ng)	RIN (RNA)	pg/cell (RNA)	amplification_method	average cDNA fragment length (bp)	sequencing platform	sequencing parameters	uniquely mapped reads
64 N3-B4-4	N3-B4	2	Neurog3 (N3)	Pou3f4 (Brn4, B4)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.8	7.1	Contech SMARTer Ultra Low -v3	379	Illumina NextSeq	Single-read 1x75	40777692
65 N3-O4-4	N3-O4	1	Neurog3 (N3)	Pou5f1 (Oct4)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	10	10.2	Contech SMARTer Ultra Low -v3	350	Illumina NextSeq	Single-read 1x75	36690754
66 N3-O4-6	N3-O4	2	Neurog3 (N3)	Pou5f1 (Oct4)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	2	9.3	2.2	Contech SMARTer Ultra Low -v3	N/A	Illumina NextSeq	Single-read 1x75	32379813
67 N3-P1-1	N3-P1	1	Neurog3 (N3)	Pou1f1 (P1)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.1	7.4	Contech SMARTer Ultra Low -v3	337	Illumina NextSeq	Single-read 1x75	62957048
68 N3-P1-6	N3-P1	2	Neurog3 (N3)	Pou1f1 (P1)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.3	5.4	Contech SMARTer Ultra Low -v3	347	Illumina NextSeq	Single-read 1x75	40918180
69 ND2-B3c-5	ND2-B3c	1	NeuroD2 (ND2)	Pou4f3 (Brn3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	9.3	11.0	Contech SMARTer Ultra Low -v3	N/A	Illumina NextSeq	Single-read 1x75	49672986
70 ND2-B3c-6	ND2-B3c	2	NeuroD2 (ND2)	Pou4f3 (Brn3c, B3c)	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	2	9	8.2	Contech SMARTer Ultra Low -v3	N/A	Illumina NextSeq	Single-read 1x75	35966970
71 MEFs-sorted-1	MEFs-sorted	1	N/A	N/A	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.7	6.3	Contech SMARTer Ultra Low -v3	348	Illumina NextSeq	Single-read 1x75	59484892
72 MEFs-sorted-2	MEFs-sorted	2	N/A	N/A	TauEGFP	MEF (E13.5)	TauEGFP-positive iNs	16 dpi	FACS	10	8.8	4.7	Contech SMARTer Ultra Low -v3	379	Illumina NextSeq	Single-read 1x75	49976705
										average	8.7	8.5		369			37559314
										std. dev	0.6	4		19			6623060

**Table 3.2 Metadata of endogenous neuron populations selected for RNA-Seq.** Table of experimental information related to cell sorting, RNA extraction, library preparation and sequencing of each endogenous neuron population selected for RNA-Seq analyses, with 2-3 biological replicates per population. RIN, RNA integrity number.

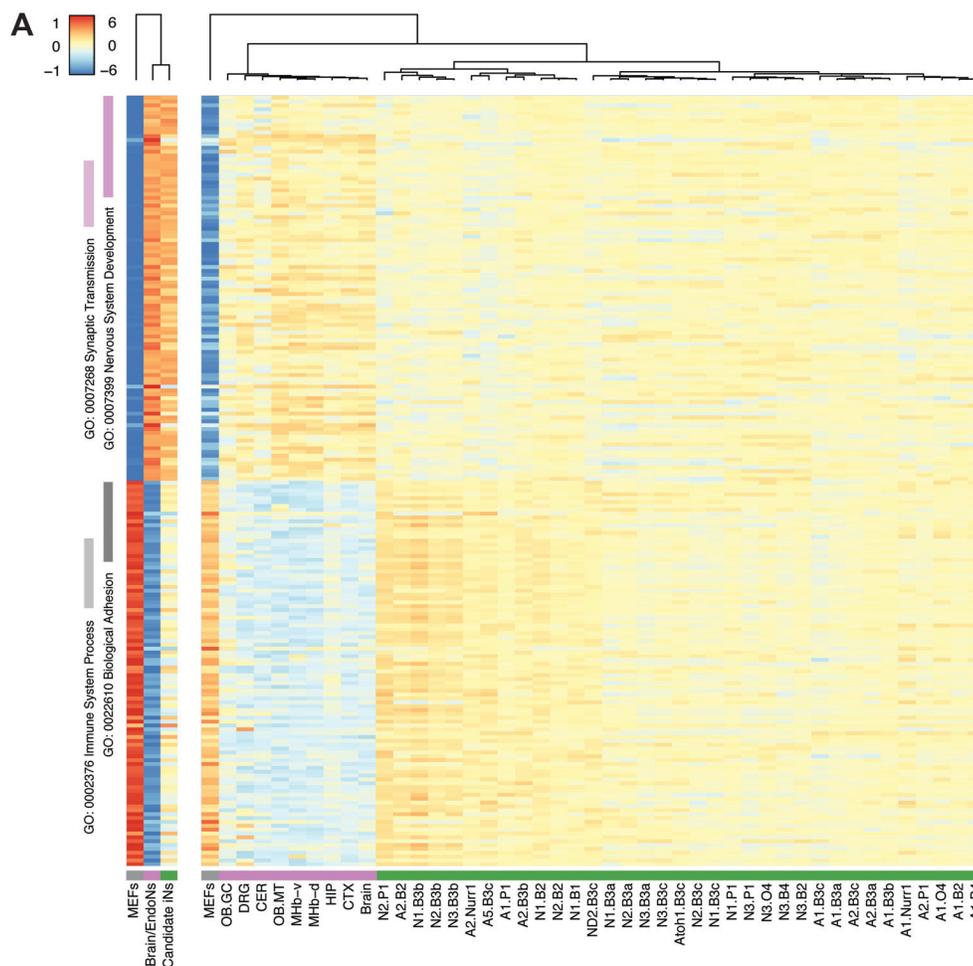
sample name	sample population	biological replicate	species	strain	area	cell_type	age	purification method	input RNA (ng)	RIN (RNA)	pg/cell (RNA)	amplification method	avg. cDNA frag. length (bp)	sequencing platform	sequencing parameters	no. uniquely mapped reads
1 DRG-1	DRG	1	mouse	TauEGFP	dorsal root ganglion	TauEGFP-positive dorsal root ganglion cells	P21	FACS	10	7.6	N/A	Clontech SMARTER Ultra Low -v3	393	Illumina NextSeq	Single-read 1x75	32781655
2 DRG-2	DRG	2	mouse	TauEGFP	dorsal root ganglion	TauEGFP-positive dorsal root ganglion cells	P21	FACS	10	7.4	N/A	Clontech SMARTER Ultra Low -v3	N/A	Illumina NextSeq	Single-read 1x75	33582852
3 MHb-d-1	MHb-d	1	mouse	Tac2-Cre/A19	habenula	High intensity Tac2-Cre-positive habenula cells	P21	FACS	7	7.8	3.7	Clontech SMARTER Ultra Low -v3	N/A	Illumina NextSeq	Single-read 1x75	42756418
4 MHb-d-2	MHb-d	2	mouse	Tac2-Cre/A19	habenula	High intensity Tac2-Cre-positive habenula cells	P21	FACS	7	8.4	3.0	Clontech SMARTER Ultra Low -v3	N/A	Illumina NextSeq	Single-read 1x75	38534871
5 MHb-v-1	MHb-v	1	mouse	Tac2-Cre/A19	habenula	Low intensity Tac2-Cre-positive habenula cells	P21	FACS	7	8	3.9	Clontech SMARTER Ultra Low -v3	N/A	Illumina NextSeq	Single-read 1x75	36764668
6 MHb-v-2	MHb-v	2	mouse	Tac2-Cre/A19	habenula	Low intensity Tac2-Cre-positive habenula cells	P21	FACS	7	9.2	3.3	Clontech SMARTER Ultra Low -v3	N/A	Illumina NextSeq	Single-read 1x75	32913868
7 OB-MT-1	OB-MT	1	mouse	Pcdh21-Cre/A19	olfactory bulb	Pcdh21-positive Mitral and Tufted cells	P21	FACS	10	8.8	3.1	Clontech SMARTER Ultra Low -HV	371	Illumina HiSeq	Single-read 1x100	22441660
8 OB-MT-2	OB-MT	2	mouse	Pcdh21-Cre/A19	olfactory bulb	Pcdh21-positive Mitral and Tufted cells	P21	FACS	10	N/A	3.7	Clontech SMARTER Ultra Low -HV	320	Illumina HiSeq	Single-read 1x100	21378473
9 OB-MT-3	OB-MT	3	mouse	Pcdh21-Cre/A19	olfactory bulb	Pcdh21-positive Mitral and Tufted cells	P21	FACS	10	N/A	3.1	Clontech SMARTER Ultra Low -HV	N/A	Illumina HiSeq	Single-read 1x100	21954875
10 OB-GC-1	OB-GC	1	mouse	Pcdh21-Cre/A19	olfactory bulb	Pcdh21-negative olfactory bulb cells	P21	FACS	10	8.9	N/A	Clontech SMARTER Ultra Low -HV	N/A	Illumina HiSeq	Single-read 1x100	253883570
11 OB-GC-2	OB-GC	2	mouse	Pcdh21-Cre/A19	olfactory bulb	Pcdh21-negative olfactory bulb cells	P21	FACS	10	8.9	N/A	Clontech SMARTER Ultra Low -HV	N/A	Illumina HiSeq	Single-read 1x100	30199541
12 OB-GC-3	OB-GC	3	mouse	Pcdh21-Cre/A19	olfactory bulb	Pcdh21-negative olfactory bulb cells	P21	FACS	10	9	N/A	Clontech SMARTER Ultra Low -HV	N/A	Illumina HiSeq	Single-read 1x100	21125674
15 HIP-1	HIP	1	mouse	Tac2-Cre/A19	hippocampus	Tac2-Cre-positive hippocampal cells	P21	FACS	5	7.5	1.7	Clontech SMARTER Ultra Low -v3	331	Illumina NextSeq	Single-read 1x75	45927781
16 HIP-2	HIP	2	mouse	Tac2-Cre/A19	hippocampus	Tac2-Cre-positive hippocampal cells	P21	FACS	5	7.3	1.8	Clontech SMARTER Ultra Low -v3	346	Illumina NextSeq	Single-read 1x75	38212008
17 CTX-1	CTX	1	mouse	TauEGFP	cortex	High intensity TauEGFP-positive cortical cells	P21	FACS	10	6.9	3.4	Clontech SMARTER Ultra Low -v3	341	Illumina NextSeq	Single-read 1x75	38655942
18 CTX-2	CTX	2	mouse	TauEGFP	cortex	High intensity TauEGFP-positive cortical cells	P21	FACS	10	7.5	1.9	Clontech SMARTER Ultra Low -v3	356	Illumina NextSeq	Single-read 1x75	40682671
21 CER-1	CER	1	mouse	TauEGFP	cerebellum	High intensity TauEGFP-positive cerebellar cells	P21	FACS	10	7.8	3.8	Clontech SMARTER Ultra Low -v3	342	Illumina NextSeq	Single-read 1x75	47873719
22 CER-2	CER	2	mouse	TauEGFP	cerebellum	High intensity TauEGFP-positive cerebellar cells	P21	FACS	10	7.7	3.5	Clontech SMARTER Ultra Low -v3	340	Illumina NextSeq	Single-read 1x75	46036439
13 Brain-1	Brain	1	mouse	WT BALB/c	whole brain	neural cells	P66.5	From Clontech SMARTER Ultra Low -HV	10	N/A	N/A	Clontech SMARTER Ultra Low -HV	358	Illumina HiSeq	Single-read 1x100	25519983
14 Brain-2	Brain	1	mouse	WT BALB/c	whole brain	neural cells	P66.5	From Clontech SMARTER Ultra Low -HV	10	N/A	N/A	Clontech SMARTER Ultra Low -v3	346	Illumina NextSeq	Single-read 1x75	45578133
									average	8.0	3.1		349.5			34415240
									std. dev	0.7	0.7		18.9			8849744



**Figure 3.10 Principal component analysis reveals similarities and differences between individual iN and endogenous neuron populations.** (A) PCA performed using DESeq2 vsd-normalized RNA-Seq data from TauEGFP positive candidate iN populations (green), MEFs (grey), endogenous neurons (purple), and whole brain samples (dark purple). Loadings for principal components 1 (PC1, 19.1%), 2 (PC2, 16.7%) and 3 (PC3, 8.7%) are reported on x-, y- and z-axes, respectively. Larger spheres represent centroid of connecting samples. (B) Candidate iNs populations (green triangles), MEFs (grey squares), endogenous neurons (purple circles), and whole brain samples (dark purple diamonds) plotted along a horizontal axis (loadings) for the first three principal components.



**Figure 3.11 Gene expression profiles of candidate iNs overlap with profiles of endogenous neurons.** (A) Top 10 biological process (BP level 5) gene ontology (GO) terms overrepresented in the TauEGFP positive candidate iN populations (green) and MEFs (grey) as determined by DAVID. Input gene lists included significantly upregulated genes ( $p$ -adjusted  $< 0.05$ ) identified by DESeq2 when comparing the 35 duplicate candidate iN populations (3,860 genes) to MEFs (3,508 genes). Data are presented as fold enrichment, with corresponding FDR q-values. (B) Volcano plot of  $\log_2$  fold change versus  $-\log_{10}$   $p$ -adjusted value outputted from DESeq2 when comparing MEFs to endogenous neuronal populations and whole brain (MEFs vs Endo/Brain). Upregulated genes for the MEFs are plotted as negative  $\log_2$  fold change and vice versa for the upregulated genes of the endogenous neuronal and whole brain populations. Significantly upregulated genes in the candidate iN populations (3,860 genes) when compared to MEFs were plotted in green onto the MEFs vs Endo/Brain volcano plot. Shared upregulated genes between candidate iN and Endo/Brain populations were plotted in orange and make up 75.5% (2,239/2,965) of the significant Endo/Brain upregulated genes. The remaining 24.5% (726/2,965) of upregulated genes in the Endo/Brain populations are plotted in purple. Red line represents  $-\log_{10}$  of 0.05  $p$ -adjusted value.



