


Spring 5-15-2018

Defining Neuronal Identity Using MicroRNA-Mediated Reprogramming

Matthew James Mccoy
Washington University in St. Louis

Follow this and additional works at: https://openscholarship.wustl.edu/art_sci_etds

 Part of the [Bioinformatics Commons](#), [Genetics Commons](#), and the [Neuroscience and Neurobiology Commons](#)

Recommended Citation

McCoey, Matthew James, "Defining Neuronal Identity Using MicroRNA-Mediated Reprogramming" (2018). *Arts & Sciences Electronic Theses and Dissertations*. 1560.
https://openscholarship.wustl.edu/art_sci_etds/1560

This Dissertation is brought to you for free and open access by the Arts & Sciences at Washington University Open Scholarship. It has been accepted for inclusion in Arts & Sciences Electronic Theses and Dissertations by an authorized administrator of Washington University Open Scholarship. For more information, please contact digital@wumail.wustl.edu.

WASHINGTON UNIVERSITY IN ST. LOUIS

Division of Biology and Biomedical Sciences
Molecular Genetics & Genomics

Dissertation Examination Committee:

Andrew S. Yoo, Chair

John R. Edwards

Sarah C. R. Elgin

James B. Skeath

Ting Wang

Defining Neuronal Identity Using MicroRNA-Mediated Reprogramming

by

Matthew J. McCoy

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May 2018
St. Louis, Missouri

© 2018, Matthew J. McCoy

Table of Contents

List of Figures	v
List of Tables	vii
Acknowledgments.....	viii
Abstract	x
Chapter 1: Introduction.....	1
Chapter 2: MicroRNA-Induced Epigenetic Remodeling Enables Neuronal Subtype-Specific Reprogramming of Adult Human Fibroblasts	6
2.1 Results	7
2.1.1 Neuronal Conversion of Human Adult Fibroblasts With miR-9/9*-124 Alone	7
2.1.2 Functional Properties and Stability of miNs	10
2.1.3 Transcriptional Profiling of miNs	11
2.1.4 Transcriptional Changes in Epigenetic Machinery	15
2.1.5 Dynamic Regulatory Events During Neuronal Reprogramming	15
2.1.6 DNA Methylation Profiling of miNs	18
2.1.7 Chromatin Remodeling in miNs	22
2.1.8 Erasure of Fibroblast Epigenetic Identity and Gain of Neuronal Chromatin Architecture.....	25
2.1.9 MicroRNA-induced Chromatin Remodeling at Heterochromatin Regions in Fibroblasts	26
2.1.10 Chromatin Remodeling is Required for Direct Conversion.....	27
2.1.11 Instructing the miRNA-induced Neurogenic State to a Motor Neuron Fate.....	28
2.1.12 Electrophysiological properties of Moto-miNs.....	32
2.1.13 Transcriptional Profiling of Moto-miNs	36
2.1.14 Transcriptional Activation of ISL1 and LHX3 Genomic Targets.....	37
2.1.15 Comparison of Moto-miNs to Endogenous Spinal Cord Motor Neurons.....	37
2.2 Author Contributions.....	39
2.3 Acknowledgements	39
2.4 Supplemental Figures	40
2.5 Materials and Methods	53
2.6 Data Availability	71
Chapter 3: LONGO—An R Package for Interactive Gene Length Dependent Analysis for Neuronal Identity	72

3.1 Methods.....	73
3.1.1 LONGO algorithm and framework.....	73
3.1.2 Measuring Jensen-Shannon Divergence.....	75
3.1.3 Developing the long gene quotient.....	75
3.1.4 Gene ontology (GO) analysis.....	76
3.1.5 Cellular reprogramming and analyses.....	77
3.2 Results.....	78
3.2.1 LGE identifies neurons upon differentiation and maturation during development.....	78
3.2.2 LGE identifies neurons successfully reprogrammed from non-neuronal cells.....	82
3.2.3 GO analysis of neuronal differentiation.....	84
3.3 Supplemental Figures.....	87
3.4 Author Contributions.....	92
3.5 Acknowledgements.....	92
3.6 Funding.....	92
3.7 Data Availability.....	92
Chapter 4: Single-cell Sequencing Reveals Transcriptomic Features of Mature Neurons.....	94
4.1 Results.....	95
4.1.1 Single-cell RNA-seq of microRNA-mediated reprogramming reveals a homogenous reprogramming trajectory.....	95
4.1.2 Long gene expression distinguishes mature from immature neuronal gene expression.....	98
4.1.3 BRG1 inhibition reduces LGE.....	102
4.1.4 Inhibiting LGE reduces spontaneous activity.....	104
4.2 Supplemental Figures.....	106
4.3 Materials and Methods.....	108
4.4 Author contributions.....	110
4.5 Acknowledgements.....	111
4.6 Funding.....	111
Chapter 5: Conclusions and Discussion.....	113
5.1 Molecular Mechanisms of MicroRNA-Mediated Reprogramming.....	114
5.1.1 Chromatin Remodeling Accompanies Cell Fate Conversion.....	114
5.1.2 A Modular Neuronal State.....	115
5.2 Long Gene Expression Identifies Neurons.....	116

5.2.1 LONGO is a Computation Tool that Measures LGE to Identify Neurons.....	116
5.2.2 Long Gene Expression Distinguishes Mature from Immature Neurons	117
5.3 References	118

List of Figures

Chapter 2:

Fig. 1. Direct Conversion of Young and Old Primary Adult Human Fibroblasts into Neurons via miRNA Overexpression.....	8
Fig. 2. Gene Expression Profiling Reveals Pan-Neuronal Identity Induced by miRNAs Alone...13	
Fig. 3. Time Series Transcriptome Analysis Reveals Early Dynamic Gene Expression Changes Followed by a Stable Transcriptome Switch	17
Fig. 4. MiR-9/9*-124 Alter DNA Methylation at Neuronal Loci.....	20
Fig. 5. MiR-9/9*-124 Globally Changes Chromatin Accessibility	23
Fig. 6. MiRNA-Induced Neuronal Competence Enables Motor Neuron Transcription Factors, ISL1 and LHX3, to Determine Motor Neuron Identity.	30
Fig. 7. Functional Properties and Gene Expression Profile of Moto-miNs	34
Fig. S1. MiRNA-Mediated Conversion into Neuronal Fate is Stable.	40
Fig. S2. Expression Changes in Epigenetic Modifiers During Reprogramming.....	42
Fig. S3. Pre-existing Heterochromatic Neuronal Loci Open in Response to miR-9/9*-124 Expression.	43
Fig. S4. Loss of BRG1 Prevents Neuronal Fate Acquisition.....	45
Fig. S5. Identification of Transcription Factors for Defining Motor Neuron Specific Conversion	46
Fig. S6. Addition of ISL1/LHX3 to miR-9/9*-124 Increases Functional Maturity and Generates Motor Neuron Transcriptional Network.	48
Fig. S7. Direct Comparison of Moto-miN Transcriptome to in vivo Mouse Motor Neurons by Translating Ribosomal Affinity Purification (TRAP) Sequencing.....	50

Chapter 3:

Fig. 1. Illustration of LONGO output using Rat BodyMap data.	74
Fig. 2. LONGO output of neuronal differentiation and maturation during development.....	81
Fig. 3. LONGO output of somatic cell direct neuronal reprogramming.	83
Fig. 4. LONGO interactive GO analysis snapshot using incorporated topGO R package.	85
Fig. S1. LGE identifies neuronal tissues, while PCA only separates tissues.....	87
Fig. S2. LONGO output of neuronal differentiation in vivo and in vitro.....	88
Fig. S3. LQ cutoff of 0.25 stringently distinguishes neurons from non-neuronal cells.....	89

Fig. S4. Medium spiny neurons reprogrammed with miR-9/9*-124+CDM from fibroblasts of two unrelated adults shows variation in the proper acquisition of neuronal morphological features.....90

Fig. S5. LONGO output of small molecule and shPTBP1-induced reprogramming.91

Chapter 4:

Fig. 1. Direct conversion of human adult fibroblasts yields a homogenous population of neurons.96

Fig. 2. Long gene expression distinguishes mature from immature neurons.100

Fig. 3. Functional neuronal BAF complex is necessary for LGE.....103

Fig. 4. Inhibition of long gene expression reduces spontaneous electrical activity of mature neurons.....105

Fig. S1. PHF10/BAF45a to DPF3/BAF45c BAF complex subunit switching during neuronal reprogramming.....106

Fig. S2. Inhibition of BRG1 expression impairs maturation of reprogramming neurons.107

List of Tables

Chapter 2:

Supplemental Table 1	52
----------------------------	----

Acknowledgments

This work was made possible with the guidance and support of my thesis advisor, Andrew Yoo. Your door was always open, and you were never too busy to discuss science, whether it was directly related to my current work or not. In this way especially, you provided a model for mentoring.

Thank you to my thesis committee: Dr. John Edwards, Dr. Sarah (Sally) Elgin, Dr. James (Jim) Skeath, and Dr. Ting Wang. You helped provide focus to my thesis and helped me accomplish my academic goals.

I would especially like to thank Jim Skeath. You are a cornerstone of the student community, always willing to make time for guidance and advice. Thank you for your mentoring and friendship—I will always be grateful for all the time we spent on the river, even though you always caught more fish than me.

Thank you to my colleagues in the Department of Developmental Biology. My conversations with Dr. Craig Micchelli and Dr. Irving Boime were always stimulating and encouraging. I would especially like to thank our Department Chair, Dr. Lilianna (Lila) Solnica-Krezel—you constantly provide a source of inspiration and a model for leadership, showing the utmost respect for every member of our community.

Thank you to all members of the Yoo lab, past and present. From you, I learned that science is more than ideas and experiments. Thank you especially to WooKyung Kim and Daniel Abernathy for your scientific contributions to my thesis.

Thank you to the various funding sources that made this work possible. I was personally supported by the Interface of Psychology, Neuroscience, and Genetics Program (IPNG; T32GM081739; Barch, PI). Thank you, Dr. Deanna Barch and other members of the IPNG

program for a broad education in this important interface of disciplines. The experiments in this dissertation were supported by the NIH Director’s Innovator Award (DP2NS083372-01), the Missouri Spinal Cord Injury/Disease Research Program (SCIDRP), the Cure Alzheimer’s Fund (CAF) and the Presidential Early Career Award for Scientists and Engineers (PECASE).

Outside the lab, a small collection of friends and family made this work possible. To Dr. Carl Sagan, you teach me that every day is a new day. To my friends Chris and Cole, you help keep me sane with regular diversions into the unknown. To my brothers Dan and Connor, I cannot tell you how proud I am to be your kin—you inspire me with the boundaries you each push. To my brother, Sean, thank you for teaching me to be present, to be grateful for what you have for the short time you have it. To my parents Jeff and Marge, I can never repay you for the success and happiness you made possible for me. Also, I literally cannot repay you—just remember, you told me I could become a scientist if it was really what I wanted.

Finally, I have to thank Amy Herbert (Dr. Amy Herbert, by the time this is accepted). You are my paralog and perfect complement. This ends our first chapter, and I can’t wait for all the others to come.

Matthew J. McCoy

Washington University in St. Louis

May 2018

ABSTRACT OF THE DISSERTATION

Defining Neuronal Identity Using MicroRNA-Mediated Neuronal Reprogramming
by

Matthew J. McCoy

Doctor of Philosophy in Biology and Biomedical Sciences

Molecular Genetics & Genomics

Washington University in St. Louis, 2018

Professor Andrew S. Yoo, Chair

Cell fate reprogramming is transforming our understanding of the establishment and maintenance of cellular identity. In addition, reprogramming holds great promise to model diseases affecting cell types that are prohibitively difficult to study, such as human neurons. Overexpression of the brain-enriched microRNAs (miRNAs), miR-9/9* and miR-124 (miR-9/9*-124) results in reprogramming human somatic cells into neurons and has recently been used to generate specific neuronal subtypes affected in neurodegenerative disorders. However, the mechanisms governing the ability of miR-9/9*-124 to generate alternative subtypes of neurons remained unknown. In this thesis, I report that overexpressing miR-9/9*-124 triggers reconfiguration of chromatin accessibility, DNA methylation, and mRNA expression to induce a default neuronal state. MiR-9/9*-124-induced neurons (miNs) are functionally excitable and are uncommitted towards specific subtypes yet possess open chromatin at neuronal subtype-specific loci, suggesting such identity can be imparted by additional lineage-specific transcription factors. Consistently, we show ISL1 and LHX3 selectively drive conversion to a highly homogenous population of human spinal cord motor neurons. This work shows that modular synergism

between miRNAs and neuronal subtype-specific transcription factors can drive lineage-specific neuronal reprogramming, thereby providing a general platform for high-efficiency generation of distinct subtypes of human neurons.

Since many neurodegenerative diseases occur after development, modeling them requires reprogramming methods capable of generating functionally mature neurons. However, few robust molecular hallmarks existed to identify such neurons, or to compare efficiencies between reprogramming methods. Recent studies demonstrated that active long genes (>100 kb from transcription start to end) are highly enriched in neurons, which provided an opportunity to identify neurons based on the expression of these long genes. We therefore worked to develop an R package, LONGO, to analyze gene expression based on gene length. We developed a systematic analysis of long gene expression (LGE) in RNA-seq or microarray data to enable validation of neuronal identity at the single-cell and population levels. By combining this conceptual advancement and statistical tool in a user-friendly and interactive software package, we intended to encourage and simplify further investigation into LGE, particularly as it applies to validating and improving neuronal differentiation and reprogramming methodologies. Using this tool, I found by single-cell RNA sequencing that microRNA-mediated neuronal reprogramming of human adult fibroblasts yields a homogenous population of mature neurons, and that LGE distinguishes mature from immature neurons. I found that LGE correlates with expression of neuronal subunits of the Swi/Snf-like (BAF) chromatin remodeling complex, such as ACTL6B/BAF53b. Finally, I found that the loss of a functional neuronal BAF complex, as well as chemical inhibition of topoisomerase I, decreases LGE and reduces spontaneous electrical activity. Together, these results provide mechanistic insights into microRNA-mediated neuronal reprogramming, and demonstrate a transcriptomic feature of functionally mature

neurons.

Chapter 1: Introduction

In part from:

MicroRNA-induced Epigenetic Remodeling During Direct Cell-Fate Conversion of Adult Human Fibroblasts

Daniel G. Abernathy*, Wookyung Kim*, Matthew J. McCoy*, Allie Lake, Rebecca Ouwenga, Xiaoyun Xing, Daofeng Li, Hyung Joo Lee, Robert O. Heuckeroth, Joseph D. Dougherty, Ting Wang, Andrew S. Yoo

*Authors contributed equally

Cell Stem Cell 21, 332-348, September 7, 2017

Copyright © 2017 Elsevier Inc. All Rights Reserved.

LONGO: An R Package for Interactive Gene Length Dependent Analysis for Neuronal Identity

Matthew J. McCoy*, Alexander J. Paul*, Matheus B. Victor, Michelle Richner, Harrison W. Gabel, Haijun Gong, Andrew S. Yoo and Tae-Hyuk Ahn

*Authors contributed equally

Bioinformatics, In Press

Copyright © 2018 Oxford University Press, All Rights Reserved

Single-cell Sequencing Reveals Transcriptomic Features of Mature Neurons

Matthew J. McCoy*, Daniel G. Abernathy*, Harrison W. Gabel, Samantha A. Morris, Andrew S. Yoo

*Authors contributed equally

Manuscript in preparation

Understanding genetic pathways that specify neuronal cell fates during development has enabled directed differentiation of pluripotent stem cells to specific neuronal subtypes (Perrier et al., 2004; Wichterle et al., 2002). This knowledge has been further leveraged to directly convert (or reprogram) non-neuronal somatic cells into neurons via ectopic expression of pro-neural transcription factors (TFs) or neurogenic miRNAs with TFs (Mertens et al., 2016). These direct conversion modalities may prove invaluable in the study of late-onset neurodegenerative diseases, as the original age of human fibroblasts is maintained in converted neurons (Huh et al., 2016; Mertens et al., 2015) in contrast to the cellular rejuvenation observed in induced pluripotent stem cells (Horvath, 2013; Miller et al., 2013). Interestingly, the miRNA-mediated reprogramming approach can achieve high conversion efficiency in adult human fibroblasts, which may provide unique opportunities in modeling neurological disorders using patient-derived neurons (Victor et al., 2014). However, despite the advantages of direct reprogramming, little is known about the epigenetic and molecular events that accompany direct cell-fate conversion.

MiRNAs regulate genetic pathways by binding to their target transcripts and repressing their expression (Pasquinelli, 2012). Target specificity is governed largely through short sequence complementarity within the 5' end of a miRNA enabling a single miRNA to target hundreds of mRNA transcripts (Boudreau et al., 2014; Chi et al., 2009). Moreover, a single mRNA can be targeted by multiple miRNAs, markedly enlarging the effect on single gene repression (Wu et al., 2010). The convergence of genetic controls by miRNAs towards a specific biological process is exemplified by miR-9/9* and miR-124 miRNAs activated at the onset of neurogenesis (Conaco et al., 2006; Makeyev et al., 2007). For example, miR-9* and miR-124 synergistically initiate subunit switching within BAF chromatin remodeling complexes (Staaahl et al., 2013; Yoo et al., 2009) while separately

repressing the neuronal cell-fate inhibitors REST, Co-REST and SCP1 (Packer et al., 2008; Visvanathan et al., 2007). These examples suggest that miR-9/9* and miR-124 target components of genetic pathways that antagonize neurogenesis to promote a neuronal identity during development.

Co-expressing miR-9/9* and miR-124 (miR-9/9*-124), with TFs enriched in the cortex and striatum directly converts primary adult human fibroblasts to cortical and striatal medium spiny neurons, respectively (Yoo et al., 2011; Victor et al., 2014). However, the same TFs without miR-9/9*-124 fail to trigger neuronal conversion (Victor et al., 2014; Yoo et al., 2011), suggesting that the miRNA-induced neuronal state is permissive to terminal selector TFs which, upon determination of a neuronal fate, initiate and advance mature subtype-identities (Stefanakis et al., 2015).

In my thesis work, my colleagues and I systematically investigated a miRNA-induced neuronal state in adult human cells. Longitudinal analyses of the transcriptome, genome-wide DNA-methylation and chromatin accessibilities revealed that miR-9/9*-124 induced extensive remodeling of the epigenome, including simultaneous activation of a pan-neuronal program and the reconfiguration of chromatin accessibilities. These changes precede the emergence of differentially methylated genomic regions. Because miR-9/9*-124 also led to the opening of genomic loci for multiple subtype-specific genes including established motor neuron markers, we postulated that motor neuron-enriched transcription factors would cooperate with miR-9/9*-124 to specify a motor neuron lineage. As such, we demonstrate that co-expressing TFs ISL1 and LHX3 along with miR-9/9*-124 generated a highly pure population of human spinal cord motor neurons. Taken together, these results demonstrate that miR-9/9*-124 opens the neurogenic potential of adult human fibroblasts and provides a platform for subtype-specific neuronal conversion of human cells.

With the growing interest in advancing these reprogramming methods, there is a need to determine efficiencies and compare outcomes across different methods. However, these efforts are hindered by the lack of reliable measures for neuronal identity. Electrical activity, one hallmark of neurons, is typically only measured in a handful of cells in a given experiment and can be biased, as cells targeted for analysis may not be representative of the entire population. Expression of a small subset of neuronal proteins measured by immunocytochemistry (ICC) is also used to mark neuronal identity but is often subjective and highly variable. Furthermore, interpretation of ICC is unreliable, given that extrinsic factors used during neuronal reprogramming may selectively increase the expression of several neuronal genes without necessarily resulting in complete cell-fate transition. Whole transcriptome data has potential to overcome these limitations, and several complex computational approaches such as iterative principal component analysis (PCA) and iterative weighted gene co-expression network analysis (WGCNA) have been successful in classifying cell types (Tasic, et al., 2016). However, these methods require a positive control and at least 15 samples for accuracy. A simpler and more easily assessable measure of neuronal identity at the transcriptome level would greatly facilitate the development and validation of neuronal reprogramming methods.

Recent studies have shown that long genes (>100 kb from transcription start to end) are selectively expressed in neurons (Gabel, et al., 2015; King, et al., 2013; Sugino, et al., 2014). This unique feature of neurons may provide an opportunity to utilize measurements of long gene expression (LGE) in transcriptome data to distinguish neurons from non-neuronal cells quickly and easily. In my thesis work, I helped developed LONGO (<https://github.com/biohpc/longo>) (an open source computational package based in R with the interactive web-supporting library, R-Shiny) to standardize measurement of LGE within

RNA-seq and microarray data formats. By providing an interactive and convenient analysis of LGE within transcriptome data, LONGO allows researchers to identify neurons, and to explore gene ontology (GO) terms associated with enriched neuronal genes. Additionally, by applying LGE analysis to single-cell RNA-seq of microRNA-mediated neuronal reprogramming, I found that the onset of LGE correlates with subunit switching of the BAF chromatin remodeling complex. This same subunit switching can be observed during differentiation of neurons from neural progenitor cells, a process which, significantly, is controlled by the expression of miR-9/9* and miR-124 in vertebrates (Yoo, et al., 2009). I also show that disrupting the functional BAF complex decreases LGE, which results in less spontaneous electrical activity. Together, this work demonstrates that LGE distinguishes fully differentiated and functionally mature neurons from immature neurons and non-neuronal cells, which will allow for the innovation and improvement of neuronal reprogramming and differentiation methods to model human disease.

Chapter 2: MicroRNA-Induced Epigenetic Remodeling Enables Neuronal Subtype-Specific Reprogramming of Adult Human Fibroblasts

Adapted from:

MicroRNA-induced Epigenetic Remodeling During Direct Cell-Fate Conversion of Adult Human Fibroblasts

Daniel G. Abernathy*, Wookyoung Kim*, Matthew J. McCoy*, Allie Lake, Rebecca Ouwenga, Xiaoyun Xing, Daofeng Li, Hyung Joo Lee, Robert O. Heuckeroth, Joseph D. Dougherty, Ting Wang, Andrew S. Yoo

*Authors contributed equally

Cell Stem Cell 21, 332-348, September 7, 2017

Copyright © 2017 Elsevier Inc. All Rights Reserved.

2.1 Results

2.1.1 Neuronal Conversion of Human Adult Fibroblasts With miR-9/9*-124 Alone

To dissect how miRNAs alone contribute to neuronal conversion, we first tested the ability of miR-9/9*-124 to convert primary human fibroblasts collected from multiple adult individuals from ages 22 to 68 into microRNAs-induced neurons (miNs). We transduced fibroblasts with lentivirus containing a doxycycline (Dox)-inducible promoter driving miR-9/9*-124 and BCL-XL expression (Victor et al., 2014) (Figures 1A and S1A). At 35 days post-transduction, we evaluated cell morphology and the expression of neuronal markers MAP2, TUBB3 and NEUN by immunocytochemistry (Figure 1B). Strikingly, miR-9/9*-124 alone converted 80% of the fibroblasts to neuronal cells displaying complex neurite outgrowth and neuronal marker expression (Figures 1B and 1C). The converted cells stained positive for voltage-gated sodium channels SCN1A and Ankyrin G, which localized at axonal initial segments with a characteristic polarized staining pattern (Figure 1D). The synaptic vesicle marker SV2 displayed defined puncta along neurites, consistent with the adoption of a neuronal fate (Figure 1D).

Figure 1

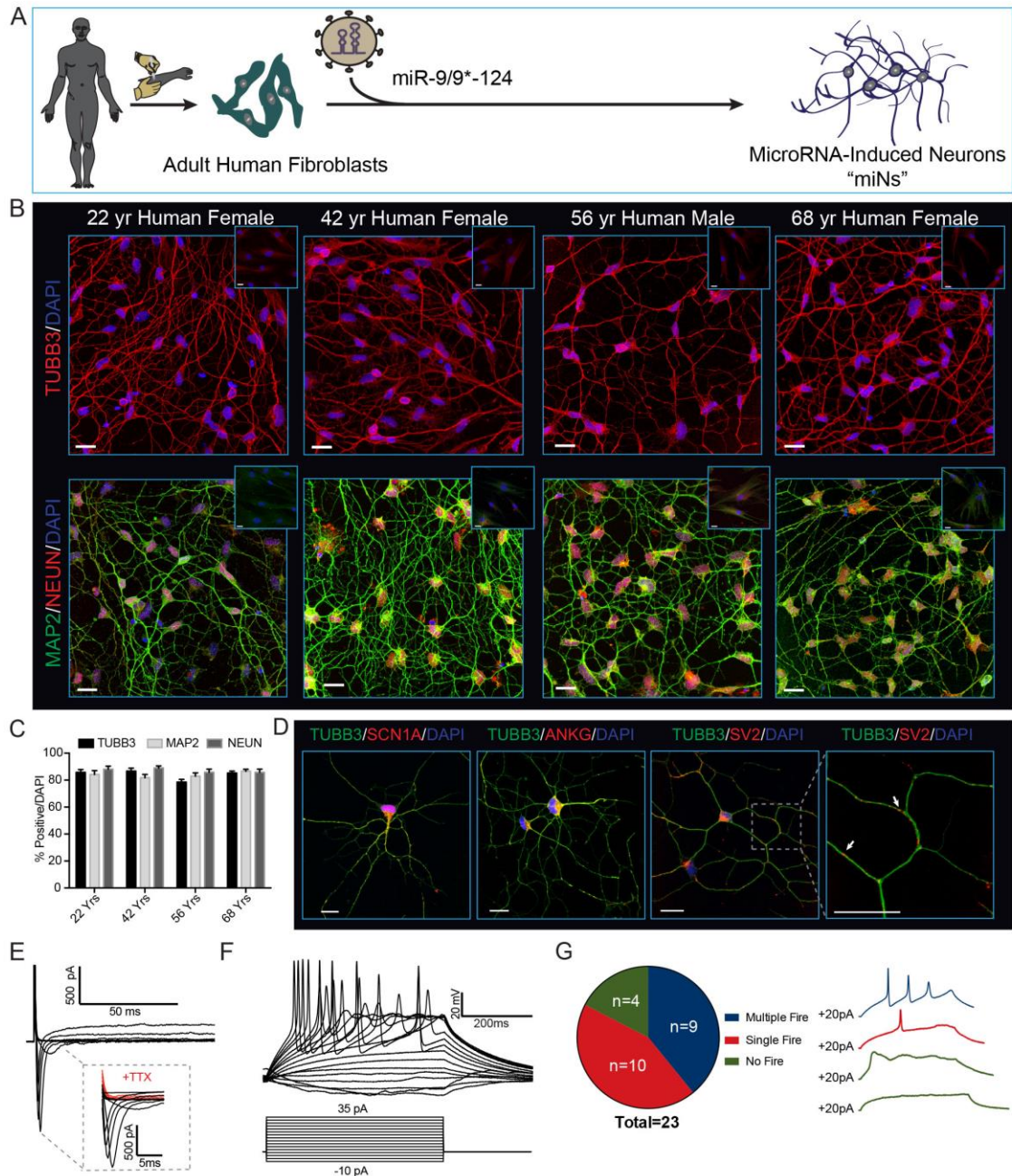


Fig. 1. Direct Conversion of Young and Old Primary Adult Human Fibroblasts into Neurons via miRNA Overexpression.

(A) Experimental scheme for miR-9/9*-124 mediated direct neuronal conversion. (B) Adult human fibroblasts ectopically expressing miR-9/9*-124 for 35 days immunostained for the pan-neuronal markers TUBB3, MAP2 and NEUN. Insets represent starting fibroblasts co-stained as negative controls. Scale bars = 20µm. (C) Quantification of TUBB3, MAP2 and NEUN positive cells over total number of cells (DAPI). For TUBB3 and MAP2, only cells

with processes at least three times the length of the soma were counted. For NEUN, only cells with proper nuclear localization were counted. Data are represented as mean \pm SEM. 22 Yr Female n = 238 cells, 42 Yr Female, n = 100 cells, 56 Yr Male n = 171 cells, and 68 Yr Female n = 216. (D) Converted neurons display hallmark sodium channel (SCN1A), axonal initial segment (ANKG) (left) and synaptic vesicle (SV2) (right) staining patterns. Scale bars = 20 μ m. (E) Representative traces of TTX-sensitive inward and potassium whole-cell currents. (F) Repetitive AP waveforms in response to 500 ms current injections recorded from converted neurons in monoculture. (G) Summary of AP firing patterns observed in 23 neurons recorded in current-clamp mode (left) and representative waveforms within each firing pattern recorded (right). See also Figure S1.

2.1.2 Functional Properties and Stability of miNs

To determine if miNs displayed membrane excitability, we performed whole-cell recording on miNs without glial co-culture. All cells recorded (23/23 cells) exhibited fast TTX-sensitive inward currents upon depolarization (Figure 1E), while 19/23 cells fired action potentials (APs) during current injections (9 cells with multiple APs and 10 with single APs) (Figures 1F and 1G). Similar current (I)-voltage (V) curve relationships were observed between cells that fired multiple or single APs (Figure S1B). All miNs had stable hyperpolarized resting membrane potentials ranging from -52.7 to -84.5 mV with a mean value of $-69.63 \text{ mV} \pm 2.15 \text{ mV}$ (S.E.M.) (Figures S1C and S1D). There was no correlation between capacitance and firing patterns suggesting this membrane property is not as an accurate measure of neuronal maturation (Figure S1E). Together, this data indicates that miNs exhibit the membrane functionality of neurons.

The minimum duration of miRNA expression required for neuronal conversion was determined by inactivating the Dox-inducible promoter at 3-day intervals by Dox removal beginning at day 9 until reprogramming day 30 (Figure S1F). Loss of fibroblast identity and gain of neuronal identity was assayed by analyzing fibroblast-specific protein (FSP1) and MAP2 expression, respectively. Surprisingly, we saw a reduction in the number of FSP1-positive cells and the appearance of MAP2-positive cells after only 9 days of miRNA expression. However, effective switching of a cell population from FSP1- to MAP2-positive cells required 30 days of miR-9/9*-124 expression (Figure S1F). These data reveal the non-synchronous process of the neuronal conversion and temporal requirements for highly efficient neuronal conversion. The stability of miRNA-induced neuronal conversion was determined by following miNs for an additional 30 days after removing miR-9/9*-124

exposure (Figure S1G, top). The majority of miNs remained as post-mitotic neurons (marked by the absence of Ki67, a cell proliferation marker) expressing MAP2, TUBB3, NeuN and NCAM, in contrast to non-converted fibroblasts (Figure S1G, left panel), indicating that the morphological and protein expression changes that accompany miR-9/9*-124-mediated conversion of adult human fibroblasts are stable after 30 days of miR-9/9*-124 expression.

2.1.3 Transcriptional Profiling of miNs

To further explore the miR-9/9*-124-mediated neuronal output, we profiled the transcriptome of starting human adult fibroblasts and miNs by RNA-Seq after 30 days of neuronal conversion. We identified 2,692 differentially expressed genes (DEGs) in miNs representing 1,251 up-regulated and 1,441 down-regulated genes in comparison to fibroblasts (log fold change ≥ 2 ; adjusted p-value < 0.01) (Figure 2A). A robust down-regulation of fibroblast-specific genes (for instance, *S100A4*, *VIM*, *FBN1* and, *COL13A1*) was accompanied an enrichment of pan-neuronal genes including *MAP2*, *SCN1A*, *SNAP25*, *NRCAM*, and *NEFM* (Figure 2A, 2B top two traces). Analysis of the top 10 gene ontology (GO) terms revealed that upregulated genes in miNs were primarily enriched with terms related to neuronal development and functionality (Figure 2C) while downregulated genes were associated with fibroblast functions (Figure 2C). Downregulated genes also included key cell-cycle components (data not shown), consistent with the previous finding that miR-9/9*-124 expression in human fibroblasts caused rapid cell cycle exit without transitioning through a neural stem cell-like state (Yoo et al., 2011). Interestingly, neuronal subtype-specific genes such as *TH* (dopaminergic neurons), *GABBR2*, *GABRI*, and *GAD2* (GABAergic neurons), *CHAT* (cholinergic neurons), or *DARPP-32* (striatal medium spiny

neurons), were not significantly enriched in miNs (Figure 2B, bottom two traces as examples). Overall, our transcriptome analyses show that miR-9/9*-124 induce a neuronal state characterized by the loss of fibroblast identity and the presence of a pan-neuronal gene expression program without commitment to a particular subtype.

