Regulatory logic of cellular diversity in the nervous system

Michael Closser

Submitted in partial fulfillment of the requirements for
the degree of Doctor of Philosophy under the Executive
Committee of the Graduate School of Arts and Sciences

COLUMBIA UNIVERSITY

2018

© 2018

Michael Closser

All Rights Reserved

Abstract

Regulatory logic of cellular diversity in the nervous system

Michael Closser

During nervous system development, thousands of distinct cell types are generated and assembled into complex circuits that control all aspects of animal cognition and behavior. Understanding what these diverse cells are, how they are generated, and what they do in the context of circuits and behavior form the fundamental efforts of the field of neuroscience. In this thesis, I investigate how the genomic organization of regulatory elements informs specific patterns of gene expression in the nervous system. In particular, I examine how distinct combinations of transcription factors interpret information encoded in the genome to control global gene expression programs in a cell-type-specific manner.

In Chapter Two, I describe the establishment of a developmentally inspired transcriptional programming system to generate spinal and cranial motor neurons directly from mouse embryonic stem cells. Programmed motor neurons acquire general characteristics that mirror their *in vivo* counterparts, providing a robust system for studying cell fate specification in the nervous system. Combinatorial expression of cell-type-specific programming factors informs context-dependent enhancer binding and acquisition of appropriate cell-type-specific molecular and functional properties.

In Chapter Three, I take advantage of this robust, experimentally accessible system to probe the chromatin-level organization and regulatory principles controlling specificity of motor neuron gene expression programs. Motor neuron genes are controlled by multiple distantly distributed enhancer constellations stretched across large regulatory domains. Using this motor neuron specification model, I discovered a unique regulatory organization controlling gene expression in the nervous system, whereby neuronal genes are controlled from uniquely complex regulatory domains acting over large distances.

In Chapter Four, I extrapolate on the insights gained from studying motor neurons at a single point in time to investigate the dynamics of the regulatory environment during neuronal maturation. We demonstrate that enhancers are highly dynamic even after postmitotic specification. The dynamic nature of enhancers is dependent on combinatorial binding with new transcriptional cofactors.

Overall, my results suggest that neuronal gene expression programs within a single cell type are regulated in a highly dynamic fashion by a complex set of enhancers. I propose that during development the immense cellular complexity of the nervous system is established and maintained by correspondingly complex repertoire of enhancers.

Table of Contents

List of Figures ..... **iii**

Acknowledgments ..... **iv**

CHAPTER 1: Introduction **1**

A. Regulatory control of cell identity and diversity in the nervous system 2
   1. Fundamental concepts of cellular identity and diversity in the nervous system
   2. Generation of generic motor neuron identity in the mammalian spinal cord
   3. Dynamic combinatorial expression of transcription factors in motor neurons
   4. Combinatorial expression of motor neuron programming factors in other neuronal cells
   5. Regulatory control of postmitotic neuron identity
   6. Lessons learned from a simple nervous system

B. Enhancer control of gene expression 14
   1. A brief overview of mammalian enhancers
   2. Transcription factor binding at enhancers
   3. Dynamic activation of enhancers during development
   4. Transcription factor control of long-range enhancer-promoter interactions
   5. Topological domains and insulated neighborhoods

C. Questions, Challenges and Approach 26

D. Figures 31

CHAPTER 2: Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity **38**

A. Introduction 39

B. Results 41
   1. Ngn2, Isl1 and Lhx3 program spinal motor neuron fate
   2. Ngn2, Isl1 and Phox2a program cranial motor neuron fate
   3. Functional characterization of induced NIL and NIP neurons
   4. Expression profiles of motor neuron programming
   5. Motor neuron programming bypasses neural progenitor stages
   6. Isl1 genome binding is dependent on programming partners
   7. Cell-specific Isl binding correlates with gene expression
   8. Sequence motifs explain differential Isl binding
   9. Lhx3 and Phox2a co-occupy Isl-binding sites

C. Discussion 50

D. Figures 53

CHAPTER 3: Neuronal gene expression is controlled by constellations of enhancers distributed across large regulatory domains **64**

A. Introduction 65

B. Results 68
   1. An integrated approach to map *in vivo* regulatory regions in motor neurons
   2. Dynamic reorganization of the active enhancer landscape in motor neurons
   3. Mapping high-resolution enhancer-promoter interactions in motor neurons
   4. IL mediated enhancer-promoter connectivity directly control MN expression program
   5. Constellations of distributed enhancers control the expression of motor neuron genes
   6. Enhancer constellations are distributed across large regulatory regions
   7. Independent regulation from enhancer constellations
   8. Gene regulatory organization in the nervous system

C. Discussion 82

1. Programming factors control the postmitotic expression program in motor neurons
2. Distributed enhancer constellations in motor neurons
3. Regulatory complexity in the nervous system
D. Figures 88

CHAPTER 4: Expression of terminal effector genes in mammalian neurons is maintained by a dynamic relay of transient enhancers 99

A. Introduction 100
B. Results 102
1. Dynamic changes in the genomic regulatory landscape in postmitotic motor neurons
2. Divergent binding of Ngn2 and Isl1 to stage-specific enhancers
3. Isl1 rapidly relocates to Onecut1-bound sites following Lhx3 downregulation
4. Stage-specific enhancers maintain expression of postmitotic motor neuron genes
5. Transient Isl1-bound enhancers control stable motor neuron gene expression
6. Isl1 is recruited to stage-specific enhancers through interactions with Onecut1 or Lhx3
7. Isl1 and p300 are preferentially recruited to clusters of Onecut1 or Lhx3
C. Discussion 111
1. Outwardly stable gene expression in a dynamic regulatory environment
2. Indirect recruitment of Isl1 to enhancers through protein-protein interactions
3. Isl1 operates as an integrator TF, functionally integrating binding patterns of anchor TFs
D. Figures 115

CHAPTER 5: General discussion and future perspectives 124

A. Cell- and stage-specific combinatorial binding of transcription factors
B. Klf factors as general activators of enhancer-promoter engagement
C. Enhancer-promoter specificity
D. A dynamic relay of enhancers initiates and maintains neuronal gene expression programs
E. Reverse engineering cell identity: implications for development and disease

Experimental procedures 133

References 140

List of figures

Introduction

Fig 1.1 Secreted signals regulate the spatial expression of transcription factors in the neural tube....... 32
Fig 1.2 LIM transcription factor codes control spinal motor neuron subtype-specific properties............ 33
Fig 1.3 Potential modes of regulatory control in the nervous system.. 34
Fig 1.4 Chromatin accessibility and histone marks at regulatory elements... 35
Fig 1.5 Genomic methods to identify regions involved in enhancer promoter interactions 36
Fig 1.6 Insulated Neighborhoods 37

Chapter 2

Fig 2.1 NIL and NIP transcription factors program spinal and cranial motor neurons, respectively........ 54
Fig 2.2 Characterization of NIL and NIP induced motor neurons... 55
Fig 2.3 Xenotransplantation of induced NIL and NIP motor neurons 57
Fig 2.4 NIL and NIP induce different transcriptomes.. 58
Fig 2.5 Induced iNIL cells bypass progenitor stages.. 60
Fig 2.6 Context-specific Isl1 genome association correlates with differential gene expression 61
Fig 2.7 Sequence motifs occupied by Isl1 in iNIL and iNIP cells... 62

Chapter 3

Fig 3.1 A comparison of ATAC-seq based chromatin accessibility in *in vitro* and *in vivo* motor neurons 88
Fig 3.2 Isl/Lhx3 binding drives dynamic reorganization of active enhancer landscape in motor neurons 89
Fig 3.3 Isl/Lhx3 bound enhancers are pioneered during motor neuron specification............................. 90
Fig 3.4 Isl/Lhx3 binding drives dynamic reorganization of active enhancer landscape in motor neurons 91
Fig 3.5 IL specifically induce MN gene expression program through E-P interaction 92
Fig 3.6 Regulation of MN expression program from distantly distributed enhancer constellations......... 93
Fig 3.7 Functional dissection of distributed enhancers in motor neurons 95
Fig 3.8 Neuronal genes are associated with increased regulatory complexity in large gene poor regulatory domains.. 96

Chapter 4

Fig 4.1 Dynamic changes in enhancer organization and gene expression during MN differentiation... 115
Fig 4.2 Ephemeral nature of enhancers established by motor neuron programming TFs 116
Fig 4.3 Ephemeral nature of Isl1 bound enhancers in postmitotic neurons 117
Fig 4.4 Motor neuron effector genes are associated with stage-specific Isl1-bound enhancers........... 118
Fig 4.5 Spatiotemporal specificity of transient Isl1-bound enhancers in the developing neural tube 119
Fig 4.6 Transient Isl1-bound enhancers are necessary for a stable expression of MN genes............. 120
Fig 4.7 High-resolution mapping of Isl1 and Onecut1 binding sites in maturing MN enhancers........... 121
Fig 4.8 Isl1 is recruited to DNA through interactions with Lhx3 in nascent motor neurons 122
Fig 4.9 Isl1 is preferentially recruited to enhancers containing clusters of Onecut1 or Lhx3 TFs 123

General Discussion and Future Perspectives

Fig 5.1 Klf factors are enriched at interacting enhancers 132

Acknowledgments

I would like to thank my PhD advisor, Hynek Wichterle, for giving me the opportunity and freedom to pursue my interests, allowing me to find my way as an independent scientist, and providing critical feedback, whether I asked for it or not. His enthusiasm and support has always made it easy to continue on during the obstacles one faces in graduate school. I will be forever thankful for the lessons I have learned during this process.

I would like to thank the members of my thesis committee: Tom Maniatis, Oliver Hobert, Stavros Lomvardas and Lori Sussel, who were all instrumental during my time in graduate school and provided valuable feedback and support. Their advice and criticism always helped me maintain perspective on what was important. I would also like to thank Jeremy Dasen for joining my thesis committee as my external examiner.

Thank you to my undergraduate advisors at UCSD, Richard Kolodner and Christopher Putnam, who first taught me how to design and think critically about experiments. It was in Richard's lab where I first learned what it was like to be an independent scientist and my time there was fundamental to my success in graduate school and as a scientist.

Thank you to everyone in the Wichterle lab, past and present. In particular, Esteban Mazzoni who provided important guidance in my early stages and throughout graduate school. Phuong Hoang for always being there when I had a question about how development works or how to write something that was remotely articulate. Rachel Kopunova for being a fantastic undergraduate student over the years and helping with many important experiments. Martin Jacko for interesting conversations about science, crazy ideas not related to science and even the meaning of life.

Most importantly, I want to thank the person who made this project possible: Yuchun Guo has been a close collaborator for my entire time in graduate school and taught me everything I know about computational biology. Together we've spent years trying to make sense of how transcription factors interpret information encoded in the genome to control complex neuronal gene expression programs and we finally have. Finally, thanks to Dave Gifford and everyone in his lab for support with analyzing sequencing data over the years.

Chapter 1: Introduction

Over one hundred years ago, the heroic work of Ramon y Cajal unveiled a fundamental principle of the nervous system which is that it is composed of a morphologically diverse set of cells that form the functional building blocks of neural circuits (Cajal 1899). Understanding what these diverse cells are, how they are generated, and what they do in the context of circuits and behavior form the fundamental efforts of the field of neuroscience. Here, I will focus on understanding the molecular mechanisms controlling specification and control of distinct cell identities in the nervous system.

The mammalian nervous system consists of billions of neurons, composed of thousands of distinct types, making trillions of synapses (Luo et al., 2008). The proper development of such a complex nervous system is a highly orchestrated process that is dependent on the precise control of gene expression programs in space and time. Individual neurons express distinct combinations of genes, which together define their identity. Although considerable information exists with respect to the genetic programs that specify cell-type identity of neurons, the regulatory mechanisms controlling neuron specific gene expression programs remain poorly understood. Given the cellular complexity of the nervous system, a major question that arises is what is the regulatory organization facilitating the control of specific patterns of gene expression?

Taking a reductionist point of view, we can say that all neurons share the same genetic information. Thus, the answer to regulation of cellular diversity likely resides in the genome. The recent sequencing of animal genomes has uncovered an interesting discrepancy between the organisms with simple nervous systems like the roundworm, *C. elegans,* and those with more complex nervous systems including mice and humans. In general, more complex organisms do not have more genes but instead have much larger genomes consisting largely of non-coding DNA, suggesting the interesting idea that cellular complexity could be controlled by increased genomic complexity (Bulger and Groudine, 2011; de Laat and Duboule, 2013).

In this introductory chapter, I will first describe how the combinatorial expression of transcription factors results in the specification of distinct cell identities in the nervous system, with a particular focus on motor neurons in the spinal cord. Next, I will describe what is currently known about the regulation of

gene expression in mammals. In particular, I focus on the dynamic nature of enhancer activity and the role of genome organization in the precise control of gene expression programs. Finally, I will describe the approach I take in this thesis to dissect the molecular mechanisms controlling the specification and maintenance of specific cell identities in the mammalian nervous system.

A. Regulatory control of cell identity and diversity in the nervous system

Fundamental concepts of cellular identity and diversity in the nervous system

In order to study the mechanisms controlling specification of cellular diversity we need to first understand what defines a particular neuron's identity. Classically, distinct cell identities have been defined based on gross phenotypic characteristics including morphology, physiological properties and physical locations within the nervous system (Fishell and Heintz, 2013; Jessell, 2000a; Livesey and Cepko, 2001; Molyneaux et al., 2007). However, recent technological advances in genetic labeling and molecular profiling with genomic approaches allow for more robust classification of cell identity in the mammalian nervous system. In a simplistic view the classical definitions of neuronal identity can be reduced to define each neurons identity as the sum its gene expression program. That is individual properties which define types of neurons are fundamentally a reflection of differential gene expression.

The mammalian nervous system is made up of thousands of distinct neuron types with even more subtypes begging the question how such complexity in cell identity and by extension, complexity in their gene expression programs is established. One important concept to consider is that in general, neurons are not defined by the specific expression of individual genes, but by their global expression program. That is many neurons share expression of genes expressed throughout the nervous system, but have unique expression of cell specific genes. For example, in the case of neurotransmitter identity, motor neurons in the spinal cord and neurons in the basal forebrain are both cholinergic, yet have very different cellular identities (morphology, axon projections, synaptic partners). Together these concepts introduce a few fundamental questions that remain to be answered. What are the gene regulatory mechanisms that allow for the specific control of cell identity in the nervous system? Are all genes in a given neuron regulated in a common way? Are cell specific genes regulated differently than pan-neuronal

genes? In the next portion of this chapter, I will describe years of work in spinal motor neurons that helped to uncover the genetic mechanisms controlling aspects of cell identity.

Generation of generic motor neuron identity in the mammalian spinal cord

The spinal cord has classically served as a model to study acquisition of cell identity in the CNS. In particular, the specification of ventral cell fates in the spinal cord has provided a relatively simple model for studying molecular and cellular mechanisms of cell fate specification. Ventral spinal cord identity is generated via the intersection of signaling pathways along the dorsoventral (D-V) and rostrocaudal (R-C) axes of the developing neural tube. Sonic Hedgehog (Shh) is secreted from the notochord and floor plate, forming a concentration gradient along the D-V axis of the ventral neural tube to generate different classes of ventral neurons in a concentration-dependent manner (Briscoe et al., 2000). Additionally, retinoids secreted from paraxial mesoderm induce a caudal character to differentiate spinal neural structures from the more rostral forebrain, midbrain and hindbrain. Together, the RA and Shh signaling pathways are chiefly responsible for creating cartesian-like coordinates that converge along the dorsoventral and rostrocaudal axes to generate ventral spinal cord identity (Jessell, 2000a).

The ventral spinal cord is comprised of multiple classes of neurons including motor neurons and interneurons (Arber 2012; Catela, Shin, and Dasen 2015). Specific neural identities are generated in response to a gradient of Shh signaling which induces transcription factor networks of homeodomain and basic helix loop helix (bHLH) classes of transcription factors that establish a set of cross-repressive interactions that ultimately set up discrete progenitor domains (p0-p3 and pMN) in the ventral spinal cord (Figure 1.1A, B). Each progenitor domain gives rise to a distinct class of ventral neurons: from ventral to dorsal, V3 interneuron, motor neuron, V2 interneuron, V1 interneuron and V0 interneurons (Figure 1.1B) (Briscoe et al., 2000; Dessaud et al., 2008; Jessell, 2000a; Novitch et al., 2001).

Motor neurons are born from the second most ventral domain, the pMN domain, marked by expression of the bHLH transcription factor Olig2. Within this domain, a transcriptional cascade is initially established whereby Nkx6 family members induce expression of the bHLH transcription factor Olig2 in order to initiate the generation of motor neuron progenitor identity (Figure 1.1B,C) (Briscoe et al., 2000; Novitch et al., 2001). Olig2 promotes motor neuron identity over competing ventral fates, V3 interneurons,

which arise from the Nkx2.2 expressing p3 domain; as well as the more dorsal V2 interneurons arising from the Irx3-expressing p2 domain. Indeed, misexpression of Olig2 in spinal progenitors leads to differentiation of more dorsal progenitors into motor neurons (Novitch et al., 2001). Olig2 expression is followed by induction of the proneural bHLH transcription factor Neurogenin-2 (Ngn2) that drives motor neuron progenitors to exit the cell cycle and acquire expression of generic neuronal traits (Figure1.1C) (Mizuguchi et al., 2001; Novitch et al., 2001). Indeed, Ngn2 is required to induce the neurogenic program in motor neurons and induce expression of pan-neuronal genes as loss of function of Ngn2 leads to a defect in motor neuron generation (Scardigli et al., 2001; Tan et al., 2016). Interestingly, despite being expressed in many cell types throughout the nervous system, Ngn2 seems to have some specificity. In particular, replacement of Ngn2 with another proneural transcription factor, Mash1, results in maintenance of postmitotic neuron exit from the cell cycle and neurogenesis, but fails to accurately specify motor neurons (Parras et al., 2002). Taken together, these studies have uncovered the intricate mechanisms by which spatial signaling cues are interpreted and converted into gene regulatory networks, which facilitate the generation of distinct cell types within the developing nervous system. However, all of these transcriptional events take place prior to the exit of the cell cycle, and are thus unlikely to be responsible for directly controlling the postmitotic features that define motor neuron identity. What then are the molecular determinants of acquisition of postmitotic motor neuron identity?

Upon exit from the cell cycle and terminal differentiation the postmitotic identity of all spinal cord motor neurons is defined by the expression of a specific combination of homeodomain transcription factors, in particular Hb9, Isl1/2 and Lhx3/4 (Figure1.1C) (Dasen and Jessell, 2009; Tanabe et al., 1998; Tsuchida et al., 1994). Hb9 is a transcription factor whose expression in the nervous system is almost exclusively restricted to spinal motor neurons (Arber et al., 1999; Pfaff et al., 1996; Thaler et al., 1999). Interestingly, despite being the most specific marker of motor neurons in the spinal cord, Hb9 is not required for their initial specification. Instead, it has been shown to be important for the stable maintenance of motor neuron identity. Hb9 mutant mice exhibit normal birth timing and generation of motor neurons, but gain ectopic expression of the transcription factor Chx10, normally expressed only in V2 interneurons arising from the p2 domain. Additionally, motor neuron topographic organization within the spinal cord and their axonal projections outside of the spinal cord are affected (Arber et al., 1999;

Thaler et al., 1999). These studies implicate Hb9 as functioning primarily as a transcriptional repressor. This model is supported by a study in chick where ectopic expression of MNR2, a chick homolog of Hb9, in the developing spinal cord was sufficient to suppress the generation of V2 interneurons (Tanabe et al., 1998). Taken together, these studies show that despite not being required to generate motor neuron identity, Hb9 is integral to the normal progression of specification, including the suppression of genes associated with alternative spinal interneuron fates. Thus, Hb9 may primarily act to refine and consolidate motor neuron identity in concert with activators that function to induce the expression of effector genes. Given that Hb9 does not act to directly specify motor neuron identity, what then are the transcription factors more directly responsible for motor neuron specification?

