SMALL MOLECULES MODULATE CHROMATIN ACCESSIBILITY TO PROMOTE NEUROG2-MEDIATED FIBROBLAST-TO-NEURON REPROGRAMMING

APPROVED BY SUPERVISORY COMMITTEE

Mentor: Chun-Li Zhang, Ph.D.

Committee Chair: Jane E. Johnson, Ph.D.

Committee Member: Tae-Kyung Kim, Ph.D.

Committee Member: Eric Olson, Ph.D.

DEDICATION

This work is dedicated to my parents in honor of their unwavering support and encouragement.

SMALL MOLECULES MODULATE CHROMATIN ACCESSIBILITY TO PROMOTE NEUROG2-MEDIATED FIBROBLAST-TO-NEURON REPROGRAMMING

by

DEREK KURTIS SMITH

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center

Dallas, Texas

August 12, 2016

SMALL MOLECULES MODULATE CHROMATIN ACCESSIBILITY TO PROMOTE NEUROG2-MEDIATED FIBROBLAST-TO-NEURON REPROGRAMMING

Publication No. _____

Derek Kurtis Smith, Ph.D.

The University of Texas Southwestern Medical Center, 2016

Supervising Professor: Chun-Li Zhang, Ph.D.

The activity of pro-neural signaling molecules and transcription factors is sufficient to induce the transdifferentiation of lineage-restricted fibroblasts into functional neurons; however, a mechanistic model of the immediate-early events that catalyze this conversion has not been well defined. We utilized a high-efficiency reprogramming system of NEUROG2, forskolin (F), and dorsomorphin (D) to characterize the genetic and epigenetic events that initiate an acquisition of neuronal identity in fetal human fibroblasts. NEUROG2 immediately activates a neurogenic program, but is only sufficient to impart a functional identity in the presence of FD. These small molecules promote NEUROG2 and CREB1 co-transcription, induce *SOX4* expression, and promote SOX4-dependent chromatin remodeling. Genome-wide occupancy analysis revealed that SOX4 targets numerous SWI/SNF complex subunits and co-binds with NEUROG2 to enhance the expression of diverse neurogenic factors. The overexpression of SWI/SNF chromatin remodeling factors or treatment with small molecules that modify chromatin accessibility enhanced NEUROG2-mediated neuronal reprogramming of adult human skin fibroblasts. This work represents the first comprehensive mechanism for the immediate events that catalyze neuronal transdifferentiation.

ACKNOWLEDGEMENTS

I owe my sincerest gratitude to Dr. Chun-Li Zhang for his guidance during my doctoral research. He instilled in me the meaning of scientific integrity, the value of being self-critical, and the importance of methodical experimentation. His invaluable friendship, guidance, persistence, knowledge, and work ethic were fundamental to my success and growth as a scientist. For these reasons and many more, I truly appreciate his contributions to my training.

I sincerely thank Dr. Meng-Lu Liu for jumpstarting this research project. Without his vast expertise, feedback, and willingness to constantly answer questions, this work would not have been possible. I also owe a debt of gratitude to my friends and co-workers in the Zhang laboratory. I specifically want to thank Dr. Yu Tang, Dr. Jianjing Yang, Dr. Leilei Wang, and Yuhua Zou for performing experiments or providing assistance directly related to this work.

I greatly appreciate the generous feedback and guidance provided by my dissertation committee members Dr. Jane Johnson, Dr. Tae-Kyung Kim, and Dr. Eric Olson. I also want to acknowledge The University of Texas Southwestern Medical Center Graduate Program for providing this unique opportunity to grow in both knowledge and skill as a research scientist.

Lastly, I thank my family for their encouragement throughout my graduate training. I am particularly grateful to my parents for their commitment to my education and happiness. It was their constant support and love that carried me through the most challenging obstacles in my life. Without their unwavering support, I would not be where I am today.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Smith DK, Yang J, Liu M-L, Zhang C-L (2016) Small molecules modulate chromatin accessibility to promote fibroblast-to-neuron reprogramming. *In review*.

Smith DK, Wang L, Zhang C-L (2016) Physiological, pathological, and engineered cell identity reprogramming in the central nervous system. WIREs Dev Biol, 324.

Smith DK, He M, Zhang C-L, Zheng J (2016) The therapeutic potential of cell identity reprogramming for the treatment of aging-related neurodegenerative disorders. Prog Neurobiol 15:30067-30084.

Islam MM, **Smith DK**, Niu W, Fang S, Iqbal N, Sun G, Shi Y, Zhang C-L (2015) Enhancer analysis unveils genetic interactions between TLX and SOX2 in neural stem cells and in vivo reprogramming. Stem Cell Reports 5:805-815.

Smith DK, Niu W, Zhang C-L (2015) Induced neurogenesis as a mechanism for adult central nervous system regeneration. Neural Stem Cells in Health and Disease, Chapter 12:337-357.

Smith DK, Zhang C-L (2015) Regeneration through reprogramming adult cell identity in vivo. Am J Pathol 185:2619-2628.

Niu W, Zang T, **Smith DK**, Vue TY, Zou Y, Bachoo R, Johnson JE, Zhang C-L (2015) SOX2 reprograms resident astrocytes into neural progenitors in the adult brain. Stem Cell Rep 4:780-794.

Qin S, Niu W, Iqbal N, **Smith DK**, Zhang C-L (2014) Orphan nuclear receptor TLX regulates astrogenesis by modulating BMP signaling. Front Neurosci 8:74.

Niu W, Zang T, Zou Y, Fang S, **Smith DK**, Bachoo R, Zhang C-L (2013) In vivo reprogramming of astrocytes to neuroblasts in the adult brain. Nat Cell Biol 15:1164-1175.

LIST OF FIGURES

LIST OF TABLES

LIST OF APPENDICES

LIST OF DEFINED ABBREVIATIONS

°C	degree(s) Celsius
× g	relative centrifugal force
µg	microgram(s)
µl	microliter(s)
µm	micrometer(s)
µM	micromolar
5-azacitidine	4-amino-1-β-D-ribofuranosyl-1,3,5-triazin-2(1*H*)-one
AP1	activator protein 1
ARID1A	AT-rich interaction domain 1A
ASCL1	achaete-scute family bHLH transcription factor 1
ATAC-seq	assay for transposase-accessible chromatin using sequencing
BCL11B	B-cell CLL/lymphoma 11B
BCL2	B-cell CLL/lymphoma 2
BDNF	brain-derived neurotrophic factor
bHLH	basic helix-loop-helix
BMP	bone morphogenetic protein
BSA	bovine serum albumin
BSD	blasticidin
cAMP	cyclic adenosine monophosphate
caPRKACA	constitutively active protein kinase A catalytic subunit
CD44	CD44 molecule (Indian blood group)

cDNA	complementary DNA
CHAT	choline O-acetyltransferase
ChIP	chromatin immunoprecipitation
ChIP-seq	chromatin immunoprecipitation sequencing
CHIR99021	6-[[2-[[4-(2,4-dichlorophenyl)-5-(5-methyl-1*H*-imidazol-2-yl)-2-pyrimidinyl]amino]ethyl]amino]-3-pyridinecarbonitrile
CHRNA3	cholinergic receptor nicotinic alpha 3 subunit
cm	centimeter(s)
CMV	cytomegalovirus
CREB1	cyclic AMP responsive element binding protein 1
CREB1 pS133	cyclic AMP responsive element binding protein 1 phosphorylated serine 133
D	dorsomorphin, 6-[4-(2-Piperidin-1-ylethoxy)phenyl]-3-pyridin-4-ylpyrazolo [1,5-a]pyrimidine
DAPT	*N*-[(3,5-difluorophenyl)acetyl]-L-alanyl-2-phenyl]glycine-1,1-dimethylethyl ester
DCX	doublecortin
decitabine	4-amino-1-(2-deoxy-β-D-*erythro*-pentofuranosyl)-1,3,5-triazin-2(1*H*)-one
DLL1	delta-like 1
DLL3	delta-like 3
DLX1	distal-less homeobox 1
DLX2	distal-less homeobox 2
DLX5	distal-less homeobox 5
DMEM	Dulbecco's modified eagle medium
DNA	deoxyribonucleic acid

DTT	dithiothreitol
DZNep	(1*S*,2*R*,5*R*)-5-(4-Amino-1*H*-imidazo[4,5-*c*]pyridin-1-yl)-3-(hydroxymethyl)-3-cyclopentene-1,2-diol hydrochloride
EDTA	ethylenediaminetetraacetic acid
EMP1	epithelial membrane protein 1
EMSA	electrophoretic mobility shift assay
EPC1	enhancer of polycomb homolog 1
EPZ5676	9H-purin-6-amine,9-[5-deoxy-5-[[cis-3-[2-[6-(1,1-dimethylethyl)-1H-benzimidazol-2-yl]ethyl]cyclobutyl](1-methylethyl)amino]-β-D-ribofuranosyl]
ESC	embryonic stem cell
F	forskolin, 7β-Acetoxy-8,13-epoxy-1α,6β,9α-trihydroxylabd-14-en-11-one
FGF	fibroblast growth factor
FGF2	fibroblast growth factor 2
FK228	cyclo[(2*Z*)-2-amino-2-butenoyl-L-valyl-(3*S*,4*E*)-3-hydroxy-7-mercapto-4-heptenoyl-D-valyl-D-cysteinyl], cyclic (3-5) disulfide
FPKM	fragments per kilobase of exon per million mapped fragments
FOXA2	forkhead box A2
FOXG1	forkhead box G1
G	gauge
GABA	gamma-aminobutyric acid
GADD45A	growth arrest and DNA-damage-inducible, alpha
GADD45B	growth arrest and DNA-damage-inducible, beta
GADD45G	growth arrest and DNA-damage-inducible, gamma

GAPDH	glyceraldehyde-3-phosphate dehydrogenase
GDF5	growth differentiation factor 1
GDNF	glial cell-derived neurotrophic factor
GFAP	glial fibrillary acidic protein
GFP	green fluorescent protein
GO6983	3-[1-[3-(dimethylamino)propyl]-5-methoxy-1*H*-indol-3-yl]-4-(1*H*-indol-3-yl)-1*H*-pyrrole-2,5-dione
GSK126	1-(S)-sec-butyl-N-((4,6-dimethyl-2-oxo-1,2-dihydropyridin-3-yl)methyl)-3-methyl-6-(6-(piperazin-1-yl)pyridin-3-yl)-1H-indole-4-carboxamide
GSK3	glycogen synthase kinase 3
H3K27	histone 3 lysine 27
H3K4	histone 3 lysine 4
H3K9	histone 3 lysine 9
HA	hemagglutinin tag
HCl	hydrochloric acid
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HES6	hes family bHLH transcription factor 6
HMG	high mobility group
HPRT1	hypoxanthine phosphoribosyltransferase 1
HRP	horseradish peroxidase
I-BET151	7-(3,5-dimethyl-4-isoxazolyl)-8-(methyloxy)-1-[(1R)-1-(2-pyridinyl)ethyl]-1,3-dihydro-2H-imidazo[4,5-c]quinolin-2-one

I-BET762	4H-[1,2,4]Triazolo[4,3-a][1,4]benzodiazepine-4-acetamide, 6-(4-chlorophenyl)-N-ethyl-8-methoxy-1-methyl-, (4S)
IGF1	insulin-like growth factor 1
IgG	immunoglobulin G
INSM1	insulinoma associated 1
IRES	internal ribosome entry site
ISL1	ISL LIM homeobox 1
JNK	c-Jun N-terminal kinase
JQ1	(6*S*)-4-(4-Chlorophenyl)-2,3,9-trimethyl-6*H*-thieno[3,2-*f*][1,2,4]triazolo[4,3-*a*][1,4]diazepine-6-acetic acid 1,1-dimethylethyl ester
KCl	potassium chloride
K_d	dissociation constant
KLF4	Kruppel-like factor 4
LDN193189	4-[6-[4-(1-piperazinyl)phenyl]pyrazolo[1,5-a]pyrimidin-3-yl]-quinoline hydrochloride
LHX3	LIM homeobox 3
LHX6	LIM homeobox 6
LiCl	lithium chloride
liproxstatin-1	N-[(3-chlorophenyl)methyl]-spiro[piperidine-4,2′(1′H)-quinoxalin]-3′-amine
LMO2	LIM domain only 2
LMO4	LIM domain only 4
LMX1A	LIM homeobox transcription factor 1 alpha
ln	natural log

M	molar
MAP2	microtubule-associated protein 2
MAPK14	mitogen-activated protein kinase 14
MAPT	microtubule protein associated tau
MEF2A	myocyte enhancer factor 2A
methanol	CH_3OH
miR	microRNA
ml	milliliter(s)
MNX1	motor neuron and pancreas homeobox 1
MOPS	3-(*N*-morpholino)-propanesulfonic acid
mRNA	messenger RNA
MTF2	metal response element binding transcription factor 2
MYC	v-myc avian myelocytomatosis viral oncogene homolog
MYOD1	myogenic differentiation 1
MYT1L	myelin transcription factor 1 like
NCAM	neural cell adhesion molecule 1
NES	nestin
NEUROD1	neuronal differentiation 1
NEUROD2	neuronal differentiation 2
NEUROD4	neuronal differentiation 4
NEUROG2	neurogenin 2
NFD	neurogenin 2, forskolin, and dorsomorphin
NF-Y	nuclear transcription factor Y

ng	nanogram(s)
NHLH1	nescient helix-loop-helix 1
NR4A1	nuclear receptor subfamily 4, group A, member 1
NR4A2	nuclear receptor subfamily 4, group A, member 2
NT-3	neurotrophin 3
ORF	open reading frame
PAX6	paired box 6
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PHOX2A	paired-like homeobox 2A
PITX3	paired-like homeodomain 3
PKC	protein kinase C
POU3F2	POU class 3 homeobox 2
POU3F3	POU class 3 homeobox 3
POU5F1	POU class 5 homeobox 1
purmorphamine	9-cyclohexyl-*N*-[4-(4-morpholinyl)phenyl]-2-(1-naphthalenyloxy)-9*H*-purin-6-amine
PRKACA	protein kinase A catalytic subunit
RBFOX3	RNA binding protein fox-1 homolog 3
RepSox	2-[3-(6-methyl-2-pyridinyl)-1H-pyrazol-4-yl]-1,5-naphthyridine
REST	RE1 silencing transcription factor
RFX1	regulatory factor X1
RG108	*N*-phthalyl-L-tryptophan

RNA	ribonucleic acid
RNA-seq	RNA sequencing
RND2	Rho family GTPase 2
rpm	rotations per minute
SAHA	*N*-hydroxy-*N*'-phenyloctanediamide
SB431542	4-[4-(1,3-benzodioxol-5-yl)-5-(2-pyridinyl)-1*H*-imidazol-2-yl]benzamide
SDCBP	syndecan binding protein
SDS	sodium lauryl sulfate
SEMA3C	semaphorin 3C
SGI1027	*N*-[4-[(2-Amino-6-methyl-4-pyrimidinyl)amino]phenyl]-4-(4-quinolinylamino)benzamide
SHH	sonic hedgehog
shRNA	microRNA30-based short hairpin RNA
SIK1	salt-inducible kinase 1
SLC17A6	solute carrier family 17 member 6
SLC17A7	solute carrier family 17 member 7
SLC18A3	solute carrier family 18 member 3
SMARCA2	SWI/SNF related actin dependent regulator of chromatin, subfamily a member 2
SMARCA4	SWI/SNF related actin dependent regulator of chromatin, subfamily a member 4
SMARCB1	SWI/SNF related actin dependent regulator of chromatin, subfamily b member 1
SMARCC2	SWI/SNF related actin dependent regulator of chromatin, subfamily c member 2
SMARCD1	SWI/SNF related actin dependent regulator of chromatin, subfamily d member 1
SMARCE1	SWI/SNF related actin dependent regulator of chromatin, subfamily e member 1

SOX2	sex determining region Y-box 2
SOX4	sex determining region Y-box 4
SOX9	sex determining region Y-box 9
SOX11	sex determining region Y-box 11
SOX12	sex determining region Y-box 12
SP600125	1,9-pyrazoloanthrone
SSTR2	somatostatin receptor 2
STAT3	signal transducer and activator of transcription 3
SWI/SNF	switch/sucrose non-fermentable
SYN1	synapsin 1
SYT1	synaptotagmin 1
T2A	*Thosea asigna* 2A peptide
TBR1	T-box, brain 1
TBR2	T-box, brain 2
TBST	tris-buffered saline and tween 20
TCF12	transcription factor 12
TCF3	transcription factor 3
TCF4	transcription factor 4
TCFL5	transcription factor-like 5
TCF/LEF	T-cell factor/lymphoid enhancer factor
TEAD2	TEA domain transcription factor 2
TEK	TEK tyrosine kinase, endothelial
TET1	ten-eleven translocation methylcytosine dioxygenase 1

TGFB1	transforming growth factor beta 1
thiazovivin	*N*-benzyl-2-(pyrimidin-4-ylamino)thiazole-4-carboxamide
TP53	tumor protein p53
Tris	2-amino-2-hydroxymethyl-propane-1,3-diol
TSS	transcription start site
TTNPB	4-[(E)-2-(5,6,7,8-Tetrahydro-5,5,8,8-tetramethyl-2-naphthalenyl)-1-propenyl]benzoic acid
TUBB3	tubulin, beta 3 class III
UNC669	(2-(phenylamino)-1,4-phenylene)bis((4-(pyrrolidin-1-yl)piperidin-1-yl) methanone)
v/v	volume per volume
valproic acid	2-propylpentanoic acid
w/v	weight per volume
Y-27632	*trans*-4-[(1*R*)-1-Aminoethyl]-*N*-4-pyridinylcyclohexanecarboxamide dihydrochloride
ZBTB18	zinc finger and BTB domain containing 18
zebularine	1-β-D-ribofuranosyl-2(1*H*)-pyrimidinone

CHAPTER 1

INTRODUCTION

Cell identity is defined by cellular function and this function is a product of time and controlled transcription. Throughout fibroblast development, the transcriptome is defined by diverse pro-fibroblast signaling cues that modulate the activity of transcription factors. Targeted changes in the expression or activity of these transcription factors have enabled researchers to re-define the identity of developmentally differentiated fibroblasts. In a remarkable example of cellular engineering, the overexpression of only three pro-neural transcription factors is sufficient to induce functional neurons from mouse fibroblasts (Vierbuchen et al. 2010). Broadening this core set of transcription factors and integrating small molecules has enabled the production of diverse neuron subtypes such as cholinergic, dopaminergic, GABAergic, glutamatergic, and motor neurons (Table 1). While the activity of pro-neural transcription factors and signaling molecules has been used to induce transdifferentiation, a mechanistic model of the early molecular events that catalyze this conversion has not been well defined.

Recently, we demonstrated that the single transcription factor Neurogenin 2 (NEUROG2) combined with the small molecules forskolin (F) and dorsomorphin (D) efficiently transdifferentiate human fetal lung fibroblasts into functional cholinergic neurons (Liu et al. 2013). This model's reduced complexity enabled us to define the critical early molecular events that catalyze this cell identity change. This dissertation research highlights the coordinated transcriptional programs and chromatin remodeling events that underpin the genetic and epigenetic reprogramming of fibroblast-derived cholinergic neurons.

1.1. Developmental and Engineered Cell Identity

1.1.1. Neural Development

The human central nervous system is an intricately structured network of diverse cell types that cooperatively function to process and relay information. Primary neurulation is one of the earliest events to occur in the developing human embryo preceded only by fertilization, pro-nuclei fusion, and blastocyst implantation. Cascades of bone morphogenetic protein (BMP), fibroblast growth factor (FGF), sonic hedgehog, and Wnt proteins pattern the emergent neural plate opposite the primitive streak. Invagination of this region closes the neural tube and establishes the birthplace of all neural cell types (Stiles et al. 2010). Proliferative neuroepithelial cells residing in the neural tube divide asymmetrically to produce the neural stem cells responsible for neurogenesis and gliogenesis. As these stem cell populations generate billions of neurons throughout early development, local circuits and complex regional interconnections are established by secreted guidance and survival cues. Neurogenesis slows in the maturing postnatantal nervous system and non-neurogenic cell populations ultimately predominate the adult brain and spinal cord. This limits the physiological capacity for functional regeneration of adult neural circuits affected by aging-related neurodegenerative disorders such as Alzheimer disease, Huntington disease, and Parkinson disease. Therefore, a central goal of reprogramming-based regenerative therapies is the production of integration-competent functional neurons driven by the principles of neural development and multipotency.

1.1.2. Cellular Potency and Neural Differentiation

Potency reflects the lineage-differentiation potential of stem and progenitor cells. Pioneering nuclear transplantation studies used *Rana pipens* oocytes and nuclei from advanced blastula cells

to demonstrate that cell nuclei regulate differentiation potential (Briggs et al. 1952). Reinforcing this concept, nuclei transplanted from *Xenopus laevis* epithelial cells into enucleated oocytes yielded viable embryos that developed into mature frogs (Gurdon et al. 1970). This remarkable discovery confirmed that nuclei from differentiated somatic cells retain pluripotent potential and suggested reversible genetic regulation underpins cell lineage identity (Smith et al. 2016).

The first *in vitro* experimental model of cellular pluripotency relied on mouse teratocarcinoma cells with properties reminiscent of stem cells in the early mouse embryo (Kleinsmith et al 1964, Stevens 1967). While these multipotential cells offered an unprecedented opportunity to investigate the mechanisms regulating cell identity, these cell lines exhibited limited differentiation potential relative to stem cells directly derived from totipotent embryos. The subsequent isolation of embryonic stem cells (ESCs) from proliferating mouse (Evans et al. 1981), primate (Thomson et al. 1995), and human (Thomson et al. 1998) blastocysts established a new paradigm in pluripotency and cell identity research. With the capacity to generate diverse cell types of all three germ layers (Figure 1), these pluripotent cells enabled targeted investigations into the molecular mechanisms of neural differentiation.

Pro-neural effector molecules bias mouse and human pluripotent stem cells toward a neural identity *in vitro*. For instance, FGF2 activates a nestin-positive progenitor state differentiable to neuron and glial fates (Okabe et al. 1996), while retinoic acid promotes GABAergic and cholinergic identities (Bain et al. 1995, Fraichard et al. 1995). *In vivo* transplantation of stem cell-derived neural progenitors has produced subtype-specific dopaminergic neurons (Deacon et al. 1998), serotonergic neurons (Deacon et al. 1998), spinal motor neurons (Wichterle et al. 2002), and striatal medium spiny neurons (Aubry et al. 2008). While an advantageous model, over proliferation and ethical debate restrict the therapeutic value of stem cell-derived neurons.

1.1.3 Progenitor-Based Engineered Reprogramming

The viral overexpression of KLF4, MYC, POU5F1, and SOX2 in embryonic and adult mouse fibroblasts catalyzed an *in vitro* reversion to pluripotent identity that redefined the concept of terminal differentiation (Takahashi et al. 2006). Induced pluripotent stem cells (iPSC) were similarly generated from fetal and adult human fibroblasts using these factors (Park et al. 2008). Further, cultured fibroblasts exposed to a cocktail of small molecules acquired pluripotent features highlighting the integral role for intracellular signaling in cell identity maintenance (Hou et al. 2013). While patient-specific fibroblast-derived iPSCs offer a potentially novel platform for regenerative therapeutics, a high propensity for teratoma formation inherently limits the *in vivo* utility of these cells (Abad et al. 2013).

Cardiac, neural, pancreatic, and lineage-specific progenitor cells differentiated from iPSCs have provided fundamental insights into lineage-specific transcription programs (Figure 1). For instance, temporal restriction of POU5F1 expression to the first five days of iPSC reprogramming induced fibroblasts to adopt a neural progenitor-like identity with properties similar to tripotent endogenous neural stem cells (Their et al. 2012). Further, the transplantation of iPSC-derived neural progenitors into the mouse brain (Wernig et al. 2008) and non-human primate spinal cord (Emborg et al. 2013) demonstrated engraftment and neuronal differentiation without tumor formation. While the tumorigenicity of pluripotent stem cells hinders therapeutic application, these results suggest that engineered lineage-committed progenitor cells might be effective clinical tools.

The direct *in vitro* reprogramming of fibroblasts into tripotent neural progenitors without passing through a pluripotent intermediate was achieved using viral expression of transcription factors (Han et al. 2012, Lujan et al. 2012) or small molecules in a hypoxic environment (Cheng

et al. 2014). SOX2 promotes self-renewing neural stem cell identity in cultured human fibroblasts and yields neurospheres that differentiate into mature functional neurons following cortical transplantation in the mouse (Ring et al. 2012). Moreover, SOX2 induces *in vivo* reprogramming of mouse brain and spinal cord astrocytes into neuroblasts poised to differentiate into functional neurons upon valproic acid exposure (Niu et al. 2013, Su et al. 2014). Acquired pluripotency and multipotency experiments established a direct link between cell identity and transcriptional regulation suggesting lineage-specific differentiation factors might also catalyze cell identity reprogramming.

1.1.4. Transdifferentiation

In a milestone discovery, MYOD1 overexpression induced the transdifferentiation of fibroblasts into myogenic cells (Tapscott et al. 1988). This reprogramming event underscores the pivotal role for direct regulators of lineage-specific transcription in the specification of cell identity. Expanding this concept, the pioneer activity of ASCL1 together with MYT1L and POU3F2 catalyzed mouse embryonic and postnatal fibroblasts to transdifferentiate into functional neurons *in vitro* (Vierbuchen et al. 2010, Wapinski et al. 2013). The addition of NEUROD1 to this reprogramming cocktail facilitated the neuronal conversion of human fibroblasts (Pang et al. 2011).

Further reinforcing the notion that targeted changes in transcription can directly transform cell identity, neurons were generated from human adult fibroblasts using pools of pro-neural transcription factors and microRNAs (Ambasudhan et al. 2011, Yoo et al. 2011), as well as, transcriptome-modifying chemical cocktails (Hu et al. 2015, Zhang et al. 2015). Diverse protocols for the generation of subtype-specific cholinergic, dopaminergic, GABAergic,

glutamatergic, and motor neurons quickly followed (Table 1). The *in vivo* transdifferentiation of glia to functional neurons (Torper et al. 2013), cardiac fibroblasts to cardiomyocytes (Ieda et al. 2012, Song et al. 2012), and pancreatic exocrine cells to insulin-producing β-cells (Zhou et al. 2008) using exogenous transcription factors broadly highlights the transformative effects of modified transcription on cellular physiology (Figure 1).

1.2. Engineered Neurons

1.2.1. Lineage-Specific Neuron Identity

The functional complexity of neural circuits in the brain relies on diverse neuron lineages with characteristic regional and neurotransmitter properties. The region-specific degeneration of these dopaminergic, GABAergic, glutamatergic, and motor neuron populations underlies prevalent neurodegenerative disorders such as Parkinson disease, Huntington disease, Alzheimer disease, and amyotrophic lateral sclerosis (ALS). Therefore, targeted genetic engineering strategies are being explored in an effort to generate functional lineage-specific neurons for *in vitro* disease modeling and *in vivo* transplantation.

The high-efficiency *in vitro* transdifferentiation of human fetal lung fibroblasts into functional cholinergic neurons was demonstrated using NEUROG2 and the small molecules forskolin and dorsomorphin (Liu et al. 2013). The efficient reprogramming of human adult skin fibroblasts required these factors, FGF2, and SOX11 (Liu et al. 2013). Broadening this core set of factors to include ISL1 and LHX3 enabled transdifferentiation to a functional motor neuron identity and provided a platform for screening pro-survival compounds in ALS models using neurons derived from the fibroblasts of ALS-afflicted individuals (Liu et al. 2016). Engineered motor neurons transdifferentiated from mouse fibroblasts (Son et al. 2011) and fibroblast-derived

iPSCs (Meyer et al. 2013, Kiskinis et al. 2014, Yoshida et al. 2015) have enabled molecular investigations into the cause of motor neuron degeneration and the screening-based identification of pro-survival small molecules.

The degeneration of striatal dopaminergic neurons is the primary underlying cause of Parkinson disease symptoms. The transplantation of induced dopaminergic neurons generated by iPSC differentiation (Wernig et al. 2008) and fibroblast transdifferentiation (Kim et al. 2011) modestly improved mouse motor function. In addition to *in vitro* transdifferentiation and electrophysiological analyses of fibroblast-derived dopaminergic neurons (Caiazzo et al. 2011, Pfisterer et al. 2011, Liu et al. 2012, Jiang et al. 2015), mouse astrocytes transduced with ASCL1, LMX1B, and NR4A2 have been shown to directly adopt dopaminergic identity (Addis et al. 2011). Further, human fibroblasts carrying inducible pro-dopaminergic reprogramming factors were transplanted into the mouse neural microenviroment and *in vivo* differentiated into synapse-forming tyrosine hydroxylase-positive neurons (Torper et al. 2013). In an effort to move away from viral integration as a delivery system for exogenous reprogramming factors, CRE recombinase-based excision of exogenous reprogramming factors was performed following dopaminergic differentiation of iPSC lines (Soldner et al. 2009).

In addition to cholinergic and dopaminergic identities, engineered GABAergic and glutamatergic neurons have been generated by both iPSC differentiation (Brennand et al. 2011) and transdifferentiation (Vierbuchen et al. 2010). The *in vitro* transdifferentiation of human fibroblasts (Victor et al. 2014, Colasante et al. 2015), mouse astrocytes (Heinrich et al. 2010), and mouse NG2 glia (Guo et al. 2014) into GABAergic neurons demonstrated that dominant pro-neural transcription factors can reprogram diverse cell types. *In vivo*, NEUROD1 drives heterogeneous GABAergic and glutamatergic identities in NG2 glia (Guo et al. 2014). In

contrast, NEUROG2 and NEUROD1 reprogram mouse striatal glia and cortical astrocytes into functional glutamatergic neurons, respectively (Grande et al. 2013, Guo et al. 2014). This suggests that the cell-of-origin genetic landscape primes adoption of a specific identity post-reprogramming; for instance, NEUROG2-transduced mouse cortical astrocytes and NEUROD1-transduced human cortical astrocytes adopted exclusively glutamatergic identity (Heinrich et al. 2010, Guo et al. 2014). Moreover, human cortical astrocytes treated with a pro-neural small molecule cocktail *in vitro* transdifferentiated into functional glutamatergic neurons (Zhang et al. 2015).

1.2.2. Genetic Properties of Engineered Neurons

The genetic and epigenetic maturation of reprogrammed neurons requires the acquisition of RNA and protein expression patterns, chromatin condensation patterns, epigenetic signatures, and cytoskeletal architecture resembling endogenous neurons. To determine whether the gene regulatory networks and transcription in reprogrammed neurons parallels primary neurons of the developed nervous system, the bioinformatics algorithm CellNet was used for genome-scale comparison of ESC, fibroblast, primary neuron, ESC-derived neuron, and transdifferentiated fibroblast-derived neuron transcription profiles (Cahan et al. 2014). Highly similar profiles were observed for primary and ESC-derived neurons; however, transdifferentiated dopaminergic neurons shared meaningful overlap with fibroblast genetic regulatory networks interspersed with the activation of general neural transcription (Cahan et al. 2014, Caiazzo et al. 2011). This hybrid genetic background was previously detected by hierarchical clustering of microarray datasets indicating transdifferentiated dopaminergic neurons resemble midbrain dopaminergic neurons but do not faithfully recapitulate the global genetic background of primary neurons (Caiazzo et

al. 2011). The targeted activation of pro-neural transcription in reprogrammed dopaminergic neurons has been demonstrated by quantitative real-time PCR enrichment of *Foxa2*, *Lmx1a*, *Pitx3*, *Slc18a2*, and similar lineage-specific factors (Addis et al. 2011, Kim et al. 2011). A similar analysis of dopaminergic neurons induced by either two or six transcription factors revealed that immature dopaminergic identity is evoked by ASCL1 and PITX3 overexpression; however, transdifferentiation toward a state representative of physiological identity required four additional transcription factors.

While some reprogrammed neurons exhibit a heterogeneous genetic background, others adopt lineage-specific identities with remarkable fidelity. Fibroblast-derived GABAergic medium spiny neurons exhibit expression profiles strongly correlated to primary human striatal neurons isolated by laser capture microdissection from post-mortem neural tissue (Victor et al. 2014). Unlike engineered cortical GABAergic neurons characterized by RNA-seq (Colasante et al. 2015), induced medium spiny neurons adopted a striatal-specific identity unique from other GABAergic interneuron lineages (Victor et al. 2014).

In an effort to define how individual reprogramming factors contribute to neural specification, chromatin immunoprecipitation sequencing (ChIP-seq) was used to identify the genome-wide targets of ASCL1 and POU3F2 in neural precursor cells and reprogrammed neurons (Wapinski et al. 2013). Interestingly, in both neural precursors and neurons ASCL1 genetic targets directly overlapped, while POU3F2 binding was primarily localized to non-physiological ASCL1 occupancy sites in reprogrammed fibroblasts (Wapinski et al. 2013). This prominent difference suggests the epigenetic environment of embryonic mouse fibroblasts disrupts the transcriptional activity of non-pioneer transcription factors and supports the concept that active pioneer factors such as ASCL1 are required to jumpstart pro-neural transcription.

In addition to transcriptional profiling, the characterization of engineered neurons relies on immunocytochemical reactivity and measurements of electrical function. General hallmarks of neuron identity such as MAP2, MAPT, RBFOX3, SYN1, and TUBB3 have been extensively used to identify reprogramming and quantify efficiency (Vierbuchen et al. 2011, Hu et al. 2015). Excitatory differentiation toward cortical glutamatergic identity (Guo et al. 2014), deep layer forebrain and hippocampal identity (Zhang et al. 2015), and cholinergic motor neurons (Son et al. 2011, Liu et al. 2016) was detected by immunostaining for BCL11B, MNX1, SLC17A7, and other lineage-determinative markers. Accompanying these genetic hallmarks, neurons derived from the fibroblasts of individuals afflicted with Alzheimer disease exhibit aberrations in Aβ peptide production that might have long-term deleterious physiological consequences reminiscent of progressive neurodegeneration (Hu et al. 2015). Together, these findings demonstrate that specific transcriptional networks control cell identity and targeted modifications to these networks can promote a functional neuronal identity similar to primary neurons.

1.2.3. Functional Properties of Engineered Neurons

Neurons exhibit specialized membrane properties that enable the rapid transmission of ionic current in response to stimuli. While fibroblasts and astrocytes lack the membrane structures required to establish the ion gradients essential to this neurotransmission, exposure to pro-neural reprogramming factors promotes the expression of membrane channels, kinases, and neurotransmitter release machinery essential to neurotransmission. Neurons generated from mouse and human fibroblasts acquire the ability to fire repetitive action potentials, exhibit fast activating and inactivating sodium ion currents, respond to extracellular GABA and glutamate, and exhibit spontaneous postsynaptic currents (Vierbuchen et al. 2011, Pang et al. 2011, Hu et al.

2015). While NEUROG2 and DLX2 both reprogram astrocytes toward neuronal morphology *in vitro*, NEUROG2 is a significantly more potent driver of functional identity with 58% of induced neurons firing action potentials relative to 6% of DLX2-induced neurons (Heinrich et al. 2010). This functional characterization demonstrates how partial reprogramming via upregulated neural transcription is not an indicator of physiological fidelity.

In addition to the basic membrane potential properties acquired during reprogramming, specialized neuron lineages synthesize specific neurotransmitters and exhibit brain-region specific functional activity. Reverse-phase high performance liquid chromatography confirmed the production of dopamine by induced dopaminergic neurons (Kim et al. 2011). Moreover, *in vitro* reprogramed dopaminergic neurons exhibited calcium oscillations and pacemaker activity representative of primary midbrain dopaminergic neurons (Addis et al. 2011). Similar functional analyses demonstrated that induced mouse and human cholinergic motor neurons formed neuromuscular junctions with co-cultured C2C12 myotubes and elicited rhythmic contraction in these cellular structures (Son et al. 2011, Liu et al. 2016).

The transplantation and engraftment of GABAergic medium spiny neurons in the mouse striatum enabled direct functional comparisons of endogenous primary neurons and engineered cells (Victor et al. 2014). Remarkably, transplanted neurons survived in the striatal microenvironment and exhibited functional properties highly similar to endogenous local neurons (Victor et al. 2014). Likewise, transplantation of iPSC-derived neural progenitor cells into the embryonic mouse brain confirmed the neural environment is amenable to *in vivo* differentiation of functional neurons (Wernig et al. 2008).

In vivo astrocyte-to-neuron transdifferentiation by NEUROD1 yields functional integration-competent neurons that fire repetitive actions and evoke spontaneous synaptic events during

cortical slice recordings (Guo et al. 2014). The detailed functional characterization of *in vivo* reprogrammed neurons by SOX2 and valproic acid revealed four predominant types of neurons responsive to neurotransmitter stimulation and capable of firing repetitive action potentials (Niu et al. 2015). These neurons exhibited diverse resting membrane potentials, action potential firing rates, and broad spontaneous postsynaptic current frequencies suggesting individual astroglia respond differently to exogenous reprogramming factors (Niu et al. 2015).

1.2.4. Network Integration of Engineered Neurons and Applications

The therapeutic value of cell identity engineering relies on the capacity of transplanted or *in vivo* generated neurons to survive and functionally integrate into endogenous neural networks. Immunogold labeling followed by transmission electron microscopy confirmed that primary hippocampal neurons form direct synapses with GABAergic reprogrammed neurons *in vitro* (Colasante et al. 2015). When transplanted into the hippocampus, these GABAergic neurons integrated into hippocampal neuronal circuits (Colasante et al. 2015). Similarly, neurons generated by chemical-mediated reprogramming survive transplantation and effectively integrate into neuronal circuits (Zhang et al. 2015). Cortical and striatal neurons generated *in vivo* outgrow dendritic branches and axons that form synaptic connections with endogenous interneurons (Guo et al. 2014, Niu et al. 2013). In the developing spinal cord, induced motor neurons grafted into the ventral horn recognized endogenous path finding cues and extended axons toward musculature (Song et al. 2011).

With respect to models of neurodegeneration, the transplantation of induced dopaminergic neurons into 6-hydroxydopamine lesioned rat brains reduced amphetamine-induced rotational bias indicating that transplanted neurons can directly improve performance in motor coordination

tests (Wernig et al. 2008). Moreover, fibroblast-derived dopaminergic neurons transplanted into a mouse model of Parkinson disease improved basic motor function suggesting that reprogrammed neurons might effectively remodel degenerated neural circuits in a functionally meaningful manner (Kim et al. 2011).

1.3. Genetic and Molecular Mechanisms of Neuronal Transdifferentiation

The transdifferentiation of a somatic or glial cell into a specialized neuron requires large-scale modifications to the transcriptome, proteome, DNA methylome, histone landscape, chromatin structure, cytoskeleton, and membrane polarization properties. Remarkably, small pools of transcription factors and pro-neural chemicals are sufficient to rapidly catalyze these broad molecular changes to induce transdifferentiation.

1.3.1. Chemical-Mediated Transdifferentiation

Human cortical and midbrain astrocytes adopt glutamatergic deep layer forebrain and hippocampal neuron identity *in vitro* when sequentially exposed to combinations of signaling molecules (Zhang et al. 2015). Initial exposure to LDN193189, SB431542, thiazovivin, and TTNPB primes cells for conversion. Neuronal differentiation occurs upon exposure to CHIR99021, DAPT, purmorphamine, smoothened agonist, and valproic acid with maturation enhanced by BDNF, IGF1, and NT-3 (Zhang et al. 2015). However, human spinal astrocytes and mouse astrocytes treated with this same chemical reprogramming protocol do not adopt neuronal identity highlighting the dynamic roles for intracellular signaling pathways in the maintenance of unique regional cellular identities (Zhang et al. 2015).