**Figure 3.12 Candidate iNs express both neuronal and fibroblasts genes.** (A) Heat map shows expression patterns of the top 200 significantly different genes (rows, p-adjusted < 0.05) equally split between MEFs and endogenous neuronal populations and whole brain, as determined by DESeq2. Expression levels are defined as DESeq2 vsd-normalized RNA-Seq counts. The mean gene expression for each class of cells (MEFs, Endo/Brain, and candidate iN populations) is shown on the left. Dendrogram represents hierarchical clustering based on correlation distance. GO terms associated with a subset of genes are highlighted along the y-axis axis. OB-GC, olfactory bulb- granule cells; DRG, dorsal root ganglion; CER, cerebellum; HIP, hippocampus; MHb-d, medial habenula- dorsal; MHb-v, medial habenula- ventral; Brain, whole brain; CTX, cortex; OB-MT, olfactory bulb- mitral and tufted cells; A1, *Ascl1*; A2, *Ascl2*; N1, *Ng1*; N2, *Ng2*; N3, *Ng3*; ND2, *NeuroD2*; B2, *Brn2*; B4, *Brn4*; B3a, *Brn3a*; B3b, *Brn3b*; B3c, *Brn3c*; O4, *Oct4*; P1, *Pit1*.

### 3.4 Discussion

Through screening 598 distinct pairings of neuronal-related bHLH and POU factors, we identified 72 novel transcription factor combinations capable of reprogramming fibroblasts into cells that expressed neuronal markers *Tuj1*, *Map2* and *Tau* and exhibited characteristic neuronal morphology. Further transcriptomic analysis of 35 of these candidate iN populations revealed their marked similarity to endogenous neuronal populations.

The high percentage of novel reprogramming pairs discovered in the screen highlights the nature of bHLH proneural genes to function as “master regulators” in direct neuronal reprogramming. Of the 16 bHLH factors capable of generating candidate iNs when paired with a POU factor, 14 are members of the proneural *Ascl*, *Neurogenin*, and *Neurogenic differentiation* gene families, although not all individual factors are widely considered proneural (Huang et al., 2014). Interestingly, bHLH factors *Myf5* and *Ptf1a*, which are more commonly known to be involved in skeletal myogenesis (Rudnicki et al., 1993) and pancreatic development (Masui et al., 2007) respectively, generated *Tuj1* expressing cells with neuronal morphology when paired with *Brn3* factors. *Ptf1a* has been previously observed to be essential for proper specification of neurons in the cerebellum, spinal cord and retina (Meredith et al., 2009; Yamada et al., 2007), while the function of *Myf5* expression in neurons remains unknown (Francetic and Li, 2011). Therefore, conducting a “semi-unbiased” screen in which we included as many neuron-expressing factors regardless of proneural status, revealed unexpected combinations worth future examination in reprogramming and development. In fact, we may be underreporting the number of successful reprogramming pairs since viral titer variability may account for false negatives. Additional reprogramming pairs may be identified with increased viral titer or the addition of small molecules or factors previously reported to increase efficiency of conversion from fibroblast to neuron (Gascon et al., 2015; Ladewig et al., 2012).

The high number of identified pairs also demonstrates that there may be multiple transcriptional pathways that converge on a common neuronal “identity” in induced neurons. Whole transcriptome analysis revealed thousands of upregulated genes shared between the candidate iN and endogenous neuronal populations when both populations were compared to fibroblasts. These shared genes were

enriched for biological processes related to synaptic transmission, neurogenesis, and neuron projection development (data not shown) and could represent a “core” set of neuronal genes that induced neurons express regardless of the reprogramming factors used to generate them. To elucidate the transcriptional programs that result in “core” neuronal gene expression in the candidate iN populations, we could conduct chromatin immunoprecipitation followed by high-throughput sequencing (ChIP-Seq) to identify the occupied genomic loci of the reprogramming factors. Using this technique, *Ascl1* has been shown to be an “on-target pioneering factor” in that *Ascl1* alone binds to its cognate lineage-specific targets in fibroblasts regardless of permissibility (Wapinski et al., 2013). Wapinski et al. (2013) also demonstrated that a “trivalent” chromatin state could predict the accessibility of *Ascl1* and effectiveness of *Ascl1*-driven iN reprogramming in different cell sources. Therefore, it would be interesting to determine if the novel reprogramming factor combinations we identified have similar “on-target” capabilities and preferentially bind to the same targets of *Ascl1*. By transcription profiling at multiple time points during iN reprogramming, similar to iPSC reprogramming studies (Tanaka et al., 2015), we could also gain insights into how and when the transcriptional programs generated from different reprogramming pairs differ or converge.

It is important to note that despite the abundance of neuron-related genes expressed in the iN populations compared to MEFs, we do see residual intermediate expression of fibroblast genes in the iN population RNA-Seq datasets. As mentioned previously, this may be due to either the presence of fibroblasts and/or intermediate, transitioning cells in the RNA-Seq populations or the persistence of fibroblast gene expression in the candidate iNs. To determine which of these possibilities it may be, we could conduct single cell RNA-Seq or quantitative RT-PCR. If fibroblast- and neuron-related genes are expressed in the same TauEGFP positive, candidate iN cell, then the question is whether additional factors or other improvements in the reprogramming protocols are needed to further reprogram these candidate iNs so they more closely resemble endogenous neurons. In the original paper describing iN reprogramming from fibroblasts, they reported that although *Ascl1* with *Brn2* can generate iN cells capable of firing action potentials with complex neuronal morphology, the addition of *Myt1l* produced iNs that exhibited repetitive action potentials and more complex morphology (Vierbuchen et al., 2010). This published result

emphasizes the need to conduct electrophysiological assays to determine if the candidate iNs are truly functional iNs. Simultaneously, we can determine if the individual iN populations differ in functional properties and address whether these differences are reflected in their transcriptional profiles. Transcriptional and functional diversity of the iN populations will be addressed in Chapter 4.

Chapter 3, in part, is currently being prepared for submission for publication of the material. Tsunemoto, R.K., Lee, S., Szűcs, A., Chubukov, P., Blanchard, J.W., Eade, K.T., Torkamani, A., Sanna, P.P., and Baldwin, K.K. The dissertation author was the primary investigator and author of this paper.

### References:

Blanchard, J.W., Eade, K.T., Szucs, A., Lo Sardo, V., Tsunemoto, R.K., Williams, D., Sanna, P.P., and Baldwin, K.K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nature neuroscience* *18*, 25-35.

Castro, D.S., Skowronska-Krawczyk, D., Armant, O., Donaldson, I.J., Parras, C., Hunt, C., Critchley, J.A., Nguyen, L., Gossler, A., Gottgens, B., Matter, J.M., and Guillemot, F. (2006). Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. *Developmental cell* *11*, 831-844.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. *Genome biology* *4*, P3.

Fode, C., Ma, Q., Casarosa, S., Ang, S.L., Anderson, D.J., and Guillemot, F. (2000). A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. *Genes & Development* *14*, 67-80.

Francetic, T., and Li, Q. (2011). Skeletal myogenesis and Myf5 activation. *Transcription* *2*, 109-114.

Gascon, S., Murenu, E., Masserdotti, G., Ortega, F., Russo, G.L., Petrik, D., Deshpande, A., Heinrich, C., Karow, M., Robertson, S.P., Schroeder, T., Beckers, J., Irmeler, M., Berndt, C., Angeli, J.P., Conrad, M., Berninger, B., and Gotz, M. (2015). Identification and Successful Negotiation of a Metabolic Checkpoint in Direct Neuronal Reprogramming. *Cell Stem Cell*.

He, X., Treacy, M.N., Simmons, D.M., Ingraham, H.A., Swanson, L.W., and Rosenfeld, M.G. (1989). Expression of a large family of POU-domain regulatory genes in mammalian brain development. *Nature* *340*, 35-42.

Huang, C., Chan, J.A., and Schuurmans, C. (2014). Proneural bHLH genes in development and disease. *Current topics in developmental biology* *110*, 75-127.

Josephson, R., Muller, T., Pickel, J., Okabe, S., Reynolds, K., Turner, P.A., Zimmer, A., and McKay, R.D.G. (1998). POU transcription factors control expression of CNS stem cell-specific genes. *Development* *125*, 3087-3100.

Ladewig, J., Mertens, J., Kesavan, J., Doerr, J., Poppe, D., Glaue, F., Herms, S., Wernet, P., Kogler, G., Muller, F.J., Koch, P., and Brustle, O. (2012). Small molecules enable highly efficient neuronal conversion of human fibroblasts. *Nat Methods* *9*, 575-578.

Lee, S.K., and Pfaff, S.L. (2003). Synchronization of neurogenesis and motor neuron specification by direct coupling of bHLH and homeodomain transcription factors. *Neuron* *38*, 731-745.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* *15*, 550.

Ma, Q., Sommer, L., Cserjesi, P., and Anderson, D.J. (1997). Mash1 and neurogenin1 expression patterns define complementary domains of neuroepithelium in the developing CNS and are correlated with regions expressing notch ligands. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *17*, 3644-3652.

Masui, T., Long, Q., Beres, T.M., Magnuson, M.A., and MacDonald, R.J. (2007). Early pancreatic development requires the vertebrate Suppressor of Hairless (RBPJ) in the PTF1 bHLH complex. *Genes & Development* *21*, 2629-2643.

Meredith, D.M., Masui, T., Swift, G.H., MacDonald, R.J., and Johnson, J.E. (2009). Multiple transcriptional mechanisms control Ptf1a levels during neural development including autoregulation by the PTF1-J complex. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *29*, 11139-11148.

Osorio, J., Mueller, T., Retaux, S., Vernier, P., and Wullimann, M.F. (2010). Phylotypic expression of the bHLH genes Neurogenin2, Neurod, and Mash1 in the mouse embryonic forebrain. *The Journal of comparative neurology* *518*, 851-871.

Philippidou, P., and Dasen, J.S. (2013). Hox genes: choreographers in neural development, architects of circuit organization. *Neuron* *80*, 12-34.

Rudnicki, M.A., Schnegelsberg, P.N., Stead, R.H., Braun, T., Arnold, H.H., and Jaenisch, R. (1993). MyoD or Myf-5 is required for the formation of skeletal muscle. *Cell* *75*, 1351-1359.

Soufi, A., Garcia, M.F., Jaroszewicz, A., Osman, N., Pellegrini, M., and Zaret, K.S. (2015). Pioneer transcription factors target partial DNA motifs on nucleosomes to initiate reprogramming. *Cell* *161*, 555-568.

Tanaka, Y., Hysolli, E., Su, J., Xiang, Y., Kim, K.Y., Zhong, M., Li, Y., Heydari, K., Euskirchen, G., Snyder, M.P., Pan, X., Weissman, S.M., and Park, I.H. (2015). Transcriptome Signature and Regulation in Human Somatic Cell Reprogramming. *Stem cell reports* *4*, 1125-1139.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* *463*, 1035-1041.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., Drechsel, D., Martynoga, B., Castro, D.S., Webb, A.E., Sudhof, T.C., Brunet, A., Guillemot, F., Chang, H.Y., and Wernig, M. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* *155*, 621-635.

Yamada, M., Terao, M., Terashima, T., Fujiyama, T., Kawaguchi, Y., Nabeshima, Y., and Hoshino, M. (2007). Origin of climbing fiber neurons and their developmental dependence on Ptf1a. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *27*, 10924-10934.

## **Chapter 4:**

### **Functional and transcriptional diversity of induced neuron populations**

#### **4.1 Introduction**

In Chapter 3, we identified 76 bHLH and POU transcription factor pairs capable of generating Tuj1-expressing cells with neuronal morphology directly from fibroblasts, which we classified as candidate induced neurons (iNs). 72 were novel reprogramming combinations not previously published. Whole transcriptome analysis of 35 of the candidate iN populations revealed gene expression profiles similar to endogenous neurons. The question remained whether transcriptional similarities between the candidate iN and endogenous neuron populations would be reflected in similar functional capabilities. Previously published reprogramming papers established neuronal functionality in their iNs by demonstrating action potential firing elicited by membrane depolarization (Blanchard et al., 2015; Kim et al., 2011; Marro et al., 2011; Pang et al., 2011; Son et al., 2011; Vierbuchen et al., 2010). Since the candidate iNs share many features (e.g. morphology and gene expression profiles) with previously published iNs and endogenous neurons, we predicted that the candidate iNs would also be functional neurons capable of firing action potentials.

Additionally, in Chapter 3, we hypothesized that different bHLH and POU pairings would result in markedly diverse iN populations. Such diversity in the candidate iN populations may be reflected in electrophysiological differences. In the brain, many different neuronal subtypes exhibit unique electrophysiological properties, such as the fast-spiking properties seen in inhibitory basket cells of the cerebellum, hippocampus and cerebral cortex (Tripathy et al., 2015). Chapter 4 will investigate the functional and transcriptomic diversity among 35 different iN populations.