Figure 2

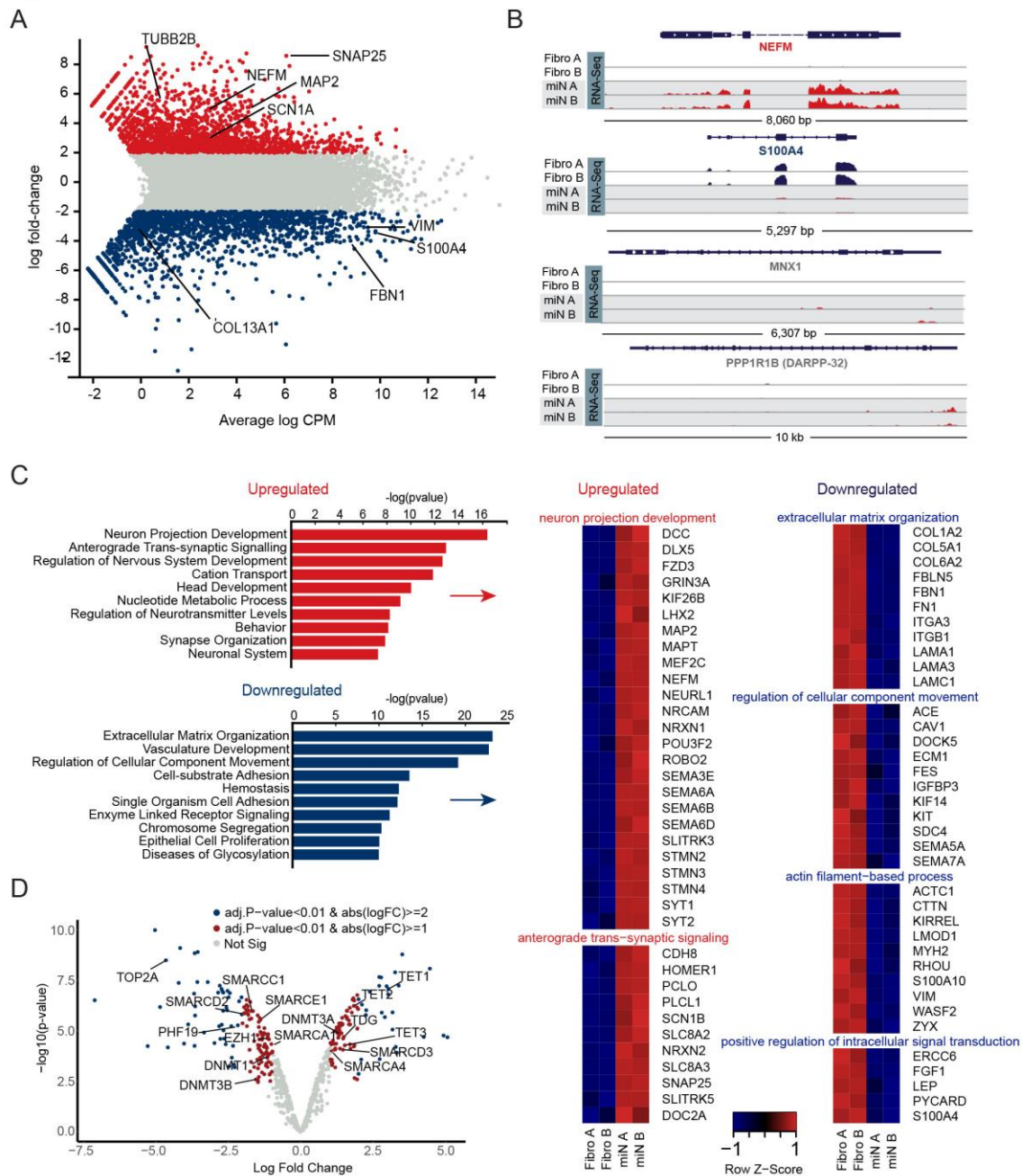


Fig. 2. Gene Expression Profiling Reveals Pan-Neuronal Identity Induced by miRNAs Alone.

(A) Genome wide expression analysis of miNs and starting fibroblasts by RNA-seq at day 30 of reprogramming. Plot shows the relationship between average gene expression (logCPM) and log fold-change of miNs compared to fibroblasts. A selection of pan-neuronal and fibroblast-specific genes are highlighted in black text. Blue = fold change < -2 $\log_2 p < 0.01$ (more abundant in fibroblasts), grey = fold change > -2 $\log_2 p > 0.01$, and red = fold change > 2 $\log_2 p < 0.01$ (more abundant in miNs). FDR < 0.01 .

(B) Representative genome browser snapshots demonstrating increased expression for a pan-neuronal gene (*NEFM*), loss of fibroblast marker gene expression (*S100A4*), and absence of neuronal subtype marker gene expression (*MNX1*, motor neuron marker; *DARPP-32*, medium spiny neuron marker).

(C) Gene ontology (GO) terms associated with genes upregulated in miNs (red) and GO terms associated with genes downregulated in miNs (blue). Right, examples of genes that fall within top GO categories listed.

(D) Volcano plot representing chromatin modifier genes differentially expressed between fibroblasts and miNs. Blue dot, $\text{abs}(\log\text{FC}) > 2$ and $p < 0.01$, red dot, $\text{abs}(\log\text{FC}) > 1$ and $p < 0.01$, grey dot, no significant difference. See also Figure S2.

2.1.4 Transcriptional Changes in Epigenetic Machinery

Epigenetic modifications can markedly affect gene expression and developmental programs (Cantone and Fisher, 2013). Our gene expression studies showed that when compared to fibroblasts, miNs had markedly altered expression of genes involved in DNA methylation, histone modifications, chromatin remodeling and chromatin compaction (Figures 2D and S2). For instance, the TET family of proteins, key mediators of DNA-demethylation (Wu and Zhang, 2011) were upregulated along with the brain-enriched *de novo* DNA-methyltransferase *DNMT3A* (Lister et al., 2013), while *DNMT3B* (Okano et al., 1999) mRNA levels were reduced in miNs compared to fibroblasts (Figures 2D and S2). Transcripts encoding histones and histone variants were altered (Figure S2) suggesting that changes in histone composition may accompany neuronal conversion of human fibroblasts. Genes encoding chromatin remodelers important for neurogenesis like *CHD5*, *CHD7* and components of the BAF chromatin remodeling complex were expressed at higher levels in miNs than in fibroblasts (Egan et al., 2013; Feng et al., 2013; Lessard et al., 2007) (Figures 2D and S2). Additionally, the main DNA topoisomerase 2 family member expressed in miNs is *TOP2B*, which replaces the non-neuronal *TOP2A*, a switch that has been observed during normal neuronal differentiation (Tiwari et al., 2012) (Figures 2D and S2). In sum, dynamic changes and switches within diverse epigenetic modifiers coincide with neuronal differentiation and appear to be recapitulated in direct neuronal conversion of fibroblasts by miR-9/9*-124.

2.1.5 Dynamic Regulatory Events During Neuronal Reprogramming

Because transcriptome profiling at day 30 only provided a snapshot of the functional output of neuronal reprogramming, we explored transcriptome dynamics by profiling

intermediary timepoints (days 3, 6, 10, and 20) by RNA-seq. The Dynamic Regulatory Events Miner (DREM) (Schulz et al., 2012) reports 13 paths of co-regulated, differentially expressed genes during the first 20 days of neuronal conversion (Figure 3A). Combining DREM with predicted TF-gene binding interactions (Ernst et al., 2010) revealed several potential TFs associated with major regulatory events (bifurcations in each path; Figure 3A). Altogether, major regulatory events were observed before day 10, suggesting genetic networks were established within 10 days of miR-9/9*-124 expression. Thereafter, the directionality of gene expression stayed the same but transcript levels markedly changed. This transcriptional maturation over time may explain why the acquisition of functional neuronal characteristics requires 30 days of culture. GO analyses of each path revealed enrichment of neuronal terms in the most upregulated path, whereas down-regulated paths were enriched for cell cycle and extracellular matrix-related terms (Figure 3B), consistent with results for day 30 transcriptome profiling (Figure 2C). Other dynamically regulated paths included terms such as extracellular matrix, cell adhesion, and regulation of cytoskeleton, which are broadly reflective of the extensive morphological changes that occur during reprogramming. The transcriptome switch from a fibroblast to a neuronal program occurred as early as day 10 of miR-9/9*-124 expression, as fibroblast genes were repressed (e.g. *MFAP5* and *S100A4*) and neuronal genes steadily increased in expression (e.g. *SNAP25*) while other genes associated with synaptic functionality appeared at day 20 (e.g. *HOOK1*) (Figure 3C).

Finally, there were no significant changes in *ASCL1* or *SOX2* (Figure S3F), TFs that have been used to reprogram somatic cells into neurons (Niu et al., 2013; Pang et al., 2011), suggesting miR-9/9*-124-induced neuronal conversion activates a neuronal program through mechanisms distinct from those previously reported.

Figure 3

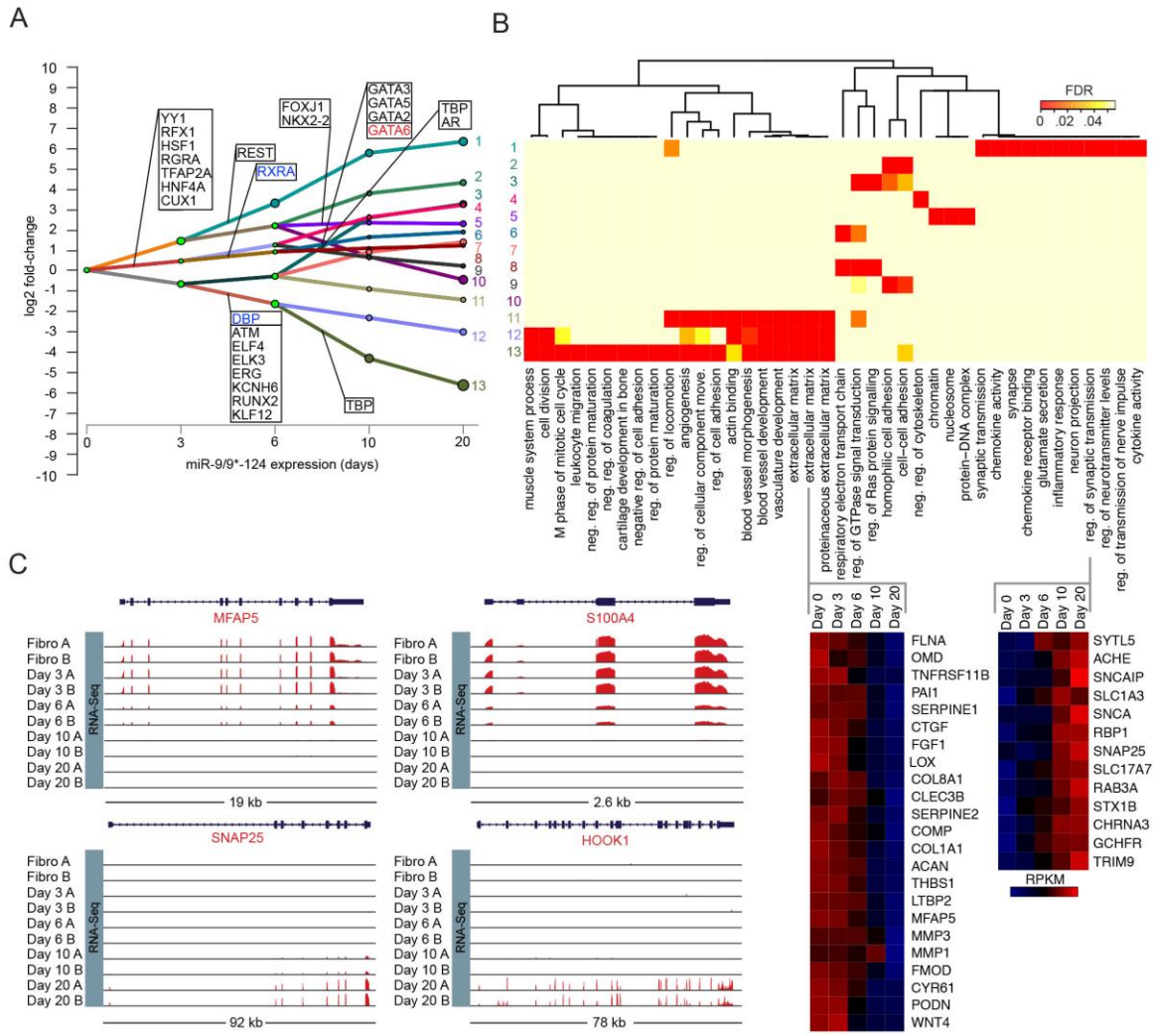


Fig. 3. Time Series Transcriptome Analysis Reveals Early Dynamic Gene Expression Changes Followed by a Stable Transcriptome Switch

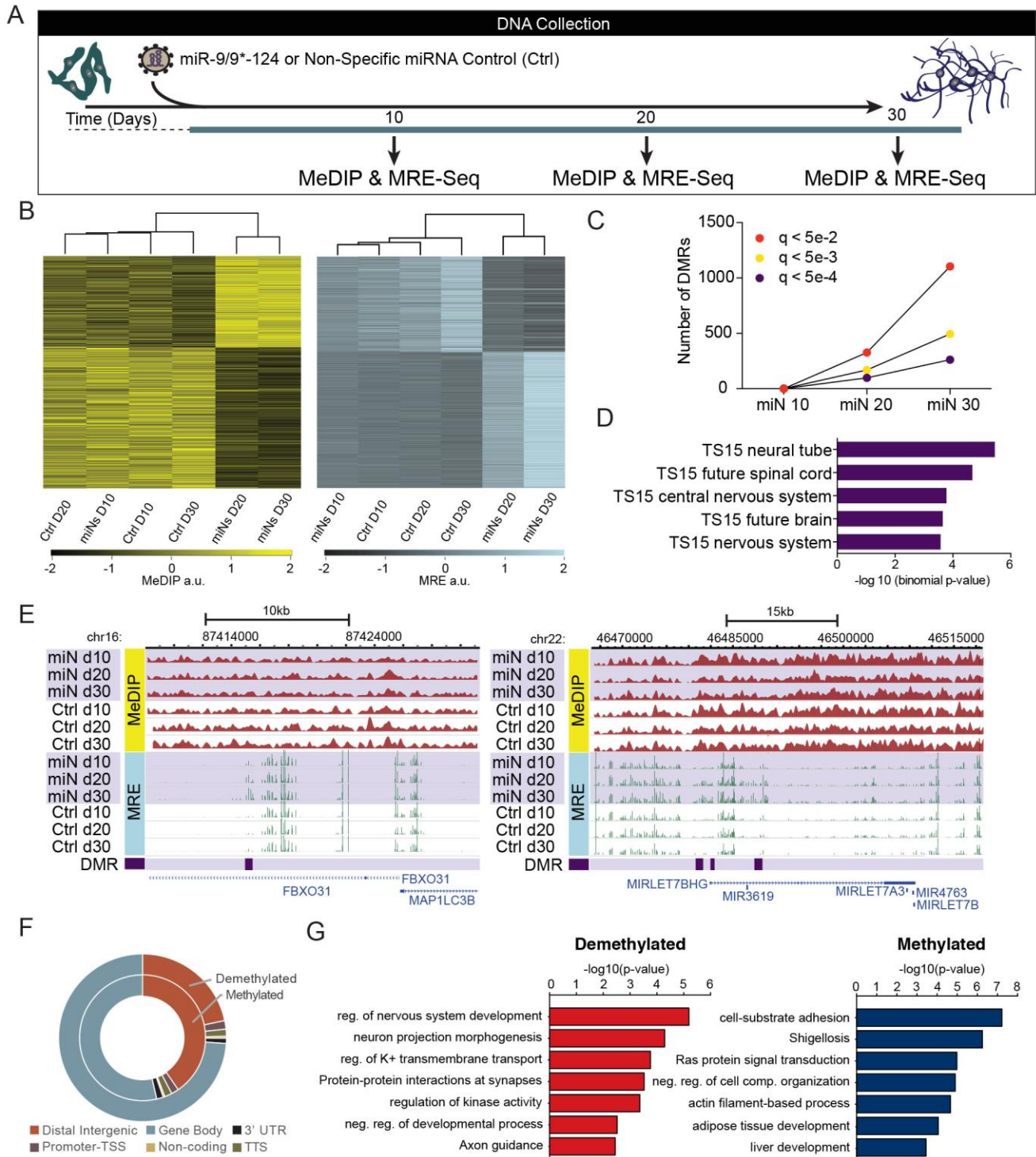
(A) Dynamic Regulatory Event Miner (DREM) analysis reveals multiple paths of co-regulated genes emerge over time (days). Genes shown in each rectangle represent predicted transcription factors that may underlie the transcriptome changes detected. (B) GO terms enriched in each path (top) and heat maps of genes within the extracellular matrix and regulation of synaptic transmission terms (bottom). Z-score normalized RPKM. (C) Representative genome browser snapshots demonstrating the time-dependent loss of fibroblast gene expression (top), emergence of pan-neuronal marker gene expression, and synaptic component expression (bottom).

2.1.6 DNA Methylation Profiling of miNs

After observing numerous changes in DNA methylation machinery we assessed genome-wide DNA methylation at an early (day 10), intermediate (day 20), and late stage (day 30) of neuronal reprogramming by combining methylated DNA immunoprecipitation sequencing (MeDIP-seq) and methylation sensitive restriction enzyme sequencing (MRE-seq; Figure 4A) (Zhang et al., 2013). No significant changes in DNA methylation were detected at day 10 between cells that were exposed to miRNAs and non-treated control (Ctrl) cells. In contrast, we identified 1,540 differentially methylated regions (DMRs) at day 20 (miN day 20 vs. Ctrl day 20) that overlap with DMRs at day 30 (miN day 30 vs. Ctrl day 30; Figures 4B and 4C). The difference in DNA methylation at day 30 between treated and control cells was more dramatic than changes observed at day 20 with most DNA regions undergoing demethylation. Gene ontology (GO) analysis of Mouse Genome Informatics (MGI) expression using the GREAT (McLean et al., 2010) tool to characterize the top overlapping DMRs in miNs enriched only for neuronal tissue developmental processes (Figure 4D). Interestingly, these differentially methylated regions undergo changes at TS15 in mouse development (equivalent to mouse E9/10), which is when *miR-9* expression is first detected in developing mouse telencephalon (Shibata et al., 2008). Two examples of top DMRs within genes important for neuronal development and function are shown in Figure 4E (Gehrke et al., 2010; Vadhvani et al., 2013). Nearly 64% of DMRs were located in introns and 28% of DMRs are in intergenic regions (Figure 4F). Comparison of data from day 30 demethylated and methylated DMRs and RNA-seq revealed 882 differentially expressed genes in total. GO analysis of top demethylated DMRs associated with upregulated genes (> 2.5 logFC) was enriched for neuronal terms. In contrast, top

methylated DMRs associated with down-regulated genes ($< 2.5 \log_{2}FC$) did not match GO terms involved in neural development (Figure 4G; data not shown). Furthermore, since miR-9/9*-124 expression quickly induces cell cycle exit, these changes in DNA methylation must necessarily occur via active processes, such as those catalyzed by TET and TDG enzymes (Kohli and Zhang, 2013), rather than via failure to remethylate DNA after DNA replication. We should note that due to the need to isolate sufficient DNA, profiling of DNA methylation was performed using human neonatal fibroblasts, which grow faster and senesce much later than the human adult fibroblasts used throughout this manuscript, and could differ in some ways from results generated using adult cells.

Figure 4



virus expressing miR-9/9*-124 or a non-specific (NS) control (Ctrl) virus at day 0. Samples were collected at day 10, day 20, and day 30.

(B) Biclustering analysis of DMRs. Heatmaps based on MeDIP-seq RPKM (left) and MRE-seq RPKM (right) showing overlapping DMRs at days 20 and day 30.

(C) Quantification of DMRs at multiple q -value cutoffs ($q < 5e-2$ in red; $q < 5e-3$ in yellow; $q < 5e-4$ in purple) across all time points: miN 10 (miN day 10 vs. Ctrl day 10), miN 20 (miN day 20 vs. Ctrl day 20), and miN 30 (miN day 30 vs. Ctrl day 30).

(D) Tissue development enrichment of top overlapping DMRs at day 20 and day 30 show neuronal tissue terms at TS15 (~E9/10 in mouse development).

(E) WashU Epigenome Browser screenshots of two DMRs: *FBXO31* (left) and *MIRLET7BHG* (right) loci are shown with MeDIP-seq tracks (red; top), MRE-seq tracks (green; middle), and DMR positions (purple; bottom).

(F) Genomic distribution of differentially methylated and demethylated regions.

(G) Functional enrichment of top demethylated and upregulated (red; left) or methylated and downregulated (blue; right) DMRs overlapping at day 20 and day 30 compared with RNA-seq at day 30.

2.1.7 Chromatin Remodeling in miNs

Extensive expression changes in diverse chromatin remodeling genes during neuronal reprogramming (Figures 2D and S2) suggest that miR-9/9*-124 may also alter chromatin accessibility. We therefore performed Assay for Transpose-Accessible Chromatin followed by high throughput sequencing (ATAC-seq) 10 and 20 days after initiating miR-9/9*-124 induced neuronal conversion. A high correlation between replicates confirmed the reproducibility of our analyses (Figure 5A). We obtained 154,406 total peaks across all samples and identified 59,200 differential peaks (Figure 5B). Of the total peaks detected, 20,712 became more accessible (opened) and 38,882 peaks became inaccessible (closed). Most differential peaks in miNs at day 10 overlapped with the peaks at day 20 and the signal intensity of peaks gradually increased or decreased during the conversion (Figure 5C), suggesting gradual transition of chromatin accessibility during reprogramming (Figure 5C). Interestingly, the ratio of open intragenic to distal intergenic regions increased during miR-9/9*-124-mediated conversion (Figure 5D).

Figure 5

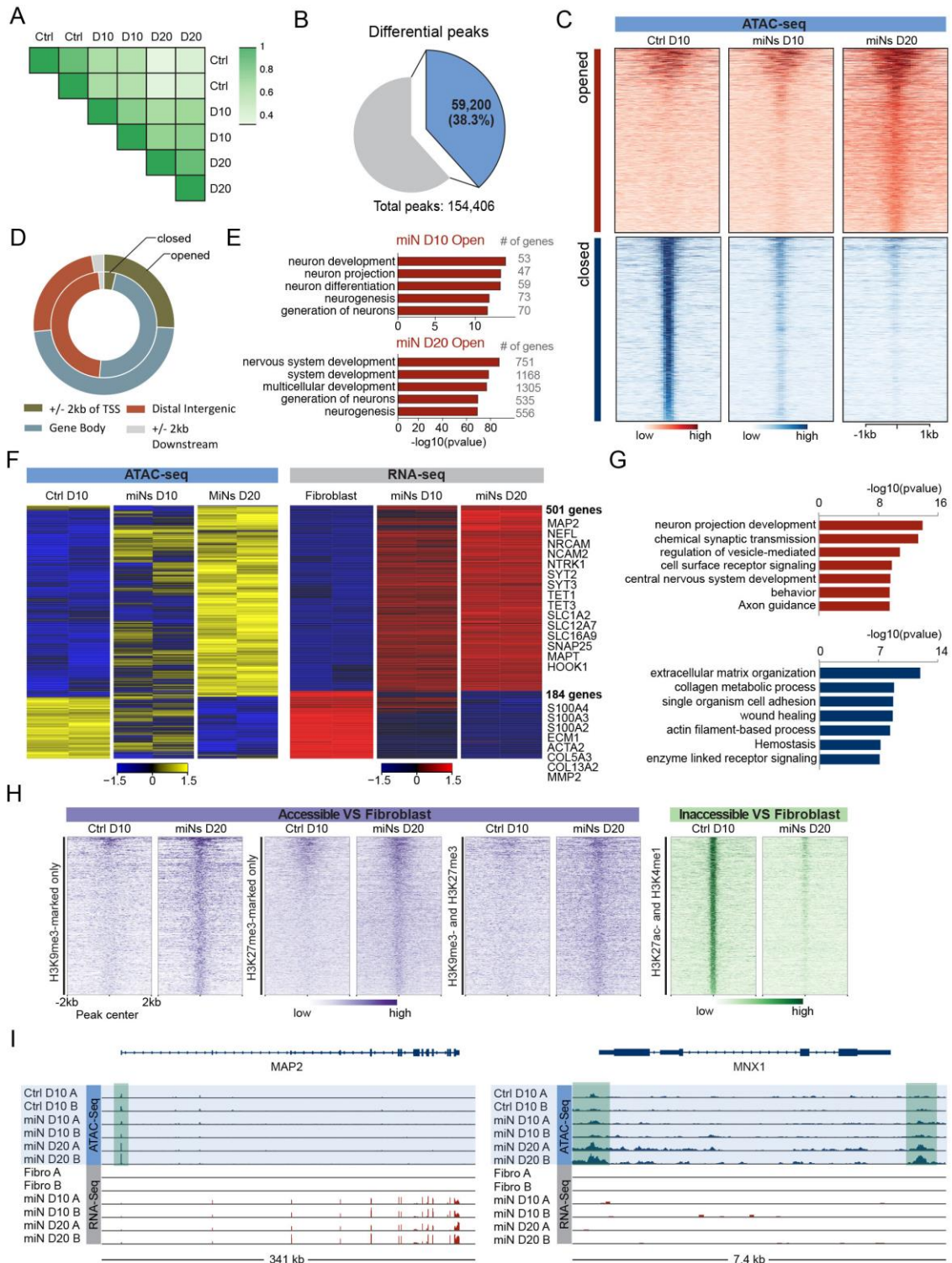


Fig. 5. MiR-9/9*-124 Globally Changes Chromatin Accessibility

(A) Two-dimensional correlation plot of samples. Pearson's correlation coefficient: 0.90 for Ctrl D10; 0.83 for miNs D10 (D10); 0.90 for miNs D20 (D20).

(B) Pie chart showing the proportion of differential peaks and total peaks. Differential peaks were obtained by combining all significant peaks (Ctrl D10 vs miNs D10, miNs D10 vs miNs D20).

(C) Heatmaps showing signal intensity in opened and closed chromatin peaks across all time points. All opened and closed chromatin regions were ranked according to maximum intensity across all samples.

(D) The genomic distribution of opened and closed chromatin regions.

(E) Comparison of GO terms for genes with opened chromatin regions at promoters in miNs at day 10 and 20.

(F) Heatmaps showing gene expression levels for DEGs positively correlated with ATAC-seq signal intensities in their promoter regions. Signal intensity is based on normalized counts per million (CPM) values. Z-score normalized logCPM.

(G) Top GO terms associated with DEGs that correlate with ATAC-seq signal intensity in promoter regions. Top (red): opened and upregulated genes. Bottom (blue): closed and downregulated genes.

(H) Heatmaps showing ATAC signal intensities in the opened and closed chromatin regions in response to miR-9/9*-124, which overlapped with closed/heterochromatin and enhancer regions of fibroblasts.

I) Integrated genomics viewer (IGV) screenshots showing two different examples of ATAC-seq and RNA-seq integration, with genes being transcribed from left to right. An example of ATAC-seq and RNA-seq peaks within a pan-neuronal gene: *MAP2* (left) and an example of ATAC-seq peaks (shaded green) within a subtype-specific locus without gene expression changes: *MNX1* (right). See also Figures S3 and S4.

2.1.8 Erasure of Fibroblast Epigenetic Identity and Gain of Neuronal Chromatin Architecture

To gain insight into biological relevance of changes in chromatin accessibility after miR-9/9*-124 expression, we performed gene enrichment analysis on genes with differential ATAC signals around the transcription start site \pm 2Kb (TSS). We identified 4,915 genes with gradual increases and 1,763 genes with gradual decreases in ATAC signals during conversion. Top GO terms associated with genes with increased ATAC signals are enriched in neuronal terms (Figure 5E), while closed regions were not (Figure S3A). The number of genes in open regions associated with neuronal terms gradually increased from day 10 to 20 consistent with our transcriptional profiling and coinciding with *bona fide* neuronal commitment of miNs (Figure 5E). Consistent with this hypothesis, chromatin accessibility for fibroblast marker genes like *S100A4*, *S100A10*, *VIM* and *COL13A1* gradually decreased between 10 and 20 days into neuronal conversion. Collectively, our ATAC-seq analyses demonstrate miR-9/9*-124-induced chromatin remodeling events are characterized by concurrent closing of fibroblast-related genomic loci and opening of neuronal gene loci.

Next, we examined whether miR-9/9*-124-induced chromatin accessibility correlated with changes in mRNA levels. We compared DEGs ($\log_{2}FC > 2$ or < -2 adjusted p-value < 0.01) to genes with altered chromatin accessibility around the TSS and identified 501 upregulated and 184 down-regulated genes that coincide with open and close regions, respectively (Figure 5F). GO enrichment analysis revealed that upregulated genes from opened regions associated with neurogenic terms, while downregulated genes from closed regions were connected with terms important for fibroblast function (Figure 5G, see S3E for example tracks). This demonstrates the concordant regulation of transcription and

chromatin accessibility mediated by miR-9/9*-124. Interestingly, some opened genomic regions that displayed no changes in gene expression contained genes that are specifically expressed in various neuronal subtypes, including those enriched in: dopaminergic neuron markers (*TH* and *SLC6A3*), serotonergic neuron markers (*FEV*, *LMX1B* and *SLC6A4*), GABAergic neuron markers (*SLC6A1*, *SLC32A1*, *GAD2*), striatal medium spiny neuron marker (*PPP1R1B*), glutamatergic neuron markers (*GLUL* and *SLC1A6*), and cholinergic or motor neuron markers (*MNX1*, *CHAT*, and *SLC5A7*) (Figure S3G for example tracks). These results suggest that miR-9/9*-124 poise the chromatin to accept additional inputs from subtype-specifying determinants without activating subtype-specific programs. In all, the chromatin dynamics observed during miRNA mediated neuronal conversion are consistent with time-dependent suppression of fibroblast identity concurrent with the opening of neuronal loci and activation of pan-neuronal gene expression.

2.1.9 MicroRNA-induced Chromatin Remodeling at Heterochromatin Regions in Fibroblasts

To gain a more complete understanding of the epigenetic architecture within opened and closed chromatin sites, we examined the relationship of these regions to pre-existing histone marks present in fibroblasts. We hypothesized that regions that close during reprogramming would overlap with the active enhancer/euchromatin marks H3K27ac and H3K4me1 in fibroblasts. Conversely, regions that open during neuronal reprogramming may overlap with heterochromatic H3K9me3 and H3K27me3 signatures pre-existing in fibroblasts.

We selected 70,661 regions commonly marked by H3K27ac and H3K4me1 (enhancer/euchromatin) and 5,843 regions commonly marked by H3K9me3 and H3K27me3 (heterochromatin/closed) in human fibroblasts based on the Roadmap Epigenome database

(Roadmap Epigenomics et al., 2015). Strikingly, we found that 1,128 ATAC signal peaks present in day 20 miNs that displayed increased accessibility during the neuronal conversion overlapped with regions of fibroblasts marked by H3K9me3/H3K27me3, heterochromatic marks in fibroblasts, whereas H3K27ac/H3K4me1 euchromatic regions in fibroblasts overlapped with 16,207 peaks that were closed in miNs (Figure 5H). GO enrichment analysis of genes associated with open chromatin regions in miNs that were marked by H3K9me3 or H3K27me3 in fibroblasts resulted in neuronal differentiation and function-related terms (Figure S3B). In contrast, regions with decreased chromatin accessibility in miNs were enriched with non-neuronal terms (Figure S3D). These results demonstrate the surprising potency of miR-9/9*-124 to open heterochromatin regions needed for neuronal fate and to close enhancer regions associated with fibroblast fate. Our results collectively demonstrate that microRNAs can change chromatin architecture to promote neuronal and repress fibroblast fates during the direct conversion of human fibroblasts to neurons.

2.1.10 Chromatin Remodeling is Required for Direct Conversion

To further determine if chromatin changes were necessary for generating miNs, we knocked down the expression of BRG1, a core component of BAF chromatin remodeling complex whose reduced function has been shown to collapse the overall chromatin architecture (Kadoch et al., 2017) (Figure S4A). After 20 days of neuronal conversion, loss of BRG1 markedly decreased the number of MAP2-positive cells when compared to a control shRNA (Figure S4B and C). ATAC-seq revealed regions which failed to open in response to miR-9/9*-124 in the absence of BRG1 (Figure S4D and E). These regions were associated with neuronal GO terms in contrast to the fibroblast-related GO terms associated

with regions that failed to close (Figure S4F). These data demonstrate the requirement of chromatin remodeling during miR-9/9*-124-mediated neuronal conversion.

2.1.11 Instructing the miRNA-induced Neurogenic State to a Motor Neuron Fate

The deposition and removal of nucleosomes along regulatory elements within DNA inhibits or enables the binding of TFs, simultaneously facilitating and reinforcing cell-type specific gene expression programs (Jiang and Pugh, 2009). In our previous analyses we noted chromatin regions in miNs with enhanced accessibility were proximal to *MNX1* and choline acetyl transferase *CHAT*, two of the hallmark genes expressed by motor neurons (Fonnum, 1973; Tanabe et al., 1998) (Figure 5I). Neither *MNX1* nor *CHAT* mRNA levels were elevated in day 30 miNs after miR-9/9*-124 directed neuronal conversion, but we hypothesized the open chromatin would facilitate expression of these genes in response to motor neuron-specific TFs. To test this hypothesis, we expressed a panel of TFs known to promote motor neuron identity (Figure S5A) in combination with ectopic miR-9/9*-124 expression, and assayed for MAP2 and MNX1 expression by immunostaining (data not shown). While all combinations resulted in MAP2 positive cells, only ISL1 and LHX3 robustly led to the generation of MNX1-positive cells. Co-expressing LHX3 and ISL1 with miR-9/9*-124 in adult human fibroblasts from 22-, 42-, 56- and 68-year-old donors resulted in MAP2, TUBB3 and NCAM-positive cells with complex neuronal morphologies (Figures 6A and 6B). Approximately 80% of cells positive for DAPI staining were positive for TUBB3, MAP2 and NCAM expression (Figure 6C). The majority of converted cells displayed nuclear staining of MNX1 (~85% of TUBB3 positive cells in each line, Figures 6D and 6E). Similarly, cytoplasmic CHAT protein and SMI-32 (a neurofilament protein found in motor neurons) were detected in ~80% of TUBB3 positive cells within each age

group (Figure 6D and 6E).