The transdifferentiation of human fibroblasts from healthy and Alzheimer disease-afflicted individuals into glutamatergic neurons was also achieved using sequential chemical exposures (Hu et al. 2015). Initial treatment with CHIR99021, forskolin, GO6983, RepSox, SP600125, valproic acid, and Y-27632 induces morphological conversion and MAP2 expression (Hu et al. 2015). Survival and functional maturation are enhanced following mid-conversion exposure to BDNF, CHIR99021, dorsomorphin, forskolin, GDNF, and NT-3 (Hu et al. 2015). Similar work demonstrated the induction of heterogeneous GABAergic and glutamatergic neurons from mouse fibroblast treated with CHIR99021, forskolin, I-BET151, and isoxazole 9 (Li et al. 2015). These results demonstrate the potent roles for signaling pathways in the specification and maintenance of cellular identity.

1.3.2. Transcription Factor-Mediated Transdifferentiation

1.3.2.1. Pioneer Transcription Factors

The potency of transcription factors used to reprogram cell identity depends on accessibility to genetic regulatory regions and an ability to remodel occluded target sites. Lineage-restrictive pioneer transcription factors passively bind closed heterochromatin and a subset of these factors initiate local remodeling to promote the recruitment of cooperative transcriptional complexes (Zaret et al. 2011). For instance, NEUROD1 displaces transcriptional repressors and converts Polycomb-associated histone methylation to activating acetylation marks (Pataskar et al. 2015). This pro-neural remodeling activity likely contributes to the rapid conversion of astrocytes into neuron-like cells upon viral overexpression of NEUROD1 (Guo et al. 2014). ASCL1 similarly pioneers high-density nucleosome conformations in pro-neural enhancer and promoter sequences to drive the *in vitro* transdifferentiation of embryonic mouse fibroblasts into functional neurons

(Wapinski et al. 2013). In constrast, the non-pioneer factor POU3F2 generally occupies nucleosome-depleted regions; however, ASCL1 recruits this factor to non-physiological binding sites when co-expressed in fibroblasts (Wapinski et al. 2013). This demonstrates the dominant role for ASCL1 pioneer activity in resetting the cellular transcriptome during transdifferentiation. Active pioneer factors such as the NF-Y family also exhibit nucleosome repositioning and cooperative binding properties during developmental pluripotency and neuron differentiation (Oldfield et al. 2014).

1.3.2.2. Secondary Transcriptional Programs and Gene Regulatory Networks

Pioneer factors are often high-level lineage-specific transcriptional regulators that drive secondary networks of hierarchical transcription. Exogenous ASCL1 upregulates the transcription factors *Lmo2*, *Rfx1*, *Tcfl5*, and *Zbtb18* during fibroblast-to-neuron reprogramming (Wapinski et al. 2013). The co-overexpression of ZBTB18 and MYT1L in mouse fibroblasts induces a functionally immature MAP2-positive neuron-like identity (Wapinski et al. 2013). The inclusion of LMO2, RFX1 or TCFL5 with these two factors significantly improves reprogramming efficiency demonstrating that the parallel action of gene regulatory networks activated by ASCL1 drive neuronal reprogramming (Wapinski et al. 2013).

Repurposing the differentiation mechanisms that sculpt the early nervous system, chemicals-mediated astrocyte-to-neuron transdifferentiation utilizes intracellular signaling pathways to downregulate hallmarks of astrocyte identity such as *Gfap* and upregulate the neuronal-lineage transcription factors *Neurod1* and *Neurog2* (Zhang et al. 2015). NEUROG2 competes with the transcriptional repressor REST for occupancy of the *Neurod4* promoter during reprogramming (Masserdotti et al. 2015). NEUROD4 and INSM1, both induced by NEUROG2, catalyze

reprogramming in mouse astrocytes, human astrocytes, and mouse embryonic fibroblasts (Masserdotti et al. 2015). The dominant roles for single transcription factors and their associated gene regulatory networks in cell identity re-specification led to the creation of a computational algorithm that identifies master transcription factors with a high transdifferentiation probability (Rackham et al. 2016). This algorithm has successfully predicted transdifferentiation factors for cardiomyocyte, hepatocyte, keratinocyte, macrophage, and other cell lineages (Rackham et al. 2016). Similar programs will likely become fundamental tools for neuron subtype engineering.

1.3.2.3. Lineage-Specific Transcriptional Programs

Lineage-specific transcription factors define three germ layers in the developing embryo. These master regulators activate hierarchical networks of transcription in ectodermal cells to promote glial and lineage-specific neuronal differentiation (Molyneaux et al. 2007, Schuurmans et al. 2004). Interestingly, the neuron identity induced by ASCL1, NEUROD1, and NEUROG2 during transdifferentiation is cell-of-origin dependent. NEUROG2 drives glutamatergic identity in mouse astrocytes (Heinrich et al. 2010), but cholinergic identity in human fibroblasts (Liu et al. 2013). While ASCL1 drives astrocytes toward GABAergic identity (Heinrich et al. 2010), embryonic mouse fibroblasts adopt a primarily glutamatergic identity (Wapinski et al. 2013). Further, exogenous NEUROD1 converts astrocytes into glutamatergic neurons and NG2 glia into both GABAergic and glutamatergic neurons (Guo et al. 2014).

ASCL1 and NEUROG2 activate distinct transcriptional programs in astrocyte-derived neurons with only 3% overlap in target gene expression (Masserdotti et al. 2015). The inclusion of coordinating transcription factors with either ASCL1 or NEUROG2 enables the generation of brain region-specific neural subtypes. For instance, ASCL1, GBX1, and LHX8 induce

embryonic stem cells and human fibroblasts to adopt basal forebrain cholinergic neuron identity (Bissonnette et al. 2011, Liu unpublished data). In contrast, NEUROG2 combined with ISL1, LHX3, SOX11, growth factors, and small molecules convert adult skin fibroblasts into functional motor neuron identity (Liu et al. 2016, Mazzoni et al. 2013). Further, the differentiation of specialized spinal and cranial motor neurons is regulated by the expression of either LHX3 or PHOX2A, respectively (Mazzoni et al. 2013). These results highlight how established epigenetic landscapes, lineage-restrictive transcription factors, and coordinated genetic programs converge to define cell identity during transdifferentiation.

1.3.2.4. NEUROG2 Transcriptional Program

NEUROG2 is one of three conserved Neurogenin transcription factors with diverse roles in neural development and neuron differentiation (Gradwohl et al. 1996, Ma et al. 1996, Sommer et al. 1996). As a basic helix-loop-helix (bHLH) transcription factor, NEUROG2 contains a basic DNA binding domain immediately followed by two alpha helices that promote bHLH heterodimerization and association with 5'-CANNTG-3' enhancer box DNA sequences (Murre et al. 1989). Phylogenetic analyses have categorized bHLH proteins into six high-order clades collectively containing 44 orthologous families (Ledent et al. 2002). NEUROG2 is classified as a member of the Atonal superfamily and is most closely related to the NEUROD family of pro-neural transcription factors (Ledent et al. 2002).

The developing mouse telencephalon is patterned by dorsal NEUROG2 and ventral ASCL1 expression (Fode et al. 2000, Parras et al. 2002, Britz et al. 2006). Neural cells of both the ventricular and subventricular zones of the developing dorsal telencephalon transcribe *Neurog2* mRNA in high quantities (Fode et al. 2000, Schuurmans et al. 2004, Osório et al. 2010). These

progenitor cells primarily differentiate into excitatory glutamatergic pyramidal neurons with unipolar dendritic morphology (Hand et al. 2005). In addition to deep-layer cortical neurons (Schuurmans et al. 2004), NEUROG2 instructs the differentiation of spinal motor neurons (Mizuguchi et al. 2001, Ma et al. 2008) and cranial sensory neurons in the peripheral nervous system (Fode et al. 2000).

The transcriptional activity of NEUROG2 is tightly regulated by post-translational phosphorylation and co-factor interactions. Cell cycle-related cyclin dependent kinases directly target NEUROG2 for multi-site phosphorylation to negatively regulate the DNA affinity of NEUROG2 heterodimers (Ali et al. 2011, Hindley et al. 2012). Site-specific phosphorylation of tyrosine 241 similarly modulates NEUROG2-chromatin interactions to promote radial migration of cortical neurons in the developing brain (Hand et al. 2005). Additional phosphorylation by GSK3 represses NEUROG2 transcription during late corticogenesis in response to declining Wnt signaling (Li et al. 2012). GSK3-mediated phosphorylation of serines 231 and 234 promotes NEUROG2 association with the LIM homeodomain transcription factors ISL1 and LHX3 (Ma et al. 2008). These phosphorylation-dependent co-factor interactions are essential to the specification of motor neurons in the mouse spinal cord (Lee et al. 2003, Ma et al. 2008). Moreover, the retinoic acid receptor recruits histone acetyltransferase to NEUROG2 target sites in spinal motor neurons (Lee et al. 2009) and LMO4 acts as a co-activator in the developing cortex (Asprer et al. 2011). Computational identification of adjacent binding motifs at known NEUROG2 genomic target sites indicates NEUROG2 might directly associate with CREB1, MEF2A, POU-domain containing transcription factors, TCF/LEF transcription factors, and SOX9 (Gohlke et al. 2008).

Dimerization is a prerequisite to NEUROG2 function. Importantly, NEUROG2 homodimers do not efficiently bind DNA (Bronicki et al. 2012) and, while NEUROG2 has been shown to heterodimerize with ASCL1 and TCF3, NEUROG2-TCF3 heterodimers are much stronger transcriptional activators than NEUROG2-ASCL1 heterodimers (Gradwohl et al. 1996, Henke et al. 2009, Bronicki et al. 2012). In addition to protein-protein interaction, NEUROG2 indirectly regulates ASCL1 expression (Kovach et al. 2013). NEUROG2 also regulates the Notch ligands *Dll1* (Fode et al. 1998) and *Dll3* (Henke et al. 2009), the neurogenic transcription factors *Neurod1* and *Neurod4* (Seo et al. 2007), the neuronal markers *Dcx*, *Ncam*, and *Sema3c* (Mattar et al. 2004), and lineage-specific factors *Rnd2*, *Slc17A6*, *Slc17A7*, *Tbr1*, *Tbr2*, and *Vglut2* (Li et al. 2012, Schuurmans et al. 2004). Interestingly, mouse knockout and overexpression studies revealed that the reciprocal regulation of PAX6 and NEUROG2 expression guides early motor neuron differentiation (Scardigli et al. 2001).

1.3.2.5. SOX4 Transcriptional Program

The identification of sex-determining region Y marked the discovery of the first of at least 20 SOX transcription factors containing a high mobility group (HMG) DNA binding domain (Gubbay et al. 1990, Sinclair et al. 1990). Homology-based classification distributes SOX factors into six classes with SOXC representing SOX4, SOX11, and SOX12 (Schepers et al. 2002). The SOX family of transcription factors associates with the degenerate DNA sequence 5'-(A/T)(A/T)CAA(A/T)G-3' (Harley et al. 1994) and SOX4 specifically targets the central DNA heptamer 5'-AACAAAG-3' (van Beest et al. 2000). Nucleotide contacts in the minor groove of the DNA helix enable SOX4 to induce a sharp bend in conformation and enhance accessibility to DNA-binding co-factors (van de Wetering et al. 1992, van Houte et al. 1995).

Sox4 is broadly expressed in the mouse heart, lungs, pancreas, thymus, and developing central nervous system (van de Wetering et al. 1993, Cheung et al. 2000). SOX4 is transcriptionally active in cells of the mouse uterus and mammary glands (Hunt et al. 1999), adult B and pre-T lymphocytes (van de Wetering et al. 1993), pancreatic islet cells (Lioubinski et al. 2003), retinal ganglion cells (Jiang et al. 2013), and mechanosensory hair cells of the inner ear (Gnedeva et al. 2015). Loss of *Sox4* expression in the developing endocardial ridge results in cardiac defects and malformation of the semilunar valves (Schilham et al. 1996). In addition to these diverse developmental roles, *Sox4* has been implicated in the survival and proliferation of human breast cancer (Graham et al. 1999), prostate cancer (Liu et al. 2006), bladder cancer (Aaboe et al. 2006), and multiple other cancer subtypes (Vervoort et al. 2013).

In the developing nervous system, *Sox4* and *Sox11* are expressed in similar patterns during neurulation (Cheung et al. 2000). While independent knockout of either factor does not significantly affect neural development, double knockout results in severe hippocampal deformation, disrupted cortical lamination, and reduced neuron survival (Cheung et al. 2000, Miller et al. 2013). In post-mitotic neural precursors, NEUROG2 outcompetes REST in the *Sox4* promoter to drive *Sox4* expression and neuronal differentiation (Bergsland et al. 2006). As the central nervous system matures, *Sox4* becomes primarily localized to the dorsal region of the neural tube with strong expression in the cortical plate (Cheung et al. 2000). This expression pattern suggests SOX4 is an important lineage-specification factor for cortical neurons and late pyramidal neurons of the hippocampus (Cheung et al. 2000). *Sox4* expression is downregulated as neurons mature and prolonged overexpression results in cerebellar defects (Hoser et al. 2007) and disrupted neuronal myelination (Potzner et al. 2007). Importantly, SOX4 functions as a

pro-survival factor in late-stage sympathetic ganglia (Potzner et al. 2010) and mouse spinal motor neurons (Thein et al. 2010).

SOXC factors often function as transcriptional activators in complex with other proteins. While these factors share 94% homology in the C-terminal transactivation domain, domain swap experiments demonstrated that SOX4 exhibits reduced transcriptional activation strength relative to SOX11, but much higher DNA affinity (Dy et al. 2008). Further, this domain is critical to association with neuronal POU3F2 (Dy et al. 2008) and POU3F3 in oligodendrocyte-lineage cells (Kuhlbrodt et al. 1998). SOX4 also directly interacts with β-catenin to promote Wnt signaling (Sinner et al. 2007), NEUROG2 (Chen et al. 2015), SDCBP (Geijsen et al. 2001), TBR2 (Chen et al. 2015), and TCF4 (Sinner et al. 2007).

Coordinated binding of SOX4 and POU3F2 upregulate *Nes* expression in neural primordial cells (Tanaka et al. 2004). SOX4 similarly drives pro-neural expression of *Tubb3* (Bergsland et al. 2006) and targets *Tead2* of the HIPPO signaling pathway (Bhattaram et al. 2010). When overexpressed with NEUROG2, SOX4 efficiently reprograms cultured mouse astrocytes into TUBB3-positive neurons with complex neurite outgrowth (Mu et al. 2012). This transdifferentiation activity highlights how SOX4 enhances pro-neural transcription in the development and specification of early neurons of the central nervous system.

1.3.3. Enzymatic Activity, Intracellular Signaling, and Metabolic Pathways

Cell identity reprogramming involves broad changes in metabolism, kinase activity, cell cycle regulation, and intracellular signaling. The rapid changes in transcription induced by reprogramming factors often dramatically increase energy expenditure and generate metabolic byproducts that modify DNA methylation and histone marks (Ryall et al. 2015). Prolonged

exposure to high concentrations of reactive metabolites can become toxic during reprogramming. Neuron induction efficiency and survival is significantly improved with ferroptosis and lipid peroxidation inhibitors, as well as, overexpression of the anti-apoptotic protein BCL2 (Gascón et al. 2016). The small molecule forskolin boosts reprogramming efficiency and neuron survival by considerably reducing lipid peroxidation in astrocytes, fibroblasts, and pericytes during transdifferentiation (Gascón et al. 2016, Hu et al. 2015, Li et al. 2015, Liu et al. 2013). Further, G1 phase cell cycle arrest significantly increases the efficiency of TET1-dependent fibroblast reprogramming and demonstrates how cell cycle kinetics and metabolism potently regulate changes in cellular identity (Jiang et al. 2015).

The mechanistic roles for intracellular signaling and active kinases during neuronal specification have been demonstrated through chemical-mediated reprogramming strategies (Hu et al. 2015, Li et al. 2015, Zhang et al. 2015). For instance, the GSK3 inhibitor CHIR99021 promotes neural induction, while thiazovivin and Y-27632 promote cell survival by inhibiting Rho-associated kinase (Hu et al. 2015, Li et al. 2015, Zhang et al. 2015). GO6983 and SP600125 promote neuronal conversion by inhibiting of PKC and JNKs, respectively (Hu et al. 2015). While dorsomorphin reduces the kinase activity of AMPK, this small molecule and LDN193189 also inhibit BMP type I receptor signaling to repress the activation of glial and somatic lineage identities (Hu et al. 2015, Liu et al. 2013).

BMP, Notch, SHH, and TGFB1 signaling pathways each exert significant influence over transcription and cellular function. DAPT, an indirect inhibitor of Notch, in combination with the SHH activators purmorphamine and smoothed agonist promote neuronal differentiation of cultured astrocytes (Zhang et al. 2015). Inhibition of the pro-glial activin receptor and TFGB1 signal transduction cascade with RepSox or SB431542 enhances the adoption of neuronal

identity during transdifferentiation (Hu et al. 2015, Zhang et al. 2015). Exposure to neurotrophic growth factors like BDNF, FGF2, GDNF, and NT-3 during the maturation stage of neuronal differentiation significantly improves the functional properties and survival of transdifferentiated neurons (Jiao et al. 2013, Liu et al. 2013, Hu et al. 2015). While pro-neural transcription factors and miRNAs often jumpstart reprogramming, it is these signal transduction pathways and neurotrophic survival cues that drive neurogenic identity to completion.

1.3.4. Epigenetic Signatures

The epigenome of DNA methylation, histone marks, and chromatin condensation directly influence the stability of cellular identity. The histone deacetylase inhibitor valproic acid has been shown to induce neuroblast differentiation (Niu et al. 2013) and enhance human astrocyte and fibroblast transdifferentiation efficiency (Hu et al. 2015, Zhang et al. 2015). Mechanistically, histone acetylation and methylation patterns modulate transcription factor affinity for specific chromatin sites to indirectly activate or repress the transcription of lineage-specific genes.

ASCL1 preferentially occupies chromatin sites containing high levels of H3K4 mono-methylation and H3K27 acetylation, as well as, low-to-intermediate enrichment of H3K9 tri-methylation (Wapinski et al. 2013). Similarly, MYT1L and POU3F2 primarily associate with methylation and acetylation markers of active transcription. This dynamic interplay suggests histone posttranslational modifications act as an atlas that directs pioneer and non-pioneer transcription factors to specific genomic regulatory regions. During reprogramming, chromatin modifiers activated by signal transduction pathways and recruited by pro-neural transcription factors re-write this atlas. For instance, H3K27 acetylation was strongly enriched in the *Neurog2*

promoter during the first eight days of astrocyte-to-neuron reprogramming and H3K4 tri-methylation replaced H3K27 tri-methylation at the transcription start site (Zhang et al. 2015).

Coordinated histone mark and DNA methylation profiles stabilize cell identity and provide a multi-layered regulatory system to prevent aberrant transcription. The differentiation of ESC-derived neural precursor cells into neurons requires parallel changes in Polycomb group-mediated histone tri-methylation and DNA methylation (Mohn et al. 2008). In the process of astrocyte conversion, chemical reprogramming catalyzes broad methylation of the *Gfap* promoter and targeted hypermethylation at AP1 and STAT3 binding sites to significantly downregulate *Gfap* expression (Zhang et al. 2015). Similarly, fibroblast transdifferentiation is dependent on the global 5-methylcytosine to 5-hydroxymethylcytosine DNA demethylation activity of TET1 (Jiang et al. 2015). Collectively, these dynamic changes in DNA methylation combined with multi-site histone modifications remodel the chromatin landscape and enable mediators of pro-neural transcription to ultimately redefine cell identity.

1.3.5. Hybrid Genetic Background and Permanence of Cellular Identity

Differentiation collectively involves millions of changes in the proteome, kinome, transcriptome, DNA methylome, histone landscape, chromatin architecture, and even extracellular environment. Remodeling each of these diverse cellular properties during reprogramming in an effort to mimic a developmentally differentiated cell is a monumental task. A systematic computational comparison of fibroblast, endogenous neuron, and induced neuron transcriptomes demonstrated that induced neurons adopt a functioning neuronal identity, yet retain significant gene expression and epigenetic hallmarks specific to the fibroblast lineage (Cahan et al. 2014). This failure to completely silence fibroblast gene regulatory networks in

reprogrammed neurons results in the generation of cells with heterogeneous genetic and epigenetic background that resemble functional neurons (Wapinski et al. 2013). This cell-of-origin background effect is even observed in the transcriptome (Polo et al. 2010) and DNA methylome (Ohi et al. 2011) upon reprogramming to pluripotency. This partial reset in the epigenetic state produces adult-derived neurons with aging hallmarks and nucleocytoplasmic deficits that limit therapeutic utility (Mertens et al. 2015). While some reprogramming and transdifferentiation strategies are more efficient at removing cell-of-origin hallmarks, these genetic elements remain a barrier to both induced neuron production and clinical translation.

1.4. Dissertation Research

1.4.1. Experimental Model

The high-efficiency *in vitro* transdifferentiation of human fetal lung and adult skin fibroblasts into functional cholinergic neurons revealed important roles for cell-of-origin identity in the design of reprogramming strategies (Liu et al. 2013). NEUROG2 and the small molecules forskolin (F) and dorsomorphin (D) are sufficient to catalyze the rapid transdifferentiation of fetal lung fibroblasts into neurons with greater than 90% efficiency (Liu et al. 2013). The addition of SOX11 and FGF2 to this cocktail was required to promote the efficient reprogramming of adult skin fibroblasts (Liu et al. 2013).

NEUROG2 and FD (NFD)-induced neurons express the mature neuronal proteins TUBB3, MAP2, MAPT, RBFOX3, and SYT1, as well as, the cholinergic markers CHAT, ISL1, and SLC18A3 (Liu et al. 2013). These neurons fire action potentials, exhibit tetrodotoxin-sensitive inward sodium current, respond to excitatory and inhibitory neurotransmitters, and exhibit spontaneous synaptic currents *in vitro* (Liu et al. 2013). Further, NFD-induced neurons form

functional neuromuscular junctions when co-cultured with C2C12 myotubes (Liu et al. 2013). The overexpression of ISL1 and LHX3 significantly improves this motor neuron identity in neurons derived from adult fibroblasts (Liu et al. 2016).

The reduced complexity of this single transcription factor transdifferentiation model enabled us to define the critical immediate-early molecular events that catalyze this change in cellular identity using high-through sequencing approaches.

1.4.2. Research Objective and Specific Aims

The objective of this research is to define the essential immediate-early molecular events that catalyze fibroblast-to-neuron transdifferentiation. Three specific aims comprise this objective.

Aim 1: The focus of this aim is to characterize NEUROG2 pioneer activity, define NEUROG2 chromatin occupancy genome-wide, and quantify the transcriptional changes induced by NEUROG2 overexpression in fibroblasts.

Aim 2: The focus of this aim is the identification of signal transduction pathways activated by FD and changes in transcription induced by these small molecules that improve reprogramming efficiency or neuron survival.

Aim 3: The focus of this aim is the identification of synergistic transcription and epigenetic modifications induced by NFD. This includes FD-induced modifications to NEUROG2 and heterodimeric partners that modulate DNA binding activity, the induction of co-factors that associate with NEUROG2 transcription complexes, FD-driven transcription factor activity or silencing, genome-wide changes in histone acetylation or methylation, and FD-driven changes in DNA accessibility that promote pro-neural transcription.

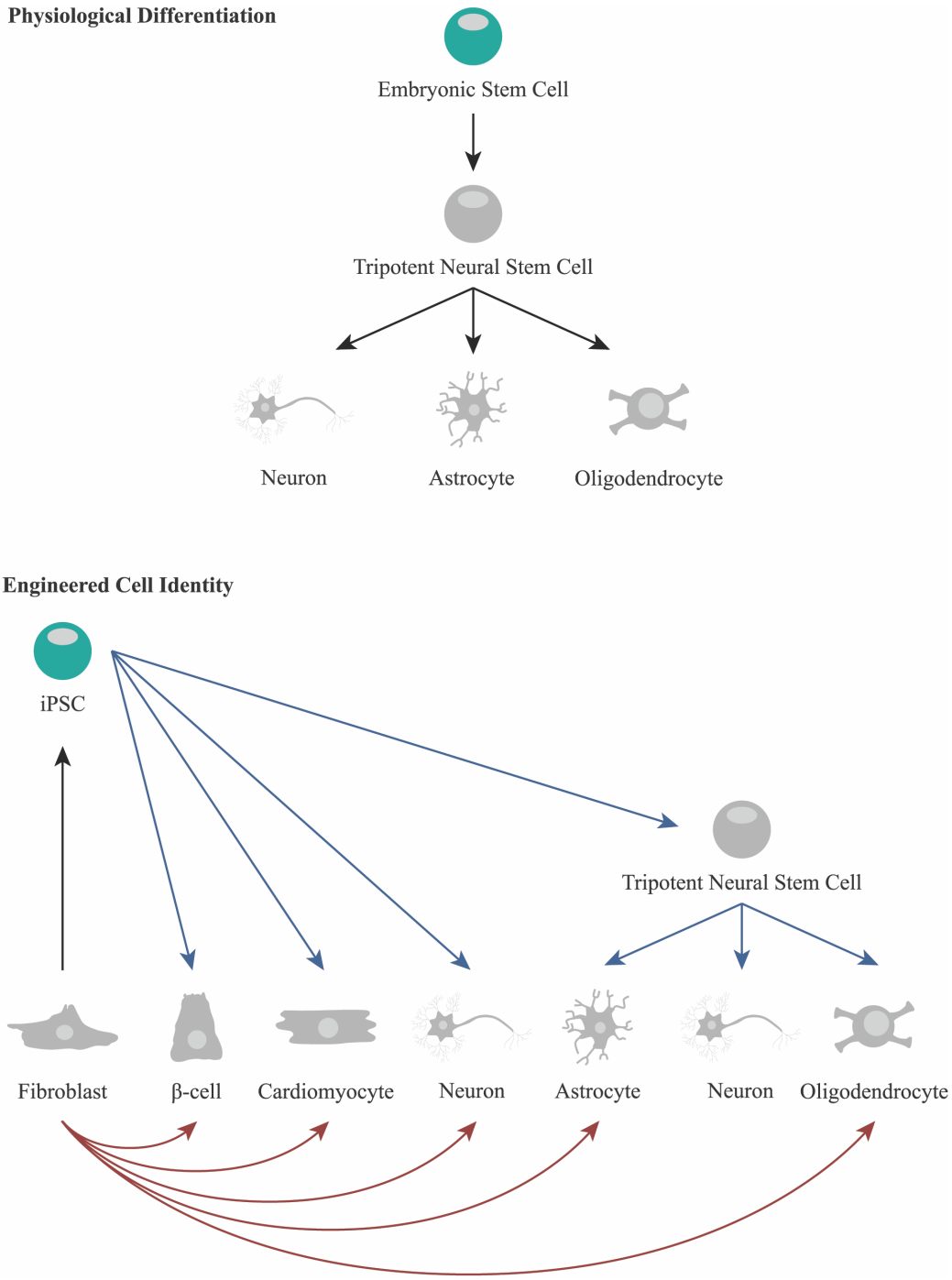


Figure 1: Physiological and Engineered Cell Identity

Physiological differentiation proceeds from pluripotency to multipotency to defined cell identity. Engineered cell identity is achieved through de-differentiation to a pluripotent or multipotent intermediate (blue arrows) or direct transdifferentiation (red arrows).

Table 1: Transdifferentiation of Subtype-Specific Human Neurons

Cell of Origin	Neuron Subtype	Reprogramming Factors	Reference
Cortical Astrocyte	Undefined	NEUROD4	Masserdotti et al. 2015
Fibroblast	Undefined	ASCL1, MYT1L, NEUROD1, POU3F2	Pang et al. 2011
Fibroblast	Undefined	miR-124, MYT1L, POU3F2	Ambasudhan et al. 2011
Fibroblast	Undefined	ASCL1, miR-124, miR-9/9*, MYT1L, NEUROD2	Yoo et al. 2011
Fibroblast	Undefined	7 Small Molecules	Hu et al. 2015
Fibroblast, Adult	Cholinergic	Dorsomorphin, FGF2, Forskolin, NEUROG2, SOX11	Liu et al. 2013
Fibroblast, Fetal	Cholinergic	Dorsomorphin, Forskolin, NEUROG2	Liu et al. 2013
Fibroblast	Dopaminergic	ASCL1, LMX1A, NR4A2	Caiazzo et al. 2011
Fibroblast	Dopaminergic	ASCL1, LMX1A, miR-124, NR4A2, TP53-shRNA	Jiang et al. 2015
Fibroblast	Dopaminergic	ASCL1, NEUROG2, NR4A2, PITX3, SOX2	Liu et al. 2012
Fibroblast	Dopaminergic	ASCL1, FOXA2, LMX1A, MYT1L, POU3F2	Pfisterer et al. 2011
Fibroblast	GABAergic	ASCL1, DLX5, FOXG1, LHX6, SOX2	Colasante et al. 2015
Fibroblast	GABAergic Striatal Medium Spiny	BCL11B, DLX1, DLX2, miR9/9*, miR-124, MYT1L	Victor et al. 2014
Cortical Astrocyte	Glutamatergic	NEUROD1	Guo et al. 2014
Cortical Astrocyte	Glutamatergic	9 Small Molecules	Zhang et al. 2015
Fibroblast	Motor	Dorsomorphin, Forskolin, FGF2, NEUROG2, SOX11, ISL1, LHX3	Liu et al. 2016
Fibroblast	Motor	ASCL1, ISL1, LHX3, MNX1, MYT1L, NEUROD1, NEUROG2, POU3F2	Son et al. 2011

CHAPTER 2

EXPERIMENTAL METHODS

2.1. Cell Culture, Lentivirus Production, and Neuron Induction

2.1.1. Cell Culture

293T/17 and human fibroblast cell lines (Appendix 2) were cultured in Dulbecco's modified eagle medium (DMEM, GE Healthcare, SH30243.01) supplemented with 10% (v/v) (293T/17) or 15% (v/v) (fibroblast cell lines) fetal bovine serum (Corning, 35-010-CV) and 1% (v/v) penicillin-streptomycin (GE Healthcare, SV30010). Cells were maintained at 37°C in a humidified incubator with 5% CO_2.

2.1.2. Lentivirus Production

293T/17 cells were seeded at a density of 3×10^6 cells in a 10 cm polystyrene dish 16 hours prior to transient transfection. 293T/17 cultures were treated with fresh medium 45 minutes prior to transfection. Polyethylenimine (Polysciences, 23966) was used to transfect cells with third-generation lentiviral packaging, envelope, transfer, and expression vectors (Appendix 3). Cultures were treated with fresh medium 16 hours post transfection and replication-deficient lentivirus was collected 24 and 48 hours following this medium change. Lentivirus-containing medium was syringe filtered through a 0.22 μm polyvinylidene fluoride membrane (EMD Millipore, SLGV033RS) and stored at 4°C.

2.1.3. Neuron Induction

2.1.3.1. NEUROG2, Forskolin, and Dorsomorphin

MRC-5 fibroblasts were seeded at a density of 0.8×10^6 cells in a 10 cm polystyrene dish or 2×10^4 cells on standard 24-well glass coverslips sequentially treated with 0.1% (w/v) gelatin (Bio-Rad Laboratories, 170-6537) for 10 minutes at 24°C and Matrigel Basement Membrane Matrix (BD Biosciences, 356234) diluted 200-fold in DMEM for 16 hours at 37°C. Fibroblasts were transduced with *NEUROG2*-encoding lentivirus 24 hours after plating. The optimal lentivirus titer was empirically determined as 4-fold dilution for 10 ml-capacity 10 cm dishes and 10-fold dilution for 1 ml-capacity 24-well plates. These dilutions consistently yielded ≥90% TUBB3-positive and MAP2-positive neurons. Transduced-fibroblast cultures were treated with fresh medium 24 hours post infection. Cultures were transitioned to neuron induction medium (DMEM, Ham's F12 nutrient mixture (GE Healthcare, SH30026.02), neurobasal medium (Life Technologies, 21103-049), N-2 supplement (Life Technologies, 17502-048), B-27 supplement (Life Technologies, 17504-044), and penicillin-streptomycin at 1 : 1 : 0.5 : 0.02 : 0.01 : 0.025) supplemented with 10 μM forskolin (Sigma Aldrich, F6886) and 1 μM dorsomorphin (Millipore, 171260) 48 hours post infection. Four days post infection half of the total volume of medium was removed and replaced by one full volume of medium (for example, 500 μl removed and 1 ml added). Six days post infection half of the total volume of medium was removed and replaced with half of the original volume of neuron induction medium supplemented with 10 μM forskolin (for example, 750 μl removed and 500 μl added). Neuron induction medium supplemented with 10 μM forskolin was then half changed every two days until replating (Liu et al. 2013). Immunocytochemical staining, gene expression profiling, and electrophysiological characterization of induced neurons was previously demonstrated (Liu et al. 2013). Neuron

induction efficiency was calculated as the total number of cells expressing a neuron-specific marker (TUBB3 or MAP2) with characteristic neuron morphology (rounded soma and neurites at least five times the soma length) relative to the total number of viable GFP-expressing cells.

2.1.3.2. Forskolin Replacement Screen

MRC-5 fibroblasts were cultured, transduced with *NEUROG2*-encoding lentivirus, and treated with neuron induction medium along the above-described time course. Neuron induction medium was supplemented with 1 μM dorsomorphin and either 10 μM forskolin, 10 μM cyclic AMP (cAMP, Sigma Aldrich, A6885) or 10 μM dibutyryl cAMP (Sigma Aldrich, D-0627). TUBB3 immunocytochemical staining was performed in triplicate 12 days post infection and reprogramming efficiency quantified using 10 random fields from each replicate.

2.1.3.3. NEUROG2, Constitutively Active PRKACA, and Dorsomorphin

A constitutively active form of the human protein kinase A catalytic subunit (*caPRKACA*) was generated by site-directed mutagenesis of the wild-type catalytic subunit (Addgene, 23495) to introduce H88Q and W197R amino acid substitutions (Orellana et al. 1992) (Appendix 3). MRC-5 fibroblasts were cultured, transduced with a *NEUROG2+caPRKACA*-encoding lentivirus, and treated with neuron induction medium supplemented with 1 μM dorsomorphin along the above-described time course. Immunocytochemical staining for TUBB3 and MAP2 was performed to quantify reprogramming efficiency.

2.1.3.4. NEUROG2 Phosphomutant Screen

A multi-species alignment of the *NEUROG2* gene identified nine potential GSK3- or

PRKACA-targeted phosphorylation sites. Phosphomimetic and phospho-deficient NEUROG2 constructs were generated for these sites by consecutive rounds of site-directed mutagenesis of wild-type *NEUROG2* to introduce phosphomimetic (S24D, S91D, S193D, S207D, S209D, S219D, S232D, S239D, S242D) and phospho-deficient (S24A, S91A, S193A, S207A, S209A, S219A, S232A, S239A, S242A) mutations (Appendix 3). MRC-5 fibroblasts were cultured, transduced with phosphomutant *NEUROG2*-encoding lentivirus, and treated with neuron induction medium or neuron induction medium supplemented with 10 μM forskolin and 1 μM dorsomorphin along the above-described time course. Immunocytochemical staining for TUBB3 and MAP2 was performed to quantify reprogramming efficiency.

2.1.3.5. NEUROG2 Deletion Construct

A NEUROG2 deletion construct (*NEUROG2* coding sequence positions 328 to 651 or NEUROG2 residue positions 110 to 217) was cloned into pCSC-SP-PW-IRES/GFP using AgeI and PstI restriction sites. MRC-5 fibroblasts were cultured, transduced with *NEUROG2* deletion construct-encoding lentivirus (10-fold dilution for 1 ml-capacity 24-well plates), and treated with neuron induction medium with or without 10 μM forskolin and 1 μM dorsomorphin along the above-described time course. MAP2 immunocytochemical staining was performed in triplicate 14 days post infection and reprogramming efficiency quantified using three random fields from each replicate.

2.1.3.6. Adult Fibroblast Transdifferentiation: NEUROG2 and SOX4

Adult fibroblast lines (AG05811, AG09969, ND29563, and ND39027) were plated at a density of 0.3×10^6 cells in a 6 cm polystyrene dish or 1×10^4 cells in a 48-well plate treated with

500 µl DMEM containing 100-fold diluted Matrigel Basement Membrane Matrix for 16 hours at 37°C. Fibroblasts were transduced with *NEUROG2*- or *NEUROG2*+*SOX4*-encoding lentivirus 24 hours after plating. The optimal lentivirus titer was empirically determined for each cell line. Transduced-fibroblast cultures were treated with fresh medium 24 hours post infection. Cultures were transitioned to neuron induction medium supplemented with 10 µM forskolin, 1 µM dorsomorphin, and 20 ng/ml FGF2 48 hours post infection. Four days post infection half of the total volume of medium was removed and replaced by one full volume of medium (for example, 250 µl removed and 500 µl added). Six days post infection half of the total volume of medium was removed and replaced with half of the original volume of neuron induction medium containing supplements (for example, 375 µl removed and 250 µl added). Neuron induction medium with supplements was then half changed every two days until analysis of neuron induction efficiency. Efficiency was calculated as the total number of GFP- and TUBB3-expressing cells with rounded morphology and at least one neurite relative to the total number of viable GFP-expressing cells. Alternatively, neurons were replated 14 days post infection. Replated neurons were washed twice with 2 ml phosphate-buffered saline (PBS, GE Healthcare, SH30028.02), incubated in 1 ml of 10-fold diluted trypsin (GE Healthcare, SH30042.01) for 10 minutes at 37°C, and suspended in 6 ml 15% (v/v) fetal bovine serum-containing medium. Cells were transferred to a 10 cm polystyrene dish pre-treated with 6 ml of 0.1% gelatin for 10 minutes at room temperature and incubated for 30 minutes at 37°C. Medium was collected without disturbing adherent fibroblasts and neurons were collected by centrifugation (500 × g, 3 minutes, 23°C). Neurons were suspended in 100 µl neuron maturation medium (DMEM, Ham's F12 nutrient mixture, neurobasal medium, N-2 supplement, B-27 supplement, and penicillin-streptomycin at 1 : 1 : 0.5 : 0.02 : 0.01 : 0.025) supplemented with 5 µM forskolin, 20

ng/ml BDNF (PeproTech, 450-02), 20 ng/ml GDNF (PeproTech, 450-10), and 20 ng/ml NT-3 (PeproTech, 450-03). Neuron suspensions were plated in 1 ml neuron maturation medium on standard 24-well glass coverslips treated with 500 µl coating solution (5 µg fibronectin (ThermoFisher Scientific, CB-40008A), 5 µg laminin (ThermoFisher Scientific, CB-40232), and Matrigel Basement Membrane Matrix diluted 500-fold in DMEM for 48 hours at 37°C then washed once with PBS before plating). Medium was half changed every 48 hours for the remainder of neuron culture.

2.1.3.7. Adult Fibroblast Transdifferentiation: NEUROG2 and SWI/SNF Factors

AG05811 adult skin fibroblasts were plated at a density of 1×10^4 cells in a 48-well plate treated with 100 µl DMEM containing 500-fold diluted Matrigel Basement Membrane Matrix, 1 µg fibronectin, and 1 µg laminin for 48 hours at 37°C. Fibroblasts were transduced with *NEUROG2*- and *SMARCA4*- or *NEUROG2*-, *SMARCA4*- *SMARCB1*-, and *SMARCC2*-encoding lentiviruses 24 hours after plating. Transduced-fibroblast cultures were treated with fresh medium 24 hours post infection then transitioned to neuron induction medium supplemented with 10 µM forskolin, 1 µM dorsomorphin, and 20 ng/ml FGF2 48 hours post infection. Reprogramming was performed in both the presence and absence of 20 ng/ml BDNF, 20 ng/ml GDNF, and 20 ng/ml NT-3. Neuron induction medium was changed along the above-described time course until TUBB3 immunocytochemical staining 10 days post infection.