#### **4.2 Methods**

##### **4.2.1 Functional analysis using whole-cell patch clamp technique**

To determine if the candidate iNs had functional membrane properties similar to endogenous neurons, we used whole-cell patch clamp recordings of TauEGFP positive and synapsin expressing cells with neuronal morphology on days 16 through 24 after induction. To identify TauEGFP positive cells that

expressed synapsin, we transduced the candidate iNs with lentivirus encoding the fluorescent red protein, TdTomato, under the control of a *SYN1* promoter. Constant, rectangular, current steps of 350 milliseconds (ms) duration were delivered intracellularly in 2 or 5 picoamp (pA) increments in successive cycles of stimulation at a rate of 1 Hertz (Hz). Firing responses and multiple physiological properties were assessed in five populations of candidate iNs (*Ngn3/Pit1*, *Ngn3/Oct4*, *Ascl2/Brn3c*, *NeuroD2/Brn3c*, and *Atoh1/Brn3c*) to determine functional diversity in the iN populations. Detailed recording methods are provided in Appendix A.3.

#### **4.2.2 Weighted gene coexpression network analysis**

To assess transcriptomic diversity between the candidate iN populations, we conducted weighted gene coexpression network analysis (WGCNA) (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). WGCNA has been previously shown to reveal biologically meaningful patterns in transcriptome data related to brain anatomy and function (Hawrylycz et al., 2015; Hawrylycz et al., 2012). Therefore, to conduct WGCNA on the iN and MEF datasets, we used a similar approach to Hawrylycz et al. (2012, 2015) using the user-friendly WGCNA R package (Langfelder and Horvath, 2008). Detailed methods regarding WGCNA are provided in Appendix A.3.

### **4.3 Results**

#### **4.3.1 iN populations exhibit functional and diverse electrophysiological properties**

Of the 76 bHLH and POU transcription factors, three (*Ngn1/Brn3a*, *Ngn2/Brn3a*, and *Ascl1/Brn2*) have already been shown to produce functional iNs directly from fibroblasts using the whole-cell patch clamp technique (Blanchard et al., 2015; Vierbuchen and Wernig, 2011). Therefore, for functional analysis, we selected five representative combinations of the remaining 73 combinations that utilized transcription factors not previously used in published reprogramming experiments; this included *Ngn3/Pit1*, *Ngn3/Oct4*, *Ascl2/Brn3c*, *NeuroD2/Brn3c*, and *Atoh1/Brn3c*.

To characterize intrinsic excitability and multiple physiological properties of the five candidate iN populations, we used the whole-cell patch clamp technique with injection of rectangular current pulses. For

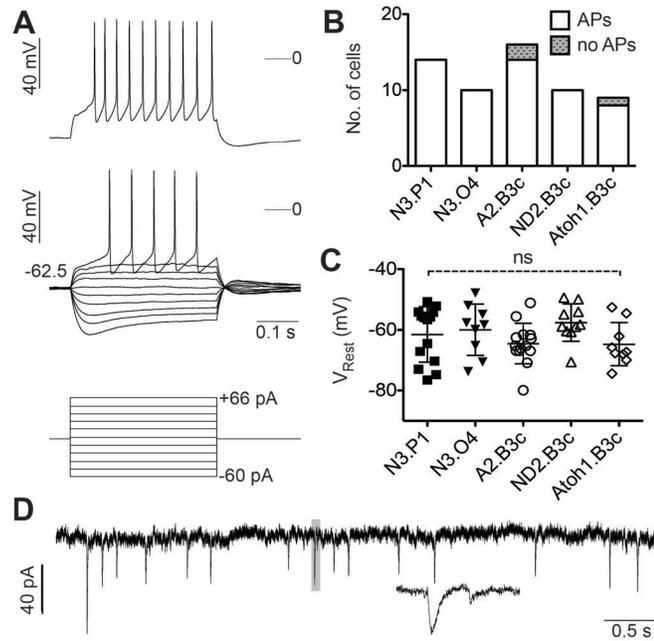
each population, we selected cells that exhibited neuronal morphology and expressed both TauEGFP and the synapsin reporter at 16 to 24 days post-induction. Under the patch clamp protocol, the great majority of cells recorded exhibited behaviors characteristic to mature endogenous neurons. Of the 59 total TauEGFP and synapsin positive cells recorded, 56 (95%) displayed both resting membrane potentials at levels seen in endogenous neurons ( $-61.8 \pm 7.8$  mV) and healthy action potentials upon depolarization (Figure 4.1A-C). This was independent of the transcription factor combination used during reprogramming and provided evidence that at least these five candidate iN populations can be classified as bona-fide functional iNs. We also observed excitatory post-synaptic currents (EPSCs) in five individual recorded candidate iNs, which is indicative of functional synapse formation in the cultures of fibroblasts and candidate iNs (Figure 4.1D). Cells that expressed neither TauEGFP nor TdTomato and exhibited fibroblast-like morphology did not display these electrophysiological properties (data not shown).

Current step experiments also revealed physiological properties that were not uniform across the iN populations. We found combination-specific differences in the membrane input resistance (Figure 4.2A-C) and voltage sag slope parameters (Figure 4.2E-F). iNs generated using *Ngn3/Pit1*, on average, had significantly greater membrane resistance ( $1001.9 \pm 501.4$  M $\Omega$ , n = 14) than iNs generated from *Atoh1/Brn3c* ( $459.8 \pm 257.4$  M $\Omega$ , n = 9) and *Ascl2/Brn3c* ( $399.4 \pm 212.6$  M $\Omega$ , n = 14) (Figure 4.2C). The membrane input resistance of each cell was calculated using either the voltage sampled at the most negative part of the response or just before the end of the 350 ms current step (steady-state resistance). Resistance values were then plotted against input current and linear extrapolation was used to obtain the input resistance at rest ( $I = 0$ ) (Figure 4.2B). Although there were significant differences in membrane input resistance, there was no significant differences in the amount of current needed to elicit an action potential, or rheobase (Figure 4.2D).

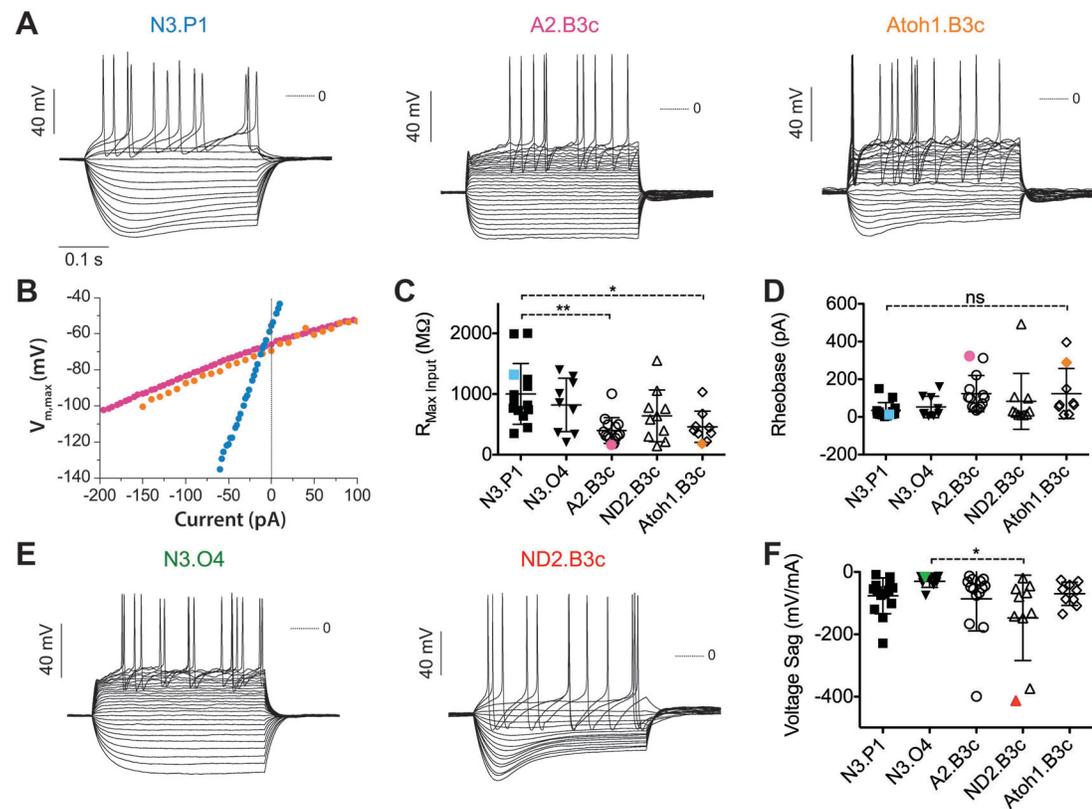
On the other hand, the slope of current-voltage sag relationship, on average, was significantly stronger in iNs generated with *NeuroD2/Brn3c* ( $-147.2 \pm 136.9$  mV/nA, n = 10) compared to iNs generated from *Ngn3/Oct4* ( $-29.9 \pm 19.4$  mV/nA, n = 9) (Figure 4.2E-F). Similar to calculating membrane resistance, voltage sag values were plotted against the amplitude of the injected current and curves were fitted with

linear functions to extract slope parameters. The slope is indicative of the amount of hyperpolarization-activated cation currents ( $I_h$ ) in the cell.

Taken together, cells from all five of the iN populations selected for functional analysis exhibited electrophysiological properties of neurons. This suggests that the other candidate iN populations, which share similar gene expression profiles, are also functional neurons. Additionally, we observed differences in multiple electrophysiological properties across the five iN populations, which provides evidence that different reprogramming pairs give rise to diverse iNs.



**Figure 4.1 iN populations exhibit electrophysiological properties of endogenous neurons.** (A) Representative membrane voltage responses from a TauEGFP-, synapsin-positive cell with neuronal morphology generated with *Ascl2/Brn3c* under whole-cell patch clamp conditions at max current injection (top) and current steps until the first induction of action potentials (middle). Current step traces shown below voltage traces. (B) Quantification of number of cells exhibiting current-induced action potentials per condition. (C) Quantification of resting membrane potentials for cells that exhibited current-induced action potentials per condition. Data are presented mean  $\pm$  SD. (D) Representative current trace showing EPSCs from one TauEGFP-, synapsin-positive cell generated with *Ngn3/Oct4*.



**Figure 4.2 Different reprogramming pairs generate iNs with varying membrane input resistances and hyperpolarization-activated cation currents.** (A) Representative membrane voltage responses to depolarizing current steps of TauEGFP-, synapsin-positive cells with neuronal morphology generated with *Ngn3/Pit1* (left), *Ascl2/Brn3c* (middle), and *Atoh1/Brn3c* (right). (B) Current-voltage plot of cells shown in (A): *Ngn3/Pit1* (blue), *Ascl2/Brn3c* (pink), and *Atoh1/Brn3c* (orange). Linear extrapolation was used to obtain the input resistance at rest ( $I = 0$ ). (C) Quantification of membrane input resistance for recorded cells per condition. Data points obtained from cells shown in (A) labeled with appropriate color. (D) Quantification of rheobase for recorded cells per condition. Data points obtained from cells shown in (A) labeled with appropriate color. (E) Representative membrane voltage responses to depolarizing current steps of TauEGFP-, synapsin-positive cells with neuronal morphology generated with *Ngn3/Oct4* (left) and *NeuroD2/Brn3c* (right). Voltage sag depicted in trace of cell generated with *NeuroD2/Brn3c* (right). (F) Quantification of voltage sag for recorded cells per condition. Data points obtained from cells shown in (E) labeled with appropriate color. Data are presented mean  $\pm$  SD. \*\*\*, \*\* and \* represents a p-value  $< 0.001$ ,  $0.01$ , and  $0.05$ , respectively (Bonferroni's Multiple Comparison Test). A2, *Ascl2*; N3, *Ngn3*; ND2, *NeuroD2*; B3c, *Brn3c*; O4, *Oct4*; P1, *Pit1*.

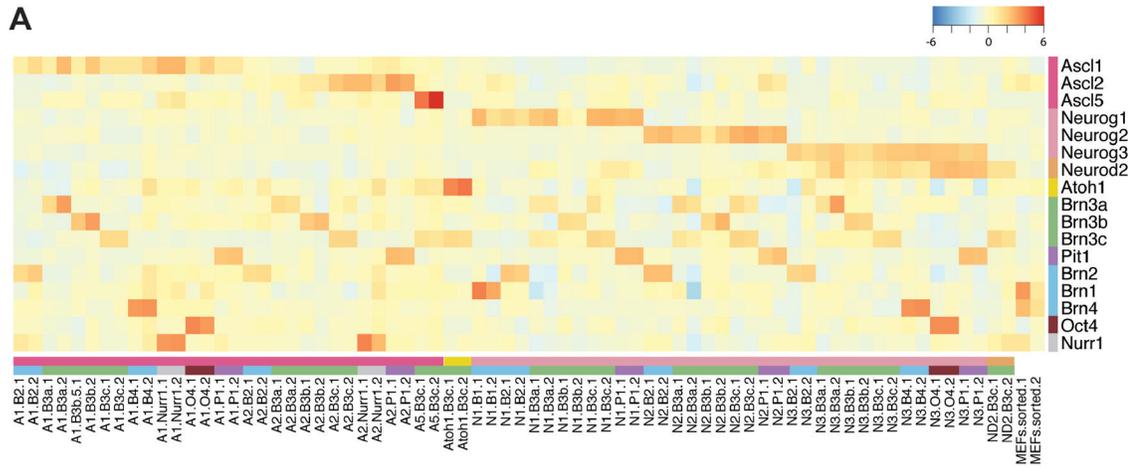
### 4.3.2 Identification of genes selectively expressed in individual iN populations

In addition to a very high percentage of the iNs proved to be functional neurons, there were subtle differences in electrophysiological properties observed between the representative iN populations. This suggests that beyond the “core” neuronal genes shared in the majority of iN populations, as described in Chapter 3, there are specific genes differentially expressed by discrete groups of iNs. Therefore, we aimed to identify genes with significant differences in expression between iNs generated with different reprogramming factor pairs and cluster these genes based on meaningful expression patterns.