Given their expression pattern, the LIM homeodomain (Lim-HD) containing transcription factors Isl1/2 and Lhx3/4 are likely at least partially responsible for the acquisition of the generic phenotypic properties of motor neurons (Ericson et al., 1992; Tsuchida et al., 1994). Isl1 expression is apparent soon after the final mitotic division of motor neuron progenitors and precedes the acquisition of other characteristics of motor neuron identity, including their patterns of axon projection and formation of columnar identity. Additionally, Isl1 is initially expressed in all motor neurons independent of columnar organization within the spinal cord (Ericson et al., 1992; Pfaff et al., 1996; Tsuchida et al., 1994). Together these data implicate Isl1 as a major regulator of motor neuron identity. Fittingly, loss of function studies in both mouse and chick embryos revealed that Isl1 is indeed required for the specification of motor neurons. Mice mutant for Isl1 fail to generate motor neurons in the spinal cord and lack axonal projections from the ventral roots, a hallmark of motor neuron identity. Interestingly, these studies also revealed a loss of expression of other motor neuron specific transcription factors such as Hb9, Isl2, and Lhx3, placing Isl1 at the top of a regulatory hierarchy of motor neuron identity (Pfaff et al., 1996).

The role of Isl1 as the major activator of the spinal motor neuron program and Hb9 as an apparent repressor of alternative programs of spinal neuron fate seemingly covers both the activation and repression programs required to accurately specify and maintain motor neuron identity. What role then do Lhx3/Lhx4 play in the motor neuron program? In the static view of motor neuron development, Lhx3 was initially observed to be expressed in a subset of motor neurons in the median motor column (MMC) targeting axial musculature, suggesting it might play a subtype-specific role. It is worth noting that its

expression was also observed in a domain dorsal to motor neurons where the cell bodies of V2 interneurons reside (Tsuchida et al., 1994). However, lineage-tracing experiments in mice revealed that Lhx3 and its homolog Lhx4 have a surprisingly dynamic expression pattern. Early during motor neuron generation, both Lhx3 and Lhx4 are expressed in all developing motor neurons in the neural tube. Following motor neuron birth, both factors are downregulated in all but one motor column, the median motor column. The dynamic expression of Lhx3 during motor neuron specification suggests that Lhx3 might play dual roles during motor neuron specification, once during initial specification and again in a subtype-specific manner, providing the first evidence that the process of motor neuron specification is perhaps highly dynamic even after exit from the cell cycle and terminal differentiation (Pfaff et al. 1996; Sharma et al. 1998; Tsuchida et al. 1994).

Together, these studies have uncovered, on a genetic level, the transcriptional programs that control specification of a generic postmitotic motor neuron identity. To summarize - during motor neuron specification, the pMN progenitor domain is established by Olig2 expression, followed by induction of proneural transcription factors such as Ngn2 acting to drive cell cycle exit and acquisition of postmitotic fate. Finally, the postmitotic motor neuron identity is driven by expression of specific combinations of homeodomain transcription factors, including Hb9, Isl1/2, and Lhx3/4. While progenitor specific transcription factors such as Olig2 are important for generating discrete domains that restrict the type of neurons that are born, they are unlikely to actively drive the specific features, which define the postmitotic neuronal identity. Instead, it is the combination of a somewhat generic proneural program driven by Ngn2 working and the postmitotically expressed Isl1 and Lhx3/4 to induce a generic motor neuron identity (Dubreuil et al., 2002; Mizuguchi et al., 2001; Novitch et al., 2001; Scardigli et al., 2001).

These two seemingly separate pathways are activated at the time of motor neuron birth suggesting a potential synchronization between acquisition of generic neural identity and motor neuron specification. The most convincing support of this idea comes from a set of overexpression experiments in the developing chick neural tube. Overexpression of both Isl1 and Lhx3 was sufficient to induce ectopic generation of motor neurons in dorsal progenitors. Interestingly, Isl1 or Lhx3 alone had little to no ability to generate ectopic motor neurons suggesting the combinatorial requirement for both factors to generate motor neurons (Lee and Pfaff 2003). Interestingly, co-electroporation of any of the proneural factors

normally expressed in motor neurons including Ngn2, NeuroD or NeuroM greatly enhanced the ectopic generation of Hb9-expressing motor neurons. These experiments established the sufficiency of the three factors for the expression of motor neuron effector genes (Lee and Pfaff 2003). Additionally, they suggest that proneural factors are able to functionally synergize with LIM-HD proteins to efficiently induce specific characteristics of motor neuron identity. However, it isn't clear at what level the synergism is accomplished. For example, does Ngn2 activate a parallel pathway to Isl1 and Lhx3, which simply increases the efficiency of Hb9 induction, or do they functional synergize on the same pathway?

Initial investigations suggested that, at least in the case of Hb9 expression, the proneural and LIM-HD factors functionally converge. Genetic dissection of an enhancer for Hb9 identified closely spaced LIM motifs recognized by Isl1 and Lhx3 and E-box motifs recognized by NeuroM. Chromatin immunoprecipitation (ChIP) revealed that both sets of factors bind and activate expression of an enhancer reporter construct and when co-expressed they functionally synergize to drive expression form the Hb9 enhancer. Together the authors concluded that there is functional synchrony between proneural and homeodomain transcription factors to efficiently drive motor neuron specification. While these studies were incredibly insightful and uncovered the basic transcriptional requirement for acquisition of generic motor neuron identify they nevertheless, raised multiple questions about how this generic identity is generate. Interestingly, during development Ngn2 protein expression is highly dynamic and never actually overlaps with Isl1 protein expression suggesting regulation of the Hb9 enhancer might occur in a dynamic fashion with sequential waves of activation. Additionally, given it's functional importance to motor neuron identity, it is not clear whether this functional synergy is a common principle of regulation or a specialized mechanism to robustly activate expression of Hb9 (Arber et al. 1999; Lee and Pfaff 2003; Lee et al. 2004).

Dynamic combinatorial expression of transcription factors in motor neurons

The studies of LIM-HD expression patterns and function in the spinal cord have uncovered several interesting findings. Following the acquisition of generic motor neuron identity, including cholinergic neurotransmitter identity and ventral projection of axons, expression of the programming factor Lhx3 is rapidly downregulated in all motor neurons except those that make up the median motor

column (MMC), and project to the axial musculature. Importantly, despite the loss of Lhx3 expression, all motor neurons maintain their recently acquired generic motor neuron identity and acquire unique subtype specific identities (Sharma et al. 1998; Tsuchida et al. 1994). These studies reveal a surprising concept with respect to neuronal identity, that despite becoming postmitotic, neurons maintain a dynamic regulatory environment, which acts to facilitate the acquisition of subtype-specific characteristics to increase cellular diversity within the motor neuron lineage.

Indeed, the dynamic expression of Lhx3/4 in motor neurons suggests a potentially dynamic function of transcription factors during neuronal specification, with both a generic early role and subsequently a more subtype-specific role. This concept has been directly investigated in loss of function experiments in which mice deficient for both Lhx3 and Lhx4 were observed to generate normal numbers of motor neurons yet these motor neurons displayed substantial defects in their axonal projections. Rather than projecting through ventral exit points in the spinal cord, motor neuron axons exited more dorsally. Interestingly, this projection pattern is consistent with dorsally projecting visceral motor neurons, which lack expression of Lhx3 and Lhx4 during specification, suggesting that at least one role played by Lhx3/4 is the specification of generic ventral motor projections. Consistent with this, ectopic expression of Lhx3 in the rostral neural tube, where dorsally projecting motor neurons normally arise, was sufficient to reroute axons to project ventrally. Additionally, when Lhx3 was ectopically expressed in non-MMC motor neuron subtypes, motor neuron axons were subsequently rerouted to phenocopy MMC projections to target axial muscles (Sharma et al. 1998; Sharma et al. 2000).

This dynamic reorganization of combinatorial expression of LIM-HD transcription factors is not restricted to Lhx3/4, but is a common theme within motor neuron subtypes (Figure 1.2). In most cases the downregulation of Lhx3/4 coincides with the upregulation of a different LIM-HD transcription factor. However, in hypaxial motor column (HMC) motor neurons no such LIM-HD has been identified suggesting there could be even more complex combinatorial logic than a simple combination of LIM factors (Figure 1.2C) (Dasen, 2009). In the most extreme case, the lateral subset of limb projecting lateral motor column (LMCl), motor neurons lose both Lhx3 and Isl1 expression shortly after their postmitotic transition (Figure 1.2D). This is a remarkable observation considering these factors are both induced postmitotically and are both necessary and sufficient for generation of the generic motor neuron program. The extinguished

expression of Isl1 and Lhx3/4 in LMCl motor neurons is accompanied by induction of Isl2 and Lim1 (Lhx1) transcription factors, the latter of which acts to control the projection of axons into the dorsal limb (Figure 1.2D) (Dasen 2009; Kania, Johnson, and Jessell 2000). Together, these studies beautifully exemplify the mechanisms by which the dynamic combinatorial expression of transcription factors can robustly generate cellular diversity within a single class of neurons, even while retaining generic motor neuron properties such as cholinergic neurotransmitter identity and ventral axonal projections outside the central nervous system. Based on these studies of LIM-HD expression in postmitotic motor neuron subtypes a so-called "LIM code" has been proposed whereby the specific combination of LIM-HD factors expressed acts to define specific properties of motor neurons, particularly the axonal projection patterns establishing motor neuron connections to their appropriate muscle targets (Dasen, 2009; Kania and Jessell, 2003; Kania et al., 2000).

Overall, these studies have cemented the idea that motor neuron identity is defined by a molecular code under the control of specific combinations of transcription factors. These transcription factors function (1) in a combinatorial manner, (2) are dynamically expressed even in postmitotic neurons, and (3) control both generic and subtype specific aspects of motor neuron identity. Yet on a mechanistic level it remains almost entirely unclear how these transcription factors work to establish motor neuron identity (Dasen, 2009). Interestingly, even neurons of different identity share the expression of the motor neuron LIM-HD transcription factors, suggesting these regulatory principles are potentially far reaching. For example, Lhx3 is also initially expressed in all V2 spinal interneurons then maintained in a subpopulation of these cells. Additionally, Isl1 is expressed in sensory neurons as well as a more dorsal population of interneurons in the spinal cord (Bui et al., 2013; Dasen, 2009). However, this introduces a conundrum whereby the same transcription factors maintain cell type specific identity in motor neurons as well as non-motor neuron types.

Combinatorial expression of motor neuron programming factors in other neuronal cells

How does overlapping expression of transcription factors control distinct neuronal identities? One can imagine two distinct mechanisms. Either the factors act on the same set of genes in each cell type or

they could control distinct sets of genes dependent on their cellular context. In other words, they either perform context-independent or context-dependent regulation of gene modules in each type of neuron.

The first insights into the molecular mechanisms that govern the cell type specific actions of LIM-HD factors in the spinal cord comes from a biochemical study of Lhx3 complexes in motor neurons and V2 interneurons (Thaler et al., 2002). Despite their prominent role in motor neuron specification, the expression of LIM-HD factors is not restricted to motor neurons. As mentioned previously, Lhx3/4 are also highly expressed in V2 interneurons, which develop from the p2 progenitor domain just dorsal to motor neurons, but lack expression of Isl1 (Ericson et al., 1992; Sharma et al., 1998; Tsuchida et al., 1994). Thus, the authors proposed a simple LIM code defined by the presence or absence of Isl1. This model is supported by overexpression studies whereby Lhx3 alone is sufficient to induce ectopic V2 specification while overexpression of Lhx3 and Isl1 in combination induces the expression of a motor marker (Tanabe et al., 1998). Indeed, misexpression studies in the developing chick neural tube in combination with *in vitro* biochemistry revealed that Lhx3's context-dependent protein-protein interactions acts drive specific cell fates. In particular, in V2 interneurons, Lhx3 forms a tetrameric complex with the LIM accessory factor NLI, in which 2 NLI factors bind with 2 Lhx3 factors. In contrast, expression of Isl1 in motor neurons serves as a molecular switch to form a hexameric complex with Lhx3 and NLI in a 2:2:2 ratio. These in vitro studies were further supported in the primary sequence discovered within distinct enhancers that respond specifically to the tetramer complex of V2 interneurons or the hexamer complex of motor neurons. Together, these experiments suggest a model, at least in the few isolated examples tested that facilitate cell type specific control of transcription factor activity dependent on cofactor binding to determine affinity for specific DNA sequences around cell type specific genes.

These studies, conducted over the past few decades, have uncovered a remarkable amount of insight regarding the molecular logic of cell type specification in the ventral spinal cord. Interestingly, they uncover a complex and dynamic transcriptional regulatory code that involves combinatorial expression of transcription factors to impart distinct cell identities despite shared expression of transcription factors. Furthermore, factors which initiate and are required for postmitotic identity display dynamic extinction of expression, suggesting a somewhat counterintuitive and poorly appreciated model that challenges the classic idea that cell identity is fixed when progenitors exit the cell cycle (Fishell and Heintz, 2013).

Instead, specification and maintenance of postmitotic cell identity might take place in a dynamic regulatory environment despite the outwardly stable state of a postmitotic neuron.

Regulatory control of postmitotic neuron identity

The studies described so far have been instrumental in our understanding of developmental signaling pathways that induce transcription factor combinations to encode molecular programs that generate specific cell fates in the nervous system. However, the underlying mechanisms by which these transcription factors control cell fate specification remain poorly understood. Concomitant with exit from the cell cycle, postmitotic neurons express distinct batteries of genes to acquire distinct biophysical, morphological and anatomical properties. In general, most neurons express overlapping sets of these genes, including axon guidance molecules, neurotransmitter synthesis enzymes, neurotransmitter receptors, synaptic components and ion channels. Thus, it is the combinatorial expression of distinct sets of these genes, which define a specific neuronal identity.

Given the cellular complexity within the nervous system and the dynamic nature of specification and subtype diversification as described previously, a mechanism needs to exist to facilitate specific control of these identity genes in an immense number of different cell types. One could imagine two extreme regulatory mechanisms. In the first, each gene or set of genes associated with a specific aspect of neuronal identity would be under the control of the same transcription factor(s) in all cell types. In this scenario, the genes encoding the components of the cholinergic biosynthesis pathway would be regulated in the same way in cholinergic motor neurons of the spinal cord and cholinergic interneurons in the basal ganglia. That is, a specific transcription factor or combination of transcription factors would regulate the cholinergic gene modules in all cholinergic cells. This would be extended to all genes so that each cell would have a large battery of transcription factors to specifically control expression of their respective effector genes (Figure 1.3A). Alternatively, in a second scenario, there could be a set of master regulator transcription factors that control all or most genes in a given cell type. In this example, cholinergic motor neurons and striatal interneurons would each have their own unique programming transcription factor code that would regulate expression of both cholinergic pathway genes shared by the two cell types as well as all other genes that make up their unique identity (Figure 1.3B). Based on these

two extreme models, understanding the specification and control of the identity of given neuron is dependent on understanding the regulatory mechanisms that control expression of its specific gene expression program. What then is the regulatory organization that facilitates both specific and broad expression of genes within the diversity of the nervous system?

Lessons learned from a simple nervous system

Over the years, model systems with simpler nervous systems have provided great insight into the regulation of neuronal identity and diversity (Hobert, 2011; Pearson and Doe, 2004). In particular, *Caenorhabditis elegans*, which has a relatively simple nervous system whose entire cellular composition of only 302 neurons has been thoroughly mapped, has provided a remarkably robust system for studying gene regulatory mechanisms controlling cell type specific gene expression programs. This has been achieved primarily through classic “promoter bashing” techniques of transgenic GFP reporters driven by neuron type specific genes. The application of this approach to a diverse set of neurons has uncovered a vast set of cis-regulatory elements and their regulators that are involved in the regulatory logic of neuronal identity.

For example, in cholinergic motor neurons in *C. elegans*, the genes that make up the entirety of the cholinergic identity, including acetylcholine biosynthesis, transport, and degradation, all contain at least one, if not more, instances of a shared motif in their *cis*-regulatory regions. This motif is for the COE-type transcription factor (Collier, Olf, Ebf), which was identified to be Unc-3. Unc-3 controls all aspects of cholinergic identity, as well as some other features of motor neuron identity (Kratsios et al., 2012). In another example, dopaminergic identity was dissected and a similar result was observed. In particular, the genes encoding components of the dopamine synthesis pathway, as well as dopamine transporters all share a common *cis*-regulatory motif for ETS family member Ast-1. Interestingly, the Ast-1 homolog, Etv1, was tested in mice and shown to also regulate at least some aspect of dopaminergic identity (Flames and Hobert, 2009). Together, these studies and others have uncovered a general principle of *cis*-regulatory control in *C. elegans*. Genes associated with a given neuron type tend to share a common *cis*-regulatory signature composed of at least one but more often multiple copies of motifs for the same trans-acting factors, suggesting that within a given type of neuron genes are co-regulated by a common

transcription factor or combination of transcription factors (Hobert, 2011). However, many of these studies are centered on neurotransmitter identity and therefore do not necessarily distinguish between the two models described previously. That is, it is not entirely clear whether each neuron type uses the same transcription factors to control neurotransmitter identity or if different cell types use different factors.

To distinguish between these two models, we can take advantage of the fact that worms, similar to mice, have multiple distinct neuron types that use cholinergic neurotransmitter providing a simplified test of the model of regulation. Interestingly, the evidence from *C. elegans* suggests a general model of cell type specific co-regulation. Each neuron type tends to have its own set of transcription factors, which control gene batteries in a coordinated fashion. For example, as mentioned earlier, *unc-3* controls cholinergic identity of ventral nerve cord motor neurons, while a combination of *ttx-3/ceh-10* controls the cholinergic identity of AIY interneurons. This cell type specific control is achieved through utilization of distinct binding motifs in the proximal regulatory region of the cholinergic genes. Additionally, *unc-3* and *ttx-3/ceh10* were shown to regulate the expression of additional genes that distinguish motor neuron and interneuron identity in *C. elegans* (Kratsios et al., 2012; Wenick and Hobert, 2004). This elegant series of genetic studies in *C. elegans* shed light on the mechanisms by which postmitotic cell identity is controlled in the nervous system. Together, they support a model whereby the particular features of a given postmitotic neuron are co-regulated by a shared cis-regulatory signature under the control of a common trans-acting factor(s), coined as a “terminal selector”. In particular, these cis-regulatory regions are composed of a series of recognition motifs to allow for cell type specific control of broadly expressed factors by specific combinations of terminal selector transcription factors.