2.1.3.8. Adult Fibroblast Transdifferentiation: NEUROG2 and Chemicals

AG05811 adult fibroblasts were plated at a density of 1×10^4 cells in a 48-well plate treated with 100 µl DMEM containing 500-fold diluted Matrigel Basement Membrane Matrix, 1 µg

fibronectin, and 1 μg laminin for 48 hours at 37°C. Fibroblasts were transduced with 25 μl *NEUROG2*-encoding lentivirus 24 hours after plating. Fibroblast cultures were treated with fresh medium 24 hours post infection then transitioned to neuron induction medium supplemented with 10 μM forskolin, 1 μM dorsomorphin, 20 ng/ml FGF2, and one or a combination of the following small molecules: 15 μM 5-azacitidine (Selleck Chemicals, L1700), 32 μM decitabine (Selleck Chemicals, L1700), 0.5 μM DZNep (Selleck Chemicals, L1700), 1 μM EPZ5676 (Selleck Chemicals, L1700), 1 μM FK228 (Selleck Chemicals, L1700), 2 μM (EMD Millipore, 500580), 0.5 μM I-BET151 (Selleck Chemicals, L1700), 0.5 μM I-BET762 (Selleck Chemicals, L1700), 0.5 μM JQ1 (Selleck Chemicals, L1700), 1 μM liproxstatin-1 (Sigma Aldrich, SML1414), 25 μM RepSox (Selleck Chemicals, L1700), 50 μM RG108 (Selleck Chemicals, L1700), 2.5 μM SAHA (Selleck Chemicals, L1700), 2.5 μM SGI1027 (Selleck Chemicals, L1700), 10 μM UNC669 (Selleck Chemicals, L1700), 1 mM valproic acid (Sigma Aldrich, P4543) or 10 μM zebularine (Selleck Chemicals, L1700) 48 hours post infection. Small molecule-enhanced reprogramming was performed in both the presence and absence of 20 ng/ml BDNF, 20 ng/ml GDNF, and 20 ng/ml NT-3. Neuron induction medium was changed along the above-described time course until immunocytochemical staining 10 days post infection. The optimization of FK228 treatment was performed using 1 μM, 0.5 μM, 0.1 μM, and 0.01 μM concentrations for 7 hour and 24 hour exposure periods with and without 1 μM kenpaullone.

2.2. Chromatin Immunoprecipitation and Next Generation Sequencing

2.2.1 Assay for Transposase-Accessible Chromatin Using Sequencing

MRC-5 fibroblasts were cultured, transduced with *GFP*-encoding, *NEUROG2*-encoding or *NEUROG2*+*SOX4*-shRNA-encoding lentivirus, and treated with DMEM containing 15% fetal

bovine serum, neuron induction medium or neuron induction medium supplemented with 10 μM forskolin and 1 μM dorsomorphin for 4 days, respectively. Approximately 5×10^6 cells were treated with Accutase cell detachment solution (Innovative Cell Technologies) for 3 minutes and collected in ice-cold PBS by centrifugation (525 × g, 5 minutes, 4°C). Cells were resuspended by repeated gentle pipetting in 2 ml ice-cold resuspension buffer (PBS containing 25 mM HEPES (pH 7.0) and 5 mM EDTA). GFP-expressing viable cells were isolated using fluorescence-based sorting on a MoFlo platform (Beckman Coulter). Exactly 50,000 cells were sorted into lysis buffer (10 mM Tris (pH 7.5), 10 mM NaCl, 3 mM $MgCl_2$, and 0.1% (v/v) IGEPAL CA-630) at 4°C then collected by centrifugation (21,130 × g, 2 minutes, 4°C). The pellet was resuspended in 50 μl transposition mix (25 μl 2X TD buffer, 22.5 μl nuclease-free H_2O, and 2.5 μl Nextera Tn5 transposase (Illumina, FC-121-1030)) and incubated exactly 30 minutes at 37°C with gentle agitation (400 rpm). Chromatin was purified with a MinElute PCR purification kit (Qiagen, 28004) and twice eluted in 10 μl elution buffer. Amplification reactions of 20 μl eluted chromatin, 2.5 μl universal ATAC primer (25 μM, Appendix 4), 2.5 μl barcoded ATAC primer (25 μM, Appendix 4), and 25 μl NEBNext High-Fidelity 2X PCR Master Mix (New England Biolabs, M0541S) were amplified for 1 cycle (5 minutes, 72°C) followed by 5 cycles (10 seconds, 98°C; 30 seconds, 63°C; 1 minute, 72°C). To determine the number of additional amplification cycles required, quantitative real-time PCR using 5 μl ATAC library reaction, 2.5 μl NEBNext High-Fidelity 2X PCR Master Mix, 0.125 μl universal ATAC primer (25 μM), 0.125 μl barcoded ATAC primer (25 μM), and 2 μl 5X SYBR Green I (ThermoFisher Scientific, S7563) was performed for 1 cycle (30 seconds, 98°C) followed by 20 cycles (10 seconds, 98°C; 30 seconds, 63°C; 1 minute, 72°C). The appropriate number of additional amplification cycles was determined for each library (11 total cycles) and amplified libraries were purified with a

MinElute PCR purification kit then eluted in 20 µl elution buffer. Right side size selection with 0.4X Agencourt AMPure XP beads (Beckman Coulter, A63881) was used to reduce fragments larger than 1,000 nucleotides and left side size selection with 1X Agencourt AMPure XP beads was used to eliminate fragments smaller than 150 nucleotides. Fragment length and size selection were evaluated after library amplification using a high-sensitivity DNA analysis kit (Agilent Technologies, 5067-4626). Library quantification prior to flow cell loading was performed using bioanalyzer traces and a Quant-iT PicoGreen dsDNA assay kit (ThermoFisher Scientific, P11496).

2.2.2. Crosslinking Chromatin Immunoprecipitation

MRC-5 fibroblasts were cultured, transduced with lentivirus (*HA-NEUROG2*-encoding or *NEUROG2+HA-SOX4*-encoding), and treated with neuron induction medium or neuron induction medium supplemented with 10 µM forskolin and 1 µM dorsomorphin along the above-described time course. AG05811 fibroblasts were cultured, transduced with *NEUROG2+HA-SOX4*-encoding lentivirus, and treated with neuron induction medium supplemented with 10 µM forskolin, 1 µM dorsomorphin, and 20 ng/ml FGF2 along the above-described time course. Approximately 1×10^7 cells were treated with 16% (w/v) methanol-free formaldehyde (Thermo Scientific, 28908) at a final concentration of 1% (v/v) for 8 minutes at room temperature with gentle rotation following 0.5, 1, 2 or 4 days in neuron induction medium. Crosslinking was quenched with 1.375 M glycine at a final concentration of 0.125 M for 5 minutes at room temperature with gentle rotation. Cells were immediately placed on ice, washed twice with 10 ml ice-cold PBS, scraped from the dish surface, and collected by centrifugation (525 × g, 5 minutes, 4°C). Pelleted cells were resuspended in 5 ml lysis buffer

(100 mM HEPES (pH 8.0), 85 mM KCl, 1% (v/v) IGEPAL CA-630, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail (Roche, 11873580001)) and incubated on ice for 20 minutes. Cells were Dounce homogenized (15 repetitions) on ice and nuclei collected by centrifugation (525 × g, 5 minutes, 4°C). Pelleted nuclei were resuspended in 260 µl briefly chilled shearing buffer (50 mM HEPES (pH 8.0), 10 mM EDTA (pH 8.0), 1% (w/v) SDS, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail) and transferred to a TPX microtube (Diagenode, C-30010010-50) on ice. Chromatin was sheared for 45 minutes (sonication for 30 seconds followed by a 30 second pause for 45 cycles, power setting: high) using a Bioruptor (Diagenode, B01010002) equipped with a 4°C refrigerated water bath. Sheared chromatin was purified by centrifugation (21,130 × g, 10 minutes, 4°C) to remove precipitated detergent and insoluble debris. Chromatin concentration was measured using a microvolume spectrophotometer (DeNovix, DS-11+) and 150 µg (MRC-5) or 200 µg (AG05811) was 5-fold diluted in ice-cold immunoprecipitation buffer (50 mM HEPES (pH 8.0), 20 mM NaCl, 1 mM EDTA (pH 8.0), 0.1% (v/v) Triton X-100, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail). CREB1-bound DNA fragments were immunoprecipitated with 5 µg rabbit anti-phospho-CREB1 monoclonal antibody (Cell Signaling Technology, 9198S), HA-NEUROG2-bound and HA-SOX4-bound DNA fragments were immunoprecipitated with 5 µg rabbit anti-HA polyclonal antibody (Abcam, AB9110), NEUROD1-bound DNA fragments were immunoprecipitated with 5 µg rabbit anti-NEUROD1 monoclonal antibody (Abcam, AB109224), and NEUROD4-bound DNA fragments were immunoprecipitated with 5 µg rabbit anti-NEUROD4 polyclonal antibody (Abcam, AB90484) for 18 hours at 4°C with gentle nutation (Labnet, S0500). One aliquot of 10 µg sheared chromatin was stored at -20°C as input. Another aliquot of 10 µg sheared chromatin was sequentially treated with 10 µg RNase A

(Roche, 10109142001) for 1 hour at 37°C, 40 µg Proteinase K (ThermoFisher Scientific, BP1700-50) for 2 hours at 55°C, and 2 µl 5 M NaCl for 16 hours at 67°C. This sample was loaded to a 1% (w/v) agarose gel, electrophoresed at 100 volts for approximately 2 hours, and stained with ethidium bromide to confirm fragmentation within a range of 100-500 nucleotides. The following day, 100 µl Magnetic Protein G Dynabeads (Life Technologies, 10003D) were washed three times with 1 ml ice-cold PBS containing 0.2% (v/v) Tween 20 then transferred to a 1.5 ml microtube on a MagneSphere separation stand (Promega, Z5332) at 4°C. The wash solution was removed and the beads were resuspended in immunoprecipitation samples for 2 hours at 4°C with gentle nutation. The bead complexes were magnetically isolated, twice washed by resuspension in ice-cold immunoprecipitation buffer for 2 minutes, twice washed by resuspension in ice-cold wash buffer (100 mM Tris-HCl (pH 9.0), 500 mM LiCl, 1% (v/v) IGEPAL CA-630, 1% (w/v) deoxycholic acid, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail) for 1 minute, and finally resuspended in ice-cold high-salt wash buffer (wash buffer containing 150 mM NaCl) for transfer to a 1.5 ml microtube on ice followed by bead collection. Chromatin was eluted from collected beads by addition of 100 µl elution buffer (1% (w/v) SDS and 50 mM $NaHCO_3$) for 30 minutes at 27°C with vigorous agitation (1,400 rpm, Eppendorf Thermomixer). Beads were magnetically collected and eluted chromatin was transferred to a sterile 1.5 ml microtube. Input chromatin was thawed on ice and treated with 85 µl elution buffer for 30 minutes at 27°C. Input and immunoprecipitated chromatin were sequentially treated with 10 µg RNase A for 1 hour at 37°C, 80 µg Proteinase K for 2 hours at 55°C, and 15 µl 5 M NaCl for 18 hours at 67°C. Chromatin was purified with a QIAquick PCR purification kit (Qiagen, 28104). Each sample was mixed with 1.4 ml PB binding buffer, isolated on a QIAquick column matrix by centrifugation (15,800 × g, 1 minute), twice washed with

750 µl PE wash buffer, and eluted with 50 µl elution buffer into 1.5 ml DNA LoBind microtubes (Eppendorf, 022431021). Enrichment of CREB1-, NEUROD1-, NEUROD4-, NEUROG2-, and SOX4-targeted chromatin sites was evaluated by quantitative real-time PCR. Chromatin concentration was determined using a Qubit fluorometer and Qubit dsDNA HS assay kit (Life Technologies, Q32850).

2.2.3. Crosslinking Chromatin Co-Immunoprecipitation

MRC-5 fibroblasts were cultured, transduced with *NEUROG2*-encoding lentivirus, and treated with neuron induction medium supplemented with 10 µM forskolin and 1 µM dorsomorphin along the above-described time course. Chromatin was isolated from approximately 1×10^7 cells, sheared into 100-500 nucleotide fragments, immunoprecipitated with appropriate antibodies (Appendix 1), captured, washed, and eluted as described for crosslinking chromatin immunoprecipitation experiments. Input (25 µg) and immunoprecipitation (20% of eluted volume) samples were treated with Laemmli denaturing buffer, electrophoresed on 4-15% Mini-PROTEAN TGX precast gels, transferred to Immobilon-P polyvinylidene difluoride membrane (EMD Millipore, IPVH00010), incubated with appropriate antibodies (Appendix 1), and treated with enhanced chemiluminescence horseradish peroxidase substrate for detection (Life Technologies, 32106).

2.2.4. Flow Cytometry and Low-Cell Native Chromatin Immunoprecipitation

MRC-5 fibroblasts were cultured, transduced with *NEUROG2*-encoding lentivirus, and treated with neuron induction medium or neuron induction medium supplemented with 10 µM forskolin and 1 µM dorsomorphin for 2 days. Approximately 1×10^7 cells were treated with

Accutase cell detachment solution (Innovative Cell Technologies) for 3 minutes and collected in ice-cold PBS by centrifugation (525 × g, 5 minutes, 4°C). Cells were resuspended by repeated gentle pipetting in 4 ml ice-cold resuspension buffer (PBS containing 25 mM HEPES (pH 7.0) and 5 mM EDTA). GFP-expressing viable cells were isolated using fluorescence-based sorting on a MoFlo platform (Beckman Coulter). Approximately 500,000 cells were collected by centrifugation (525 × g, 5 minutes, 4°C) then washed with 1 ml ice-cold PBS for chromatin digestion (Gilfillan et al. 2012). Cells were resuspended in 95 µl Micrococcal Nuclease digestion buffer (New England Biolabs, B0247S) supplemented with 0.01 µg BSA (New England Biolabs, B9001S), 0.2% (v/v) Triton X-100, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail. Resuspended cells were treated with 100 gel units Micrococcal Nuclease (New England Biolabs, M0247S) at 37°C for 5 minutes. Cells were immediately transferred to ice, treated with 10 µl ice-cold quench buffer (100 mM HEPES (pH 8.0) and 55 mM EDTA), and sonicated in a 1.5 ml TPX microtube for 60 seconds (power setting: high) using a Bioruptor equipped with a 4°C refrigerated water bath. Chromatin was diluted with 110 µl ice-cold immunoprecipitation buffer (50 mM HEPES (pH 8.0), 40 mM NaCl, 5 mM EDTA (pH 8.0), 0.2% (v/v) Triton X-100, 0.2% (w/v) SDS, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail), cell debris removed by centrifugation (21,130 × g, 10 minutes, 4°C), and 100 µl of supernatant transferred into two 0.2 ml microtubes (VWR, 732-0547). Chromatin was incubated with 1 µg rabbit polyclonal histone H3 acetyl K27 (Abcam, ab4729) or mouse monoclonal histone H3 tri-methyl K27 (Abcam, ab6002) antibody for 18 hours at 4°C with gentle nutation. The remaining 20 µl of supernatant was stored at -20°C as input. The following day, 20 µl Magnetic Protein G Dynabeads were washed three times with 200 µl ice-cold PBS containing 0.2% (v/v) Tween 20 on a MagneSphere separation stand at 4°C. The wash solution was

removed and the beads were resuspended in 10 µl immunoprecipitation buffer then added to immunopreciptation samples for 2 hours at 4°C with gentle nutation. The bead complexes were magnetically isolated, twice washed by resuspension in 150 µl ice-cold immunoprecipitation buffer for 2 minutes, twice washed by resuspension in 150 µl ice-cold wash buffer (100 mM Tris-HCl (pH 9.0), 500 mM LiCl, 1% (v/v) IGEPAL CA-630, 1% (w/v) deoxycholic acid, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail) for 1 minute, and finally resuspended in 150 µl ice-cold high-salt wash buffer (wash buffer containing 150 mM NaCl) for transfer to a 1.5 ml microtube on ice followed by bead collection. Chromatin was eluted from collected beads in 50 µl elution buffer (1% (w/v) SDS and 50 mM $NaHCO_3$) by two sequential additions of 25 µl elution buffer for 30 minutes at 27°C with vigorous agitation (1,400 rpm). Beads were magnetically collected and eluted chromatin was transferred to a sterile 1.5 ml microtube. Input chromatin was thawed on ice and treated with 50 µl elution buffer for 30 minutes at 27°C. Input and immunoprecipitated chromatin were sequentially treated with 10 µg RNase A for 15 minutes at 37°C and 80 µg Proteinase K for 3 hours at 55°C. Chromatin was purified with a QIAquick PCR purification kit and concentration was determined using a Qubit fluorometer and Qubit dsDNA HS assay kit as described above.

2.2.5. Massively Parallel DNA Sequencing

ChIP-seq libraries were synthesized from 20 ng purified input chromatin and 3.5-10 ng purified immunoprecipitated chromatin using a NEBNext ChIP-Seq Library Prep Master Mix Set for Illumina (New England Biolabs, E6240S) with NEBNext Multiplex Oligos for Illumina (New England Biolabs, E7335S). Replicate libraries for each experimental condition and time point were prepared from independent immunoprecipitations. Condition-dependent replicates

were not performed for ATAC-seq libraries. Single-end 50-base length sequencing reads were generated on an Illumina HiSeq 2500 System. Reads were aligned to the human reference sequence GRCh37/hg19 with the Bowtie algorithm (version 1.0.0) (Langmead et al. 2009). Peak calling, peak-gene annotation, motif discovery (parameter: 200 nucleotide window from peak center), and generation of heatmap matrices were performed using HOMER (version 4.7) (Heinz et al. 2010). Heatmaps were generated by hierarchical clustering using Cluster (version 3.0) (Eisen et al. 1998) and Java TreeView (version 1.1.6) (Saldanha, 2004). Gene ontology classification was performed using HOMER and Genomic Regions Enrichment of Annotations Tool (version 2.0.2) (McLean et al. 2010). The UCSC Genome Browser was used to visualize tag densities and multi-experiment datasets (Kent et al. 2002). Sequences under peaks annotated to the 100 greatest upregulated and downregulated genes were extracted from the UCSC Genome Browser for enhancer box motif compositional analysis (NEUROG2 peak height ≥4 in the NFD condition, genomic sequence was defined between two points where the peak height equals 2 units, peak assignment required a maximum distance of 100 kilobases from the annotated gene and a minimum of 15 kilobases from any neighboring gene with exception for peaks within 10 kilobases of the annotated gene). Super-enhancer analysis was performed with HOMER using a 12.5 kilobase stitching window and NEUROG2 ChIP-seq and ATAC-seq normalized tag count cutoffs of 10.1 (Whyte et al. 2013). Positional analysis of histone 3 lysine 27 acetylation and tri-methylation reads was performed with NGSplot (Shen et al. 2014).

2.2.6. Quantitative Real-Time PCR

Enrichment of transcription factor-targeted chromatin sites was validated for immunoprecipitation experiments on a 7900HT Fast Real-Time PCR System (Applied

Biosystems, 4329002) using FastStart Universal SYBR Green Master Mix (ROX) (Roche, 04913922001) and primers specific to target genes containing enhancer box, high mobility group box or cAMP response element motifs, as well as, controls for non-targeted chromatin and chromatin shearing efficiency (Appendix 4). DNA amplification was performed for 2 minutes at 50°C, 10 minutes at 95°C, 40 cycles of 14 seconds at 95°C, and 60 seconds at 58°C. Amplification specificity was confirmed using a heat dissociation protocol during the final cycle.

2.3. Electrophoretic Mobility Shift Assay

PRKACA, *NEUROG2*, and N-terminal HA-tagged *TCF3* (E12 isoform, Harvard Medical School Plasmid Repository), *TCF3* (E47 isoform, Addgene, 16059), *TCF4* (Harvard Medical School Plasmid Repository), and *TCF12* (Harvard Medical School Plasmid Repository) were cloned into bacterial expression vectors. An in-frame GFP-flexible-linker-NEUROG2 fusion protein was generated in a bacterial expression vector using a 22 amino acid linker peptide (Neuhold et al. 1993). Recombinant protein was generated using a TNT T7 Quick Coupled Transcription/Translation System (Promega, L1170). Single stranded oligos corresponding to *DLL3* (chr19:39,988,668-39,988,693) or *DLL3* with destroyed enhancer box (AAGCAA) were end-labeled using Titanium Taq DNA polymerase (Takara, 639242) with ^{32}P-dCTP (PerkinElmer, BLU513A250UC) and annealed for 5 minutes at 95°C then 68°C for 15 minutes. Double stranded oligos were purified using a 1 ml hand-packed column containing Sephadex G-25 resin (Sigma Aldrich, S5772) pre-soaked in TE buffer (10 mM Tris (pH 8.0) and 1 mM EDTA) for 24 hours at 4°C. Labeling efficiency was determined using a scintillation counter (Beckman). Recombinant proteins were incubated with 12,500 units PRKACA (New England Biolabs, P6000S) supplemented with protein kinases buffer (New England Biolabs, B6022S) and

200 μM adenosine 5'-triphosphate magnesium salt (Sigma Aldrich, A9187) for 45 minutes at 30°C. Binding reactions containing recombinant protein (1, 2 or 3 μl TCF3 isoform E47 and GFP-NEUROG2 or NEUROG2), 2 μl reaction buffer (100 mM Tris (pH 7.5), 500 mM NaCl, 10 mM DTT, 10 mM EDTA, 50% glycerol), 0.1 nM ^{32}P-labeled double stranded oligo, and 2 μg Poly(dI-dC) (Sigma Aldrich, P4929) were mixed, incubated at 30°C for 60 minutes, and electrophoresed on a 1.5-hour pre-run 5% native polyacrylamide gel at 4°C for 2 hours at 200 volts. Each polyacrylamide gel was transferred to 3 mm chromatography paper (GE Healthcare, 3030-6461) and vacuum dried for 1.5 hours at 80°C. Protein-DNA interactions were visualized following 30 minute to 12 hour phosphorscreen exposure using a STORM 820 phosphorimager (GE Healthcare).

2.4. Luciferase Assay

An analysis of NEUROG2 transcriptional repression was performed using chromatin immunoprecipitation sequencing-identified genomic sequences annotated to significantly downregulated genes (*CD44*, chr11:35,166,898-35,167,156; *EMP1*, chr12:13,428,228-13,428,875; *GDF5*, chr20:34,029,331-34,029,429; *TEK*, chr20:34,029,331-34,029,429). These sequences were cloned into *CS-CDF-CG-PRE* engineered to replace the *CMV* promoter with a minimal promoter sequence derived from *pGL4.23[luc2/minP]*. MRC-5 fibroblasts were seeded at a density of 2×10^4 cells in a 24-well plate and infected with 150 μl reporter sequence-encoding lentivirus combined with *NEUROG2*-encoding lentivirus or *GFP*-encoding lentivirus. Transfected-fibroblast cultures were treated with neuronal induction medium 24 hours post transfection. Transfected cells were lysed in 50 μl Tropix Lysis solution 48 hours post

transfection and 10 µl lysate assayed using a Dual-Light Luciferase Gene Assay System (ThermoFisher Scientific, T1004) on a FLUOStar Optima platform (BMG LABTECH).

2.5. Mass Spectrometry

MRC-5 fibroblasts were cultured, transduced with *NEUROG2*-encoding lentivirus, and treated with neuron induction medium or neuron induction medium supplemented with 10 µM forskolin and 1 µM dorsomorphin along the above-described time course. Approximately 1.25×10^7 cells were placed on ice, washed twice with 10 ml ice-cold PBS, scraped from the dish surface into ice-cold PBS, and collected by centrifugation (525 × g, 5 minutes, 4°C). Cells were suspended in 500 µl ice-cold lysis buffer (100 mM Tris (pH 7.5), 100 mM KCl, 5 mM $MgCl_2$, 0.2% (v/v) Triton X-100, 0.1% (v/v) Tween 20, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail) and homogenized for 10 repetitions using a 25 G × 5/8 inch sterile PrecisionGlide needle (BD Biosciences, 305122). Cell lysate was treated with 1,000 units of Pierce Universal Nuclease (ThermoFisher Scientific, 88701) and incubated on ice for 30 minutes with 3-second bursts of vortexing every 10 minutes. Cell debris was removed by centrifugation (21,130 × g, 10 minutes, 4°C). In parallel, 100 µl Pierce Anti-HA magnetic beads (ThermoFisher Scientific, 88836) were washed with 1 ml ice-cold lysis buffer then transferred to a 1.5 ml microtube on a MagneSphere separation stand at 4°C. The wash solution was removed and the beads were resuspended in 500 µl cell lysate for 16 hours at 4°C with gentle nutation. The following day, bead complexes were magnetically isolated, washed three times by resuspension in ice-cold lysis buffer for 2 minutes, and protein eluted with 20 µl glycine (pH 2) at 27°C and 1,000 rpm shaking for 7 minutes. Eluate was transferred to a 1.5 ml microtube, neutralized with 4 µl 1 M Tris (pH 8), mixed with 24 µl Laemmli denaturing buffer (Sigma Aldrich, S3401), and

incubated at 100°C for 5 minutes. Protein was electrophoresed 1 cm into a 10% Mini-PROTEAN TGX precast gel (Bio-Rad Laboratories, 456-1033) using NuPAGE MOPS SDS Running Buffer (Life Technologies, NP0001). Protein was visualized with blue silver stain (10% (v/v) phosphoric acid, 10% (w/v) ammonium sulfate, 0.12% (w/v) Brilliant blue G 250 (Sigma Aldrich, 27815), and 20% (v/v) anhydrous methanol) for 12 hours at room temperature with gentle rotation (Candiano et al. 2004). The gel was washed for 2 hours in HyPure Molecular Biology Grade Water (GE Healthcare, SH30538.02) then gel bands were removed by sterile excision and transferred into a 1.5 ml microfuge tube rinsed one time with 50% ethanol. Samples were reduced and alkylated with DTT (Sigma Aldrich, 9779) and iodoacetamide (Sigma Aldrich, I6125) then digested overnight with trypsin (Promega, V5280). Protein was purified by solid-phase extraction using an Oasis HLB plate (Waters, WAT058951) then analyzed by liquid chromatography-mass spectrometry-mass spectrometry using an Ultimate 3000 RSLC-Nano liquid chromatography system (Dionex) coupled to a Q Exactive mass spectrometer (Thermo Electron). Raw data files were converted to peak-list format and analyzed using the central proteomics facilities pipeline (version 2.0.3) (Trudgian et al. 2010, Trudgain et al. 2012). Peptide identification was performed using the X!Tandem (Craig et al. 2004) and open mass spectrometry search algorithm (Geer et al. 2004) engines against the Uniprot human protein database with common contaminants and reversed decoy sequences flagged (Elias et al. 2007). Fragment and precursor tolerances of 20 parts per million and 0.1 Daltons were specified with three missed cleavages permitted. Label-free quantitation of proteins was performed using SINQ normalized spectral index software (Trudgian et al. 2011). Three replicates from NEUROG2-transduced fibroblasts and two replicates from NFD-treated fibroblasts were used for analysis. Classification as a NEUROG2 interactor protein required spectral counts ≥4 in at least

two replicates of one condition with the ratio of normalized spectral counts indicating no detection in the untreated fibroblast control sample. Interactors were assigned as unique if undetected in any replicate of the opposite condition.

2.6. Microscale Thermophoresis

GFP-NEUROG2 fusion and N-terminal HA-tagged TCF3 (E12 isoform), TCF3 (E47 isoform), TCF4, and TCF12 recombinant proteins were prepared as described above (see electrophoretic mobility shift assay). Single stranded oligos corresponding to *DLL3* (chr19:39,988,668-39,988,693) or *DLL3* with mutated enhancer box (CAGATG or CATCTG) were annealed in thermophoresis buffer (25 mM HEPES (pH 7.3), 50 mM NaCl, 2.5 mM $MgCl_2$, 0.025% IGEPAL CA-630, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail) using a 98°C to 4°C gradient at a cooling rate of 0.5°C per 20 seconds. GFP-NEUROG2 was mixed at a 1:1 (v/v) ratio with each E protein and incubated at 30°C for 30 minutes. In parallel, 1:1 (v/v) NEUROG2-E protein mixtures were incubated with 12,500 units PRKACA in the presence of 200 μM adenosine 5'-triphosphate magnesium salt for 30 minutes at 30°C. Recombinant protein heterodimer mixtures were added to double stranded oligos (1:7 (v/v) dilution) in a titration series from 208 μM to 6 nM. Binding reactions were loaded into standard treatment capillaries (NanoTemper Technologies, MO-K002) and thermophoretic mobility quantified using a Monolith NT.115 Platform (NanoTemper Technologies). Thermophoretic traces were collected from three positions in each capillary using 100% light emitting diode power, 40% infrared laser power, and standard thermophoresis program (fluorescence before: 5 seconds, laser on: 30 seconds, fluorescence after: 5 seconds, delay: 25 seconds). NTControl (version 2.2.1) and

PALMIST (Chad Brautigam, Ph.D., The University of Texas Southwestern Medical Center) were used to derive average binding curves and apparent K_d values.

2.7. PRKACA Phosphorylation Assay

Escherichia coli Rosetta (DE3) cells (EMD Millipore, 70954) were transformed with *GFP-NEUROG2*, *NEUROG2*, *HA-TCF3* (isoform E12), *HA-TCF3* (isoform E47), *HA-TCF4*, and *HA-TCF12* constructs independently or co-transformed with *PRKACA*. Single colonies were used to inoculate 2 ml Luria broth containing 100 μg/ml ampicillin (Fisher Scientific, BP1760-25). Co-transfected cells were dual-selected with 25 μg/ml kanamycin (Fisher Scientific, BP906-5). Cultures were grown 16 hours at 37°C with shaking (225 rpm) then 500 μl was transferred to 50 ml antibiotic-containing Luria broth and grown to an optical density$_{600}$ of 0.6. Protein overexpression was induced with 1 mM isopropyl-β-D-thiogalactopyranoside (Fisher Scientific, BP1620-10) for 16 hours at 30°C with shaking (225 rpm). Protein-expressing cells were collected by centrifugation (525 × g, 5 minutes, 4°C), lysed in 250 μl ice-cold lysis buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 1% (v/v) Triton X-100, 1% (w/v) deoxycholic acid, 0.1% (w/v) SDS, and 1% (v/v) cOmplete EDTA-free Protease Inhibitor Cocktail), and incubated on ice for 30 minutes with 3 second vortex bursts every 10 minutes. Lysed cell debris was removed by centrifugation (21,130 × g, 10 minutes, 4°C) and protein concentration quantified by colorimetric assay (Bio-Rad Laboratories, 500-0006). Protein extracts were treated with calf intestinal alkaline phosphatase (New England BioLabs, M0290S) for 2 hours at 37°C. Protein extracts were treated with Laemmli denaturing buffer for 5 minutes at 100°C then electrophoresed on an 8% denaturing polyacrylamide gel with and without 50 μM Phos-tag reagent (NARD Institute, 300-93523) and 100 μM $MnCl_2$ in NuPAGE MOPS SDS Running

Buffer. Each gel was rinsed two times with 10 ml transfer buffer (250 mM Tris, 2 M glycine, 20% (v/v) methanol) then consecutively incubated two times in 15 ml transfer buffer containing 10 mM EDTA for 10 minutes with gentle rotation followed by one 15 minute incubation in transfer buffer without EDTA. Protein was transferred to Immobilon-P polyvinylidene difluoride membrane, blocked with 5% (w/v) non-fat dry milk solution (5% (w/v) non-fat dry milk dissolved in TBST (100 mM Tris (pH 7.4), 150 mM NaCl, and 0.1% (v/v) Tween 20) for 1 hour at room temperature, incubated with anti-HA polyclonal antibody (Appendix 1) for 18 hours at 4°C with gentle rotation, washed five times with TBST for 5 minutes per wash, incubated with secondary antibody for 1 hour at room temperature with gentle rotation, washed three times with TBST for 5 minutes per wash, and treated with enhanced chemiluminescence horseradish peroxidase substrate for detection.

2.8. Reverse Transcription and Quantitative Real-Time PCR

Total RNA was isolated from cells using TRIzol (Life Technologies, 15596-018) followed by chloroform extraction and column-based purification (Zymo Research, R1016). Reserve transcription was performed using 1.5 μg purified RNA and SuperScript III Reverse Transcriptase (Invitrogen, 18080-093) with random primers. Amplification was performed for 2 hours at 42 °C then 15 minutes at 72°C. Samples were cooled to 4°C then analyzed by quantitative real-time PCR using target-specific primers (Appendix 4) as described above.

2.9. RNA Sequencing

Total RNA was isolated in triplicate from fibroblasts transduced with *GFP*-encoding lentivirus (control), fibroblasts exposed to FD, fibroblasts transduced with *NEUROG2*-encoding

lentivirus, and fibroblasts transduced with *NEUROG2*-encoding lentivirus and exposed to FD. RNA isolation was performed 0.5, 1, and 2 days post transition to neuron induction medium using TRIzol (Life Technologies, 15596-018) followed by chloroform extraction and column-based purification (Zymo Research, R1016). Libraries were synthesized from 4 μg purified RNA using a TruSeq Stranded Total RNA Sample Preparation Kit (Illumina, RS-122-2201). Single-end 50-base length sequencing reads were generated on an Illumina HiSeq 2500 System. Reads were aligned to the human reference sequence GRCh37/hg19 with the TopHat (version 1.4.1) algorithm and transcript assembly performed with Cufflinks (version 2.1.0) (Trapnell et al. 2012). Expression levels of RefSeq-annotated genes were calculated in units of fragments per kilobase of exon per million mapped fragments (FPKM) and differential expression analysis was performed with Cuffdiff (version 2.1.0). Genes were defined as significant if the following three criteria were satisfied: Student's t-test p value ≥ 0.05, at least 2-fold gene expression change relative to the GFP-transduced control, and the triplicate average of ln(FPKM) ≥ 1 in either the experimental condition or fibroblast control. Gene ontology analysis was performed using the DAVID Functional Annotation Tool (version 6.7) (Huang et al. 2009a; Huang et al. 2009b).

2.10. Short Hairpin RNA-Mediated Gene Knockdown

Multiple rounds of PCR were used to introduce miR30 regulatory sequences at the 3' end of a *NEUROG2-IRES-GFP* lentiviral construct. Two restriction sites, XhoI and PstI, were engineered into these regulatory sequences at optimized processing locations to permit efficient generation of shRNAs targeting *CREB1*, *GADD45A*, *GADD45G*, *SOX4*, and *SOX11* (Fellmann et al. 2013) (Appendix 4). MRC-5 fibroblasts were cultured, transduced with *NEUROG2*+shRNA-encoding lentivirus, and treated with neuron induction medium supplemented with 10 μM forskolin and

1 μM dorsomorphin along the above-described time course. Approximately 6×10^5 transduced cells were scraped from the dish surface in ice-cold PBS and collected by centrifugation (525 × g, 5 minutes, 4°C). Cells were suspended in 75 μl ice-cold lysis buffer (50 mM Tris (pH 7.4), 150 mM NaCl, 1% (v/v) Triton X-100, 1% (w/v) deoxycholic acid, and 0.1% (w/v) SDS) and incubated on ice for 30 minutes with 3 second vortex bursts every 10 minutes. Lysed cell debris was removed by centrifugation (21,130 × g, 10 minutes, 4°C) and protein concentration quantified by colorimetric assay. Protein extracts (25 μg) were treated with Laemmli denaturing buffer for 5 minutes at 100°C then electrophoresed on 4-15% Mini-PROTEAN TGX precast gels (Bio-Rad Laboratories, 456-1083) and transferred to Immobilon-P polyvinylidene difluoride membrane. Membranes were blocked with 5% (w/v) non-fat dry milk or 5% (w/v) bovine serum albumin dissolved in TBST for 1 hour at room temperature, treated with target-specific antibodies (Appendix 1) for 18 hours at 4°C with gentle rotation, washed five times with TBST for 5 minutes per wash, treated with corresponding secondary antibodies (Appendix 1) for 1.5 hours at room temperature with gentle rotation, washed three times with TBST for 5 minutes per wash, and treated with enhanced chemiluminescence horseradish peroxidase substrate for detection. In the absence of a functioning antibody, RNA was isolated as described above for reverse transcription and quantitative real-time PCR analysis. In addition, MAP2 immunocytochemical staining was performed in triplicate 12 days post infection and reprogramming efficiency quantified using 10 random fields from each replicate.

CHAPTER 3

NEUROG2 CHROMATIN OCCUPANCY AND TRANSCRIPTIONAL REGULATION

3.1. Small Molecules Modulate NEUROG2 Chromatin Occupancy

MRC-5 fibroblasts transduced with NEUROG2 do not exhibit neuronal morphology or express hallmark neuronal proteins. In contrast, NEUROG2-transduced fibroblasts exposed to controlled dosages of FD undergo rapid morphological remodeling and adopt functional neuronal morphology within 50 days of treatment (Liu et al. 2013). The neuron induction protocol used to generate these neurons follows a stepwise procedure of 24-hour lentiviral infection, 24-hour recovery in FBS-containing media, and long-term exposure to neuronal induction media containing small molecules to jumpstart and maintain conversion. Fibroblasts immediately respond to FD with elongated morphology and significant transgene upregulation visible by enhanced GFP fluorescence. Within three days, fibroblasts elongate and outgrow low-complexity neurites. TUBB3- and MAP2-expressing high-complexity neurons emerge within 10-12 days of initial small molecules treatment.

To investigate how FD catalyze the reprogramming process, we performed ChIP-seq using crosslinked chromatin from *NEUROG2*-transduced fibroblasts in both the presence and absence of FD at 0.5, 1, and 2 DPT (Figure 2A). Replicate datasets were generated for each time point and condition with no individual dataset exhibiting less than 83% enrichment for the canonical enhancer box motif (Figure 2B). A comparative analysis of binding events in the presence and absence of FD at each time point revealed an immediate and sustained increase in NEUROG2 binding upon FD exposure (Figure 2C-F, Figure 3A). For instance, 12 hours following FD treatment 21,697 peaks were detected in the NFD condition, while only 8,164 of those peaks

were detected in the absence of small molecules (Figure 2C). This trend for enhanced occupancy continued at 1 DPT and 2 DPT (Figure 2D,E). Interestingly, the total number of NFD binding sites rapidly increased from 21,697 to 29,166 during the 0.5 DPT to 1 DPT transition indicating that these datasets collected information during the establishment phase of NEUROG2 action (Figure 2C,D). An NFD condition-dependent comparison of binding events in 0.5, 1, and 2 DPT datasets revealed a common core of 10,993 shared targets across all time points with smaller populations of dynamic binding events distributed over each time point (Figure 3A). This conditional analysis confirmed the genome-wide FD-enhanced NEUROG2 binding detected in the time point-specific analyses (Figure 2) and similarly detected thousands of NFD-specific binding events (Figure 3B).