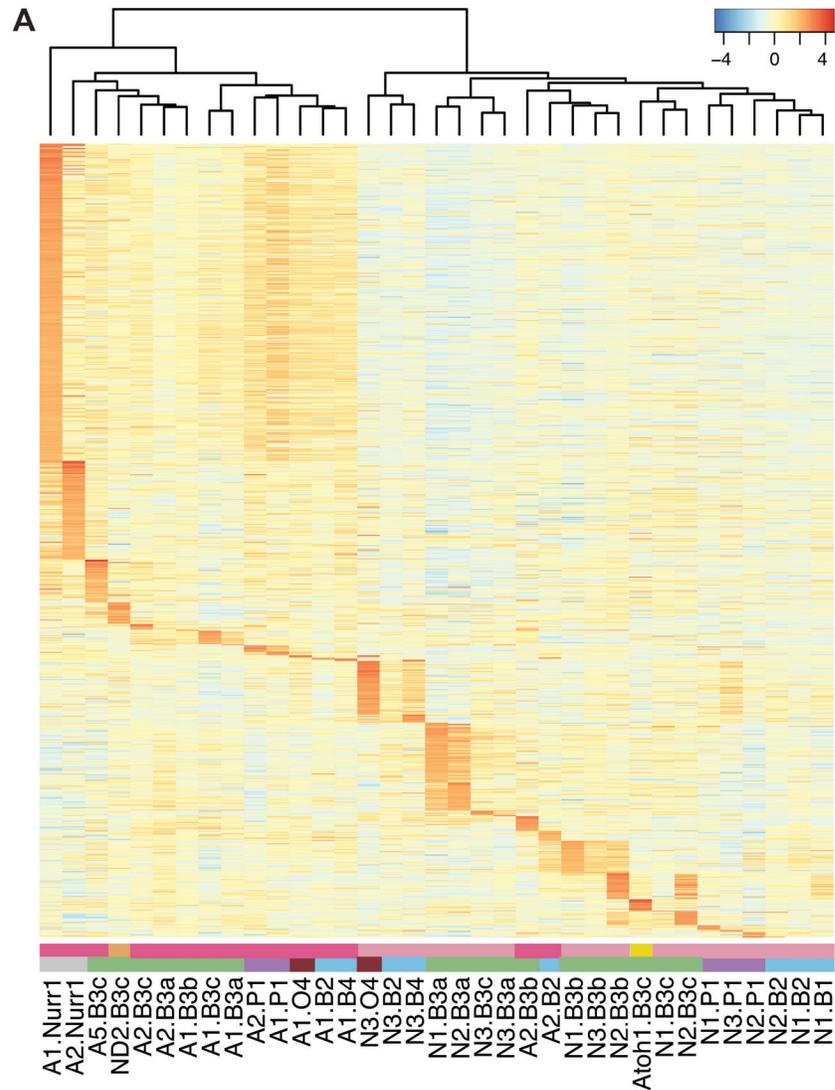
First, we confirmed that the appropriate reprogramming pairs were expressed in each of the sorted iN populations (Figure 4.3A). Although, we turned off expression of the reprogramming transgenes with the removal of doxycycline at day 8 of reprogramming, expression of the reprogramming factors remained detectable. This could be explained by either residual expression of the transgenes or the activation of the endogenous genes. Interestingly, we also observed that expression of specific reprogramming pairs upregulated the expression of other reprogramming factors. For example, in iN populations containing *Ngn3*, the expression of *NeuroD2* is upregulated compared to the expression in other iN populations (Figure 4.3A). Additionally, when *Ngn1* and *Ngn2* were paired with *Brn3a*, *Brn3c* was upregulated, and vice versa. This pattern was not observed when *Brn3a* or *Brn3c* was paired with *Ascl1* or *Ascl2* (Figure 4.3A). This provides evidence that synergy between different bHLH and POU pairings yields different patterns of gene expression in the iNs.

To discern gene expression patterns unique to each reprogramming pair, we used DESeq2 (Love et al., 2014) to identify genes that were significantly upregulated (p-adjusted value < 0.05) in each iN population versus all other iN populations. Of those upregulated genes, we filtered out genes whose expression levels were not significantly different from MEFs. In total, we identified 706 genes that met this criteria (Figure 4.4A). 67 of those genes, which included reprogramming factors *Ascl2*, *Ngn1*, *Ngn2*, *Brn3a*, *Oct4*, *Brn4*, and *Pit1*, were significantly upregulated in more than one population. The iN populations with the largest number of selectively expressed genes were generated with *Ascl1/Nurr1* (n = 282) and *Ascl2/Nurr1* (n = 101), which is not surprising since they are the only iN populations generated using the non-POU factor, *Nurr1*. iN populations generated with combinations that included both a bHLH

and POU factor had, on average,  $12 \pm 15$  selectively expressed genes. Given the low number of genes selectively expressed in an individual iN population, we hypothesized that many genes are selectively upregulated in groups of iN populations. Therefore, we next aimed to identify the genes differentially expressed between different groups of iN populations.



**Figure 4.3 Appropriate reprogramming pairs are expressed in the sequenced iN populations.** (A) Heat map shows expression of reprogramming factors in all replicate iN and MEF populations. Expression levels are defined as DESeq2 vsd-normalized RNA-Seq counts and scaled by row (gene). A1, *Ascl1*; A2, *Ascl2*; A5, *Ascl5*; N1, *Ngn1*; N2, *Ngn2*; N3, *Ngn3*; ND2, *NeuroD2*; B2, *Brn2*; B4, *Brn4*; B3a, *Brn3a*; B3b, *Brn3b*; B3c, *Brn3c*; O4, *Oct4*; P1, *Pit1*.



**Figure 4.4 Select genes expressed in discrete iN populations.** (A) Heat map shows expression of genes significantly upregulated (p-adjusted value < 0.05) in each iN population versus all other iN populations and MEFs as determined by DESeq2. Expression levels are defined as DESeq2 vsd-normalized RNA-Seq counts. Dendrogram represents hierarchical clustering based on Euclidean distance. A1, *Ascl1*; A2, *Ascl2*; A5, *Ascl5*; N1, *Ngn1*; N2, *Ngn2*; N3, *Ngn3*; ND2, *NeuroD2*; B2, *Brn2*; B4, *Brn4*; B3a, *Brn3a*; B3b, *Brn3b*; B3c, *Brn3c*; O4, *Oct4*; P1, *Pit1*.

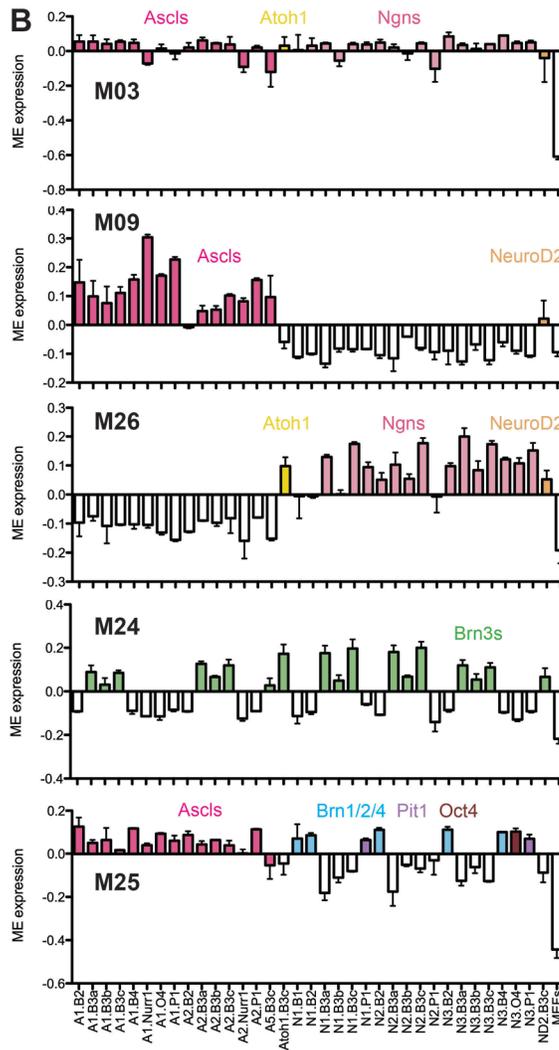
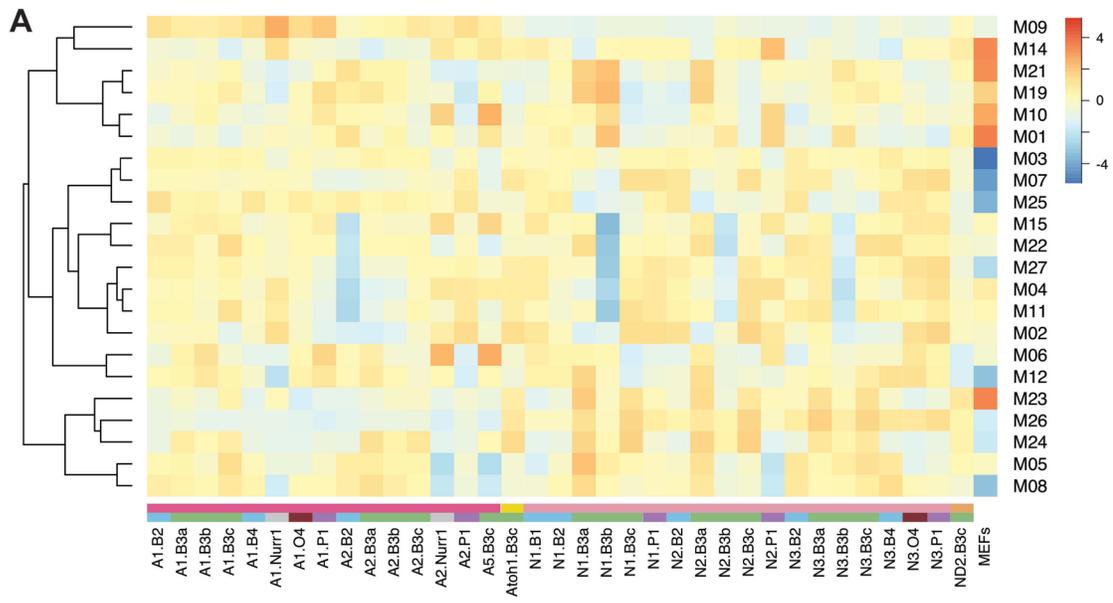
### 4.3.3 Transcriptional patterning in iN populations

Identifying clusters of genes coexpressed across different neuronal populations can elucidate cell type- and anatomic-specific transcriptional patterning related to brain organization and function (Hawrylycz et al., 2015; Hawrylycz et al., 2012). As an example of how to identify biologically meaningful gene coexpression patterns that describe the differences between the iN populations, we utilized WGCNA (Langfelder and Horvath, 2008; Zhang and Horvath, 2005). First, in efforts to reduce the noise from low expressing genes in the dataset, we only included genes in which the non-normalized counts were greater than 200 in at least one iN or MEF population, in both replicates ( $n = 12,549$ ). The resulting network analysis (see Appendix A.3 for exact parameters) identified 27 gene coexpression patterns, known as modules, which we collapsed into 22 main modules. Each main module exhibited a distinct expression pattern across the iN and MEF samples, as described by the module eigengene (ME), the first principal component of the module (Figure 4.5A).

As expected, many modules were enriched in the iN populations versus MEFs (Figure 4.5B-C). Module 3 (M03), for example, was enriched in most iN populations and included genes associated with neurogenesis, dendritic spine morphology and synaptic transmission, as determined by PANTHER Overrepresentation Test (Mi et al., 2013; Mi et al., 2016). The representative hub genes, selected for their high correlation with the respective ME, were also related to neuronal function (Figure 4.5C). Additionally, some of these modules were selectively enriched in distinct groups of iN populations based on the identities of their reprogramming factors. M09 was enriched in iN populations generated with either a member of the *Ascl* family or *NeuroD2*. Interestingly, although M09 contained genes associated with neuronal function, many were also related to muscle development and function. In contrast, M26 was enriched in iN populations generated with either a member of the *Neurogenin* family, *Atoh1* or *NeuroD2*, but not members of the *Ascl* family. Some modules were also selective for iN populations generated with particular POU transcription factor subclasses. M24 was enriched in iN populations generated with Class IV (*Brn3*) POU transcription factors, while M25 was enriched in generally all iN populations except those generated with Class IV (*Brn3*) POU factors when specifically paired with *Atoh1*, *NeuroD2*, or a member of the *Neurogenin* family.