Interestingly, ISL1 and LHX3 alone were not sufficient to induce neuronal conversion when co-expressed with a non-specific miRNA (miR-NS) (Figure S5B), supporting the notion that miR-9/9*-124 is necessary for the subtype-specifying activities of ISL1 and LHX3. We also found the motor neuron conversion induced by miR-9/9*-124 with ISL1 and LHX3 to be stable, displaying neuronal morphologies, cell cycle exit, and motor neuron marker expression for 30 days after Dox removal (Figure 6F).

Figure 6

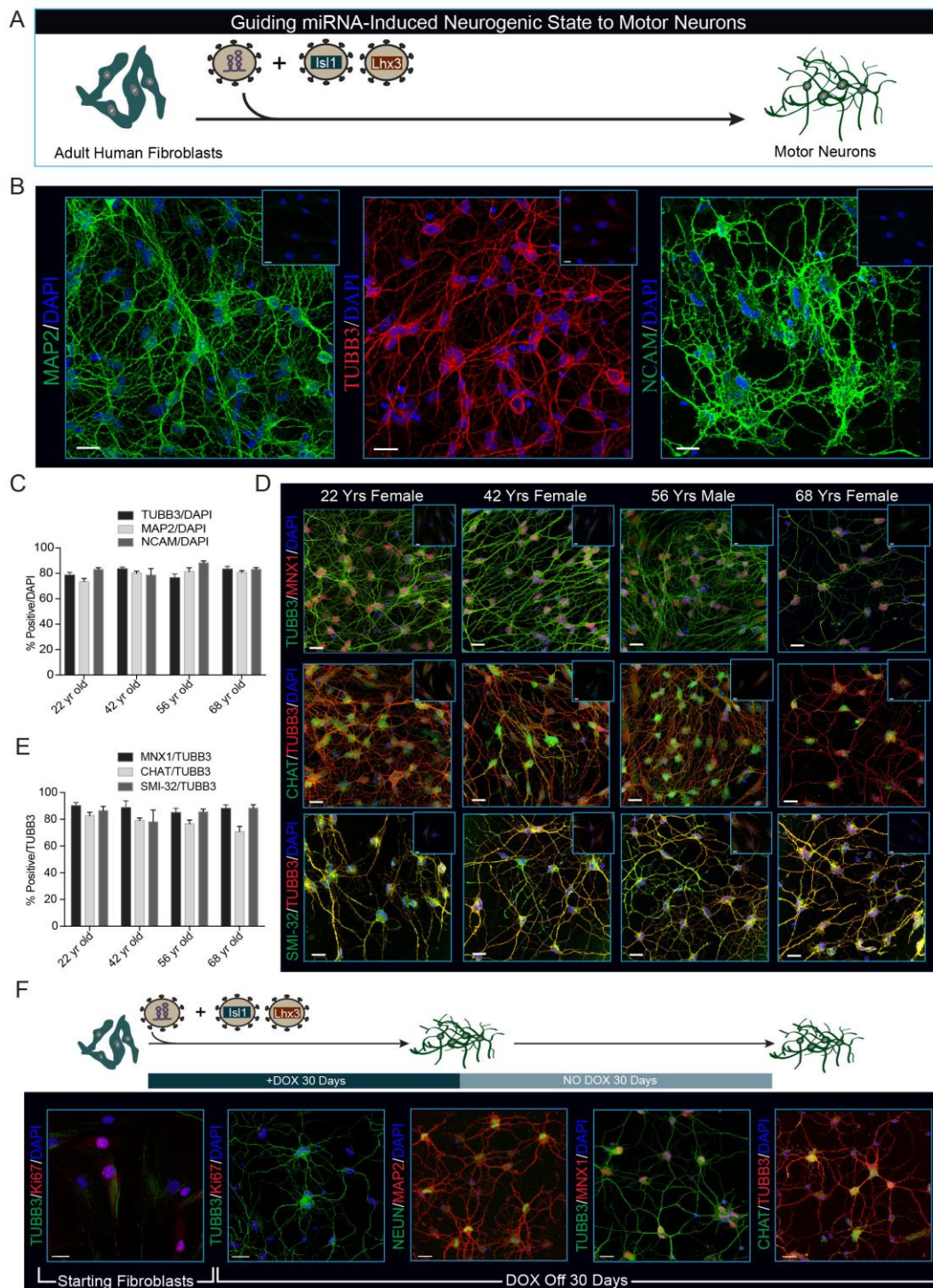


Fig. 6. MiRNA-Induced Neuronal Competence Enables Motor Neuron Transcription Factors, ISL1 and LHX3, to Determine Motor Neuron Identity.

(A) Schematic of neuronal induction paradigm using miR-9/9*-124 plus *ISL1* and *LHX3*.
(B) Representative immunohistochemistry for pan-neuronal markers in neurons generated from fibroblasts through 35 days of miR-9/9*-124, *ISL1*, and *LHX3* co-expression. Fibroblasts were isolated from a 22-year-old female donor. Scale bars = 20 μ m.
(C) Quantification of 4 independent primary human fibroblast samples from male and female donors stained with TUBB3, MAP2 and NCAM. Percentages represent total number of positive cells over all cells (DAPI) and are represented as mean \pm SEM. Cells (N) analyzed: 22 yr old N=TUBB3 325, MAP2 219, NCAM 275; 42 yr old N=TUBB3 304, MAP2 236, NCAM 129; 56 yr old N=TUBB3 275, MAP2 279, NCAM 213; 68 yr old N=TUBB3 282, MAP2 234, NCAM 190.
(D) Expression and correct localization of motor neuron markers in neurons converted by miR-9/9*-124 and *ISL1/LHX3* as demonstrated by immunohistochemistry. MNX1, (top), CHAT (middle) and SMI-32 (bottom). Scale bars = 20 μ m.
(E) Quantification of (D) represents the total percentage of MNX1, CHAT and SMI-32-positive cells over TUBB3-positive cells. Data are represented as mean \pm SEM. Cells analyzed: 22 yr old N=MNX1 256, CHAT 256, SMI-32 113; 42 yr old N= MNX1 151, CHAT 151, SMI-32 283; 56 yr old N= MNX1 207, CHAT 207, SMI-32 174; 68 yr old N= MNX1 151, CHAT 151, SMI-32 96.
(F) After 30 days of neuronal conversion by ectopic miR-9/9*-124 expression, Dox was removed and cells were cultured for an additional 30 days. Immunocytochemistry showing motor neurons produced by miR-9/9*-124 plus *ISL1* and *LHX3* (Moto-miNs) remain Ki-67 negative (2nd panel), retain expression and localization of the neuronal proteins TUBB3, NEUN, and MAP2 (2nd and 3rd panel), and express the motor neuron proteins MNX1 and CHAT (4th and 5th panel). Scale bars = 20 μ m.
See also Figure S5.

2.1.12 Electrophysiological properties of Moto-miNs

Motor neurons produced from fibroblasts by co-expression of miR-9/9*-124, ISL1 and LHX3 (Moto-miNs) demonstrated robust inward and outward currents in response to depolarizing steps (Figure 7A) and displayed AP trains through the injection of step-wise depolarizations (Figures 7B and S5C). The visualization of single traces recorded at individual current steps revealed the characteristic hyperpolarization following each action AP seen in mature neurons (Figure 7C). To assess the percentage of functionally mature Moto-miNs, we patched 45 randomly chosen Moto-miNs from 22-year-old and 68-year-old donors. All Moto-miNs fired APs in which most cells fired multiple APs (80% n = 20 and 74% n = 25) while single APs were observed in ~20-25% of the patched cells (Figure 7D). Gap-free recordings in current clamp mode revealed cells capable of firing spontaneous APs (Figure 7E), demonstrating the excitability of the converted Moto-miNs. Similar I-V curve relationships were observed between donors and firing patterns (Figure 7F). Peak inward currents were substantially higher than those observed in miNs (Figure S6A). Lastly, resting membrane potentials in all Moto-miNs tested were hyperpolarized (Figure 7G; 22 yr old, $-67.2 \text{ mV} \pm 3.3 \text{ mV}$; 68 yr old, $-72.8 \text{ mV} \pm 2.0 \text{ mV}$, S.E.M.). Coupled with the increased proportion of cells that fire multiple APs, these data suggest that the addition of ISL1 and LHX3 to miR-9/9*-124 produced more mature neurons than exposure to miR-9/9*-124 alone.

The ability of motor neurons to control voluntary muscle movement stems from their ability to form neuromuscular junctions (NMJs), unique synapses formed between motor neurons and muscle cells. The formation of NMJs was visualized through the co-localization of EGFP-labeled Moto-miNs, Alexa-fluor-594 Bungarotoxin (BTX, a toxin that

binds to the nicotinic acetylcholine receptor (AChR) of NMJs), and myosin heavy chain. BTX puncta were not observed in the absence of Moto-miNs (Figure 7H, left inset). In contrast, Moto-miNs were able to induce characteristic BTX-clustering in close apposition with EGFP labeled neurons and myotubes (Figure 7H) indicating the formation of putative NMJs.

Figure 7

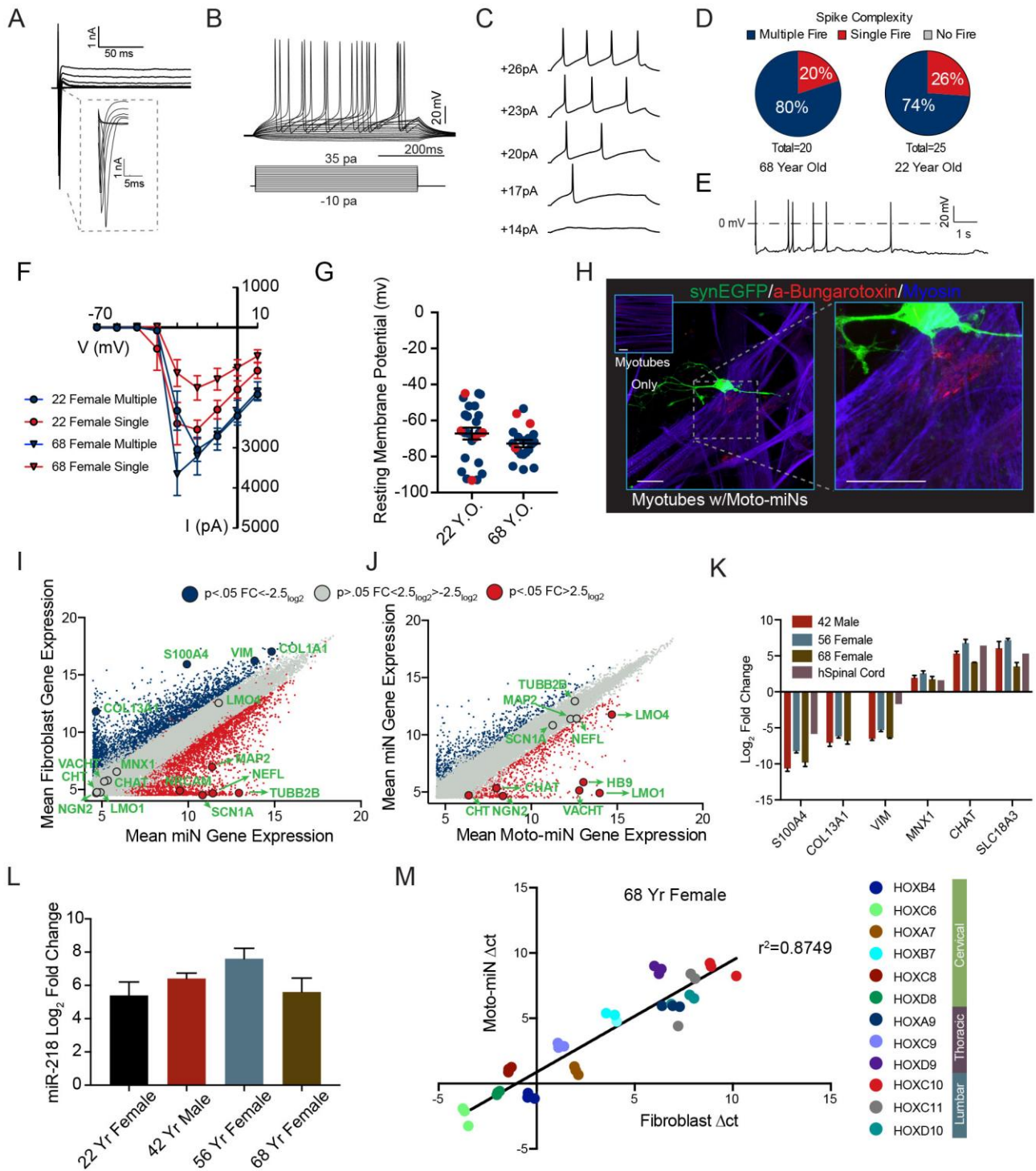


Fig. 7. Functional Properties and Gene Expression Profile of Moto-miNs

(A) Representative traces of inward sodium and outward potassium whole-cell currents.

- (B) Repetitive AP waveforms in response to 500 ms current injections recorded from Moto-miNs in monoculture.
- (C) Representative waveforms of APs with increasing current injections.
- (D) Summary of firing patterns observed in Moto-miNs converted from both old and young donors. 68 yr old 80% multiple (N=20), 22 yr old 74% multiple (N=25).
- (E) Spontaneous AP generation observed in a small percentage of Moto-miNs (3/20).
- (F) Combined plot of the current (I) - voltage (V) relationship for all Moto-miNs recorded. Data are represented as mean \pm SEM.
- (G) Moto-miNs converted from both young and old donors are hyperpolarized, demonstrating mean resting membrane potentials of -67.2mV and -72.8mV, respectively. Data are represented as mean \pm SEM.
- (H) Staining of Moto-miNs cultured with differentiated human myotubes. Moto-miNs were labeled with synapsin-EGFP via viral transduction and then plated onto human myotubes. Myotube only cultures did not have α -Bungarotoxin-594 (red) puncta (top left inset). Scale bar = 20 μ m.
- (I) Scatterplot comparing the mean gene expression between starting fibroblasts from a 22-year-old donor (y-axis) and miNs generated from the same individual (x-axis). Plot highlights a selection of pan-neuronal and fibroblast-specific genes in green text. Blue = $\log_2FC < -2.5$ and $p < 0.05$, (more abundant in fibroblasts) grey = $\log_2FC > -2.5$ and < 2.5 $p > 0.05$ (no significant difference), and red = $\log_2FC > 2.5$ and $p < 0.05$ (more abundant in miNs).
- (J) Scatterplot comparing the mean gene expression between miNs from a 22-year-old donor (y-axis) and Moto-miNs generated from the same individual (x-axis). Plot highlights a selection of pan-neuronal and motor neuron-specific genes in green text. Blue = $\log_2FC < -2.5$ and $p < 0.05$, (more abundant in miNs) grey = $\log_2FC > -2.5$ and < 2.5 $p > 0.05$ (no significant difference), and red = $\log_2FC > 2.5$ and $p < 0.05$ (more abundant in Moto-miNs).
- (K) qRT-PCR validation of fibroblast and motor neuron-specific genes. Moto-miNs were analyzed by qRT-PCR 35 days post-transduction. Human spinal cord RNA served as a positive control (normalized to 42 yr fibroblasts, $\Delta\Delta Ct$ method). Data are represented as mean \pm SEM.
- (L) Moto-miNs express the motor neuron specific miRNA, miR-218. RNA was isolated from fibroblasts and Moto-miNs 35 days post-transduction and analyzed by qRT-PCR. Data are represented as mean \pm SEM.
- (M) Moto-miNs retain HOX gene expression pattern of donor fibroblasts as demonstrated by qRT-PCR. ΔCt method. Data represent ΔCt values for each biological replicate (3 separate Moto-miN conversions). See also Figure S6 and S7.

2.1.13 Transcriptional Profiling of Moto-miNs

To fully characterize the acquisition of a motor neuron fate and assess the contribution of ISL1 and LHX3, we profiled the transcriptome of 22-yr-old starting fibroblasts, miNs and Moto-miNs by microarray. We again observed the loss of fibroblast gene expression (for example, *S100A4*, *VIM* and *COL13A1*), and the gain of a pan-neuronal identity (for example, *MAP2*, *NEFL*, *SNAP25* and *SCN1A*) 35 days after the expression of miR-9/9*-124 (Figure 7I). Expression of motor neuron-specific genes was not significantly different between starting fibroblasts and miNs. While the addition of ISL1 and LHX3 to miR-9/9*-124 did not significantly change the expression of pan-neuronal genes when compared to miNs (Figure 7J), ISL1 and LHX3 selectively activated key motor neuron genes including *MNX1*, *CHAT*, *VACHT*, *LMO1* and *LMO4* (Figure 7J). We further validated the loss of fibroblast identity and gain of motor-neuron identity in Moto-miNs derived from 42-, 56- and 68-year-old donors by qRT-PCR using RNA from human spinal cord as a positive control (Figure 7K). In addition, we observed dramatic upregulation of miR-218, a recently identified motor neuron-specific miRNA (Amin et al., 2015; Thiebes et al., 2015) in Moto-miNs (Figure 7L).

Next, we used the cell type-specific enrichment analysis tool (CSEA) (Xu et al., 2014) to test whether the gene expression profile within each population of miNs and Moto-miNs would be associated with distinct subtypes of *in vivo* neurons. When queried with a gene list, CSEA identifies neuronal subtypes that show significant enrichment in genes from the input list through curated transcriptomic data. Our CSEA analysis of the 100 most enriched genes within Moto-miNs identified two subtypes, brainstem motor neurons and spinal motor neurons, to be significantly associated. No subtype specificity was assigned to

the miN transcriptome (Figure S6B). This unbiased bioinformatics approach further supports the motor neuron identity of converted Moto-miNs.

Lastly, HOX gene expression patterns were compared by qRT-PCR between starting fibroblasts and Moto-miNs. Interestingly, we observed a high correlation ($R^2 = 0.88$) between the expression levels of HOX genes before and after conversion within each of the defined spinal cord regions tested, indicating Moto-miNs retain the positional identity that existed in original fibroblasts (Figure 7M, S6C).

2.1.14 Transcriptional Activation of ISL1 and LHX3 Genomic Targets

An alternative approach for generating motor neurons is forced expression of NGN2, ISL1, and LHX3 in human embryonic stem cells (ESCs) (Mazzoni et al., 2013). We compared the genomic targets of ISL1 and LHX3 identified by Mazzoni et al. through ChIP-seq to genes whose expression increases in Moto-miNs compared to miNs. Surprisingly, we identified a large cohort of overlapping genes (323) that include numerous hallmark motor neuron markers (Figure S7A), suggesting that ISL1 and LHX3 activate expression of a core gene regulatory network that specifies the miR-9/9*-124-mediated neuronal conversion towards motor neurons.

2.1.15 Comparison of Moto-miNs to Endogenous Spinal Cord Motor Neurons

The inability to obtain a pure population of spinal cord motor neurons from human adults prevents direct transcriptional comparisons between Moto-miNs and their human *in vivo* counterparts. Therefore, we directly compared Moto-miNs to fully differentiated *in vivo* mouse spinal cord motor neurons. To interrogate the gene expression of motor neurons within the large heterogeneity of cell-types present in the spinal cord, we performed

Translating Ribosomal Affinity Purification (TRAP) followed by RNA-seq (Figure S7B). The use of two mouse lines expressing EGFP tagged ribosomes (one line under the pan-neuronal *SNAP25* promoter, and the other through the *CHAT* promoter), enabled the enrichment and subsequent sequencing of actively transcribed mRNA in all neurons and motor neurons within the spinal cord. We also profiled the transcriptome of the entire spinal cord as an additional ‘pre-IP’ control. Comparisons between CHAT pre-IP controls and CHAT-TRAP transcripts confirmed significant enrichment of motor neuron markers (Figure S7C, left) by the TRAP procedure. To further separate motor neuron enriched genes from common neuronal genes we first normalized CHAT-TRAP and SNAP-25 datasets to their pre-IP controls, then directly compared the normalized gene expression values. This analysis revealed the co-expression of pan-neuronal genes such as *MAP2*, *NRCAM*, *SCN1A* and *TUBB3* and highly enriched for motor neurons transcripts such as *SLC18A3* (*VACHT*), *CHAT* and *MNX1* (Figure S7C, right). This dataset also serves as a resource for genes enriched in spinal cord neurons over all *SNAP25* expressing neurons within the spinal cord. To that end, we compared differentially expressed mouse spinal cord motor neuron genes to genes enriched in Moto-miNs over miNs. We observed a significantly larger overlap between Moto-miNs and mouse motor neurons than expected by chance (Fisher’s exact test $p = 3.522e-06$), indicating that Moto-miNs and endogenous mouse motor neurons utilize similar genetic networks. This includes expression of canonical motor neuron markers such as, *SLIT2* and *SLIT3*, host genes for the motor neuron specific miRNA, miR-218 (Amin et al., 2015) (Figure S7D). Altogether, our in-depth analyses of Moto-miNs at cellular, functional, and transcriptomic levels confirm that co-expression of miR-9/9*-124 and *ISL1/LHX3* can directly convert adult fibroblasts into spinal cord motor neurons.

2.2 Author Contributions

ASY, DGA, MJM and WK conceived the concept for this study. ASY and DGA designed and performed experiments determining the neurogenic potential of miR-9/9*-124. DGA and WK designed and performed gene expression studies. MJM performed timecourse RNA-seq analyses. ASY, TW, HL and MJM designed DNA-methylation studies, and MJM generated and analyzed DNA-methylation data with assistance from XX and DL. ASY, DGA and WK designed and performed ATAC-Seq experiments and analyses. ASY and DGA designed and performed motor neuron experiments. JDD, ROH, AL and RO designed and analyzed TRAP-Seq experiments. ASY and DGA wrote the manuscript with input from all authors.

2.3 Acknowledgements

DGA is supported by the Philip and Sima Needleman Graduate Student Fellowship. MJM and RO are supported by the NIH-funded Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Predoctoral Fellowship (T32GM081739; Barch, PI). J.D. and R.O.H. are supported by the Children's Discovery Institute and J.D. is supported by NIH 1U01MH1091330. ASY is supported by NIH Director's Innovator Award (DP2NS083372-01), Missouri Spinal Cord Injury/Disease Research Program (SCIDRP), Cure Alzheimer's Fund (CAF) and Presidential Early Career Award for Scientists and Engineers (PECASE).

2.4 Supplemental Figures

Figure S1

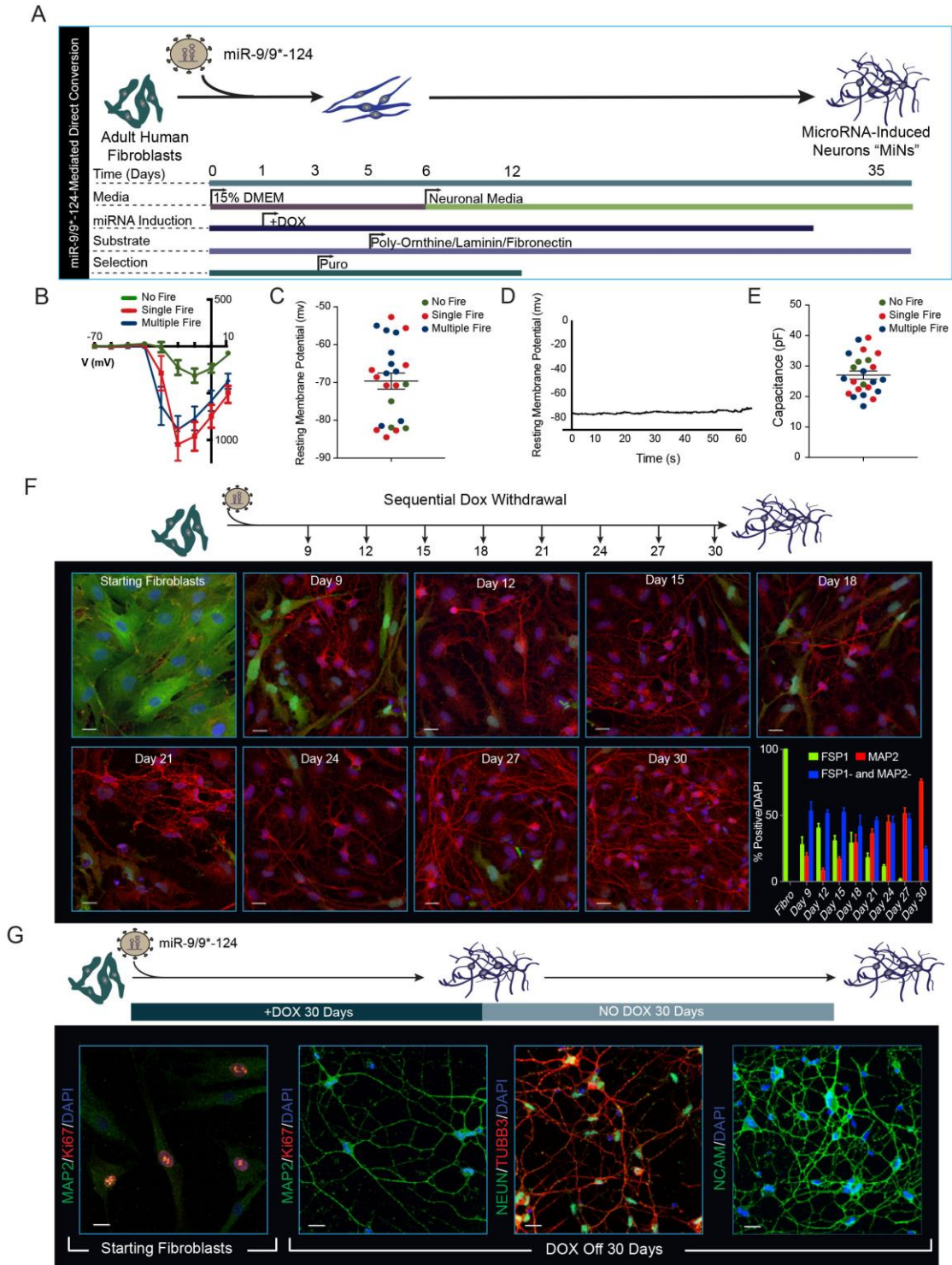


Fig. S1. MiRNA-Mediated Conversion into Neuronal Fate is Stable.

- (A) Detailed schematic of miR-9/9*-124 direct conversion protocol.
- (B) Combined plot of the current (I) - voltage (V) relationship for all neurons recorded.
- (C) Tabulated values of resting membrane potentials in miNs.
- (D) Gap-free recording of resting membrane potential in a miN.
- (E) Tabulated values of capacitance values in miNs.
- (F) MiR-9/9*-124 was ectopically expressed under a doxycycline (DOX) inducible promoter in 22 year old human fibroblasts for 9 days then DOX was removed from the media at 3 day intervals. Cells were cultured until reprogramming day 30. Neuronal fate was assayed by immunostaining for the pan-neuronal marker MAP2 and the fibroblast identity was determined by FSP1 expression. Scale bar = 20 μ m. All data are represented as mean \pm SEM.
- (G) MiR-9/9*-124 was ectopically expressed under a DOX-inducible promoter in 22 year old human fibroblasts for 30 days then DOX was removed from the media and cells were cultured for an additional 30 days. Stable adoption of the neuronal fate was assayed by immunostaining for the pan-neuronal markers MAP2, NEUN, TUBB3 and NCAM. To assay if cells remained post-mitotic cells were stained with the proliferative marker ki-67. Scale bar = 20 μ m. All data are represented as mean \pm SEM.

Figure S2

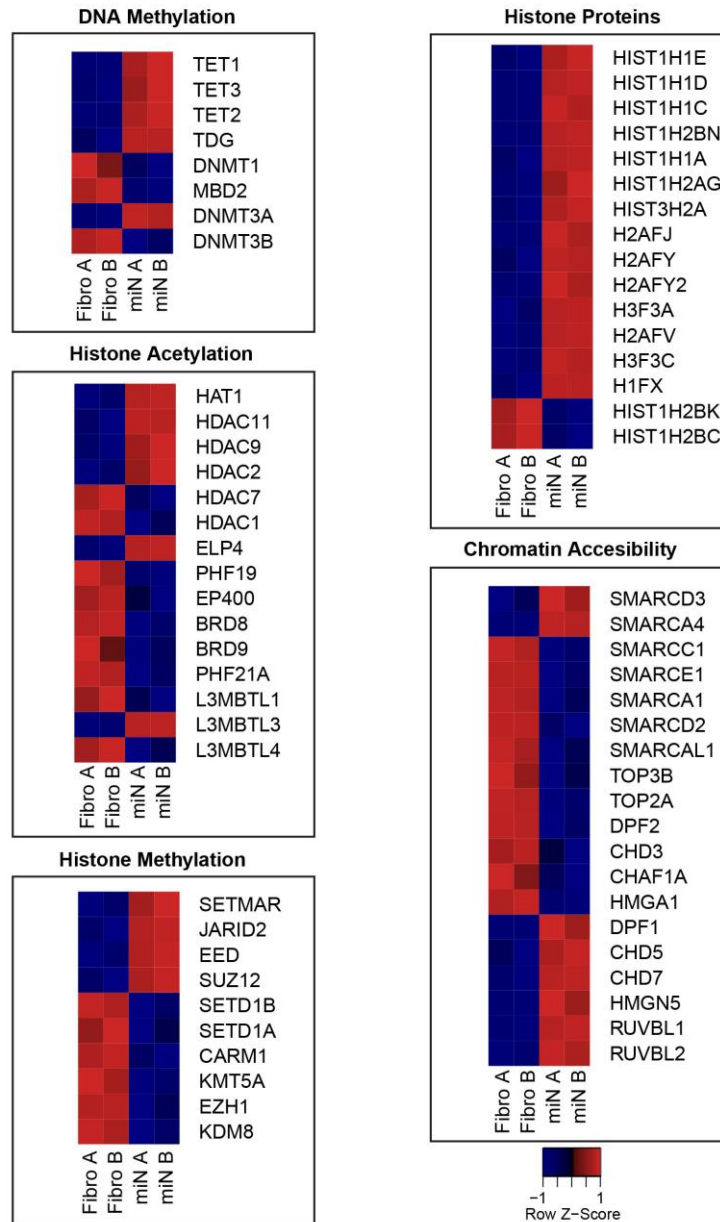


Fig. S2. Expression Changes in Epigenetic Modifiers During Reprogramming.

Heatmaps show expression changes observed between miNs and fibroblasts within a subset of proteins that influence chromatin state through various modes.