In addition to thousands of new binding events, FD enhanced NEUROG2 signal intensity for 17% of targets detected at 0.5 DPT and 5.4% of targets detected at 2 DPT (Figure 2F). This suggests FD both facilitates NEUROG2 binding at novel genomic sites and enhances NEUROG2 binding at specific genomic targets. Independent of time and condition, NEUROG2 occupied 5,376 chromatin sites in transduced fibroblasts with an additional 1,393 sites specific to the NFD condition (Figure 3C). The genomic distribution of this shared and unique NEUROG2 occupancy relative to annotated genes indicates a bias toward intergenic and intronic sites. Although FD modestly enhanced promoter and transcription start site occupancy, NEUROG2 seems to dominantly regulate transcription via distal regulatory elements (Figure 3D).

3.2. NEUROG2 Functions as a Pioneer Factor and FD Induce Chromatin Remodeling

With ChIP-seq analyses identifying thousands of genomic binding sites for NEUROG2 and an interesting enhancement of this binding upon FD exposure, we next investigated the relative

accessibility of NEUROG2 binding sites in transdifferentiating fibroblasts using an assay for transposase-accessible chromatin coupled with massively-parallel sequencing (ATAC-seq). The genomic locations of the 1,000 highest scored peaks from 2 DPT NEUROG2 ChIP-seq datasets were probed for enrichment in 4 DPT ATAC-seq datasets collected from MRC-5 fibroblasts, *NEUROG2*-transduced fibroblasts, and *NEUROG2*-transduced fibroblast also treated with FD. Interestingly, the majority of these binding events occurred in the absence of ATAC-seq peaks indicating that NEUROG2 functions as a pioneer factor to access these sites in closed chromatin (Figure 4A, Zaret et al. 2011). While only a modest increase in open chromatin is detected at these sites upon NEUROG2 expression, a significant increase is observed upon addition of FD (Figure 4A). Narrowing this analysis to the single gene level, it becomes clear that NEUROG2 binds closed chromatin in the *INSRR-NTRK1*, *LRRC10B-SYT7*, and *GADD45G* loci highlighting the pioneer function of this transcription factor (Figure 4B). This local view also highlights the dramatic increase in chromatin accessibility induced by FD treatment (Figure 4B). Strong enrichment of FD-induced open chromatin is detected within the *INSRR* and *SYT7* gene bodies, as well as, the *GADD45G* enhancer demonstrating that these small molecules directly influence chromatin accessibility during transdifferentiation (Figure 4B). Fate- and lineage-determinative transcription factors like NEUROG2 have been shown to bind in regions of high density histone modifications, known as super-enhancers, to regulate gene expression (Whyte et al. 2013). Analysis of H3K27 acetylation (H3K27AC) ChIP-seq datasets from NFD-treated fibroblasts at 1 DPT identified 454 potential super-enhancers containing at least one NEUROG2 bound site, of which, 277 were localized to regions without ATAC-seq open chromatin signatures (Figure 4C). While an increase from 50 to 118 open chromatin sites was detected upon FD exposure, the vast

majority of super-enhancers remained localized to regions of compact chromatin. This highlights the functional importance of NEUROG2 pioneer activity during reprogramming (Figure 4C).

A genome-wide comparison of MRC-5 and NFD ATAC-seq datasets demonstrated that FD treatment substantially increased the number of open chromatin sites in MRC-5 fibroblasts. An analysis of the peaks uniquely detected upon treatment with NEUROG2 and FD revealed that newly open chromatin was primarily localized to intergenic and intronic regions, while approximately 8% of events occurred at promoter-TSS sites (Figure 4D). This distribution of open chromatin mimics the distribution of ChIP-detected NEUROG2 binding events that occur at distal regulatory elements (Figure 3D). A functional classification of genes annotated to newly detected NFD ATAC-seq peaks revealed a strong enrichment of differentiation and neurogenesis related genes (Figure 4E). Importantly, these peaks were enriched for the enhancer box motif with 50.1% of peaks containing the motif, as well as, CRE half-site and HMG box motifs with 4,672 peaks and 2,075 peaks containing each motif, respectively (Figure 4E). This motif-based search approach hints that FD might improve the accessibility of genetic regulatory elements via activation of transcription factors critical to transdifferentiation. To determine the potential transcriptional effects of this newly open chromatin signature, we performed a pair-wise analysis of RNA-seq expression datasets for genes annotated to NFD-unique ATAC-seq peaks (Section 3.3, Figure 4F). This analysis revealed a significant upregulation of pro-neural genes involved in neurogenic transcription, the Notch and Wnt signaling pathways, and chromatin remodeling genes (Figure 4F). Collectively, these results demonstrate that NEUROG2 functions as a pioneer transcription factor, FD promote significant changes chromatin structure during reprogramming, and these chromatin changes are directly correlated to the upregulation of pro-neural gene ontologies. In an effort to systematically characterize the effects of NEUROG2 and FD on

transcription during transdifferentiation, we performed a multi-time point analysis of gene regulation by whole transcriptome RNA-seq.

3.3. NEUROG2 and FD Synergize to Activate Pro-Neural Transcription

With evidence that NEUROG2 acts as a pioneer transcription factor and small molecules modulate NEUROG2 chromatin interactions, we next investigated how NEUROG2 and FD synergize to promote pro-neural transcription. Whole transcriptome RNA-seq was performed at 0.5, 1, and 2 DPT in parallel with NEUROG2 ChIP-seq. This design enabled us to capture profiles of both NEUROG2 genomic occupancy and transcription at identical time points in early converting neurons. Sequence reads from RNA libraries isolated from FD-treated, *NEUROG2*-transduced, and NFD-treated fibroblasts were normalized to RNA from time point-specific GFP-transduced fibroblasts for differential expression analysis.

NEUROG2 immediately induced a dramatic upward shift in the expression of neuronal genes such as *DLL3* and *HES6*, while FD independently activated relatively few neural-specific genes (Figure 5A). However, in combination, NEUROG2 and FD synergistically modified the expression of 1,398 additional genes unaltered by individual factors at 2 DPT (Figure 5B,C). The synergistic induction of *NEUROD1*, *NEUROD4*, and *SOX4* highlights the significance of this gene pool to the survival and maturation of engineered neurons.

Functional classification of NFD-regulated genes revealed an immediate repression of mitotic cell cycle and cell motion-related genes (Figure 5D). Similar classification of upregulated genes revealed a progressive enrichment of factors related to nervous system development and neuron differentiation (Figure 5E). Induction of the brain region-specific forebrain markers *NTRK1* and *SSTR2*, as well as, the cholinergic receptor *CHRNA3* and vesicular transporter

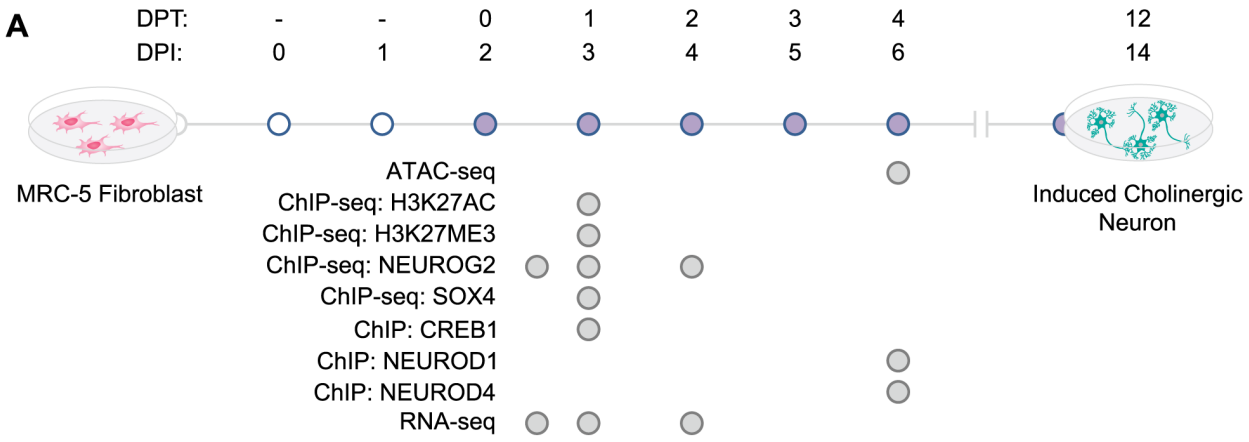
SLC18A3 suggests these early neurons are initially biased toward an excitatory cholinergic forebrain fate at 2 DPT (Figure 6A).

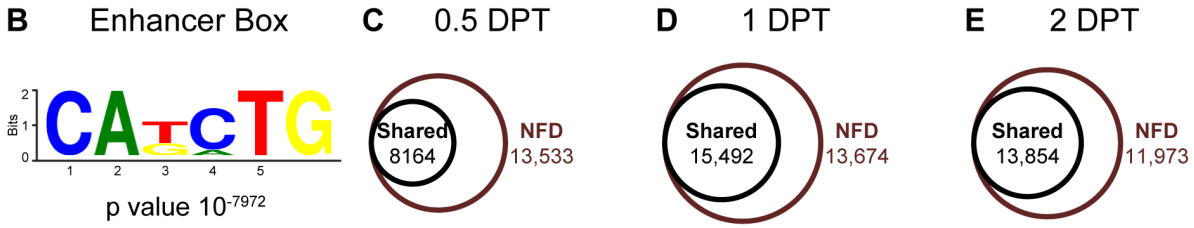
A temporal analysis of gene expression during active reprogramming underscores the rapid rate of change induced by NFD (Figure 5C). Synergistic regulation accounted for 53% of the total change in gene expression at 2 DPT and the total number of these genes exhibiting ≥5-fold change increased from 4% to 10% within the 36-hour window from 0.5 to 2 DPT (Figure 5C). This quantitative analysis demonstrates that NEUROG2 is the primary driver of pro-neural gene expression and FD compliment this neurogenic program by modulating transcription in a NEUROG2-sensitive manner.

Integrating ChIP-seq and RNA-seq datasets, peak-to-gene assignment by nearest-neighbor correlation followed by functional categorization of annotated genes revealed direct overlap with RNA expression changes ranging from high-level pan-neuronal gene ontologies to single gene targets such as *DLL3* and *SOX4* (Figure 6B). To identify how specific binding events correlate to the gene expression changes detected by RNA-seq, proximity-based peak-to-gene annotation was used to identify differentially expressed genes targeted by NEUROG2. FD exposure immediately enhanced the transcriptional activation of 679 genes with 13% exhibiting greater than 5-fold upregulation relative to NEUROG2 activation alone (Figure 6C). Surprisingly, FD similarly enhanced the downregulation of 898 NEUROG2-target genes at 2 DPT (Figure 6C). To determine whether NEUROG2 acts as a direct repressor of fibroblast-associated genes, we generated four lentiviral luciferase reporter constructs using ChIP-identified sequences derived from promoter, intronic, and distal regulatory regions of *GDF5*, *CD44*, *TEK*, and *EMP1*. Although these genes were significantly downregulated during reprogramming, NEUROG2 was not observed to repress luciferase expression via these genetic elements, which suggests that

NEUROG2 is not a direct transcriptional repressor. NEUROG2 might instead occupy these sites to competitively inhibit activation by fibroblast-lineage factors or recruit active repressor complexes to adjacent sites (Kovach et al. 2013). Additionally, proximity-based gene annotation does not reflect the multi-dimensional architecture of folded chromatin and the linear assignment binding sites to annotated genes might mask higher-complexity spatial regulatory functions.

In addition to enhanced expression and repression, FD initiate the transcription of NEUROG2-target genes unaffected by *NEUROG2*-transduction alone (Figure 6D). To determine how FD drives these changes in NEUROG2-mediated transcription, we next investigated whether FD exposure modifies NEUROG2 phosphorylation, heterodimer composition, sequence-specific enhancer box affinity, and association with co-activating transcription factors.





F **Small Molecules Modulate NEUROG2 Chromatin Occupancy**

	0.5 DPT		1 DPT		2 DPT	
NFD Unique	13,533 Peaks		13,674 Peaks		11,973 Peaks	
≥5 Fold Increase	0% (0 Peaks)	8,164 Shared Peaks	0% (0 Peaks)	15,492 Shared Peaks	0% (0 Peaks)	13,854 Shared Peaks
3-5 Fold Increase	1.1% (89 Peaks)		0.5% (77 Peaks)		0.3% (48 Peaks)	
2-3 Fold Increase	15.9% (1,297 Peaks)		5.9% (920 Peaks)		5.1% (711 Peaks)	
No Change	83% (6,776 Peaks)		93.1% (14,416 Peaks)		94% (13,016 Peaks)	
2-3 Fold Decrease	0.02% (2 Peaks)		0.5% (78 Peaks)		0.6% (79 Peaks)	
3-5 Fold Decrease	0% (0 Peaks)		0.01% (1 Peak)		0% (0 Peaks)	
≥5 Fold Decrease	0% (0 Peaks)		0% (0 Peaks)		0% (0 peaks)	

Figure 2: Small Molecules Enhance NEUROG2 Chromatin Occupancy

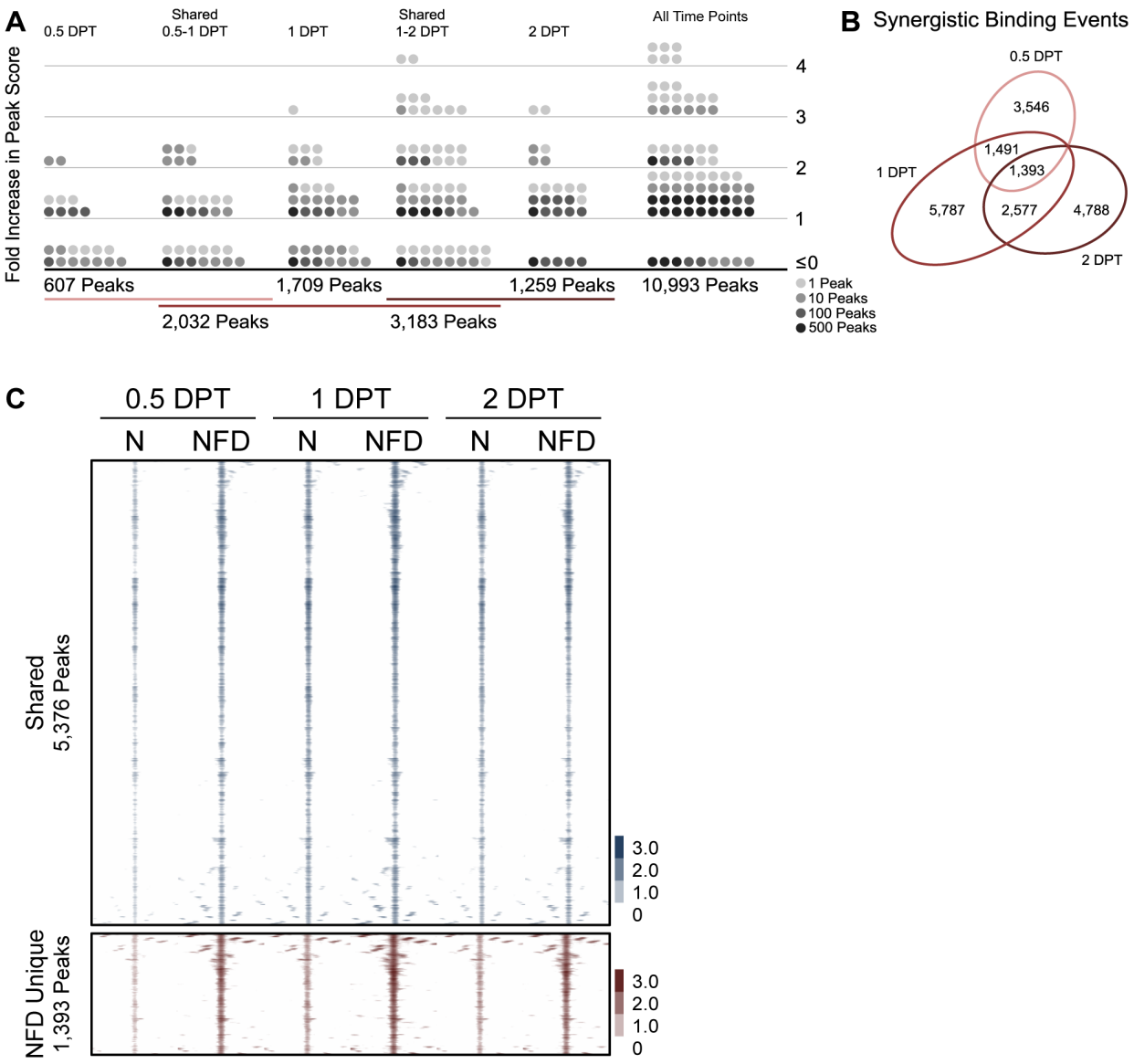
(A) Experimental design and time course of neuron induction (DPI, days post infection; DPT, days post treatment with FD).

(B) The enhancer box motif is significantly enriched at NEUROG2 occupancy sites in NFD-treated fibroblasts (2 DPT, 91% of targets). This motif is similarly enriched in ChIP-seq datasets across all conditions and time points.

(C-E) Condition-dependent analysis using area-proportional Venn diagrams to depict the number of NEUROG2 binding events at 0.5, 1, and 2 DPT. Shared events are detected in both the presence and absence of FD, while NFD unique events are detected only in the presence of FD.

(F) The effect of FD exposure on the intensity of NEUROG2 binding events detected each time point. The change in intensity is represented as a ratio of the peak score for NEUROG2 binding events in the presence and absence of FD. This analysis represents only peaks shared between both NEUROG2 and NFD conditions at the indicated time point.

*Portions of figure reproduced from Smith et al. 2016



D **Genomic Distribution of NEUROG2 Binding Events**

	0.5 DPT			1 DPT			2 DPT		
Location	NFD	Shared	Unique	NFD	Shared	Unique	NFD	Shared	Unique
3' UTR	1.4%	0.2%	1.3%	1.3%	1.4%	1.2%	1.5%	1.5%	1.6%
5' UTR	0.5%	0.3%	0.7%	0.9%	0.6%	1.2%	0.9%	0.6%	1.3%
Exon	2.3%	2.2%	2.3%	3.0%	2.5%	3.6%	3.0%	2.4%	3.8%
Intron	49.7%	49.8%	49.7%	49.1%	49.6%	48.6%	49.6%	49.2%	50.0%
Intergenic	38.1%	40.6%	36.5%	35.9%	38.7%	32.9%	35.8%	39.2%	31.9%
Noncoding	0.6%	0.5%	0.7%	0.7%	0.6%	0.5%	0.7%	0.6%	0.9%
Promoter/TSS	7.4%	5.0%	8.8%	9.0%	6.6%	11.7%	8.4%	6.5%	10.5%

Figure 3: Condition-Dependent NEUROG2 DNA Binding and Genomic Distribution

(A) Analysis of NEUROG2 binding events detected at 0.5, 1, and 2 DPT following NFD treatment and the effects of FD exposure on peak score. The change in intensity is represented as a ratio of the peak score for NEUROG2 binding events in the presence and absence of FD. This analysis represents all peaks detected following NFD treatment for all time points.

(B) An area-proportional Venn diagram that depicts the number of NEUROG2 binding events unique to the NFD-condition at 0.5, 1, and 2 DPT.

(C) Heatmaps representing NEUROG2 genome-wide occupancy profiles in the presence and absence of FD. Treatment with FD prompts 1,393 unique NEUROG2 binding events shared across all early time points. Input-normalized tag densities are plotted in $\log_2$ scale and centered in a 4-kilobase window around the peak apex (n=2 biological replicates for each condition and each time point; N, *NEUROG2*-transduced fibroblasts; NFD, *NEUROG2*-transduced fibroblasts exposed to FD).

(D) The genome-wide distribution of NEUROG2 binding events relative to annotated gene structures (TSS, transcription start site; UTR, untranslated region).

*Portions of figure reproduced from Smith et al. 2016

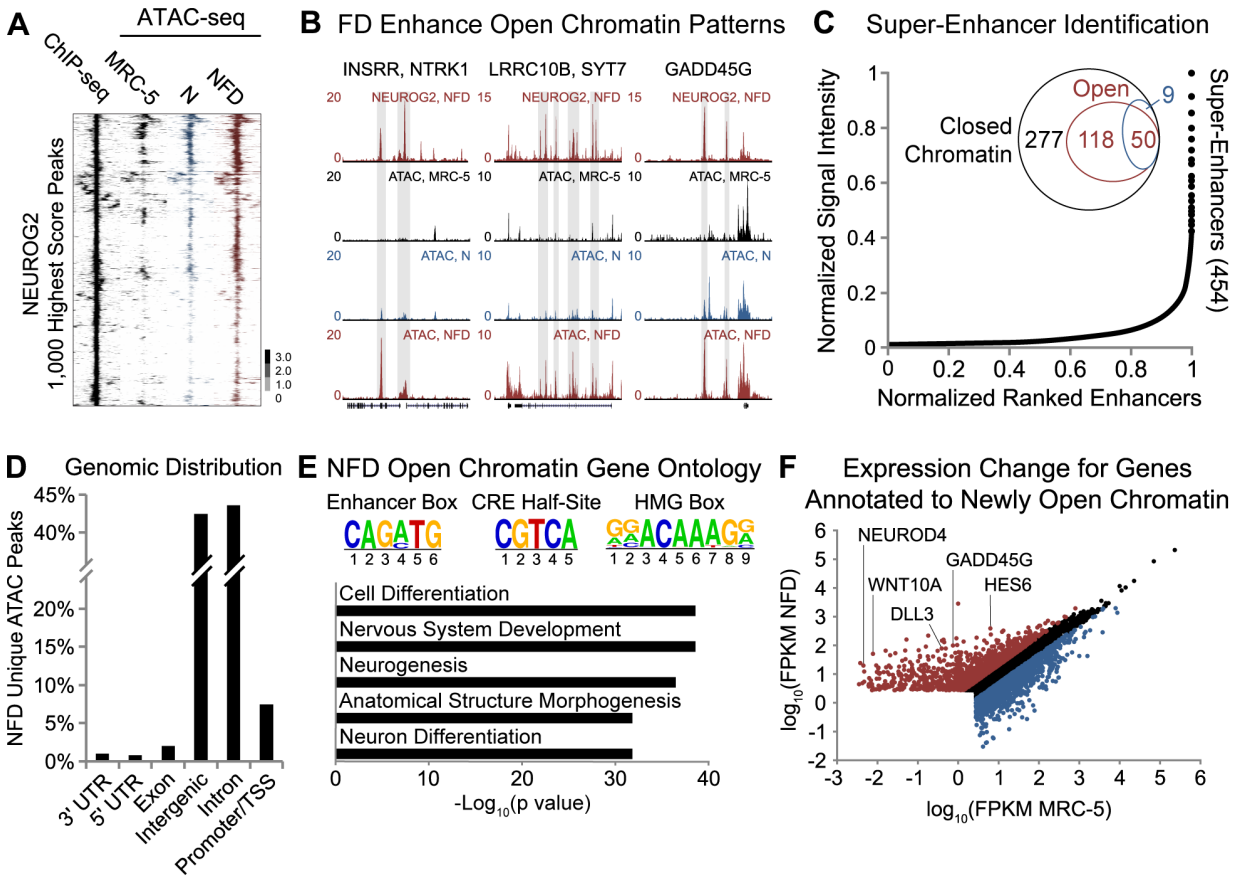

A
ATAC-seq
ChIP-seq
MRC-5
N
NFD
NEUROG2
1,000 Highest Score Peaks
3.0
2.0
1.0
0
B
FD Enhance Open Chromatin Patterns
INSRR, NTRK1
LRRC10B, SYT7
GADD45G
NEUROG2, NFD
ATAC, MRC-5
ATAC, N
ATAC, NFD
C
Super-Enhancer Identification
Open
Closed
Chromatin
277
118
50
9
Normalized Signal Intensity
Normalized Ranked Enhancers
Super-Enhancers (454)
D
Genomic Distribution
NFD Unique ATAC Peaks
3' UTR
5' UTR
Exon
Intergenic
Intron
Promoter/TSS
E
NFD Open Chromatin Gene Ontology
Enhancer Box
CRE Half-Site
HMG Box
Cell Differentiation
Nervous System Development
Neurogenesis
Anatomical Structure Morphogenesis
Neuron Differentiation
-Log10(p value)
F
Expression Change for Genes
Annotated to Newly Open Chromatin
NEUROD4
GADD45G
WNT10A
HES6
DLL3
log10(FPKM NFD)
log10(FPKM MRC-5)

Figure 4: NEUROG2 Functions as a Pioneer Factor and FD Opens Chromatin

(A) Heatmaps representing ATAC-seq detected open chromatin events at NEUROG2 binding sites (2 DPT ChIP-seq) in MRC-5 fibroblasts (black), MRC-5 fibroblasts treated with NEUROG2 (N, blue), and MRC-5 fibroblasts treated with NEUROG2 and FD (NFD, rightmost red). Input-normalized tag densities are plotted in $\log_2$ scale and centered in a 4-kilobase window around each peak.

(B) NEUROG2 ChIP-seq track (2 DPT NFD) and ATAC-seq tracks generated from fibroblasts (MRC-5), *NEUROG2*-transduced fibroblasts (N), and *NEUROG2*-transduced fibroblasts exposed to FD (NFD) for *INSRR*/*NTRK1*, *LRRC10B*/*SYT7*, and *GADD45G*. Gray shading highlights regions representing NEUROG2 pioneer activity or FD-induced chromatin remodeling.

(C) Plot representing the 454 NEUROG2-targeted super-enhancers identified from 1 DPT H3K27AC and NEUROG2 ChIP-seq datasets. A Venn diagram depicts the distribution of super-enhancers in open (NFD, red or N, blue) and closed (black) chromatin in each condition.

(D) The genomic distribution of ATAC-seq peaks newly detected upon NFD treatment is biased toward intergenic and intronic regions.

(E) Gene ontology analysis revealed enrichment of neurogenesis- and differentiation-related classifications for open chromatin peaks newly detected following NFD treatment. Additionally, the enhancer box, cAMP regulatory element (CRE) half-site, and HMG box motifs were detected in this population of open chromatin sites.

(F) Pair-wise scatter plot representing the change in expression of genes annotated to ATAC-seq open chromatin peaks newly detected upon NFD treatment. Upregulated genes (red) were enriched for pro-neuronal identity, while downregulated genes (blue) had non-neural function.

*Figure reproduced from Smith et al. 2016

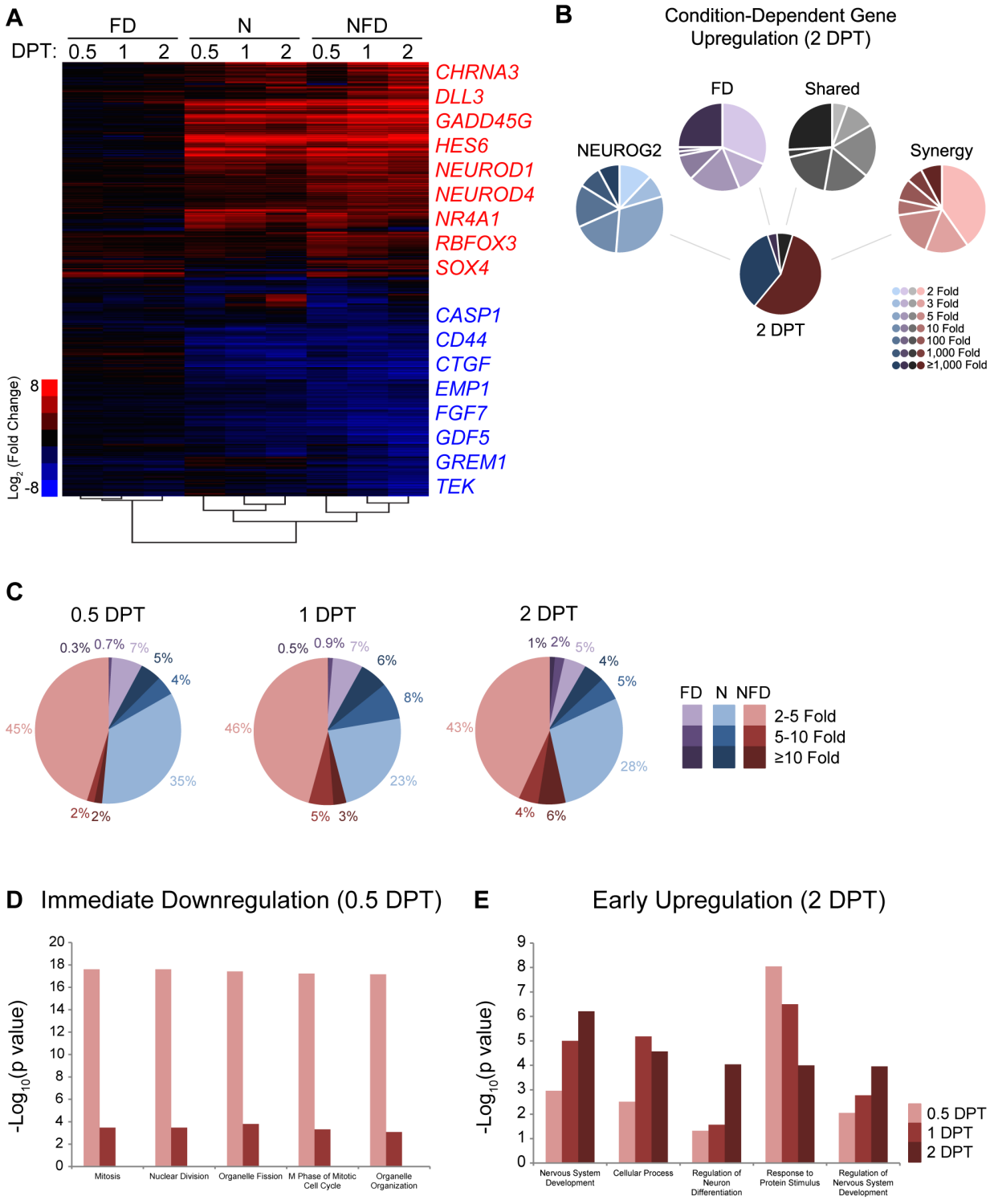

A
FD
N
NFD
DPT: 0.5 1 2 0.5 1 2 0.5 1 2
CHRNA3
DLL3
GADD45G
HES6
NEUROD1
NEUROD4
NR4A1
RBFOX3
SOX4
CASP1
CD44
CTGF
EMP1
FGF7
GDF5
GREM1
TEK
Log2 (Fold Change)
8
-8
B
Condition-Dependent Gene Upregulation (2 DPT)
FD
Shared
NEUROG2
Synergy
2 DPT
2 Fold
3 Fold
5 Fold
10 Fold
100 Fold
1,000 Fold
≥1,000 Fold
C
0.5 DPT
1 DPT
2 DPT
0.3% 0.7% 7% 5% 4% 45% 35% 2% 2%
0.5% 0.9% 7% 6% 8% 46% 23% 5% 3%
1% 2% 5% 4% 5% 43% 28% 4% 6%
FD N NFD
2-5 Fold
5-10 Fold
≥10 Fold
D Immediate Downregulation (0.5 DPT)
-Log10(p value)
20
18
16
14
12
10
8
6
4
2
0
Mitosis
Nuclear Division
Organelle Fission
M Phase of Mitotic Cell Cycle
Organelle Organization
E Early Upregulation (2 DPT)
-Log10(p value)
9
8
7
6
5
4
3
2
1
0
Nervous System Development
Cellular Process
Regulation of Neuron Differentiation
Response to Protein Stimulus
Regulation of Nervous System Development
0.5 DPT
1 DPT
2 DPT

Figure 5: Immediate-Early Changes in Global Transcription Induced by NFD

(A) Hierarchical clustering and heatmap of changes in genome-wide transcription detected by RNA-seq relative to GFP-transduced control fibroblasts (n=3 biological replicates for each condition and each time point; FD, fibroblasts exposed to FD; N, *NEUROG2*-transduced fibroblasts; NFD, *NEUROG2*-transduced fibroblasts exposed to FD).

(B) Pie diagram representing the respective number of genes induced by NEUROG2, FD, and synergistic action in actively converting cells at 2 DPT. Synergy represents genes induced only in the NFD condition.

(C) Time-course representation of the change in transcript expression intensity for genes regulated by FD, NEUROG2 or NFD synergy.

(D) The five most significant gene ontology terms that represent the genes sets immediately repressed by NFD treatment at 0.5 DPT. The significances of each functional classification at 0.5, 1, and 2 DPT are compared to represent the progressive or transient enrichment of individual ontologies.

(E) The five most significant gene ontology terms representing the genes sets upregulated by NFD treatment at 2 DPT.

*Portions of figure reproduced from Smith et al. 2016

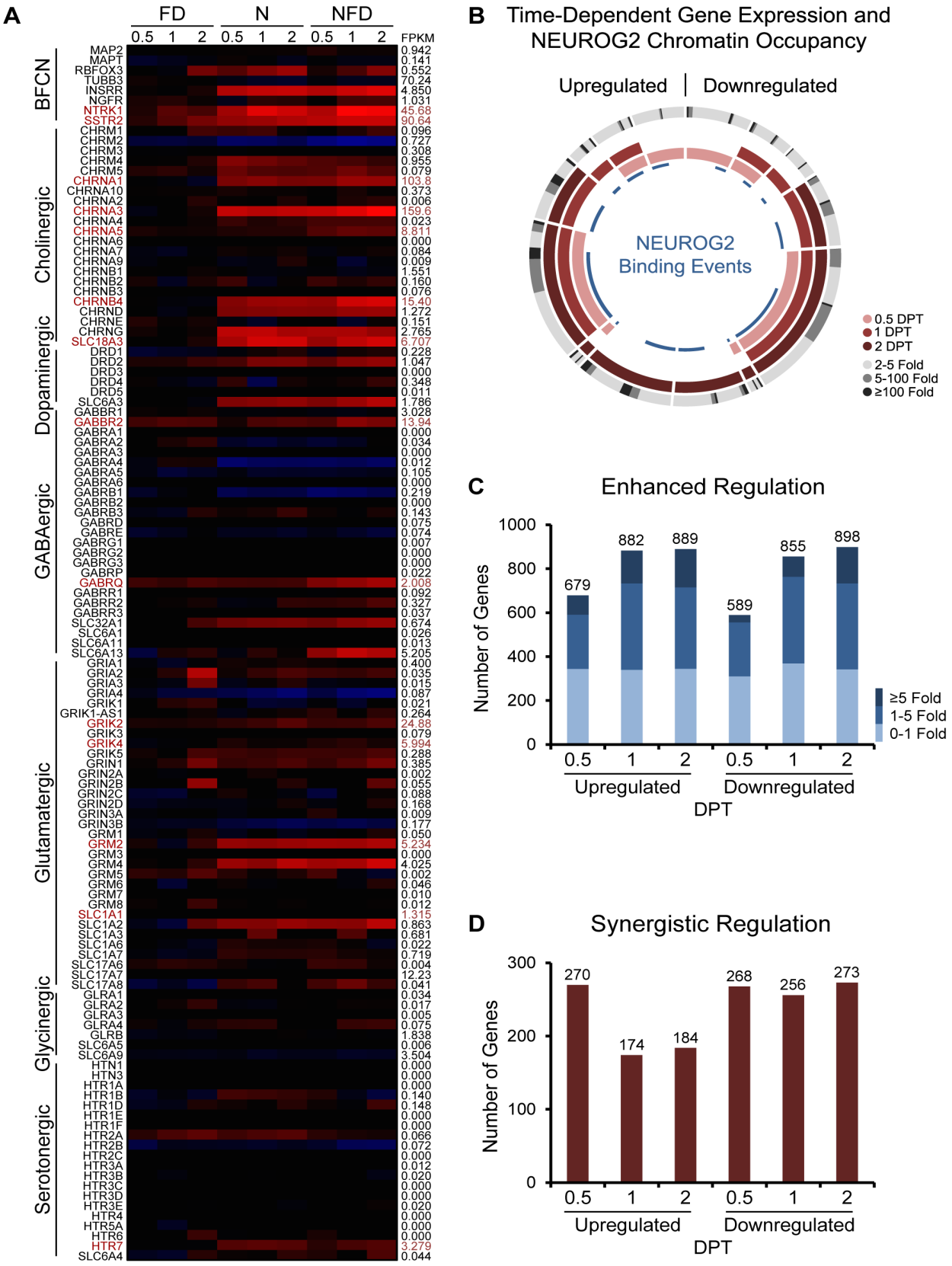

A
FD
N
NFD
0.5 1 2 0.5 1 2 0.5 1 2
FPKM
BFCN
Cholinergic
Dopaminergic
GABAergic
Glutamatergic
Glycinergic
Serotonergic
MAP2 0.942
MAPT 0.141
RBFOX3 0.552
TUBB3 70.24
INSRR 4.850
NGFR 1.031
NTRK1 45.68
SSTR2 90.64
CHRM1 0.096
CHRM2 0.727
CHRM3 0.308
CHRM4 0.955
CHRM5 0.079
CHRNA1 103.8
CHRNA10 0.373
CHRNA2 0.006
CHRNA3 159.6
CHRNA4 0.023
CHRNA5 8.811
CHRNA6 0.000
CHRNA7 0.084
CHRNA9 0.009
CHRNB1 1.551
CHRNB2 0.160
CHRNB3 0.076
CHRNB4 15.40
CHRND 1.272
CHRNE 0.151
CHRNG 2.765
SLC18A3 6.707
DRD1 0.228
DRD2 1.047
DRD3 0.000
DRD4 0.348
DRD5 0.011
SLC6A3 1.786
GABBR1 3.028
GABBR2 13.94
GABRA1 0.000
GABRA2 0.034
GABRA3 0.000
GABRA4 0.012
GABRA5 0.105
GABRA6 0.000
GABRB1 0.219
GABRB2 0.000
GABRB3 0.143
GABRD 0.075
GABRE 0.074
GABRG1 0.007
GABRG2 0.000
GABRG3 0.000
GABRP 0.022
GABRQ 2.008
GABRR1 0.092
GABRR2 0.327
GABRR3 0.037
SLC32A1 0.674
SLC6A1 0.026
SLC6A11 0.013
SLC6A13 5.205
GRIA1 0.400
GRIA2 0.035
GRIA3 0.015
GRIA4 0.087
GRIK1 0.021
GRIK1-AS1 0.264
GRIK2 24.88
GRIK3 0.079
GRIK4 5.994
GRIK5 0.288
GRIN1 0.385
GRIN2A 0.002
GRIN2B 0.055
GRIN2C 0.088
GRIN2D 0.168
GRIN3A 0.009
GRIN3B 0.177
GRM1 0.050
GRM2 5.234
GRM3 0.000
GRM4 4.025
GRM5 0.002
GRM6 0.046
GRM7 0.010
GRM8 0.012
SLC1A1 1.315
SLC1A2 0.863
SLC1A3 0.681
SLC1A6 0.022
SLC1A7 0.719
SLC17A6 0.004
SLC17A7 12.23
SLC17A8 0.041
GLRA1 0.034
GLRA2 0.017
GLRA3 0.005
GLRA4 0.075
GLRB 1.838
SLC6A5 0.006
SLC6A9 3.504
HTN1 0.000
HTN3 0.000
HTR1A 0.000
HTR1B 0.140
HTR1D 0.148
HTR1E 0.000
HTR1F 0.000
HTR2A 0.066
HTR2B 0.072
HTR2C 0.000
HTR3A 0.012
HTR3B 0.020
HTR3C 0.000
HTR3D 0.000
HTR3E 0.020
HTR4 0.000
HTR5A 0.000
HTR6 0.000
HTR7 3.279
SLC6A4 0.044
B
Time-Dependent Gene Expression and NEUROG2 Chromatin Occupancy
Upregulated
Downregulated
NEUROG2 Binding Events
0.5 DPT
1 DPT
2 DPT
2-5 Fold
5-100 Fold
≥100 Fold
C
Enhanced Regulation
Number of Genes
1000
800
600
400
200
0
679
882
889
589
855
898
0.5 1 2 0.5 1 2
Upregulated
Downregulated
DPT
≥5 Fold
1-5 Fold
0-1 Fold
D
Synergistic Regulation
Number of Genes
300
200
100
0
270
174
184
268
256
273
0.5 1 2 0.5 1 2
Upregulated
Downregulated
DPT

Figure 6: Enhanced and Synergistic Transcription Promotes Neuron Identity

(A) Heatmap representing the change in expression of subtype-specific neuron genes detected by RNA-seq relative to GFP-transduced control fibroblasts. Genes exhibiting at least 2-fold enrichment, $\log_2$(NFD FPKM) ≥ 1, and p value ≤ 0.05 are labeled red. (BFCN, basal forebrain cholinergic neuron; FD, fibroblasts exposed to FD; FPKM, fragments per kilobase per million reads for NFD 2 DPT RNA-seq replicates; N, NEUROG2-transduced fibroblasts; NFD, *NEUROG2*-transduced fibroblasts exposed to FD).