The studies discussed thus far provide insight into how a small subset of genes is regulated within the simple nervous system of *C. elegans*. The experimental accessibility of *C. elegans* allowed for exhaustive dissection of *cis*-regulatory regions that are nearly impossible in mice. Control of neuronal identity genes, particularly neurotransmitter identity, is under the combinatorial control of transcription factors in a cell type specific manner. However, the relative simplicity of the *C. elegans* nervous system begs the question how neuronal identity in a much more diverse nervous system like that of mice is regulated. Above I discussed a series of studies that suggest mammalian neurons maintain a dynamic regulatory environment even after becoming postmitotic. The dynamic combinatorial expression of

programming transcription factors in motor neurons could potentially work in a similar fashion to that or worms, by utilizing different combinations of *cis*-regulatory motifs around neuronal identity genes. However, given the complexity of the mammalian nervous system there is likely a more complicated regulatory logic to specifically control gene expression programs during development. In the next section I will discuss recent progress in understanding mammalian gene regulation that will provide insight into how this specificity in gene expression might be obtained.

B. Enhancer control of gene expression

In the previous section I discussed a series of genetic studies that historically have been used to exemplify how transcription factors are able to "program" functional properties of cell identity in the nervous system, with a particular focus on spinal motor neurons. These functional transformations of cell identity are driven by alterations of core gene expression networks under the control of programming transcription factors (also called master regulators), yet the mechanisms by which these factors induce such transformations to specify properties of individual nerve cells remain poorly understood. Additionally, I discussed the confounding findings that the same transcription factors are often used in a highly dynamic fashion and by many different cell types, both functionally related and unrelated. This raises an interesting question of what mechanisms enable commonly expressed transcription factors to induce unique gene expression programs in different cell types and at the same time maintain specificity in cell specific gene expression programs.

Cell type specification and differentiation during development is a dynamic process requiring precise spatiotemporal control of gene expression programs. Indeed, gene expression is a highly coordinated process, which is under the control of multiple distinct regulatory elements spanning a large region of genomic space that includes core promoters, promoter-proximal enhancers, promoter-distal enhancers and insulator elements (Gorkin et al., 2014; Levine et al. 2010; Shlyueva et al., 2014). In eukaryotes, enhancers are thought to be the principle drivers of cell specific gene expression programs. Understanding enhancer elements is of great importance as they are thought to play an essential role in normal development, evolution and disease. Despite great interest in recent years, our global

understanding of enhancers remains rudimentary. In this section I will review what is known about the cell type specific control of gene expression in mammals by transcriptional enhancers. In particular, I will first describe fundamental concepts and mechanisms of enhancer control of gene expression, then discuss the dynamic process of enhancer activation during development, and finally recent insight into higher order chromatin organization and the mechanisms that control specificity in enhancer-promoter communication to activate specific patterns of gene expression.

A brief overview of mammalian enhancers

Enhancers are cis-regulatory elements first described more than 30 years ago when a short 72-base pair DNA sequence from the SV40 viral genome was shown to drive transcription of a reporter gene in mammalian cell lines. Shortly after this, the sequence specific transcription factors AP1 and NF-kB were shown to bind to discrete locations within the SV40 enhancer in order to synergistically activate gene expression (Banerji et al., 1981; Benoist and Chambon, 1981; Gruss et al., 1981; Lee et al. 1987). These pioneering studies were the first to dissect the general mechanisms of enhancer control of gene expression, uncovering key general principles that can be extended to all enhancers (Levine, 2010; Shlyueva et al., 2014). In particular, enhancers are generally thought to share the following set of key features: (1) they are relatively short pieces of DNA sequence, usually spanning from ~100 base pairs up to one kilobase in the genome; (2) they harbor DNA binding motifs for multiple sequence-specific transcription factors, including both activators and repressors which recruit co-activator and co-repressor complexes, respectively; and (3) they function in isolation to recapitulate gene expression patterns matching those of the endogenous gene expression independent of distance or orientation to their target genes (Levine, 2010; Shlyueva, Stampfel, & Stark, 2014).

Interestingly, in mammals, these enhancer elements are located throughout the genome, ranging from a few thousand to millions of base pairs from their target genes making their identification and functional dissection daunting (Gorkin et al., 2014; Hnisz et al., 2016). They are enriched in proximally spaced sequence motifs, functioning as platforms for the binding of sequence specific transcription factors. Transcription factors govern enhancer activity by binding in a combinatorial fashion to local regions of DNA accessibility and recruiting generic co-activator complexes and histone modifiers (Calo

and Wysocka, 2013; Roeder, 2005). These co-activator complexes include, but are not limited to, histone acetyltransferases like p300/CBP, which function to acetylate histone tails to facilitate increased transcription factor binding and enhancer activity; and the ~30 subunit mediator complex, which promotes long-range "communication" between distal enhancers and the pre-initiation complex (PIC) at the promoter of target genes. It is worth noting that RNA polymerase II is also present at enhancers, but its role is still controversial and poorly understood (Heintzman et al., 2007; Kagey et al., 2010a; Kim et al., 2010b; Zhang et al., 2013). Interestingly, the binding of all of these cofactors are independent of DNA sequence, supporting a model that their recruitment is dependent on the transcription factors bound.

In recent years, immense technological advances in high-throughput genomics have facilitated progress towards generating nearly comprehensive maps of putative enhancers throughout the genome. By taking advantage of the known characteristics of enhancers described above, new genome-wide applications are theoretically able to map the distal locations of all putative enhancers in the genome. In large part due to the Encyclopedia of DNA Elements (ENCODE) consortium, the locations and patterns of activity for hundreds of thousands of enhancers have been identified in a series of tissues and cell types (Heintzman et al., 2007, 2009; Shen et al., 2012; Visel et al., 2009). Interestingly, these datasets indicate that the numbers of distal regulatory elements in the genome greatly outnumber the number of identified genes. The majority of these experiments were performed in whole tissues, which on their own are composed diverse cell types. This raised multiple questions, including which enhancers are active in a given cell type at a given point in time? How are the patterns of cell type specific enhancer activity regulated to control specific gene expression patterns?

Transcription factor binding at enhancers

Fundamentally, enhancer activity is under the control of transcription factor binding. Each transcription factor recognizes its own relatively conserved short (6-12bp) consensus binding motif and thus if a motif for a given factor is known the entire genome can be scanned and one can computationally predict all possible locations that a factor might occupy. This conceptually simple idea is enticing in principle, but it tends to work only in limited genomic regions. Considering the fact that transcription factor binding motifs are relatively short and somewhat degenerate when scaled to the entire genome these

methods can predict up to millions of binding sites. For example, if we assume a transcription factor recognizes a strict 6 base pair motif that occurs randomly then it would be predicted to occur ever 4096 (4^6) base pairs or more than 700,000 times in the genome, making it nearly impossible to know which motifs are relevant. In fact, individual enhancers often contain an array of transcription factor binding motifs, suggesting that there are likely more complex rules for transcription factor recruitment and occupancy, and by extension enhancer activity, than a simple model of individual transcription factors binding to their cognate motifs. Indeed, previous studies indicate that in general, transcription factors work in a combinatorial manner with additional factors (Levine, 2010; Tjian and Maniatis, 1994). However, mechanistic understanding of how different transcription factors work in concert and the contribution of each factor provides remains rudimentary.

The combinatorial mechanisms of transcription factor function can be somewhat inferred from primary enhancer sequence. For example, enhancers often contain motifs for multiple transcription factors that are spaced and orientated in specific patterns, which can inform potential combinations of regulators of activity. In addition to motif composition, motif positioning within an enhancer also plays an important role transcription factor function. The specific presence and location of combinations of motifs within an enhancer facilitates optimal spacing of transcription factors to promote protein-protein interactions and cooperative binding. Together the motif makeups of enhancers support a model combinatorial of transcription factor binding to enhancers, which results in the formation of large multi-protein complexes to generate synergistic activity. Moreover, this combinatorial and synergistic model of DNA binding at enhancers explains how only specific subsets of binding sites might be active.

The enhanceosome model at the interferon-β enhancer is perhaps the clearest example of these two concepts and their role in enhancer function. Eight transcription factors occupy a 55bp element in a highly structured manner (Panne et al., 2007; Thanos and Maniatis, 1995). Individual transcription factors are recruited in an ordered fashion to these enhancers, with binding of the first transcription factor promoting binding of the second and so forth, leading eventually to a highly ordered protein complex with switch-like activation of the IFN-β promoter (Thanos and Maniatis, 1995). This step-wise recruitment leading up to eventual activation of IFN-β expression by the enhanceosome provides a remarkably strict example of combinatorial transcription factor binding. Whether there is such a common level stringency in

activity remains poorly understood on a global level. Never the less, combinatorial binding seems to be a general principle of enhancer regulation and control of cell specific gene expression (Spitz and Furlong, 2012; Whyte et al., 2013; Young, 2011).

Over recent years mouse embryonic stem cells have provided a robust system for studying transcription factor control of cell identity. The transcription factors important for maintenance of ESC identity in pluripotency have been identified and studied extensively. Oct4, Sox2, and Nanog have been identified as the master regulators of ESC identity. Genome-wide binding of the three factors revealed they collaborate to regulate a large majority of genes that are expressed specifically in ESCs (Chen et al., 2008; Kagey et al., 2010b; Marson et al., 2008; Whyte et al., 2013). These three factors co-occupy thousands of binding sites throughout the genome. These co-occupied genomic regions are highly enriched for enhancer-associated co-activators including the HAT, p300 and mediator complexes. Additionally, the OSN factors form sub complexes with many other factors that are suggested to regulate specific subsets of genes important for ESC identity (Chen et al., 2008; Kim et al., 2008; Young, 2011). Together, these data support a global model of synergistic binding of master transcription factors to generate large multi-protein enhanceosome-like complexes at enhancers to control cell specific gene expression programs.

The dynamic activation of the IFN-B enhancer and the global binding patterns of OSN in regulating the mESC expression program introduce an interesting conundrum. Evidence from the regulation of IFN-b suggests that the synergistic assembly of the enhanceosome happens in an ordered fashion. However, experiments studying OSN binding in ESCs provide a static view of enhancers at a single point in time and reveal the combinatorial binding of the large OSN containing complexes, but information on the dynamics of complex assembly and enhancer activation are lacking. These studies introduce a few interesting questions. How do these developmental enhancers form in time? Do they follow the same step-wise principles as the enhanceosome? For example, in the context of motor neuron specification a dynamic wave of combinatorial transcription factor expression drives the activation of specific effector gene expression programs. Starting from Ngn2 expression to activation of Isl1 and Lhx3 in nascent motor neurons and finally dynamic changes in LIM-HD factor expression in specific motor

neuron subtypes. Could there be a hierarchical sequence of transcription factor recruitment to the core motor neuron enhancers to activate and maintain expression of cell identity genes?

Dynamic activation of enhancers during development

The general characteristics of enhancers provided by the studies discussed above have provided a static view of enhancers in time. However, at any given point in time during a development an enhancer exists in any one of the following states, silent, poised, primed or active (Figure 1.4). These different states of enhancers are dependent on the ordered activity of combinations of transcription factors bound at a specific point in time. Enhancers do not act simply as binary switches, but instead progress through an ordered set of events that are under the control of multiple steps of transcription factor activity. Below I will start with enhancers in the silent state and discuss the events that lead to their progressive activation.

In its silent state an enhancer is usually in relatively inaccessible nucleosome occupied chromatin and free of transcription factor binding (Figure 1.4A). During development binding of so-called “pioneer” transcription factors must first make the enhancer accessible (Figure 1.4A). Transcription factors with pioneering activity were first described more than 30 years ago in yeast and subsequently in mammals years later (Almer et al., 1986; Spitz and Furlong, 2012; Zaret and Carroll, 2011). Pioneer factors are able to bind their DNA motifs in the context of the nucleosome, which generally are inaccessible to most transcription factors. Binding of the pioneer transcription factor induces local chromatin remodeling events to reposition nucleosomes that increases local DNA accessibility. By increasing local accessibility pioneer factors along with chromatin remodeling enzymes expose secondary DNA motifs to facilitate cooperative transcription factor binding (Cirillo et al. 2002; Ghisletti et al. 2010; Zaret and Carroll 2011). For example, FOXA1, a classic example of a transcription factor with pioneer activity, is able to bind its target motif within the context of compact nucleosome-associated DNA and promote chromatin remodeling and increase the accessibility of the associated enhancer during development (Lee et al. 2005; Zaret and Carroll 2011).

Interestingly, binding of these pioneer factors is often not sufficient to generate enhancer activity or influence gene expression (Ghisletti et al., 2010). Instead, they primarily play an essential role in early establishment of the enhancer and facilitate the binding of future transcription factors. This suggests that

they play an initial role to facilitate future combinatorial binding at enhancers. Indeed, recent studies on the mechanisms by which Oct4, Sox2, Klf4 and Myc (OSKM) factors are able to reprogram fibroblasts to induced pluripotent cells have supported this model. Oct4 and Sox2 display much higher affinity for nucleosome-associated motifs than Klf4 and Myc indicating they might have more pioneer like activity and potentially facilitate Klf4 and Myc binding (Soufi et al., 2012, 2015). Together, these studies provide strong evidence for the role of pioneer transcription factors in the establishment of enhancers throughout the genome during development. Importantly, not all transcription factors exhibit pioneer activity. Instead, the binding of pioneer transcription factors is seemingly acting to demarcate the locations of enhancers that potentially guide future binding of cofactors to create combinatorial output. By binding to their nucleosome protected motifs they inform the recruitment of generic histone remodelers in a cell type specific manner.

Following this initial establishment by pioneer transcription factors, enhancers are thought to go through a phase of priming where repositioned nucleosomes flanking local regions of DNA accessibility acquire specific modifications on histone tails. In particular, primed enhancers are flanked by nucleosomes that have a specific histone modification, monomethylation of lysine 4 of histone H3 (H3K4me1), the first enhancer associated histone modification identified by the ENCODE consortium in early genomic studies across a small portion of the genome (Figure1.4D) (Heintzman et al., 2007). H3K4me1 at enhancers seemingly acts as a priming signal as its presence precedes substantial nucleosome depletion or recruitment of co-activator complexes and activation associated histone modifications. In mouse embryonic stem cells a subset of developmental enhancers are “primed” and are marked by H3K4me1 modifications to demarcate putative enhancers for future activation (Creyghton et al., 2010; Rada-Iglesias et al., 2011). Together, these studies suggest that like the pioneering activity required to establish enhancers from inaccessible chromatin, H3K4me1 modifications are not directly linked to enhancer activity, but instead provide a facilitative environment that allows for binding of activating transcription factors and efficient recruitment of co-activator complexes (Calo and Wysocka, 2013; Zaret and Carroll, 2011). This suggests there are downstream steps, which take advantage of the primed state to drive activation of enhancers. Such a mechanism could potentially allow for dynamic

enhancer activity during differentiation and specification to output spatiotemporally precise patterns of gene expression.

Progression of enhancers from a primed to active state is dependent on the recruitment of co-activator proteins including histone acetyltransferases (HATs). Like proximal regulatory regions in yeast where Gcn5 is recruited to acetylate histone and facilitate transcription, in mammals p300 and the closely related CBP are recruited to distal enhancers throughout the genome by combinatorial binding of activating transcription factors (Ghisletti et al., 2010; Heintzman et al., 2009; Kim et al., 2010; Rada-Iglesias et al., 2011; Visel et al., 2009). Interestingly, CBP and p300 are ubiquitously expressed in all tissues supporting the idea that tissue specific combinations of transcription factor binding to enhancers provide specificity to generic co-activator activity. The main acetyltransferase activity of p300/CBP is to acetylate lysine 27 on histone H3 and acquisition of this specific histone modification distinguishes active enhancers from poised or primed enhancers marked only by H3K4me1 (Creyghton et al., 2010; Heintzman et al., 2009; Rada-Iglesias et al., 2011). Accordingly, genome-wide mapping of p300 and/or H3K27ac is able to predict the dynamic activity of enhancers *in vivo* with both tissue and temporal specificity (Figure 1.4B) (Nord et al., 2013; Visel et al., 2009).

Together, these studies demonstrate the dynamic nature of enhancer activation in developing tissues. Enhancers go through a step-wise process of activation, from inaccessible nucleosome-occupied and silent, established and accessible by pioneer transcription factors, primed by acquisition of H3K4me1, and finally, activated by p300 recruitment and H3K27ac. The underlying theme of all these steps is the requirement for sequential binding of sequence specific transcription factors, which impart specificity onto ubiquitously expressed histone remodelers and co-activators. While these studies focus on the activations of distal enhancers, they do not explain how enhancer activation induces expression of target genes. In the next section, I will discuss a series of studies that uncovered mechanisms by which distal enhancers control gene expression.

Transcription factor control of long-range enhancer-promoter interactions

A major differentiating feature of the vertebrate genome is the relatively large distances separating enhancers from their target genes. Nevertheless, enhancers are able to recapitulate

endogenous gene expression in a position-independent manner (Levine, 2010; Pennacchio et al., 2006; Spitz and Furlong, 2012; Visel et al., 2009). This property led to the early idea that enhancers act from a distance to activate transcription by physically interacting with their target promoters. It was long assumed that enhancers would regulate genes in their proximity; however mechanistic support was lacking.

Recent advances in the development of technologies that allow for the detection of chromosomal contacts between distally separated locations in the genome have begun to facilitate mapping of distal enhancers to their target genes. Chromosome Conformation Capture (3C) and its high-throughput derivatives each rely on a similar proximity-ligation based method to ligate pieces of DNA that are in close spatial proximity. 3C was the first method and allowed for mapping of interactions in a single region at time. Modifications of the method have sense been adapted to unbiased genome-wide mapping of all chromatin interactions (Hi-C) and to interactions occupied by a protein of interest (ChIA-PET) (Dekker, 2002; Fullwood et al., 2009). Hi-C and ChIA-PET are currently the two most widely used methods for mapping global enhancer-promoter interactions. In general, Hi-C is entirely unbiased and can identify all forms of chromatin interactions, while ChIA-PET is biased to only a subset of interactions. However, ChIA-PET tends to provide higher resolution maps of the subsets of interactions it identifies (Figure 1.5B) (Dekker, Marti-Renom, and Mirny 2013).

Taking advantage of these new methods and its high-throughput derivatives many distal enhancer-promoter interactions have been identified to support a model for enhancer-promoter communication in the activation of target gene expression. Long-range communication is generally intrachromosomal and proposed to work via the formation of chromatin interactions, which loop out intervening sequences in order to bring the enhancer and promoter in spatial proximity (Bulger and Groudine, 1999; de Laat et al., 2008). Interestingly, enhancer-promoter contacts have been detected between different chromosomes, but these findings seem to be the exception to the general rule of intrachromosomal enhancer-promoter organization (Apostolou and Thanos, 2008; Lomvardas et al., 2006).

Over the past several years, the mouse β-globin locus has a served as an excellent model for studying chromatin interactions and their role in gene expression. The locus control region (LCR), located approximately 50 kilobases upstream of the gene, loops to engage the promoters of β-globin locus in a

tissue and differentiation specific manner where it specifically interacts in erythroid cells, but not in other cell types (Carter et al. 2002; Tolhuis et al. 2002). Subsequent analysis has revealed that transcription itself is not required for loop formation, but looping is critical for induction of transcription as tethering of the LCR to the promoter was shown to be sufficient to induce transcription (Hnisz et al., 2016; de Laat and Duboule, 2013). These studies established the role of transcription factor-mediated looping in target gene activation, but has left open the as-yet unresolved questions of what actually controls and facilitates the looping of distantly located sequences. Furthermore, it was unknown if transcription factors alone promote looping or whether they coordinate looping with other proteins potentially acting as architectural or structural support.