Figure S3

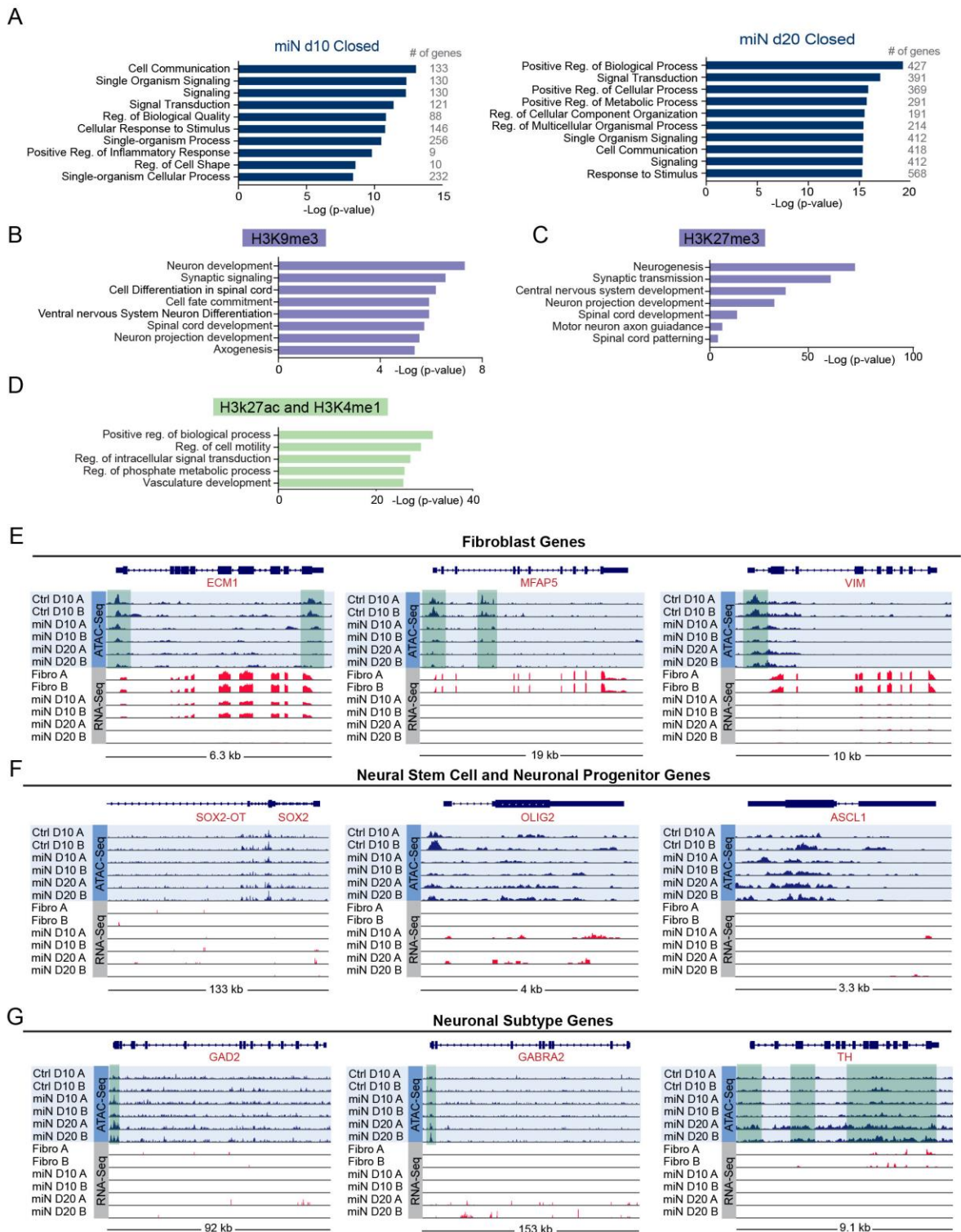


Fig. S3. Pre-existing Heterochromatic Neuronal Loci Open in Response to miR-9/9*-124 Expression.

- (A) Top GO terms associated with promoter regions that close during reprogramming from fibroblasts to miNs.
- (B) Closed regions in fibroblasts marked by H3K9me3 that open during neuronal reprogramming are enriched for neuronal GO terms.
- (C) Closed regions in fibroblasts marked by H3K27me3 that open during reprogramming are also enriched for neuronal GO terms.
- (D) Pre-existing distal H3K27ac and H3K4me1 marks within fibroblasts that close during neuronal reprogramming show GO terms related to general biological processes.
- (E) Genome browser snapshots demonstrating closing and loss of fibroblast gene expression (ECM1, MFAP5, and VIM).
- (F) Genome browser snapshots demonstrating neither opening or activation of progenitor genes (SOX2, OLIG2 and ASCL1).
- (G) Genome browser snapshots demonstrating neuronal subtype gene loci that open, but do not show gene expression changes (GAD2 and GABRA2, GABAergic markers; TH, dopaminergic neuron marker). Shaded green: location of differential ATAC peaks detected.

Figure S4

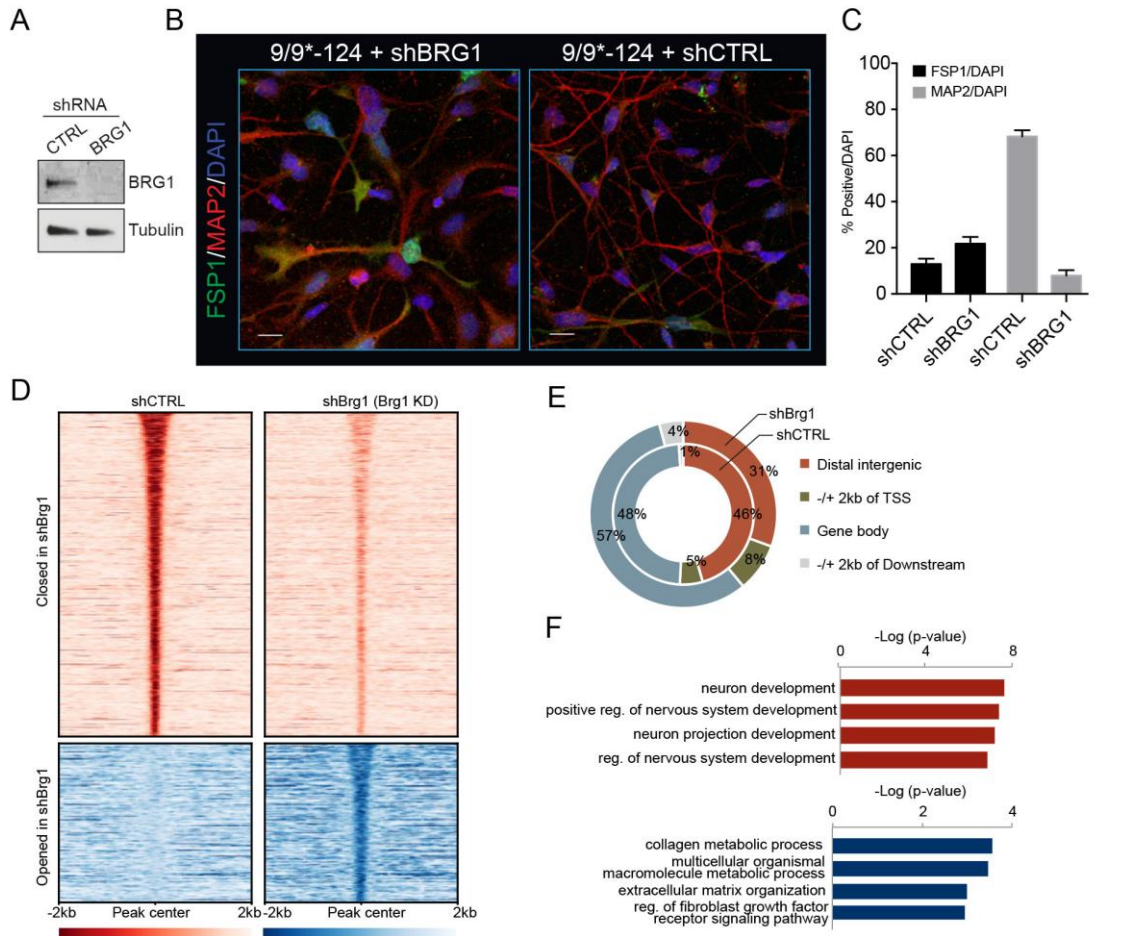


Fig. S4. Loss of BRG1 Prevents Neuronal Fate Acquisition

(A) BRG1 expression in human fibroblasts expressing shCTRL or shBRG1

(B) Adult human fibroblasts ectopically expressing miR-9/9*-124 and shBRG1 or shCTRL for 20 days, immunostained for the pan-neuronal marker MAP2 and fibroblast marker FSP1. Scale bars = 20 μ m.

(C) Quantification of (B) MAP2 and FSP1 positive cells over total number of cells (DAPI). Scoring for MAP2-positive cells, only cells with processes at least three times the length of the soma were counted. Data are represented as mean \pm SEM. shBRG1 n = 155 cells, shCTRL, n = 173 cells.

(D) Heatmaps showing ATAC signal intensity in open and close chromatin peaks in shBRG1 and shCTRL. All open and closed chromatin regions were ranked according to maximum intensity across all samples.

(E) The genomic distribution of ATAC signals in cells expressing miR-9/9*-124 and shCTRL or shBRG1.

(F) GO terms showing biological function of closed (lost in shBRG1, red) and open (retained in shBRG1, blue) regions.

Figure S5

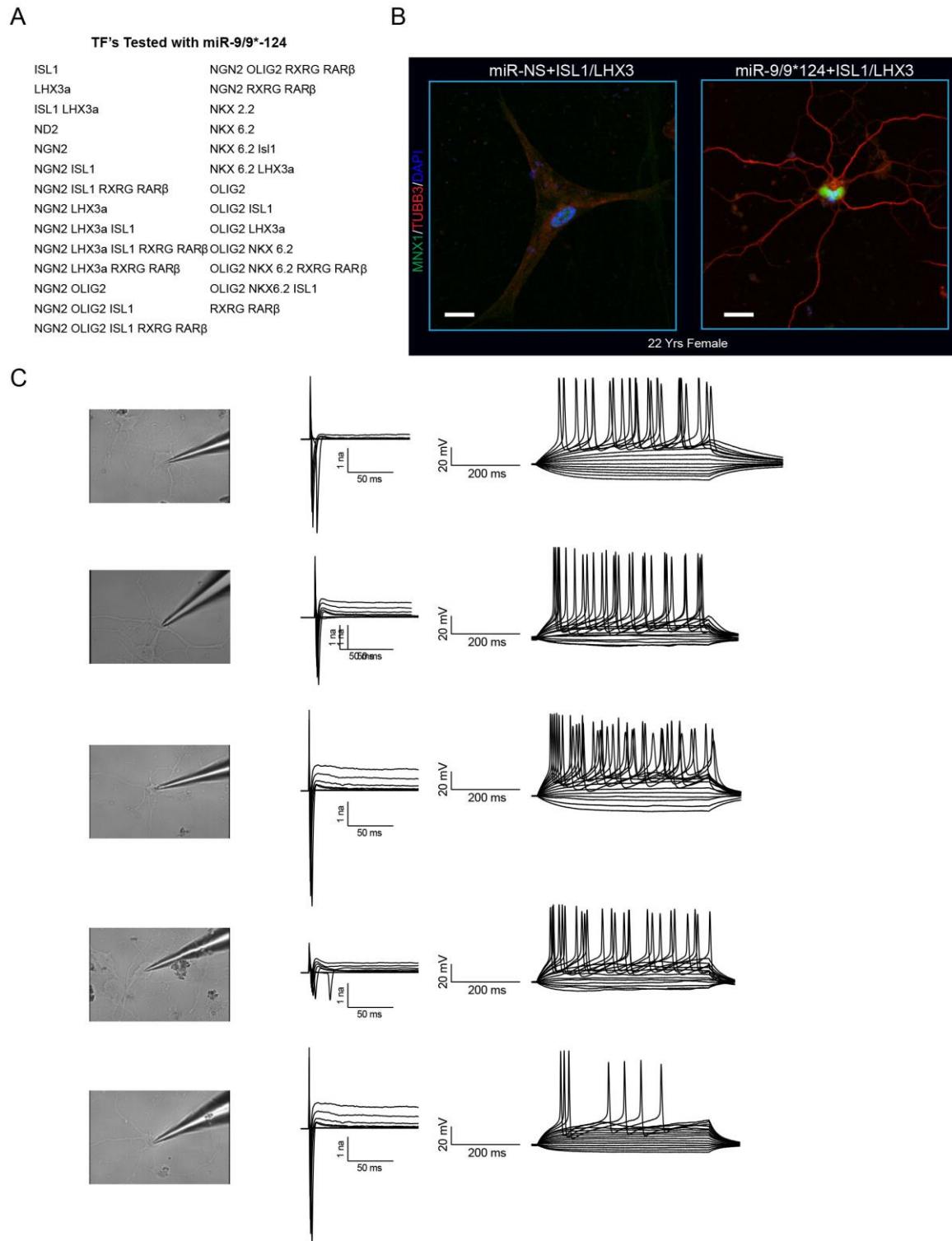


Fig. S5. Identification of Transcription Factors for Defining Motor Neuron Specific Conversion

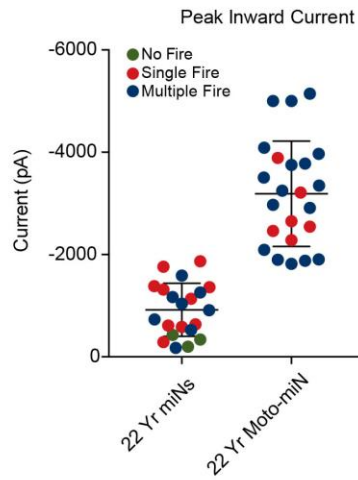
(A) Candidate motor neuron transcription factors and combinations co-expressed with miR-9/9*-124 in human fibroblasts.

(B) Immunocytochemistry of adult human fibroblasts overexpressing a non-specific miRNA (miR-NS) and ISL1/LHX3 (left) or miR-9/9*-124 and ISL1/LHX3 (right) for 35 days. Images demonstrate the necessity of miR-9/9*-124 for opening the neurogenic potential of human fibroblasts. Scale bar = 20 μ m.

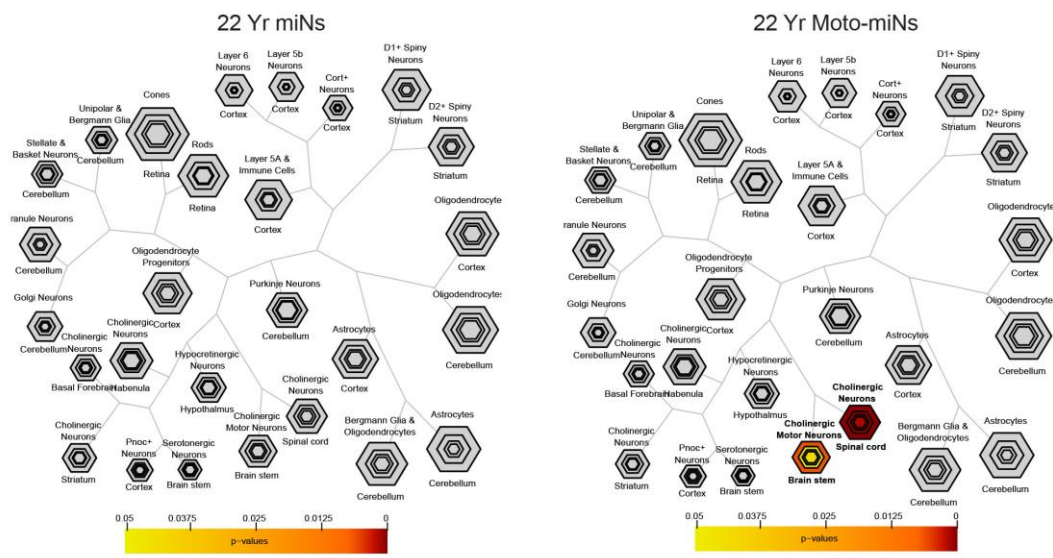
(C) Additional representative inward/outward whole-cell currents and repetitive AP waveforms generated from whole cell patch clamp recordings of Moto-miNs. Images show the patch clamped cells.

Figure S6

A



B



C

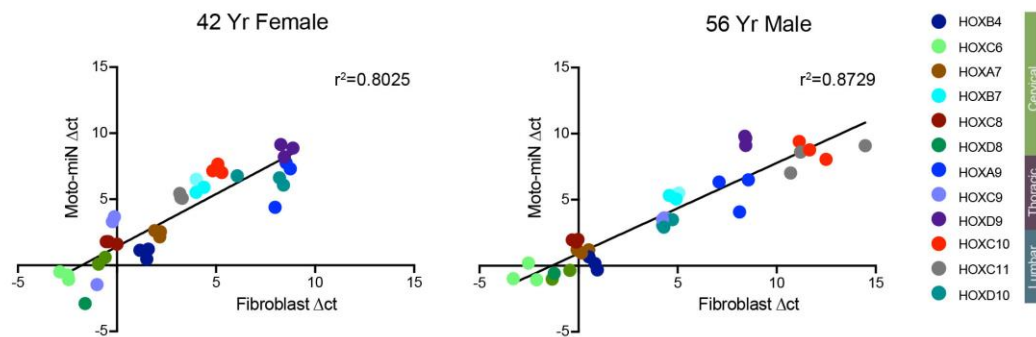


Fig. S6. Addition of ISL1/LHX3 to miR-9/9*-124 Increases Functional Maturity and Generates Motor Neuron Transcriptional Network.

(A) Peak inward current measured during voltage clamp mode of miNs and Moto-miNs reveals increased peak inward current in Moto-miNs ($-3,189 \text{ pA} \pm 214 \text{ pA}$) compared to miNs ($-919 \text{ pA} \pm 113 \text{ pA}$). Data are represented as mean \pm SEM.

(B) Cell Type-specific Enrichment Analysis (CSEA) tool reveals the top 100 most significantly expressed genes in miNs do not enrich for defined neuronal subtypes (left), while the top 100 most significantly expressed genes in Moto-miNs are enriched in cholinergic motor neurons in the brain stem and spinal cord.

(C) HOX gene expression analysis by qRT-PCR in 42-year-old female and 56-year-old male donor fibroblasts before and after conversion confirms that Moto-miNs retain donor fibroblast HOX gene expression patterns. Data represent Δ CT values for each biological replicate (3 separate Moto-miN conversions).

Figure S7

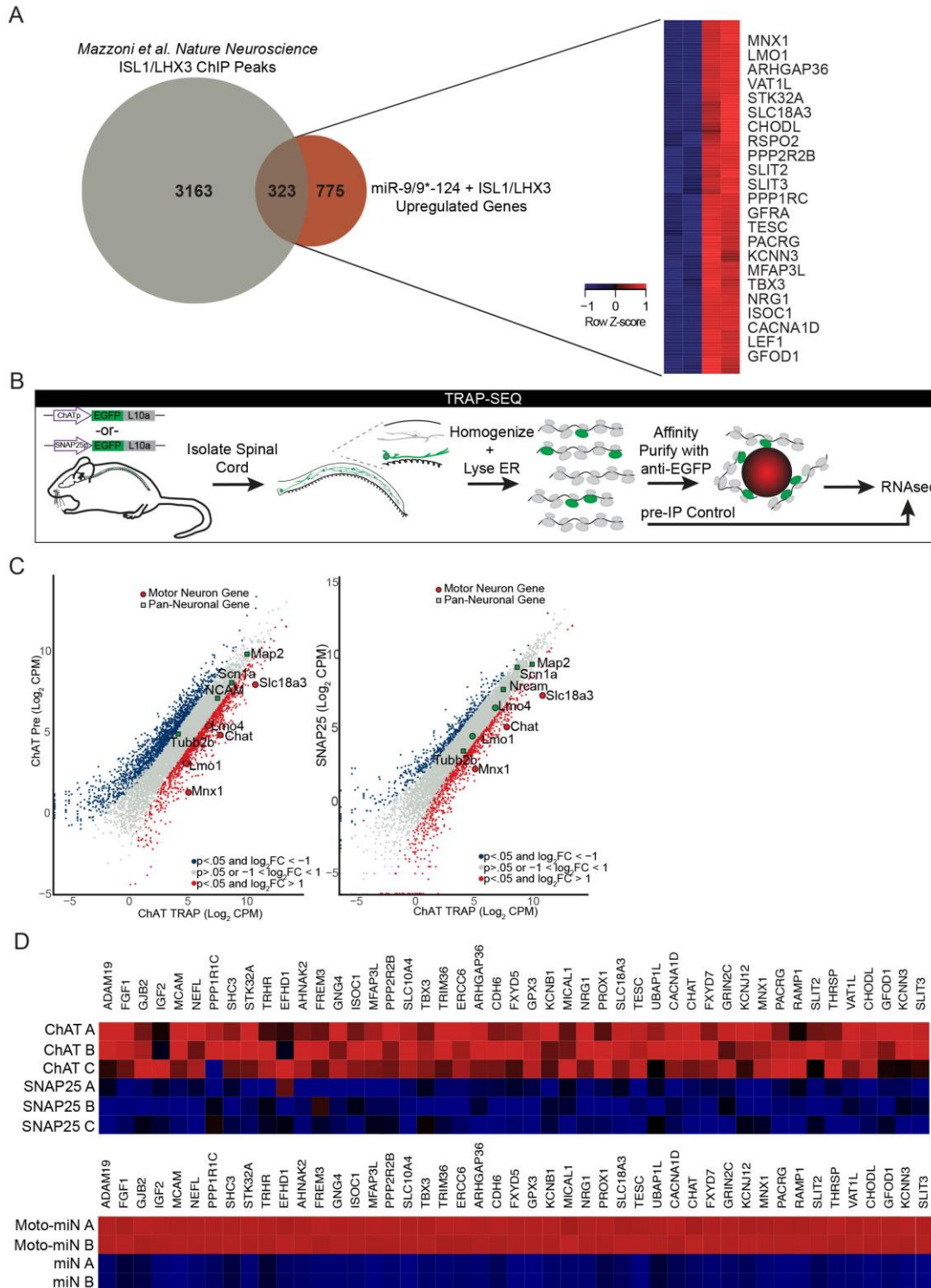


Fig. S7. Direct Comparison of Moto-miN Transcriptome to in vivo Mouse Motor Neurons by Translating Ribosomal Affinity Purification (TRAP) Sequencing.

- (A) Venn Diagram depicting the number of ISL1/LHX3 ChIP-seq peaks identified by Mazzone et al. during ISL1/LHX3 directed ES to motor neuron differentiation (3,486) and genes enriched in Moto-miN transcriptome (775). Heatmap shows overlapping activated genes (323) include hallmark motor neuron markers.
- (B) Schematic of TRAP-Seq strategy used to identify transcripts in all neurons (SNAP-25 genetic driver) and motor neurons (CHAT genetic driver) in mouse spinal cord. TRAP is a method to precipitate actively translated mRNA bound to ribosomes using an antibody to EGFP-L10A.
- (C) Pairwise comparisons between mean expression values in CHAT IP v. Pre-IP (left) and CHAT IP v. SNAP25 IP (right). Differentially expressed genes are shown in red ($\log_{2}FC > 1$ and $p < 0.05$) and blue ($\log_{2}FC < -1$ and $p < 0.05$).
- (D) Example mean expression values of overlapping genes between human (Moto-miNs versus miNs) and mouse (all spinal cord neurons SNAP25-TRAP) and motor neurons (CHAT-TRAP) datasets.

Supplemental Table 1

Related to STAR Methods section. Primer sequences used for qRT PCR analysis

Primer Name	Sequence (5'-3')
S100A4 FWD	GATGAGCAACTTGGACAGCAA
S100A4 REV	CTGGGCTGCTTATCTGGGAAG
VIM FWD	AGTCCACTGAGTACCGGAGAC
VIM REV	CATTTACGCATCTGGCGTTC
COL13A1 FWD	GGAGACGGCTATTTTGGGACG
COL13A1 REV	TCCTTGAGTGGAGCTTCCATT
ChAT FWD	TCAATCATGTCCAGCGAGTC
ChAT REV	AACGAGGACGAGCGTTTG
HB9 FWD	CTCCTACTCGTACCCGCAG
HB9 REV	TTGAAGTCGGGCATCTTAGGC
SLC18A3 FWD	TTCGCCTCTACAGTCCTGTTC
SLC18A3 REV	GCTCCTCCGGGTACTIONTATCG
HOXB4 FWD	CGTGAGCACGGTAAACCCC
HOXB4 REV	CGAGCGGATCTTGGTGTTG
HOXC6 FWD	ACAGACCTCAATCGCTCAGGA
HOXC6 REV	AGGGGTAAATCTGGATACTGGC
HOXA7 FWD	CGTTCCGGGCTTATAACAATGT
HOXA7 REV	CTCGTCCGTCTTGTCGCAG
HOXB7 FWD	TTCCCAGAACAACTTCTTGTGC
HOXB7 REV	GCATGTTGAAGGAACTCGGCT
HOXC8 FWD	ACCGGCCTATTACGACTGC
HOXC8 REV	TGCTGGTAGCCTGAGTTGGA
HOXD8 FWD	GGAAGACAAACCTACAGTCGC
HOXD8 REV	TCCTGGTCAGATAGGGGTAAAA
HOXA9 FWD	TACGTGGACTCGTTCCTGCT
HOXA9 REV	CGTCGCCTTGGACTGGAAG
HOXC9 FWD	ACTCGCTCATCTCTCACGACA
HOXC9 REV	GACGGAAAATCGCTACAGTCC
HOXD9 FWD	GGACTCGCTTATAGGCCATGA

HOXD9 REV	GCAAAACTACACGAGGCGAA
HOXC10 FWD	ACATGCCCTCGCAATGTAAC
HOXC10 REV	GAGAGGTAGGACGGATAGGTG
HOXC11 FWD	ATGTTTAACTCGGTCAACCTGG
HOXC11 REV	GCATGTAGTAAGTGCAACTGGG
HOXD11 FWD	TCTCCGAGTCCTCGTGGGGA
HOXD11 REV	GCAAAACACCAGCGCCTTCTA
HPRT FWD	TCCTTGGTCAGGCAGTATAATCC
HPRT REV	GTCAAGGGCATATCCTACAACAAA

2.5 Materials and Methods

Primary Cell Cultures

Primary fibroblasts utilized in this study: 1-year-old male (PCS-201-010, ATCC), 22-year-old female (GM02171, NIGMS Human Genetic Cell Repository at the Coriell Institute for Medical Research), 42-year-old female (F09-238, Washington University in St. Louis School of Medicine iPSC core facility), 56-year-old male (AG04148, NIA Aging Cell Repository at the Coriell Institute for Medical Research), and 68-year-old female (ND34769, NINDS Cell Line Repository at the Coriell Institute for Medical Research). Human skeletal myoblasts (HSMM) were obtained from a commercial source (Lonza, CC-2580). The sex of the human skeletal myoblasts was not provided; however, these cells were not used for neuronal conversion and therefore the sex of HSMM would not change the conclusions drawn from the current study.

Mice

Mouse lines Tg(Chat-EGFP/Rpl10a)DW167Htz (JAX: 030250) and Tg(Snap25-EGFP/Rpl10a)JD362Jdd (JAX: 030273) were purchased from Jackson Laboratory. All the mice were housed under pathogen-free conditions in the East McDonnell animal housing facility of Washington University in St. Louis under a 12 hr light-dark cycle. 21-day-old pooled male and female mice were used in the TRAP-seq studies; see Method Details for TRAP procedure. All

the subjects were not involved in any previous procedures.

Ethics statement:

The Washington University Animal Studies Committee approved all procedures used for the mouse experiments described in the present study. Overall care of the animals was consistent with *The Guide for the Care and Use of Laboratory Animals* from the National Research Council and the USDA *Animal Care Resource Guide*.

Cell Culture

Adult and neonatal human fibroblasts were maintained in fibroblast media comprised of Dulbecco's Modified Eagle Medium (Invitrogen) supplemented with 15% fetal bovine serum (Life Technologies), 0.01% β - mercaptoethanol (Life Technologies), 1% non-essential amino acids, 1% sodium pyruvate, 1% GlutaMAX, 1% 1M HEPES buffer solution, and 1% penicillin/streptomycin solution (all from Invitrogen). Cells were never passaged more than 15 times.

Plasmid construction and virus production

Complementary cDNA was generated from adult human spinal cord (Clontech) from which individual motor neuron transcription factors (Figure S5A) were subcloned into the N174 and N106 lentiviral vectors using standard techniques. Lentivirus was produced in Lenti-X 293T (293LE; Clontech, cat. no. 632180) cells. 293LE cells plated in 10cm dishes (6.5×10^6 cells per dish) via polyethyleneimine (48 μ L of 2 mg/mL, Polysciences) assisted transfection of 2nd generation packaging vectors (1.5 μ g pMD2.G and 4.5 μ g psPAX2), and 6 μ g of lentiviral backbone plasmid (e.g. pT-BCL-9/9*-124) 16 hours after initial plating. Media was changed the next day. After 2 days, media was collected, filtered through a 0.45 μ m polyethersulfone (PES) syringe filter and concentrated by centrifugation at 70,000xG for 2 hours at 4°C. Virus collected

from a single 10 cm dish was resuspended in 1 mL of sterile PBS then aliquoted and stored at -80°C. Before each transduction, virus aliquots were spun at 5,000xG for 5 minutes at 4°C to remove debris. Control vector expressing non-specific (NS) miRNA and BCL-XL was generated previously (Victor et al., 2014). Lentiviral BRG1 shRNA was generated by annealing then ligating the following sequences into the pLKO vector (Addgene, 8453), shBRG1 FWD 5' CCGGGGTGCTCAACACGCACTATGTCTCGAGACATAGTGCGTGTTGAGCACCTTTTT G 3', shBRG1 REV 5' AATTCAAAAAGGTGCTCAACACGCACTATGTCTCGAGACATAGTGCGTGTTGAGCAC C 3'.

Direct Neuronal Conversion

To initiate direct conversion, 1.8×10^6 cells were seeded onto Costar 6-well cell culture plates (Corning; 300,000 cells/well). The following day, each plate was transduced with the following reprogramming cocktail: 750 μ L of concentrated lentivirus containing the reverse tetracycline-controlled transactivator (rtTA; Addgene, 66810) and 500 μ L of virus containing pT-BclXL-9/9*-124 or pT-BclXL-9/9*-124 plus 500 μ L of each individual TF driven by the EF1 α promoter in the presence of polybrene (8 μ g/mL; Sigma-Aldrich) all diluted up to 18 mL (3 mL per well) then spininfected at 37°C for 30 minutes at 1,000xG using a swinging bucket rotor. The following day media was changed to fresh fibroblast media (2 mL per well) supplemented with doxycycline (Dox; Sigma Aldrich, 1 μ g/mL). After 2 days, fresh fibroblast media was changed and supplemented with Dox and antibiotics for respective vectors (Puromycin, 3 μ g/mL; Blastidicin 5 μ g/mL; Geneticin, 400 μ g/mL; all from Invitrogen). Five days post-transduction cells were replated on to poly-ornithine/laminin/fibronectin (PLF) coated glass coverslips. Before PLF coating, glass coverslips were acid treated as previously

reported (Richner et al., 2015). To transfer cells, for each well of a 6 well plate, cells were first washed 2x with 1 mL sterile PBS. Then 320 μ L of 0.25% Trypsin (Gibco) was added to each well then placed in an incubator. Cells were monitored every 2 minutes. As soon as cells began to detach (no more than 6 minutes), 1 ml of fibroblast media supplemented with 1 μ g/mL Dox was added to each well. One by one, each well was gently triturated three times to remove remaining attached cells then transferred to a sterile 1.5 mL Eppendorf tube. Cells were then spun at 200xG for 5 minutes at 37°C. The supernatant was aspirated and cells were gently resuspended in 300 μ L fibroblast media supplemented with Dox. Cells were then drop-plated onto either 18 mm (150 μ L per coverslip; placed in 12 well plate) or 12mm (60 μ L per coverslip; placed in 24 well plate) coverslips. Cells were left to settle for 15 minutes in an incubator then each well was flooded with fibroblast media supplemented with 1 μ g/mL Dox. The following day media was then changed to Neuronal Media (ScienCell, 1521) supplemented with Dox, valproic acid (1 mM; EMD Millipore) dibutyryl cAMP (200 μ M; Sigma-Aldrich), BDNF, NT-3, CNTF, GDNF (all 10 ng/mL, Peprotech), and Retinoic Acid (1 μ M; Sigma-Aldrich) and antibiotics for each vector. Dox was replenished every two days and half the media was changed every 4 days. Drug selection was stopped 14 days into conversion. A diagram of the reprogramming protocol is available in Figure S1.

For time-course studies, the above reprogramming protocol was utilized with the following exception. In 3-day intervals starting on reprogramming day 9, a full media change excluding Dox was used to effectively remove Dox from the well at each timepoint.

For DNA methylation profiling, 1.8×10^6 human neonatal fibroblasts were seeded onto 10 cm plates (Corning). To achieve the cell number required for meDIP and MRE-seq human neonatal fibroblasts were utilized as adult fibroblasts were unable to expand to sufficient

quantities. The following day, each plate was transduced with 10 ml of un-concentrated lentivirus containing media with a doxycycline inducible miR-9/9*-124 vector (Victor et al., 2014) and polybrene (8 µg/mL; Sigma-Aldrich). The following day media was changed to fresh fibroblast media (2 mL per well) supplemented with Dox. After 2 days, fresh fibroblast media was changed and supplemented with Dox and antibiotics for respective vectors (see Table 1). Seven days post-transduction, cells were first washed 2x with 3 mL sterile PBS. Then 1 ml of 0.25% Trypsin (Gibco) was added to each plate then placed in a 37°C tissue culture incubator. Cells were monitored every 2 minutes and as soon as cells began to detach (no more than 6 minutes), 4 ml of fibroblast media supplemented with 1 µg/mL Dox was added to each plate. Cells were transferred to Primaria modified 10 cm plates (Corning) and 5 ml fresh fibroblast media supplemented with 1 µg/mL Dox was added to a final volume of 10 ml. The following day media was changed to Neuronal Media (Sciencell) supplemented with Dox, valproic acid (1 mM; EMD Millipore) dibutyryl cAMP (200 µM; Sigma-Aldrich), BDNF, and NT-3 (all 10 ng/mL, Peprotech), Retinoic Acid (1 µM; Sigma-Aldrich), and 4% FBS. Dox was replenished every two days and half the media was changed every 4 days.