(B) Sunburst diagram depicting upregulated and downregulated genes at each time point, the intensity of expression change at 2 DPT, and corresponding NEUROG2 ChIP-seq annotation data. Each segment of the sunburst plot represents one collection of time points. For example, the section with pink, red, and deep red segments represents genes with significant differential expression at all time points, while the segment with pink alone represents genes differentially expressed only at 0.5 DPT. The outer gradient represents the intensity of change in expression observed for the genes in each time segment. The inner blue marks represent the fraction of genes in each time segment with annotated NEUROG2 binding events.

(C) Time-course representation of GREAT-annotated NEUROG2-bound genes that exhibit enhanced transcriptional upregulation or downregulation upon FD exposure relative to *NEUROG2*-transduced fibroblasts not exposed to FD. The vertical axis indicates the total number of genes with enhanced regulation determined from RNA-seq datasets. Each bar is subdivided to illustrate the intensity of change in transcription upon FD exposure.

(D) The total number of GREAT-annotated NEUROG2-bound genes induced or repressed only upon FD exposure.

*Portions of figure reproduced from Smith et al. 2016

CHAPTER 4

NEUROG2 HETERODIMER COMPOSITION AND TRANSCRIPTIONAL CO-ACTIVATORS

The observation that small molecules initiate potent changes in transcription and significantly enhance NEUROG2 chromatin binding led us to propose four hypothetical mechanisms underlying NEUROG2-FD synergy (Figure 7). These mechanisms involve FD-activated kinases that target NEUROG2 or NEUROG2 heterodimers to modify DNA binding kinetics, the activation of NEUROG2 co-factors or coordinating transcription factors, and modification of the epigenetic landscape to promote accessibility to pro-neural genetic elements.

4.1. Cyclic AMP and Protein Kinase A

Forskolin is a well-established activator of cAMP synthesis. To evaluate whether increased intracellular cAMP promotes NEUROG2-mediated reprogramming, forskolin was replaced during transdifferentiation with cAMP or dibutyryl cAMP, a cell-permeable PRKACA activator. These treatments similarly promoted NEUROG2-mediated reprogramming (Figure 4A,B); however, these molecules generated a significantly smaller population of neurons potentially due to lower thermostability in culture medium. Higher concentrations and frequencies of cAMP infusion enhanced neuron survival but were not sufficient to induce transdifferentiation on the same scale as forskolin.

High intracellular concentrations of cAMP stimulate protein kinase A activity by catalyzing the release of PRKACA from an associated regulatory subunit. To define the roles for this kinase in reprogramming, NEUROG2 was co-expressed with a constitutively active isoform of PRKACA in MRC-5 fibroblasts in the presence of dorsomorphin. This combination induced

rapid neurite outgrowth and produced TUBB3- and MAP2-positive cells within 12 days (Figure 4C). Importantly, these neurons exhibited slightly larger cell bodies than neurons generated by forskolin and total population of neurons generated by this co-expression method was significantly smaller.

4.2. NEUROG2 Phosphorylation

NEUROG2 phosphorylation by cyclin dependent kinases and GSK3 has been demonstrated to regulate NEUROG2 DNA affinity (Ali et al. 2011), neuron differentiation (Hindley et al. 2012), and motor neuron specification (Ma et al. 2008). To identify whether FD activates these kinases or PRKACA to increase NEUROG2 phosphorylation during reprogramming, we performed site-directed mutagenesis to introduce phospho-deficient and phosphomimetic amino acids at serine residues with a high probability for phosphorylation (Figure 8D,E). Surprisingly, alanine or aspartate replacement of S24, S91, S193, S207, S209, S219, S232, S239 or S242 had no effect on reprogramming efficiency or speed of conversion. Even fibroblasts transduced with lentivirus encoding a *NEUROG2* construct with seven combined substitutions exhibited no differences in reprogramming efficiency, conversion speed, morphological complexity or TUBB3 immunoreactivity. Although not an exhaustive analysis, this mutagenesis screen suggests that FD-induced phosphorylation of NEUROG2 is not a likely mechanism for the enhanced NEUROG2 chromatin binding observed in prior ChIP-seq datasets.

4.3. NEUROG2 Heterodimer Composition, Phosphorylation, and Chromatin Affinity

Homodimeric and heterodimeric NEUROG2 complexes exhibit variable chromatin affinity and functional stability (Bronicki et al. 2012, Henke et al. 2009, Li et al. 2012). Secondary

associations with LIM domain factors (Lee and Pfaff, 2003, Asprer et al. 2011), retinoic acid receptors (Lee et al. 2009), and other transcriptional activators fine-tune NEUROG2-target gene expression to define neuronal subtype identity (Ma et al. 2008).

To evaluate whether any of these interactions enhance immediate-early neuronal transcription in fibroblasts, we generated a minimal NEUROG2 truncation construct containing only the DNA binding domain and a single transactivation domain (Li et al. 2012). MRC-5 fibroblasts transduced with *NEUROG2*$_{110\text{-}228}$-encoding lentivirus and exposed to FD exhibited significant reductions in reprogramming efficiency (Figure 9A). Importantly, this reduced structure did not entirely abolish the capacity of NEUROG2 to induce neuron-like MAP2-positive cells indicating NEUROG2$_{110\text{-}228}$ is a functional but inefficient transcriptional activator.

We next used mass spectrometry to identify NEUROG2-associated co-factors and the effects of FD exposure on NEUROG2 complex composition at 1 DPT (Figure 9B). This screen revealed direct interaction between NEUROG2 and TCF3, which has been shown to modulate NEUROG2 chromatin binding dynamics in a phosphorylation-dependent manner (Li et al. 2012). To investigate whether forskolin-induced PRKACA activity might target factors that heterodimerize with NEUROG2, we co-expressed PRKACA with NEUROG2, TCF3 isoform E12, TCF3 isoform E47, TCF4, and TCF12. Electrophoretic analysis revealed that PRKACA catalyzes phosphatase-sensitive TCF3, TCF4, and TCF12 phosphorylation, but does not target recombinant NEUROG2 (Figure 9C).

We next performed a comprehensive analysis of phosphorylation-dependent changes in NEUROG2 heterodimer affinity for sequence-specific enhancer boxes. We utilized RNA-seq datasets from NFD-treated fibroblasts to identify the 100 most upregulated and downregulated genes with NEUROG2-annotated ChIP-seq binding events. Genomic sequences were extracted

from 284 occupancy sites annotated to upregulated genes and 192 occupancy sites annotated to downregulated genes then directionally searched to identify the distribution of enhancer box sequences at these sites (Figure 9D). Three unique sequences, 5'-CAGCTG-3', 5'-CAGATG-3', and 5'-CATCTG-3', were identified with similar distributions of enhancer box motifs detected for upregulated and downregulated genes.

Phosphorylation-dependent binding of NEUROG2 heterodimers to each of these enhancer box sequences was next investigated using high-throughput microscale thermophoresis (Timofeeva et al. 2012). DNA binding-induced changes in the thermophoresis of recombinant amino-terminal GFP-labeled NEUROG2 enabled us to quantitatively define the affinity of NEUROG2 and TCF3 (E47 isoform), TCF3 (E12 isoform), TCF4 or TCF12 heterodimers for sequence-specific enhancer box motifs. A preliminary electrophoretic mobility shift assay (EMSA) was performed to ensure GFP fusion did not affect the ability of NEUROG2 to heterodimerize or associate with DNA (Figure 9E). This foundational assay confirmed that NEUROG2 homodimers do not bind DNA (Bronicki et al. 2012), GFP fusion does not affect NEUROG2 association with TCF3 (E47 isoform), and NEUROG2 efficiently binds a ChIP-identified *DLL3* enhancer box *in vitro*. Next, thermophoretic traces were collected for each NEUROG2-E protein heterodimer using a genomic *DLL3* 5'-CAGCTG-3' enhancer box sequence and this same sequence with the internal dinucleotide pair mutagenized to encode 5'-CAGATG-3' and 5'-CATCTG-3' sequences. Interestingly, NEUROG2-TCF3 (E47 isoform) heterodimers associated with all three enhancer box sequences with modest differences in sequence-specific affinity (Figure 10). TCF3 (E12 isoform) preferentially associated with 5'-CAGATG-3' and 5'-CATCTG-3' enhancer boxes, TCF4 exhibited low affinity (K_d: 20-40 μM) for 5'-CATCTG-3' upon PRKACA-induced phosphorylation, and association of NEUROG2-

TCF12 heterodimers was not detected for any of the three sequences. This suggests NEUROG2 selectively heterodimerizes with TCF3 to bind DNA.

To investigate whether TCF3 phosphorylation affects heterodimer affinity for DNA, PRKACA-phosphorylated recombinant TCF3 was incubated with NEUROG2 and DNA binding assessed by microscale thermophoresis. Surprisingly, phosphorylation induced sequence-specific reductions in DNA affinity with greater than 4-fold lower affinity for the 5'-CAGCTG-3' enhancer box motif (Figure 10). Among several possibilities, this result suggests FD-induced TCF3 phosphorylation might be a mechanism for enhancing the specificity and resolution of NEUROG2 binding in a sequence-dependent manner requiring coordination with additional factors to stabilize NEUROG2 heterodimer association at specific motifs (Figure 13F related).

4.4. CREB1 Transcriptional Program and Chromatin Co-Occupancy with NEUROG2

To investigate whether PRKACA might also function to change the expression of reprogramming-induced genes not targeted by NEUROG2, we performed sequence-based discovery of regulons using RNA-seq datasets for upregulated genes without NEUROG2 ChIP-seq annotation (Janky et al. 2014). Within a 10-kilobase window centered at the transcription start site, cAMP responsive element motifs were detected adjacent to 79 of 111 FD-induced genes. CREB1 is an established target of PRKACA phosphorylation and interacts with this FD-enriched motif (Gonzalez et al. 1989). High-efficiency short hairpin RNA (shRNA)-mediated knockdown of CREB1 significantly reduced the survival of NFD-treated fibroblasts undergoing active reprogramming (Figure 11A,B). This suggests forskolin initiates a CREB1-driven transcriptional program that promotes neuron survival.

To identify the genetic targets of activated CREB1, we performed ChIP using a CREB1 antibody specific for PRKACA-mediated phosphorylation at serine 133 (Figure 11C). We confirmed functional IP using the established CREB1 target genes *NR4A1* (Impey et al. 2004) and *SIK1*, which also exhibited FD-induced upregulation in RNA-seq datasets. Exposure to FD enhanced CREB1 enrichment at these sites and, more interestingly, at CRE elements in direct overlap with NEUROG2 occupancy sites in the enhancer regions of *GADD45G* and *SOX4* (Figure 11D). To determine whether these co-binding events occur on a genome-wide scale, the co-binding of NEUROG2 and CREB1 was assessed using chromatin co-IP (Figure 11E). Immunoblot of NEUROG2-bound chromatin fragments exposed positive for CREB1 protein and CREB1-bound chromatin fragments validated this result with NEUROG2 detected in NFD-treated cells but not control fibroblasts.

Growth arrest and DNA damage inducible (GADD45) alpha, beta, and gamma are three related proteins with diverse roles in p38/JNK-mediated apoptosis (Takekawa et al. 1998), DNA demethylation, and neurogenesis (Ma et al. 2009). CREB1 co-binding potentiates NEUROG2-driven *GADD45A* and *GADD45G* expression to 4-fold and 139-fold induction at 0.5 DPT, respectively (Figure 12A,B). Therefore, we performed gene knockdown using efficient *GADD45A*- (Figure 12C) and *GADD45G*-targeted (Figure 12D) shRNA constructs, which even in combination did not affect transdifferentiation efficiency, time of conversion or morphological complexity of induced neurons (Figure 12E,F). While the GADD45 family executes critical activity-dependent roles in functionally mature neurons, these knockdown results suggest that the earliest events in neuron specification might rely on key modulators of transcription such as SOX4 (Figure 5).

4.5. SOX4 Chromatin Occupancy Analysis and Roles in Chromatin Remodeling

To identify whether SOX4 upregulation is essential to early reprogramming, we performed gene knockdown using two SOX4-specific shRNA constructs expressed in NFD-treated fibroblasts (Figure 13A). Disrupted SOX4 expression nearly abolished reprogramming potential with only 7±1% of fibroblasts activating MAP2 expression in the most stringent condition (Figure 13B,C). In addition, striking defects in morphological conversion occurred in the absence of SOX4 (Figure 13C).

To define the genetic targets of SOX4, we performed ChIP-seq using crosslinked chromatin isolated in triplicate at 1 DPT from NFD-treated MRC-5 fibroblasts overexpressing SOX4. A comparison of genome-wide SOX4 and NEUROG2 binding profiles revealed broad overlap with 413 of 1,119 SOX4 binding events also detected in NEUROG2 ChIP-seq datasets (Figure 13D,E). Sequence-specific analysis of the enhancer boxes neighboring SOX4-bound HMG motifs revealed a strong enrichment of CAGCTG motifs suggesting NEUROG2-TCF3(E47) heterodimers preferentially associate with SOX4-bound chromatin (Figure 13F).

To identify whether co-binding events affect gene expression, peak-to-gene annotation was used to identify upregulated genes targeted by SOX4 and NEUROG2. Relative to NEUROG2 alone, FD-induced SOX4 binding enhanced transcription for 81% of co-bound genes such as *NEUROD1* and *NEUROD4* (Figure 13E). A functional classification of these targets revealed an enrichment of genes associated with RNA transcription, cell cycle regulation, and chromatin remodeling (Figure 13G). The SWI/SNF complex is a multiunit nucleosome remodeling complex essential to neuron specification and maturation (Kwon et al. 1994, Son et al. 2004). SOX4 targets at least nine components of this complex including *ARID1A*, *ARID1B*, *SMARCA2*, *SMARCAD1*, *SMARCC1*, *SMARCC2*, *SMARCD1*, *SMARCD2*, and *SMARCE1*. This not only

provides SOX4 with genetic regulatory functions via direct chromatin binding, but additional indirect chromatin remodeling functions which increase accessibility to neurogenic heterochromatin regions for pro-neural transcription factors targeted by NEUROG2 and SOX4.

These indirect remodeling functions were intensively investigated using *SOX4* knockdown coupled with ATAC-seq in NFD-treated fibroblasts. NEUROG2 binding and FD-induced chromatin decondensation at the *NEUROD1* and *NEUROD4* loci drives strong upregulation of these genes during reprogramming. However, knockdown of *SOX4* in NFD-treated fibroblasts significantly reduced open chromatin signatures in the regulatory regions of both genes (Figure 14A). Expanding this analysis to a genome-wide view, *SOX4* knockdown results in the loss of 51,123 ATAC-seq peaks or approximately 59% of total open chromatin sites (Figure 14B). In addition, 22% of peaks detected in both conditions have an increased peak score in the presence of SOX4 (Figure 14B). To define the importance of newly open chromatin sites during reprogramming, ChIP-qPCR analysis of potential NEUROD1 and NEUROD4 binding sites was used to probe bidning within or neighboring regions of newly opened chromatin (Figure 14C). To define putative binding sites, publicly available NEUROD1 ChIP-seq datasets from mouse cortical tissue were analyzed to identify potential regulatory sites in direct overlap with human NEUROG2 binding sites or ATAC-seq identified open chromatin (Pataskar et al. 2016). For instance, NEUROD1 enrichment was observed at NEUROG2 occupancy sites in the regulatory regions of *DLL3*, *HES6*, and *NHLH1* (Figure 14C). Interestingly, NEUROD1 also exhibited modest enrichment in the NEUROD4 promoter at an open chromatin site lost upon *SOX4* knockdown. Similarly, NEUROD4 was modestly enriched at these neurogenic genetic elements, but to a significantly lower extent than NEUROD1 (Figure 14C).

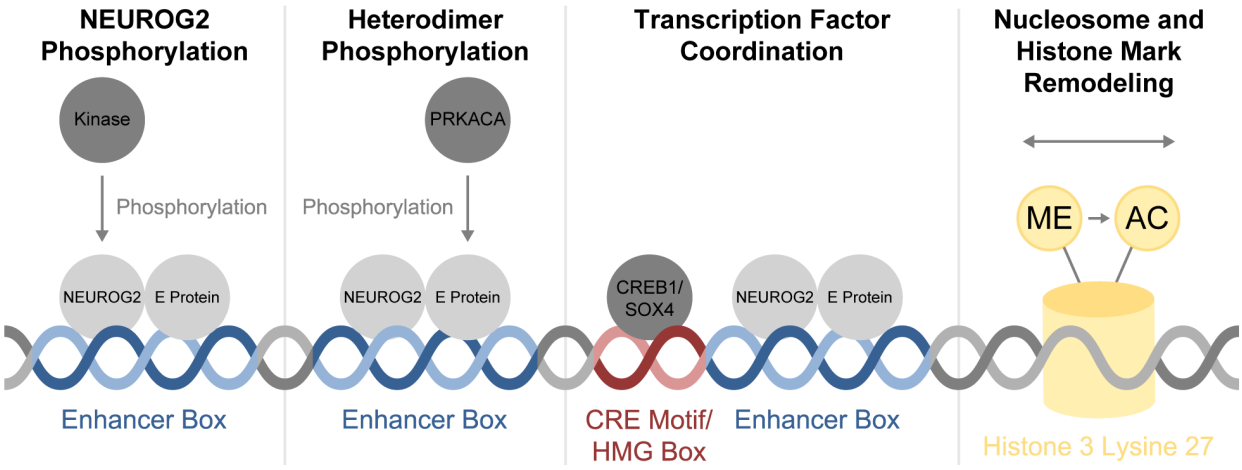


Figure 7: Hypothetical Mechanisms for NEUROG2-FD Synergy

Four hypothetical mechanisms might underlie NEUROG2-FD synergy. First, FD-activated kinases might target NEUROG2 to modify chromatin association kinetics or, second, might similarly target NEUROG2 heterodimers. Third, FD might induce the expression or activation of NEUROG2 co-factors or coordinating transcription factors that promote neurogenic transcription. Finally, FD might modify the epigenetic landscape to promote accessibility at pro-neural genetic elements.

*Figure reproduced from Smith et al. 2016

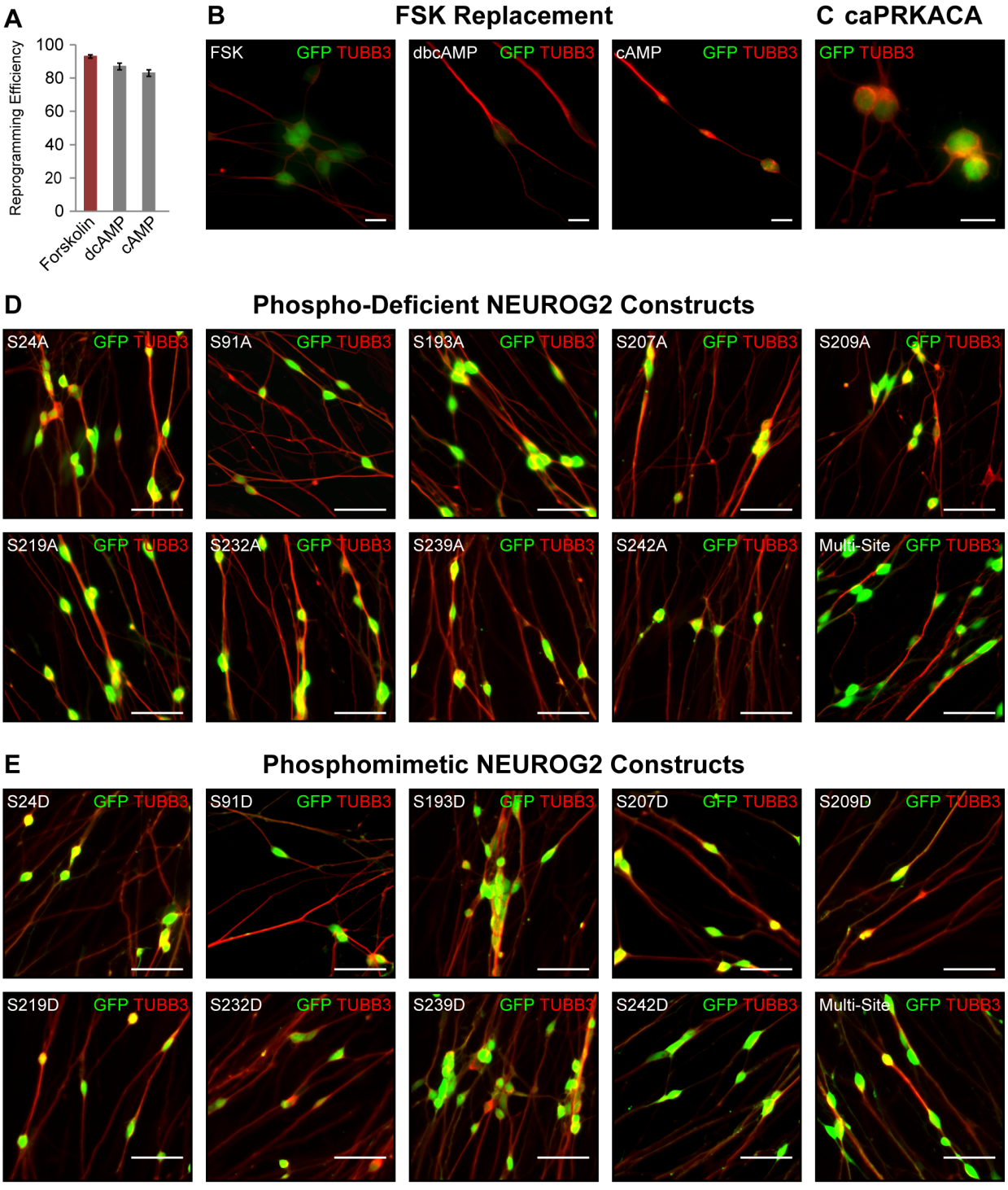

A
Reprogramming Efficiency
100
80
60
40
20
0
Forskolin
dcAMP
cAMP
B
FSK Replacement
C caPRKACA
FSK
GFP TUBB3
dbcAMP
GFP TUBB3
cAMP
GFP TUBB3
GFP TUBB3
D
Phospho-Deficient NEUROG2 Constructs
S24A
GFP TUBB3
S91A
GFP TUBB3
S193A
GFP TUBB3
S207A
GFP TUBB3
S209A
GFP TUBB3
S219A
GFP TUBB3
S232A
GFP TUBB3
S239A
GFP TUBB3
S242A
GFP TUBB3
Multi-Site
GFP TUBB3
E
Phosphomimetic NEUROG2 Constructs
S24D
GFP TUBB3
S91D
GFP TUBB3
S193D
GFP TUBB3
S207D
GFP TUBB3
S209D
GFP TUBB3
S219D
GFP TUBB3
S232D
GFP TUBB3
S239D
GFP TUBB3
S242D
GFP TUBB3
Multi-Site
GFP TUBB3

Figure 8: Forskolin Replacement and NEUROG2 Phosphorylation

(A) Reprogramming efficiency for *NEUROG2*-transduced cells treated with dorsomorphin and cAMP synthesis activators. Efficiency is calculated as (GFP-positive TUBB3-positive cells) / (total GFP-positive cell population).

(B) Dorsomorphin-treated *NEUROG2*-transduced fibroblasts reprogram to TUBB3-positive neurons when treated with 10 μM forskolin, 10 μM dibutyryl cAMP, and 10 μM cAMP for 12 days. Reductions in reprogramming efficiency and diminished morphological complexity might be the result of reduced stability of dibutyryl cAMP and cAMP in neuronal culture media relative to forskolin, scale: 20 μm.

(C) Representative TUBB3 staining of fibroblast-derived neurons 12 days after lentiviral delivery of NEUROG2 and caPRKACA, scale: 25 μm.

(D-E) Site-directed mutagenesis of NEUROG2 at putative GSK3- and PRKACA-targeted phosphorylation sites. (D) Alanine-substitution models permanent dephosphorylation, while (E) phosphomimetic aspartate-substitution models constitutive phosphorylation. Multi-site designates a NEUROG2 construct containing substitutions at S193, S207, S209, S219, S232, S239, and S242. None of the serine substitutions modified the ability of NEUROG2 to induce TUBB3-positive neurons from fibroblasts at 14 DPI, scale: 50 μm.

*Portions of figure reproduced from Smith et al. 2016

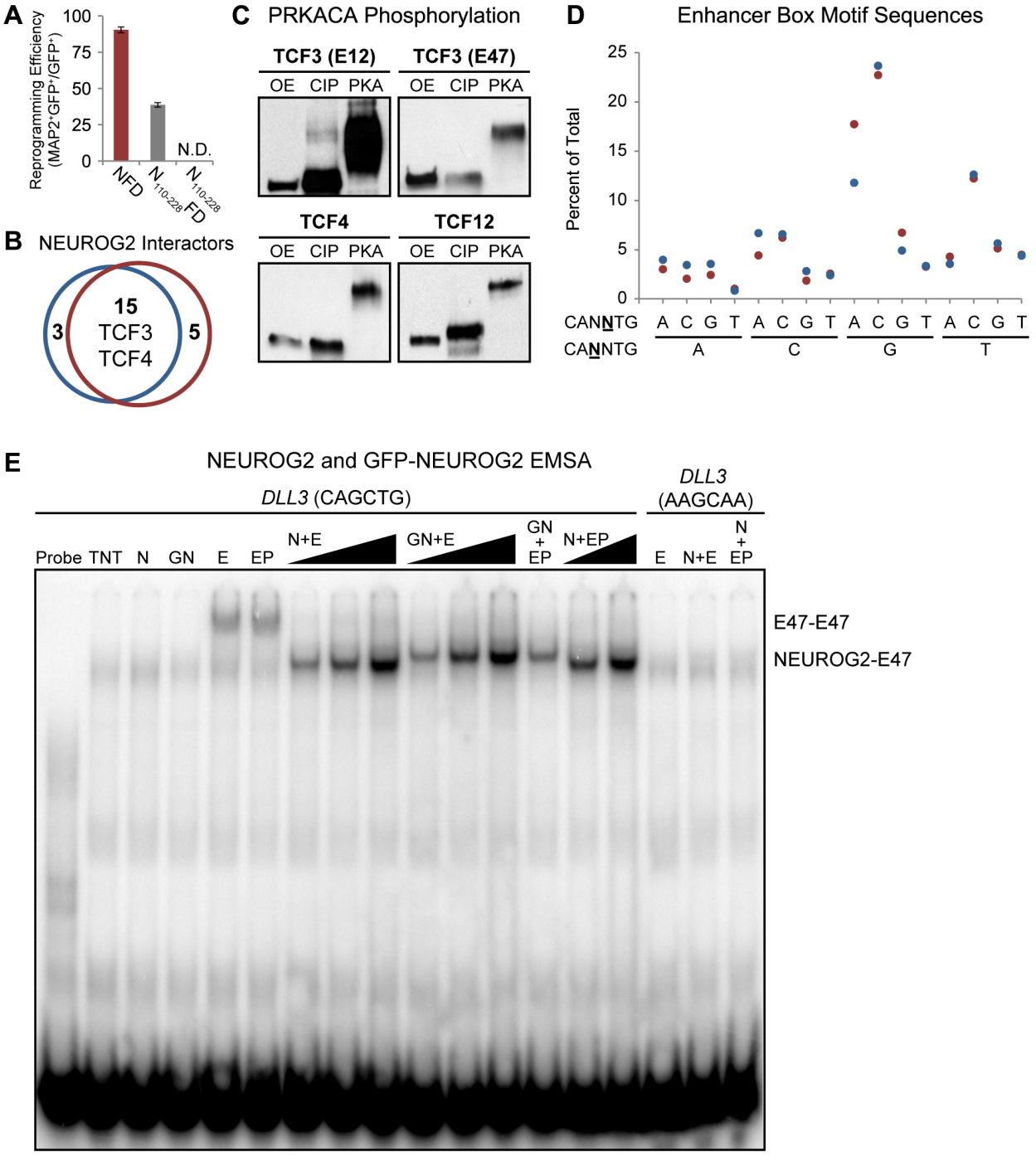

A
Reprogramming Efficiency (MAP2+GFP+/GFP+)
100
75
50
25
0
N.D.
NFD
N 110-228 FD
N 110-228
B
NEUROG2 Interactors
15
TCF3
TCF4
3
5
C
PRKACA Phosphorylation
TCF3 (E12)
TCF3 (E47)
TCF4
TCF12
OE CIP PKA
D
Enhancer Box Motif Sequences
Percent of Total
25
20
15
10
5
0
CANNTG A C G T A C G T A C G T A C G T
CANNTG A C G T
E
NEUROG2 and GFP-NEUROG2 EMSA
DLL3 (CAGCTG)
DLL3 (AAGCAA)
Probe TNT N GN E EP N+E GN+E GN + EP N+EP E N+E N + EP
E47-E47
NEUROG2-E47

Figure 9: PRKACA Phosphorylates Co-Factors Essential to NEUROG2 Function

(A) Reprogramming efficiency for *NEUROG2*$_{110\text{-}228}$-transduced fibroblasts. Efficiency was calculated as (GFP-positive MAP2-positive cells) / (total GFP-positive cell population).

(B) NEUROG2-interacting proteins identified in the absence (blue) and presence (red) of FD by mass spectrometry.

(C) Analysis of PRKACA-mediated E protein phosphorylation by denaturing polyacrylamide gel electrophoresis. TCF3 (E12 isoform), TCF3 (E47 isoform), TCF4, and TCF12 were overexpressed (OE), co-overexpressed with PRKACA (PKA), and co-overexpressed with PRKACA then incubated with calf intestinal alkaline phosphatase (CIP). TCF3 isoform E12 is moderately overexposed to demonstrate partial CIP dephosphorylation and the corresponding within sample mobility shift.

(D) Distribution of enhancer box motif sequences identified within NEUROG2-target sequences for the 100 most significantly upregulated genes (red, 284 binding events) and 100 most significantly downregulated genes (blue, 192 binding events).

(E) EMSA illustrating NEUROG2 and GFP-NEUROG2 binding at an enhancer box upstream of *DLL3* (chr19:39,988,668-39,988,693). Probe, ^{32}P-labeled double-stranded probe only; TNT, T7 reticulocyte lysate expressing control vector; N, recombinant NEUROG2; GN, recombinant GFP-NEUROG2; E, recombinant TCF3 isoform E47; EP, recombinant TCF3 isoform E47 incubated with PRKACA. Filled ramps represent increased recombinant protein per reaction.

*Figure modified from Smith et al. 2016

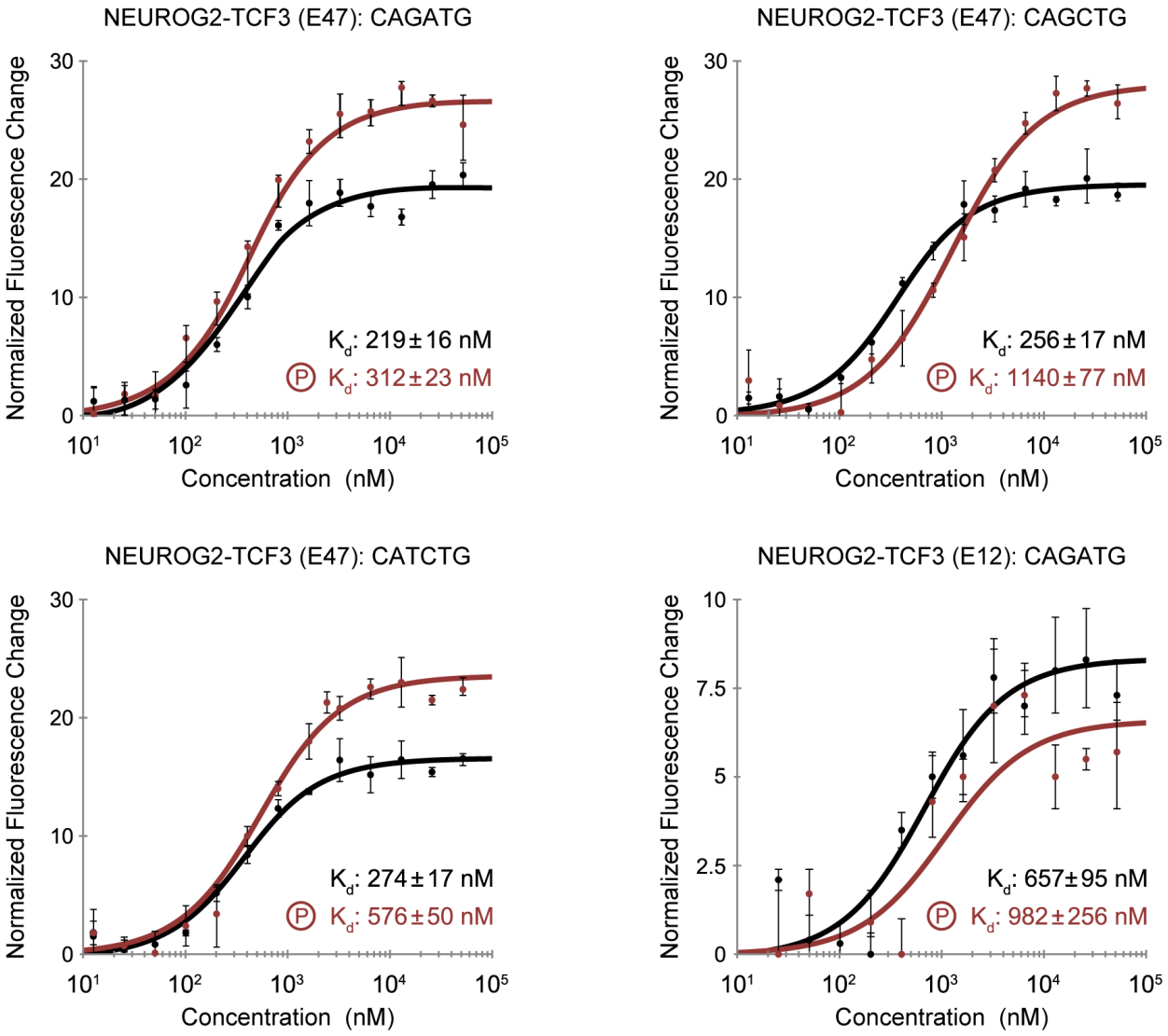


Figure 10: NEUROG2 Heterodimers Exhibit Sequence-Specific Chromatin Affinity

Thermophoretic binding curves depicting the effect of phosphorylation (red) on the affinity of GFP-NEUROG2 heterodimers for sequence-specific enhancer box motifs.

*Figure modified from Smith et al. 2016

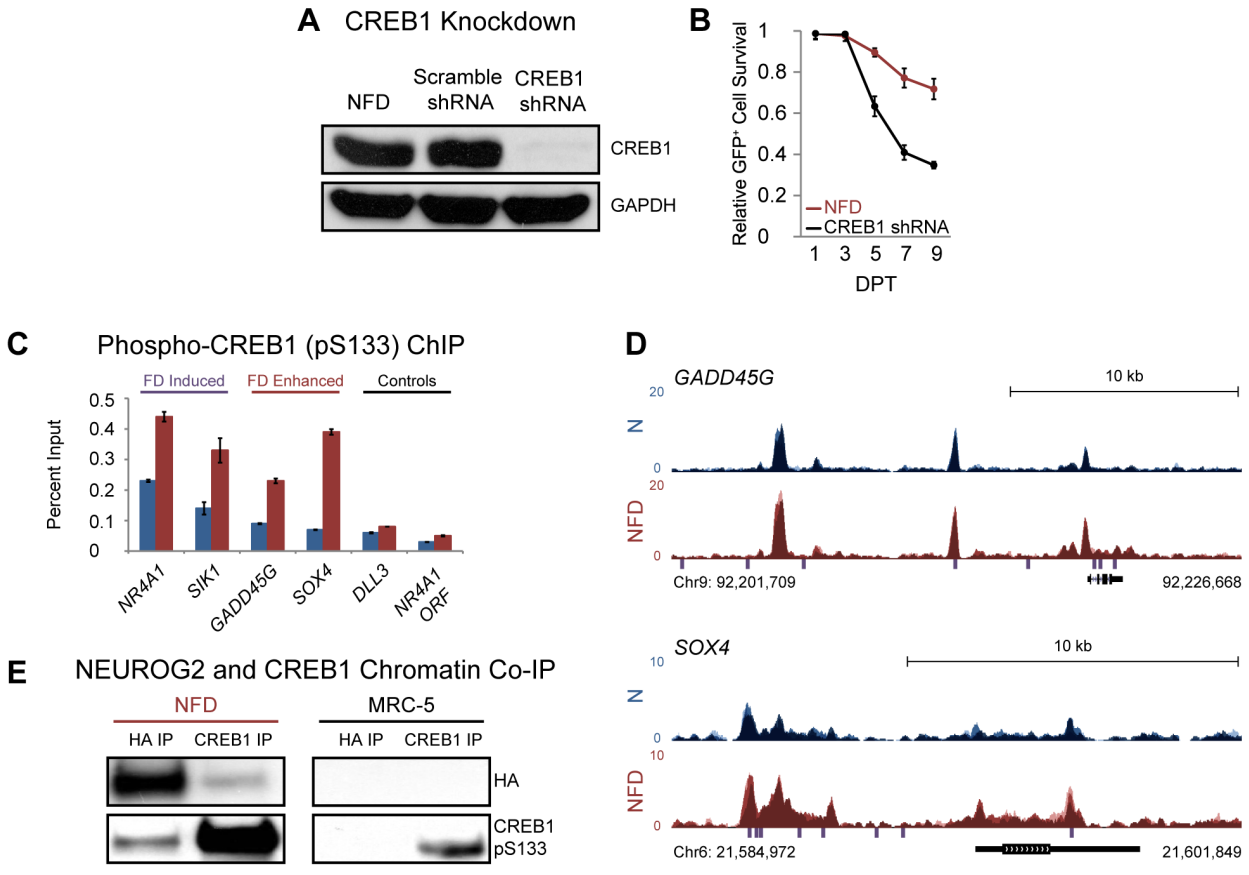

A
CREB1 Knockdown
NFD
Scramble shRNA
CREB1 shRNA
CREB1
GAPDH
B
Relative GFP+ Cell Survival
1
0.8
0.6
0.4
0.2
0
NFD
CREB1 shRNA
1 3 5 7 9
DPT
C
Phospho-CREB1 (pS133) ChIP
FD Induced
FD Enhanced
Controls
Percent Input
0.5
0.4
0.3
0.2
0.1
0
NR4A1
SIK1
GADD45G
SOX4
DLL3
NR4A1 ORF
D
GADD45G
10 kb
N
NFD
20
0
Chr9: 92,201,709
92,226,668
SOX4
10 kb
10
0
Chr6: 21,584,972
21,601,849
E
NEUROG2 and CREB1 Chromatin Co-IP
NFD
MRC-5
HA IP
CREB1 IP
HA
CREB1 pS133

Figure 11: CREB1 Promotes Neuron Survival and Co-Binds with NEUROG2

(A) Immunoblot validation of shRNA-mediated CREB1 knockdown.