Initial insights to these questions came from an short-hairpin RNA (shRNA) screen in mouse embryonic stem (ES) cells for regulators of transcription of the pluripotency gene Pou5f1 (Oct4) (Kagey et al., 2010b). Somewhat expectedly, the screen identified sequence specific transcription factors with established roles in pluripotency. However, more interestingly, multiple large protein complexes with essential roles in chromatin organization and gene expression including the cohesin and mediator complexes were also identified in this screen. In particular, cohesin has been shown to play a key role in sister chromatid organization during mitosis (Dorsett, 2011; Gorkin et al., 2014). Moreover, the mediator complex's role in transcriptional activation has been known for decades (Ge et al., 2002; Taatjes, 2002; Zhang et al., 2005). In this study, genome-wide maps of mediator and cohesin complexes revealed that they co-occupy both enhancers and promoters of actively transcribed genes with established roles in ES cell biology. The distal enhancers are co-occupied by and highly enriched for the master regulators of ESC identity, Oct4, Sox2, and Nanog suggesting that their recruitment is likely depending on the combinatorial binding of the sequence specific transcription factors. These results suggest the model that cohesion and mediator complexes act together as architectural proteins to control the formation of chromatin loops, with mediator acting as a bridge between the transcription factor bound enhancer and pre-initiation complex at the promoter while cohesin acts as a stabilizer of interaction. Additionally, this model suggests that both are required for efficient gene expression, a prediction confirmed by knockdown studies of individual subunits of the mediator and cohesin complexes (Figure 1.5A) (Kagey et al., 2010).

Interestingly, while the majority of mediator-bound regions are co-occupied by cohesin, not all cohesin-bound sites are co-occupied by mediator, suggesting additional roles for cohesin. Accordingly, cohesin co-occupies a large proportion of the binding sites of another structural protein, CTCF. Interestingly, the majority of these CTCF sites are not enhancers and lack mediator binding (Parelho et al., 2008; Wendt et al., 2008). Taken together, these studies suggest that cohesin seems to be generically involved in promoting chromatin interactions whose functions are dependent on the additional architectural proteins it binds with. This idea is supported by recent experiments where chromatin interactions were mapped at multiple loci in mouse embryonic stem cells and differentiated neural progenitors. Specific combinations of architectural proteins were associated with specific types of interactions. Dynamic cell type-specific interactions were facilitated by co-occupancy of mediator and cohesion while stable interactions shared between cell types were associated with CTCF and cohesion (Phillips-Cremins et al., 2013).

Together these studies uncovered the mechanism by which distally located enhancers are able to activate target gene expression. Combinatorial binding of transcription factors leads to recruitment of architectural proteins, including mediator and cohesion that facilitate long-range communication between distal regulatory elements and their target genes. Mediator dependent interactions are largely cell type specific and thus likely associated with cell identity genes. These results of these studies suggest that enhancers are able to specifically engage in interactions of their target genes to activate expression despite their distal location. In the next section I will discuss current ideas about how such specificity is generated.

Topological domains and insulated neighborhoods

Considering the fact that enhancers are generally located far away from their target genes and often do not affect expression of the closest gene it is not currently clear how enhancers specifically activate their target genes (Gorkin et al., 2014; Montavon et al., 2011). Recent discoveries with regard to the basic organization of the genome have begun to uncover potential mechanisms underlying this process. Chromosomes are organized into repeating structural units called topologically associating domains (TADs) (Dixon et al., 2012; Nora et al., 2012). TADs were discovered from genome-wide

chromosome conformation (Hi-C) data where local regions in the genome were observed to have increased contact frequencies that are separated by sharp boundaries (Figure 1.6A). The mammalian genome contains approximately 2000 TADs with an average size of one megabase, varying from a few hundred kilobases to several megabases. TADs seem to be a fundamental unit of chromosome organization as they have been discovered in all mammalian (and many other organisms) cells tested. Moreover, the boundaries between individual TADs are highly conserved across cell types and thought to be relatively stable and in many cases bound CTCF, a transcription factor with an established role at insulators (Bell et al., 1999). Together these findings implicate TADs as an enticing method of restricting enhancer-promoter communication because the majority of DNA contacts occur within the same TAD (Dixon et al., 2012, 2015). Additionally, this also explains why, with a few exceptions, enhancer-promoter interactions tend to be intrachromosomal and within approximately a megabase. These initial experiments provide correlative support for insulation of enhancer-promoter interactions that could control specificity of gene expression however the mechanism by which TADs restrict interactions remained unclear.

The mechanistic basis of restricted interactions within TADs were recently uncovered with the discovery of “insulated neighborhoods” (Figure 1.6A, B). Based on the findings that cohesin plays an architectural role and a general facilitator of chromatin interactions, genome-wide maps of cohesin-associated interactions were mapped in mouse embryonic stem cells with ChIA-PET. By mapping cohesin-anchored interactions throughout the genome both stable CTCF-associated and dynamic mediator-associated interactions were recovered. Remarkably, CTCF-associated interactions were found to be the basis of TADs. Pairs of CTCF bound locations at the boundaries of a TAD act as anchors that are connected via cohesin (Figure 1.6A, B). The large CTCF-cohesin loops are proposed to form an insulated domain within which dynamic enhancer-promoter interactions are restricted (Dowen 2014, Ji 2016). Accordingly, in two separate studies, in mouse embryonic stem cells and T lymphocytes, ninety percent of enhancer-promoter interactions were restricted within these insulated domains. Finally, mutation of CTCF binding sites led to a decrease in endogenous target gene expression and coincided with an aberrant increase in expression of genes in neighboring domains. Importantly, this is not a mechanism that is restricted to mouse embryonic stem cells as alteration of the domain boundaries

around the protocadherin locus in both mouse brain and cell lines resulted in a similar phenotype (Dowen et al., 2014; Guo et al., 2015).

The studies discussed in this section reveal the immense complexity involved in specifically regulating gene expression in mammals. Enhancers must be dynamically activated in a temporally precise manner to activate gene expression of target genes. This dynamic enhancer activation is dependent on the expression of cell-specific combinations of transcription factors at a given point in time. Synergistic binding by these cell-specific combinations of transcription factors recruits co-activators, which in turn leads to activation of target promoters through long distance enhancer-promoter interactions within insulated regulatory domains. These fundamentals of gene regulation provide a basis from which we can investigate the regulatory mechanisms controlling cell identity in the nervous system. In the next section, I will describe the major questions, which I aim to address in this thesis and the approach I will take to do so.

C. Questions, Challenges and Approach

In the previous two sections I have reviewed years of work in two seemingly separate fields, neuronal specification and diversification and enhancer control of gene expression. Yet from these conceptually separate fields comes a unifying principle, combinatorial control by programming transcription factors. From a genetic perspective, specific combinations of transcription factors are able to impart unique properties of cell identity within the nervous system and similarly, specific combinations of transcription factors are able to generate specific patterns of enhancer activity. Together, this unifying concept provides a potential path to understanding the mechanisms controlling cellular identity and diversity in the nervous system.

While significant progress has been made in identifying the genetic requirement of specific combinations of transcription factors during neuronal specification, mechanistic understanding of how combinatorial codes of transcription factor activity actually control identity remains minimal. Given its cellular complexity, the mammalian nervous system has remained particularly elusive to mechanistic analysis of single neuronal cell types. Obtaining pure cell populations of sufficient numbers required for

biochemical analysis has kept the mechanistic basis of cellular diversity guarded from investigation. Considering the fact that the general principles of gene regulation are shared amongst cell types tested to date, access to a single neuronal cell type for mechanistic dissection would likely uncover the general principles of regulatory control of cell identity in the nervous system.

In recent years, significant advances in stem cell biology have begun to provide an entry point to study neuronal specification on a biochemical level. Directed differentiation of pluripotent stem cells into a diverse set of cell types throughout the nervous system have begun to provide a system to dissect the regulatory mechanisms controlling cell fate specification in the nervous system (Gaspard et al., 2009; Maroof et al., 2013; Peljto and Wichterle, 2011; Wichterle et al., 2002). Taking advantage of the well-established developmentally relevant patterning signals both mouse and human pluripotent stem cells can be directed to differentiate into specific neuronal cell types. Importantly, the process of directed differentiation has been shown to closely follow the temporal and molecular patterns of development in vivo. For example, the functional differentiation of a specific class of neuron was first described over 15 years ago when spinal motor neurons were successfully generated from embryonic stem cells (Wichterle et al., 2002). In addition to the acquisition of generic motor neuron identity, stem cells can be directed to differentiate into specific subtypes of motor neurons expressing distinct combinations of LIM-HD transcription factors previously discussed in this chapter (Peljto et al., 2010; Tan et al., 2016). For example, as opposed to the rostral cervical identity induced by RA and Shh mediated differentiation, ES-derived motor neurons can by coaxed into more caudal identities by modulating signaling with Wnt and FGF (Peljto et al., 2010). Additionally, by modulating notch signaling during differentiation motor neurons can be directed into HMC subtypes at the expense of MMC (Tan et al., 2016). These studies provided the first insight that in vitro-derived neurons could provide a model for studying specification in the nervous system in a relatively pure setting. Using stem cell differentiation as a model for development, a nearly unlimited supply of motor neurons can be generated to allow for biochemical dissection of regulatory control by specific combinations of transcription factors.

In addition to directed differentiation, direct cellular programming and reprogramming has emerged as a highly useful model for studying transcriptional control of cell identity. The concept was pioneered approximately 30 years ago when expression of the muscle transcription factor, MyoD, was

able to convert fibroblasts into cells that resemble muscle cells with elongated morphology (Davis et al., 1987). Years later, building upon the discovery of MyoD's ability to program specific features of muscle identity into a different cell type, a groundbreaking discovery was made that showed overexpression of a specific cocktail of transcription factors (Oct4, Sox2, Klf4, c-Myc (OSKM)) was sufficient to reprogram fibroblasts into induced pluripotent stem cells (iPSCs) that resemble mouse embryonic stem cells (Takahashi and Yamanaka, 2006). In a final example, taking a similar approach to iPSC reprogramming, it was shown that fibroblasts can be directly converted into "induced neurons" by the misexpression of a unique cocktail of transcription factors including a proneural transcription factor, Ascl1 (Mash1) and two factors expressed in postmitotic neurons in the brain, Brn2 and Myt1l (Vierbuchen et al., 2010). Together, these experiments beautifully exhibit the power of transcription factor misexpression as a model for studying control of cell identity and fate specification.

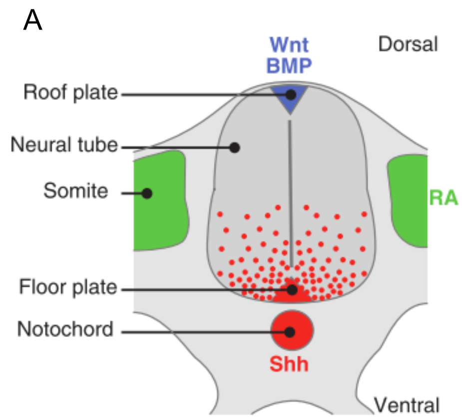
The history of programming cell types by misexpression of transcription factors have provided an initial feasibility for generation of desired cells of interest as well as providing a robust experimental paradigm for studying the mechanisms controlling specification of cellular identity. Directed differentiation and direct programming each provide their own unique advantages and caveats. In general, directed differentiation follows patterns of *in vivo* development allowing for the study of the dynamic temporal regulation by transcription factors. However, for mechanistic studies some experiments are difficult to interpret due to a lack of purity and temporal precision within the heterogeneous culture. In contrast, direct programming of one cell type to another by misexpression of transcription factors allows for the mechanistic dissection of only the factors of interest at a static point in time.

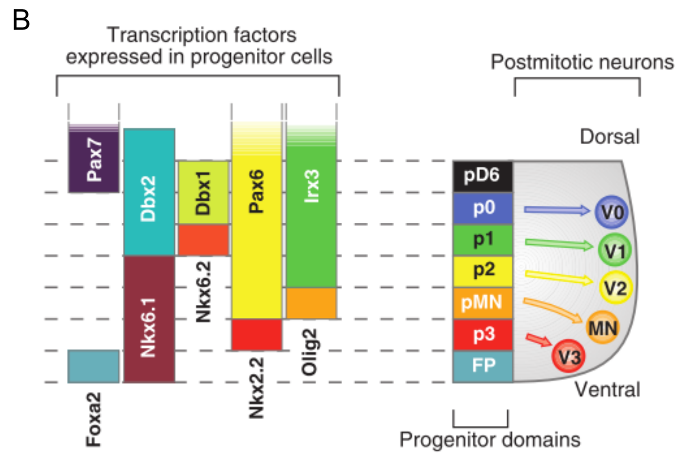
In this thesis, I will use motor neurons as proxy for understanding the regulatory mechanisms that control the activation and maintenance of specific gene expression programs in the mammalian nervous system. In order to study specification within the complexity of the nervous system one needs to first understand the basic principles of regulatory control of a given cell type. In other words, the best approach is to dissect neuronal identity one cell at a time. I take advantage of an approach integrating information from transcriptionally programmed, *in vitro* differentiated and *in vivo* purified motor neurons to dissect the regulatory logic of motor neuron identity. By combining three different sources of motor neurons we are strategically placed to perform in depth dissection of the general gene regulatory

principles controlling cell identity and diversity in the nervous. In particular, the questions I aim to address are: (1) How does the combinatorial expression of transcription factors specify distinct cell types in the nervous system? (2) What is the *cis*-regulatory organization that allows for specificity in gene expression within the cellular complexity of the nervous system? (3) What are the regulatory mechanisms that maintain stable neuronal gene expression programs in the face of dynamic changes in combinatorial transcription factor codes?

In Chapter 2, I describe our work to program motor neurons directly from mouse embryonic stem cells. Taking advantage of the transcription factor codes controlling motor neuron specification described in this introduction, we design transcriptional programming modules to directly specify motor neuron identity from mouse embryonic stem cells. Using developmentally inspired transcription factor codes we are able to robustly program two distinct motor neuron types with high efficiency. These cells resemble those of their *in vivo* counterparts and provide an entry point into studying the molecular mechanisms controlling cell identity in the nervous system. Initial results suggest that expression of distinct combinations of transcription factors informs context dependent genomic binding and can to some degree explain regulatory control of cell-specific and shared properties of motor neuron identity. Next, in Chapter 3, I take advantage of this robust motor neuron programming system to study the regulatory mechanisms, which control motor neuron fate specification. Using this unlimited supply of pure motor neurons, I investigate the chromatin and genomic mechanisms controlling motor neuron identity. We generate genome-wide maps of chromatin modifications, co-factor binding and enhancer-promoter interactions to directly identify the regulatory targets under the control of programming transcription factors. Interestingly, by studying a single cell type in isolation we are able to uncover a unique mode of regulatory organization around neuronal genes and make predictions about how gene expression is specifically controlled within the complexity of the nervous system. And finally, in Chapter 4, we take the mechanisms we learned from the previous chapters to ask what happens when motor neurons mature and downregulate the programming factors which initiate their specification. We find that enhancers remain highly dynamic even in post-mitotic motor neurons and the control cell identity genes is initiated and maintained by a dynamic relay of stage-specific transcription factors activating stage-specific enhancers. Altogether, the work

presented in this thesis reveal the gene regulatory mechanisms controlling the precise specification and maintenance of distinct cell identities within the complexity of the mammalian nervous system.


A
Wnt
BMP
Dorsal
Roof plate
Neural tube
Somite
RA
Floor plate
Notochord
Shh
Ventral


B
Transcription factors
expressed in progenitor cells
Postmitotic neurons
Pax7
Dbx2
Dbx1
Pax6
Irx3
Nkx6.1
Nkx6.2
Nkx2.2
Olig2
Foxa2
Dorsal
pD6
p0
p1
p2
pMN
p3
FP
V0
V1
V2
MN
V3
Ventral
Progenitor domains

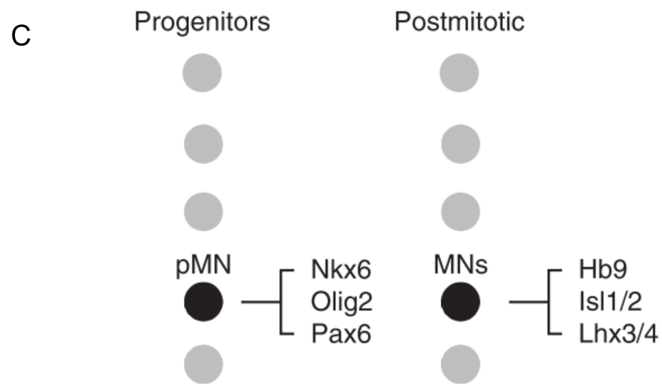

C
Progenitors
Postmitotic
pMN
Nkx6
Olig2
Pax6
MNs
Hb9
Isl1/2
Lhx3/4

Figure 1.1. Secreted signals regulate the spatial expression of transcription factors in the neural tube. **(A)** Within the spinal cord, functionally distinct neurons are generated in a spatially segregated manner in response to signals emanating from within the neural tube and surrounding tissue. The key signals include Shh (red), secreted by the notochord and floor plate; retinoic acid (RA, green), produced by the somites that flank the neural tube; and BMP and Wnt family members (blue), which are produced dorsally. The spread of Shh from ventral to dorsal establishes a gradient of activity within the ventral neural tube (red dots). **(B)** The ventral gradient of Shh activity controls position identity by regulating the expression, in neural progenitors, of a set of transcription factors. The differential response of these genes to graded Shh signaling establishes distinct dorsal and ventral boundaries of expression for each factor. The combinatorial expression of the transcription factors defines domains of progenitors (p). From the ventral pole, these are termed FP (floor plate), p3, pMN and p2-p0. Each progenitor domain is identified by its transcription factor code, and this code determines the neuronal subtype progeny the progenitors produce. Each progenitor domain generates different ventral (V) interneuron subtypes (V0-V3) or motor neurons (MN). **(C)** Motor neurons are generated from the motor neuron progenitor domain (pMN) in the ventral spinal cord. Progenitor and postmitotic motor neurons express distinct combinatorial code of transcription factors, predominantly homeodomain (e.g. Nkx6, Pax6, Hb9), basic helix-loop-helix (bHLH, e.g. Olig2) and LIM (e.g. Isl1/2, Lhx3/4) transcription factors. *A and B adapted from Dessaud and Briscoe (2008); C adapted from Dasen and Jessell (2009)*.