Myotube Differentiation

Human myotubes were generated by differentiating human myoblasts using defined culture conditions (Steinbeck et al., 2016). Briefly, human skeletal myoblasts were cultured according to manufacturers recommendations (HSMM; Lonza, CC-2580) then were plated on matrigel (0.1mg/mL) coated 12mm glass coverslips at a density of 80,000 cells/well. The following day HSMM's were differentiated by switching media to skeletal muscle differentiation media comprised of a 1:1 mixture of DMEM F12 (Gibco) and Complete Neuronal Media + 2% Horse Serum (Gibco). Every 2 days 75% of the media was replaced with fresh differentiation

media. After 10-14 days of differentiation, Moto-miNs 14 days into conversion and labeled with synapsin-eGFP via lentiviral transduction were replated onto the established myotubes at a 1:1 ratio (i.e. one 12mm Moto-miN coverslip was replated on top of a 12mm myotube coverslip). The following day media was changed to complete neuronal media and cells were cultured for 2 weeks. Dox was replenished every two days and half the media was changed every 4 days. After two weeks cells were fixed with 4% paraformaldehyde and processed for immunocytochemistry and α -Bungarotoxin staining.

Immunocytochemistry

Cells were fixed using 4% formaldehyde for 18 minutes at room temperature (RT) then blocked and permeabilized for one hour at room temperature in PBS containing 0.3% Triton-X100, 5% bovine serum albumin (Sigma-Aldrich), and 2% of either goat or donkey serum (Sigma-Aldrich). Primary antibodies were incubated overnight at 4°C in blocking buffer. Cells were then washed 3x and incubated with secondary antibodies conjugated to either Alexa-488, -594 or -647, for one hour at room temperature. α -Bungarotoxin was incubated with secondary antibodies at a concentration of 1:200. Images were obtained on a Leica SP-2 Confocal Microscope.

Electrophysiology

Whole-cell patch-clamp recordings were performed 35-40 days post-transduction. Data was acquired using pCLAMP 10 software with multiclamp 700B amplifier and Digidata 1550 digitizer (Molecular Devices). Electrode pipettes were pulled from borosilicate glass (World Precision Instruments) and typically ranged between 5–8 M Ω resistance. Intrinsic neuronal properties were studied using the following solutions (in mM): Extracellular: 140 NaCl, 3 KCl, 10 Glucose, 10 HEPES, 2 CaCl₂ and 1 MgCl₂ (pH adjusted to 7.25 with NaOH). Intracellular:

130 K- Gluconate, 4 NaCl, 2 MgCl₂, 1 EGTA, 10 HEPES, 2 Na-ATP, 0.3 Na-GTP, 5 Creatine phosphate (pH adjusted to 7.5 with KOH). Membrane potentials were typically kept at -65 mV. In voltage-clamp mode, currents were recorded with voltage steps ranging from -20 mV to +90 mV. In current-clamp mode, action potentials were elicited by injection of step currents that modulated membrane potential from -10 mV to +35 mV. Data was collected in Clampex and initially analyzed in Clampfit (Molecular Devices). Further analysis was done in GraphPad Prism 7 (GraphPad Software). Liquid junction potential was calculated to be 15.0 mV and corrected in calculating resting membrane potential according to previously published methods (Barry, 1994).

RNA-Sequencing

Total RNA from day 30 miNs and starting human adult fibroblasts (22 yr old) were extracted by RNeasy plus micro kit (Qiagen). The RNA samples with > 9.5 of RIN based on a 2100 Bioanalyzer were used for RNA-Seq library preparation. Library preparation and sequencing were performed by Genome Technology Access Center in Washington University School in St. Louis. Briefly mRNA was isolated by using SMARTer Ultra Low RNA Kit for Illumina sequencing (Clontech). All cDNA libraries, based on two biological replicates for each condition, were sequenced on Illumina Hi-Seq 2500 with single-end 50 bp read length. For RNA-seq time series data in Figure 3, total RNA was extracted from starting human adult fibroblasts (22 yr old) and day 3, 6, 10, and 20 miNs by RNeasy plus micro kit (Qiagen). The RNA samples with > 9.5 of RIN based on a 2100 Bioanalyzer were used for RNA-Seq library preparation. Total RNA was processed for library construction by Cofactor Genomics (<http://cofactorgenomics.com>, St. Louis, MO) according to the following procedure: Briefly, total RNA was reverse-transcribed using an Oligo(dT) primer, and limited cDNA amplification

was performed using the SMARTer® Ultra® Low Input RNA Kit for Sequencing – v4 (Takara Bio USA, Inc., Mountain View, CA). The resulting full-length cDNA was fragmented and tagged, followed by limited PCR enrichment to generate the final cDNA sequencing library (Nextera® XT DNA Library Prep, Illumina, San Diego, CA). Libraries were sequenced as single-end 75 base pair reads on an Illumina NextSeq500 following the manufacturer's instructions by Cofactor Genomics.

DREM Analysis

The Dynamic Regulatory Events Miner (DREM) and cDREM were used to integrate time series gene expression data with predicted TF-gene binding interactions (top 100 genes per PWM from Ernst et al., 2010 to identify patterns of temporal gene expression patterns and the associated regulators. The log fold change of 7310 DE genes (fold-change > 10; q-value < 0.1) was used to generate all paths in Figure 3.

MicroArray

Total RNA was extracted from miNs and Moto-miNs derived from 22 yr old donor fibroblasts alongside corresponding starting fibroblast controls using TRIzol (Thermo Fisher Scientific, Waltham, MA) according to the manufacturer's instruction followed by extraction using chloroform and then ethanol precipitation. RNA quality was determined by the ratio of absorbance at 260 nm and 280 nm to be approximately 2.0. Samples for RNA microarray were then standardly prepped and labeled with Illumina TotalPrep kits (Thermo Fisher Scientific, Waltham, MA) for Agilent Human 4x44Kv1. Standard hybridization and image scanning procedure were performed according to the manufacturer's protocol at Genome Technology Access Center at Washington University School of Medicine, St. Louis.

Methylated DNA immunoprecipitation sequencing

MeDIP-seq was performed as in Maunakea *et al.* (Maunakea et al., 2010). Five micrograms of genomic DNA was sonicated to a fragment size of ~100-400 bp using the Bioruptor sonicator (Diagenode). End-repair, addition of 3'-A bases and PE adapter ligation with 2 µg of sonicated DNA was performed according to the Illumina Genomic DNA Sample Prep Kit protocol. Adapter-ligated DNA fragments were size selected to 166-366 bp and purified by gel electrophoresis. DNA was heat denatured and then immunoprecipitated with 5-methylcytidine antibody (Eurogentec; 1 µg of antibody per 1 µg of DNA) in 500 µl of immunoprecipitation buffer (10 µM sodium phosphate, pH 7.0, 140 mM sodium chloride and 0.05% Triton X-100) overnight at 4 °C. Antibody/DNA complexes were isolated by addition of 1 µl of rabbit anti-mouse IgG secondary antibody (2.4 mg ml⁻¹, Jackson ImmunoResearch) and 100 µl protein A/G agarose beads (Pierce Biotechnology) for 2 h at 4 °C. Beads were washed nine times with immunoprecipitation buffer and then DNA was eluted in TE buffer with 0.25% SDS and 0.25 mg/mL of proteinase K for 2 h at 50 °C. DNA was then purified with the Qiagen QIAquick kit and eluted in 30 µl EB buffer. Ten microliters of DNA was used for a PCR-enrichment reaction with PCR PE Primers 1.0 and 2.0. PCR products were size selected (220-420 bp) and purified by gel electrophoresis. Methylated DNA enrichment was confirmed by PCR on known methylated (SNRPN and MAGEA1 promoters) and unmethylated (a CpG-less sequence on chromosome 15 and glyceraldehyde 3-phosphate dehydrogenase promoter) sequences. DNA libraries were checked for quality by Nanodrop (Thermo Scientific) and Agilent DNA Bioanalyzer (Agilent). Reads were aligned to hg19 using BWA and pre-processed using methylQA (an unpublished C program; available at <http://methylqa.sourceforge.net/>). Detailed library construction protocols for MRE-seq and MeDIP-seq are publically available at the NIH Roadmap Epigenomics project website

[\(http://www.roadmapepigenomics.org/protocols/type/experimental/\)](http://www.roadmapepigenomics.org/protocols/type/experimental/).

Me-sensitive restriction enzyme sequencing

Methylation (Me)-sensitive enzyme sequencing (MRE-seq) was performed as in Maunakea *et al.* (Maunakea et al., 2010), with modifications as detailed below. Five parallel restriction enzyme digestions (*HpaII*, *Bsh1236I*, *SsiI(AciI)* and *Hin6I* (Fermentas), and *HpyCH4IV* (NEB)) were performed, each using 1 µg of DNA per digest for each of the samples. Five units of enzyme were initially incubated with DNA for 3 h and then an additional five units of enzyme were added to the digestion for a total of 6 h of digestion time. DNA was purified by phenol/chloroform/isoamyl alcohol extraction, followed by chloroform extraction using phase lock gels. Digested DNA from the different reactions was combined and precipitated with one-tenth volume of 3 M sodium acetate (pH 5.2) and 2.5 volumes of ethanol. The purified DNA was size selected and purified (50-300 bp) by gel electrophoresis and Qiagen MinElute extraction. Library construction was performed as per the Illumina Genomic DNA Sample Prep Kit protocol with the following modifications. During the end-repair reaction, T4 DNA polymerase and T4 PNK were excluded and 1 µl of 1:5 diluted Klenow DNA polymerase was used. For the adapter ligation reaction, 1 µl of 1:10 diluted PE adapter oligo mix was used. Ten microliters from the 30 µl of purified adapter ligated DNA was used for the PCR enrichment reaction with PCR PE Primers 1.0 and 2.0. PCR products were size selected and purified (170-420 bp) by gel electrophoresis and Qiagen QIAquick extraction. DNA libraries were checked for quality by Nanodrop (Thermo Scientific) and Agilent DNA Bioanalyzer (Agilent). Reads were aligned to hg19 using BWA and pre-processed using methylQA. MRE reads were normalized to account for differing enzyme efficiencies and methylation values were determined by counting reads with CpGs at fragment ends (Maunakea et al., 2010).

Genomic features

DMRs from day 30 (miN day 30 vs. Ctrl day 30) were segregated into exons, introns, intergenic regions, 3' UTRs, 5' UTRs, non-coding regions, promoter-TSSs, and TTSs by using the annotatePeaks program provided by HOMER (Heinz et al., 2010).

ATAC-sequencing

ATAC-seq was performed as previously described (Buenrostro et al., 2013). Briefly, 50,000 cells were collected for ATAC-seq library preparation at Ctrl day 10, miNs day 10, miNs day 20, shBRG1 day 20 and shCTRL day 20. Transposition reaction was carried out with Nextera Tn5 Transposase for 30 min at 37 °C. Library fragments were amplified for optimal amplification condition. Final libraries were purified using AMPure XP beads (Ampure) and sequenced with 50 bp paired-end reads on Illumina HiSeq 2500.

Translating ribosome affinity purification

Translating ribosome affinity purification (TRAP) (Heiman et al., 2014) was performed on spinal cord dissections pooled from 3-4 mice 21 days post birth that were positive for the eGFP-L10A fusion ribosomal marker protein under either the Chat promoter (Tg(Chat-EGFP/Rpl10a)DW167Htz) or the Snap25 promoter (Tg(Snap25-EGFP/Rpl10a)JD362Jdd). TRAP samples underwent immunopurification for four hours at 4°C. Both TRAP and pre-immunopurification control RNA samples were extracted through TRIzol purification, DNase treatment, and Qiagen RNeasy Mini columns (74104). Quality and quantity of RNA was assessed using a Bioanalyzer 2100 RNA Pico Chip. Sequencing libraries were amplified using Nugen Amplification Kit Ovation® RNA Seq System V2 (7102). Genome Technology Access Center at Washington University in St. Louis performed adapter ligation and sequencing of the libraries on the Illumina HiSeq2500. Three replicates of this procedure were analyzed.

Immunoblot analysis

Human fibroblasts (ATCC) expressing shBRG1 or shCtrl were lysed 7 days post-transduction in RIPA buffer (Thermo Scientific) supplemented with protease inhibitors (Roche). The concentrations of whole cell lysates were measured using the Pierce BCA protein assay kit (Thermo Scientific). Equal amounts of whole cell lysates were resolved by SDS-PAGE and transferred to a nitrocellulose membrane (GE Healthcare Life Sciences) using a transfer apparatus according to the manufacturer's protocols (Bio-rad). After incubation with 5% BSA in TBS containing 0.1% Tween-20 (TBST) for 30 min, the membrane was incubated with primary antibodies at 4°C overnight. Following incubation, membranes were incubated with a horseradish peroxidase-conjugated anti-mouse or anti-rabbit antibody for 1 hr. Blots were developed with the ECL system (Thermo Scientific) according to the manufacturer's protocols.

Quantitative PCR

Total RNA was extracted using TRIzol (Invitrogen, USA) according to the manufacturer's instruction. Reverse-transcribed complementary DNA (cDNA) was synthesized from 500ng of RNA with SuperScript III First-Strand Synthesis SuperMix (Invitrogen, USA) or from 10 ng of RNA for microRNA expression analyses using specific stem-loop primer probes from TaqMan MicroRNA Assays (Invitrogen, USA). Subsequently, the cDNA was analyzed on a StepOnePlus Real-Time PCR System (AB Applied Biosystems, Germany). The following primers were utilized:

miRNA qRT-PCR primers

hsa-miR-218 (Thermo Fisher Cat. # 4427975)

RNU-44 (Thermo Fisher Cat. # 4427975)

For primers utilized for mRNA qRT PCR see Table S1.

Overlap with ISL1/LHX3 ChIP Seq

To identify the genes regulated by LHX3 and ISL1 in motor neurons, ISL1- and LHX3-ChIP sequencing data (Mazzoni et al., 2013) were used. We selected the regions co-occupied by ISL and LHX3 during ES to motor neuron differentiation, accounting for 84.2% of peak regions called in each ChIP-seq data. Based on the peaks co-occupied by ISL1 and LHX3, we annotated 3,486 closest genes with peaks located within 5Kb upstream of TSS and intragenic regions. Comparing those annotated genes with genes selectively enriched in Moto-miNs vs miNs (\log_2 fold change ≥ 2.5 , $p < 0.01$), identified 323 genes co-occupied by ISL1 and LHX3 that are also upregulated when miR-9/9*-124 is co-expressed with ISL1/LHX3.

Experimental Design

For all experiments performed on miNs and Moto-miNs including molecular and functional characterizations of these cells, we do not specifically include or exclude samples or data. At minimum, biological duplicates were utilized for each experiment except DNA methylation profiling, where day 20 and day 30 conditions were treated as replicates to identify final DMRs. No statistical methods were used to predetermine sample size. Further specifics about the replicates (experimental n number) and performing procedures for each experiment are available in the Figure Legends and Star Methods.

Analysis of converted cells

The quantification of TUBB3, MAP2, NCAM and NEUN-positive cells over total number of cells (DAPI) shown in Figure 1C, Figure 6C, Figure S1F, and Figure S4C was done as follows: for TUBB3, MAP2, and NCAM only cells with processes at least three times the length of the soma were counted. For NEUN, only cells with proper nuclear localization were counted.

Figure 1C data are represented as mean \pm SEM. N = number of cells analyzed in at least 6 random fields of view from biological replicates. 22 Yr Female N = 238 cells, 42 Yr Female, N = 100 cells, 56 Yr Male N = 171 cells, and 68 Yr Female N = 216 cells. Figure 6C data are represented as mean \pm SEM. Cells (N) analyzed: 22 yr old N=TUBB3 325, MAP2 219, NCAM 275; 42 yr old N=TUBB3 304, MAP2 236, NCAM 129; 56 yr old N=TUBB3 275, MAP2 279, NCAM 213; 68 yr old N=TUBB3 282, MAP2 234, NCAM 190. Quantification shown in Figure 6E represents the total percentage of MNX1, CHAT and SMI-32-positive cells over TUBB3-positive cells. Data are represented as mean \pm SEM. Cells (N) analyzed: 22 yr old N=MNX1 256, CHAT 256, SMI-32 113; 42 yr old N= MNX1 151, CHAT 151, SMI-32 283; 56 yr old N= MNX1 207, CHAT 207, SMI-32 174; 68 yr old N= MNX1 151, CHAT 151, SMI-32 96. Figure S1F data are represented as mean \pm SEM. Cells (N) analyzed: Starting Fibroblasts N= DAPI 155, FSP1 155, MAP2 0; Day 9 N= DAPI 96, FSP1 26, MAP2 20; Day 12 N= DAPI 290, FSP1 106, MAP2 29; Day 15 N= DAPI 313, FSP1 88, MAP2 56; Day 18 N= DAPI 88, FSP1 27, MAP2 26; Day 21 N= DAPI 175, FSP1 31, MAP2 62; Day 24 N= DAPI 230, FSP1 27, MAP2 95; Day 27 N= DAPI 290, FSP1 5, MAP2 143; Day 30 N= DAPI 203, FSP1 0, MAP2 155. Figure S4C data are represented as mean \pm SEM. Cells (N) analyzed: Day 20 shBRG1 N= DAPI 155, FSP1 32, MAP2 11; Day 20 sh CTRL N= DAPI 173, FSP1 22, MAP2 117

RNA-Seq data analyses

More than 35 million reads of each RNA-seq sample were aligned to human genome assembly GRCh 37. For differential expression analysis, edgeR and limma were used. Genes with low read counts, regarded as genes not expressed at a biologically meaningful level were filtered out before read normalization. The cut-off for low read count was counts per million (CPM) < 1 in at least any two samples across the experiment. Reads for each sample were

normalized by the edgeR method of trimmed mean of M-values (TMM). The quantitative difference of read counts between miNs and starting fibroblast samples were evaluated by carrying out limma and graphically represented by Glimma. Gene enrichment analysis for differentially expressed genes was performed using Metascape Gene Annotation and Analysis Resource tool.

For RNA-seq time series data in Figure 3, quality control, alignment, clustering, normalization, and expression comparison were performed by Cofactor Genomics (<http://cofactorgenomics.com>, Saint Louis, Missouri, USA). Raw sequence data in FASTQ format were assessed for quality (FastQC, <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>) and ribosomal RNA content (sortmeRNA, <http://bioinfo.lifl.fr/RNA/sortmerna/>). NovoAlign (Novocraft, <http://novocraft.com>) was used to align reads to a set of transcript sequences. NovoAlign parameters were set to allow multiple alignments to the transcriptome set to allow for isoforms. Alignments to the genome were performed using STAR (<https://github.com/alexdobin/STAR>). Only unique alignments to the genome were allowed. The genome alignment loci from all samples were combined and clustered to generate genomic loci (“patches”) with contiguous read coverage. Patches overlapping reference genome annotation loci were labelled as such. For each transcript or patch, the RPKM expression value was calculated for each sample. These RPKM values are the basis for expression comparison and statistics generation. For each replicate group, the mean and coefficient of variation for each transcript or patch were calculated across the expression values for the samples in that group. These means were considered to be the expression values for the replicate group. Adjusted P-values were calculated between the means of each pair of replicate groups using a Welch’s t-test corrected for false discovery rate (FDR) by

the method of Benjamini-Hochberg.

MicroArray analyses

The intensity of the probes was imported into Partek and quantile normalized. Differentially expressed genes were identified using Partek with a cut-off of adjusted p-value < 0.05 and over 2.5 log₂ fold expression change.

Quantitative-PCR analyses

Expression data were normalized to housekeeping genes *HPRT1* and *RNU44* for coding genes and microRNAs, respectively, and analyzed using the $2^{-\Delta\Delta CT}$ relative quantification method. Data are represented as mean +/- S.E.M. from 3 biological replicates each individually performed in technical triplicates. For primers utilized for mRNA qRT PCR see Table S1.

Differential DNA-methylated region analysis

The M&M statistical model (Zhang et al., 2013), which integrates MeDIP-seq and MRE-seq data to identify differentially methylated regions between two samples was implemented with a window size of 500 bp and a *q*-value (FDR-corrected *P*-value) cutoff of 5e-2. This cutoff was determined from Figure 3C, where only 1 DMR was detected at day 10 (miN day 10 vs. Ctrl day 10). For Figure 3B, only regions that were considered DMRs (*q*-value < 1e-5) at both day 20 (miN day 20 vs. Ctrl day 20) and day 30 (miN day 30 vs. Ctrl day 30) are displayed.

GO enrichment analyses

DNA methylation GO analyses of MGI (Mouse Genome Informatics) expression (Smith et al., 2014) presented in Figure 3D were performed using the GREAT package (McLean et al., 2010). Gene regulatory domains were defined by default as the regions spanning 5 kb upstream and 1 kb downstream of the TSS (regardless of other nearby genes). Gene regulatory domains were extended in both directions to the nearest gene's basal domain, but no more than a maximum extension in one direction. The top 100 most significant overlapping DMRs from day

20 (miN day 20 vs. Ctrl day 20) and day 30 (miN day 30 vs. Ctrl day 30) were used as input. For Figure 3G, GO analyses were performed using Metascape (Tripathi et al., 2015) with a minimum enrichment of 1.5, a minimum overlap of 3, and a *p*-value cutoff of 0.01 using all demethylated or methylated DMRs at day 30 (miN day 30 vs. Ctrl day 30) that were either up- or down-regulated, respectively, by RNA-seq at day 35 using a cutoff of 2.5 logFC.

ATAC-Seq analyses

More than 50 million ATAC-seq reads per sample were trimmed for Nextera adapter sequences using TrimGalore and aligned to hg19 human genome assembly using bowtie2 with parameters `--very-sensitive --maxins 2000 --no-discordant --no-mixed`. Duplicate reads were discarded with Picard and uniquely mapped reads were used for downstream analysis. Peaks were called using Homer with parameters `findPeaks -region -size 150 -minDist 300`. Peaks called from all the samples were combined and raw reads mapped on the combined peaks were counted using HTSeq-count.

Differential peaks between any two different samples of Ctrl day 10, miNs day 10, miNs day 20 were identified using edgeR with a cut-off of fold-change > 1.5 and FDR < 0.01 .

Differential peaks were regarded as peaks gained or lost at each time point. Gained peaks at miNs day 10 and day 20 were combined and defined as open chromatin regions. Conversely, all lost peaks at miNs day 10 and day 20 were defined as close chromatin regions. The genomic features in the differential open and closed chromatin regions were distributed by CEAS software (Shin et al., 2009). Differential peaks between shBRG1 day 20 and shEGFP day 20 were defined using an independent cut-off (\log_2 -fold-change ≥ 2 , FDR < 0.05).

We annotated Ref-seq genes nearest differential peaks with Homer `annotatePeaks` command. Based on the genomic distributions and peak annotations, we defined the promotor

regions (-/+ 2 kb from TSS) and distal regions (all peak positions except the promoter regions). GO enrichment analyses were performed by Metascape or the Gene Ontology. All heatmaps were made based on normalized signal intensity values (i.e. \log_2 CPM) of each sample on relevant specific regions.

All histone mark ChIP-seq data were obtained from Roadmap Epigenome database of human fibroblasts (Roadmap Epigenomics et al., 2015). To identify histone mark-occupied chromatin accessibility during reprogramming, we compared each histone ChIP-seq data with open and closed chromatin regions based on ATAC-seq. We confirmed that most open and closed chromatin peaks overlapped with histone mark-ChIP peaks were found outside promoter regions (+/- 2 kb from TSS). Open and closed chromatin regions excluding promoter regions were then compared with histone ChIP-seq data and analyzed for GO enrichment.

Analysis of TRAP RNA-Seq data

RNA-Seq reads were mapped to Ensembl release 76 using STAR (analysis performed by Genome Technology Access Center at Washington University in St. Louis). For downstream analyses, only those genes with CPM > 1 in at least 3 samples, with an Ensembl gene biotype of "protein_coding," were retained. For gene symbols mapping to multiple Ensembl gene IDs, only the ID with the highest number of mapped reads was retained, resulting in a total of 14,009 genes used for downstream analyses. Using edgeR, read counts were fit to a negative binomial generalized log-linear model, and a likelihood ratio test was done to determine differential expression.

Comparative analysis of RNA-seq and microarray data

For comparative analysis, only probes with a detected call in at least 1 of 6 samples was retained, resulting in 23,775 probes mapping to a gene symbol. Expression level was then

averaged over all probes for each gene, resulting in a total of 15,333 genes that were used for comparative analysis, 10,736 of which were also present in the gene set retained from the RNA-seq dataset (described above) after CPM filtering. Within the genes retained in both datasets, the top differentially expressed genes between motor neurons and controls - CHAT IP vs. SNAP25 IP ($\log_{2}FC > 1$ and $p < 0.05$) in the RNA-seq dataset, and Moto-miN vs. miN ($\log_{2}FC > 2.5$ and adjusted $p < 0.05$) in the microarray dataset - were assessed for significant contingency using a one-tailed Fisher's exact test.

2.6 Data Availability

The RNA-seq, ATAC-seq, MeDIP and MRE-seq, and microarray data generated in this study are: 14 samples, single-end RNA-seq libraries from starting fibroblasts and miNs throughout direct conversion (Days 3, 6, 10, 20 and 30), 2 replicates per timepoint; 10 paired-end ATAC-seq libraries from day 10 Ctrl, day 10 and day 20 miNs, day 20 miNs with shBRG1 and day 20 miNs with an shCtrl, 2 replicates for each timepoint; 6 microarrays from starting fibroblasts, day 35 miNs, and day 35 moto-miNs, 2 replicates per condition; 6 samples of paired-end MeDIP libraries and 6 samples of MRE single-end libraries from day 10, 20, 30 miNs and corresponding Ctrl. The GEO accession number for the RNA-seq, ATAC-seq, MeDIP and MRE-seq, and microarray data reported in this paper is GSE102035. The GEO accession number for the TRAP-seq data is GSE93412

Chapter 3: LONGO—An R Package for Interactive Gene Length Dependent Analysis for Neuronal Identity

Adapted from:

LONGO: An R Package for Interactive Gene Length Dependent Analysis for Neuronal Identity

Matthew J. McCoy*, Alexander J. Paul*, Matheus B. Victor, Michelle Richner, Harrison
W. Gabel, Haijun Gong, Andrew S. Yoo and Tae-Hyuk Ahn

*Authors contributed equally

Bioinformatics, In Press

Copyright © 2018 Oxford University Press, All Rights Reserved

3.1 Methods

3.1.1 LONGO algorithm and framework

We developed LONGO, an R package that takes gene expression data from transcriptome experiments (e.g. RNA-seq or RNA microarray) as inputs from species represented on the ENSEMBL BioMart project (Smedley, et al., 2015). Figure 1 shows an overview of the LONGO algorithm using Rat BodyMap data (Yu, et al., 2014). The length of genes from the user-specified species is calculated from gene start and stop positions, which are retrieved automatically using the biomaRt package (Durinck, et al., 2009). Gene-length dependent expression of samples is then calculated using a sliding window of genes sorted in ascending order by the length of the genes. LONGO quantile normalizes and filters input data (optional), calculates the median expression (or mean expression optionally) of genes binned by length (default genes per bin: 200), and then consecutively calculates the expression of the next bin (default step size: 40 genes). LONGO comes in two versions: one with a local HTML GUI for interactive work using the R Shiny package (LONGO) and one without a GUI for batch processing of data files (LONGOcmd). One strength of LONGO is provided through interactive plots—users have the option to modify parameters such as bin size, step size, sliding window mean or median, graph scales, as well as which samples are used as controls for downstream analyses. The significant difference of gene-length dependent expression is statistically tested by the Jensen-Shannon divergence (JSD) and the long gene quotient (LQ), which we develop herein. In addition, LONGO shows a plot of p-values generated using the Wilcox test comparing the binned gene expression values to the control.

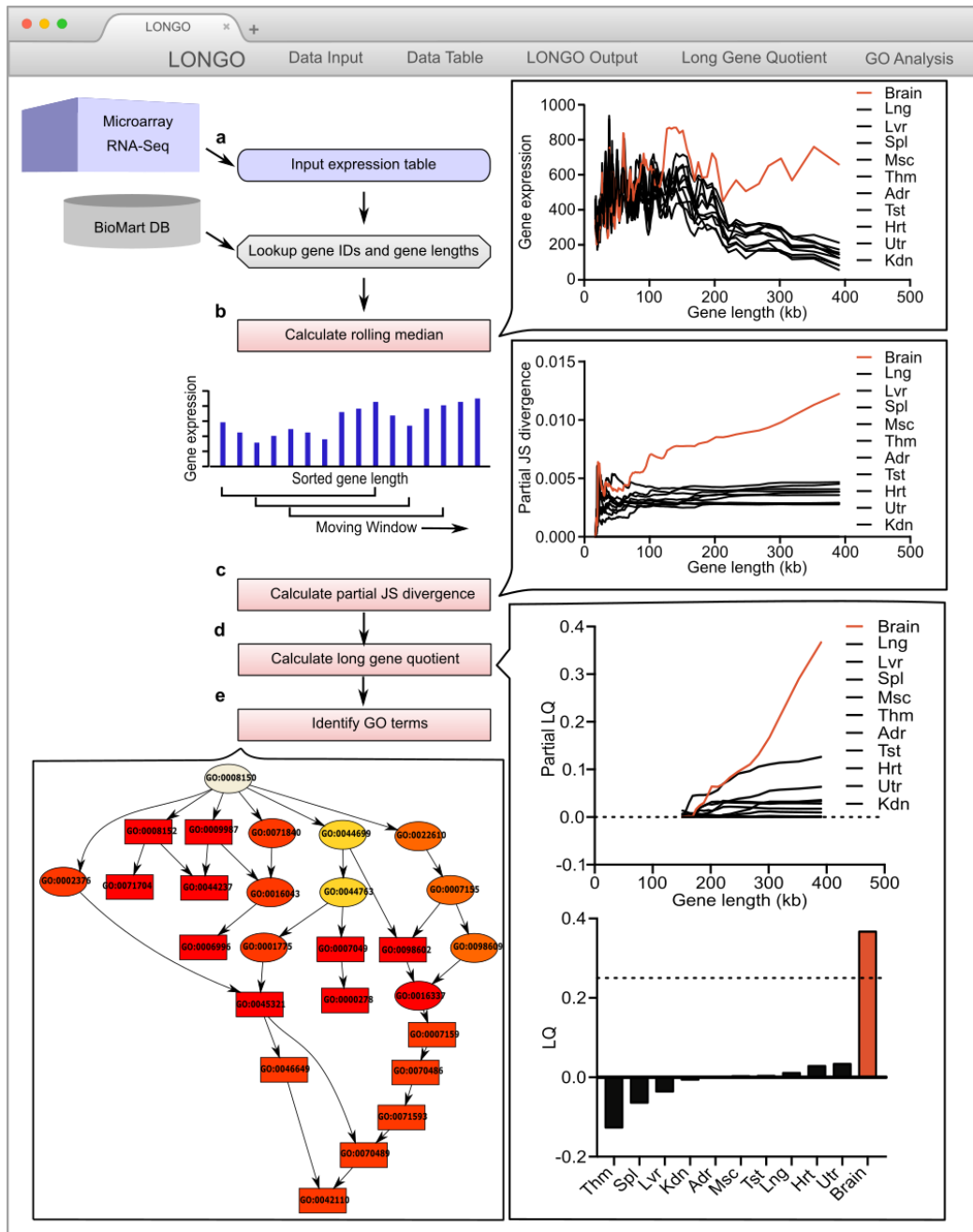


Fig. 1. Illustration of LONGO output using Rat BodyMap data.

RNA-seq of rat tissues from Rat BodyMap (Yu, et al., 2014): non-neuronal tissues (black) and brain tissues (red). (a) User inputs gene expression table into LONGO, which associates gene lengths with each gene. (b) LONGO calculates rolling median per user-defined parameters (For this figure: 200 gene bins, 40 gene step). (c) LONGO calculates partial increasing JSD from short genes to long genes between each sample and a user-defined control sample. (d) LONGO calculates a partial LQ (left) and final LQ (right). (e) LONGO identifies GO terms.

3.1.2 Measuring Jensen-Shannon Divergence

The Kullback-Leibler divergence (KLD) is a non-commutative measure of the difference between two probability distributions P and Q , typically P representing the “true” distribution and Q representing an arbitrary probability distribution. One symmetrized and smoothed version of the KLD is the JSD (Endres and Schindelin, 2003), defined as

$$D_{JS}(P||Q) = \frac{1}{2}D_{KL}(P||M) + \frac{1}{2}D_{KL}(Q||M)$$

where $M = 0.5(P + Q)$ and the KLD is defined to be

$$D_{KL}(P||Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$$

where $Q(i) \neq 0$, and the summation is taken over histogram bins. Therefore, for gene expression data ordered by gene length, we set P as the distribution from the control sample, and Q the distribution obtained from each of the testing samples. Smaller values of KLD represent more similar distributions. LONGO first measures the JSD for the shortest genes in two samples, then iteratively re-computes the JSD after adding the next longest until all genes in respective datasets have been added. The final JSD is the JSD between all expressed genes.