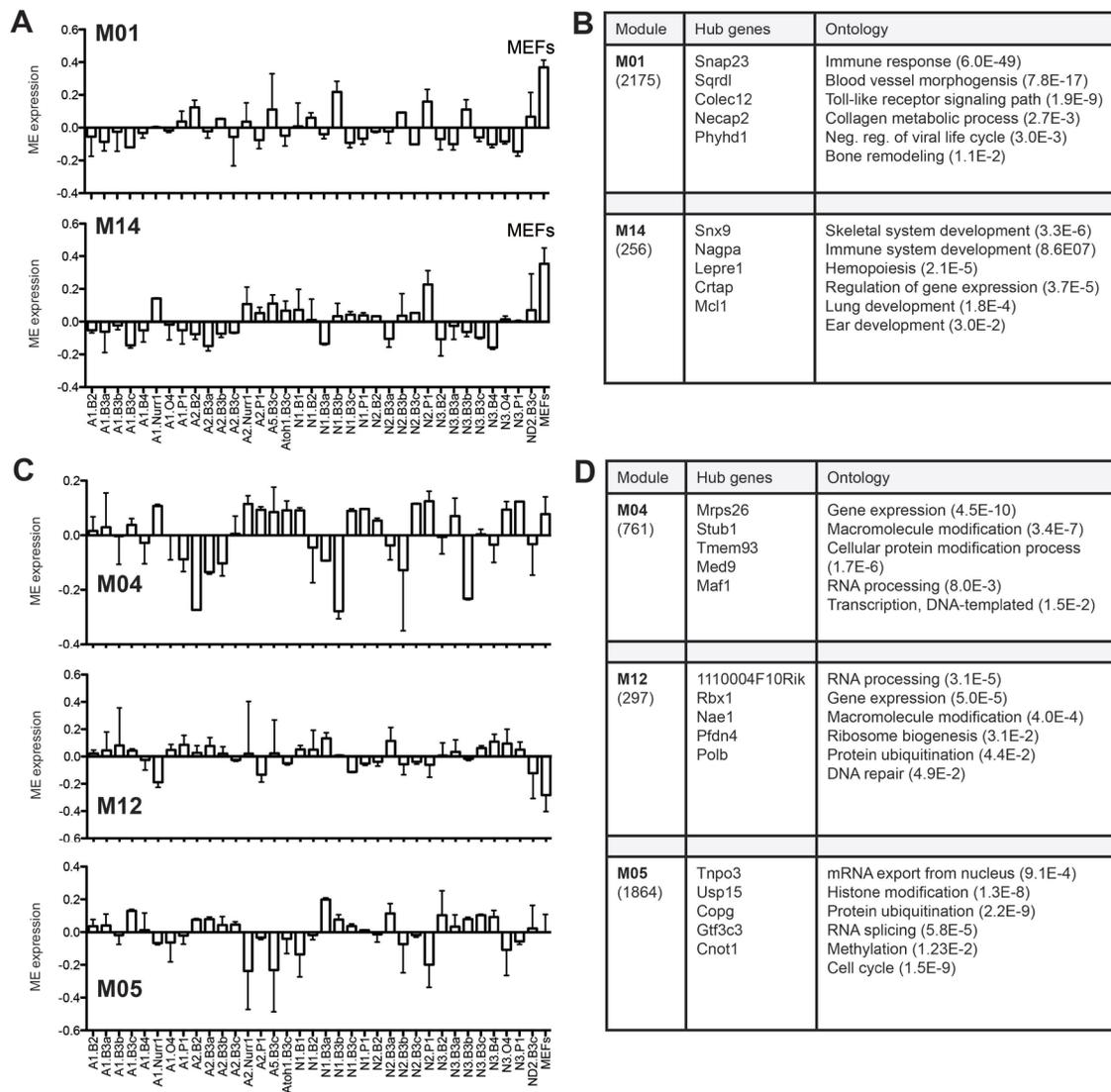
On the other hand, other modules exhibited more complex expression patterning (Figure 4.6). For example, modules M01 and M14 were enriched in MEFs and various iN populations with genes associated with immune response and the development of the skeletal system, blood, and lungs (Figure 4.6A-B). M03, M12, and M05, which had varied expression across the iN populations, contained genes associated with gene expression, methylation and histone modification (Figure 4.6C-D).

**Figure 4.5 Transcription factor specificity of WGCNA module expression.** (A) Hierarchical clustering of merged modules M1-M27 based on correlation distance with iN populations fixed based on bHLH transcription factor identity. Modules were determined using individual replicates of all iN populations and MEFs, but for clarity, replicates were collapsed in the heat map. (B) Bar plots of average module eigengene expression of five representative modules depicting transcription factor specificity. Data are presented as mean  $\pm$  SD. Colors highlight iNs populations generated with shared transcription factors. (C) Table of number of module genes, representative hub genes and associated GO Terms determined using PANTHER Overrepresentation Test with adjusted p-values (Bonferroni correction).



**C**

Module	Hub genes	Ontology
<b>M03</b> (1029)	Prkcz Syt11 Mcf2l Sgip1 Nap112	Neurogenesis (4.2E-32) Synaptic transmission (2.9E-25) Neuronal action potential (7.4E-4) Pos. reg. of EPSPs (1.43E-4) Reg. of dendritic spine morpho. (2.7E-3) Startle response (2.4E-2)
<b>M09</b> (477)	Ldb3 Ttn Actn2 Tnnt3 Tnni1	Muscle structure development (1.1E-44) Muscle system process (3.4E-41) Muscle contraction (2.8E-35) Striated muscle tissue development (8.5E-33) Modulation of synaptic transmission (4.5E-3) Neuron development (1.2E-2)
<b>M26</b> (104)	Rgnef 2610109H07R-ik Slco4a1 Irs2 Mad2l2	Nervous system development (6.3E-3) Regulation of signaling (5.4E-3) Intracellular signal transduction (1.1E-2) Regulation of neurogenesis (1.5E-2) Brain development (2.1E-2) Generation of neurons (2.7E-2)
<b>M24</b> (137)	Cacnb4 Tanc1 Snca Syt7 Trpm8	Behavior (7.3E-3)
<b>M25</b> (126)	Trank1 Slc7a14 Mdga2 Iqsec3 Grin2b	Adult behavior (8.4E-5) Learning or memory (2.1E-3) Neuron-neuron synaptic transmission (1.7E-2) Regulation of synapse assembly (3.6E-2) Associative learning (3.9E-2) Action potential (4.2E-2)



**Figure 4.6 Complex gene coexpression patterning across iN populations.** (A) Bar plots of average module eigengene expression of two representative modules enriched in MEF and various iN populations. (B) Table of number of module genes, representative hub genes and associated GO Terms for modules shown in (A). (C) Bar plots of average module eigengene expression of three representative modules with complex expression patterns across iN populations. (D) Table of number of module genes, representative hub genes and associated GO Terms for modules shown in (C). Data in (A) and (C) are presented mean  $\pm$  SD. Associated GO Terms in (B) and (D) determined using PANTHER Overrepresentation Test with adjusted p-values (Bonferroni correction).

#### 4.4 Discussion

As predicted, the cells selected for functional analysis (*Ngn3/Pit1*-, *Ngn3/Oct4*-, *Ascl2/Brn3c*-, *NeuroD2/Brn3c*-, and *Atoh1/Brn3c*-generated iNs) exhibited resting membrane potentials similar to endogenous neurons and fired action potentials upon depolarization. These experiments suggest that screening for cells that coexpress Tuj1, Map2 and TauEGFP upon induction of transient expression of pairs of transcription factors is a powerful means to identify cells with electrophysiological properties that strongly resemble those of endogenous neurons. Although we did not record cells from the remaining 68 candidate iN populations discovered, we expect that most, if not all, of these induced neuronal populations will contain functional neurons. In particular, based on RNA-Seq analyses, we would predict that the 35 profiled populations would contain functional iNs because of their similar transcriptional profiles with the five combinations we profiled electrophysiologically.

Of the cells we did record, the high percentage that fired action potentials as early as day 16 post-induction is evidence for the robustness of direct reprogramming. Additionally, Pang et al. (2011) found that direct reprogramming with transcription factors yielded human iNs that all generated action potentials at day 6 and day 34-35 when made from human ESCs and fetal fibroblasts, respectively (Pang et al., 2011). Collectively, it is evident that direct reprogramming provides a means to rapidly generate functional neurons from both mouse and human.

The observed difference in electrophysiological properties is also evidence for diversity in the iN populations. The voltage sag was significantly different in the *NeuroD2/Brn3c*- versus *Ngn3/Oct4*- iN population. This indicated the presence of hyperpolarization-activated mixed cation current ( $I_h$ , h-current) in the iNs generated with *NeuroD2/Brn3c*. In the mouse brain, the presence of prominent h-current helps distinguish different neuronal subtypes, even within the same brain region (Cooper and Stanford, 2000; Sheets et al., 2011). For example, within the motor cortex, corticospinal neurons have high  $I_h$  expression, while corticostriatal and corticocortical neurons have low  $I_h$  expression. This difference is reflected in higher mRNA expression of the hyperpolarization-activated and cyclic nucleotide-gated (HCN) channel, *Hcn1*, in corticospinal versus corticostriatal neurons (Sheets et al., 2011). In the central nervous system, *Hcn1* has high expression in neurons of the cortex, hippocampus, and superior colliculus. Other isoforms

include *Hcn2*, which is expressed at high levels throughout the brain, *Hcn3* which is expressed at low levels throughout the brain, and *Hcn4*, which has restricted patterns of expression (Biel et al., 2009; Franz et al., 2000; Kase and Imoto, 2012; Moosmang et al., 1999).

Interestingly, in the five recorded populations, RNA-Seq results revealed iNs generated with *Ngn3/Oct4*, which is the population we observed low/absent voltage sag, had the lowest expression of *Hcn1* (data not shown). If *Hcn1* expression correlated with the presence of Ih in the iN populations, we would expect the *NeuroD2/Brn3c* population, in which we observed significantly high levels of voltage sag, to have the highest *Hcn1* expression. However, according to RNA-Seq, *Hcn1* expression was highest in the *Atoh1/Brn3c* population (data not shown). Intriguingly, the *Atoh1/Brn3c* population also had the highest expression of a modulator of HCN channel trafficking and channel gating, *Pex5l/Trip8b* (data not shown) (Lewis et al., 2009; Santoro et al., 2009; Zolles et al., 2009), which could explain the lower levels of voltage sag detected. Therefore, in the iN populations, the functional diversity we observed was reflected in the varying expression of relevant genes and supports the use of the RNA-Seq data to help predict the functionality of the different populations. However, further functional testing should be conducted to confirm these predictions if the generated iNs are to be used for these properties.

To further explore the transcriptional diversity in the iN populations on a larger scale, we conducted DESeq2 and WGCNA analyses to extract meaningful patterns of differential gene expression. DESeq2 revealed a limited number of uniquely expressed genes in single iN populations; therefore, we used WGCNA to identify patterns of gene expression shared across multiple iN populations. Here we discovered modules of genes that were enriched in all iN populations versus MEFs and modules enriched in iN populations based on the identities of their reprogramming factors. Interestingly, *Hcn1* was one of the genes in module 24 (M24), a module selectively enriched in iN populations generated with members of the Class IV (*Brn3*) family of POU factors. Functionally, of the populations we recorded, we observed h-currents in iNs generated with *Brn3c*. Similarly, *TrpM8*, a gene encoding a cold-temperature and menthol receptor, was also in M24. Blanchard et al. (2015) demonstrated using calcium imaging that iNs generated with *Brn3a* and *Ngn1* or *Ngn2* were responsive to menthol. Taken together, these results suggest that iNs generated with members of the Class IV (*Brn3*) family of POU factors would exhibit h-currents and

menthol responses. This again demonstrates that the expression of different bHLH and POU transcription factors gives rise to transcriptional and functional diversity in the resulting iN populations. Additionally, this highlights the use of WGCNA as a discovery tool in discerning meaningful patterns of gene expression across multiple iN populations which can help predict that functional responses of those iNs.

Some modules identified by WGCNA revealed more complex patterns of gene expression that varied across the iN populations and were associated with fibroblast function, gene expression, methylation, and histone modification. These modules may reflect differences in the stages of transitioning from fibroblast to neuron across the iN populations. Although this could be influenced by the specific reprogramming factors, alternatively, these differences could reflect the epigenetic differences in the multiple sources of MEFs. Further analysis will need to be conducted to determine the driver of these complex gene coexpression patterns. In fact, altering the set of input genes or adjusting the parameters of WGCNA might reveal additional modules driving disparate features of neuronal identity.

Chapter 4, in part, is currently being prepared for submission for publication of the material. Tsunemoto, R.K., Lee, S., Szűcs, A., Chubukov, P., Blanchard, J.W., Eade, K.T., Torkamani, A., Sanna, P.P., and Baldwin, K.K. The dissertation author was the primary investigator and author of this paper.