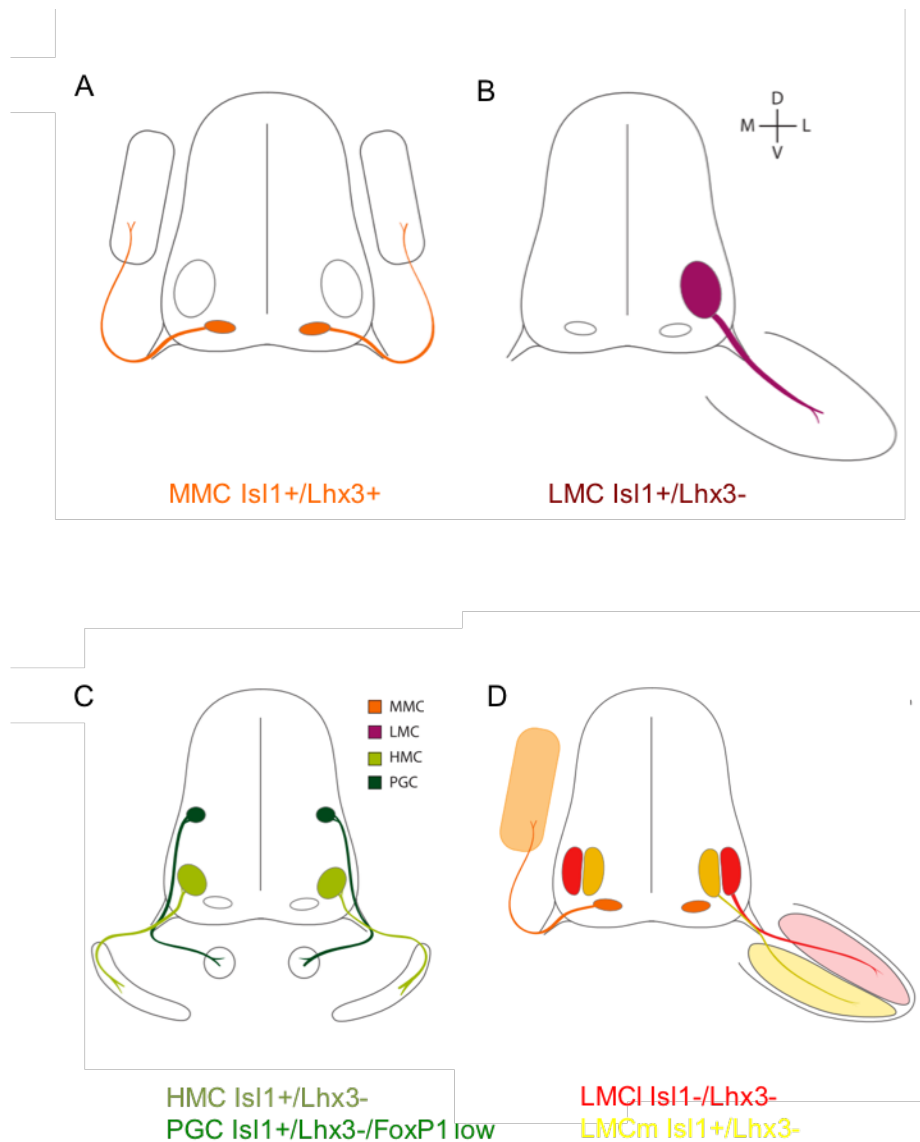


Figure 1.2. LIM transcription factor codes control spinal motor neuron subtype-specific properties. Different motor neuron subtypes express different combinations of LIM transcription factors which specify their unique subtype properties, including settling position in the spinal cord (columnar organization) and axonal projections. **(A)** Medial motor column (MMC) motor neurons express Isl1 and Lhx3 in combination and project dorsally to axial muscles. **(B)** Lateral motor column (LMC) motor neurons expressing Isl1 but not Lhx3 project along a ventrolateral route toward limb muscles. **(C)** Hypaxial motor column (HMC, light green) and preganglionic motor column (PGC) motor neurons (dark green) project towards hypaxial muscles and sympathetic chain ganglia, respectively. HMC and PGC motor neurons both express. Isl1 without Lhx3, with differentiation of PGC motor neurons by their low FoxP1 expression compared to HMC. **(D)** LMC motor neurons can be further divided into lateral (LMCl) and medial (LMCm) divisions; LMCl neurons express Lhx1 and Isl2 while downregulating Isl1 while LMCm neurons lack Lhx1 but retain Isl1/2 expression. *Adapted from Catela et al. (2015).*

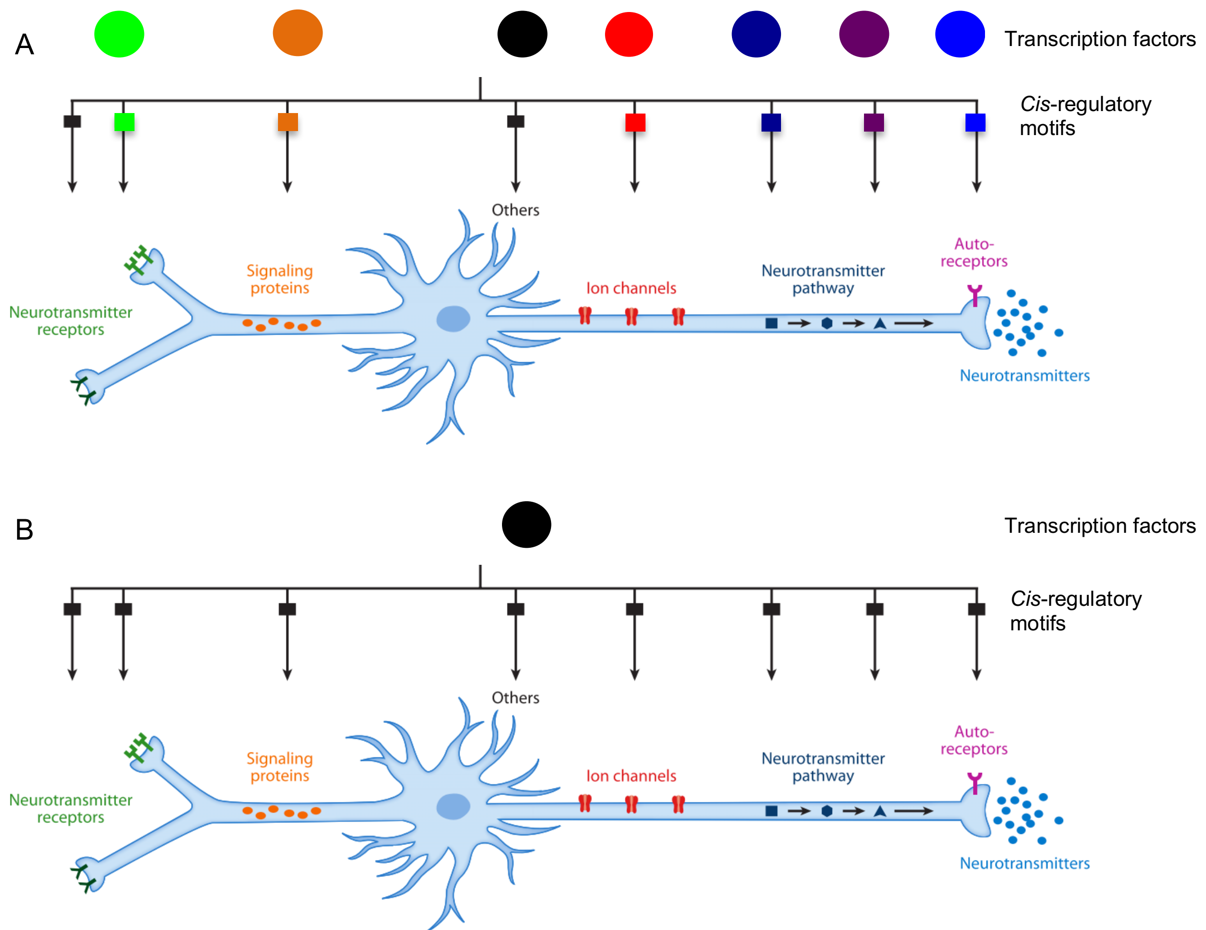


Figure 1.3. Potential models of regulatory control in the nervous system. (A) In this model, each type of gene making up the complete neuronal identity would be under the control of specific sets of regulatory factors. For example, expression of neurotransmitter biosynthesis genes (navy) and ion channels (red) would be under the control of a distinct transcription factor or combination of transcription factors. **(B)** In this model, a single or specific combination of master regulatory transcription factor(s) (black) would control all aspects of a distinct neuron's identity. *Adapted from Hobert O. (2011).*

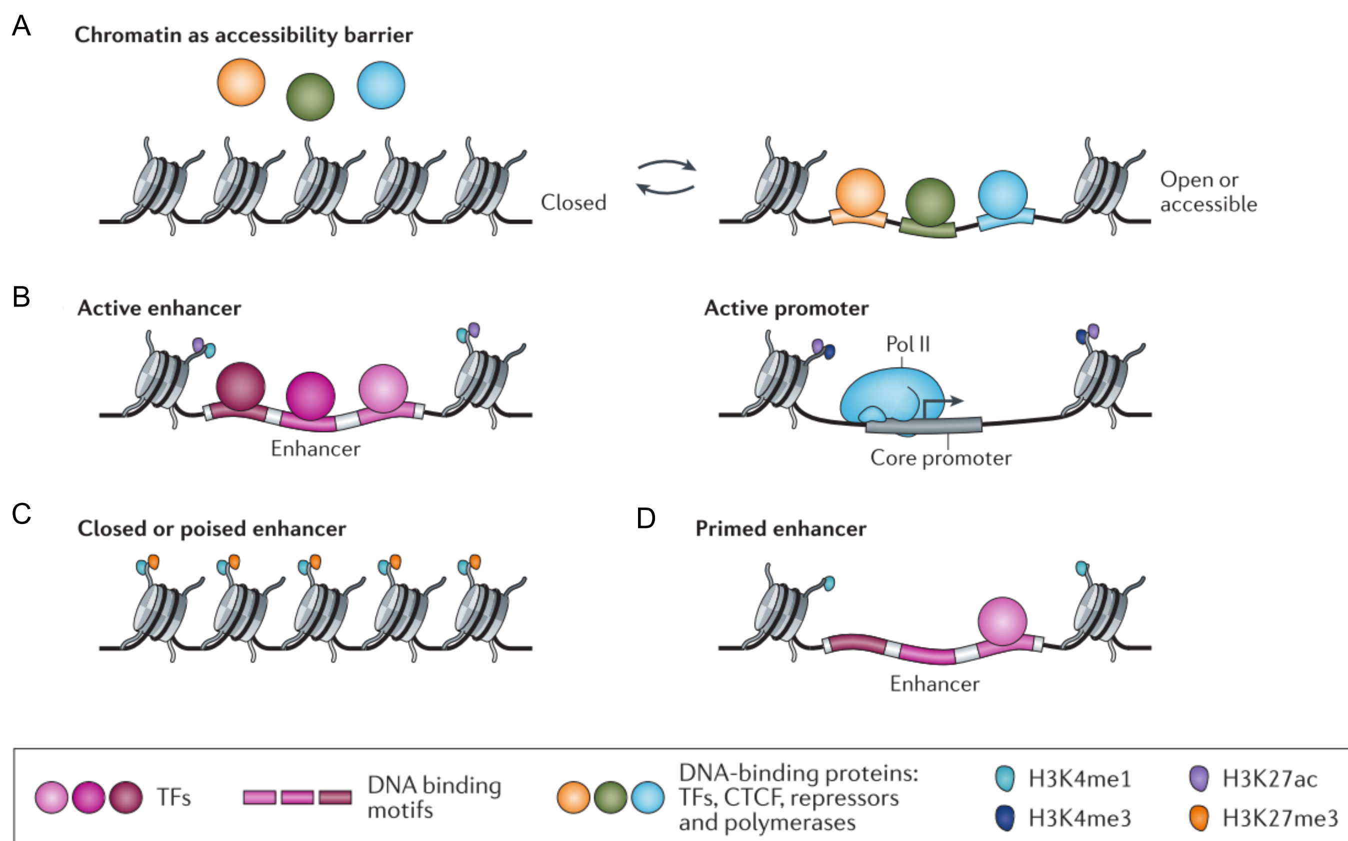


Figure 1.4. Chromatin accessibility and histone marks at regulatory elements. (A) Chromatin is shown as a 'gatekeeper' for transcription factor (TF) binding and enhancer activity. Densely positioned nucleosomes can restrict access for transcription factors (both activators and repressors), CCCTC-binding factor (CTCF), RNA polymerase II (Pol II) and other proteins. Accessible (that is, nucleosome-free) regions can be bound by these proteins, which define and mediate the identity of a region (for example, active enhancers, repressors or core promoters). The transition from 'open' to 'closed' chromatin, and vice versa, is determined by regulatory proteins, including pioneer transcription factors. Insulator proteins (for example, CTCF) and other architectural proteins also bind to open regions, and they make up a substantial proportion of sites that are accessible across multiple cell types (for example, 10%). **(B)** Histones that flank active enhancers are often marked by histone H3 acetylated at lysine 27 (H3K27ac) and H3 monomethylated at lysine 4 (H3K4me1). **(C)** Active promoters (depicted as Pol II bound) are flanked by nucleosomes with H3K27ac and H3K4me3 modifications. **(D)** Some closed or poised enhancer regions can bear the active H3K4me1 and the repressive Polycomb protein-associated H3K27me3 marks. **(E)** Enhancers that are not yet active but that are primed for activation either at a later developmental time point or in response to external stimuli can be pre-marked by H3K4me1. *Adapted from Shlyueva et al. (2014).*

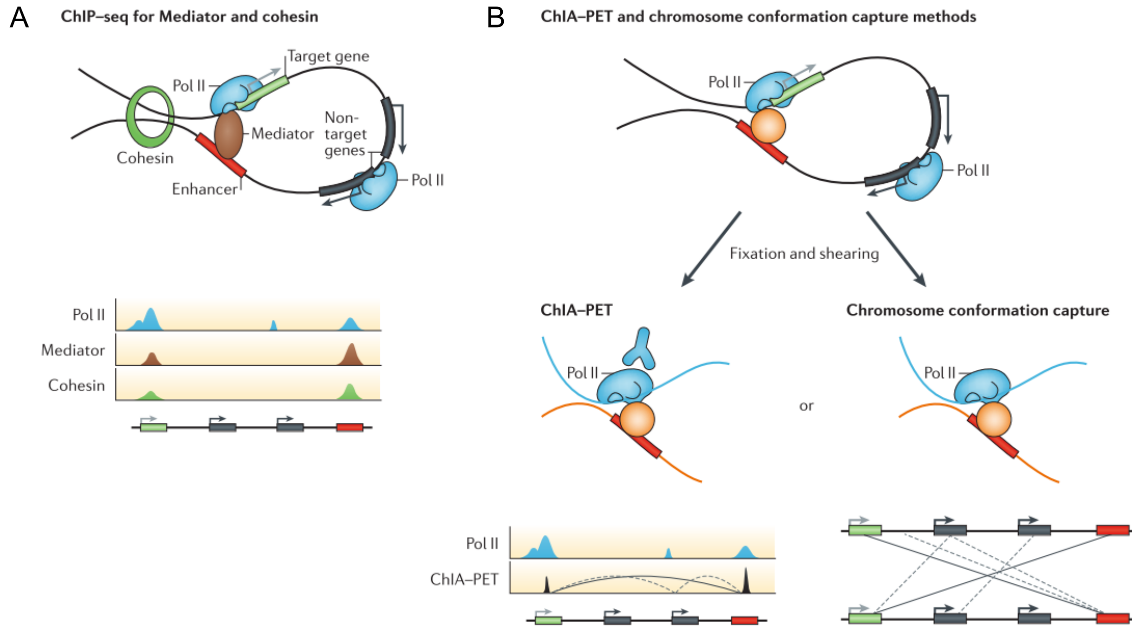


Figure 1.5. Genomic methods to identify regions involved in enhancer promoter interactions. (A) Enhancers are brought into close proximity of their respective target promoters through the formation of chromatin loops, which are thought to be established by cohesin and Mediator complexes. ChIP–seq can detect the contact points of cohesin and Mediator at promoters and enhancers, and has been used to predict enhancers. **(B)** Chromatin interaction analysis with paired-end tag sequencing (ChIA–PET) and chromosome conformation capture (3C)-based methods preserve and detect spatial contacts by crosslinking, DNA fragmentation, DNA fragment ligation and deep sequencing. ChIA– PET includes a ChIP step to enrich for complexes that contain a specific protein, such as RNA polymerase II (Pol II). In contrast to ChIP–seq (part d), both ChIA–PET and 3C-based methods detect not only the contact points but also the pairwise connections between these points. The thin, solid lines indicate that pairwise connections between spatial contact points are captured in ChIA–PET and 3C-based methods. For 3C-based methods a schematic output of a chromosome conformation capture carbon copy (5C) or Hi-C experiment is shown; this method probes all interactions between defined genomic loci for their spatial proximity and physical contacts, which is similar to ChIA–PET in that it might (solid lines) or might not (dashed lines) correspond to regulatory interactions. *Adapted from Shlyueva et al. (2014).*

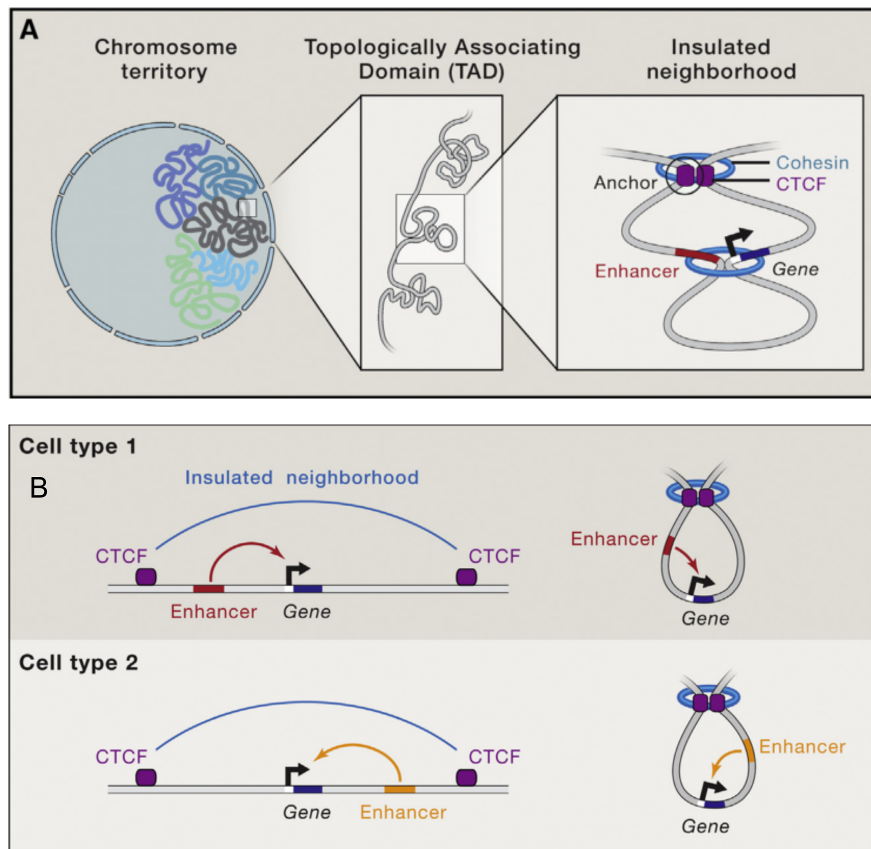


Figure 1.6. Insulated Neighborhoods. (A) Hierarchy of chromosome structures: chromosome territories, TADs, and insulated neighborhoods. Anchor refers to the CTCF-bound site interacting with another CTCF-bound sites, both co-bound by a cohesin ring. **(B)** Cell-specific enhancer-gene interactions occur within insulated neighborhoods that are generally maintained in different cell types. Left side displays a linear model of a genomic region encompassing a gene associated with cell-type-specific enhancers while the right side displays the insulated neighborhood model of the locus. *Adapted from Hnisz et al. (2016).*

Chapter 2

Synergistic binding of transcription factors to cell-specific enhancers programs motor neuron identity

Esteban O Mazzoni, Shaun Mahony, Michael Closser, Carolyn A Morrison, Stephane Nedelec, Damian J Williams, Disi An, David K Gifford, Hynek Wichterle

The data in this chapter are published in Mazzoni et al., 2013. Esteban Mazzoni and Shaun Mahony conceived and led this project. Esteban Mazzoni and Shaun Mahony wrote the majority of the manuscript. I was involved in writing and editing the portions related to experiments I performed.