3.1.3 Developing the long gene quotient

To systematically discover disproportionately elevated long versus short gene expression between two samples, we adapted the JS divergence to generate the long gene quotient. LONGO first removes all genes whose gene length is less than the median gene length (typically ~30 kb)

for the dataset, which allows a clearer assessment of differences in long gene expression. Then LONGO calculates the partial LQ (PLQ) for any testing sample Q_i ($i = 1, \dots, n$), which we define as

$$PLQ_i = \frac{D'_{JS}(P||Q_i)}{\max\{D_{JS}(P||Q_j), j = 1, \dots, n\}}$$

where P represents the distribution of the control sample and D'_{JS} only considers long genes (>150 kb). The PLQ describes the relative change in JSD of each sample compared with the maximum changes of JSD across all samples, so its magnitude ranges from 0 to 1.

Because the underlying JSD does not distinguish the directionality of changes, we define the final LQ as $LQ = S(PLQ)$, where $S = -1$ if $\text{mean}(Q) < \text{mean}(P)$. This ensures that LQ is positive only if LGE is higher in the testing sample than in the control sample. The LQ is dependent on the control sample selected, as well as the sample with the most extensive LGE. Therefore, while LQ is effective at distinguishing neuronal from non-neuronal cells in the absence of a positive control, we strongly encourage including a positive control neuronal sample. The LQ is also dependent on the sliding window bin and step size; for the default of 200 genes per bin with a 40 gene-step, using the median, neuronal cells typically exceed an LQ of 0.25.

In addition to LQ, the correlation between binned gene expression and median length of bins (default genes per bin: 200; default step size: 40 genes) for genes longer than 100 kb also identifies neuronal sample when highly positive (**Fig. S1**). Together with LQ, these metrics reliably distinguish neurons from non-neuronal samples.

3.1.4 Gene ontology (GO) analysis

Gene Ontology (GO) analysis allows understanding of the overall differences of gene

expression among multiple samples. This is accomplished by utilizing the GO database. The GO database contains information that links genes to GO terms. These GO terms can be broad, encompassing a large variety of biological functions, or narrow, including only a few specific functions. This allows the overall expression patterns of the genes to be categorized and then quantitatively measured. The hierarchies of GO terms provide an overview of the intersection of genes to biological functions. This intersection provides an insight into the different cellular mechanisms by simplifying the data. One problem with using GO analysis is that as new information is discovered the genes for GO terms can be changed. Using an up-to-date database of GO terms can be used to avoid this problem but may lead to different results over time. Since the GO terms are determined by previous knowledge it can limit the potential to discover new features of genes.

The LONGO package handles GO analysis by utilizing topGO (Alexa and Rahnenfuhrer, 2016). LONGO allows multiple parameters in the GO analysis step. The main two are the statistical test and the method for graphing. The statistical test is used to determine the significance of the biological functions and the method for graphing is used to determine how the significant nodes should be graphed.

3.1.5 Cellular reprogramming and analyses

Human adult dermal fibroblasts from healthy individuals were acquired from the Coriell Institute for Medical Research: ND34769 (female, 68 years old at sampling; WT4) and AG04148 (male, 56 years old; WT2). Cells were reprogrammed as described in Victor et al. (2014) using a lentiviral cocktail of rtTA, pTight-9-124-BclxL, CTIP2, MYT1L, DLX1, and DLX2.

Immunocytochemistry was also performed as described in Victor et al. (2014) using primary antibody of rabbit anti- β -III tubulin (BioLegend, 1:2,000) and secondary antibody of anti-rabbit

IgG conjugated with Alexa-488 (Invitrogen, 1:1000). Images were captured using a Leica SP5X white light laser confocal system with Leica Application Suite Advanced Fluorescence 2.7.3.9723. RNA-seq raw data was recently published (Victor et al., 2018). Briefly, RNA was extracted from converted neurons and isolated with TRIzol reagent (Thermo Fisher Scientific) per manufacturer's instructions. After treating samples with Ribo-Zero kit (Illumina), cDNA library was sequenced in Illumina HiSeq 2500. Sequence reads were aligned to the human genome (hg38) with STAR v2.4.2a. Gene counts were derived from the number of uniquely aligned unambiguous reads by Subread:featureCount, version 1.4.6, with GENCODE gene annotation (V23).

3.2 Results

3.2.1 LGE identifies neurons upon differentiation and maturation during development

To demonstrate LGE as an indicator of neuronal identity, we first analyzed RNA-seq gene expression profiles at the tissue level from the Rat BodyMap database (Yu, et al., 2014), which profiled 32 rats across 10 different organs (i.e. adrenal gland, brain, heart, kidney, liver, lung, muscle, spleen, thymus, and testis or uterus), using LONGO. Consistent with previous reports of LGE in brain tissues (Gabel, et al., 2015), we found that LONGO clearly distinguishes the brain from all other non-neuronal tissues with an LQ of 0.31 (**Fig. 1b**). Additionally, by measuring the collective levels of long gene expression, neuronal samples become readily identifiable in a population of non-neuronal samples without depending on the individual expression levels of known neuronal markers (**Fig. S1**).

To further evaluate LGE in assessing post-mitotic neurons during neural development, we used LONGO to analyze transcriptome datasets collected from distinct regions of the

developing human cortex (Miller, et al., 2014). Interestingly, LGE analysis distinguished cortical and subcortical regions (i.e. intermediate zone, subcortical plate, cortical plate, subpial granular zone, and marginal zone; max LQ of 0.55) from other zones in which proliferative neural progenitors are prominent (i.e. ganglionic eminences, ventricular and sub-ventricular zones; max LQ of 0.13) (**Fig. 2a**). Our finding is consistent with the neuronal populations observed in cortical and subcortical regions (Miller, et al., 2014), and the increase in LGE likely reflects the number of differentiated neurons within each layer. To further pinpoint the cellular source of differential LGE, we analyzed LGE in different cell types isolated by fluorescence activated cell sorting (FACS) from the cortex (Zhang, et al., 2016), and found that only neurons displayed significantly enriched LGE in contrast to other cell types including astrocytes and oligodendrocytes, suggesting that increased LGE in whole cortex is likely due to neurons (**Fig. S2a**). Similarly, LGE analysis of single-cell RNA-seq of mouse visual cortex (Tasic, et al., 2016) identified increased LGE and LQ only for neurons, further validating LGE as a unique feature of neurons (**Fig. 2b**). Furthermore, by applying LONGO to data generated by Habib *et al.* (Habib, et al., 2016), where single-nucleus RNA-seq was combined with pulse-labeling of proliferating cells by EdU to track transcriptional dynamics of newborn neurons within the neurogenic niche of the adult hippocampus, we found that LGE increased as neural progenitor cells (NPCs) exited the cell cycle and continued to differentiate into neurons and mature *in vivo* (**Fig. 2c**), perfectly matching the maturation trajectory determined by Habib *et al.*, and validating LGE as a reliable marker for post-mitotic mature neurons. In order to address whether increased LGE is also a feature of cultured neurons, we applied LONGO to the transcriptome data of motor neurons differentiated from mouse embryonic stem cells (ESCs) (Mahony, et al., 2011). Although minor differences in LGE were observed between embryoid body formation and induction of NPCs by

retinoic acid, the largest increase in LGE occurred when progenitors differentiated into post-mitotic motor neurons (**Fig. S2b**). This finding is consistent with the notion that onset of LGE occurs during differentiation of NPCs to neurons, which is also apparent in tissue culture conditions. Together, our findings demonstrate LGE as a hallmark of neuronal development assayed at the single-cell and population levels.

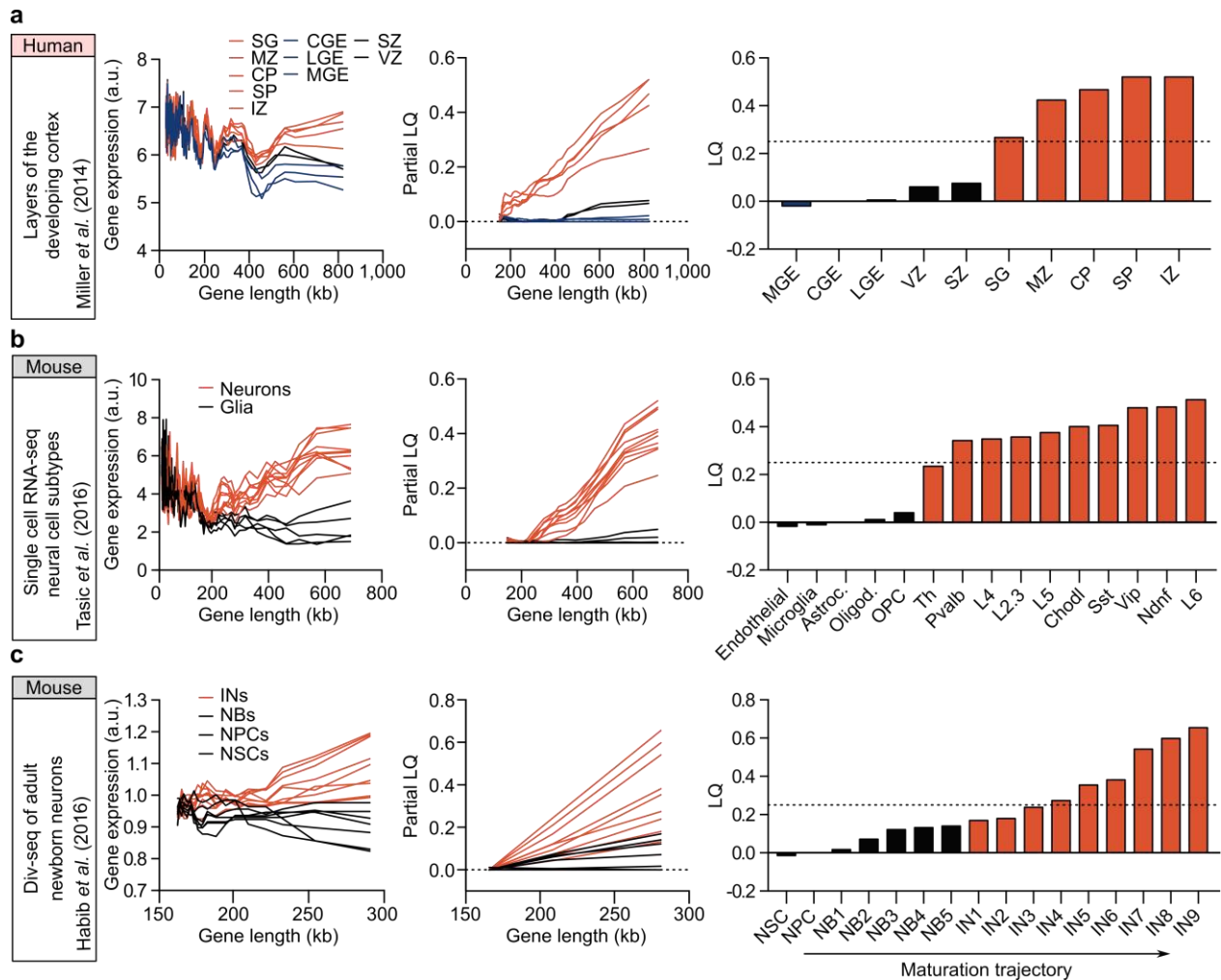


Fig. 2. LONGO output of neuronal differentiation and maturation during development.

From left to right, reference for input data, rolling media of gene expression versus length (200 gene bins, 40 gene step), partial LQ, and final LQ. **(a)** Layers of the developing human cortex (Miller, et al., 2014): ganglionic eminences (MGE, LGE, CGE; blue); ventricular zones (VZ and SVZ; black); and post-mitotic zones (SG, MZ, CP, SP, and IZ; red). **(b)** scRNA-seq neural cell subtypes from mouse visual cortex (Tasic, et al., 2016): glial cells (endothelial cells, microglia, astrocytes, oligodendrocytes, and OPCs) shown in black; neuronal subtypes (Th, Pvalb, L4, L2.3, L5, Chodl, Sst, Vip, Ndnf, and L6) shown in red. **(c)** Div-seq of adult newborn neurons in the neurogenic niche of mouse hippo-campus (Habib, et al., 2016): neural stem cells (NSCs), neural progenitor cells (NPCs), and neuroblasts (NBs) shown in black; immature neurons (IN) shown in red. Samples are numbered according to their maturity (Habib, et al., 2016).

3.2.2 LGE identifies neurons successfully reprogrammed from non-neuronal cells

Based on the authenticity of LGE as a transcriptomic feature of neurons both *in vivo* and in tissue culture, we applied LONGO to evaluate cells generated by neuronal conversion using datasets from transcription factor-based (Colasante, et al., 2015; Treutlein, et al., 2016), microRNA-based (Victor, et al., 2014), and small molecule-based (Hu, et al., 2015) approaches to convert mouse or human fibroblasts to neurons. Because we routinely observe LQ exceeding 0.25 in neurons (**Fig. 2 and Fig. S3**), we use this value as the threshold to indicate successful neuronal conversion. First, we used a transcriptome dataset in which neuronal conversion of mouse embryonic fibroblasts (MEFs) was monitored at a single cell level (Treutlein, et al., 2016) and applied LONGO to analyze differential LGE between successfully reprogrammed cells and cells that failed to convert (**Fig. 3a**). Only successfully converted neurons showed increased LGE and an LQ value substantially above 0.25, validating LGE analysis as a reliable approach to assess neuronal conversion. To test microRNA-based neuronal conversion (Victor, et al., 2014), we prepared RNA from an unpurified population of human striatal medium spiny neurons (MSNs) converted from fibroblasts and performed RNA-seq. When we applied LONGO to converted MSNs, we observed a dramatic increase in LGE similar to ESC-derived human neurons (**Fig. 3b**), despite that samples were unpurified. Furthermore, we find that LONGO detects variable conversion efficiencies inherent in different fibroblast samples—one fibroblast cell line (HAF2) that we previously found to display a lower conversion efficiency (**Fig. S4**) yielded an average LQ barely reaching 0.25, whereas the other

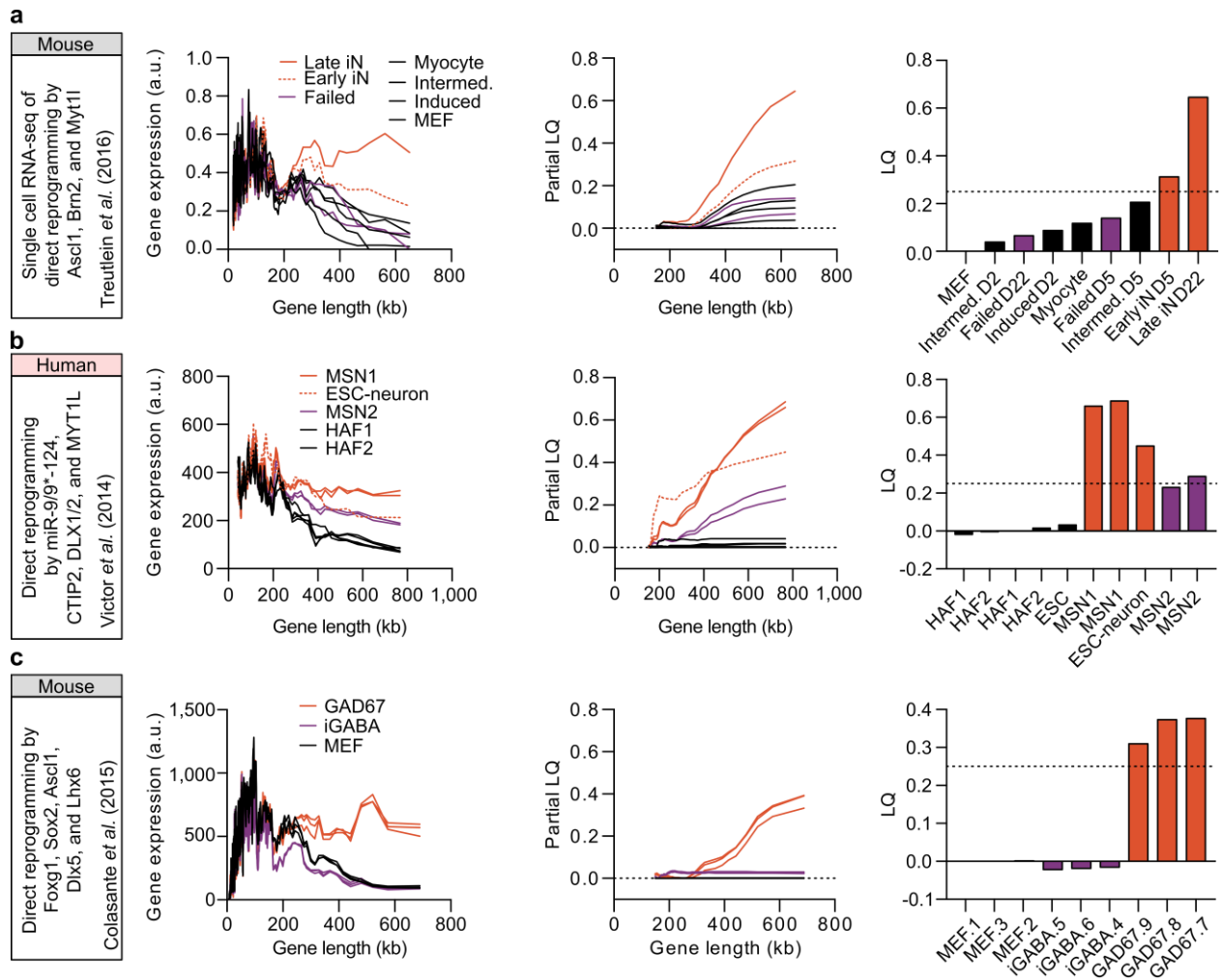


Fig. 3. LONGO output of somatic cell direct neuronal reprogramming.

From left to right, reference for input data, rolling median of gene expression versus length (200 gene bins, 40 gene step), partial LQ, and final LQ. **(a)** Direct reprogramming of mouse embryonic fibroblasts (MEFs) to neurons by transcription factors (Treutlein, et al., 2016): MEFs, myocytes, intermediate reprogramming (D2 and D5), and induced MEFs (D2) shown in black; failed reprogramming (D5 and D22) shown in purple; early (D5) and late (D22; dotted-line) successful reprogramming in red. **(b)** Direct reprogramming of human adult fibroblasts (HAFs) to striatal medium spiny neurons (MSNs) by microRNAs: HAFs from two different cell lines (black); MSNs derived from HAF2 (MSN2; purple); ESC-derived neurons (Gill, et al., 2016) (dotted-line) and MSNs derived from HAF1 (MSN1) shown in red. **(c)** Direct reprogramming of MEFs to iGABA by *Foxg1*, *Sox2*, *Ascl1*, *Dlx5*, and *Lhx6* (Colasante, et al., 2015): MEFs (black); iGABA (dark purple); GAD67+ interneurons (red).

fibroblast cell line (HAF1) yielded an average LQ of 0.67 (**Fig. 3b**).

Analysis of direct neuronal reprogramming of human fibroblasts by small molecules (Hu, et al., 2015) also revealed increased LGE approaching the levels obtained by ESC-derived neurons by prolonged treatment with small molecules (**Fig. S5a**). However, knockdown of PTBP1, which reportedly generates neuronal-like cells from HAFs (Xue, et al., 2016), does not increase LGE at the population level (**Fig. S5b**). Under these conditions, Xue *et al.* reported little expression of neuronal markers, such as MAP2 and NeuN, and the absence of neuronal electrical activity (Xue, et al., 2016), which is consistent with our finding of unaltered LGE. It will be interesting to apply LGE analysis to sequential knockdown of PTBP1 and PTBP2, which reportedly generated more functional neurons (Xue, et al., 2016). Finally, we analyzed RNA-seq data from MEFs overexpressing neuronal transcription factors Foxg1, Sox2, Ascl1, Dlx5, and Lhx6, which has been previously reported to generate GABAergic interneurons (Colasante, et al., 2015). Cells were purified by FACS based on the expression of GAD67, a marker for GABAergic interneurons, but were not electrically active at this time-point (Colassante *et al.* reported that cells became electrically active only after 4 weeks of co-culture with rat hippocampal neurons). Supporting LGE as a marker for mature neurons, we only detected increased LGE for control interneurons (GAD67), not for induced GABAergic interneurons (iGABA) in monoculture (**Fig. 3c**).

3.2.3 GO analysis of neuronal differentiation

GO enrichment analysis of mouse ESC differentiation to motor neurons in vitro (Mahony, et al., 2011) is shown in **Fig. 4**. The experiment has multiple GO terms identified as being significant. Rectangle color represents the relative significance, ranging from dark red (most significant) to bright yellow (least significant). The two lines show the GO identifier and a

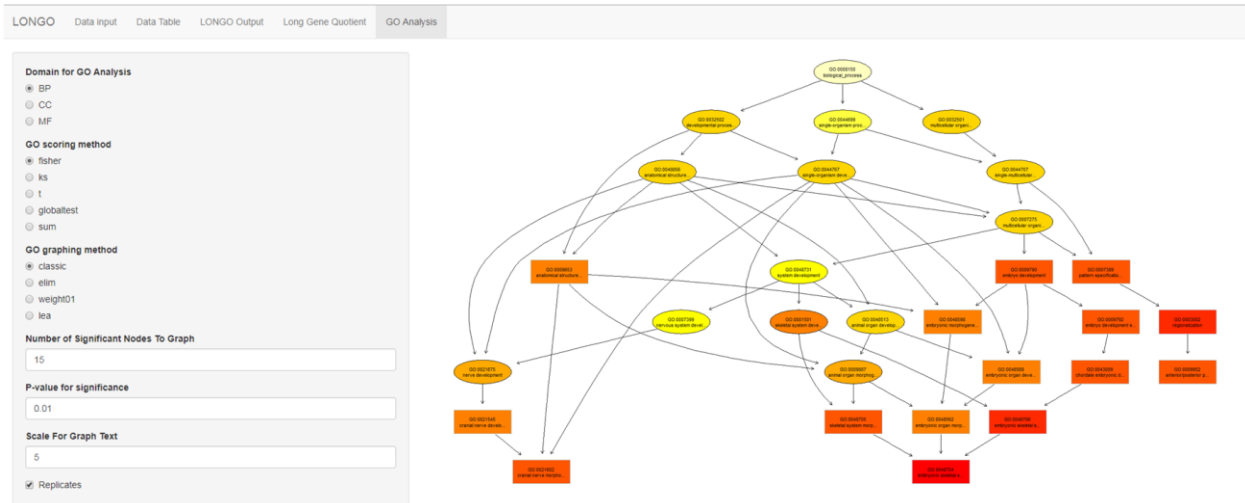


Fig. 4. LONGO interactive GO analysis snapshot using incorporated topGO R package.

Enrichment analysis of mouse ESC differentiation to motor neurons *in vitro* (Mahony, et al., 2011). Rectangle color represents the relative significance, ranging from dark red (most significant) to bright yellow (least significant). The two lines show the GO identifier and a trimmed GO name.

trimmed GO name.

The two categories identified are embryogenesis and neurogenesis. These results are in line with the source of the data set that was an experimental differentiation of embryonic stem cells to neurons. As the cells differentiate, the genes they express are altered providing the enriched GO terms.

3.3 Supplemental Figures

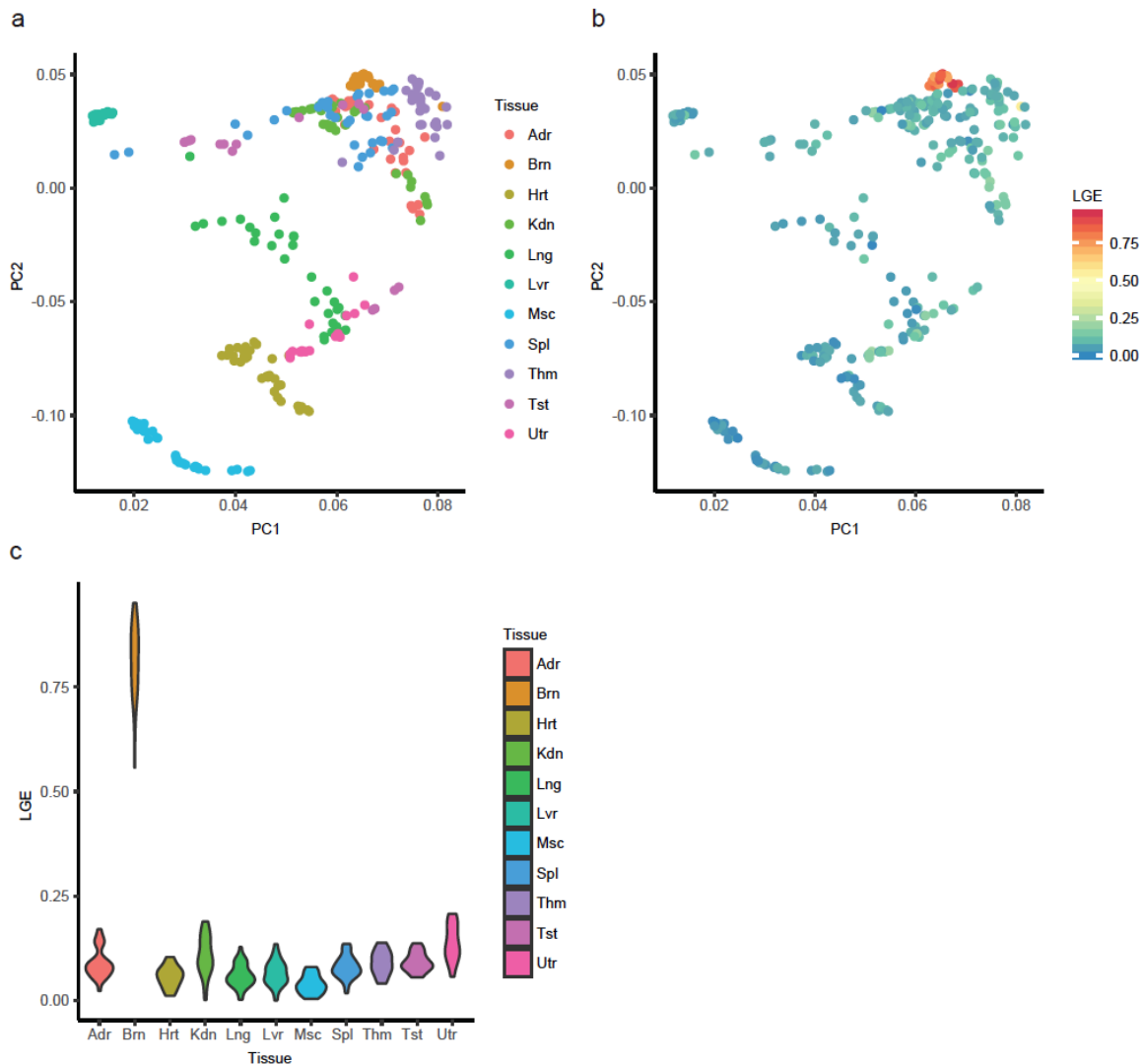


Fig. S1. LGE identifies neuronal tissues, while PCA only separates tissues.

(a) Principle components 1 and 2 of Rat BodyMap principle component analysis (PCA), with each sample colored by tissue of origin. (b) The same PCA as in (a), but now each sample is colored by the scaled correlation of binned gene expression (200 genes per bin; 40 genes per step) versus median gene length, here termed LGE. (c) Violin plot of the LGE values plotted in (b) for each Tissue.

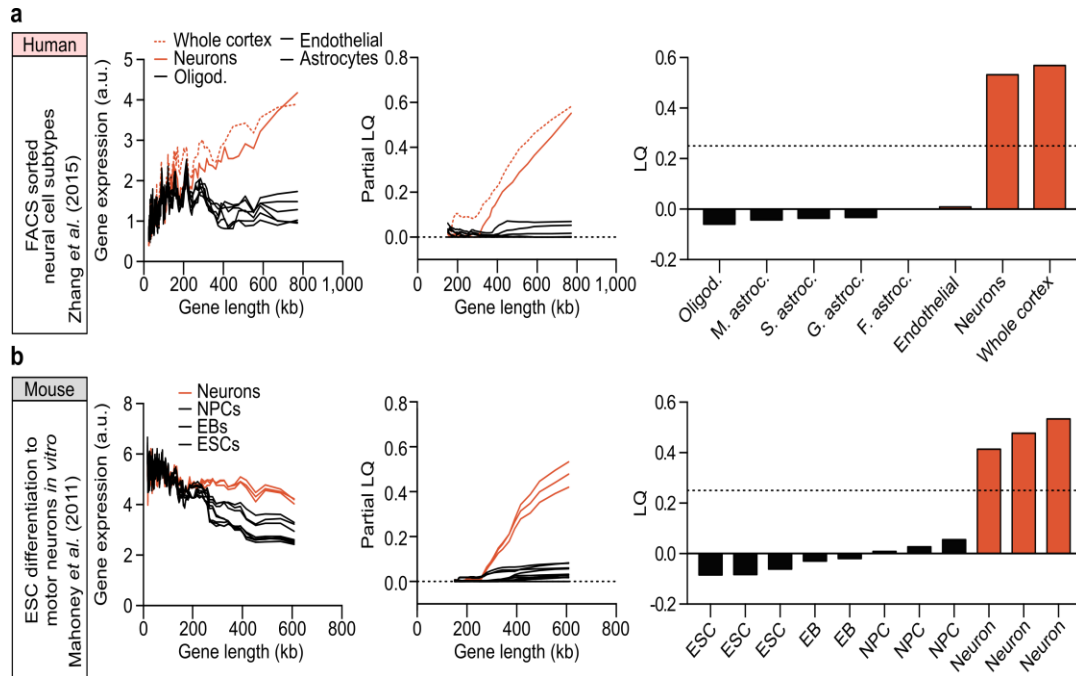


Fig. S2. LONGO output of neuronal differentiation in vivo and in vitro.

From left to right, reference for input data, rolling media of gene expression versus length (200 gene bins, 40 gene step), partial LQ, and final LQ. (a) FACS sorted human neural cell subtypes (Zhang, et al., 2016): glial cells (oligodendrocytes, endothelial cells, S. astrocytes (sclerotic hippocampi astrocytes), G. astrocytes (GBM peri tumor astrocytes), F. astrocytes (fetal astrocytes), and M. astrocytes (mature astrocytes) shown in black; neurons and whole cortex (dotted-line) shown in red. (b) Mouse ESC differentiation to motor neurons *in vitro* (Mahoney, et al., 2011): ESCs (D0), embryoid bodies (EBs; D2), and NPCs (D4) shown in black; motor neurons (D7) shown in red. In all analyzed data, LQ typically exceeds 0.25 in neurons.

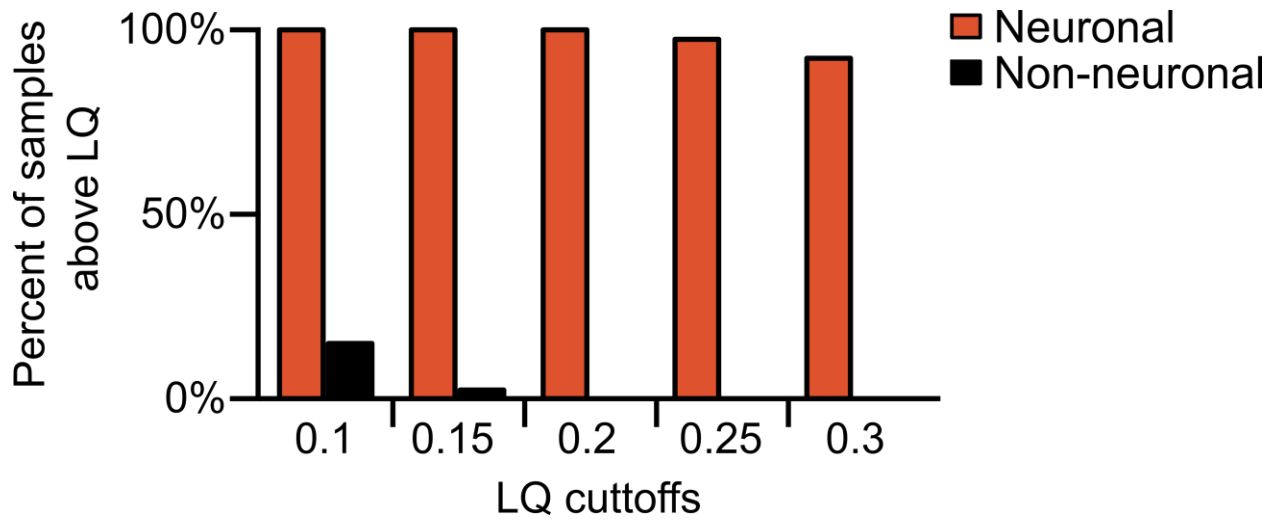


Fig. S3. LQ cutoff of 0.25 stringently distinguishes neurons from non-neuronal cells.

Percentage of samples that pass various LQ cutoffs from LQ data in Figures 2b,2c, and 2e for neuronal cells (red bar; n = 39) and non-neuronal cells (black bar; n = 40).