#### References:

- Biel, M., Wahl-Schott, C., Michalakis, S., and Zong, X. (2009). Hyperpolarization-activated cation channels: from genes to function. *Physiol Rev* 89, 847-885.
- Blanchard, J.W., Eade, K.T., Szucs, A., Lo Sardo, V., Tsunemoto, R.K., Williams, D., Sanna, P.P., and Baldwin, K.K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nature neuroscience* 18, 25-35.
- Cooper, A.J., and Stanford, I.M. (2000). Electrophysiological and morphological characteristics of three subtypes of rat globus pallidus neurone in vitro. *The Journal of physiology* 527, 291-304.
- Franz, O., Liss, B., Neu, A., and Roeper, J. (2000). Single-cell mRNA expression of HCN1 correlates with a fast gating phenotype of hyperpolarization-activated cyclic nucleotide-gated ion channels (Ih) in central neurons. *The European journal of neuroscience* 12, 2685-2693.
- Hawrylycz, M., Miller, J.A., Menon, V., Feng, D., Dolbeare, T., Guillozet-Bongaarts, A.L., Jegg, A.G., Aronow, B.J., Lee, C.K., Bernard, A., Glasser, M.F., Dierker, D.L., Menche, J., Szafer, A., Collman, F., Grange, P., Berman, K.A., Mihalas, S., Yao, Z., Stewart, L., Barabasi, A.L., Schulkin, J., Phillips, J., Ng, L., Dang, C., Haynor, D.R., Jones, A., Van Essen, D.C., Koch, C., and Lein, E. (2015). Canonical genetic signatures of the adult human brain. *Nature neuroscience* 18, 1832-1844.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., Abajian, C., Beckmann, C.F., Bernard, A., Bertagnolli, D., Boe, A.F., Cartagena, P.M., Chakravarty, M.M., Chapin, M., Chong, J., Dalley, R.A., Daly, B.D., Dang, C., Datta, S., Dee, N., Dolbeare, T.A., Faber, V., Feng, D., Fowler, D.R., Goldy, J., Gregor, B.W., Haradon, Z., Haynor, D.R., Hohmann, J.G., Horvath, S., Howard, R.E., Jeromin, A., Jochim, J.M., Kinnunen, M., Lau, C., Lazarz, E.T., Lee, C., Lemon, T.A., Li, L., Li, Y., Morris, J.A., Overly, C.C., Parker, P.D., Parry, S.E., Reding, M., Royall, J.J., Schulkin, J., Sequeira, P.A., Slaughterbeck, C.R., Smith, S.C., Sodt, A.J., Sunkin, S.M., Swanson, B.E., Vawter, M.P., Williams, D., Wohnoutka, P., Zielke, H.R., Geschwind, D.H., Hof, P.R., Smith, S.M., Koch, C., Grant, S.G., and Jones, A.R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* *489*, 391-399.

Kase, D., and Imoto, K. (2012). The Role of HCN Channels on Membrane Excitability in the Nervous System. *J Signal Transduct* *2012*, 619747.

Kim, J., Su, S.C., Wang, H., Cheng, A.W., Cassady, J.P., Lodato, M.A., Lengner, C.J., Chung, C.Y., Dawlaty, M.M., Tsai, L.H., and Jaenisch, R. (2011). Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. *Cell Stem Cell* *9*, 413-419.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* *9*, 559.

Lewis, A.S., Schwartz, E., Chan, C.S., Noam, Y., Shin, M., Wadman, W.J., Surmeier, D.J., Baram, T.Z., Macdonald, R.L., and Chetkovich, D.M. (2009). Alternatively spliced isoforms of TRIP8b differentially control h channel trafficking and function. *The Journal of neuroscience : the official journal of the Society for Neuroscience* *29*, 6250-6265.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* *15*, 550.

Marro, S., Pang, Z.P., Yang, N., Tsai, M.C., Qu, K., Chang, H.Y., Sudhof, T.C., and Wernig, M. (2011). Direct lineage conversion of terminally differentiated hepatocytes to functional neurons. *Cell Stem Cell* *9*, 374-382.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature protocols* *8*, 1551-1566.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res* *44*, D336-342.

Moosmang, S., Biel, M., Hofmann, F., and Ludwig, A. (1999). Differential distribution of four hyperpolarization-activated cation channels in mouse brain. *Biological chemistry* *380*, 975-980.

Pang, Z.P., Yang, N., Vierbuchen, T., Ostermeier, A., Fuentes, D.R., Yang, T.Q., Citri, A., Sebastiano, V., Marro, S., Sudhof, T.C., and Wernig, M. (2011). Induction of human neuronal cells by defined transcription factors. *Nature* *476*, 220-223.

Santoro, B., Piskorowski, R.A., Pian, P., Hu, L., Liu, H., and Siegelbaum, S.A. (2009). TRIP8b splice variants form a family of auxiliary subunits that regulate gating and trafficking of HCN channels in the brain. *Neuron* *62*, 802-813.

Sheets, P.L., Suter, B.A., Kiritani, T., Chan, C.S., Surmeier, D.J., and Shepherd, G.M. (2011). Corticospinal-specific HCN expression in mouse motor cortex: I(h)-dependent synaptic integration as a candidate microcircuit mechanism involved in motor control. *Journal of neurophysiology* *106*, 2216-2231.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* 9, 205-218.

Tripathy, S.J., Burton, S.D., Geramita, M., Gerkin, R.C., and Urban, N.N. (2015). Brain-wide analysis of electrophysiological diversity yields novel categorization of mammalian neuron types. *Journal of neurophysiology* 113, 3474-3489.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035-1041.

Vierbuchen, T., and Wernig, M. (2011). Direct lineage conversions: unnatural but useful? *Nat Biotechnol* 29, 892-907.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4, Article17.

Zolles, G., Wenzel, D., Bildl, W., Schulte, U., Hofmann, A., Muller, C.S., Thumfart, J.O., Vlachos, A., Deller, T., Pfeifer, A., Fleischmann, B.K., Roeper, J., Fakler, B., and Klocker, N. (2009). Association with the auxiliary subunit PEX5R/Trip8b controls responsiveness of HCN channels to cAMP and adrenergic stimulation. *Neuron* 62, 814-825.

## Chapter 5:

### Conclusions

Since the initial discovery that fibroblasts could be directly converted into functional neurons with the overexpression of specific transcription factors (Vierbuchen et al., 2010), many groups have successfully generated diverse neuronal subtypes using similar methods (Blanchard et al., 2015; Caiazzo et al., 2011; Colasante et al., 2015; Kim et al., 2011; Son et al., 2011; Vadodaria et al., 2015; Victor et al., 2014). These studies demonstrated the robustness of transcription factors in not only establishing neuronal identity, but also specifying neuronal subtype identity. However, it was unknown whether different reprogramming combinations converged on a core set of neuronal, transcriptional feedback loops or maintained distinctly different transcriptional pathways and how the identities of the reprogramming factors influenced the transcriptional and functional phenotypes of the resulting iNs.

To address these questions, I, along with my colleagues, first determined that replacing the published neuronal reprogramming factor *Ascl1* with either *Ngn1* or *Ngn2*, in combination with *Brn2* and *Zic1*, was sufficient to generate functional iNs from fibroblasts. This swap resulted in iNs that expressed *Satb2*, a marker of upper layer cortical neurons, which was not detected in iNs derived from *Ascl1-Brn2-Zic1*. The same pattern was upheld when *Ascl1*, *Ngn1*, and *Ngn2* was paired with *Brn2* alone; iNs derived from both *Ngn1-Brn2* and *Ngn2-Brn2* had a higher percentage of *Satb2*-positive cells compared to iNs derived from *Ascl1-Brn2*. Concurrently, my colleagues discovered that *Brn3a* paired with either *Ngn1* or *Ngn2* was sufficient to generate functional iNs that greatly resembled peripheral sensory neurons of the dorsal root ganglion (Blanchard et al., 2015). Collectively, this suggested that different bHLH transcription factors (e.g. *Ascl1*, *Ngn1*, *Ngn2*) paired with different POU transcription factors (e.g. *Brn2*, *Brn3a*) give rise to iNs that have distinctly different transcriptional and functional phenotypes. Therefore, we set out to identify additional bHLH and POU transcription factor pairings capable of generating functional iNs and further investigate the distinct roles of bHLH and POU factors on establishing neuronal and neuronal subtype identity.

By conducting a large-scale screen, we identified over seventy pairs of bHLH and POU transcription factors sufficient to generate neuronal-like cells (candidate iNs) directly from fibroblasts.

These pairings were comprised of 16 bHLH, 9 POU, and 1 NR factors. RNA-Seq analysis on 35 of the candidate iN populations revealed transcriptional profiles greatly similar to endogenous neuronal populations. We suggested that the genes shared between the iN and endogenous neuron populations, but differed from fibroblasts, represented a “core” neuronal gene set that defined neuronal identity in the iN populations. Indeed, of the candidate iN populations tested for functionality, almost all cells within each population displayed resting membrane potentials within the range seen in endogenous neurons and fired action potentials upon depolarization. Therefore, we concluded that the transcriptomic similarities between the candidate iN and endogenous neurons populations were reflected in similar functional capabilities and the candidate iNs could be called genuine functional iNs.

As hypothesized, we also observed functional diversity within selected iN populations. In particular, we detected significantly greater voltage sag, which is indicative of the presence of h-currents, in the iN population generated with *NeuroD2/Brn3c* compared to *Ngn3/Oct4*. Interestingly, the gene *Hcn1*, which encodes a hyperpolarization-activated and cyclic nucleotide-gated channel known to contribute to h-currents (Biel et al., 2009), was expressed at higher levels in the RNA-Seq analyses of the *NeuroD2/Brn3c*-derived iNs compared to *Ngn3/Oct4*-derived iNs. Conducting WGCNA on all of the iN RNA-Seq datasets further confirmed the specificity of *Hcn1* expression in iN populations generated with the Class IV (*Brn3*) family of POU factors, which includes *Brn3c*. WGCNA also revealed other gene coexpression patterns, or modules, enriched in iN populations based on the class of bHLH or POU transcription factors used during reprogramming. Similar to the case of *Hcn1*, genes in these modules include those that underlie very specific neuronal functions (e.g. dopamine receptors *Drd2* and *Drd1a* in M26). Therefore, these analyses give insight into how reprogramming factor identities influence the transcriptional and functional phenotypes of the resulting iNs.

The transcriptional and functional analyses on the iN populations can also serve as a resource in selecting the appropriate reprogramming factors to generate neuronal subtypes with desired functional features. As others have previously demonstrated, subtype-specific iNs can be successfully used in disease modeling and cell replacement therapy (Dell'Anno et al., 2014; Kim et al., 2011; Wainger et al., 2015). An advantage in using iNs reprogrammed directly from fibroblasts versus neurons generated from a stem cell

source, particularly for modeling late-onset neurodegenerative diseases, is that age-dependent transcriptomic signatures are preserved after iN reprogramming (Mertens et al., 2015). In collaboration with colleagues in the Baldwin laboratory, we have demonstrated that the newly identified reprogramming factor pairs can reprogram human fibroblasts into functional neurons (data not shown). Although further analysis needs to be conducted to confirm that the reprogramming factors generate mouse and human iNs with shared transcriptional and functional profiles, we predict many features will be the same since mouse and human iNs generated with *Ngn1/Brn3a* and *Ngn2/Brn3a* have already been reported to function similarly (Blanchard et al., 2015).

Understanding how different transcription factors control transcriptional and functional phenotypes in iNs may also provide further insights into how cell fate decisions are made *in vivo*. D'Alessio et al., (2015) devised a computational approach to systematically identify candidate transcription factors that control cell identity in human cells by selecting for transcription factors that are expressed at relatively high levels in a cell-type specific fashion *in vivo* (D'Alessio et al., 2015). Here they identified potential reprogramming factors for various regions of the brain, including *Ascl1*, *Neurod1*, *Neurod6*, *Brn2*, and *Brn3a*. Similarly, Thompson et al., (2014) computationally identified a minimal set of approximately 80 transcription factors that provide unique expression signatures for many brain regions during murine brain development (Thompson et al., 2014). Included in this transcription factor set were the reprogramming factors, *Ascl1*, *Ngn2*, *Neurod6*, *Brn2*, and *Brn3c*. By cross-referencing the iN dataset against the brain regions defined by the same reprogramming factors, we may be able to discern features of these brain regions that are intrinsically driven by the specific transcription factors.

The extent to which different iN populations resemble specific endogenous neuronal subtypes found in the mouse brain remains an intriguing question. This question is difficult to address experimentally due to the large number of neuronal subtypes that can be used for comparison. This is further complicated by the fact that many neuronal subtypes have gene expression profiles that change during development. However, Blanchard et al. (2015) demonstrated that *Ngn1/Brn3a* and *Ngn2/Brn3a* generate iNs that greatly resemble peripheral sensory neurons of the DRG. In our study, PCA revealed that iNs generated with *Ngn1/Brn3a* and *Ngn2/Brn3a* were indeed proximal to DRG along principal

components 2 and 3 (data not shown). Therefore, PCA could be explored further to identify other iNs population that are similar to specific endogenous neuron populations. An alternative approach would be to use WGCNA to determine coexpression patterns shared between the iN and various endogenous neuron populations. In both cases, we would need to include additional RNA-Seq datasets from many other neuronal subtype populations to conduct a more comprehensive comparison.

The mammalian nervous system is comprised of a countless number of diverse neuronal subtypes that function together to execute complex behaviors and cognition. Understanding how such diversity arises provides insights into how circuits are formed, why certain neurological diseases target specific neuronal populations, and how to replace them for regenerative medicine. Through direct reprogramming, we elucidated the individual and synergistic roles of critical bHLH and POU transcription factors in generating transcriptional and functional diversity in iNs. Such knowledge will inform future reprogramming strategies and provide better means to generate and study diverse neuronal subtypes.