I worked on this project during my first year in the lab with the goal of characterizing the functional properties of motor neurons directly programmed from mouse embryonic stem cells. I was interested to determine the validity of the direct programming paradigm as a pure source of motor neurons to model specification in the nervous system. In particular, I performed the following experiments: (1) Quantifications of programming efficiency of induced motor neurons in dissociated cultures. (2) Quantifications of specificity of programming between spinal and cranial induced motor neurons. (3) Characterization of general neuronal features in programmed motor neurons including expression of neuronal markers, synaptic maturation, and electrophysiological maturation (with Damian Williams). (4) *In vivo* transplantation of programmed cranial motor neurons and analysis of axonal projections (with Hynek Wichterle). (5) Time course expression analysis of genes associated with motor neuron progenitor and postmitotic identity. (6) Co-immunoprecipitation experiments of cell-specific transcription factor complexes.

Introduction

Recent progress in programming cell fate by transcription factors has given hope to the goal of producing clinically relevant cell types for disease modeling and direct therapeutic transplantation. Muscle cells, pluripotent stem cells, pancreatic beta cells, hepatocytes and several types of neurons have all been created by the forced expression of combinations of transcription factors known as programming modules (Mann and Carroll, 2002; Pfisterer et al., 2011; Sekiya and Suzuki, 2011; Son et al., 2011; Takahashi and Yamanaka, 2006; Tapscott et al., 1988; Zhou et al., 2008). However, the process of transcriptional programming remains largely enigmatic. Understanding the mechanism by which programming modules convert one expression profile to another one would not only illuminate the process of cell-fate specification during normal embryonic development, but would also have important implications for the rational design of programming modules for production of cell types that are difficult to generate using available methodologies.

When considering how programming modules associate with *cis*-regulatory elements to bring about a change in cellular identity, two extreme mechanisms can be hypothesized: an independent model, in which individual programming factors bind to distinct genomic locations and control independent sub-circuits of the cell specific gene regulatory network, and a synergistic model, in which the factors jointly bind to common regulatory elements to cooperatively activate the global cell type–specific expression program (Peter and Davidson, 2011). The independent model predicts that the effect of a programming module could be estimated by additively combining the effects of each individual factor. The synergistic model, on the other hand, predicts that the DNA sequence binding preference of the programming module is encoded by the multimeric transcriptional complex and that regulatory effects of the programming module will therefore be impossible to extrapolate from effects of individual factors studied in isolation (Slattery et al., 2011).

Cellular complexity in the CNS is established during development by closely related programming modules acting in a cell type–specific manner, providing a unique opportunity to study contributions of individual factors to the specification of cell fate. Motor neurons are cholinergic cells located in the ventral and caudal CNS, whose developmental program is particularly well mapped (Jessell, 2000b). Two

cardinal types of motor neurons are present in mammals, each expressing different transcriptional programs. Spinal somatic motor neurons (referred to here as spinal motor neurons) innervating skeletal muscles are derived from the ventral spinal progenitor domain and are characterized by co-expression of Isl1, Lhx3 and Hb9 (Mnx1) at the time of their birth (Jessell, 2000b). Branchiomotor and visceromotor neurons (referred to here as cranial motor neurons) located in the ventral midbrain, hindbrain and cervical spinal cord are defined by the co-expression of Isl1, Phox2a/2b and Tbx20 (Jessell, 2000b; Song et al., 2006). The combined expression of Isl1 and Lhx3, together with the proneural gene *Ngn2* (NIL factors), is sufficient to bestow spinal motor neuron identity on dorsal spinal progenitors and on spinal progenitors derived from embryonic stem cells (ESCs) (Hester et al., 2011; Lee and Pfaff, 2003; Lee et al., 2012). Our current understanding of the mechanisms through which NIL factors program spinal motor neuron identity are based on analysis of *Isl1* and *Lhx3* mutant phenotypes and on functional mapping of a spinal motor neuron specific *Hb9* (*Mnx1*) enhancer (Lee and Pfaff, 2003; Thaler et al., 2002). Although NIL factors synergize to control cell type–specific expression of the *Hb9* gene, mutations in Isl1 and Lhx3 result in distinct phenotypes, indicating that the two transcription factors may also possess independent functions (Pfaff et al., 1996; Sharma et al., 1998). Thus, the question of whether the NIL factors act primarily synergistically or independently at the genomic level remains unanswered.

To overcome the low efficiency of cell programming that limits biochemical analysis of the process, we established inducible ESC lines that harbor the NIL programming module or a module in which *Lhx3* is replaced by the cranial motor neuron determinant *Phox2a* (the NIP programming module) (Coppola et al., 2005; Hirsch et al., 2007; Pattyn et al., 1997). We found that NIL induction in differentiating ESCs resulted in rapid and highly efficient specification of spinal motor neurons and that NIP induction in an identical cellular context programed cranial motor neuron identity. Taking advantage of these robust and efficient programming systems, we mapped genome-wide binding sites of programming factors in both inducible lines. Computational analysis of occupied *cis*-regulatory elements revealed that Isl1 directly interacts and synergizes with Lhx3 or Phox2a in the relevant cellular contexts. The Isl1-Lhx3 and Isl1-Phox2a heterodimers exhibited different DNA-sequence preferences, forming the basis of cell-specific programming module activities and indicating that synergistic interactions between programming factors underlie specification of alternate motor neuron fates.

Results

Ngn2, Isl1 and Lhx3 program spinal motor neuron fate

To study the programming of spinal and cranial motor neuron identity, we generated two doxycycline inducible ESC line: one line harboring a polycistronic expression construct in which the open reading frames of spinal motor neuron determinants *Ngn2*, *Isl1* and *Lhx3* are separated by 2A peptides (iNIL line), and a second line in which we replaced *Lhx3* with a cranial motor neuron determinant *Phox2a* (iNIP line) (Figure 2.1A) (Hester et al., 2011; Lee and Pfaff, 2003; Lee et al., 2012). NIL factors have previously been shown to activate specification of motor neuron identity in retinoic acid–treated differentiating ESCs and when overexpressed with a combination of generic neuronal programming factors in fibroblasts (Hester et al., 2011; Lee et al., 2012; Son et al., 2011; Vierbuchen and Wernig, 2012). Given the developmental potential of pluripotent stem cells we first asked whether the NIL factors were sufficient to induce expression of spinal motor neuron markers even in the absence of retinoic acid. Surprisingly, we found that NIL factors were sufficient to induce expression of spinal motor neuron markers even in the absence of retinoic acid. Doxycycline treatment of ESCs grown in media lacking neuralizing signals from retinoic acid resulted in robust induction of the tricistronic transgene and activation of an Hb9::GFP reporter gene. Notably, despite continuing doxycycline treatment, Ngn2 expression was extinguished in most cells by 48 h, consistent with its transient pattern of expression in cells transitioning from progenitors to postmitotic motor neurons (Mizuguchi et al., 2001; Novitch et al., 2001).

Induction of the Hb9::GFP reporter was indicative of motor neuron specification, but to what extent NIL cells acquire global characteristics of motor neuron identity was not clear. I reasoned that if NIL expression successfully programs motor neuron identity the cells should acquire generic neuronal identity in conjunction with motor neuron specific properties. I next performed a series of studies in dissociated cells to characterize the programming efficiency and molecular characteristics of NIL cells. NIL-expressing cells plated on laminin displayed remarkably efficient induction of Hb9, adopted a typical neuronal morphology and expressed both the neuron specific isoform of class III β-tubulin (Tubβ3,

recognized by Tuj1 antibody) and the postmitotic neuronal marker NeuN (Rbfox3) (Figure 2.1B, C). I next quantified the efficiency of programming and observed that the majority of transgenic cells (labeled by antibody to V5) co-expressed the postmitotic neuronal marker, NeuN (99.72% ± 0.27 of $V5^+$ cells express NeuN) and the motor neuron marker, Hb9 (99.82% ± 0.17 express Hb9) (Figure 2.1D, E). Importantly, I did not observe any expression of the cranial motor neuron marker Phox2b (0.24% ± 0.28 express Phox2b) (Figure 2.1F, G). Together, these initial experiments suggest that in contrast to relatively inefficient reprogramming of neuronal identity from fibroblasts, our direct ESC programming paradigm results in robust and highly specific specification of spinal motor neurons.

Ngn2, Isl1 and Phox2a program cranial motor neuron fate

Cranial motor neurons share many features with spinal motor neurons including their cholinergic identity and the expression of the transcription factor Isl1 (Pfaff et al., 1996). However, unlike ventral spinal motor neurons, they express Tbx20 instead of Hb9, and their specification depends on the paired-like homeodomain transcription factors Phox2a and Phox2b instead of the LIM homeodomain factor Lhx3 (Hasan et al., 2010; Pattyn et al., 1997). Misexpression of the Phox2a and Phox2b transcription factors in the developing spinal cord is sufficient to induce ectopic formation cranial branchiomotor neurons (Hirsch et al., 2007). However, the expression of Phox2a alone in differentiating ESCs results in only a small increase in the number of Isl1-positive cells, and most of the cells fail to acquire neuronal identity, as shown by the lack of Tubβ3 expression (Figure 2.1F, H). We reasoned that joint expression of Ngn2 and Isl1 with Phox2a (NIP module) might lead to a more robust and uniform specification of cranial motor neurons (Figure 2.1A). To investigate this, I treated iNIP cells for 48 hours with doxycycline and performed the same set of characterizations in dissociated cultures as for iNIL cells. iNIP cells acquired uniform neuronal morphology and identity (Tubβ3 and NeuN expression) and expressed Phox2b in the absence of Hb9 (Figure 2.1F-H). The high efficiency of NIP programming was comparable to that of NIL programming: 99.77% ± 0.22 $V5^+$ cells expressed NeuN, 99.03% ± 0.08 expressed Phox2b and 0.11% ± 0.11 expressed Hb9. Together, these data indicate that the replacement of Lhx3 in the programming module with Phox2a results in efficient specification of neurons that acquired molecular properties of cranial motor neurons.

Functional characterization of induced NIL and NIP neurons

Having established the basic molecular properties of transcriptionally programmed motor neurons, I next sought to investigate the functional maturation of the cells including expression of synaptic proteins with known function in motor neurons. To determine whether transcriptionally programmed cells acquired key properties of mature motor neurons, I cultured induced NIL and NIP cells on monolayers of primary cortical mouse astrocytes for 7–10 days *in vitro* (Figure 2.2A). Immunostaining of NIL and NIP cells cultured on monolayers of astrocytes revealed dense arrays of synapses marked by the synaptic vesicle marker SV2 (Figure 2.2B, C). Notably, many of the synapses exhibited accumulation of vesicular acetylcholine transporter (Vacht), a marker of mature cholinergic synapses (Figure 2.2B, C). Cholinergic identity of NIL and NIP induced cells was further documented by an approximately 70-fold increase in the levels of choline acetyl transferase (*Chat*) mRNA and by Chat immunostaining (Figure 2.2B). The expression of synaptic proteins associated with pan-neuronal identity and cholinergic markers associated with suggests functional maturation of motor neuron identity on a molecular scale, but expression of synaptic proteins does not directly test whether induced neurons display neuronal activity.

Electrophysiologically mature motor neurons fire trains of action potentials following depolarization (Gao and Ziskind-Conhaim, 1998; Miles, 2004). We next asked whether induced motor neurons respond to depolarization. Whole-cell patch current-clamp recordings of NIL and NIP induced cells cultured on astrocytes for 7 days revealed that action potentials could be evoked by 20–150-pA, 1-s current injection in all cells tested (12 NIL cells, 12 NIP cells). Furthermore, nearly all patched cells (11 of 12 NIL cells, 11 of 12 NIP cells) fired trains of action potentials, sustained for the duration of the depolarizing current step (Figure 2.2D). Together, these observations suggest that inducible expression of NIL and NIP programming modules is sufficient to differentiate ESCs into electrically mature cholinergic neurons.

Motor neurons project axons outside of the CNS to innervate peripheral synaptic targets. To examine whether induced motor neurons acquired this defining characteristic, we performed a series of xenotransplantation experiments. We implanted control, iNIL or iNIP cells into the developing cervical and brachial neural tube of developing chick embryos and analyzed their projections patterns (Figure 2.3A) (Wichterle et al., 2002, 2009). We detected robust outgrowth of motor axons, labeled by a mouse specific

NCAM antibody, exiting the spinal cord via the ventral root and extending along all major spinal motor nerves. Importantly, ventral projections outside the CNS were observed in four of five successfully transplanted. In contrast, axons of control transplants remained within the spinal cord and failed to project to the periphery (Figure 2.3B). To further test the specificity of iNIL cell axonal pathfinding, I examined projections of iNIP cells transplanted into the same region of the chick neural tube. Unlike spinal motor neurons, cranial motor neurons do not exit the CNS through the ventral horn, preferring a more dorsal exit point (Pattyn et al., 1997). Accordingly, I observed that iNIP axons did not project through ventral roots, but instead accumulated selectively at the lateral region of the developing spinal cord. Interestingly, the location of iNIP axons suggests they are coalescing with the spinal accessory nerve populated by branchiomotor cranial motor axons, which originate from the spinal accessory nucleus in the lateral cervical spinal cord (Figure 2.3B) (Dillon et al., 2005). The same axonal trajectory has been observed for ectopic cranial motor neurons formed in the developing spinal cord following misexpression of Phox2a or Phox2b (Hirsch et al., 2007). Like the iNIL projections this result was highly reproducible with all four successfully transplanted embryos displaying the same axonal projection phenotype. Together, these results indicate that induced expression of the NIL and NIP modules programs cell phenotypes that are, by all examined criteria, consistent with spinal and cranial motor neuron identities (induced cranial and spinal motor neurons).

Expression profiles of motor neuron programming

Effective programming of ESCs into motor neurons should be accompanied by a repression of the stem cell expression program and induction of the spinal or cranial motor neuron specific transcriptome. Global expression profiling (Affymetrix GeneChIP ST arrays) after 48 hours of transcriptional programming of iNIL and iNIP cells resulted in a marked change in the gene expression profile (3,185 and 1,852 genes were more than twofold differentially expressed following NIL and NIP induction, respectively, $P < 0.001$; Figure 2.4A). Induction of NIL and NIP programming modules extinguished the expression of pluripotency genes (*Oct4*, *Nanog*), upregulated generic motor neuron genes (endogenous *Isl1*, *Ebf1/3*, *Onecut1/2*), cholinergic genes (*Vacht* (also known as *Slc18a3*), *Chrnb4*) and genes encoding motor axon guidance molecules (*Nrp1*, *Robo1/2*, *Dcc*) (Figure 2.4B). Comparison of

iNIL and iNIP cells revealed significant differences between the two samples supporting the idea that the programming modules induce distinct motor neuron fates (2,731 differentially expressed more than twofold, $P < 0.001$; 1,878 genes were upregulated in iNIL cells compared with iNIP, 857 genes were upregulated in iNIP cells compared with iNIL; Figure 2.4A). Indeed, the spinal motor neuron genes *Hb9*, *Isl2*, endogenous *Lhx3* and *Slit1/2* were selectively expressed in iNIL cells, while iNIP cells upregulated expression of the cranial motor neuron markers *Tbx20*, endogenous *Phox2a*, *Phox2b*, *Rgs4* and *Gal*; Figure 2.4B) (Grillet et al., 2003; Brunet and Pattyn, 2002; Holmes and Niswander, 2001; Moore, 1989). Unsupervised clustering of expression profiles revealed that iNIP cells segregated from iNIL cells, indicating that their identities are molecularly distinct on a global scale (Figure 2.4C). Spinal motor neurons can be generated from ESCs by directed differentiation controlled by patterning signals retinoic acid (RA) and sonic hedgehog (Shh) (Wichterle et al., 2002). Directed differentiation with RA and Shh recapitulates the normal process of motor neuron development. ESCs first acquire neural progenitor identity on day 3 (Sox1+, Olig2−), followed by motor neuron progenitor identity on day 4 (Olig2+, Hb9-), and differentiate to postmitotic motor neurons (Olig2-, Hb9+) on days 5–6 (Mazzoni et al., 2011; Wichterle et al., 2002).

To examine how closely programmed neurons correspond to differentiated motor neurons, we compared expression profiles of FACS-purified Hb9-GFP+ retinoic acid- and Shh-generated motor neurons on day 5 of differentiation with Hb9-GFP+ cells purified from iNIL cultures after 48 hours of programming. The induced motor neurons were markedly similar to retinoic acid– and Shh-generated motor neurons (Figure 2.4B-D). Most genes (97.4%) were expressed at levels that were not significantly different between the two samples ($P < 0.001$), and only 1.6% of all genes exhibited divergent expression with expression induced in one cell type and repressed in the other. The similarity of induced and control motor neurons is further supported by unsupervised clustering of gene expression profiles (Figure 2.4C). Although key motor neuron–specific genes were correctly regulated, we identified a set of genes controlling rostro-caudal neural identity and motor neuron subtype identity that were differentially expressed in retinoic acid– and Shh-generated cells and induced iNIL cells (Figure 2.4E). Induced iNIL motor neurons expressed low levels of *Hox* transcription factors and high levels of rostral neural markers (*Otx1*, *Otx2*). To rectify this difference, we asked whether programmed iNIL motor neurons are responsive

to the caudalizing signal retinoic acid (Mahony et al., 2011; Wichterle et al., 2002). Treatment of iNIL cells with retinoic acid in conjunction with dox resulted in the correct specification of cervical spinal identity, marked by the expression of *Hox* genes from paralogous groups 4 and 5 and suppression of rostral markers *Otx1/2* (Figure 2.4E). Thus, programmed cells acquire generic motor neuron identity following induction of NIL factors, but specification of rostro-caudal subtype identity depends on the treatment of the cells with caudalizing patterning signals.

Motor neuron programming bypasses neural progenitor stages

Rapid activation of postmitotic motor neuron markers following NIL and NIP induction raised the question of whether transcriptionally programmed cells transit through neural progenitor stages. To capture cells during the transition from ESCs to spinal motor neurons, we profiled induced cells after 24 hours of transgene expression. At this time point, NIL factors effectively repress key stem cell genes (for example, *Oct4*, *Nanog*; Fig. 3a) and already induce expression of markers associated with postmitotic spinal motor neuron identity, such as *Hb9* (*Mnx1*), *Isl2*, *Lhx4*, *VAChT* (*Slc18a3*), *Robo2*, *Slit2* and *Nrp1* (Figure 2.5A). Notably, although proliferating cells (Ki67+) were intermixed with cells expressing *Hb9*, none of the diving cells expressed the motor neuron progenitor marker *Olig2* (Figure 2.5B). Neither NIL nor NIP induced expression of genes associated with progenitor stages (*Sox1*, *Olig2* and *Ngn2*) that are highly expressed at 24 and 48 hours following retinoic acid and Shh treatment (days 3 and 4; Figure 2.5C). To completely rule out that idea that programmed cells go through a progenitor phase that is missed at our 24 and 48-hour time points we monitored Sox1 and Oilg2 expression in a more temporally precise time series. We measured expression every twelve hours during NIL and NIP induction and again observed a complete absence of Sox1 and Olig2 progenitor gene expression (Figure 2.5C). These results indicate that programming modules initiate a state transition from ESCs to postmitotic motor neurons that bypasses key steps in the normal motor neuron developmental program.