(B) CREB1 knockdown reduces the total number of surviving induced neurons. Relative survival was calculated in triplicate as (GFP-positive Hoechst 33342-positive cells) / (total GFP-positive Hoechst 33342-positive cells at 1 DPT).

(C) Phospho-CREB1 (pS133) ChIP-qPCR at known CREB1-regulated genes (*NR4A1*, *SIK1*), NEUROG2-target genes with enhanced expression upon FD exposure (*GADD45G*, *SOX4*), and controls (NEUROG2 target, *DLL3*; within gene, *NR4A1* open reading frame) for *NEUROG2*-transduced fibroblasts (blue) and *NEUROG2*-transduced fibroblasts exposed to FD (red) at 1 DPT. Error is represented as the standard deviation of ΔC_T.

(D) NEUROG2 occupancy around the *GADD45G* and *SOX4* loci. NEUROG2 occupancy is represented as overlays of 0.5 (light), 1 (medium), and 2 DPT (dark) binding event traces. Purple hashes indicate CRE motif sequences. The genomic scale is 10 kilobases (kb).

(E) Chromatin co-immunoprecipitation (co-IP) of NEUROG2 and CREB1 from 2 DPT hemagluttinin (HA) tagged-*NEUROG2*-transduced fibroblasts exposed to FD. NEUROG2 and CREB1 are both detectable when chromatin is immunoprecipitated with either HA or CREB1 (pS133) antibodies demonstrating chromatin co-occupancy.

*Figure modified from Smith et al. 2016

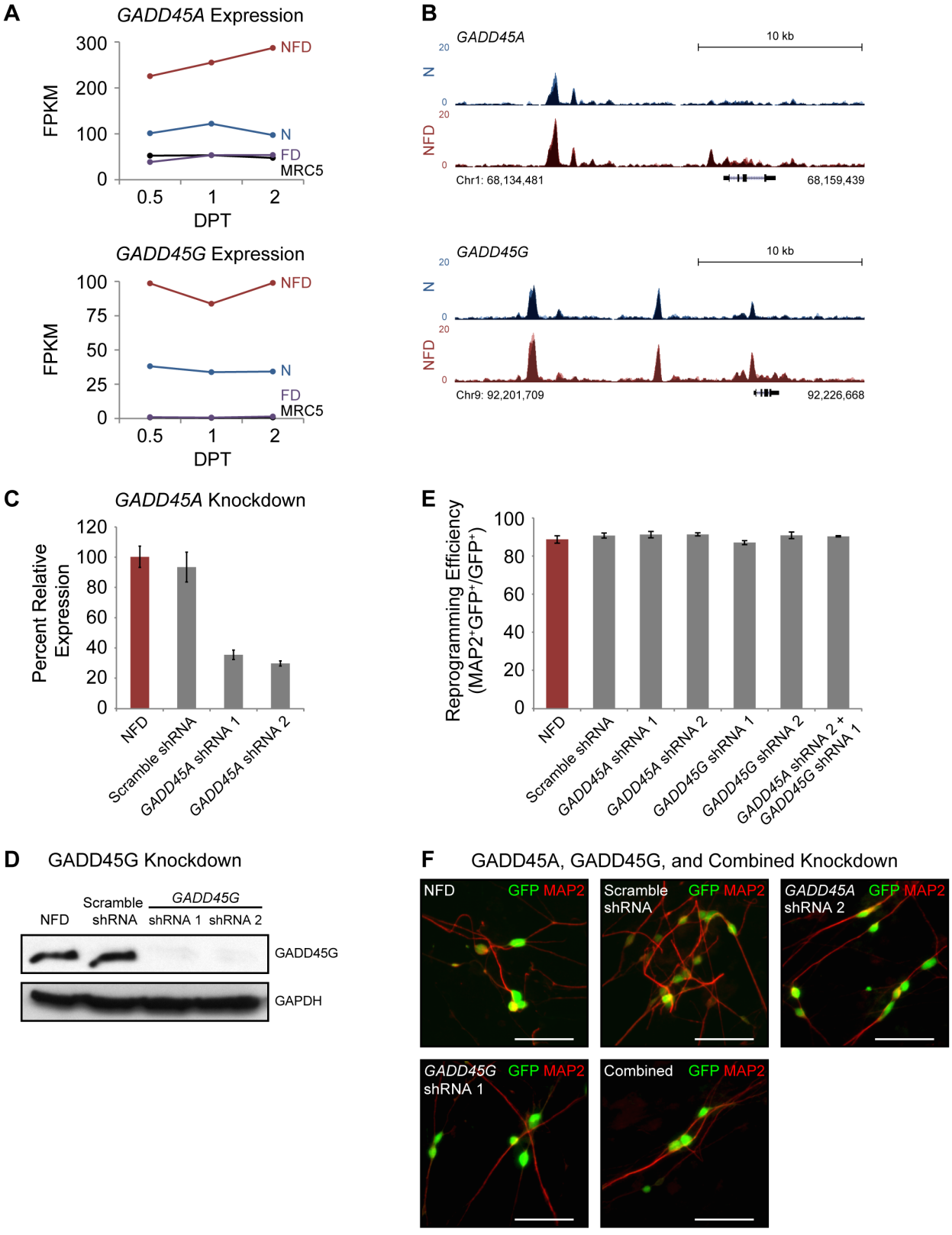

A
GADD45A Expression
FPKM
300
200
100
0
NFD
N
FD
MRC5
0.5
1
2
DPT
GADD45G Expression
FPKM
100
75
50
25
0
NFD
N
FD
MRC5
0.5
1
2
DPT
B
GADD45A
10 kb
N
NFD
20
0
Chr1: 68,134,481
68,159,439
GADD45G
10 kb
Chr9: 92,201,709
92,226,668
C
GADD45A Knockdown
Percent Relative Expression
120
100
80
60
40
20
0
NFD
Scramble shRNA
GADD45A shRNA 1
GADD45A shRNA 2
E
Reprogramming Efficiency (MAP2+GFP+/GFP+)
100
80
60
40
20
0
NFD
Scramble shRNA
GADD45A shRNA 1
GADD45A shRNA 2
GADD45G shRNA 1
GADD45G shRNA 2
GADD45A shRNA 2 + GADD45G shRNA 1
D
GADD45G Knockdown
NFD
Scramble shRNA
GADD45G
shRNA 1
shRNA 2
GADD45G
GAPDH
F
GADD45A, GADD45G, and Combined Knockdown
NFD
GFP MAP2
Scramble shRNA
GFP MAP2
GADD45A shRNA 2
GFP MAP2
GADD45G shRNA 1
GFP MAP2
Combined
GFP MAP2

Figure 12: GADD45A and GADD45G are Not Essential to Transdifferentiation

(A) FPKM of *GADD45A* and *GADD45G* transcript expression from RNA-seq datasets. *GADD45B* expression was repressed in reprogrammed neurons and was not evaluated for roles in reprogramming (FD, fibroblasts exposed to FD; N, *NEUROG2*-transduced fibroblasts; NFD, *NEUROG2*-transduced fibroblasts exposed to FD).

(B) NEUROG2 occupancy around the *GADD45A* and *GADD45G* loci. NEUROG2 occupancy in *NEUROG2*-transduced (N) and NFD-treated (NFD) fibroblasts is represented as overlays of 0.5 (light), 1 (medium), and 2 DPT (dark) binding event traces.

(C) qPCR validation of shRNA-mediated *GADD45A* knockdown in NFD-treated fibroblasts at 5 DPT. *HPRT1*-normalized knockdown efficiency was calculated from triplicate samples using the $\Delta\Delta C_T$ method.

(D) Immunoblot validation of shRNA-mediated *GADD45G* knockdown using *NEUROG2*-transfected 293T/17 cells 3 days after transfection.

(E) *GADD45A* and *GADD45G* knockdown do not affect the reprogramming potential of NFD-treated fibroblasts. Reprogramming efficiency was calculated as (GFP-positive MAP2-positive cells) / (total GFP-positive cell population) at 12 DPT. (F) Representative MAP2 staining of fibroblast-derived neurons treated with *GADD45A*-targeted shRNA and *GADD45G*-targeted shRNA. Combined knockdown of both *GADD45A* and *GADD45G* had no effect on reprogramming potential, scale: 50 μm.

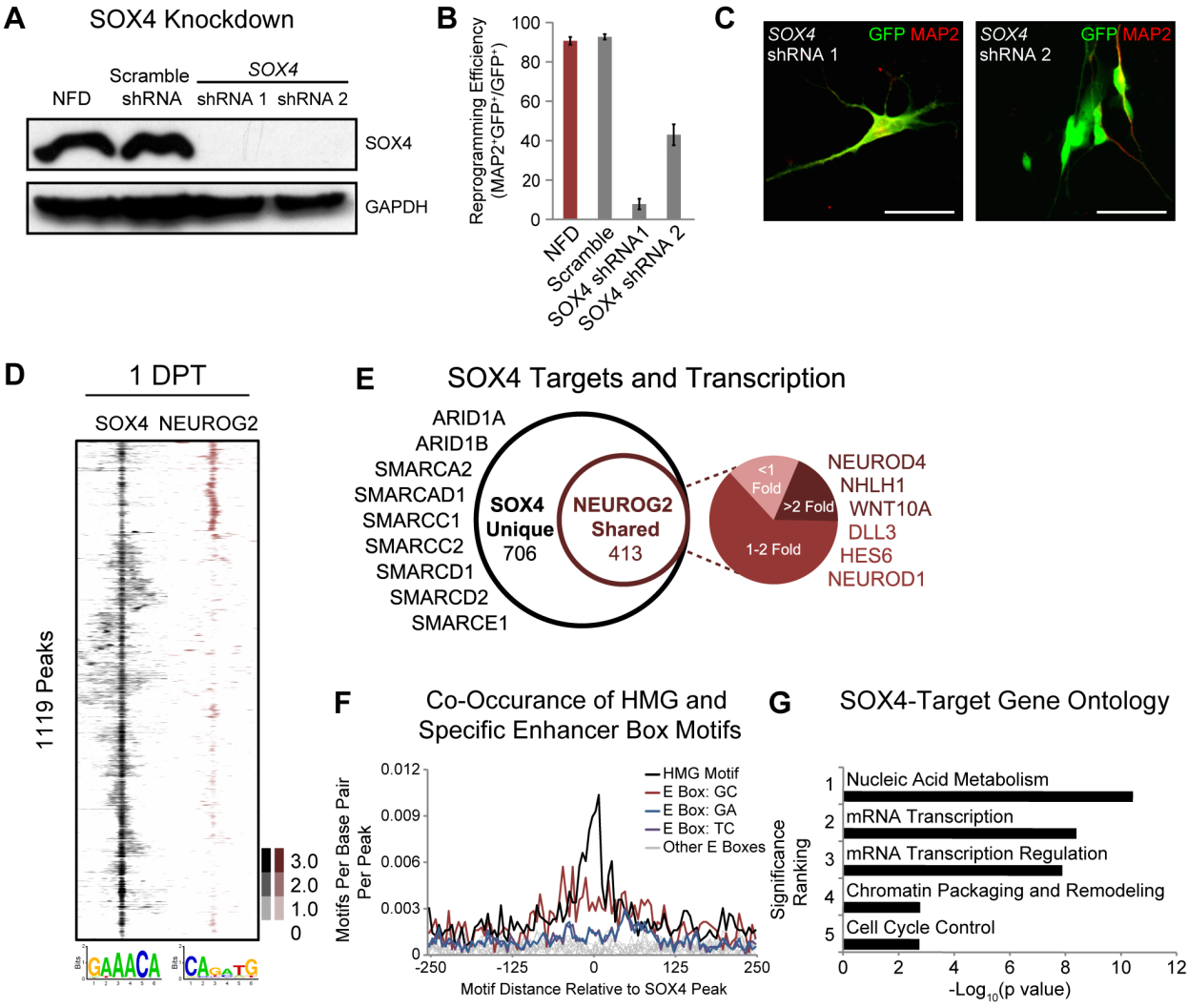

A
SOX4 Knockdown
Scramble
SOX4
NFD
shRNA
shRNA 1
shRNA 2
SOX4
GAPDH
B
Reprogramming Efficiency
(MAP2+GFP+/GFP+)
100
80
60
40
20
0
NFD
Scramble
SOX4 shRNA1
SOX4 shRNA 2
C
SOX4
shRNA 1
GFP MAP2
SOX4
shRNA 2
GFP MAP2
D
1 DPT
SOX4 NEUROG2
1119 Peaks
3.0
2.0
1.0
0
E
SOX4 Targets and Transcription
ARID1A
ARID1B
SMARCA2
SMARCAD1
SMARCC1
SMARCC2
SMARCD1
SMARCD2
SMARCE1
SOX4
Unique
706
NEUROG2
Shared
413
<1
Fold
>2 Fold
1-2 Fold
NEUROD4
NHLH1
WNT10A
DLL3
HES6
NEUROD1
F
Co-Occurance of HMG and
Specific Enhancer Box Motifs
Motifs Per Base Pair
Per Peak
0.012
0.009
0.006
0.003
0
HMG Motif
E Box: GC
E Box: GA
E Box: TC
Other E Boxes
-250
-125
0
125
250
Motif Distance Relative to SOX4 Peak
G
SOX4-Target Gene Ontology
Significance
Ranking
1
2
3
4
5
Nucleic Acid Metabolism
mRNA Transcription
mRNA Transcription Regulation
Chromatin Packaging and Remodeling
Cell Cycle Control
0
2
4
6
8
10
12
-Log10(p value)

Figure 13: SOX4 Knockdown and Genome-Wide Occupancy Analysis

(A) Immunoblot validation of shRNA-mediated *SOX4* knockdown.

(B) Reprogramming efficiency for *SOX4* shRNA-transduced fibroblasts. Efficiency was calculated as (GFP-positive MAP2-positive cells) / (total GFP-positive cell population).

(C) Representative MAP2 staining of NFD-treated fibroblasts 14 days after co-expression of *SOX4* shRNAs, scale: 50 μm.

(D) Heatmaps representing SOX4 genome-wide occupancy and corresponding NEUROG2 occupancy at those genomic coordinates. Input-normalized tag densities are plotted in $\log_2$ scale and centered in a 4-kilobase window around the peak apex (n=3 biological replicates for SOX4).

(E) Area-proportional Venn diagram that depicts the total number of SOX4 binding events and proportion of occupancy sites that directly overlap with NEUROG2 occupancy sites. Neighboring pie diagram represents the relative change in transcription of genes co-bound by SOX4 and NEUROG2 in *NEUROG2*-transduced MRC-5 fibroblasts in the presence and absence of FD. Fold increase represents the relative change in normalized transcript number detected in RNA-seq datasets.

(F) The distribution of sequence-specific enhancer box motifs neighboring HMG binding motifs within the 1119 SOX4-bound peaks.

(G) The five most significantly enriched gene ontology classifications from SOX4-targeted genes.

*Portions of figure reproduced from Smith et al. 2016

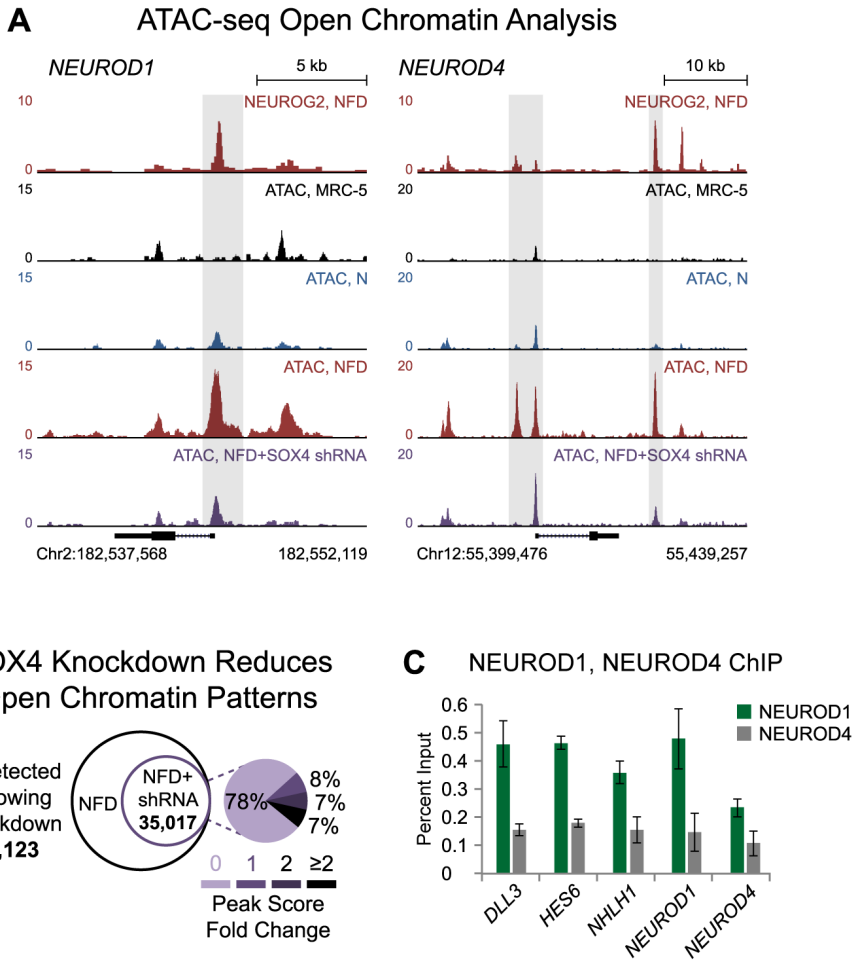

A
ATAC-seq Open Chromatin Analysis
NEUROD1
5 kb
NEUROD4
10 kb
NEUROG2, NFD
ATAC, MRC-5
ATAC, N
ATAC, NFD
ATAC, NFD+SOX4 shRNA
Chr2:182,537,568
182,552,119
Chr12:55,399,476
55,439,257
B
SOX4 Knockdown Reduces Open Chromatin Patterns
Undetected Following Knockdown 51,123
NFD
NFD+ shRNA 35,017
78%
8%
7%
7%
0 1 2 ≥2
Peak Score Fold Change
C
NEUROD1, NEUROD4 ChIP
Percent Input
NEUROD1
NEUROD4
DLL3
HES6
NHLH1
NEUROD1
NEUROD4

Figure 14: SOX4 Enhances Chromatin Accessibility During Transdifferentiation

(A) NEUROG2 ChIP-seq track (2 DPT NFD) and ATAC-seq tracks generated from fibroblasts (MRC-5), *NEUROG2*-transduced fibroblasts (N), *NEUROG2*-transduced fibroblasts exposed to FD (NFD), and fibroblasts co-transduced with *NEUROG2* and *SOX4 shRNA* then exposed to FD (NFD) at the *NEUROD1* and *NEUROD4* loci. Gray shading highlights regions representing FD-enhanced chromatin accessibility that is lost upon *SOX4* knockdown.

(B) An area-proportional Venn diagram that depicts the total number of ATAC-seq events in NFD-treated fibroblasts relative to the number of events detected at the same locations upon *SOX4* knockdown. The neighboring pie diagram represents the relative change in peak score for sites shared in both conditions (NFD peak score / NFD+*SOX4* shRNA peak score).

(C) NEUROD1 and NEUROD4 ChIP-qPCR at sites of newly open chromatin or significant transcriptional upregulation. Error is represented as the standard deviation of ΔC_T.

*Figure modified from Smith et al. 2016

CHAPTER 5

ADULT HUMAN SKIN FIBROBLAST-TO-NEURON TRANSDIFFERENTIATION

5.1. Mouse and Human SOXC Factors Exhibit Differential Reprogramming Kinetics

The time required for neuronal conversion varies significantly for adult fibroblasts transduced with mouse-derived *Sox11* (Liu et al. 2013) and human-derived *SOX4* or *SOX11*. The rate of genetic and morphological reprogramming induced by mouse SOX11 is rapid with cells adopting a rounded morphology at 3 DPT, extending branched neurites between 5-7 DPT, and high-signal TUBB3 expression detectable at 10 DPT (Liu et al. 2013). Comparatively, transdifferentiation of adult fibroblasts by human SOX4 or SOX11 is relatively slow with initial rounding (5-10 DPT), extension of low complexity neurites (10-20 DPT), and activation of the early neuronal marker TUBB3 (14 DPT) proceeding at a significantly reduced pace relative to cells expressing mouse *Sox11*. Multiple protein sequence alignment revealed 88% similarity between mouse and human SOX11 homologues with regions of divergence immediately 3' to the HMG DNA binding domain. Differences in DNA binding kinetics or recognition of these regions by repressor proteins might account for this temporal difference in conversion rate.

Early ChIP-seq and RNA-seq datasets from transdifferentiating MRC-5 fibroblasts indicate SOX11 is not significantly upregulated (Figure 15A), targeted by NEUROG2 (Figure 15B) or essential to the early reprogramming of MRC-5 fibroblasts (Figure 15C-E). This indicates overexpressed SOX11 might, in part, perform physiological roles that overlap with SOX4 function to promote neuronal transdifferentiation.

Interestingly, adult fibroblasts exhibited comparatively lower basal *SOX4* expression than fetal MRC-5 fibroblasts (Figure 16A). To investigate whether SOX4 overexpression is sufficient

to replace SOX11 in adult skin fibroblast reprogramming, adult fibroblasts were transduced with *SOX4*- and *NEUROG2*-encoding lentivirus in combination with FD and FGF2 treatment. These factors induced TUBB3 expression and slow neurite outgrowth in four adult fibroblast lines with 69% to 83% efficiency (Figure 16B,C).

A time-course analysis of SOX4 chromatin binding in AG05811 adult fibroblasts provided insight into the varied genetic landscapes of adult and fetal fibroblast lines. While significant binding events were detected by ChIP-seq at 1 DPT in MRC-5 fibroblasts, these same sites were not significantly enriched until 4 DPT in aged fibroblasts (Figure 16D). This suggests limited access to densely compact chromatin regions might inhibit the rapid activation of neurogenic transcription programs by reprogramming factors and consequently slows the transdifferentiation of adult fibroblasts. Further, NFD might not be sufficient to sustain SOX4 expression at a level required to catalyze the activation of SOX4-targeted SWI/SNF genes localized to densely compact chromatin and, therefore, necessitates SOX4 overexpression to promote successful adult fibroblast transdifferentiation.

5.2. Adult Fibroblast Transdifferentiation using NEUROG2 and SWI/SNF Factors

NEUROG2, FD, and endogenous *SOX4* upregulation should be sufficient to catalyze the transdifferentiation of adult fibroblasts if functional chromatin interactions can be achieved. Since SOX4 targets numerous subunits of the SWI/SNF complex in MRC-5 fibroblasts, *SMARCA4* and *NEUROG2* were co-expressed to determine whether chromatin remodeling could potentiate adult cell reprogramming (Figure 16E). A fraction of transduced fibroblasts rounded, outgrew neurites, and exhibited modest TUBB3 expression. However, similar to SOX4 overexpression, these fibroblasts did not adopt highly complex neuronal morphology. The

further co-expression of SWI/SNF subunits, *SMARCB1* and *SMARCC2*, with *NEUROG2* and *SMARCA4* failed to further improve conversion. As precise stoichiometry of multiple SWI/SNF complex subunits is required for efficient function, we instead turned to a chemical reprogramming approach to improve the efficiency of chromatin remodeling.

5.3. Adult Fibroblast Transdifferentiation using NEUROG2 and Chemicals

As demonstrated with FD in MRC-5 fibroblasts, small molecules are potent reprogramming factors with diverse target specificity. A chemical screen of compounds targeting chromatin modifiers was performed to identify whether enhancements in chromatin accessibility and epigenetic histone modifications could improve NEUROG2-mediated reprogramming in AG05811 fibroblasts (Figure 16E). Numerous chemicals induced rounding and modest neurite outgrowth, but only the class I histone deacetylase inhibitor FK228 was sufficient to rapidly induce neuron-like morphology in large populations of fibroblasts (Figure 16F). Remarkably, high-efficiency conversion occurred within 20-24 hours of exposure. However, cells treated with 1 μM FK228 failed to survive beyond 5 DPT in culture. The addition of BDNF, GDNF, NT-3, and 1 μM kenpaullone at the outset of reprogramming, as well as, dosage- and time-dependent optimizations improved cell survival. Treatment with neurotrophic factors, 1 μM kenpaullone , and 0.5 μM FK228 promoted strong TUBB3 expression at 3 DPT (figure 16G). This chemical approach to chromatin modification and SOX4 replacement demonstrates that NEUROG2 and small molecules are sufficient to induce neurogenic programs in fibroblasts refractory to reprogramming in the absence of epigenetic barriers to activity.

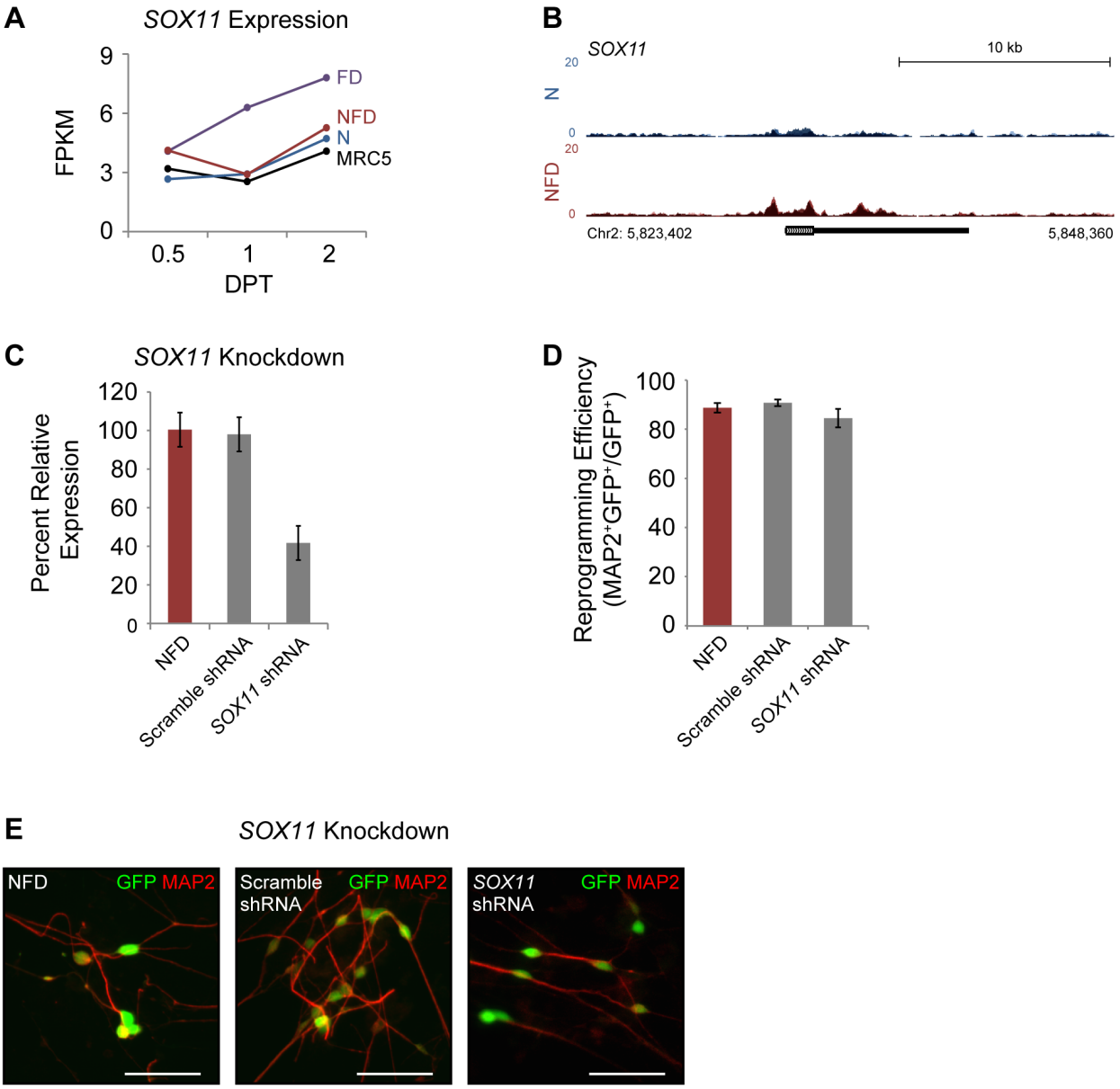

A
SOX11 Expression
FPKM
9
6
3
0
0.5
1
2
DPT
FD
NFD
N
MRC5
B
SOX11
10 kb
N
NFD
20
0
Chr2: 5,823,402
5,848,360
C
SOX11 Knockdown
Percent Relative Expression
120
100
80
60
40
20
0
NFD
Scramble shRNA
SOX11 shRNA
D
Reprogramming Efficiency (MAP2+GFP+/GFP+)
100
80
60
40
20
0
NFD
Scramble shRNA
SOX11 shRNA
E
SOX11 Knockdown
NFD
GFP MAP2
Scramble shRNA
GFP MAP2
SOX11 shRNA
GFP MAP2

Figure 15: SOX11 is Not Required for Immediate-Early Transdifferentiation

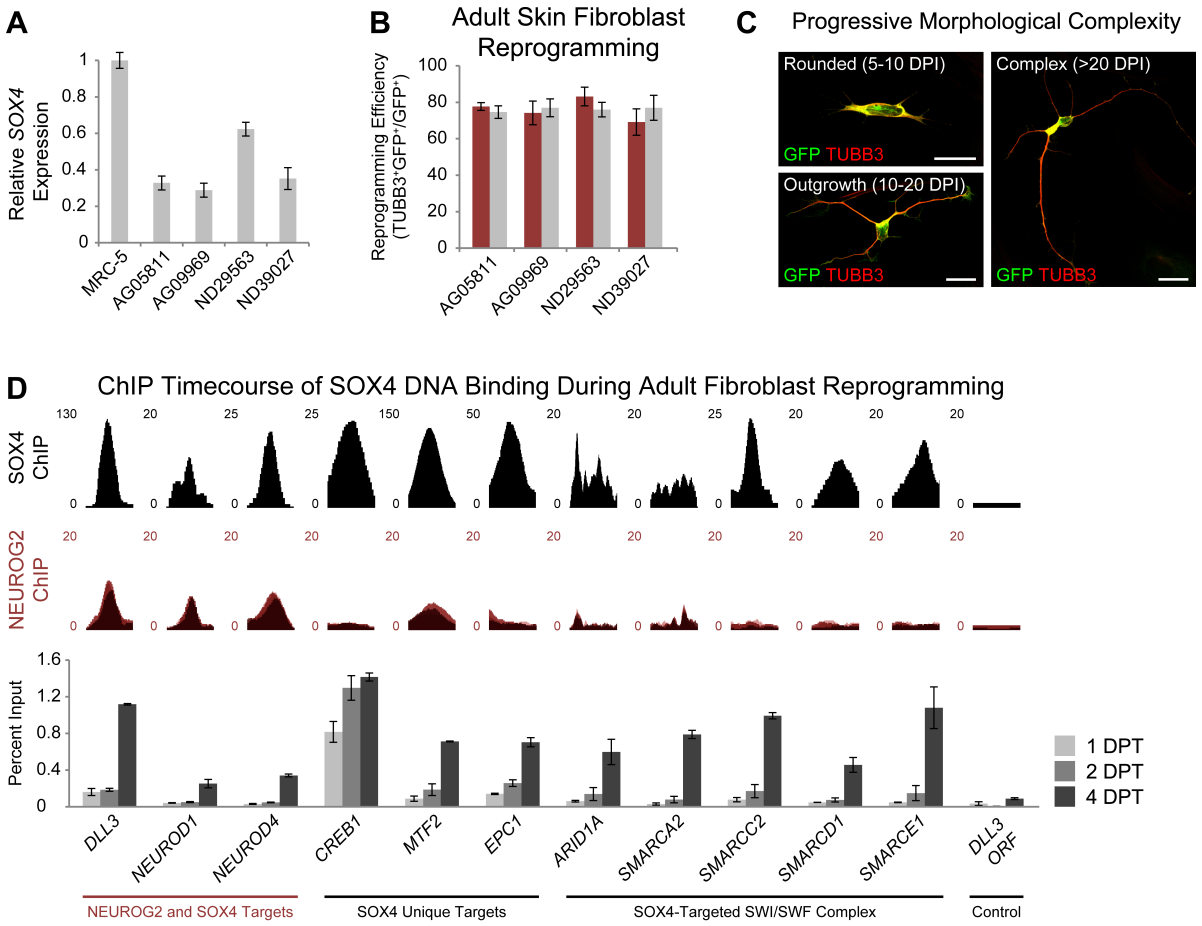
(A) *SOX11* FPKM from RNA-seq datasets (FD, fibroblasts exposed to FD; N, *NEUROG2*-transduced fibroblasts; NFD, *NEUROG2*-transduced fibroblasts exposed to FD).

(B) NEUROG2 exhibits limited occupancy around the *SOX11* locus. NEUROG2 occupancy in *NEUROG2*-transduced (N) and NFD-treated (NFD) fibroblasts is represented as overlays of 0.5 (light), 1 (medium), and 2 DPT (dark) binding event traces.

(C) qPCR validation of shRNA-mediated *SOX11* knockdown in NFD-treated fibroblasts at 5 DPT. *HPRT*-normalized knockdown efficiency was calculated from triplicate samples using the $\Delta\Delta C_T$ method.

(D) *SOX11* knockdown does not affect the reprogramming potential of NFD-treated fibroblasts. Reprogramming efficiency was calculated as (GFP-positive MAP2-positive cells) / (total GFP-positive cell population) at 12 DPT.

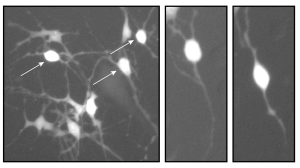
(E) Representative MAP2 staining of fibroblast-derived neurons treated with *SOX11*-targeted shRNA, scale: 50 μm.



E Chemical Screening (8 DPT)

Factor	Morphology	Rounded	Neurite	TUBB3
5-Azacitidine	-	-	-	-
Decitabine	+	+	+	-
DZNep	+	-	-	-
EPZ5676	+	+	+	-
FK228	+	+	+	*
GSK126	+	+	-	-
I-BET151	+	+	-	-
I-BET762	+	+	-	-
JQ1	+	+	+	*
Liproxstatin-1	+	+	-	-
RepSox	+	+	+	-
RG108	+	+	+	-
SAHA	+	-	-	-
SGI1027	-	-	-	-
SMARCA4	+	+	+	+
UNC669	+	+	+	-
Valproic acid	+	-	-	-
Zebularine	+	+	+	-

*Cell death prior to TUBB3 analysis



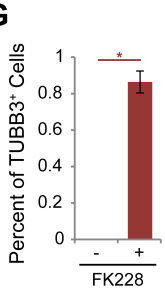


Figure 16: SOX4-Mediated Adult Fibroblast Transdifferentiation and DNA Occupancy

(A) The basal expression of *SOX4* in four adult fibroblast lines relative to MRC-5.

(B) Neuron induction efficiency for adult fibroblast lines treated with FD, FGF2, and mouse *Sox4+NEUROG2*-encoding lentivirus (red) or human *SOX4+NEUROG2*-encoding lentivirus (gray). Efficiency was calculated as (GFP-positive TUBB3-positive cells with at least one neurite) / (total GFP-positive cell population).

(C) Progressive neuronal maturation of adult fibroblasts treated with reprogramming factors. Fibroblasts adopt a rounded morphology (5-10 DPT) then outgrow low complexity neurites (10-20 DPT), which gradually mature in complexity (≥21 DPT), scale: 20 μm.

(D) ChIP-qPCR time-course analysis of SOX4 chromatin occupancy at 1, 2, and 4 DPT in adult fibroblasts treated with reprogramming factors. SOX4 (black) and NEUROG2 (red) ChIP-seq tracks depicting genomic sub-regions containing site analyzed by qPCR. Error is represented as the standard deviation of ΔC_T (ORF, open reading frame).

(E) Screen for chromatin modifying chemicals and SWI/SNF factors that promote neuronal reprogramming. The presence (+) or absence (-) of morphological change, soma rounding, neurite outgrowth, and TUBB3 expression at 8 DPT was evaluated.

(F) Exposure to FK228 for 20-24 hours induces rapid neuron-like morphology. Arrows indicate cells with advanced morphology, while partially converted cells in the bottom left are still undergoing morphological restructuring.

(G) Percent of adult fibroblasts expressing TUBB3 in the presence and absence of FK228 at 3 DPT.

*Figure modified from Smith et al. 2016

CHAPTER 6

H3K27 ACETYLATION AND TRI-METHYLATION SIGNATURES DURING TRANSDIFFERENTIATION

6.1. The Dynamics of H3K27 Acetylation and Tri-Methylation during Reprogramming

We next characterized how the intertwined genetic regulatory networks of NEUROG2, SOX4, and CREB1 collectively initiate transcription using a condition-dependent analysis of H3K27 acetylation. A quantitative analysis of H3K27AC read densities surrounding NEUROG2 genomic occupancy sites revealed that FD only modestly enhanced NEUROG2-mediated acetylation (Figure 17A). This directly correlates with DNA binding and transcription events identified in ChIP-seq and RNA-seq datasets demonstrating that NEUROG2 is sufficient to induce transcription in the absence of FD. However, FD strongly enhanced H3K27AC signal within a 400-nucleotide window centered at SOX4-targeted genomic sites (Figure 17B). This FD-induced enrichment of acetylation is even more pronounced at genomic elements for upregulated genes targeted by both NEUROG2 and SOX4 (Figure 17C). These findings suggest that NEUROG2-FD synergistic transcription could be mediated, in part, by SOX4.

Histone acetylation is primarily localized to the TSS of NEUROG2-targeted genes (Figure 17D). This promoter-centered enrichment generally precedes transcriptional activation with only the upper quartile exhibiting greater than 7-fold induction at 1 DPT (Figure 17E). Importantly, this group of early-response genes indicates that NEUROG2 acts in a targeted fashion to immediately enhance the expression of key pro-neural factors as opposed to an indiscriminate global activation of neurogenic transcription. Similar transcriptional activation schemes are also observed for SOX4 and shared gene targets.

An analysis of transcription factor binding and histone marks neighboring single genes such as *DLL3* and *SMARCC2* reveal information about both the mechanism of reprogramming, as well as, the epigenetic environment of MRC-5 fibroblasts (Figur 17F,G). SOX4 and NEUROG2 binding sites directly overlap upstream of *DLL3* and H3K27 acetylation significantly increases at this site upon NEUROG2 expression and FD treatment (Figure 17F). A simultaneous loss of H3K27 tri-methylation is also observed throughout the *DLL3* locus. Examining SMARCC2, a component of the SWI/SNF complex, we observe SOX4 binding in the promoter region in the absence of NEUROG2 (Figure 17G). Importantly, H3K27 tri-methylation is not strongly enriched an within this locus and might suggest that the epigenetic environment primes MRC-5 fibroblasts for reprogramming relative to the potentially more restrictive chromatin conformations in adult fibroblasts.

A graphical illustration of NEUROG2, SOX4, and shared target genes with corresponding transcription and histone acetylation data revealed a core set of 53 genes pivotal to the initiation of an intermediate stage of neuron specification (Figure 18). Of these, 14 genes undergo greater than 20-fold induction with pro-neural transcription factors, signaling proteins, and kinases represented. This suggests that FD activates SOX4 and CREB1 transcriptional programs that synergize with NEUROG2 to cooperatively initiate secondary neurogenic programs executed by factors such as NEUROD1, NEUROD4, and NHLH1.