Chapter 5, in part, is currently being prepared for submission for publication of the material. Tsunemoto, R.K., Lee, S., Szűcs, A., Chubukov, P., Blanchard, J.W., Eade, K.T., Torkamani, A., Sanna, P.P., and Baldwin, K.K. The dissertation author was the primary investigator and author of this paper.

## References

Biel, M., Wahl-Schott, C., Michalakis, S., and Zong, X. (2009). Hyperpolarization-activated cation channels: from genes to function. *Physiol Rev* 89, 847-885.

Blanchard, J.W., Eade, K.T., Szucs, A., Lo Sardo, V., Tsunemoto, R.K., Williams, D., Sanna, P.P., and Baldwin, K.K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nature neuroscience* 18, 25-35.

Caiazzo, M., Dell'Anno, M.T., Dvoretzkova, E., Lazarevic, D., Taverna, S., Leo, D., Sotnikova, T.D., Menegon, A., Roncaglia, P., Colciago, G., Russo, G., Carninci, P., Pezzoli, G., Gainetdinov, R.R., Gustinich, S., Dityatev, A., and Broccoli, V. (2011). Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. *Nature* 476, 224-227.

Colasante, G., Lignani, G., Rubio, A., Medrihan, L., Yekhlef, L., Sessa, A., Massimino, L., Giannelli, S.G., Sacchetti, S., Caiazzo, M., Leo, D., Alexopoulou, D., Dell'Anno, M.T., Ciabatti, E., Orlando, M., Studer, M., Dahl, A., Gainetdinov, R.R., Taverna, S., Benfenati, F., and Broccoli, V. (2015). Rapid Conversion of Fibroblasts into Functional Forebrain GABAergic Interneurons by Direct Genetic Reprogramming. *Cell Stem Cell*.

D'Alessio, A.C., Fan, Z.P., Wert, K.J., Baranov, P., Cohen, M.A., Saini, J.S., Cohick, E., Charniga, C., Dadon, D., Hannett, N.M., Young, M.J., Temple, S., Jaenisch, R., Lee, T.I., and Young, R.A. (2015). A

Systematic Approach to Identify Candidate Transcription Factors that Control Cell Identity. *Stem cell reports*.

Dell'Anno, M.T., Caiazzo, M., Leo, D., Dvoretzkova, E., Medrihan, L., Colasante, G., Giannelli, S., Theka, I., Russo, G., Mus, L., Pezzoli, G., Gainetdinov, R.R., Benfenati, F., Taverna, S., Dityatev, A., and Broccoli, V. (2014). Remote control of induced dopaminergic neurons in parkinsonian rats. *The Journal of clinical investigation* *124*, 3215-3229.

Kim, J., Su, S.C., Wang, H., Cheng, A.W., Cassady, J.P., Lodato, M.A., Lengner, C.J., Chung, C.Y., Dawlaty, M.M., Tsai, L.H., and Jaenisch, R. (2011). Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. *Cell Stem Cell* *9*, 413-419.

Mertens, J., Paquola, A.C., Ku, M., Hatch, E., Bohnke, L., Ladjevardi, S., McGrath, S., Campbell, B., Lee, H., Herdy, J.R., Goncalves, J.T., Toda, T., Kim, Y., Winkler, J., Yao, J., Hetzer, M.W., and Gage, F.H. (2015). Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem Cell* *17*, 705-718.

Son, E.Y., Ichida, J.K., Wainger, B.J., Toma, J.S., Rafuse, V.F., Woolf, C.J., and Eggan, K. (2011). Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* *9*, 205-218.

Thompson, C.L., Ng, L., Menon, V., Martinez, S., Lee, C.K., Glattfelder, K., Sunkin, S.M., Henry, A., Lau, C., Dang, C., Garcia-Lopez, R., Martinez-Ferre, A., Pombero, A., Rubenstein, J.L., Wakeman, W.B., Hohmann, J., Dee, N., Sodt, A.J., Young, R., Smith, K., Nguyen, T.N., Kidney, J., Kuan, L., Jeromin, A., Kaykas, A., Miller, J., Page, D., Orta, G., Bernard, A., Riley, Z., Smith, S., Wohnoutka, P., Hawrylycz, M.J., Puellas, L., and Jones, A.R. (2014). A high-resolution spatiotemporal atlas of gene expression of the developing mouse brain. *Neuron* *83*, 309-323.

Vadodaria, K.C., Mertens, J., Paquola, A., Bardy, C., Li, X., Jappelli, R., Fung, L., Marchetto, M.C., Hamm, M., Gorris, M., Koch, P., and Gage, F.H. (2015). Generation of functional human serotonergic neurons from fibroblasts. *Molecular Psychiatry*.

Victor, Matheus B., Richner, M., Hermansteyne, Tracey O., Ransdell, Joseph L., Sobieski, C., Deng, P.-Y., Klyachko, Vitaly A., Nerbonne, Jeanne M., and Yoo, Andrew S. (2014). Generation of Human Striatal Neurons by MicroRNA-Dependent Direct Conversion of Fibroblasts. *Neuron* *84*, 311-323.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* *463*, 1035-1041.

Wainger, B.J., Buttermore, E.D., Oliveira, J.T., Mellin, C., Lee, S., Saber, W.A., Wang, A.J., Ichida, J.K., Chiu, I.M., Barrett, L., Huebner, E.A., Bilgin, C., Tsujimoto, N., Brenneis, C., Kapur, K., Rubin, L.L., Eggan, K., and Woolf, C.J. (2015). Modeling pain in vitro using nociceptor neurons reprogrammed from fibroblasts. *Nature neuroscience* *18*, 17-24.

## **Appendix A:**

### **Supplementary Methods**

#### **A.1 Specific methods for Chapter 2**

##### **A.1.1 Mouse fibroblast isolation**

Wild-type CD1 mice and *Pchd21:CRE x Ai9* mice (Boland et al., 2009) were bred at The Scripps Research Institute animal facility. MEFs were isolated from E13.5 embryos under a dissection microscope. The head, limbs, internal organs, and spinal column were removed and discarded to eliminate cells with neurogenic potential. The remaining tissue was manually dissociated with 0.25% trypsin (Gibco) for 20 minutes at 37 °C. The trypsin was subsequently diluted with MEF media (DMEM + 10% FBS and penicillin/streptomycin) and removed via centrifugation. Pelleted cells were resuspended in MEF media and seeded on gelatin-coated (0.01%) tissue culture plates. MEFs were grown to confluence and passaged at least twice before use.

Primary TTFs were isolated from 2-4 mm-long tail tips of P3 mouse pups. Tail tips were first rinsed in 70% ethanol, washed with HBSS (Invitrogen), chopped into smaller pieces, and dissociated 0.25% trypsin for 60 minutes at 37 °C. Subsequent steps are the same as in the MEF isolation protocol.

##### **A.1.2 Molecular cloning, cell culture, and lentiviral transduction**

The cDNAs for the transcription factors used were cloned into lentiviral constructs under the control of tetracycline operator (TetO). The following primers were used: Ngn1 forward and reverse, respectively, 5'-ATGCCTGCCCTTTGGAGACC and 5'-TTCAGCGAGGGTGCAGCAACC and Ngn2 forward and reverse, respectively, 5'-ATGTTCGTCAAATCTGAGACTCTGG and 5'-AAACCAGAGCTGGTCTCCACC. Replication-incompetent VSVg-coated lentiviral particles were packaged in 293T cells, harvested 48-hours after transfection, and filtered through a 45 micron PVDF membrane before use. The reprogramming method is a modification of a previously described protocol (Blanchard et al., 2015). Passage two MEFs were infected with lentivirus in MEF media. After 12-24 hours of infection, virus-containing media was replaced with fresh MEF media. Transcription factors were induced 48 hours post infection media by switching to MEF media supplemented with 5 µM doxycycline (Sigma). 4 days after initiating induction with doxycycline, MEF media was replaced with N3 media as published in Vierbuchen et al. (2010) but using N2 supplement (Gibco) in replacement of some individual components. 7 days post-induction, doxycycline was withdrawn. 10 days post-induction, media was switched to neural maintenance media, which consisted of a 1:1 mix of N3 media and Neurobasal (Invitrogen) supplemented with B27 (minus vitamin A, Gibco) and bFGF (10 ng/ml) (N3/NB media).

##### **A.1.3 Immunohistochemistry**

For immunofluorescence staining, cells were fixed with 4% paraformaldehyde for 10 min at room temperature. Cells were then washed three times with phosphate-buffered saline (PBS) and subsequently blocked in 5% horse serum and 0.1% Triton X-100 (Sigma) for 1 h at room temperature. Primary staining was performed overnight at 4 °C in block. Cells were again washed three times and then stained with secondary antibodies diluted in block for 1 h at room temperature. The following primary antibodies and dilutions were used: Tuj1 (Sigma-Aldrich T2200, 1:500), Map2 (Sigma-Aldrich M4403, 1:500), PSA-NCAM (MAB5324, 1:500), NeuN (MAB377, 1:500), Synaptophysin (BD 611880), GluR1 (ab31232, 1:200), Satb2 (ab51502, 1:25), Brn2 (sc-6029, 1:500), Cux1 (sc-13024, 1:200), Tbr1 (ab31940, 1:500) and Ctip2 (ab18465, 1:500).

##### **A.1.4 Electrophysiology**

Electrophysiological recordings were performed using whole cell patch clamp on Pcdh21-TdTomato positive cells with neuronal morphology between days 12-21 post-induction in MEFs. Recordings of current-induced action potentials were conducted using the following current clamp protocol: cells were injected with current starting at -100 pA, with each subsequent step increasing by 50 pA, for a total of 20 steps. Current measurements of voltage-gated sodium and potassium channels were conducted using the following voltage clamp protocol: cells were held at -60 mV, with each step increasing by 10 mV from holding potential for a total of 10 steps. To record currents elicited by the application of receptor agonists, recordings were done in voltage clamp mode with cells holding at -60 mV. Saturating concentrations of AMPA (100  $\mu$ M) and GABA (200  $\mu$ M) were used to maximize the response.

### A.1.5 Calcium imaging

Calcium imaging was performed on cells 14 to 21 days post-induction. Prior to imaging, cells were transduced with a lentivirus encoding a fluorescent calcium indicator under the control of a *Map2* promoter, Map2::GCaMP5.G (Addis et al., 2011). Imaging was conducted in NRS buffer (NaCl 140 mM, KCl 5 mM, 0.8 mM MgCl<sub>2</sub>, Hepes 10 mM, CaCl 2 mM, glucose 10 mM) and perfused at a steady flow rate of 250 ml h<sup>-1</sup>. Cells were sequentially exposed to of NRS buffer, L-glutamate (1 mM) and KCl (25 mM) by direct application to the area of interest. NRS was applied first to assess whether the cells were mechanosensitive, while KCl was applied last to confirm neuronal identity and functional Map2::GCaMP5.G expression. Only Pcdh21-TdTomato positive cells with neuronal morphology that did not respond to buffer, but responded to KCl were included in our analysis. Calcium responses were calculated as the change in fluorescence intensity ( $\Delta F$ ) over the initial fluorescence intensity ( $F_0$ ) and normalized to background ( $F_B$ ):  $(\Delta F - F_B)/(F_0 - F_B)$ .  $F_0$  was calculated as the average of the first 10 intensity measurements at the start of imaging.

## A.2 Specific methods for Chapter 3

### A.2.1 Mouse fibroblast isolation

We executed the same protocol as described in Appendix A.1.1, but with the following exceptions. We used both wild-type CD1 and C57BL/6J mice and heterozygous TauEGFP mice (Jackson Laboratory, STOCK *Mapt*<sup>tm1(EGFP)Klt/J</sup>, stock number: 004779) bred at The Scripps Research Institute animal facility. Primary tail-tip fibroblasts were isolated from P5 mouse pups.

### A.2.2 Molecular cloning, cell culture, and lentiviral transduction

We followed the same protocol as described in Appendix A.1.2, except doxycycline was withdrawn 8 days post-induction. The cDNAs for the mouse transcription factors used were cloned into lentiviral constructs under the control of tetracycline operator (TetO). The cDNA for *BRN3A* was the only human factor, which has 97% homology to the mouse Brn3a peptide, and was cloned as described in Blanchard et al. (2015).

### A.2.3 Immunohistochemistry

We utilized the same protocol as described in Appendix A.1.3. The following primary antibodies and dilutions were used: Tuj1 (Sigma-Aldrich T2200, 1:500), Map2 (Sigma-Aldrich M4403, 1:500), and synapsin 1 (Synaptic Systems 106103, 1:500).

### A.2.4 FACS purification

Reprogrammed candidate iNs generated from heterozygous TauEGFP MEFs were prepped for FACS by first detaching cells from culture plate using Accutase (Innovative Cell Technologies). Accutase was subsequently diluted with neural maintenance media (N3/NB media) and removed via centrifugation.

Pelleted cells were resuspended in neural maintenance media, triturated, and strained through 35  $\mu\text{m}$  nylon mesh filter to obtain single cell suspensions. Viability markers DAPI (1  $\mu\text{M}$ ) and DRAQ5 (BioStatus DR50050, 1  $\mu\text{M}$ ) were added to the suspension at least 10 minutes prior to sorting. Appropriate gates for FACS were set based on TauEGFP, DAPI and DRAQ5 intensities to isolate live TauEGFP-positive cells as shown in Figure 3.9 using the MoFlo<sup>®</sup> Astrios<sup>™</sup> (Beckman Coulter). Isolated cells were sorted into TRIzol<sup>®</sup> LS (Invitrogen).