Isl genome binding is dependent on programming partners

Efficient and rapid programming of ESC differentiation into phenotypically distinct neurons by modules that differ only in one transcription factor provides an ideal system in which to study whether

individual transcription factors act independently or engage in synergistic interactions. If the individual factors are recruited to DNA independently, replacing Lhx3 with Phox2a in the programming module should not affect the DNA binding preference of Isl1. To test this independent model, we performed chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) analyses of Isl1 in iNIL and iNIP cells after 48 hours of transgene induction. Inducible Isl1 was not epitope tagged and we optimized ChIP using a pool of monoclonal antibodies to Isl1. As these antibodies cross-react with both Isl1 and the closely related Isl2, we refer to the data as Isl ChIP-seq. We observed extensive condition-specific Isl recruitment to genomic loci in the iNIL and iNIP induced cells (Figure 2.6A). We identified 18,187 significant Isl binding events in the two conditions, of which 38% were significantly differentially enriched ($P < 0.001$) between iNIL and iNIP lines (Figure 2.6B). In contrast, only 9.6% of the Isl binding sites were differentially enriched between iNIL cells and retinoic acid– and Shh-derived motor neurons. To further test the synergistic model, we profiled the binding of Isl when ectopically expressed alone in differentiating ESCs. The genomic occupancy of Isl was substantially different from that of Isl expressed in the context of either iNIL or iNIP cells: 67.5% of the Isl ChIP peaks were differentially enriched between iNIL and iIsl1 cells, and 48.8% were differentially enriched between iNIP and iIsl1 cells. Together, these results indicate that recruitment of Isl1 to DNA binding sites depends on the composition of programming modules, consistent with the synergistic model implicating functional interactions between programming factors.

Next we examined whether identified Isl binding sites are distributed randomly across the genome or whether their positions correlate with tissue specific *cis*-regulatory elements. We took advantage of ENCODE project data that identified putative regulatory regions in mouse ESCs, whole brain, heart, kidney, liver and spleen, defined using combinations of DNaseI hypersensitivity and enrichment in H3K4me1 and H3K27ac histone modifications (Shen et al., 2012). Of all of the tissues examined, Isl binding sites correlated best with whole brain putative regulatory regions. Notably, the overlap with regulatory regions in ESCs was as low as in unrelated tissues. These findings indicate that expressed NIL and NIP factors are not passively recruited to existing stem cell regulatory regions, but that these factors actively engage neuronal regulatory regions.

Cell-specific Isl binding correlates with gene expression

The identification of condition-specific Isl binding prompted us to examine whether differentially occupied sites in the iNIL and iNIP cells are associated with the establishment of cell type–specific gene expression profiles. We observed condition-specific Isl binding in the vicinity of developmental genes that were selectively induced by the NIL or NIP programming modules. For example, three sites that were bound by Isl in the iNIL cells, but not in the iNIP cells, were located downstream of the endogenous *Lhx3* gene; conversely, two sites that were bound in the iNIP cells, but not in the iNIL cells, were located near the Phox2b transcription start site (TSS). Meanwhile, we observed shared Isl binding sites near a subset of genes that are induced in both cell types, such as Chat (Figure 2.6A).

To extend this analysis, we asked what fraction of genes differentially expressed between the two conditions are proximal to sites that are differentially occupied by Isl. We first subdivided all induced genes that have nearby (overlapping the gene or <10 kbp upstream or downstream from the gene TSS) Isl binding sites into three categories: those that were induced in both cell lines, those that were induced selectively in iNIL cells (NIL induced) and those that were selectively induced in iNIP cells (NIP induced). Similarly, Isl binding sites were subdivided into condition-specific sites that were most differentially enriched ($P < 0.001$) in iNIL cells (5,285 sites, NIL>NIP), those most differentially enriched in iNIP cells (1,657 sites, NIP>NIL; Figure 2.6B) and condition-independent sites that were similarly enriched in the two cell lines (1,705 sites, NIL = NIP). Of all of the NIL induced genes, 57% had a nearby NIL>NIP site, 26% had a nearby NIL = NIP site and only 13% had a nearby NIP>NIL site (Figure 2.4C). Conversely, 70% of genes induced selectively in iNIP cells had a nearby NIP>NIL Isl binding site, 22% had sites similarly occupied in both cell lines and only 14% had a nearby NIL>NIP (Figure 2.4C). Based on the correlation between condition-specific Isl binding and condition-specific activation of gene expression, we propose that a subset of Isl binding sites function as context-dependent enhancers contributing to the establishment of the observed cell type–specific pattern of gene expression and to cell fate programming.

Sequence motifs explain differential Isl binding

Given that Isl genomic binding depends on the programming module context, we reasoned that Isl1 might partner with different transcription factors during NIL- and NIP-mediated cell fate programming,

resulting in a global change in its DNA binding preference. To elucidate the mechanisms underlying differential recruitment of Isl1 to genomic sites, we analyzed the DNA motifs enriched in the condition-specific and condition-independent binding sites. Motif analysis identified a monomeric sequence with consensus TAAKKR under the condition-independent (NIL = NIP) sites, which is identical to the *in vitro* binding preference characterized for Isl2 (Berger et al., 2008).

The analysis of differentially enriched sites revealed more complex dimeric motifs composed of a combination of two homeodomain binding sites (Figure 2.7A). Notably, the motifs associated with iNIL and iNIP condition-specific sites exhibited different motif grammar. Although the homeodomain half-sites formed an inverted repeat in the motif enriched under NIL-specific sites, the motif enriched under NIP-specific sites contained an everted half-site configuration (Figure 2.7B). The motifs were highly enriched under NIL- and NIP- specific Isl ChIP-seq peaks, with 60.1% of NIL and 33.5% of NIP peaks containing the NIL- and NIP-specific motif, respectively (2.5×10^{-23} false discovery rate). The marked specificity of the ordering of homeodomain binding motifs in selectively occupied sites suggests that Isl1 partners with two different homeodomain transcription factors in iNIL and iNIP cell lines. The differences in the structure of these transcription factor complexes likely underlie their sequence specific recruitment to DNA, providing a physical mechanism by which one transcription factor can regulate different targets to establish alternate cellular identities.

Lhx3 and Phox2a co-occupy Isl-binding sites

Previous analysis of the spinal motor neuron–specific *Hb9* enhancer revealed that Isl1 forms a multimeric complex with Lhx3, Ldb1 and Ngn2 or Neurod4 (Lee and Pfaff, 2003). We asked whether Lhx3 co-occupies other sites selectively bound by Isl in iNIL cell line. Taking advantage of the V5 epitope tag on the *Lhx3* transgene21, we performed ChIP-seq analysis of Lhx3 binding in the iNIL cells 48 h after doxycycline induction. We identified 47,908 Lhx3 binding sites in the genome and found that these sites were highly coincidental with the sites occupied by Isl in the iNIL cell line. We observed that only 1.7% of all sites were significantly differentially enriched ($P < 0.001$) in one experiment compared with the other (Figure 2.7A, C). These findings suggest that Isl1 and Lhx3 bind to DNA as a heterodimer during spinal motor neuron differentiation.

Although there is no prior evidence that Phox2a heterodimerizes with Isl1, we examined whether V5 epitope tagged Phox2a might pair with Isl1 in the iNIP cell line. Although ChIP-seq analysis revealed only 1,568 significant Phox2a binding events, Phox2a and Isl binding events were highly coincident and the magnitude of ChIP enrichment at the co-bound sites was also highly correlated, mirroring the co-binding of Isl and Lhx3 in iNIL cells. We observed that only 4.6% of all sites were significantly differentially enriched in one experiment compared with the other (Figure 2.7A, C). The high degree of co-binding of Isl and Phox2a raised the possibility that the two factors might be parts of the same transcriptional complex. It has been shown that purified Isl1 and Lhx3 transcription factors interact in solution, but whether they interact in vivo is unclear (Lee and Pfaff, 2003; Son et al., 2011). To test whether Isl1-Lhx3 formed a physical complex I performed co-immunoprecipitation experiments and first confirmed the Isl1-Lhx3 interaction in induced iNIL cells. I then asked whether Isl1 forms a cranial specific transcription factor complex with Phox2a. Indeed, I observed co-immunoprecipitation of Phox2a with Isl1 supporting the idea that Isl1 and Phox2a are members of the same transcriptional complex in iNIP cells (Figure 2.7D). Together, these results indicate that the alternate cellular fates produced by NIL and NIP programming modules are encoded by cooperative recruitment of Isl-Lhx3– and Isl-Phox2a–containing complexes to enhancers with distinct motif grammar.

Discussion

In this study, we exploited the differentiation potential of pluripotent ESCs to study how transcription factor modules control specification of distinct neuronal cell types. Inducible expression of two programming modules differing in one transcription factor led to a rapid and efficient specification of cells expressing key molecular and functional properties of spinal and cranial motor neurons. The transcription factor Isl1 changed its genome binding preference when expressed alone or in the context of either the NIL or NIP programming modules. Because the factors were expressed in an identical cellular context, the different binding preference of Isl is not attributed to differential chromatin accessibility or the initial presence of distinct cofactors. Instead, our data support a model in which Isl forms cell-specific transcriptional complexes with Lhx3 or Phox2a. The complexes are recruited to condition-specific

enhancers with differential motif grammar leading to activation of cell type–specific expression programs and to the specification of spinal or cranial motor neurons. These findings have broader consequences for the rational design of programming modules, as mapping an individual transcription factor's DNA binding preference is insufficient to predict its binding in specific cellular contexts and its potential for cellular programming when it is co-expressed with other cooperating programming factors. Systematic computational and experimental analysis of interactions among programming factors, along with decoding the grammar of their cooperative binding motifs, will be a fundamental step toward rational design of programming modules for predictable production of diverse cell types of interest.

The synergistic nature of the programming module's activity could explain why collections of factors are typically required to program terminal cell fate (Pfisterer et al., 2011; Sekiya and Suzuki, 2011; Son et al., 2011; Takahashi and Yamanaka, 2006; Tapscott et al., 1988; Vierbuchen et al., 2010; Zhou et al., 2008). It is of interest that Oct4, Klf4 and Sox2 (core module) co-occupy regulatory elements in ESCs, suggesting that combinatorial programming modules may be a general developmental strategy (Boyer et al., 2005; Kim et al., 2008, 2010a). A second set of transcription factors (Myc module) appears to operate in parallel to the core module in pluripotent stem cells (Kim et al., 2010a). We therefore anticipate that additional transcriptional modules besides NIL and NIP will contribute to the establishment of terminal motor neuron expression profiles. Notably, the NIL programming module does not activate expression of Hox transcription factors that control specification of motor neuron subtype identity (Dasen and Jessell, 2009). This is consistent with our recent demonstration that rostrocaudal patterning signals specify motor neuron positional identity by remodeling Hox chromatin landscape during early neural progenitor stages that are bypassed during direct programming by NIL factors (Mazzoni et al., 2013a). Thus, generic motor neuron identity can be experimentally uncoupled from the Hox-driven program controlling subtype-specific motor neuron properties. Evolution of a generic motor neuron program that operates in parallel with transcription factors controlling subtype-specific programs would provide a versatile and efficient system for diversification of generic motor neurons into distinct subtypes necessary for the assembly of a functioning motor system.

Together, the results presented here have begun to uncover the general mechanisms by which the combinatorial expression of transcription factors are able to program distinct motor neuron fates. We

observed binding of cell-specific heterodimers in close proximity (<10kb) to a subset of highly induced genes associated with both generic (Chat) and subtype specific identity in spinal and cranial motor neurons (Hb9 in spinal and Phox2a in cranial). However, binding of the programming factors was not restricted to induced genes as we also observed widespread binding in the proximity of repressed genes raising the question whether they act as activators, repressors, or both. Additionally, based on the assumption that the factors will bind in close proximity to their target genes we were only able to link a small proportion of the tens of thousands of binding sites to proximal genes suggesting that the majority of binding sites are located distal to genes. Together, these caveats present multiple important questions to be investigated about the functional role of cell-specific programming factors. What is the regulatory function of heterodimer binding? Are all sites functional and actively regulating enhancer activity and by extension target gene expression? And lastly, what are the target genes of the distal binding sites in the genome? In the next chapter, I take advantage of this robust programming system to directly address these questions in the context of spinal motor neurons.

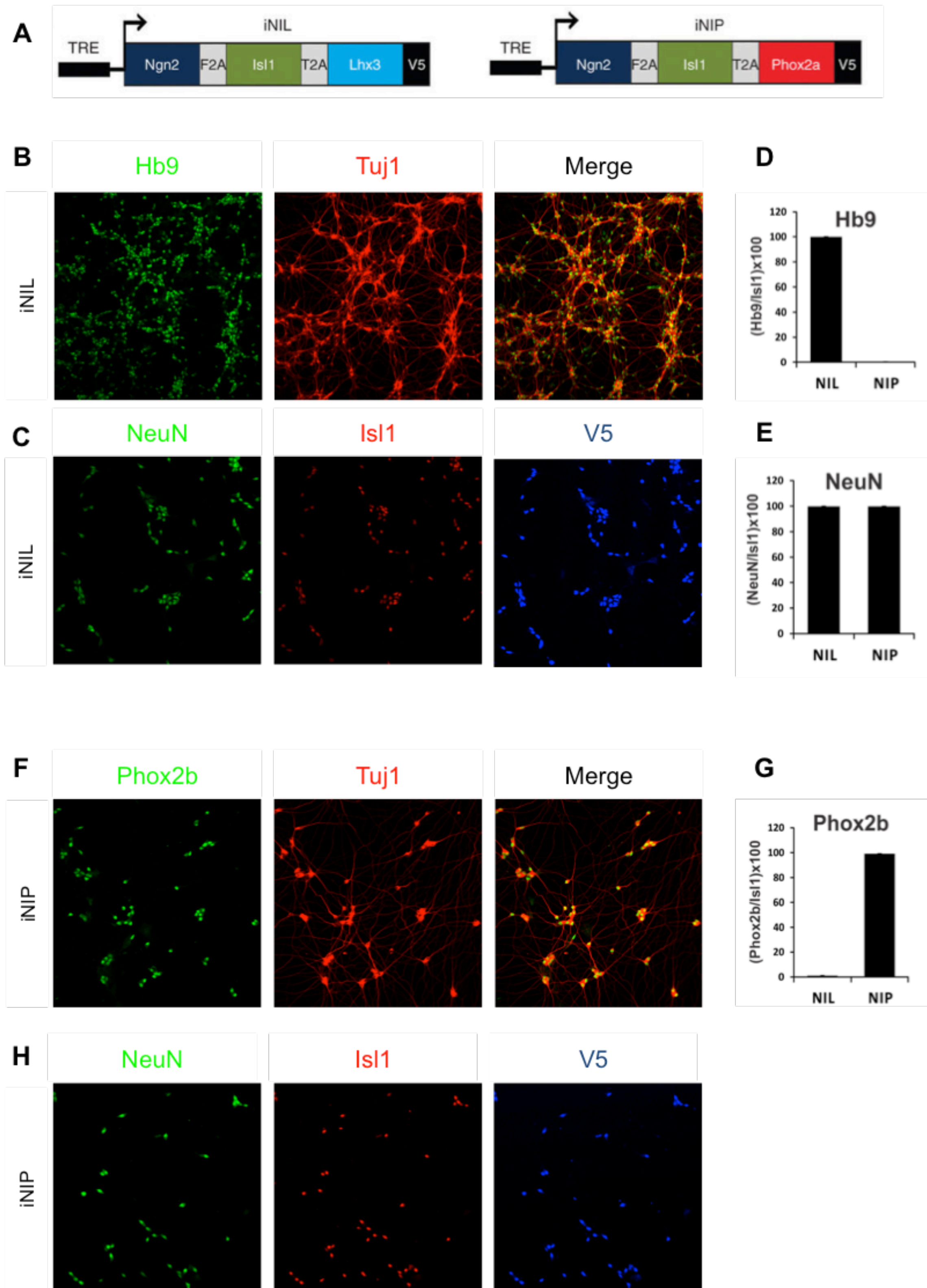
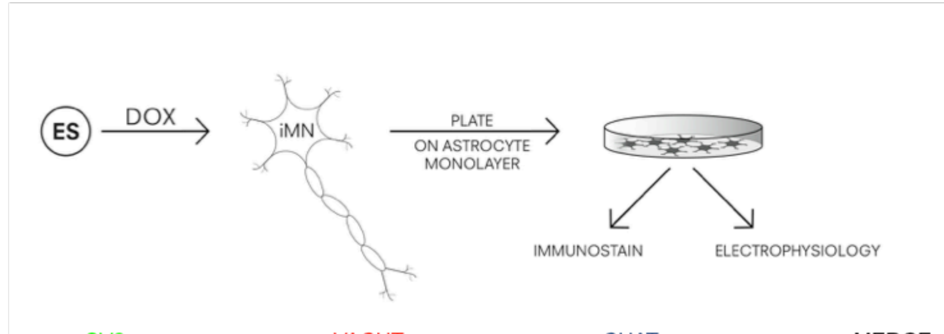
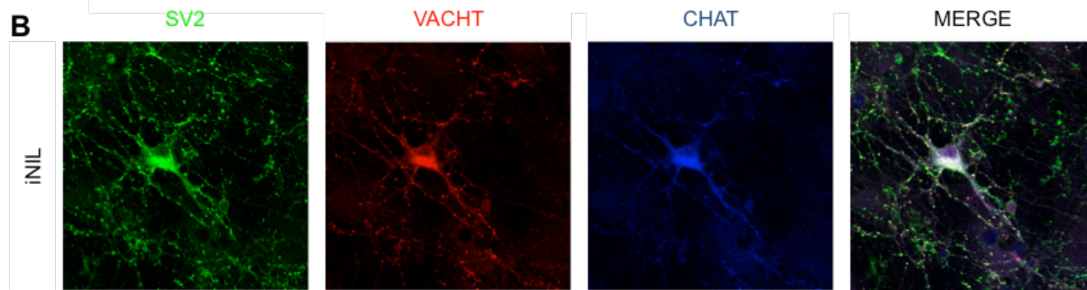
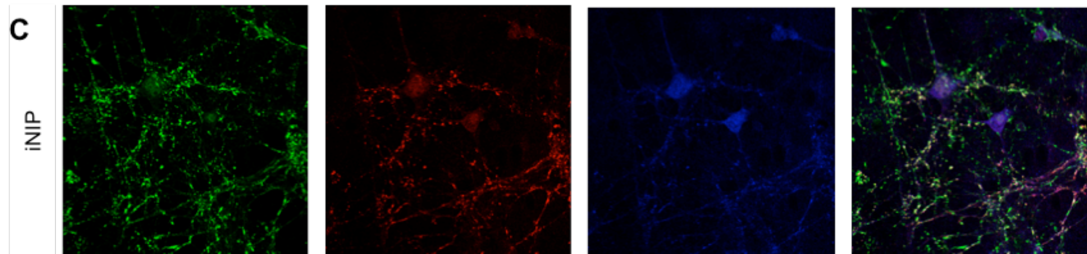

A
TRE
iNIL
Ngn2
F2A
Isl1
T2A
Lhx3
V5
TRE
iNIP
Ngn2
F2A
Isl1
T2A
Phox2a
V5
B
Hb9
Tuj1
Merge
iNIL
D
Hb9
(Hb9/Isl1)x100
NIL
NIP
C
NeuN
Isl1
V5
iNIL
E
NeuN
(NeuN/Isl1)x100
NIL
NIP
F
Phox2b
Tuj1
Merge
iNIP
G
Phox2b
(Phox2b/Isl1)x100
NIL
NIP
H
NeuN
Isl1
V5
iNIP

Figure 2.1. NIL and NIP transcription factors program spinal and cranial motor neurons, respectively. (A) Schematic representation of Dox-inducible NIL and NIP programming modules. TRE, tetracycline response element; F2A and T2A, 2A peptide sequences from foot-and-mouth disease virus. **(B)** NIL programmed spinal motor neuron in the absence of patterning signals. Induced iNIL cells exhibited neuronal morphology with multiple Tuj1-immunoreactive processes and expressed the spinal motor neuron-specific gene Hb9. Day 2 embryoid bodies treated with Dox for 48 h were dissociated, plated on laminin-coated substrate and analyzed 24 h later. Scale bars represent 50 μm. **(C)** NIL programmed neurons were plated at low density and stained with Isl1 to reveal induction of transgenes (V5) and NeuN. **(D)** Efficient induction of Hb9 by NIL, but not NIP. Shown is a quantification of the percentage of NIL or NIP induced cells ($Isl1^{+}$) expressing Hb9 (*n* = 3). Data are presented as mean ± _s.e.m. (**E**) NIL and NIP induced cells efficiently upregulate NeuN expression. Shown is a quantification of the percentage of NIL or NIP induced cells ($Isl1^{+}$) expressing NeuN (*n* = 3). Data are presented as mean ± _s.e.m. **(F)** NIP-programmed cells exhibited neuronal morphology, were Tuj1 immunoreactive and did not express Hb9, but were positive for Phox2b. **(G)** Efficient induction of Phox2b by NIP, but not NIL. Shown is a quantification of the percentage of NIL or NIP induced cells ($Isl1^{+}$) expressing Phox2b (*n* = 3). Data are presented as mean ± _s.e.m. (**H**) NIP programmed neurons were plated at low density and stained with Isl1 to reveal induction of transgenes (V5) and NeuN.