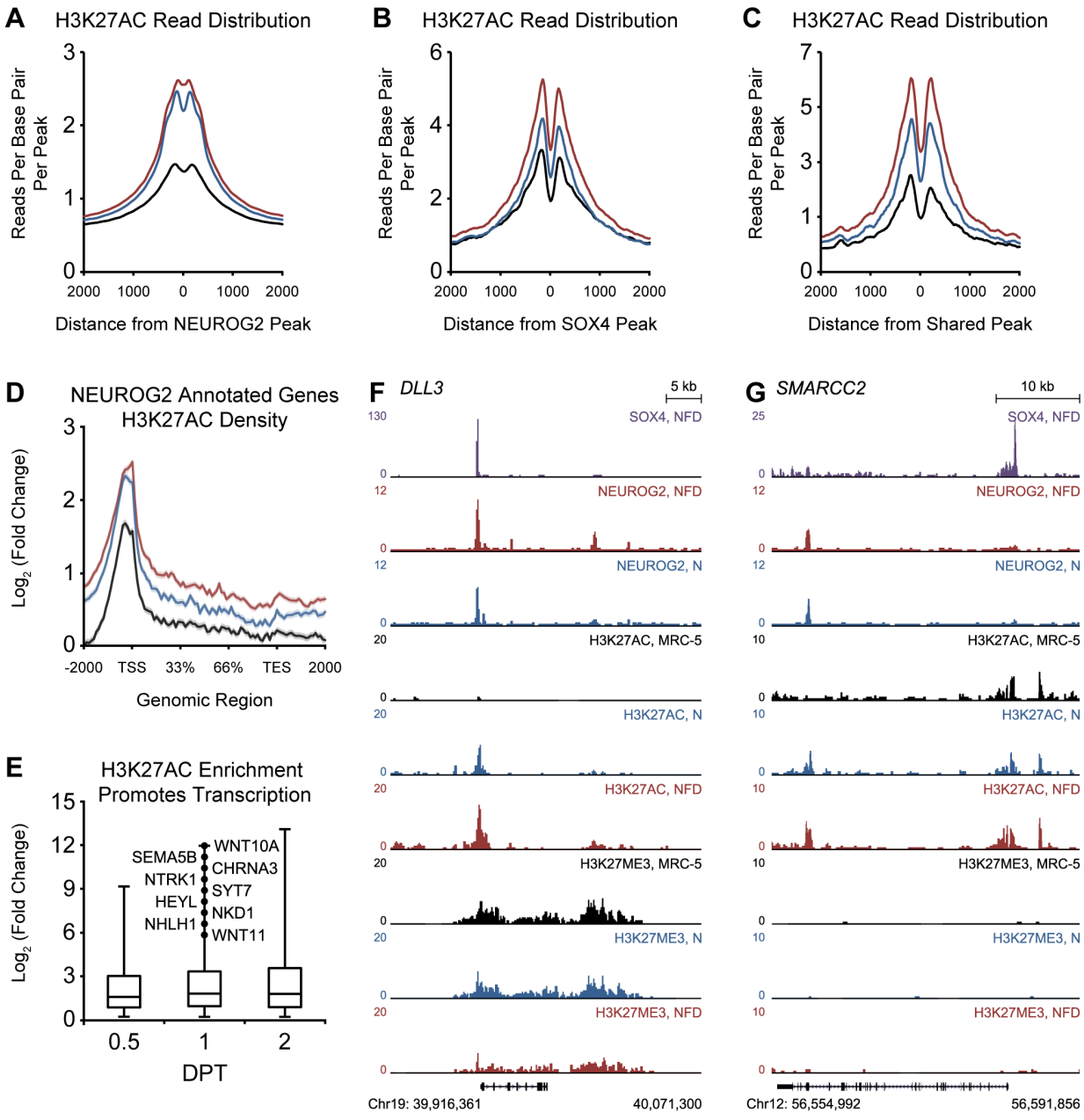

A
H3K27AC Read Distribution
Reads Per Base Pair Per Peak
Distance from NEUROG2 Peak
B
H3K27AC Read Distribution
Distance from SOX4 Peak
C
H3K27AC Read Distribution
Distance from Shared Peak
D
NEUROG2 Annotated Genes
H3K27AC Density
Log2 (Fold Change)
-2000 TSS 33% 66% TES 2000
Genomic Region
E
H3K27AC Enrichment
Promotes Transcription
SEMA5B
NTRK1
HEYL
NHLH1
WNT10A
CHRNA3
SYT7
NKD1
WNT11
DPT
F
DLL3
5 kb
SOX4, NFD
NEUROG2, NFD
NEUROG2, N
H3K27AC, MRC-5
H3K27AC, N
H3K27AC, NFD
H3K27ME3, MRC-5
H3K27ME3, N
H3K27ME3, NFD
Chr19: 39,916,361
40,071,300
G
SMARCC2
10 kb
Chr12: 56,554,992
56,591,856

Figure 17: Reprogramming Factors Modulate Histone Post-Translational Modifications

(A-C) Distribution of H3K27 acetylation (H3K27AC) ChIP-seq reads from untreated MRC-5 fibroblasts (black), *NEUROG2*-transduced fibroblasts (blue), and NFD-treated fibroblasts (red) centered around (A) NEUROG2 ChIP-seq peaks, (B) SOX4 ChIP-seq peaks, and (C) shared NEUROG2-SOX4 peaks in a 4-kilobase window.

(D) Distribution of H3K27AC ChIP-seq reads for upregulated NEUROG2-annoated genes relative to gene transcription start (TSS) and end (TES) sites for untreated fibroblasts (black), *NEUROG2*-transduced fibroblasts (blue), and NFD-treated fibroblasts (red).

(E) Box-and-whisker plot representing the fold change in expression of upregulated genes with at least 5-fold increased H3K27AC peak signal upon NFD treatment relative to control MRC-5 fibroblasts.

(F-G) ChIP-seq tracks for SOX4, NEUROG2, H3K27 acetylation, and H3K27 tri-methylation generated from fibroblasts (MRC-5), *NEUROG2*-transduced fibroblasts (N) or *NEUROG2*-transduced fibroblasts exposed to FD (NFD) encompassing the *DLL3* (F) and *SMARCC2* (G) genomic loci.

*Portions of figure reproduced from Smith et al. 2016

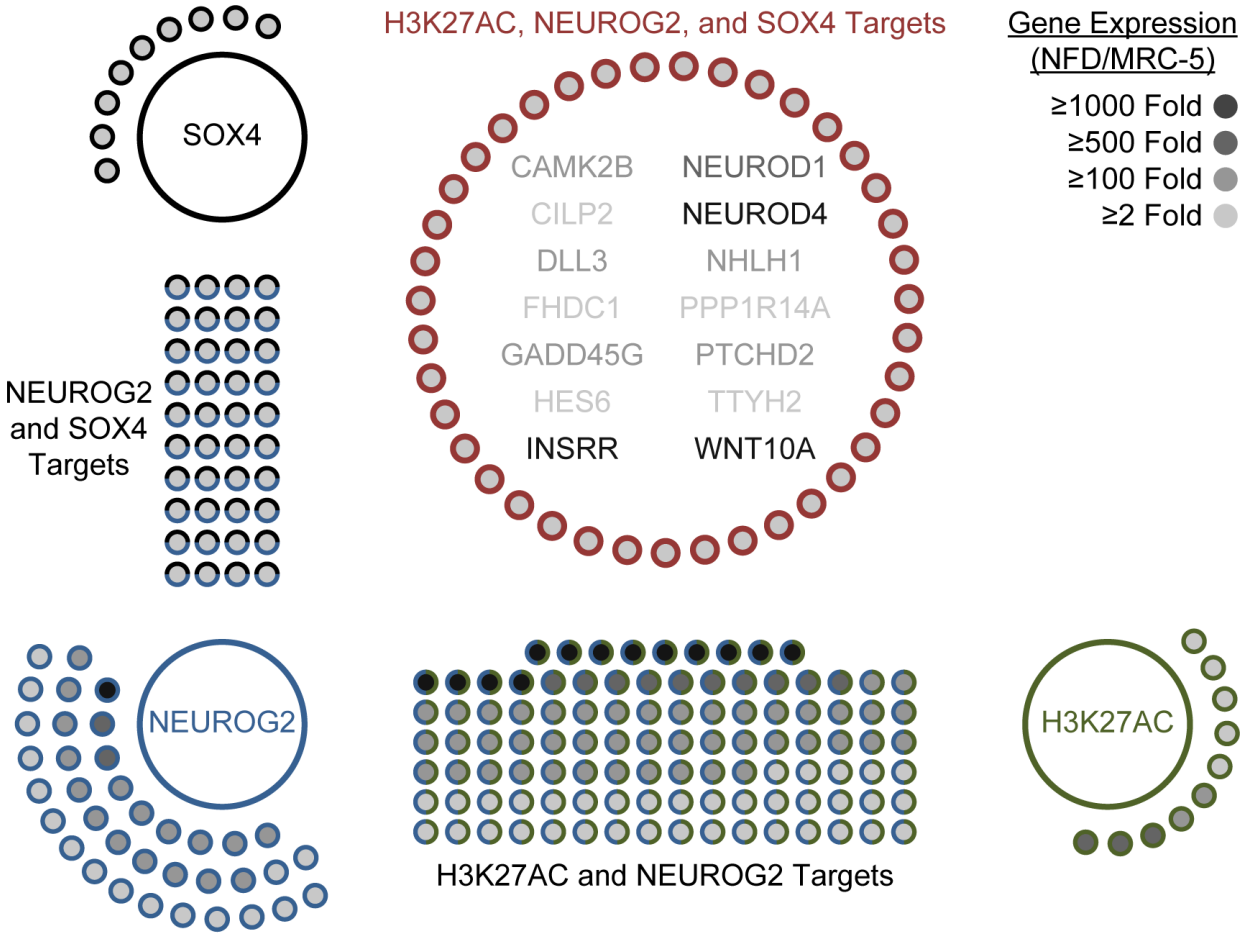
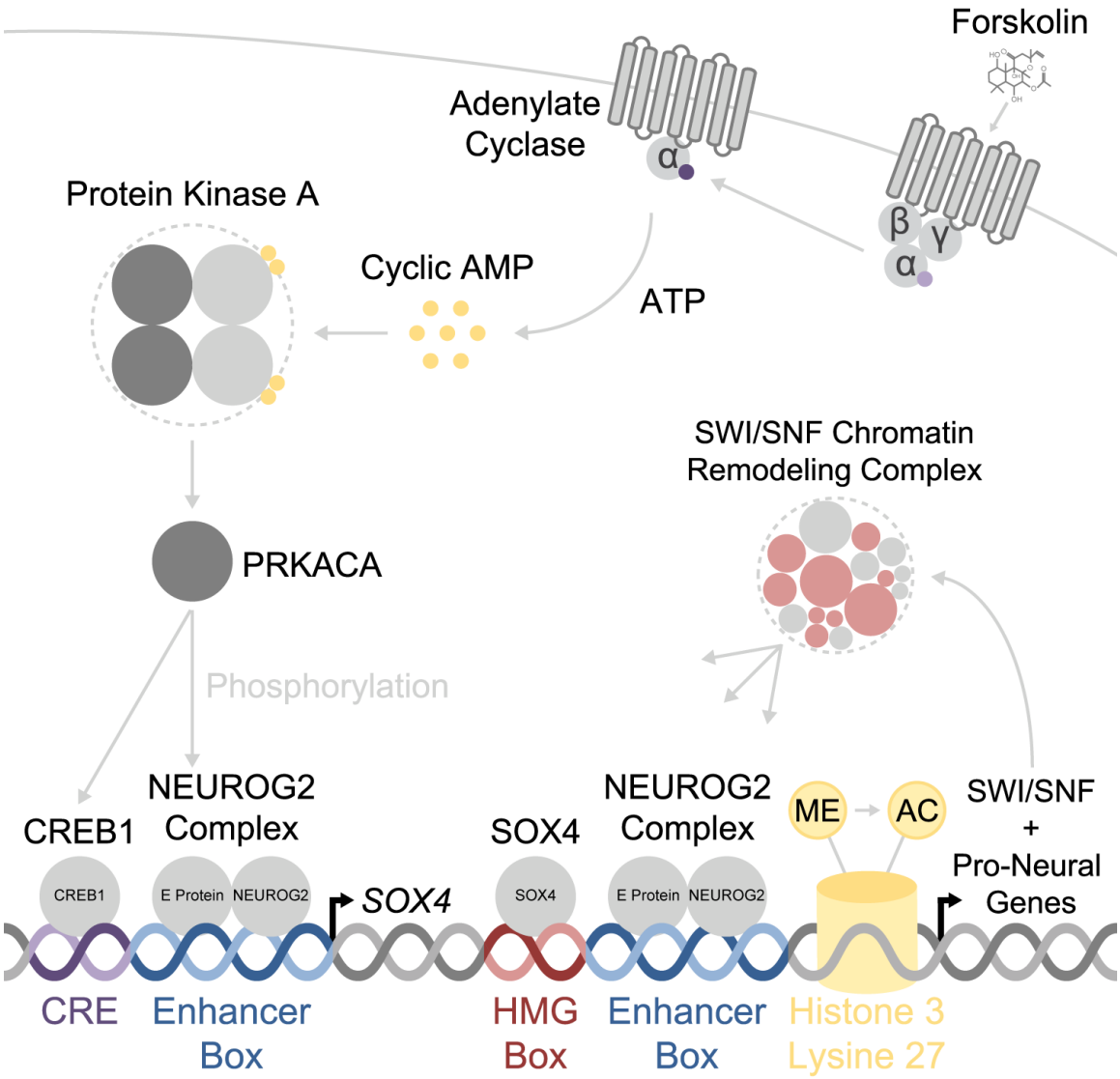


Figure 18: Integration of RNA-seq and NEUROG2, SOX4, and Histone ChIP-seq Datasets

Diagram representing H3K27AC-, NEUROG2-, and SOX4-ChIP target genes with corresponding RNA-seq expression data for NFD-treated fibroblasts relative to untreated fibroblasts at 2 DPT. Each circle represents a single gene target with the perimeter colored to indicate ChIP factors that regulate the gene (H3K27AC, green; NEUROG2, blue; SOX4, black). Genes targeted by all ChIP factors are indicated with red perimeter and genes in this category with greater than 20-fold expression change are listed alphabetically. The 150 H3K27AC-, 150 NEUROG2-, and 99 SOX4-target genes most induced at 2 DPT were used in this diagram.

CHAPTER 7

GRAPHICAL SUMMARY



CONCLUSION

This research presents the first detailed, stepwise molecular mechanism for the early events that catalyze human fibroblast-to-neuron transdifferentiation. In this work, ATAC-seq, RNA-seq, transcription factor ChIP-seq, and native histone ChIP-seq were used to establish a multi-layered view of the genetic and epigenetic mechanisms that catalyze reprogramming. We demonstrate that the small molecules forskolin and dorsomorphin act to enhance genome-wide chromatin accessibility by upregulating *SOX4*, which directly targets subunits of the SWI/SNF complex. We validated this mechanism using SOX4 to reprogram human glioblastoma and adult skin fibroblasts refractory to conversion by NEUROG2 alone. Finally, a small molecule screen targeting epigenetic modifiers resulted in the identification of FK228 as a potent activator of reprogramming when combined with NEUROG2 in adult fibroblasts.

With a model and mechanism in hand, this system is primed for further investigations into the mechanics of cellular reprogramming. Specifically, two principal lines of investigation have promising translational outlooks: (1) the termination of the cell cycle by transdifferentaiton to post-mitotic cell identity and (2) non-viral epigenetic reprogramming as a mechanism to facilitate high-efficiency conversion of cell lineages refractory to reprogramming. The findings in this work related to SOXC transcription factors and chemical-enhanced adult fibroblast reprogramming are promising leads toward solutions for these therapeutic challenges.

CHAPTER 9

REFERENCES

Aaboe M, Birkenkamp-Demtroder K, Wiuf C, Sørensen FB, Thykjaer T, Sauter G, Jensen KM-E, Dyrskjøt L, Ørntoft T (2006) SOX4 expression in bladder carcinoma: clinical aspect and in vitro functional characterization. Cancer Res 66:3434-3442.

Abad M, Mosteiro L, Pantoja C, Cañamero M, Rayon T, Ors I, Graña O, Megías D, Domínguez O, Martínez D, Manzanares M, Ortega S, Serrano M (2013) Reprogramming in vivo produces teratomas and iPS cells with totipotency features. Nature 502:340-345.

Addis RC, Hsu F-C, Wright RL, Dichter MA, Coulter DA, Gearhart JD (2011) Efficient conversion of astrocytes to functional midbrain dopaminergic neurons using a single polycistronic vector. PLoS One 6:e28719.

Ali F, Hindley C, McDowell G, Deibler R, Jones A, Kirschner M, Guillemot F, Philpott A. (2011) Cell cycle-regulated multi-site phosphorylation of Neurogenin 2 coordinates cell cycling with differentiation during neurogenesis. Development 138:4267-4277.

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

Asprer JST, Lee B, Wu C-S, Vadakkan T, Dickinson ME, Lu H-C, Lee S-K (2011) LMO4 functions as a co-activator of Neurogenin 2 in the developing cortex. Development 138:2823-2832.

Aubry L, Bugi A, Lefort N, Rousseau F, Peschanski M, Perrier AL (2008) Striatal progenitors derived from human ES cells mature into DARPP32 neurons in vitro and in quinolinic acid-lesioned rats. Proc Natl Acad Sci USA 105:16707-16712.

Bain G, Kitchens D, Yao M, Huettner JE, Gottlieb DI (1995) Embryonic stem cells express neuronal properties in vitro. Dev Biol 168:842-857.

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

Bhattaram P, Penzo-Méndez A, Sock E, Colmenares C, Kaneko KJ, Vassilev A, DePamphilis ML, Wegner M, Lefebvre V (2010) Organogenesis reies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. Nat Commun 1:9.

Bissonnette CJ, Lyass L, Bhattacharyya BJ, Belmadani A, Miller RJ, Kessler JA (2011) The controlled generation of functional basal forebrain cholinergic neurons from human embryonic stem cells. Stem Cells 29:802-811.

Brennand KJ, Simone A, Jou J, Gelboin-Burkhart C, Tran N, Sangar S, Li Y, Mu Y, Chen G, Yu D, McCarthy S, Sebat J, Gage FH (2011) Modelling schizophrenia using human induced pluripotent stem cells. Nature 473:221-225.

Briggs R, King TJ (1952) Transplantation of living nuclei from blastula cells into enucleated frogs' eggs. Proc Natl Acad Sci USA 38:455-463.

Britz O, Mattar P, Nguyen L, Langevin L-M, Zimmer C, Alam S, Guillemot F, Schuurmans C (2006) A role for proneural genes in the maturation of cortical progenitor cells. Cereb Cortex 16:138-151.

Bronicki LM, Bélanger G, Jasmin BJ (2012) Characterization of multiple exon 1 variants in mammalian HuD mRNA and neuron-specific transcriptional control via Neurogenin 2. J Neurosci 32:11164-11175.

Cahan P, Li H, Morris SA, Lummertz da Rocha E, Daley GQ, Collins JJ (2014) CellNet: network biology applied to stem cell engineering. Cell 158:903-915.

Caiazzo M, Dell'Anno MT, Dvoretskova E, Lazarevic D, Taverna S, Leo D, Sotnikova TD, Menegon A, Roncaglia P, Colciago G, Russo G, Carninci P, Pezzoli G, Gainetdinov RR, Gustincich S, Dityatev A, Broccoli V (2011) Direct generation of functional dopaminergic neurons from mouse and human fibroblasts. Nature 476:224-227.

Candiano G, Bruschi M, Musante L, Santucci L, Ghiggeri GM, Carnemolla B, Orecchia P, Zardi L, Righetti PG (2004) Blue silver: a very sensitive colloidal Coomassie G-250 staining for proteome analysis. Electrophoresis 25:1327-1333.

Chen C, Lee GA, Pourmorady A, Sock E, Donoghue MJ (2015) Orchestration of neuronal differentiation and progenitor pool expansion in the developing cortex by SoxC genes. J Neurosci 35:10629-10642.

Cheng L, Hu W, Qiu B, Zhao J, Yu Y, Guan W, Wang M, Yang W, Pei G. (2014) Generation of neural progenitor cells by chemical cocktails and hypoxia. Cell Res 24:665-679.

Cheung M, Abu-Elmagd M, Clevers H, Scotting PJ (2000) Roles of Sox4 in central nervous system development. Mol Brain Res 79:180-191.

Colasante G, Lignani G, Rubio A, Medrihan L, Yekhlef L, Sessa A, Massimino L, Giannelli SG, Sacchetti S, Caiazzo M, Leo D, Alexopoulou D, Dell'Anno MT, Ciabatti E, Orlando M, Studer M, Dahl A, Gainetdinov RR, Taverna S, Benfenati F, Broccoli V (2015) Rapid

conversion of fibroblasts into functional forebrain GABAergic interneurons by direct genetic reprogramming. Cell Stem Cell 17:719-734.

Craig R, Beavis RC (2004) TANDEM: matching proteins with tandem mass spectra. Bioinformatics 20:1466-1467.

Deacon T, Dinsmore J, Costantini LC, Ratliff J, Isacson O (1998) Blastula-stage stem cells can differentiate into dopaminergic and serotonergic neurons after transplantation. Exp Neurol 149:28-41.

Dy P, Penzo-Méndez A, Wang H, Pedraza CE, Macklin WB, Lefebvre V (2008) The three SoxC proteins – Sox4, Sox11, and Sox12 – exhibit overlapping expression patterns and molecular properties. Nucleic Acids Res 36:3101-3117.

Eisen MB, Spellman PT, Brown PO, Botstein D (1998) Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci USA 95:14863-14868.

Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. Nat Methods 4:207-214.

Emborg ME, Liu Y, Xi J, Zhang X, Yin Y, Lu J, Joers V, Swanson C, Holden JE, Zhang S-C (2013) Induced pluripotent stem cell-derived neural cells survive and mature in the nonhuman primate brain. Cell Rep 3:646-650.

Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotent cells from mouse embryos. Nature 292:154-156.

Fellmann C, Hoffmann T, Sridhar V, Hopfgartner B, Muhar M, Roth M, Lai DY, Barbosa IAM, Kwon JS, Guan Y, Sinha N, Zuber J (2013) An optimized microRNA backbone for effective single-copy RNAi. Cell Rep 5:1704-1713.

Fode C, Gradwohl G, Morin X, Dierich A, LeMeur M, Goridis C, Guillemot F (1998) The bHLH protein Neurogenin 2 is a determinant factor for epibranchial placode-derived sensory neurons. Neuron 20:483-494.

Fode C, Ma Q, Casarosa S, Ang S-L, Anderson DJ, Guillemot F (2000) A role for neural determination genes in specifying the dorsoventral identity of telencephalic neurons. Genes Dev 14:67-80.

Fraichard A, Chassande O, Bilbaut G, Dehay C, Savatier P, Samarut J (1995) In vitro differentiation of embryonic stem cells into glial cells and functional neurons. J Cell Sci 108:3181-3188.

Gascón S, Murenu E, Masserdotti G, Ortega F, Russo GL, Petrik D, Deshpande A, Heinrich C, Karow M, Robertson SP, Schroeder T, Beckers J, Irmler M, Berndt C, Angeli JPF, Conrad M, Berninger B, Götz M (2016) Identification and successful negotiation of a metabolic checkpoint in direct neuronal reprogramming. Cell Stem Cell 18:396-409.

Geer LY, Markey SP, Kowalak JA, Wagner L, Xu M, Maynard DM, Yang X, Shi W, Bryant SH (2004) Open mass spectrometry search algorithm. J Proteome Res 3:958-964.

Geijsen N, Uings IJ, Pals C, Armstrong J, McKinnon M, Raaijmakers JAM, Lammers J-WJ, Koenderman L, Coffer PJ (2001) Cytokine-specific transcriptional regulation through an IL-5Rα interacting protein. Science 293:1136-1138.

Gilfillan GD, Hughes T, Sheng Y, Hjorthaug HS, Straub T, Gervin K, Harris JR, Undlien DE, Lyle R (2012) Limitations and possibilities of low cell number ChIP-seq. BMC Genomics 12:645.

Gnedeva K, Hudspeth AJ (2015) SoxC transcription factors are essential for the development of the inner ear. Proc Natl Acad Sci USA 112:14066-14071.

Gohlke JM, Armant O, Parham FM, Smith MV, Zimmer C, Castro DS, Nguyen L, Parker JS, Gradwohl G, Portier CJ, Guillemot F (2008) Characterization of the proneural gene regulatory network during mouse telencephalon development. BMC Biol 6:15.

Gonzalez GA, Montminy MR (1989) Cyclic AMP stimulates somatostatin gene transcription by phosphorylation of CREB at serine 133. Cell 59:675-680.

Gradwohl G, Fode C, Guillemot F (1996) Restricted expression of a novel murine atonal-related bHLH protein in undifferentiated neural precursors. Dev Biol 180:227-241.

Graham JD, Hunt SMN, Tran N, Clarke CL (1999) Regulation of the expression and activity by progestins of a member of the SOX gene family of transcriptional modulators. J Mol Endocrinol 22:295-304.

Gubbay J, Collignon J, Koopman P, Capel B, Economou A, Münsterberg A, Vivian N, Goodfellow P, Lovell-Badge R (1990) A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. Nature 346:245-250.

Guo Z, Zhang L, Wu Z, Chen Y, Wang F, Chen G (2014) In vivo direct reprogramming of reactive glial cells into functional neurons after brain injury and in an Alzheimer's disease model. Cell Stem Cell 14:188-202.

Gurdon JB, Laskey RA (1970) The transplantation of nuclei from single cultured cells into enucleate frogs' eggs. J Embryol Exp Morph 24:227-248.

Han DW, Tapia N, Hermann A, Hemmer K, Höing S, Araúzo-Bravo MJ, Zaehres H, Wu G, Frank S, Moritz S, Greber B, Yang JH, Lee HT, Schwamborn JC, Storch A, Schöler HR (2012) Direct reprogramming of fibroblasts into neural stem cells by defined factors. Cell Stem Cell 10:465-472.

Hand R, Bortone D, Mattar P, Mguyen L, Heng JI-T, Guerrier S, Boutt E, Peters E, Barnes AP, Parras C, Schuurmans C, Guillemot F, Polleux F (2005) Phosphorylation of Neurogenin 2 specifies the migration properties and the dendritic morphology of pyramidal neurons in the neocortex. Neuron 48:45-62.

Harley VR, Lovell-Badge R, Goodfellow PN (1994) Definition of a consensus DNA binding site for SRY. Nucleic Acids Res 22:1500-1501.

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

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK (2010) Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol Cell 38:576-589.

Henke RM, Meredith DM, Borromeo MD, Savage TK, Johnson JE (2009) Ascl1 and Neurog2 form novel complexes and regulate Delta-like 3 (Dll3) expression in the neural tube. Dev Biol 328:529-540.

Hindley C, Ali F, McDowell G, Cheng K, Jones A, Guillemot F, Philpott A (2012) Post-translational modification of Ngn2 differentially affects transcription of distinct targets to regulate the balance between progenitor maintenance and differentiation. Development 139:1718-1723.

Hoser M, Baader SL, Bösl MR, Ihmer A, Wegner M, Sock E (2007) Prolonged glial expression of Sox4 in the CNS leads to architectural cerebellar defects and ataxia. J Neurosci 27:5495-5505.

Hou P, Li Y, Zhang X, Liu C, Guan J, Li H, Zhao T, Ye J, Yang W, Liu K, Ge J, Xu J, Zhang Q, Zhao Y, Deng H (2013) Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. Science 341:651-654.

Hu W, Qiu B, Guan W, Wang Q, Wang M, Li W, Gao L, Shen L, Huang Y, Xie G, Zhao H, Jin Y, Tang B, Yu Y, Zhao J, Pei G (2015) Direct Conversion of Normal and Alzheimer's Disease Human Fibroblasts into Neuronal Cells by Small Molecules. Cell Stem Cell 17:204-212.

Huang DW, Sherman BT, Lempicki RA (2009a) Systematic and integrative analysis of large gene lists using DAVID Bioinformatics Resources. Nat Protoc 4: 44-57.

Huang DW, Sherman BT, Lempicki RA (2009b) Bioinformatics enrichment tools: paths toward the comprehensive functional analysis of large gene list. Nucleic Acids Res 37:1-13.

Hunt SMN, Clarke CL (1999) Expression and hormonal regulation of the Sox4 gene in mouse female reproductive tissues. Biol Reprod 61:476-481.

Ieda M, Fu J-D, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG, Srivastava D (2010) Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. Cell 142:375-386.

Impey S, McCorkle SR, Cha-Molstad H, Dwyer JM, Yochum GS, Boss JM, McWeeney S, Dunn JJ, Mandel G, Goodman RH (2004) Defining the CREB regulon: a genome-wide analysis of transcription factor regulatory regions. Cell 119:1041-1054.

Janky R, Verfaillie A, Imrichová H, Van de Sande B, Standaert L, Christiaens V, Hulselmans G, Herten K, Sanchez MN, Potier D, Svetlichnyy D, Kalender Atak Z, Fiers N, Marine JC, Aerts S (2014) iRegulon: from a gene list to a gene regulatory network using large motif and track collections. PLoS Comput Biol 10:e1003731.

Jiang H, Xu Z, Zhong P, Ren Y, Liang G, Schilling HA, Hu Z, Zhang Y, Wang X, Chen S, Yan Z, Feng J (2015) Cell cycle and p53 gate the direct conversion of human fibroblasts to dopaminergic neurons. Nat Commun 6:10100.

Jiang Y, Ding Q, Xie X, Libby RT, Lefebvre V, Gan L (2013) Transcription factors SOX4 and SOX11 function redundantly to regulate the development of mouse and retinal ganglion cells. J Biol Chem 288:18429-18438.

Jiao J, Dang Y, Yang Y, Gao R, Zhang Y, Kou Z, Sun X-F, Gao S (2013) Promoting reprogramming by FGF2 reveals that the extracellular matrix is a barrier for reprogramming fibroblasts to pluripotency. Stem Cells 31:729-740.

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D (2002) The human genome browser at UCSC. Genome Res 12:996-1006.

Kim J, Su SC, Wang H, Cheng AW, Cassady JP, Lodato MA, Lengner CJ, Chung C-Y, Dawlaty MM, Tsai L-H, Jaenich R (2011) Functional integration of dopaminergic neurons directly converted from mouse fibroblasts. Cell Stem Cell 9:413-419.

Kiskinis E, Sandoe J, Williams LA, Boulting GL, Moccia R, Wainger BJ, Han S, Peng T, Thams S, Mikkilineni S, Mellin C, Merkle FT, Davis-Dusenbery BN, Ziller M, Oakley D, Ichida J, Costanzo SD, Atwater N, Maeder ML, Goodwin MJ, Nemesh J, Handsaker RE, Paull D, Noggle S, McCarroll SA, Joung JK, Woolf CJ, Brown RH, Eggan K (2014) Pathways disrupted in Human ALS motor neurons identified through genetic correction of mutant SOD1. Cell Stem Cell 14:781-795.

Kleinsmith LJ, Pierce GB Jr (1964) Multipotentiality of single embryonal carcinoma cells. Cancer Res 24:1544-1551.

Kovach C, Dixit R, Li S, Mattar P, Wilkinson G, Elsen GE, Kurrasch DM, Hevner RF, Schuurmans C (2013) Neurog2 simultaneously activates and represses alternative gene expression programs in the developing neocortex. Cereb Cortex 23:1884-1900.

Kuhlbrodt K, Herbarth B, Sock E, Enderich J, Hermans-Borgmeyer I, Wegner M (1998) Cooperative function of POU proteins and SOX proteins in glial cells. J Biol Chem 273:16050-16057.

Kwon H, Imbalzano AN, Khavari PA, Kingston RE, Green MR (1994) Nucleosome disruption and enhancement of activator binding by a human SW1/SNF complex. Nature 370:477-481.

Langmead B, Trapnell C, Pop M, Salzberg SL (2009) Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. Genome Biol 10:R25.

Ledent V, Paquet O, Vervoort M (2002) Phylogenetic analysis of the human basic helix-loop-helix proteins. Genome Biol 3:0030.1-0030.18.

Lee S, Lee B, Lee JW, Lee S-K (2009) Retinoid signaling and Neurogenin 2 function are coupled for the specification of spinal motor neurons through a chromatin modifier CBP. Neuron 62:641-654.

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

Li S, Mattar P, Zinyk D, Singh K, Chaturvedi C-P, Kovach C, Dixit R, Kurrasch DM, Ma Y-C, Chan JA, Wallace V, Dilworth FJ, Brand M, Schuurmans C (2012) GSK3 temporally regulates Neurogenin 2 proneural activity in the neocortex. J Neurosci 32:7791-7805.

Li X, Zuo X, Jing J, Ma Y, Wang J, Liu D, Zhu J, Du X, Xiong L, Du Y, Xu J, Xiao X, Wang J, Chai Z, Zhao Y, Deng H (2015) Small-molecule-derived direct reprogramming of mouse fibroblasts into functional neurons. Cell Stem Cell 17:195-203.

Lioubinski O, Müller M, Wegner M, Sander M (2003) Expression of Sox transcription factors in the developing mouse pancreas. Dev Dyn 227:402-408.

Liu M-L, Zang T, Zou Y, Chang JC, Gibson JR, Huber KM, Zhang C-L (2013) Small molecules enable Neurogenin 2 to efficiently convert human fibroblasts into cholinergic neurons. Nat Commun 4:2183.

Liu M-L, Zang T, Zhang CL. (2016) Direct Lineage Reprogramming Reveals Disease-Specific Phenotypes of Motor Neurons from Human ALS Patients. Cell Rep 14:115-128.

Liu P, Ramachandran S, Seyed MA, Scharer CD, Laycock N, Dalton WB, Williams H, Karanam S, Datta MW, Jaye DL, Moreno CS (2006) Sex-determining region Y is a transforming oncogene in human prostate cancer cells. Cancer Res 66:4011-4019.

Liu X, Li F, Stubblefield EA, Blanchard B, Richards TL, Larson GA, He Y, Huang Q, Tan AC, Zhang D, Benke TA, Sladek JR, Zahniser NR, Li C-Y (2012) Direct reprogramming of human fibroblasts into dopaminergic neuron-like cells. Cell Res 22:321-332.

Lujan E, Chanda S, Ahlenius H, Südhof TC, Wernig M (2012) Direct conversion of mouse fibroblasts to self-renewing, tripotent neural precursor cells. Proc Natl Acad Sci USA 109:2527-2532.

Ma DK, Jang M-H, Guo JU, Kitabatake Y, Chang M-L, Pow-anpongkul N, Flavell RA, Lu B, Ming G-L, Song H (2009) Neuronal activity-induced Gadd45bpromotes epigenetic DNA demethylation and adult neurogenesis. Science 323:1074-1077.

Ma Q, Kintner C, Anderson DJ (1996) Identification of Neurogenin, a vertebrate neuronal determination gene. Cell 87:43-52.

Ma Y-C, Song M-R, Park JP, Ho H-YH, Hu L, Kurtev MV, Zieg J, Ma Q, Pfaff SL, Greenberg ME (2008) Regulation of motor neuron specification by phosphorylation of Neurogenin 2. Neuron 58:65-77.

Masserdotti G, Gillotin S, Sutor B, Drechsel D, Irmler M, Jørgensen HF, Sass S, Theis FJ, Beckers J, Berninger B, Guillemot F, Götz M (2015) Transcriptional mechanisms of proneural factors and REST in regulating neuronal reprogramming of astrocytes. Cell Stem Cell 17:74-88.

Mattar P, Britz O, Johannes C, Nieto M, Ma L, Rebeyka A, Klenin N, Polleux F, Guillemot F, Schuurmans C (2004) A screen for downstream effectors of Neurogenin 2 in the embryonic neocortex. Dev Biol 273:373-389.

Mazzoni EO, Mahony S, Closser M, Morrison CA, Nedelec S, Williams DJ, An D, Gifford DK, Wichterle H (2013) Synergistic binding of transcription factors to cell-specific enhancers program motor neuron identity. Nat Neurosci 16:1219-1227.

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, Bejerano G (2010) GREAT improves functional interpretation of cis-regulatory regions. Nat Biotechnol 28:495-501.

Mertens J, Paquola ACM, Ku M, Hatch E, Böhnke L, Ladjevardi S, McGrath S, Campbell B, Lee H, Herdy JR, Gonçalves JT, Toda T, Kim Y, Winkler J, Yao J, Hetzer MW, Gage FH (2015) Directly reprogrammed human neurons retain aging-associated transcriptomic signatures and reveal age-related nucleocytoplasmic defects. Cell Stem Cell 17:705-718.

Meyer K, Ferraiuolo L, Miranda CJ, Likhite S, McElroy S, Renusch S, Ditsworth D, Lagier-Tourenne C, Smith RA, Ravits J, Burghes AH, Shaw PJ, Cleveland DW, Kolb SJ, Kaspar BK (2014) Direct conversion of patient fibroblasts demonstrates non-cell autonomous

toxicity of astrocytes to motor neurons in familial and sporadic ALS. Proc Natl Acad Sci USA 111:829-832.

Miller JA, Nathanson J, Franjic D, Shim S, Dalley RA, Shapouri S, Smith KA, Sunkin SM, Bernard A, Bennett JL, Lee C-K, Hawrylycz MJ, Jones AR, Amaral DG, Šestan N, Gage FH, Lein ES (2013) Conserved molecular signatures of neurogenesis in the hippocampal subgranular zone of rodents and primates. Development 140:4633-4644.

Mizuguchi R, Sugimori M, Takebayashi H, Kosako H, Nagao M, Yoshida S, Nabeshima Y-I, Shimamura K, Nakafuku M (2001) Combinatorial roles of Olig2 and Neurogenin 2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. Neuron 31:757-771.

Mohn F, Weber M, Rebhan M, Roloff TC, Richter J, Stadler MB, Bibel M, Schübeler D (2008) Lineage-specific polycomb targets and de novo DNA methylation define restriction and potential of neuronal progenitors. Mol Cell 30:755-766.

Molyneaux BJ, Arlotta P, Menezes JRL, Macklis JD (2007) Neuronal subtype specification in the cerebral cortex. Nat Rev Neurosci 8:427-437.

Mu L, Berti L, Masserdotti G, Covic M, Michaelidis TM, Doberauer K, Merz K, Rehfeld F, Haslinger A, Wegner M, Sock E, Lefebvre V, Couillard-Despres S, Aigner L, Berninger B, Lie DC (2012) SoxC transcription factors are required for neuronal differentiation in adult hippocampal neurogenesis. J Neurosci 32:3067-3080.

Murre C, McCaw PS, Baltimore D (1989) A new DNA binding and dimerization motif in immunoglobulin enhancer binding, daughterless, MyoD, and myc proteins. Cell 56:777-783.

Neuhold LA, Wold B (1993) HLH forced dimers: tethering MyoD to E47 generates a dominant positive myogenic factor insulated from negative regulation by Id. Cell 74:1033-1042.

Niu W, Zang T, Zou Y, Fang S, Smith DK, Bachoo R, Zhang C-L (2013) In vivo reprogramming of astrocytes to neuroblasts in the adult brain. Nat Cell Biol 15:1164-1175.