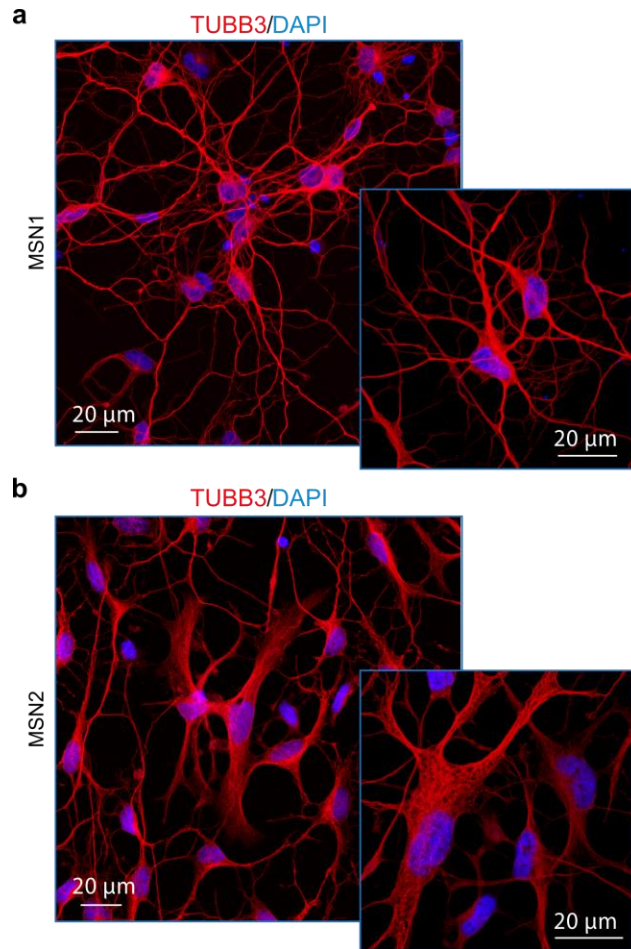


Fig. S4. Medium spiny neurons reprogrammed with miR-9/9*-124+CDM from fibroblasts of two unrelated adults shows variation in the proper acquisition of neuronal morphological features.

(a) MSN1: Female, 68 years old at sampling - Coriell: ND34769 and (b) MSN2: Male, 56 years old - Coriell: AG04148, immunostained for TUBB3 in red and the nuclear stain DAPI in blue, at 28 days post microRNA-based neuronal induction. In contrast to cells from HAF1 (MSN1), a large fraction of reprogrammed cells from HAF2 (MSN2) retains fibroblastic features such as expanded cell bodies and a large nucleus and have limited dendritic trees. Insets for each cell population display morphological characteristics in greater magnification. Our results demonstrate that TUBB3 immunostaining, a commonly used ‘neuronal marker’, does not reliably distinguish neurons from partially reprogrammed cells.

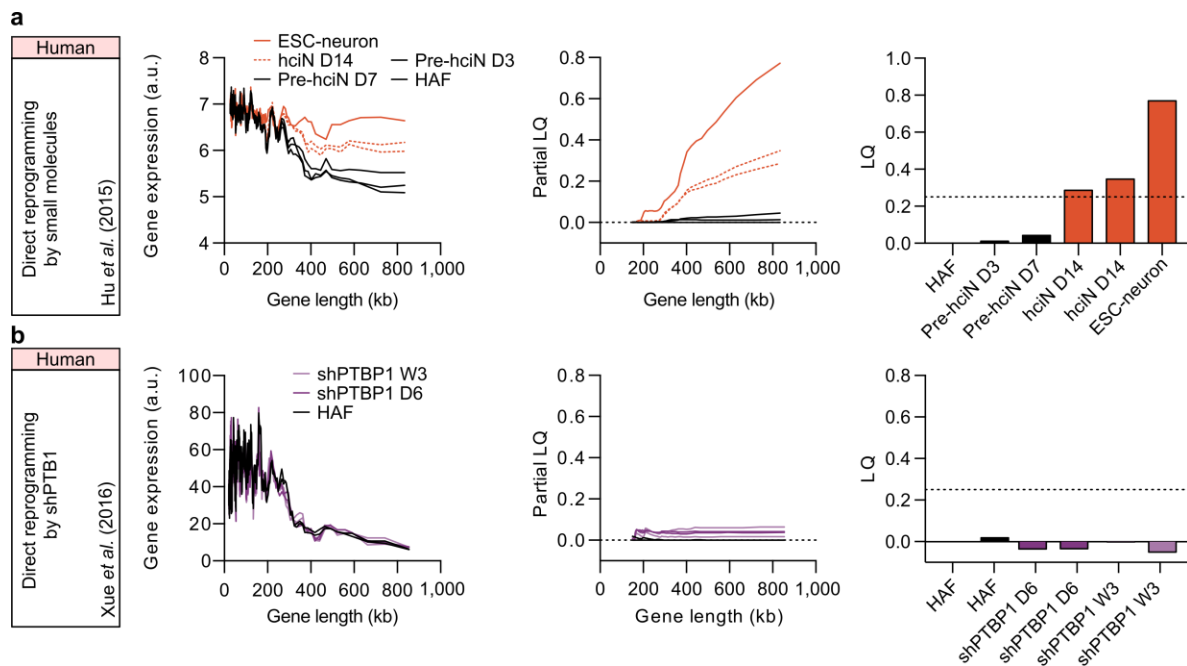


Fig. S5. LONGO output of small molecule and shPTBP1-induced reprogramming.

From left to right, reference for input data, rolling median of gene expression versus length (200 gene bins, 40 gene step), partial LQ, and final LQ. (a) Direct reprogramming of HAFs to neurons (Pre-hciN) by small molecules (Hu, et al., 2015): HAFs, Pre-hciN D3, and Pre-hciN D7 (dotted-line) shown in black; hciN D14 and ESC-derived neurons (dotted-line) shown in red. (b) Direct reprogramming of HAFs to neuron-like cells by shPTBP1 (Xue, et al., 2016): HAFs (black); shPTBP1 for 6 days (dark purple); shPTBP1 for 3 weeks (light purple).

3.4 Author Contributions

M.J.M. and A.S.Y. conceived the concept for this study. A.J.P., M.J.M. and T.-H.A. designed and constructed the LONGO package. H.G., M.J.M., and T.-H.A. conceived the long gene quotient (LQ). M.J.M. acquired and analyzed all public data. M.R. generated miRNA-mediated reprogramming data, and M.J.M. analyzed the data. M.B.V. generated and analyzed immunostaining data, as well as editing. A.J.P. and T.-H.A. incorporated GO analyses into the LONGO package, and managed R package curation. H.W.G. provided counsel throughout analyses. H.G. and T.-H.A. supervised all statistical analyses and A.S.Y. supervised the overall project. A.S.Y., M.J.M, A.J.P., and T.-H.A. wrote and all authors approved the manuscript.

3.5 Acknowledgements

We thank A. Herbert for helpful discussions and editing.

3.6 Funding

M.J.M. is supported by the NIH-funded Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Predoctoral Fellowship (T32GM081739; Barch, PI). A.S.Y. is supported by NIH Director's Innovator Award (DP2NS083372), Missouri Spinal Cord Injury/Disease Research Program (SCIDRP), Cure Alzheimer's Fund and Presidential Early Career Award for Scientists and Engineers. T.-H.A is supported by NSF CRII-1566292 and Saint Louis University President's Research Fund (PRF).

3.7 Data Availability

BioSamples: ND34769 HAF1 (SAMN05340933), ND34769 MSN1 (SAMN05340929), AG04148 HAF2 (SAMN05340931), AG04148 MSN2 (SAMN05340927).

GEO database (Edgar, et al., 2002): Mazzoni et al., 2011 (GSE19372), Wang et al., 2012 (GSE62936), Treutlein et al., 2016 (GSE67310), Hu et al., 2015 (GSE69480), Zhang et al., 2016 (GSE73721), Xue et al., 2016 (GSE77523), Colasante et al., 2015 (GSE74065).

Online resources: Tasic et al., 2016 (http://casestudies.brain-map.org/celltax/download_archive), Miller et al., 2014 (http://www.brainspan.org/api/v2/well_known_file_download/278444094), Habib et al., 2016 (http://science.sciencemag.org/highwire/filestream/682212/field_highwire_adjunct_files/5/Table_S5.xlsx)

Chapter 4: Single-cell Sequencing Reveals Transcriptomic Features of Mature Neurons

Adapted from:

Single-cell Sequencing Reveals Transcriptomic Features of Mature Neurons

Matthew J. McCoy*, Daniel G. Abernathy*, Harrison W. Gabel, Samantha A. Morris,
Andrew S. Yoo

*Authors contributed equally

Manuscript in preparation

4.1 Results

4.1.1 Single-cell RNA-seq of microRNA-mediated reprogramming reveals a homogenous reprogramming trajectory

To investigate the single-cell dynamics of miRNA-mediated neuronal reprogramming of human adult fibroblasts, we performed scRNA-seq along a time-series of reprogramming (Fig. 1A and B). Of the 16,828 cells that passed filtering during analysis, most cluster by time point (Fig. 1C). As observed in other scRNA-seq experiments of reprogramming (Treutlein, et al., 2016; Liu, et al., 2017), starting fibroblasts were heterogeneous, dividing into several distinct clusters (Fig. 1D and E). Remarkably, after day 10, virtually all cells converged onto a single reprogramming trajectory (Fig. 1D and E), which is consistent with previous reports of the high efficiency of miR-mediated reprogramming (Victor, et al., 2014; Abernathy, et al., 2017). Using quadratic programming, a comparison of these results with bulk transcriptome data from the same reprogramming conditions after day 35 (Abernathy, et al., 2017) corroborated a smooth and continuous reprogramming trajectory over time (Fig. 1F and G). Additionally, scRNA-seq identified common markers of fibroblasts and neurons (Fig. 1H and I), which have previously been demonstrated to decrease and increase, respectively, during reprogramming (Abernathy, et al., 2017). Together, these results demonstrate that miR-mediated direct reprogramming of human cells into neurons is highly efficient, with heterogeneous starting fibroblasts converging onto a homogenous reprogramming trajectory.

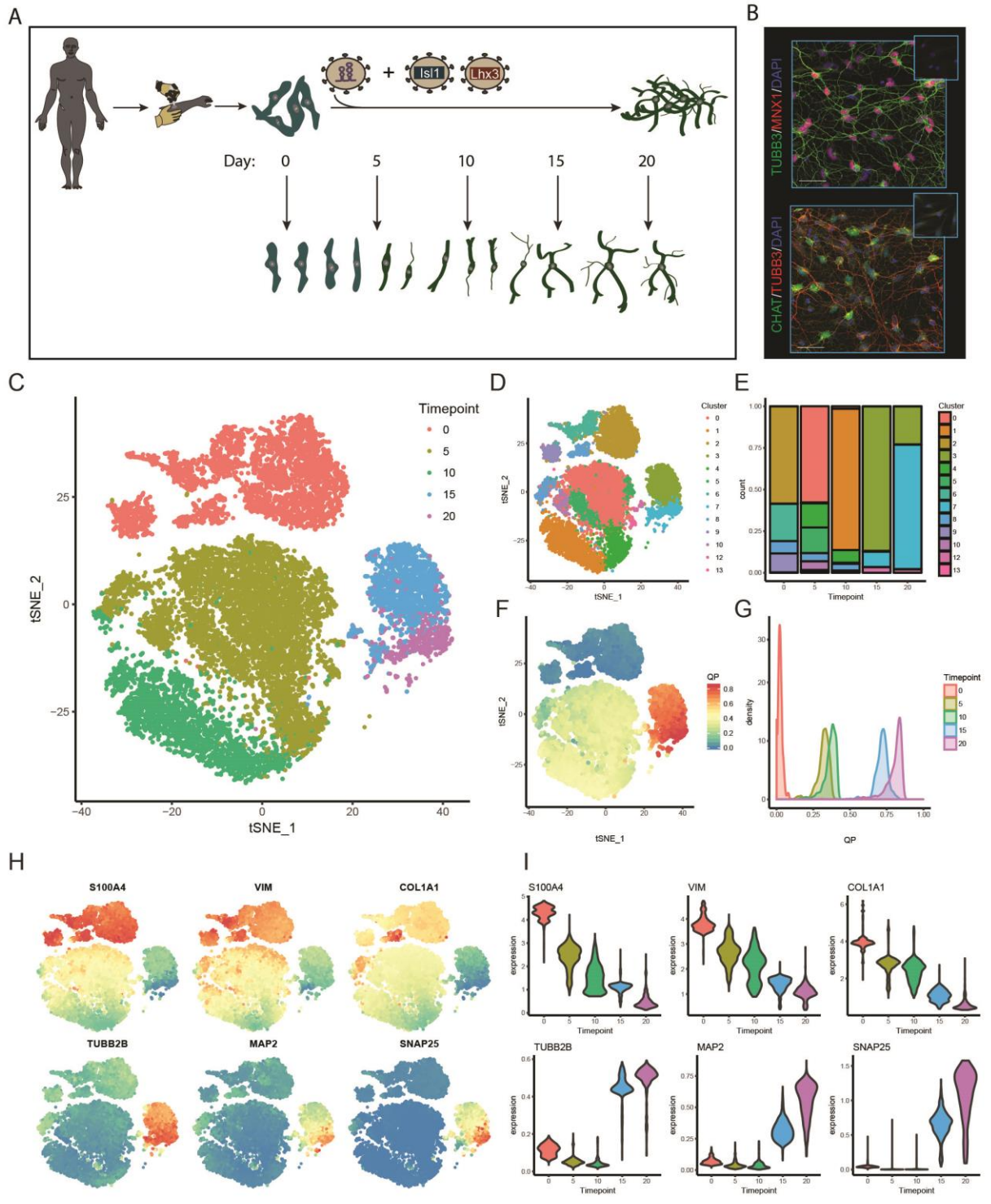


Fig. 1. Direct conversion of human adult fibroblasts yields a homogenous population of neurons.

(A) Schematic of microRNA-mediated motor neuron reprogramming and sample collection. Starting human adult fibroblasts (Day 0) were collected, followed by periodic collection of cells

after induction of reprogramming (Day 5, 10, 15, and 20). **(B)** Adult human fibroblasts ectopically expressing miR-9/9*-124 with ILS1 and LHX3 for 35 days immunostained for the early pan-neuronal marker TUBB3, and motor neurons markers MNX1 and CHAT. The insets represent starting fibroblasts co-stained as negative Controls. Scale bars, 20 μ m. **(C to I)** scRNA-seq analysis of reprogramming. **(C)** *t*-SNE projection of single cells, colored according to time point. Total cells at each time point: 4360 (Day 0), 7083 (Day 5), 3031 (Day 10), 1760 (Day 15), and 585 (Day 20). **(D)** *t*-SNE projection of single cells, colored by unsupervised clustering based on specific and distinct gene expression profiles. **(E)** Bar plot of cells per cluster per time point as a fraction of the total number of cells at each time point. **(F)** *t*-SNE projection of single cells, colored by quadratic programming scores based on overall similarity to day 35 bulk transcriptome analysis of reprogrammed motor neurons (raw data from Abernathy, et al., 2017). **(G)** Density plot of quadratic programming scores, colored by time point. **(H)** *t*-SNE projection of single cells (same axes as in previous *t*-SNE plots), colored by expression of fibroblast marker genes (top) and neuron marker genes (bottom). **(I)** Violin plot quantification of gene expression from **(H)**.

4.1.2 Long gene expression distinguishes mature from immature neuronal gene expression

Previously, we reported on the extensive epigenetic and transcriptomic remodeling that occurs early during neuronal reprogramming, which broadly reflects aspects of neurogenesis *in vivo* (Abernathy, et al., 2017). Here, we characterized the single-cell dynamics of neuronal maturation during the late stages of reprogramming. In part because single-cell RNA-seq suffers from high rates of drop-out resulting in unreliable measurement of individual gene expression, we turned to an analysis we previously developed that measures the collective expression of the top 10% of the longest genes in the genome, which we have termed Long Gene Expression [LGE (McCoy, et al., in press)]. Since LGE is specific to neurons and increases with the time since the last cellular division as a neural progenitor *in vivo* (McCoy, et al., in press), we sought to determine if LGE could be used to assess the maturity of reprogrammed neurons. We found that LGE could be detected at the single-cell level and increased over time during reprogramming (Fig. 2A-C). Additionally, the onset of LGE mirrored activation of mature neuronal markers (e.g. MAP2 and SNAP25) and occurred after the transient upregulation of genes typically enriched in late neural progenitors and immature neurons *in vivo* (e.g. GFAP and NES; Fig. 2D and E). LGE scores correlated with neuronal marker (e.g. MAP2) expression and anticorrelated with late progenitor and immature neuronal marker (e.g. GFAP) expression. (Fig. 2F). Expression of neuronal BAF (nBAF) chromatin remodeling complex subunits (e.g. ACTL6B/BAF53b and DPF3/BAF45c) were also observed in conjunction with the downregulation of the neural progenitor BAF (npBAF) complex subunits (e.g. ACTL6A/BAF53a and PHF10/BAF45a; Fig. 2D and E; fig. S1). This same subunit switching is observed *in vivo* upon differentiation of neurons from neural progenitor cells, a process which, significantly, is controlled by the

expression of miR-9/9* and miR-124 in vertebrates (Yoo, et al., 2009). This data shows that later stages of human neuronal reprogramming mirror important aspects of neuronal maturation in vivo and demonstrates that LGE is a transcriptomic feature that distinguishes mature neurons from immature neurons.

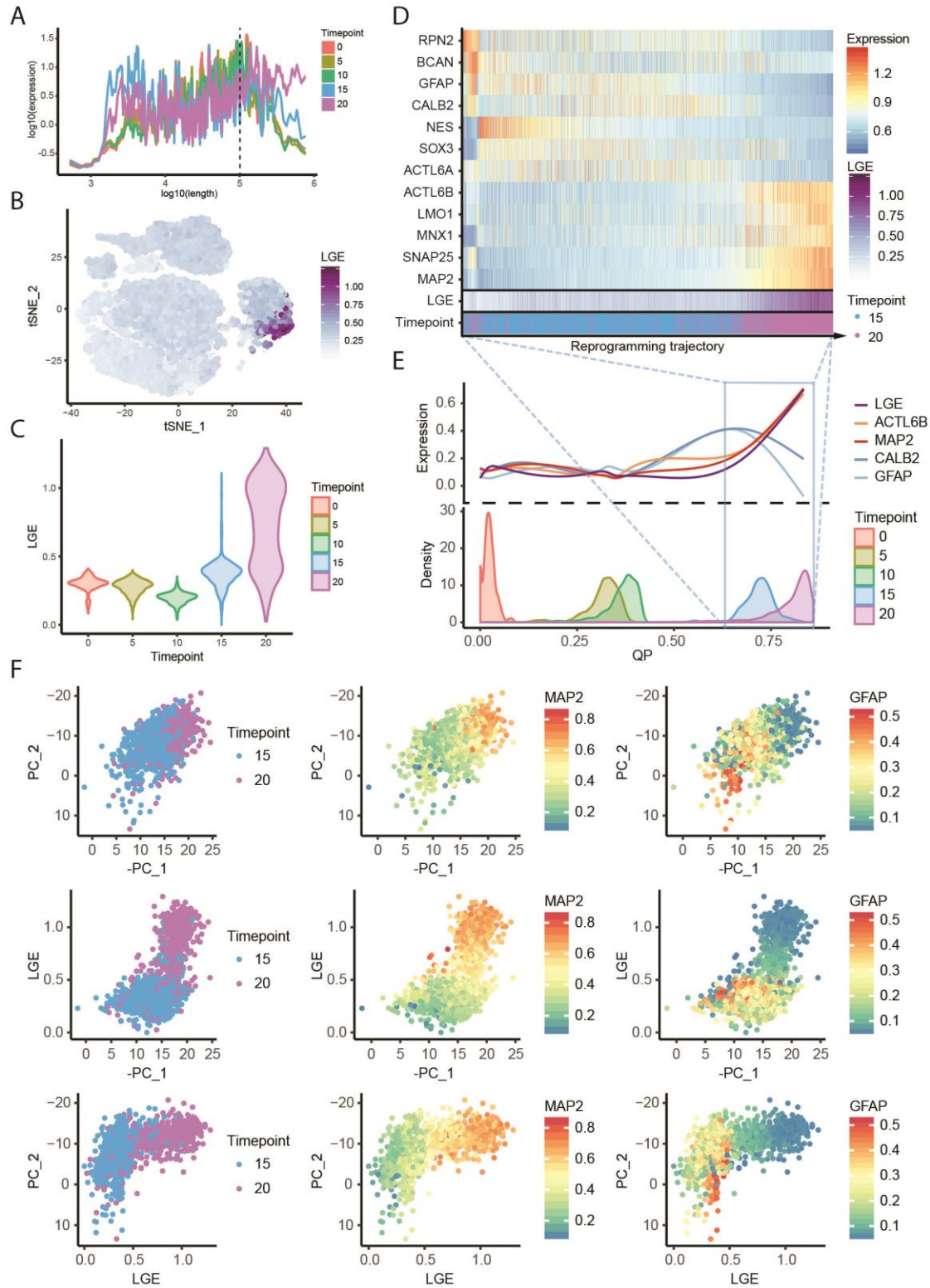


Fig. 2. Long gene expression distinguishes mature from immature neurons.

(A) Median gene expression of genes binned according to gene length (200 genes per bin, 40 genes per step). Confidence limits for the population means based on basic nonparametric bootstrapping are shown as a transparent ribbon for each time point. Vertical dashed line shows 100 kb. (B) *t*-SNE projection of single cells, colored by LGE scores (correlation of gene expression versus length of the top 10% longest genes in the human genome). (C) Violin plot of LGE scores for each time point. (D) Expression heatmap of representative late progenitor/immature neuron markers (RPN2, BCAN, GFAP, CALB2, NES, SOX3, ACTL6A/BAF53a) and mature neuron markers (ACTL6B/BAF53b, LMO1, MNX1, SNAP25, MAP2) for day 15 and day 20 cells. Cells are ordered based on quadratic programming scores, with the most neuronal cells on the right. LGE scores and time point are shown for each cell. (E) (Top) Generalized additive models for the expression of late progenitor/immature neuron markers (CALB2 and GFAP in blues), mature neuron markers (MAP2 and ACTL6B/BAF53b in reds), and LGE (purple) versus quadratic programming scores for all cells of every time point. (Bottom) Density plot of quadratic programming scores for all cells colored by time point. Light blue box indicates day 15 and 20 cells visualized in (D). (F) The first two principle components of a principle component analysis are compared against LGE using day 15 and day 20 reprogrammed neurons, with individual cells colored by time point (left column), MAP2 expression (middle column), and GFAP expression (right column).

4.1.3 BRG1 inhibition reduces LGE

BAF53b and the nBAF complex are essential for the highly conserved program of post-mitotic neural development and dendritic morphogenesis (Wu, et al., 2007; Parrish, et al., 2006). Since initiation of neuronal maturation during direct reprogramming corresponded with the switching of BAF subunits at day 20 (Fig. 2D), we predicted that functional BAF complex is necessary for neuronal maturation in reprogramming. To test this, we knocked down SMARCA4/BRG1, a core component of the BAF complex, during direct reprogramming. At day 20, we performed RNA-seq and found that LGE was significantly downregulated by BRG1 knockdown (Fig. 3). Furthermore, the switching of immature to mature gene expression patterns demonstrated in Figure 2 was not observed in BRG1 knockdown condition (fig. S2). These results are consistent with our previous findings that BRG1 knockdown reduces the number of MAP2-positive cells at day 20 of reprogramming, while also reducing chromatin accessibility of loci important for reprogramming (Abernathy, et al., 2017).

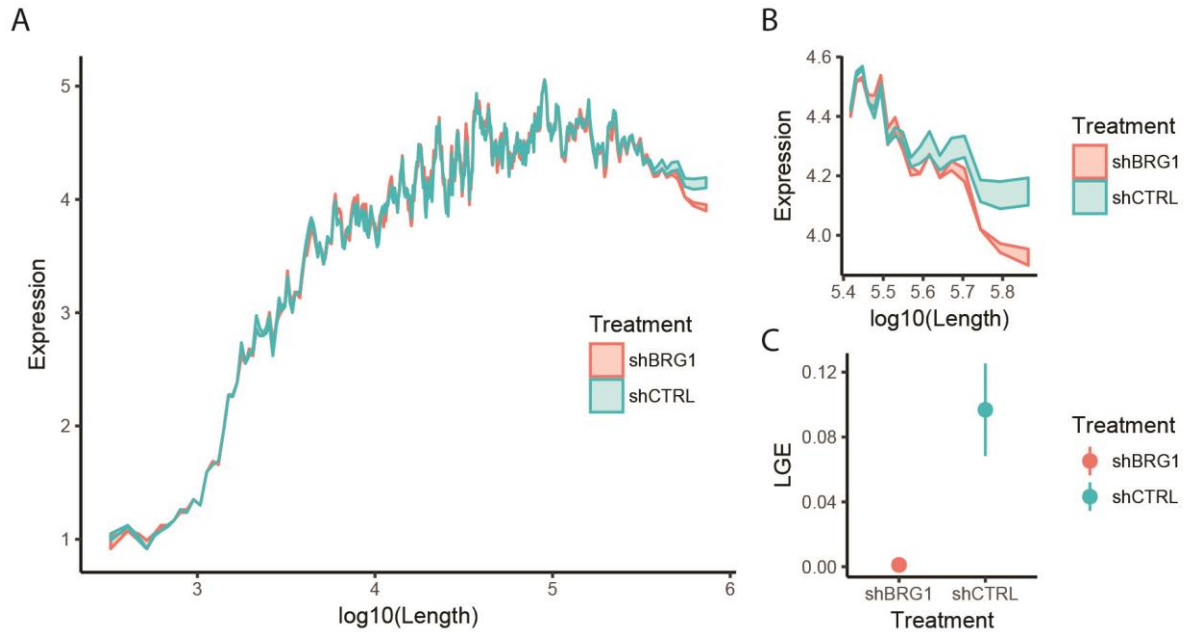


Fig. 3. Functional neuronal BAF complex is necessary for LGE.

Knockdown of BRG1 by shRNA (shBRG1; red) or control shRNA (shCTRL; blue) concurrent with microRNA-mediated reprogramming assayed by RNA-seq at day 20. (A) Median gene expression of genes binned according to gene length (200 genes per bin, 40 genes per step). Confidence limits for the population means based on basic nonparametric bootstrapping are shown as a transparent ribbon for each time point. (B) Same plot as for (A) but only with largest 10% of genes in the human genome. (C) Mean scaled correlation of gene expression versus length for top 10% of genes in the human genome. Error bars represent standard error of mean.

4.1.4 Inhibiting LGE reduces spontaneous activity

Previous studies have suggested the importance of long genes in normal neuronal development and have identified certain long genes and the pathways that regulate them (e.g. TOP1) to be mutated in several neurological disorders, such as autism spectrum disorder and Rett syndrome (King, et al., 2013; Gabel, et al., 2015). It is also known that many long genes can be artificially downregulated through inhibition of topoisomerase I (TOP1) activity, which decreases the surface expression of corresponding synaptic proteins, and thereby decreases electrical activity (Mabb, et al., 2014). Here, we treated primary rat cortical neurons with the TOP1 inhibitor, Topotecan, measured electrophysiological properties, then performed bulk RNA-seq. Consistent with previous reports, we found that treatment with Topotecan downregulated LGE and decreased spontaneous electrical activity without disrupting neuronal cell fate, as indicated by grossly normal neuronal morphology and neuronal marker expression (Fig. 4). These results are particularly interesting since the BAF complex has been implicated in recruitment of TOP1 to chromatin (Husain, et al., 2016); this association could help explain the preferential downregulation of long genes by TOP1 inhibition as well as by knockdown of BAF complex subunits.

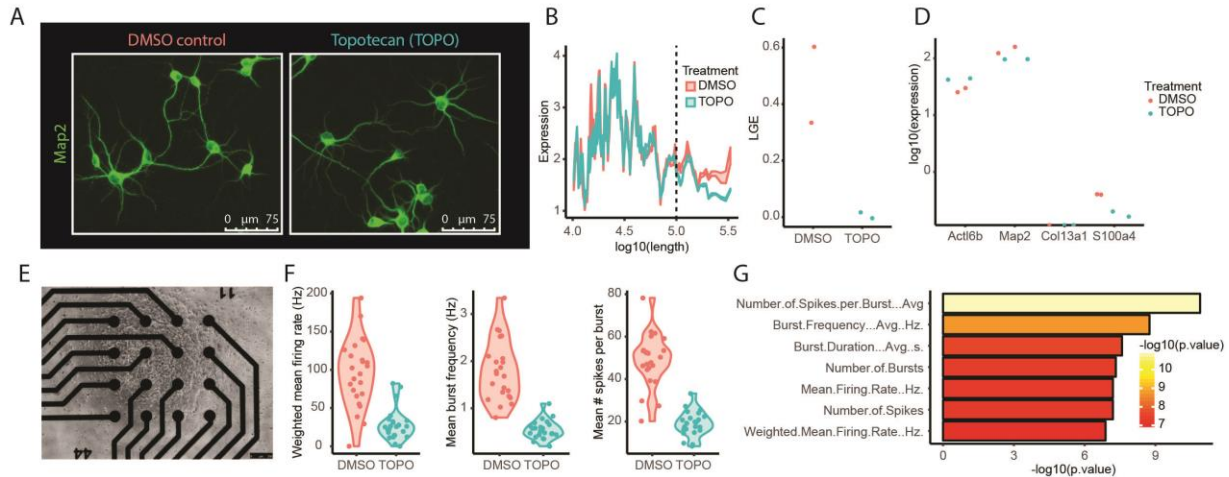


Fig. 4. Inhibition of long gene expression reduces spontaneous electrical activity of mature neurons.

(A) Map2 immunocytochemistry of primary rat cortical neurons treated with DMSO control (left) or TOP1 inhibitor, topotecan (TOPO; right) for 72 hours. (B) Median gene expression of genes binned according to length (200 genes per bin, 40 genes per step) for the top 10% longest genes in the rat genome. Top and bottom ribbon (dark lines) show confidence limits for the population means based on basic nonparametric bootstrapping. (C) Correlation values for gene expression versus length of the top 10% longest genes in the rat genome. (D) Log10 expression of two neuronal genes (*Actl6b* and *Map2*) and two non-neuronal genes (*Col13a1* and *S100a4*). (E) Phase contrast microscopy image of primary rat cortical neurons plates onto a microelectrode array (MEA). (F) Exemplary electrophysiological measurements obtained from MEA recording over a 10-minute period. (G) $-\log_{10}$ p-values for several electrophysiological measurements obtained from MEA recording.

4.2 Supplemental Figures

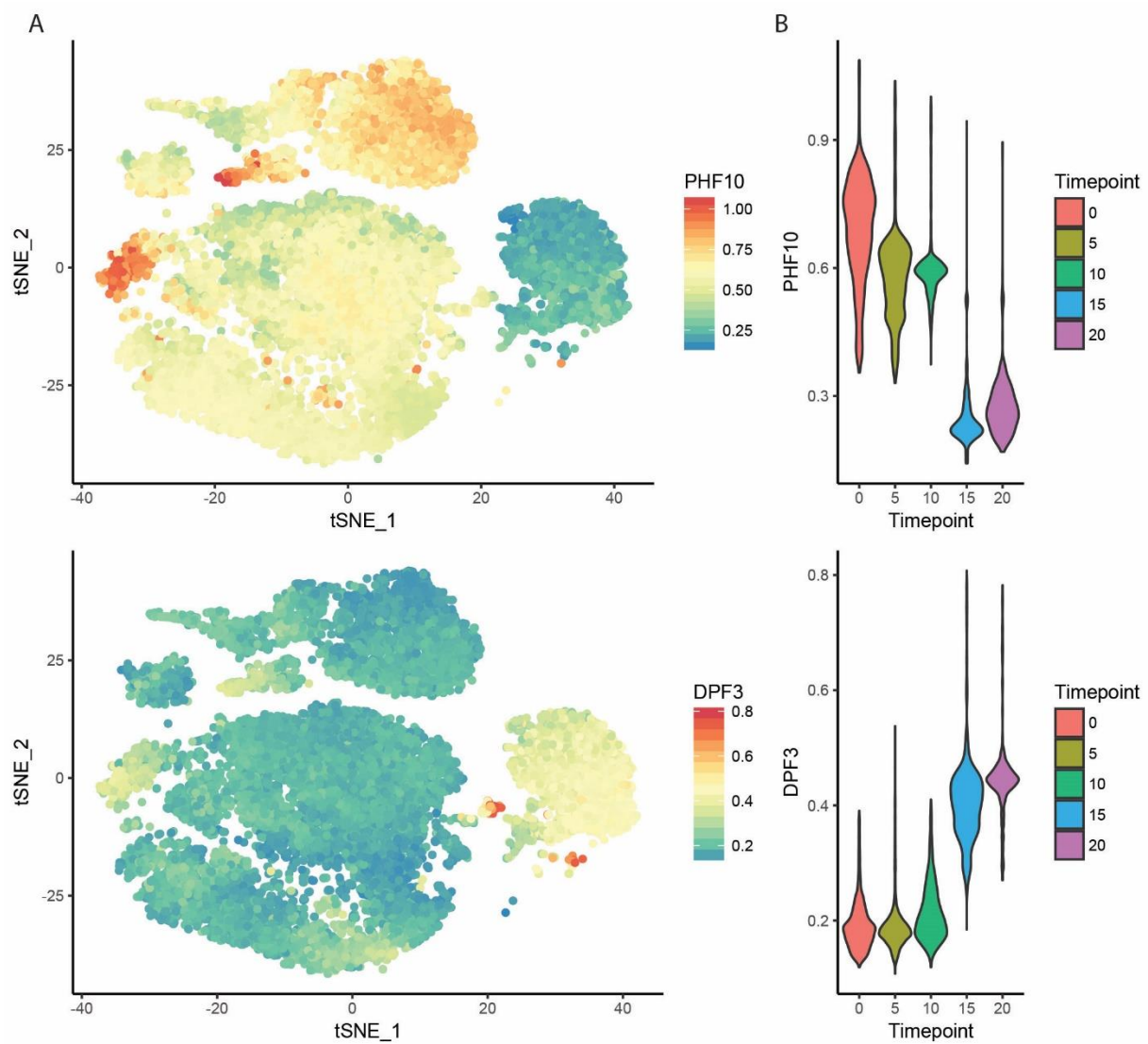


Fig. S1. PHF10/BAF45a to DPF3/BAF45c BAF complex subunit switching during neuronal reprogramming.