Similarly, endogenous neuron populations were isolated from the appropriate transgenic reporter mice at P21 (Table 3.2). Dissected tissue samples were dissociated as in Brewer and Torricelli (2007) with the following modifications. Manual homogenization was conducted with a scalpel rather than with a tissue slicer. As in Hazen et al. (2016), we also used papain-containing L-cysteine (Worthington Biochemical, PAP2 10 units/ml) because its higher activity allowed for shorter dissociation times (15 minutes total). During papain digestion, samples were triturated every 5 minutes using P1000 plastic tips instead of siliconized Pasteur glass pipettes. After centrifugation using the density gradient, we found viable neurons in the fraction containing the cell pellet and the fraction 2 mls immediately above the pellet. Both fractions were combined and washed once in 10 mls of HAGB (Hibernate-A (Gibco A1247501), 1X B-27 supplement (Gibco 12587010), 500  $\mu\text{M}$  GlutaMAX (Gibco 35050061)). After a subsequent centrifugation, pelleted cells were resuspended in HAGB, filtered and kept on ice until consequential FACS sorting. As with the candidate iNs, viability markers DAPI and DRAQ5 were added to the suspension and appropriate gates were set to purify cells into TRIzol<sup>®</sup> LS.

#### **A.2.5 RNA isolation**

Total RNA was isolated from FACS-sorted cells using Direct-zol<sup>™</sup> RNA MiniPrep (Zymo Research) according to the manufacturer's protocol, except linear acrylamide (1  $\mu\text{g}$ ) was added to each sample prior to the first step and Zymo-Spin<sup>™</sup> IC columns were used in replacement of IIC columns. RNA quality and quantity was determined with an Agilent 2100 Bioanalyzer. RNA integrity numbers (RINs) for all induced neuron samples were between 6 and 10 (median = 8.7). The amount of RNA per sorted event was between 1 and 15  $\mu\text{g}$  (median = 7.9  $\mu\text{g}$ ). Therefore, approximately 1,500 to 2,000 cells were required to yield 10  $\mu\text{g}$  RNA for library input. Similarly, for the endogenous neuron samples, RINs were between 6.9 and 9.2 (median = 7.8) and the amount of RNA per sorted event was between 1.7 and 3.9  $\mu\text{g}$  (median = 3.1  $\mu\text{g}$ ).

#### **A.2.6 RNA-Seq library preparation and sequencing**

Typically 10  $\mu\text{g}$  of purified, high quality RNA served as input for SMARTer<sup>®</sup> Ultra<sup>™</sup> Low Input RNA Kit for Sequencing – v3 (Clontech Laboratories, Inc.). A few replicate libraries were prepped from 1-7  $\mu\text{g}$  of input total RNA. Amplified cDNA was assessed for quality using High Sensitivity DNA Kit (Agilent Technologies) and sheared using the Covaris system. Sequencing libraries were subsequently prepped using NEBNext<sup>®</sup> Ultra<sup>™</sup> DNA Library Prep Kit for Illumina<sup>®</sup>. 75 base pair single end reads generated using Illumina's NextSeq platform were mapped to the mouse genome (mm10) by first removing adapters and low quality bases using Trimmomatic (v0.32, ILLUMINACLIP: TruSeq3-SE.fa:2:30:10 LEADING:3 TRAILING:3) (Bolger et al., 2014). Reads were then aligned using STAR (Dobin et al., 2013) and counts were generated using HTSeq (Anders et al., 2015). Mm10 did not include *Ascl5*, therefore, we added it to the reference GTF file in HTSeq. It is also important to note that some libraries were prepped using SMARTer<sup>®</sup> Ultra<sup>™</sup> Low Input RNA for Illumina<sup>®</sup> Sequencing – HV (Clontech Laboratories, Inc. and sequenced on Illumina's HiSeq platform, resulting in 100 base pair single reads. Details can be found in Table 3.1 and Table 3.2.

#### **A.2.7 RNA-Seq data analysis (DESeq2 and PCA)**

RNA-Seq data was analyzed using R (Team, 2011), an open source programming language and environment for statistical computing and visualization. Multiple R packages available through Bioconductor (Gentleman et al., 2004) were used during analysis. Differential gene expression analysis was conducted using DESeq2 (Love et al., 2014). Heat maps were generated using gplots (Warnes et al.,

2015). Both *rgl* (Adler, 2016) and *pca3d* (Weiner, 2015) were used to calculate and generate principal component plots.

### A.3 Specific methods for Chapter 4

#### A.3.1 Electrophysiology

TauEGFP MEFs were reprogrammed and cultured, as described in Appendix A.2.2, on Thermanox® plastic coverslips (33 mm diameter). Coverslips were placed in the recording chamber mounted on an Olympus BX51 microscope. To identify TauEGFP positive cells that expressed synapsin, we transduced our candidate iNs with lentivirus encoding the fluorescent red protein, TdTomato, under the control of a *SYN1* promoter. Spontaneous activity and evoked responses were recorded from identified cells at day 16 to 24 post-induction under whole-cell patch clamp at 33 °C. Similar to the electrophysiology protocol described in Blanchard et al. (2015), signals were amplified using a MultiClamp700B (Molecular Devices) and acquired using the data acquisition software DASyLab v.11 (National Instruments) at 20 kHz. Patch pipettes with input resistances of 6–8 MOhm were pulled from standard wall glass of 1.5-mm OD (Warner Instruments) and filled with solution containing 120 mM potassium-glutamate, 10 mM KCl, 10 mM HEPES, 10 mM EGTA, 2 mM MgATP, 0.3 mM Na3GTP at pH 7.3. The bath solution (artificial cerebrospinal fluid) was composed of 125 mM NaCl, 2.5 mM KCl, 2 mM CaCl2, 1 mM MgCl2, 1.25 mM NaH2PO4, 26 mM NaHCO3, and 25 mM glucose. To record voltage responses of the identified iNs, we used incrementing levels of constant, rectangular current steps of 350 ms duration. The initial current step level was –50 to –200 pA depending on the observed input resistance of the cell. Steps were incremented by +2 or +5 pA in successive cycles of stimulation at a rate of 1 Hertz (Hz). Analysis of the evoked responses was performed in software developed by A. Szücs (IVAnalyzer). For each cell, several physiological parameters, including the resting membrane potential, rheobase, input resistance at rest, and spike amplitude, were measured.

Spontaneous excitatory or inhibitory postsynaptic potentials were rarely observed in the recorded iNs. We performed voltage clamp recordings of postsynaptic current whenever such activity was detected (150-200 s recordings at -50 mV holding potential).

#### A.3.2 WGCNA

WGCNA has been previously described in detail (Zhang and Horvath, 2005) and also summarized in papers utilizing this technique (Hawrylycz et al., 2015; Hawrylycz et al., 2012). DESeq2 vsd-normalized counts of all iN and MEF population replicates ( $n = 72$ ) served as input into a user-friendly WGCNA R library (Langfelder and Horvath, 2008). To reduce the noise from low expressing genes in our dataset, we only included genes in which the non-normalized counts were greater than 200 in at least one iN or MEF population, in both replicates ( $n = 12,549$ ). We constructed a signed network, with a power of 12, using the default parameters except  $\text{deepSplit} = 4$  and  $\text{cutHeight} = 0.999$ . Modules were merged if their module eigengenes (ME) were correlated with  $R > 0.8$ . Module hub genes were those that had the highest module membership ( $k_{ME}$ ) for that module, which was calculated as the Pearson correlation between the gene and the corresponding ME. Associated gene ontology terms for each module were determined using PANTHER Overrepresentation Test with the PANTHER GO biological process complete annotation data set (Mi et al., 2013; Mi et al., 2016).

#### References:

Addis, R.C., Hsu, F.C., Wright, R.L., Dichter, M.A., Coulter, D.A., and Gearhart, J.D. (2011). Efficient conversion of astrocytes to functional midbrain dopaminergic neurons using a single polycistronic vector. *PLoS One* 6, e28719.

Adler, D.M., Duncan; {others} (2016). *rgl*: 3D Visualization Using OpenGL.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq—a Python framework to work with high-throughput sequencing data. *Bioinformatics (Oxford, England)* *31*, 166-169.

Blanchard, J.W., Eade, K.T., Szucs, A., Lo Sardo, V., Tsunemoto, R.K., Williams, D., Sanna, P.P., and Baldwin, K.K. (2015). Selective conversion of fibroblasts into peripheral sensory neurons. *Nature neuroscience* *18*, 25-35.

Boland, M.J., Hazen, J.L., Nazor, K.L., Rodriguez, A.R., Gifford, W., Martin, G., Kupriyanov, S., and Baldwin, K.K. (2009). Adult mice generated from induced pluripotent stem cells. *Nature* *461*, 91-94.

Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)* *30*, 2114-2120.

Brewer, G.J., and Torricelli, J.R. (2007). Isolation and culture of adult neurons and neurospheres. *Nature protocols* *2*, 1490-1498.

Dobin, A., Davis, C.A., Schlesinger, F., Drenkow, J., Zaleski, C., Jha, S., Batut, P., Chaisson, M., and Gingeras, T.R. (2013). STAR: ultrafast universal RNA-seq aligner. *Bioinformatics (Oxford, England)* *29*, 15-21.

Gentleman, R.C., Carey, V.J., Bates, D.M., Bolstad, B., Dettling, M., Dudoit, S., Ellis, B., Gautier, L., Ge, Y., Gentry, J., Hornik, K., Hothorn, T., Huber, W., Iacus, S., Irizarry, R., Leisch, F., Li, C., Maechler, M., Rossini, A.J., Sawitzki, G., Smith, C., Smyth, G., Tierney, L., Yang, J.Y., and Zhang, J. (2004). Bioconductor: open software development for computational biology and bioinformatics. *Genome biology* *5*, R80.

Hawrylycz, M., Miller, J.A., Menon, V., Feng, D., Dolbeare, T., Guillozet-Bongaarts, A.L., Jegg, A.G., Aronow, B.J., Lee, C.K., Bernard, A., Glasser, M.F., Dierker, D.L., Menche, J., Szafer, A., Collman, F., Grange, P., Berman, K.A., Mihalas, S., Yao, Z., Stewart, L., Barabasi, A.L., Schulkin, J., Phillips, J., Ng, L., Dang, C., Haynor, D.R., Jones, A., Van Essen, D.C., Koch, C., and Lein, E. (2015). Canonical genetic signatures of the adult human brain. *Nature neuroscience* *18*, 1832-1844.

Hawrylycz, M.J., Lein, E.S., Guillozet-Bongaarts, A.L., Shen, E.H., Ng, L., Miller, J.A., van de Lagemaat, L.N., Smith, K.A., Ebbert, A., Riley, Z.L., Abajian, C., Beckmann, C.F., Bernard, A., Bertagnolli, D., Boe, A.F., Cartagena, P.M., Chakravarty, M.M., Chapin, M., Chong, J., Dalley, R.A., Daly, B.D., Dang, C., Datta, S., Dee, N., Dolbeare, T.A., Faber, V., Feng, D., Fowler, D.R., Goldy, J., Gregor, B.W., Haradon, Z., Haynor, D.R., Hohmann, J.G., Horvath, S., Howard, R.E., Jeromin, A., Jochim, J.M., Kinnunen, M., Lau, C., Lazarz, E.T., Lee, C., Lemon, T.A., Li, L., Li, Y., Morris, J.A., Overly, C.C., Parker, P.D., Parry, S.E., Reding, M., Royall, J.J., Schulkin, J., Sequeira, P.A., Slaughterbeck, C.R., Smith, S.C., Sodt, A.J., Sunkin, S.M., Swanson, B.E., Vawter, M.P., Williams, D., Wohnoutka, P., Zielke, H.R., Geschwind, D.H., Hof, P.R., Smith, S.M., Koch, C., Grant, S.G., and Jones, A.R. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* *489*, 391-399.

Hazen, J.L., Faust, G.G., Rodriguez, A.R., Ferguson, W.C., Shumilina, S., Clark, R.A., Boland, M.J., Martin, G., Chubukov, P., Tsunemoto, R.K., Torkamani, A., Kupriyanov, S., Hall, I.M., and Baldwin, K.K. (2016). The Complete Genome Sequences, Unique Mutational Spectra, and Developmental Potency of Adult Neurons Revealed by Cloning. *Neuron* *89*, 1223-1236.

Langfelder, P., and Horvath, S. (2008). WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics* *9*, 559.

Love, M.I., Huber, W., and Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology* *15*, 550.

Mi, H., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2013). Large-scale gene function analysis with the PANTHER classification system. *Nature protocols* 8, 1551-1566.

Mi, H., Poudel, S., Muruganujan, A., Casagrande, J.T., and Thomas, P.D. (2016). PANTHER version 10: expanded protein families and functions, and analysis tools. *Nucleic Acids Res* 44, D336-342.

Team, R.D.C. (2011). *R: A Language and Environment for Statistical Computing*. (Vienna, Austria : the R Foundation for Statistical Computing.).

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* 463, 1035-1041.

Warnes, G.R.B., Ben; Bonebakker, Lodewijk; , Gentleman, R.H., Wolfgang; Liaw, Andy; , Lumley, T.M., Martin; Magnusson, Arni; Moeller, Steffen; , and Schwartz, M.V., Bill (2015). *gplots: Various R Programming Tools for Plotting Data*.

Weiner, J. (2015). *pca3d: Three Dimensional PCA Plots*.

Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Statistical applications in genetics and molecular biology* 4, Article17.