Ohi Y, Qin H, Hong C, Blouin L, Polo JM, Guo T, Qi Z, Downey SL, Manos PD, Rossi DJ, Yu J, Hebrok M, Hochedlinger K, Costello JF, Song JS, Ramalho-Santos M (2011) Incomplete DNA methylation underlies a transcriptional memory of somatic cells in human iPS cells. Nat Cell Biol 13:541-549.

Okabe S, Forsberg-Nilsson K, Spiro AC, Segal M, McKay RDG (1996) Development of neuronal precursor cells and functional postmitotic neurons from embryonic stem cells in vitro. Mech Dev 59:89-102.

Oldfield AJ, Yang P, Conway AE, Cinghu S, Freudenberg JM, Yellaboina S, Jothi R (2014) Histone-fold domain protein NF-Y promotes chromatin accessibility for cell type-specific master transcription factors. Mol Cell 55:708-722.

Orellana SA, McKnight GS (1992) Mutations in the catalytic subunit of cAMP-dependent protein kinase result in unregulated biological activity. Proc Natl Acad Sci USA 89:4726-4730.

Osório J, Mueller T, Rétaux S, Vernier P, Wullimann MF (2010) Phylotypic expression of the bHLH genes Neurogenin 2, NeuroD, and Mash1, in the mouse embryonic forebrain. J Comp Neurol 518:851-871.

Pang ZP, Yang N, Vierbuchen T, Ostermeier A, Fuentes DR, Yang TQ, Citri A, Sebastiano V, Marro S, Südhof TC, Wernig M (2011). Induction of human neuronal cells by defined transcription factors. Nature 476:220-223.

Park I-H, Zhao R, West JA, Yabuuchi A, Huo H, Ince TA, Lerou PH, Lensch MW, Daley GQ (2008) Reprogramming of human somatic cells to pluripotency with defined factors. Nature 451:141-146.

Parras CM, Schuurmans C, Scardigli R, Kim J, Anderson DJ, Guillemot F (2002) Divergent functions of the proneural genes Mash1 and Ngn2 in the specification of neuronal subtype identity. Genes Dev 16:324-338.

Pataskar A, Jung J, Smialowski P, Noack F, Calegari F, Straub T, Tiwari VK (2016) NeuroD1 reprograms chromatin and transcription factor landscapes to induce the neuronal program. EMBO J 35:24-45.

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

Polo JM, Liu S, Figueroa ME, Kulalert W, Eminli S, Tan KY, Apostolou E, Stadtfeld M, Li Y, Shioda T, Natesan S, Wagers AJ, Melnick A, Evans T, Hochedlinger K (2010) Cell type of origin influences the molecular and functional properties of mouse induced pluripotent stem cells. Nat Biotechnol 28:848-855.

Potzner MR, Griffel C, Lütjen-Drecoll E, Bösl MR, Wegner M, Sock E (2007) Prolonged Sox4 expression in oligodendrocytes interferes with normal myelination in the central nervous system. Mol Cell Biol 27:5316-5326.

Potzner MR, Tsarovina K, Binder E, Penzo-Méndez A, Lefebvre V, Rohrer H, Wegner M, Sock E (2010) Sequential requirement of Sox4 and Sox11 during development of the sympathetic nervous system. Development 137:775-784.

Rackham OJL, Firas J, Fang H, Oates ME, Holmes ML, Knaupp AS, The FANTOM Consortium, Suzuki H, Nefzger CM, Daub CO, Shin JW, Petretto E, Forrest ARR, Hayashizaki Y, Polo JM, Gough J (2016) A predictibe computational framework for direct reprogramming between human cell types. Nat Genet 48:331-335.

Ring KL, Tong LM, Balestra ME, Javier R, Andrews-Zwilling Y, Li G, Walker D, Zhang WR, Kreitzer AC, Huang Y (2012) Direct reprogramming of mouse and human fibroblasts into multipotent neural stem cells with a single factor. Cell Stem Cell 11:100-109.

Ryall JG, Cliff T, Dalton S, Sartorelli V (2015) Metabolic reprogramming of stem cell epigenetics. Cell Stem Cell 17:651-662.

Saldanha AJ (2004) Java Treeview-extensible visualization of microarray data. Bioinformatics 20:3246-3248.

Scardigli R, Schuurmans C, Gradwohl G, Guillemot F (2001) Crossregulation between Neurogenin 2 and pathway specifying neuronal identity in the spinal cord. Neuron 31:203-217.

Schepers G, Teasdale RD, Koopman P (2002) Twenty pairs of SOX: extent, homology, and nomenclature of the mouse and human SOX transcription factor gene families. Dev Cell 3:167-170.

Schilham MW, Oosterwegel MA, Moerer P, Ya J, de Boer PAJ, van de Wetering M, Verbeek S, Lamers WH, Kruisbeek AM, Cumano A, Clevers H (1996) Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox4. Nature 380:711-714.

Schuurmans C, Armant O, Nieto M, Stenman JM, Oliver B, Klenin N, Brown C, Langevin L-M, Seibt J, Tang H, Cunningham JM, Dyck R, Walsh C, Campbell K, Polleux F, Guillemot F

(2004) Sequentil phases of cortical specification involve Neurogenin-dependent and -independent pathways. EMBO J 23:2892-2902.

Seo S, Lim J-W, Yellajoshyula D, Chang L-W, Kroll KL (2007) Neurogenin and NeuroD direct transcriptional targets and their regulatory enhancers. EMBO J 26:5093-5108.

Shen L, Shao N, Liu X, Nestler E (2014) ngs.plot: Quick mining and visualization of next-generation sequencing data by integrating genomic databases. BMC Genomics 15:284.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf A-M, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determination region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240-244.

Sinner D, Kordich JJ, Spence JR, Opoka R, Rankin S, Lin S-CJ, Jonatan D, Zorn AM, Wells JM (2007) Sox17 and Sox4 differentially regulate β-catenin/T-cell factor activity and proliferation of colon carcinoma cells. Mol Cell Biol 27:7802-7815.

Smith DK, Yang J, Liu M-L, Zhang C-L (2016) Small molecules modulate chromatin accessibility to promote fibroblast-to-neuron reprogramming. *In review.*

Smith DK, He M, Zhang C-L, Zheng J (2016) The therapeutic potential of cell identity reprogramming for the treatment of aging-related neurodegenerative disorders. Prog Neurobiol 15:30067-30084.

Soldner F, Hockemeyer D, Beard C, Gao Q, Bell GW, Cook EG, Hargus G, Blak A, Cooper O, Mitalipova M, Isacson O, Jaenisch R (2009) Parkinson's disease patient-derived induced pluripotent stem cells free of viral reprogramming factors. Cell 136:964-977.

Sommer L, Ma Q, Anderson DJ (1996) Neurogenins, a novel family of atonal-related bHLH transcription factors, are putative mammalian neuronal determination genes that reveal progenitor cell heterogeneity in the developing CNS and PNS. Mol Cell Neurosci 8:221-241.

Son EY, Crabtree GR (2004) The role of BAF (mSWI/SNF) complexes in mammalian neural development. Am J Med Genet 166C:333-349.

Son EY, Ichida JK, Wainger BJ, Toma JS, Rafuse VF, Woolf CJ, Eggan K. (2011) Conversion of mouse and human fibroblasts into functional spinal motor neurons. Cell Stem Cell 9:205-218.

Song K, Nam Y-J, Luo X, Qi X, Tan W, Huang GN, Acharya A, Smith CL, Tallquist MD, Neilson EG, Hill JA, Bassel-Duby R, Olson EN (2012) Heart repair by reprogramming non-myocytes with cardiac transcription factors. Nature 485:599-604.

Stevens LC (1967) The biology of teratomas. Adv Morphog 6:1-31.

Stiles J, Jernigan TL (2010) The basics of brain development. Neuropsychol Rev 20:327-348.

Su Z, Niu W, Liu M-L, Zou Y, Zhang C-L (2014) In vivo conversion of astrocytes to neurons in the injured adult spinal cord. Nat Commun 5:3338.

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

Takekawa M, Saito H. (1998) A family of stress-inducible GADD45-like proteins mediate activation of the stress-responsive MTK1/MEKK4 MAPKKK. Cell 95:521-530.

Tanaka S, Kamachi Y, Tanouchi A, Hamada H, Jing N, Kondoh H (2004) Interplay of SOX and POU factors in regulation of the Nestin gene in neural primordial cells. Mol Cell Biol 24:8834-8846.

Tapscott SJ, Davis RL, Thayer MJ, Cheng PF, Weintraub H, Lassar AB (1988) MyoD1: a nuclear phosphoprotein requiring a Myc homology region to convert fibroblasts to myoblasts. Science 242:405-411.

Thein DC, Thalhammer JM, Hartwig AC, Crenshaw EB III, Lefebvre V, Wegner M, Sock E (2010) The closely related transcription factors Sox4 and Sox11 function as survival factors during spinal cord development. J Neurochem 115:131-141.

Their M, Wörsdörfer P, Lakes YB, Gorris R, Herms S, Opitz T, Seiferling D, Quandel T, Hoffmann P, Nöthen MM, Brüstle, Edenhofer F (2012) Direct conversion of fibroblasts into stably expandable neural stem cells. Cell Stem Cell 10:473-479.

Thomson JA, Kalishman J, Golos TG, Durning M, Harris CP, Becker RA, Hearn JP (1995) Isolation of a primate embryonic stem cell line. Proc Natl Acad Sci USA 92:7844-7848.

Thomson JA, Itskovitz-Eldor J, Shapiro SS, Waknitz MA, Swiergiel JJ, Marshall VS, Jones JM (1998) Embryonic stem cell lines derived from human blastocysts. Science 282:1145-1147.

Timofeeva OA, Chasovskikh S, Lonskaya I, Tarasova NI, Khavrutskii L, Tarasov SG, Zhang X, Korostyshevskiy VR, Cheema A, Zhang L, Dakshanamurthy S, Brown ML, Dritschilo A (2012) Mechanisms of unphosphorylated STAT3 transcription factor binding to DNA. J Biol Chem 287:14192-14200

Torper O, Pfisterer U, Wolf DA, Pereira M, Lau S, Jakobsson J, Björklund A, Grealish S, Parmar M (2013) Generation of induced neurons via direct conversion in vivo. Proc Natl Acad Sci USA 110:7038-7043.

Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L (2012) Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. Nat Protoc 7:562-578.

Trudgian DC, Thomas B, McGowan SJ, Kessler BM, Salek M, Acuto O (2010) CPFP: a central proteomics facilities pipeline. Bioinformatics 26:1131-1132.

Trudgian DC, Ridlova G, Fischer R, Mackeen MM, Ternette N, Acuto O, Kessler BM, Thomas B (2011) Comparative evaluation of label-free SNIQ normalized spectral index quantitation in the central proteomics facilities pipeline. Proteomics 11:2790-2797.

Trudgian DC, Mirzaei H (2012) Cloud CPFP: a shotgun proteomics data analysis pipeline using cloud and high performance computing. J Proteome Res 11:6282-6290.

van Beest M, Dooijes D, van de Wetering M, Kjaerulff S, Bonvin A, Nielsen O, Clevers H (2000) Sequence-specific HMG box factors recognize 10-12 base pair minor groove motifs. J Biol Chem 275:27266-27273.

van de Wetering M, Oosterwegel M, van Norren K, Clevers H (1993) Sox4, an Sry-like HMG box protein, is a transcriptional activator in lymphocytes. EMBO J 12:3847-3854.

van Houte LPA, Chuprina VP, van der Wetering M, Boelens R, Kaptein R, Clevers H (1995) Solution structure of the sequence-specific HMG box of the lymphocyte transcriptional activator Sox4. J Biol Chem 270:30516-30524.

Vervoort SJ, van Boxtel R, Coffer PJ (2013) The role of SRY-related HMG box transcription factor 4 (SOX4) in tumorigenesis and metastasis: friend or foe? Oncogene 32:3397-2409.

Victor MB, Richner M, Hermanstyne TO, Ransdell JL, Sobieski C, Deng PY, Klyachko VA, Nerbonne JM, Yoo AS (2014) Generation of human striatal neurons by microRNA-dependent direct conversion of fibroblasts. Neuron 84:311-323.

Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Südhof TC, Wernig M (2010) Direct conversion of fibroblasts to functional neurons by defined factors. Nature 463:1035-1041.

Wapinski OL, Vierbuchen T, Qu K, Lee QY, Chanda S, Fuentes DR, Giresi PG, Ng YH, Marro S, Neff NF, Drechsel D, Martynoga B, Castro DS, Webb AE, Südhof TC, Brunet A, Guillemot F, Chang HY, Wernig M (2013) Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell 155:621-635.

Wernig M, Zhao J-P, Pruszak J, Hedlund E, Fu D, Soldner F, Broccoli V, Constantine-Paton M, Isacson O, Jaenisch R (2008) Neurons derived from reprogrammed fibroblasts functionally integrate into the fetal brain and improve symptoms of rats with Parkinson's disease. Proc Natl Acad Sci USA 105:5856-5861.

Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, Rahl PB, Lee TI, Young RA (2013) Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell 153:307-319.

Wichterle H, Lieberam I, Porter JA, Jessell TM (2002) Directed differentiation of embryonic stem cells into motor neurons. Cell 110:385-397.

Yoo AS, Sun AX, Li L, Shcheglovitov A, Portmann T, Li Y, Lee-Messer C, Dolmetsch RE, Tsien RW, Crabtree GR (2011) MicroRNA-mediated conversion of human fibroblasts to neurons. Nature 476:228-231.

Yoshida M, Kitaoka S, Egawa N, Yamane M, Ikeda R, Tsukita K, Amano N, Watanabe A, Morimoto M, Takahashi J, Hosoi H, Nakahata T, Inoue H, Saito MK (2015) Modeling the early phenotype at the neuromuscular junction of spinal muscular atrophy using patient-derived iPSCs. Stem Cell Rep 4:561-568.

Zaret KS, Carroll JS (2011) Pioneer transcription factors: establishing competence for gene expression. Genes Dev 25:2227-2241.

Zhang L, Yin JC, Yeh H, Ma NX, Lee G, Chen XA, Wang Y, Lin L, Chen L, Jin P, Wu G-Y, Chen G (2015) Small Molecules Efficiently Reprogram Human Astroglial Cells into Functional Neurons. Cell Stem Cell 17:735-747.

Zhou Q, Brown J, Kanarek A, Rajagopal J, Melton DA (2008) In vivo reprogramming of adult pancreatic exocrine cells to β-cells. Nature 455:627-632.

CHAPTER 9

APPENDICIES

9.1. Appendix 1: Antibodies

Antigen	Conjugate	Supplier	Catalog Number	Host	Application	Dilution or Quantity
CREB1		Cell Signaling Technology	9104S	Mouse	Immunoblot	1:1000
CREB1 pS133		Cell Signaling Technology	9198S	Rabbit	ChIP	5 µg
CREB1 pS133		Cell Signaling Technology	9198S	Rabbit	Immunoblot	1:1000
GAPDH		Abcam	AB9485	Rabbit	Immunoblot	1:2500
GADD45G		Origene	TA308278	Rabbit	Immunoblot	1:1000
HA tag		Abcam	AB9110	Rabbit	ChIP	5 µg
HA tag		Covance	MMS-101P	Mouse	Immunoblot	1:1000
MAP2		Sigma	M4403	Mouse	Immunocytochemistry	1:750
Mouse IgG (H+L)	Fluor555	Life Technologies	A21203	Donkey	Immunocytochemistry	1:500
Mouse IgG (H+L)	Fluor594	Life Technologies	A21203	Donkey	Immunocytochemistry	1:500
Mouse IgG (H+L)	Fluor647	Life Technologies	A31571	Donkey	Immunocytochemistry	1:500
Mouse IgG (H+L)	HRP	Jackson ImmunoResearch	715-035-150	Donkey	Immunoblot	1:5000
NEUROD1		Abcam	AB109224	Rabbit	ChIP	5 µg
NEUROD4		Abcam	AB90484	Rabbit	ChIP	5 µg
Rabbit IgG (H+L)	Fluor647	Life Technologies	A31571	Donkey	Immunocytochemistry	1:500
Rabbit IgG (H+L)	HRP	Jackson ImmunoResearch	711-035-152	Donkey	Immunoblot	1:5000
SOX4		Abcam	AB86809	Rabbit	Immunoblot	1:1000
TUBB3		Covance	MMS-435P	Mouse	Immunocytochemistry	1:2500
TUBB3		Covance	PRB-435P	Rabbit	Immunocytochemistry	1:2500

9.2. Appendix 2: Cell Lines

Cell Line	Supplier	Catalog Number	Species	Tissue	Cell Type	Age	Disease State	Gender
293T/17	ATCC	CRL-11268	Human	Kidney	Epithelial	Fetus		
AG05811	Coriell Institute	AG05811	Human	Skin	Fibroblast	71 years	Healthy	Female
AG09969	Coriell Institute	AG09969	Human	Skin	Fibroblast	53 years	Healthy	Male
MRC-5	ATCC	CCL-171	Human	Lung	Fibroblast	14 weeks gestation	Healthy	Male
ND29563	Coriell Institute	ND29563	Human	Skin	Fibroblast	37 years	Amyotrophic lateral sclerosis	Male
ND39027	Coriell Institute	ND39027	Human	Skin	Fibroblast	50 years	Amyotrophic lateral sclerosis	Female

9.3. Appendix 3: DNA Constructs

Vector	Promoter	DNA Insert	Application(s)
pCSC-SP-PW-IRES/GFP	*CMV*	none	ChIP-seq, neuron induction, RNA-seq
pCSC-SP-PW-IRES/GFP	*CMV*	*HA-NEUROG2*	ChIP-seq, RNA-seq
pCSC-SP-PW	*CMV*	*GFP-BSD-T2A-NEUROG2-IRES-HA-SOX4*	ChIP-seq
pET-21a	*T7*	*GFP*-flexible linker-*NEUROG2*	EMSA, microscale thermophoresis, PRKACA phosphorylation assay
pET-21a	*T7*	*HA-TCF3* (*E47* isoform)	EMSA, microscale thermophoresis, PRKACA phosphorylation assay
pET-21a	*T7*	*NEUROG2*	EMSA, PRKACA phosphorylation assay
pMD2.G	*CMV*	none	Lentivirus production
pMDLg/pRRE	*CMV*	none	Lentivirus production
pRSV-REV	*CMV*	none	Lentivirus production
CS-CDF-CG-PRE	*Minimal promoter*	none	Luciferase assay
CS-CDF-CG-PRE	*Minimal promoter*	*CD44* intron (chr11:35,166,898-35,167,156)	Luciferase assay
CS-CDF-CG-PRE	*Minimal promoter*	*EMP1* 3' regulatory region (chr12:13,428,228-13,428,875)	Luciferase assay
CS-CDF-CG-PRE	*Minimal promoter*	*GDF5* 5' regulatory region (chr20:34,029,331-34,029,429)	Luciferase assay
CS-CDF-CG-PRE	*Minimal promoter*	*TEK* intron (chr20:34,029,331-34,029,429)	Luciferase assay
pCMV-LacZ	*CMV*	*LacZ*	Luciferase assay
pET-21a	*T7*	*HA-TCF3* (*E12* isoform)	Microscale thermophoresis, PRKACA phosphorylation assay
pET-21a	*T7*	*HA-TCF4*	Microscale thermophoresis, PRKACA phosphorylation assay
pET-21a	*T7*	*HA-TCF12*	Microscale thermophoresis, PRKACA phosphorylation assay
pCSC-SP-PW	*CMV*	*NEUROG2-IRES-GFP-T2A-HA-SOX4*	Neuron induction

pCSC-SP-PW	*CMV*	*NEUROG2-IRES-GFP-T2A-HA-Sox4*	Neuron induction
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2*	Neuron induction
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* deletion construct (K110 to T217)	Neuron induction
pCSC-SP-PW	*CMV*	*SMARCA4*	Neuron induction
pCSC-SP-PW-IRES/GFP	*CMV*	*SMARCB1*	Neuron induction
pCSC-SP-PW-IRES/GFP	*CMV*	*SMARCC2*	Neuron induction
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S24A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S24D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S91A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S91D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S193A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S193D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S207A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S207D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S209A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S209D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S219A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S219D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S232A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S232D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S239A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S239D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S242A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S242D)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S24/193/207/209/219/232/239/242A)	*NEUROG2* phosphomutant screen
pCSC-SP-PW-IRES/GFP	*CMV*	*NEUROG2* (S24/193/207/209/219/232/239/242D)	*NEUROG2* phosphomutant screen
pET-28a	*T7*	*PRKACA*	PRKACA phosphorylation assay
pCSC-SP-PW-IRES/GFP	*CMV*	*HA-NEUROG2-IRES-GFP-shRNA*	Short hairpin RNA knockdown

*Note: All viral plasmids are compatible with 3rd generation lentiviral packaging systems.
*Note: All genomic coordinates reference human genome version GRCh37/hg19.

9.4. Appendix 4: Oligos

Application	Identifier	Sequence
ATAC-seq	Universal ATAC primer	5'-AATGATACGGCGACCACCGAGATCTACACTCGTCGG CAGCGTCAGATGTG-3'
ATAC-seq	Barcoded ATAC primer 1	5'-CAAGCAGAAGACGGCATACGAGATTCGCCTTAGTCT CGTGGGCTCGGAGATGT-3'
ATAC-seq	Barcoded ATAC primer 2	5'-CAAGCAGAAGACGGCATACGAGATCTAGTACGGTCT CGTGGGCTCGGAGATGT-3'
ATAC-seq	Barcoded ATAC primer 3	5'-CAAGCAGAAGACGGCATACGAGATTTCTGCCTGTCT CGTGGGCTCGGAGATGT-3'
ATAC-seq	Barcoded ATAC primer 4	5'-CAAGCAGAAGACGGCATACGAGATGCTCAGGAGTCT CGTGGGCTCGGAGATGT-3'
EMSA	*DLL3* native sequence forward	5'-GCATTTACATACAGCTGGGAGCCCCCT-3'
EMSA	*DLL3* native sequence reverse	5'-GAGGGGGCTCCCAGCTGTATGTAAATG-3'
EMSA	*DLL3* enhancer box destroyed forward	5'-GCATTTACATAAAGCAAGGAGCCCCCT-3'
EMSA	*DLL3* enhancer box destroyed reverse	5'-GAGGGGGCTCCTTGCTTTATGTAAATG-3'
Microscale thermophoresis	*DLL3* native sequence forward	5'-CATTTACATACAGCTGGGAGCCCCCT-3'
Microscale thermophoresis	*DLL3* native sequence reverse	5'-AGGGGGCTCCCAGCTGTATGTAAATG-3'
Microscale thermophoresis	*DLL3* GA mutation forward	5'-CATTTACATACAGATGGGAGCCCCCT-3'
Microscale thermophoresis	*DLL3* GA mutation reverse	5'-AGGGGGCTCCCATCTGTATGTAAATG-3'
Microscale thermophoresis	*DLL3* TC mutation forward	5'-CATTTACATACATCTGGGAGCCCCCT-3'
Microscale thermophoresis	*DLL3* TC mutation reverse	5'-AGGGGGCTCCCAGATGTATGTAAATG-3'
NEUROG2 phosphomutant	*NEUROG2* (S24A) forward	5'-TCGGATCGGCC**G**CCCCCGCCTTG-3'
NEUROG2 phosphomutant	*NEUROG2* (S24A) reverse	5'-CAAGGCGGGGG**C**GGCCGATCCGA-3'
NEUROG2 phosphomutant	*NEUROG2* (S24D) forward	5'-GCTCGGATCGGCC**GA**CCCCGCCTTGGCG-3'
NEUROG2 phosphomutant	*NEUROG2* (S24D) reverse	5'-CGCCAAGGCGGGG**TC**GGCCGATCCGAGC-3'
NEUROG2 phosphomutant	*NEUROG2* (S91A) forward	5'-CGGCGCCCT**G**CCCGGGCGC-3'
NEUROG2 phosphomutant	*NEUROG2* (S91A) reverse	5'-GCGCCCGGG**C**AGGGCGCCG-3'
NEUROG2 phosphomutant	*NEUROG2* (S91D) forward	5'-CGGCGCCCT**GA**CCGGGCGCG-3'
NEUROG2 phosphomutant	*NEUROG2* (S91D) reverse	5'-CGCGCCCGG**TC**AGGGCGCCG-3'
NEUROG2 phosphomutant	*NEUROG2* (S193A) forward	5'-GGCAGTGTTGCTG**GC**CCCGGGAGGAGCC-3'

NEUROG2 phosphomutant	*NEUROG2* (S193A) reverse	5'-GGCTCCTCCCGGG**GC**CAGCAACACTGCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S193D) forward	5'-GGCAGTGTTGCTG**GA**CCCGGGAGGAGCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S193D) reverse	5'-GGCTCCTCCCGGG**TC**CAGCAACACTGCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S207A) forward	5'-AGCAGCGGAGAC**GC**CCCCTCGCCCGC-3'
NEUROG2 phosphomutant	*NEUROG2* (S207A) reverse	5'-GCGGGCGAGGGG**GC**GTCTCCGCTGCT-3'
NEUROG2 phosphomutant	*NEUROG2* (S207D) forward	5'-CAGCAGCGGAGAC**GA**CCCCTCGCCCGCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S207D) reverse	5'-GGCGGGCGAGGGG**TC**GTCTCCGCTGCTG-3'
NEUROG2 phosphomutant	*NEUROG2* (S209A) forward	5'-GGAGACAGCCCC**GC**CCCCGCCTCCACG-3'
NEUROG2 phosphomutant	*NEUROG2* (S209A) reverse	5'-CGTGGAGGCGGG**GG**CGGGGCTGTCTCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S209D) forward	5'-GCGGAGACAGCCCC**GA**CCCGCCTCCACGTG-3'
NEUROG2 phosphomutant	*NEUROG2* (S209D) reverse	5'-CACGTGGAGGCGGG**GTC**GGGGCTGTCTCCGC-3'
NEUROG2 phosphomutant	*NEUROG2* (S219A) forward	5'-GAGTTGCACCAAC**GC**CCCCGCGCCGTCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S219A) reverse	5'-GGACGGCGCGGGG**GC**GTTGGTGCAACTC-3'
NEUROG2 phosphomutant	*NEUROG2* (S219D) forward	5'-GAGTTGCACCAAC**GA**CCCCGCGCCGTCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S219D) reverse	5'-GGACGGCGCGGGG**TC**GTTGGTGCAACTC-3'
NEUROG2 phosphomutant	*NEUROG2* (S232A) forward	5'-CCTCCAATTCCACC**G**CCCCCTACAGCTGC-3'
NEUROG2 phosphomutant	*NEUROG2* (S232A) reverse	5'-GCAGCTGTAGGGGG**C**GGTGGAATTGGAGG-3'
NEUROG2 phosphomutant	*NEUROG2* (S232D) forward	5'-GTCCTCCAATTCCACC**GA**CCCCTACAGCTGCACT-3'
NEUROG2 phosphomutant	*NEUROG2* (S232D) reverse	5'-AGTGCAGCTGTAGGGG**TC**GGTGGAATTGGAGGAC-3'
NEUROG2 phosphomutant	*NEUROG2* (S239A) forward	5'-CCCCTACAGCTGCACTTTA**GC**CCCCGCCAGCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S239A) reverse	5'-GGCTGGCGGG**GG**CTAAAGTGCAGCTGTAGGGG-3'
NEUROG2 phosphomutant	*NEUROG2* (S239D) forward	5'-TCCCCCTACAGCTGCACTTTA**GA**CCCCGCCAGCCC-3'
NEUROG2 phosphomutant	*NEUROG2* (S239D) reverse	5'-GGGCTGGCGGGG**TC**TAAAGTGCAGCTGTAGGGGGA-3'
NEUROG2 phosphomutant	*NEUROG2* (S242A) forward	5'-TTATCGCCCGCC**GC**CCCGGCCGGGTC-3'
NEUROG2 phosphomutant	*NEUROG2* (S242A) reverse	5'-GACCCGGCCGGG**GC**GGCGGGCGATAA-3'
NEUROG2 phosphomutant	*NEUROG2* (S242D) forward	5'-TTTATCGCCCGCC**GA**CCCGGCCGGGTCA-3'
NEUROG2 phosphomutant	*NEUROG2* (S242D) reverse	5'-TGACCCGGCCGGG**TC**GGCGGGCGATAAA-3'
Neuron induction: *NEUROG2, caPRKACA*, and dosromorphin	*PRKACA* (H88Q) forward	5'-GAAACAGATCGAACA**A**ACCCTGAATGAAAAG-3'

Neuron induction: *NEUROG2, caPRKACA*, and dosromorphin	*PRKACA* (H88Q) reverse	5'-CTTTTCATTCAGGGT**T**TGTTCGATCTGTTTC-3'
Neuron induction: *NEUROG2, caPRKACA*, and dosromorphin	*PRKACA* (W197R) forward	5'-GAAGGGCCGCACT**C**GGACCTTGTGCG-3'
Neuron induction: *NEUROG2, caPRKACA*, and dosromorphin	*PRKACA* (W197R) reverse	5'-CGCACAAGGTCC**G**AGTGCGGCCCTTC-3'
Quantitative real-time PCR	*ARID1A* genomic forward	5'-CCTAGCGAAAGCGGCTGAGG-3'
Quantitative real-time PCR	*ARID1A* genomic reverse	5'-CGCAGGCCGTGGAACGGAC-3'
Quantitative real-time PCR	*CREB1* genomic forward	5'-CTTCCCCTCGCAGGAGGG-3'
Quantitative real-time PCR	*CREB1* genomic reverse	5'-CCGCTGGCTCCAGTCCGC-3'
Quantitative real-time PCR	*DLL3* cDNA forward	5'-CAGACTGTGATCCTAGCGCTC-3'
Quantitative real-time PCR	*DLL3* cDNA reverse	5'-GGATCTGCAGCTCGAAGACG-3'
Quantitative real-time PCR	*DLL3* genomic forward (CREB1 ChIP)	5'-GCTGGGACATCTGCTCCTAA-3'
Quantitative real-time PCR	*DLL3* genomic reverse (CREB1 ChIP)	5'-GAGCAAATGGGCAGGAAAT-3'
Quantitative real-time PCR	*DLL3* genomic forward (NEUROD ChIP)	5'-CATTTGCTCCTCACACCTTCC-3'
Quantitative real-time PCR	*DLL3* genomic forward (NEUROD ChIP)	5'-CTCCCAGCTGTATGTAAATGCC-3'
Quantitative real-time PCR	*DLL3* genomic forward (SOX4 ChIP)	5'-CCTGTATGTGAAGACGGAATTTC-3'
Quantitative real-time PCR	*DLL3* genomic reverse (SOX4 ChIP)	5'-CCCTTCGGGCCGCAGCTG-3'
Quantitative real-time PCR	*DLL3 ORF* genomic forward (SOX4 ChIP)	5'-GGTGTGCCGAGCAGGCTGC-3'
Quantitative real-time PCR	*DLL3 ORF* genomic reverse (SOX4 ChIP)	5'-GCATCCGGTGGTAGCAGAGG-3'
Quantitative real-time PCR	*EPC1* genomic forward	5'-GGATGCACAAGGCCCAACCC-3'
Quantitative real-time PCR	*EPC1* genomic reverse	5'-CGGGTCCCGCGGAGTTCG-3'
Quantitative real-time PCR	*GADD45A* cDNA forward	5'-CTCTTGGAGACCGACGCTG-3'
Quantitative real-time PCR	*GADD45A* cDNA reverse	5'-GGCAGGATCCTTCCATTGAGA-3'
Quantitative real-time PCR	*GADD45G* genomic forward	5'-GATCGCACTATGACTCTGGAAG-3'
Quantitative real-time PCR	*GADD45G* genomic reverse	5'-CGGATTCCCACCCTGGAAAAC-3'
Quantitative real-time PCR	*HES6* genomic forward	5'-CTGGTCCTGTAACTTGATGCC-3'

Quantitative real-time PCR	*HES6* genomic reverse	5'-GCCAACACCTGTCGCTCTTG-3'
Quantitative real-time PCR	*HPRT1* cDNA forward	5'-GCTTTCCTTGGTCAGGCAGTA-3'
Quantitative real-time PCR	*HPRT1* cDNA reverse	5'-GTCTGGCTTATATCCAACACTTCGT-3'
Quantitative real-time PCR	*MTF2* genomic forward	5'-CGGAAGGACTGAGCCTCACT-3'
Quantitative real-time PCR	*MTF2* genomic reverse	5'-CAGGGACTGCGCGGCGAG-3'
Quantitative real-time PCR	*NEUROD1* genomic forward (NEUROD ChIP)	5'-GCATGCGCCATATGGTCTTCC-3'
Quantitative real-time PCR	*NEUROD1* genomic reverse (NEUROD ChIP)	5'-CAGATGGGCCACTTTCTTCTG-3'
Quantitative real-time PCR	*NEUROD1* genomic forward (SOX4 ChIP)	5'-CAGCCAAGAGCCTGGAACC-3'
Quantitative real-time PCR	*NEUROD1* genomic reverse (SOX4 ChIP)	5'-CTGGCGACAGATGGGCCAC-3'
Quantitative real-time PCR	*NEUROD4* cDNA forward	5'-CTGAAGAGCATGACAGTATTGAG-3'
Quantitative real-time PCR	*NEUROD4* cDNA reverse	5'-CCTGAATCTCTCAAGGCGAGC-3'
Quantitative real-time PCR	*NEUROD4* genomic forward (NEUROD ChIP)	5'-GCATGCAAATGCCTTGTATTCTC-3'
Quantitative real-time PCR	*NEUROD4* genomic reverse (NEUROD ChIP)	5'-ACAGTTGTACAGTTGGCTGTTC-3'
Quantitative real-time PCR	*NEUROD4* genomic forward (SOX4 ChIP)	5'-GACGTGGCCCTATTGTCCTTC-3'
Quantitative real-time PCR	*NEUROD4* genomic reverse (SOX4 ChIP)	5'-TCAGAGGCAGATGCCATTTCC-3'
Quantitative real-time PCR	*NHLH1* genomic forward	5'-GAGTAAGCCAGAGCTGGTG-3'
Quantitative real-time PCR	*NHLH1* genomic reverse	5'-CCCAACCTCTTCTCTTCCCTC-3'
Quantitative real-time PCR	*NR4A1* genomic forward	5'-TGCGTCAATGGAACCCC-3'
Quantitative real-time PCR	*NR4A1* genomic reverse	5'-GCGCTCCGTGACGCACGG-3'
Quantitative real-time PCR	*NR4A1* ORF genomic forward	5'-GAGTGCTAGAGCTGCTGGTC-3'
Quantitative real-time PCR	*NR4A1* ORF genomic reverse	5'-GGTCAGCAAGAGCCCAAAATAG-3'
Quantitative real-time PCR	*SIK1* genomic forward	5'-CCATTGACGTCGCTTTGACGC-3'
Quantitative real-time PCR	*SIK1* genomic reverse	5'-CCTGACGTCACCGGCCCG-3'
Quantitative real-time PCR	*SMARCA2* genomic forward	5'-CCATCCCAACCTGGTTTACCC-3'

Quantitative real-time PCR	*SMARCA2* genomic reverse	5'-GGGAAGTGCATCGCGACACC-3'
Quantitative real-time PCR	*SMARCC2* genomic forward	5'-CGGATTGCAGTGCGGACCC-3'
Quantitative real-time PCR	*SMARCC2* genomic reverse	5'-CGAGTCCAACGCCTTAACCAC-3'
Quantitative real-time PCR	*SMARCD1* genomic forward	5'-GGAGTTCCGGTTCCGGTTCT-3'
Quantitative real-time PCR	*SMARCD1* genomic reverse	5'-CGCTTGGAGCCACAGACTGG-3'
Quantitative real-time PCR	*SMARCE1* genomic forward	5'-CACTTGGAAACACTCACCCGC-3'
Quantitative real-time PCR	*SMARCE1* genomic reverse	5'-GGAAAGTGCGAGTGCGGAAG-3'
Quantitative real-time PCR	*SOX4* genomic forward	5'-CCCCAAAGCTGTCTGCGCC-3'
Quantitative real-time PCR	*SOX4* genomic reverse	5'-CGGAGACCGCATTGGCTGCTAAG-3'
Quantitative real-time PCR	*SOX4* cDNA forward	5'-GCTGTACAAGGCGCGGACTC-3'
Quantitative real-time PCR	*SOX4* cDNA reverse	5'-CGCGCTTCACCTTCTTCT-3'
Quantitative real-time PCR	*SOX11* cDNA forward	5'-CATGTAGACTAATGCAGCCATTGG-3'
Quantitative real-time PCR	*SOX11* cDNA reverse	5'-CACGGAGCACGTGTCAATTG-3'
shRNA knockdown	shRNA forward	5'-CTACTCGAGAAGGTATATTGCTGTTG-3'
shRNA knockdown	shRNA reverse	5'-CTACTGCAGTAGCCCCTTGAAGTCCGAGGCAGTAGG-3'
shRNA knockdown	Template *CREB1* shRNA	5'-TGCTGTTGACAGTGAGCGAAGCAACCAAGTTGTTGTTCAATAGTGAAGCCACAGATGTA*TTGAACAACAACTTGGTTGCTG*TGCCTACTGCCTCGGA-3'
shRNA knockdown	Template Scramble shRNA	5'-TGCTGTTGACAGTGAGCGCAGGAATTATAATGCTTATCTATAGTGAAGCCACAGATGTA*TAGATAAGCATTATAATTCCTA*TGCCTACTGCCTCGGA-3'
shRNA knockdown	Template *GADD45A* shRNA 1	5'-TGCTGTTGACAGTGAGCGCACCAAATATGTTAAAGTTTAATAGTGAAGCCACAGATGTA*TTAAACTTTAACATATTTGGTT*TGCCTACTGCCTCGGA-3'
shRNA knockdown	Template *GADD45A* shRNA 2	5'-TGCTGTTGACAGTGAGCGCAACCAAATATGTTAAAGTTTATAGTGAAGCCACAGATGTA*TAAACTTTAACATATTTGGTTT*TGCCTACTGCCTCGGA-3'
shRNA knockdown	Template *GADD45G* shRNA 1	5'-TGCTGTTGACAGTGAGCGACGGGCAATGCTTCGTTTTCTATAGTGAAGCCACAGATGTA*TAGAAAACGAAGCATTGCCCGG*TGCCTACTGCCTCGGA-3'

shRNA knockdown	Template *GADD45G* shRNA 2	5'-TGCTGTTGACAGTGAGCGCTCGCACTATGACTCTGGAAGATAGTGAAGCCACAGATGTA*TCTTCCAGAGTCATAGTGCGAT*TGCCTACTGCCTCGGA-3'
shRNA knockdown	Template *SOX4* shRNA 1	5'-TGCTGTTGACAGTGAGCGCAAGGACAGACGAAGAGTTTAATAGTGAAGCCACAGATGTA*TTAAACTCTTCGTCTGTCCTTT*TGCCTACTGCCTCGGA-3'
shRNA knockdown	Template *SOX4* shRNA 2	5'-TGCTGTTGACAGTGAGCGCGACGAAGAGTTTAAAGAGAAATAGTGAAGCCACAGATGTA*TTTCTCTTTAAACTCTTCGTCT*TGCCTACTGCCTCGGA-3'
shRNA knockdown	Template *SOX11* shRNA	5'-TGCTGTTGACAGTGAGCG*ACAGGAGAGAATTCTACATTTA*TAGTGAAGCCACAGATGTA*TAAATGTAGAATTCTCTCCTGC*TGCCTACTGCCTCGGA-3'

*Note: Sense shRNA sequences are underlined. Antisense shRNA sequences are underlined and italicized. Nucleotide substitutions for site-directed mutagenesis are bold.