(A) t-SNE projection of single-cells, colored by PHF10 expression (top) and DPF3 expression (bottom). (B) Violin plots quantifying PHF10 expression (top) and DPF3 expression (bottom) at each time point during reprogramming.

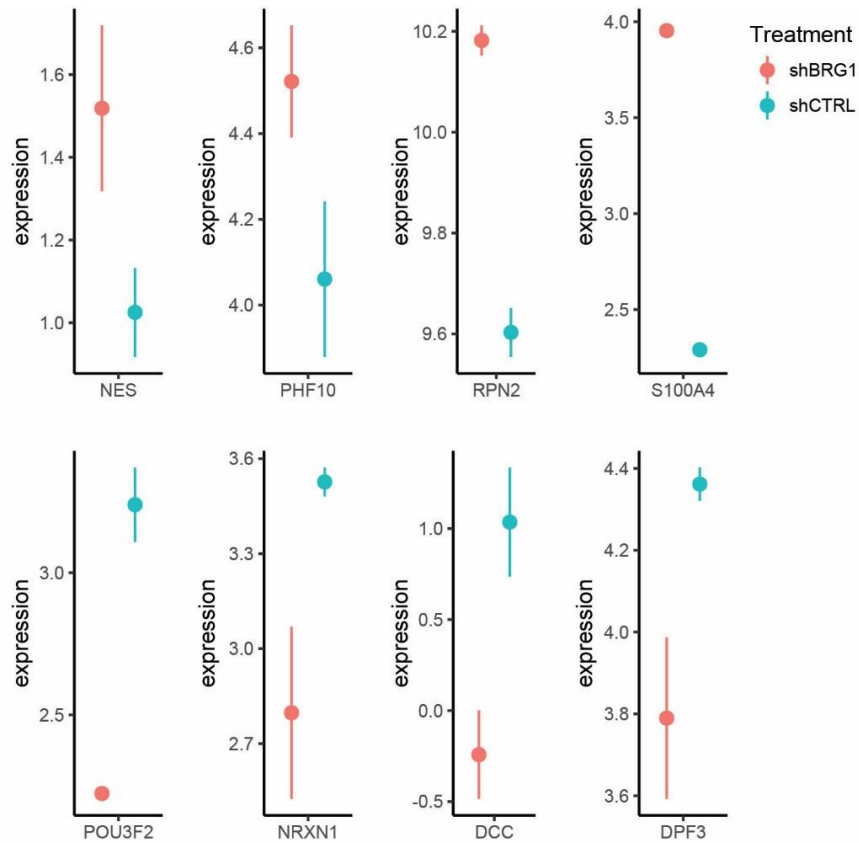


Fig. S2. Inhibition of BRG1 expression impairs maturation of reprogramming neurons.

RNA-seq analysis of microRNA-mediated reprogramming of human adult fibroblasts into neurons with Brg1 knockdown, with immature neuronal markers (NES and RPN2), neural progenitor BAF complex subunit (PHF10), fibroblast marker (S100A4), neuronal markers (POU3F2, NRXN1, and DCC), and neuronal BAF complex subunit (DPF3).

4.3 Materials and Methods

Direct Conversion and Cell Collection

Primary 22 year old female adult human fibroblasts (NIGMS, GM02171) were seeded onto Costar 6 well cell culture vessels (Corning) at a density of 300,000 cells/well. The following day, each plate was transduced with the following reprogramming cocktail: 750 uL of concentrated lentivirus containing the reverse tetracycline-controlled transactivator (rtTA; Addgene, 66810), 500 μ L of virus containing pT-BCL-9/9*-124 (Addgene, 60859), and 500 uL of each virus containing Lhx3 and Isl1 under the control of the EF1 α promoter in the presence of polybrene (8 μ g/mL ; Sigma-Aldrich) all diluted up to 18 mL with MEF media. 3 mLs of the virus/media solution was added to each well then spininfected at 37°C for 30 minutes at 1,000xG using a swinging bucket rotor.

One day post-transduction media was changed to fresh fibroblast media (2 mL per well) supplemented with doxycycline (Dox; Sigma Aldrich, 1 μ g/mL). After 2 days, fresh fibroblast media was changed and supplemented with Dox and antibiotics for respective vectors (Puromycin, 3 μ g/ml; Geneticin, 400 ug/mL; all from Invitrogen). Five days post-transduction cells were replated on to poly-ornithine/laminin/fibrobnectin (PLF) coated glass coverslips. Before PLF coating, glass coverslips were acid treated according to (Richner et al., 2015). To transfer cells, for each well of a 6 well plate, cells were first washed 2x with 1 mL sterile PBS. Then 320 μ L of 0.25% Trypsin (Gibco) was added to each well then placed in an incubator. Cells were monitored every 2 minutes, as soon as cells began to detach (no more than 6 minutes) 1 ml of MEF media supplemented with 1 μ g/mL Dox was added to each well. One by one, each well was gently triturated three times to remove remaining attached cells then transferred to a sterile

1.5 mL Eppendorf tube. Cells were then spun at 200xG for 5 minutes at 37°C. The supernatant was aspirated and cells were gently resuspended in 300 µL MEF media supplemented with Dox. Cells were then drop-plated onto 12mm (60 µL per c.s.; placed in 24 well plate) coverslips. Cells were left to settle for 15 minutes in an incubator then each well was flooded with MEF media supplemented with 1 µg/mL Dox. The following day media was then changed to Neuronal Media (Sciencell) supplemented with Dox, valproic acid (1 mM; EMD Millipore) dibutyryl cAMP (200 µM; Sigma-Aldrich), BDNF, NT-3, CNTF, GDNF (all 10 ng/ml, Peprotech), and Retinoic Acid (1 uM; Sigma-Aldrich) and antibiotics for each vector. Dox was replenished every two days and half the media was changed every 4 days. Drug selection was halted 14 days into conversion. At 0, 5, 10, 15 and 20 days into reprogramming, media was aspirated out of 2 wells containing cells being converted, then these cells were washed with 500uL of sterile PBS and then 200uL of .25% Trypsin was added to each well. Cells were closely monitored and upon detachment from the culture vessel. Upon detachment each well was flooded with complete neuronal media supplemented with 5% FBS. Cells were then counted using a hemocytometer followed centrifugation at 200xG for 5 minutes at 37°C. Cell pellets were finally resuspended in Cryostar CS10 cryopreservation medium (StemCell Technologies, #07955) at a concentration of 100 cells per uL and 100 uL aliquots were placed in a cryogenic freezing container (ThermoFisher Scientific, # 5100-0001) and stored at -80°C until all timepoints were collected.

RNA-seq data acquisition, quality control, and processing for primary rat cortical neurons treated with topotecan

RNA-seq reads were aligned to the Ensembl release 76 top-level assembly with STAR version 2.0.4b. Gene counts were derived from the number of uniquely aligned unambiguous

reads by Subread:featureCount version 1.4.5. Transcript counts were produced by Sailfish version 0.6.3. Sequencing performance was assessed for total number of aligned reads, total number of uniquely aligned reads, genes and transcripts detected, ribosomal fraction known junction saturation and read distribution over known gene models with RSeQC version 2.3.

All gene-level and transcript counts were then imported into the R/Bioconductor package EdgeR and TMM normalization size factors were calculated to adjust samples for differences in library size. Genes or transcripts not expressed in any sample or less than one count-per-million in the minimum group size minus one were excluded from further analysis. The TMM size factors and the matrix of counts were then imported into R/Bioconductor package Limma and weighted likelihoods based on the observed mean-variance relationship of every gene/transcript and sample were then calculated for all samples with the `voomWithQualityWeights` function. Performance of the samples was assessed with a spearman correlation matrix and multi-dimensional scaling plots. Gene/transcript performance was assessed with plots of residual standard deviation of every gene to their average log-count with a robustly fitted trend line of the residuals. Generalized linear models were then created to test for gene/transcript level differential expression. Differentially expressed genes and transcripts were then filtered for FDR adjusted p-values less than or equal to 0.05.

To enhance the biological interpretation of the large set of transcripts, grouping of genes/transcripts based on functional similarity was achieved using *the R/Bioconductor packages GAGE and Pathview*. GAGE and Pathview were also used to generate pathway maps on known signaling and metabolism pathways curated by KEGG.

4.4 Author contributions

M.J.M, D.G.A, S.A.M and A.S.Y. conceived and designed scRNA-seq project, which

A.S.Y. and S.A.M. supervised. M.J.M., H.W.G., and A.S.Y. conceived and designed LGE analysis, which A.S.Y. supervised. D.G.A. reprogrammed and collected cells, performed knockdown of BRG1 during reprogramming, and performed immunostaining of day 20 and day 35 reprogrammed neurons. S.A.M. performed scRNA-seq. S.A.M. and M.J.M. performed data annotation and scrubbing of scRNA-seq data. M.J.M. performed computational analyses. H.W.G. provided critical review of LGE analysis. M.J.M. performed chemical inhibition and electrophysiology of rat cortical neurons. M.J.M. and A.S.Y. wrote the initial manuscript with input from all authors.

4.5 Acknowledgements

We thank A. Herbert for helpful discussions and editing. We thank the Genome Technology Access Center in the Department of Genetics at Washington University School of Medicine for help with genomic analysis. The Center is partially supported by NCI Cancer Center Support Grant #P30 CA91842 to the Siteman Cancer Center and by ICTS/CTSA Grant# UL1TR002345 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. This publication is solely the responsibility of the authors and does not necessarily represent the official view of NCRR or NIH.

4.6 Funding

M.J.M. is supported by the NIH-funded Ruth L. Kirschstein National Research Service Award (NRSA) Institutional Predoctoral Fellowship (T32GM081739; Barch, PI). A.S.Y. is supported by NIH Director's Innovator Award (DP2NS083372-01), Missouri Spinal Cord Injury/Disease Research Program (SCIDRP), Cure Alzheimer's Fund (CAF) and Presidential

Early Career Award for Scientists and Engineers (PECASE).

Chapter 5: Conclusions and Discussion

In part from:

MicroRNA-induced Epigenetic Remodeling During Direct Cell-Fate Conversion of Adult Human Fibroblasts

Daniel G. Abernathy*, Wookyoung Kim*, Matthew J. McCoy*, Allie Lake, Rebecca Ouwenga, Xiaoyun Xing, Daofeng Li, Hyung Joo Lee, Robert O. Heuckeroth, Joseph D. Dougherty, Ting Wang, Andrew S. Yoo

*Authors contributed equally

Cell Stem Cell 21, 332-348, September 7, 2017

Copyright © 2017 Elsevier Inc. All Rights Reserved.

LONGO: An R Package for Interactive Gene Length Dependent Analysis for Neuronal Identity

Matthew J. McCoy*, Alexander J. Paul*, Matheus B. Victor, Michelle Richner, Harrison W. Gabel, Haijun Gong, Andrew S. Yoo and Tae-Hyuk Ahn

*Authors contributed equally

Bioinformatics, In Press

Copyright © 2018 Oxford University Press, All Rights Reserved

Single-cell Sequencing Reveals Transcriptomic Features of Mature Neurons

Matthew J. McCoy*, Daniel G. Abernathy*, Harrison W. Gabel, Samantha A. Morris, Andrew S. Yoo

*Authors contributed equally

Manuscript in preparation

5.1 Molecular Mechanisms of MicroRNA-Mediated Reprogramming

MiRNA-mediated neuronal conversion appears to be distinct from current models of cell fate reprogramming. Two models of lineage reprogramming have been proposed: one based on transcription factor cooperativity and positive feedback loops (Jaenisch and Young, 2008; Soufi et al., 2012; Vierbuchen and Wernig, 2012), and the other based on the “on-target” pioneer activity of a TF as shown in neuronal conversion of mouse embryonic fibroblasts (Wapinski et al., 2013). In contrast, canonical gene regulation by miRNAs presumably utilizes the removal of gene products through translational repression and transcript degradation for the target genes. This mode of repression, in conjunction with the multitude of anti-neurogenic genes targeted by miR-9/9*-124, suggests that miRNA-mediated reprogramming acts through an alternative mechanism. We propose miR-9/9*-124 expression in non-neuronal somatic cells initiates gradual, yet active changes in the activities of multiple chromatin modifiers while simultaneously repressing anti-neuronal genes and activating neuronal genes culminating in a binary cell-fate switch. This model is supported by the rapid cell cycle exit observed upon ectopic miR-9/9*-124 expression, the subsequent neuronal switching among chromatin modifiers, steady increase in epigenetic and transcriptional changes, and the time scale in which conversion takes place.

5.1.1 Chromatin Remodeling Accompanies Cell Fate Conversion

In this study, we reveal the surprising potency of miR-9/9*-124 for remodeling chromatin and altering DNA methylation. Surprisingly, preexisting neuronal loci within the heterochromatic regions in human fibroblasts opened in response to miR-9/9*-124. These data

suggest that the robustness of miRNA-mediated reprogramming observed in human adult fibroblasts could stem from the ability of these miRNAs to induce epigenetic changes. Cellular processes and identity are governed by the cumulative action of multiple levels of genome regulation, and it is unlikely that a single genetic component downstream of miR-9/9*-124 would mediate such drastic changes in cellular identities. For example, almost every level of epigenetic remodeling participates in the induction of pluripotency (Takahashi and Yamanaka, 2016), and instructions operating through multiple levels of genetic and epigenetic regulation are likely required for true cell-fate conversion. As such, our thorough analyses of miR-9/9*-124 induced transdifferentiation of human fibroblasts into functional neurons highlights molecular processes that are critical to cell fate conversion.

5.1.2 A Modular Neuronal State

The miRNA-mediated neuronal platform presented here affords modularity to direct the conversion to specific neuronal subtypes. Numerous studies in developmental neuroscience have identified subtype-specific TFs or terminal selector genes that could be incorporated in neuronal reprogramming technologies. Yet, identifying molecules capable of overcoming the cell-fate barrier present in human somatic cells and eliciting a permissive environment in which terminal selector genes can act has proven to be challenging. In this study, we demonstrate that miR-9/9*-124 are capable of evoking a neuronal state through extensive epigenetic remodeling that can lead to generation of a neuronal population highly enriched with spinal cord motor neurons on addition of ISL1 and LHX3. Because motor neurons are the major neuronal subtype affected in Amyotrophic Lateral Sclerosis (ALS) and Spinal Muscular Atrophy (SMA), the robustness and specificity of neuronal conversion employing miRNAs and motor neuron TFs may pave the way towards generating patient-specific MNs for disease modeling.

5.2 Long Gene Expression Identifies Neurons

5.2.1 LONGO is a Computation Tool that Measures LGE to Identify Neurons

As interest in modeling neuronal cell fate acquisition and modeling neurological diseases using reprogrammed neurons continues to grow, so does the need to validate and improve methodologies for neuronal reprogramming. Toward this end, we have developed a method to identify neurons from non-neuronal cells based on the collective expression of the longest genes in the genome. By incorporating this method into an R computation package, LONGO, we hope to facilitate this analysis for other researchers.

For investigators interested in adopting an existing reprogramming technique in their research programs, LONGO provides a metric for evaluating the efficiency of different techniques. The robustness of LGE analysis provided by LONGO and the steadily declining cost of transcriptome analyses lead us to strongly advise investigators who perform neuronal reprogramming or differentiation to routinely generate transcriptome data and analyze LGE to confirm neuronal identity.

One of the known correlates of gene length is the levels of alternative splicing (McGuire et al., 2008), which is a feature also enriched in vertebrate and invertebrate nervous systems (Barbosa-Morais et al., 2012; Jelen et al., 2007; Merkin et al., 2012). Some of these long genes encode over 20,000 different protein isoforms, and are uniquely expressed in individual neurons, endowing each neuron with a unique molecular profile (Miura et al., 2013). So, while each neuron expresses copies of long genes, these copies are not identical to even those from neighboring neurons of the same tissue. Our method allows us to measure the expression of long genes collectively to determine if a cell or tissue is neuronal, but does not typically provide the resolution to determine differences between one type of neuron versus another, or one type of

neural tissue versus another.

5.2.2 Long Gene Expression Distinguishes Mature from Immature Neurons

By performing single-cell RNA-seq during microRNA-mediated neuronal reprogramming, I discovered that this process results in a surprisingly homogenous reprogramming trajectory. Beyond demonstrating the robustness of microRNA-mediated neuronal reprogramming, I discovered that LGE can be applied at the single-cell level, and successfully identifies reprogrammed neurons. After the transient upregulation of immature neuron gene expression, LGE continues to increase alongside mature neuron gene expression and acquisition of spontaneous electrical activity. One of the markers of mature neuronal gene expression patterns is subunit switching from the neural progenitor BAF to the neuronal BAF complex, which we observe during neuronal reprogramming. On disrupting the functional BAF complex during reprogramming, we find that LGE is decreased, as is the acquisition of neuronal marker expression. The close relationship between LGE onset and subunit switching of the BAF complex provides insight into the gene expression programs initiated during neuronal maturation, and suggests that gene length plays an important role in the evolution and/or regulation of these programs in the nervous system. Particularly since the BAF complex is deeply conserved, determining the conservation of LGE may provide important insights into the evolution of the nervous system. Further, by disrupting LGE via TOP1 inhibition, we show that LGE is necessary for spontaneous electrical activity in mature neurons. We propose that LGE is a cumulative measure of the synaptic and adhesion molecules necessary for the electrophysiological properties of mature neurons.

5.3 References

- Abernathy, D.G., Kim, W.K., McCoy, M.J., et al. MicroRNAs Induce a Permissive Chromatin Environment that Enables Neuronal Subtype-Specific Reprogramming of Adult Human Fibroblasts. *Cell Stem Cell* 2017;21:332-348.
- Alexa, A. and Rahnenfuhrer, J. topGO: Enrichment Analysis for Gene Ontology. R package version 2.30.0. 2016.
- Amin, N.D., et al. Loss of motoneuron-specific microRNA-218 causes systemic neuromuscular failure. *Science* 2015;350:1525-1529.
- Anders, S., Pyl, P.T., and Huber, W. HTSeq--a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;31:166-169.
- Artegiani et al. A Single-Cell RNA Sequencing Study Reveals Cellular and Molecular Dynamics of the Hippocampal Neurogenic Niche. *Cell Reports* 2017;21:3271–3284.
- Barbosa-Morais, et al. The Evolutionary Landscape of Alternative Splicing in Vertebrate Species. *Science* 2012;338:1587.
- Boudreau, R.L., et al. Transcriptome-wide discovery of microRNA binding sites in human brain. *Neuron* 2014;81:294-305.
- Cantone, I., and Fisher, A.G. Epigenetic programming and reprogramming during development. *Nat Struct Mol Biol* 2013;20:282-289.
- Chi, S.W., et al. Argonaute HITS-CLIP decodes microRNA-mRNA interaction maps. *Nature* 2009;460:479-486.
- Colasante, G., et al. Rapid Conversion of Fibroblasts into Functional Forebrain GABAergic Interneurons by Direct Genetic Reprogramming. *Cell Stem Cell* 2015;17(6):719-734.
- Conaco, C., et al. Reciprocal actions of REST and a microRNA promote neuronal identity. *Proc Natl Acad Sci* 2006;103:2422-2427.
- Dobin, A., et al. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* 2013;29:15-21.
- Durinck, S., et al. Mapping identifiers for the integration of genomic datasets with the R/Bioconductor package biomaRt. *Nat Protoc* 2009;4(8):1184-1191.
- Edgar, R., Domrachev, M. and Lash, A.E. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* 2002;30(1):207-210.
- Egan, C.M., et al., CHD5 is required for neurogenesis and has a dual role in facilitating gene expression and polycomb gene repression. *Dev Cell* 2013;26:223-236.

Endres, D.M. and Schindelin, J.E. A new metric for probability distributions. *IEEE Transactions on Information Theory* 2003;49(7):1858-1860.

Ernst, J., et al. Integrating multiple evidence sources to predict transcription factor binding in the human genome. *Genome Res* 2010;20:526-536.

Feng, W., et al. The chromatin remodeler CHD7 regulates adult neurogenesis via activation of SoxC transcription factors. *Cell Stem Cell* 2013;13:62-72.

Fonnum, F. Recent developments in biochemical investigations of cholinergic transmission. *Brain Res* 1973;62:497-507.

Gabel, H.W., et al. Disruption of DNA-methylation-dependent long gene repression in Rett syndrome. *Nature* 2015;522(7554):89-93.

Gehrke, S., et al. Pathogenic LRRK2 negatively regulates microRNA-mediated translational repression. *Nature* 2010;466:637-641.

Gill, K.P., et al. Enriched retinal ganglion cells derived from human embryonic stem cells. *Sci Rep* 2016;6:30552.

Habib, N., et al. Div-Seq: Single-nucleus RNA-Seq reveals dynamics of rare adult newborn neurons. *Science* 2016;353(6302):925-928.

Heiman, M., et al. Cell type-specific mRNA purification by translating ribosome affinity purification (TRAP). *Nat Protoc* 2014;9:1282-1291.

Heinz, S., et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell* 2010;38:576-589.

Herring, C.A., et al. Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. *Cell Systems* 2017;6(1):37-51.

Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol* 2013;14:R115.

Hu, W., et al. Direct Conversion of Normal and Alzheimer's Disease Human Fibroblasts into Neuronal Cells by Small Molecules. *Cell Stem Cell* 2015;17(2):204-212.

Huh, C.J., et al. Maintenance of age in human neurons generated by microRNA-based neuronal conversion of fibroblasts. *Elife* 2016:5.

Husain, A., et al. Chromatin remodeler SMARCA4 recruits topoisomerase 1 and suppresses transcription-associated genomic instability. *Nature Communications* 2016;7:10549.

Jaenisch, R., and Young, R. Stem cells, the molecular circuitry of pluripotency and nuclear reprogramming. *Cell* 2008;132:567-582.

- Jelen, N., et al. Evolution of Nova-dependent splicing regulation in the brain. *PLoS Genet* 2007;3:1838-1847.
- Jiang, C., and Pugh, B.F. Nucleosome positioning and gene regulation: advances through genomics. *Nat Rev Genet* 2009;10:161-172.
- Kadoch, C., et al. Dynamics of BAF-Polycomb complex opposition on heterochromatin in normal and oncogenic states. *Nat Genet* 2017;49:213-222.
- King, I.F., et al. Topoisomerases facilitate transcription of long genes linked to autism. *Nature* 2013;501(7465):58-62.
- Kohli, R.M., and Zhang, Y. TET enzymes, TDG and the dynamics of DNA demethylation. *Nature* 2013;502:472-479.
- Langmead, B., and Salzberg, S.L. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 2012;9:357-359.
- Lessard, J., et al. An essential switch in subunit composition of a chromatin remodeling complex during neural development. *Neuron* 2007;55:201-215.
- Lister, R., et al. Global epigenomic reconfiguration during mammalian brain development. *Science* 2013;341:1237905.
- Liu, Z., et al. Single-cell transcriptomics reconstructs fate conversion from fibroblast to cardiomyocyte. *Nature* 2017;551:100–104.
- Mabb, A.M., et al. Topoisomerase 1 inhibition reversibly impairs synaptic function. *PNAS* 2014;111(48):17290-17295.
- Mahony, S., et al. Ligand-dependent dynamics of retinoic acid receptor binding during early neurogenesis. *Genome Biol* 2011;12(1):R2.
- Makeyev, E.V., et al. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol Cell* 2007;27:435-448.
- Maunakea, A.K., et al. Conserved role of intragenic DNA methylation in regulating alternative promoters. *Nature* 2010;466:253-257.
- Mazzoni, E.O., et al. Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity. *Nat Neurosci* 2013;16:1219-1227.
- McCoy et al., LONGO: An R Package for Interactive Gene Length Dependent Analysis for Neuronal Identity. *Bioinformatics* In press.
- McGuire, A.M., et al. Cross-kingdom patterns of alternative splicing and splice recognition. *Genome Biology* 2008;9:R50-R50.

McLean, C.Y., et al. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol* 2010;28:495-501.

Merkin, J., et al. Evolutionary Dynamics of Gene and Isoform Regulation in Mammalian Tissues. *Science* 2012;338:1593.

Mertens, J., et al. Directly Reprogrammed Human Neurons Retain Aging-Associated Transcriptomic Signatures and Reveal Age-Related Nucleocytoplasmic Defects. *Cell Stem Cell* 2015;17:705-718.

Mertens, J., et al. Evaluating cell reprogramming, differentiation and conversion technologies in neuroscience. *Nat Rev Neurosci* 2016;17:424-437.

Miller, J.A., et al. Transcriptional landscape of the prenatal human brain. *Nature* 2014;508(7495):199-206.

Miller, J.D., et al. Human iPSC-based modeling of late-onset disease via progerin-induced aging. *Cell Stem Cell* 2013;13:691-705.

Miura, S.K., et al. Probabilistic Splicing of Dscam1 Establishes Identity at the Level of Single Neurons. *Cell* 2013;155(5):1166-1177.

Niu, W., et al. In vivo reprogramming of astrocytes to neuroblasts in the adult brain. *Nat Cell Biol* 2013;15:1164-1175.

Okano, M., et al. DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* 1999;99:247-257.

Packer, A.N., et al. The bifunctional microRNA miR-9/miR-9* regulates REST and CoREST and is downregulated in Huntington's disease. *J Neurosci* 2008;28:14341-14346.

Pang, Z.P., et al. Induction of human neuronal cells by defined transcription factors. *Nature* 2011;476:220-223.

Parish, J.Z. et al. Genome-wide analyses identify transcription factors required for proper morphogenesis of Drosophila sensory neuron dendrites. *Genes Dev* 2006;20(7):820-35.

Pasquinelli, A.E. MicroRNAs and their targets: recognition, regulation and an emerging reciprocal relationship. *Nat Rev Genet* 2012;13:271-282.

Perrier, A.L., et al. Derivation of midbrain dopamine neurons from human embryonic stem cells. *Proc Natl Acad Sci* 2004; U S A 101, 12543-12548.

Richner, M., Victor, M.B., et al. MicroRNA-based conversion of human fibroblasts into striatal medium spiny neurons. *Nat Protocols* 2015;10:1543-1555.

Ritchie, M.E., et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* 2015;43:e47.

Roadmap Epigenomics, C., et al. Integrative analysis of 111 reference human epigenomes. *Nature* 2015;518:317-330.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 2010;26:139-140.

Schulz, M.H., et al. DREM 2.0: Improved reconstruction of dynamic regulatory networks from time-series expression data. *BMC Syst Biol* 2012;6:104.

Shibata, M., et al. MicroRNA-9 modulates Cajal-Retzius cell differentiation by suppressing Foxg1 expression in mouse medial pallium. *J Neurosci* 2008;28:10415-10421.

Shin, H., et al. CEAS: cis-regulatory element annotation system. *Bioinformatics* 2009;25:2605-2606.

Smedley, D., et al. The BioMart community portal: an innovative alternative to large, centralized data repositories. *Nucleic Acids Res* 2015;43(W1):W589-598.

Smith, C.M., et al. The mouse Gene Expression Database (GXD): 2014 update. *Nucleic Acids Res* 2014;42:D818-824.

Soufi, A., Donahue, G., and Zaret, K.S. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell* 2012;151:994-1004.

Staahl, B.T., et al. Kinetic analysis of npBAF to nBAF switching reveals exchange of SS18 with CREST and integration with neural developmental pathways. *J Neurosci* 2013;33:10348-10361.

Stefanakis, N., Carrera, I., and Hobert, O. Regulatory Logic of Pan-Neuronal Gene Expression in *C. elegans*. *Neuron* 2015; 87:733-750.

Steinbeck, J.A., et al. Functional Connectivity under Optogenetic Control Allows Modeling of Human Neuromuscular Disease. *Cell Stem Cell* 2016;18:134-143.

Sugino, K., et al. Cell-type-specific repression by methyl-CpG-binding protein 2 is biased toward long genes. *J Neurosci* 2014;34(38):12877-12883.

Takahashi, K., and Yamanaka, S. A decade of transcription factor-mediated reprogramming to pluripotency. *Nat Rev Mol Cell Biol* 2016;17:183-193.

Tanabe, Y., William, C., and Jessell, T.M. Specification of motor neuron identity by the MNR2 homeodomain protein. *Cell* 1998;95:67-80.

Tasic, B., et al. Adult mouse cortical cell taxonomy revealed by single cell transcriptomics. *Nature neuroscience* 2016;19(2):335-346.

Thiebes, K.P., et al. miR-218 is essential to establish motor neuron fate as a downstream effector of Isl1-Lhx3. *Nat Commun* 2015;6:7718.

Tiwari, V.K., et al. Target genes of Topoisomerase IIbeta regulate neuronal survival and are defined by their chromatin state. *Proc Natl Acad Sci U S A* 2012;109:E934-943.

Treutlein, B., et al. Dissecting direct reprogramming from fibroblast to neuron using single-cell RNA-seq. *Nature* 2016;534(7607):391-395.

Tripathi, S., et al. Meta- and Orthogonal Integration of Influenza "OMICs" Data Defines a Role for UBR4 in Virus Budding. *Cell Host Microbe* 2015;18:723-735.

Vadhvani, M., et al. The centrosomal E3 ubiquitin ligase FBXO31-SCF regulates neuronal morphogenesis and migration. *PLoS One* 2013;8:e57530.

Victor, M.B. et al. Striatal neurons directly converted from Huntington's disease patient fibroblasts recapitulate age-associated disease phenotypes. *Nature Neuroscience* 2018;21:341-352.

Victor, M.B., Richner, M., et al. Generation of human striatal neurons by microRNA-dependent direct conversion of fibroblasts. *Neuron* 2014;84(2):311-323.

Vierbuchen, T., and Wernig, M. Molecular roadblocks for cellular reprogramming. *Mol Cell* 2012;47:827-838.

Visvanathan, J., et al. The microRNA miR-124 antagonizes the anti-neural REST/SCP1 pathway during embryonic CNS development. *Genes Dev* 2007;21:744-749.

Wapinski, O.L., et al. Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. *Cell* 2013;155:621-635.

Wichterle, H., et al. Directed differentiation of embryonic stem cells into motor neurons. *Cell* 2002;110:385-397.

Wise, A., and Bar-Joseph, Z. cDREM: inferring dynamic combinatorial gene regulation. *J Comput Biol* 2015;22:324-333.

Wu, H., and Zhang, Y. Mechanisms and functions of Tet protein-mediated 5-methylcytosine oxidation. *Genes Dev* 2011;25:2436-2452.

Wu, J.I., et al. Regulation of Dendritic Development by Neuron-Specific Chromatin Remodeling Complexes. *Neuron* 2007;56(1):94-108.

Wu, S., et al. Multiple microRNAs modulate p21Cip1/Waf1 expression by directly targeting its 3' untranslated region. *Oncogene* 2010;29:2302-2308.

Xu, X., et al. Cell type-specific expression analysis to identify putative cellular mechanisms for neurogenetic disorders. *J Neurosci* 2014;34:1420-1431.

Xue, Y., et al. Sequential regulatory loops as key gatekeepers for neuronal reprogramming in human cells. *Nature neuroscience* 2016;19(6):807-815.

Yoo, A.S., et al. MicroRNA-mediated conversion of human fibroblasts to neurons. *Nature* 2011;476:228-231.

Yoo, A.S., et al. MicroRNA-mediated switching of chromatin-remodelling complexes in neural development. *Nature* 2009;460:642-646.

Yu, Y., et al. A rat RNA-Seq transcriptomic BodyMap across 11 organs and 4 developmental stages. *Nat Commun* 2014;5:3230.

Zhang, B., et al. Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Res* 2013;23:1522-1540

Zhang, Y., et al. Purification and Characterization of Progenitor and Mature Human Astrocytes Reveals Transcriptional and Functional Differences with Mouse. *Neuron* 2016;89(1):37-53.