A
ES
DOX
iMN
PLATE
ON ASTROCYTE
MONOLAYER
IMMUNOSTAIN
ELECTROPHYSIOLOGY


B
SV2
VACHT
CHAT
MERGE
iNIL


C
iNIP

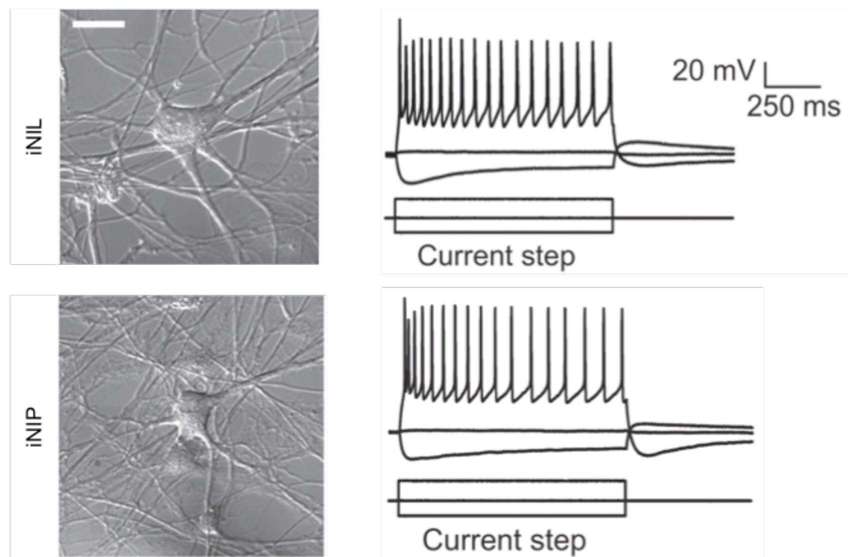

D
iNIL
20 mV
250 ms
Current step
iNIP
Current step

Figure 2.2. Characterization of NIL and NIP induced motor neurons. (A) Schematic of induction and culture protocol. ES cells were treated with Dox for 48 h to induce NIL or NIP transgenes, then dissociated and plated on astrocyte monolayer. Induced NIL and NIP motor neurons were then fixed for immunostaining or prepared for electrophysiology experiments. **(B, C)** NIL and NIP programmed cells contain cholinergic synaptic vesicles. Dissociated iNIL and iNIP cells induced with Dox were cultured on astrocyte monolayers for 7 d and stained with SV2, Vacht and Chat. Scale bars represent 10 μm. **(D)** NIL and NIP programmed neurons fired repetitive action potentials. Dox-induced iNIL and iNIP cells cultured for 7 d on astrocyte monolayers were analyzed by electrophysiology. Representative (11 of 12 cells for each cell line) current-clamp recordings are shown. The current command step that induced repetitive action potentials is shown under the membrane traces. Scale bar represents 25 μm.

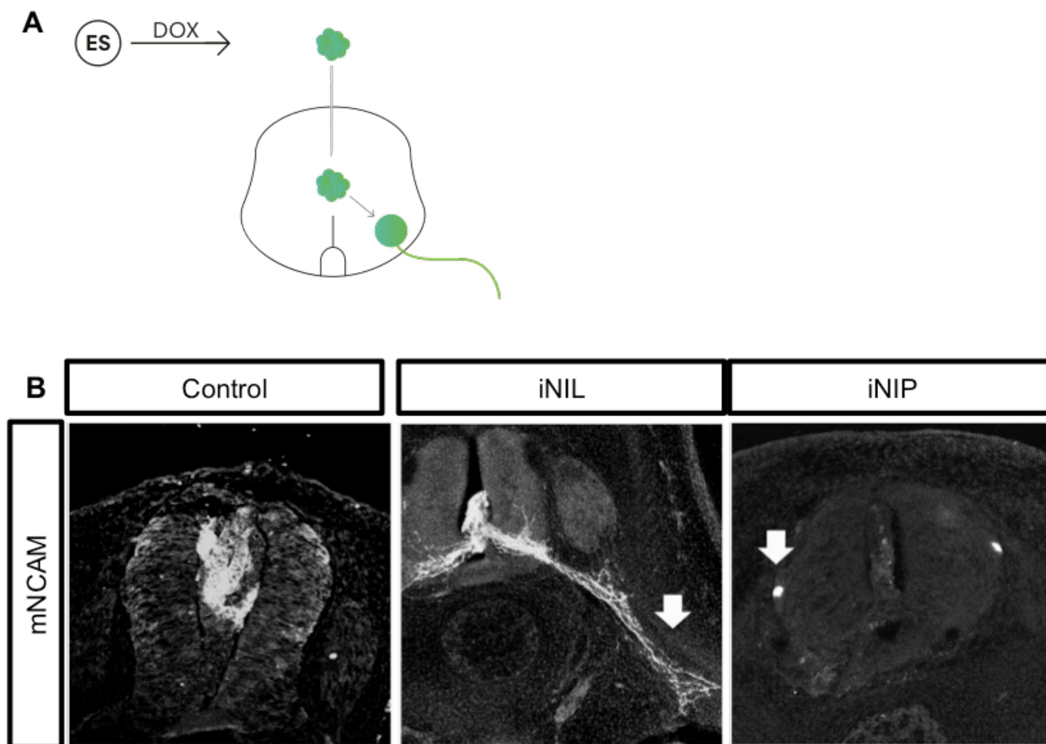
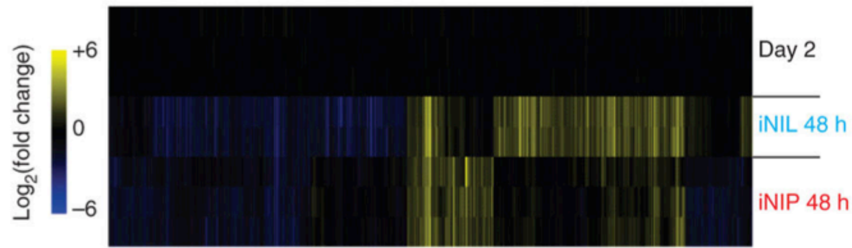
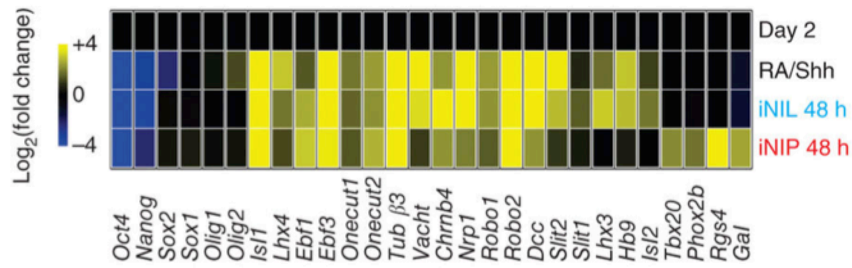


Figure 2.3. Xenotransplantation of induced NIL and NIP motor neurons. (A) Schematic of transplantation of ES-derived motor neurons into developing chick spinal cord. **(B)** NIL- and NIP-induced motor neurons projected axons toward different targets. Control and Dox-induced day 4 embryoid bodies were implanted into the Hamburger Hamilton stage 16 developing chick cervical spinal cord *in vivo*. Embryos were fixed 2 d later, sectioned and stained with a mouse-specific NCAM antibody. Dense bundles of axons emanating from NIL-induced transplants were observed in the ventral root and in axial and limb (arrow) nerve branches (four of five successfully transplanted embryos). In contrast, NIP-induced cells projected axons dorso-laterally toward the spinal accessory nerve (arrow) (four of four successfully transplanted embryos). Axons of control transplants remained in the confines of the neural tube. Scale bars represent 100 μm.

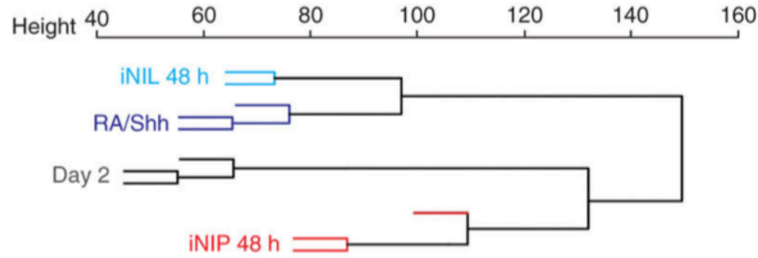
A



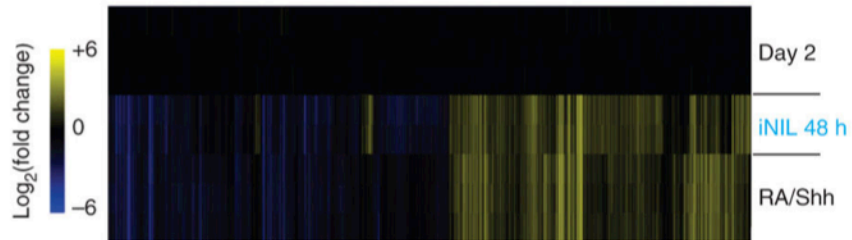
B



C



D



E

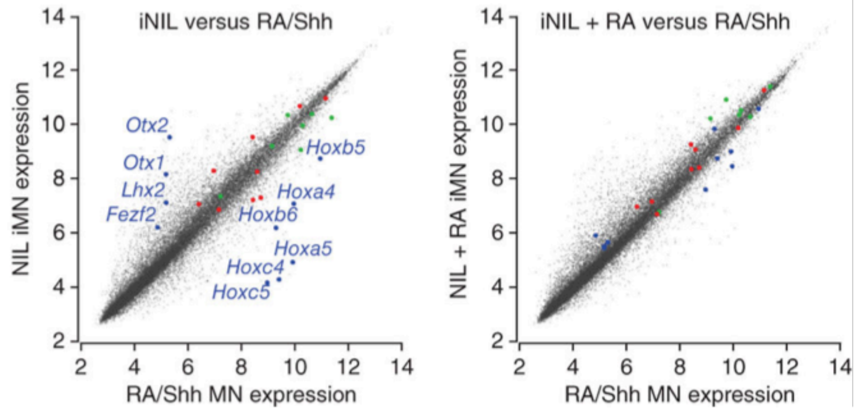


Figure 2.4. NIL and NIP induce different transcriptomes. (A) NIL and NIP factors induced different transcriptomes. Shown is a clustergram of all differentially expressed genes in day 2 embryoid bodies (before Dox induction, three replicates) and in iNIL (two replicates) and iNIP (three replicates) cells induced for 48 h with Dox. **(B)** The expression of a relevant subset of genes revealed the identity of NIL- and NIP-programmed cells. Shown is a heat map of average expression of genes associated with spinal and cranial motor neuron identity in day 2 embryoid bodies, retinoic acid– and Shh-derived (RA/Shh) spinal motor neurons (day 6 FACS-purified spinal motor neurons following 4 d of differentiation by retinoic acid and Shh treatment), and NIP- and NIL-programmed neurons induced for 48 h with Dox. **(C)** Unsupervised clustering of individual expression profiles revealed that NIL-induced cells clustered with retinoic acid– and Shh-differentiated spinal motor neurons, whereas the NIP programming module induced a different cell type that clustered apart from spinal motor neurons. **(D)** NIL expression induced a spinal motor neuron–specific transcriptome. Shown is a clustergram of all differentially expressed genes in day 2 embryoid bodies, Dox-treated iNIL cells and retinoic acid– and Shh-differentiated motor neurons. **(E)** Retinoic acid imposed cervical identity onto NIL-programmed spinal motor neurons. Left, scatter plot of mRNA expression intensities in Dox-induced iNIL cells versus retinoic acid– and Shh-differentiated spinal motor neurons. Right, scatter plot of mRNA expression intensities in Dox-induced iNIL cells treated with 1 μM retinoic acid for 48 h versus retinoic acid– and Shh-differentiated spinal motor neurons. Rostro-caudal patterning genes are shown in blue, spinal motor neuron associated transcription factors in red, and spinal motor neuron associated receptors and enzymes in green. iMN, induced motor neuron.

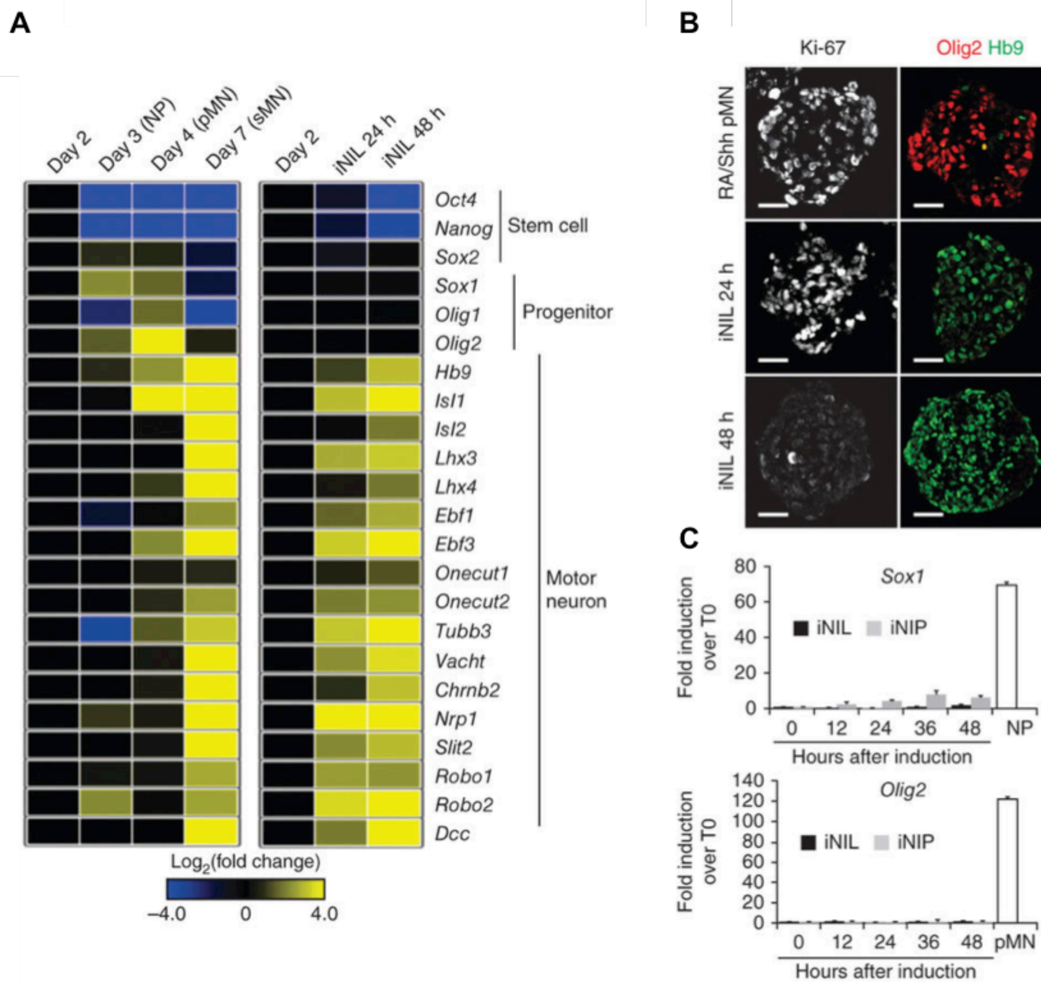


Figure 2.5. Induced iNIL cells bypass progenitor stages. (A) NIL-programmed spinal motor neurons directly induced terminal genes. Shown are expression profiles of genes associated with pluripotency (stem cell), neural and motor neuron progenitors (progenitor, NP and pMN), and postmitotic spinal motor neuron (motor neuron, sMN) identities. Clustergram of gene expression changed in a time series of embryoid bodies treated with retinoic acid and Shh from days 2 to 7 (left) and iNIL embryoid bodies treated with Dox from days 2 to 4 (right). **(B)** NIL-programmed spinal motor neurons did not express Olig2. Retinoic acid– and Shh-induced motor neuron progenitors and NIL-programmed spinal motor neurons (24 and 48 h of Dox treatment) were stained for Olig2 and Hb9. By 48 h after NIL induction, cells had exited the cell cycle, as revealed by the lack of Ki-67 staining. Scale bars represent 20 μm (top four panels) and 25 μm (bottom two panels). **(C)** Neither NIL nor NIP transcription factors induced expression of progenitor markers Sox1 or Olig2. Shown is a time series of *Sox1* and *Olig2* mRNA levels analyzed by quantitative PCR after the Dox treatment of iNIL and iNIP cells. Neural progenitors (NP;24 h after retinoic acid and Shh treatment of day 2 embryoid bodies) and motor neuron progenitors (pMN;48 h after retinoic acid and Shh treatment of day 2 embryoid bodies) served as positive controls for *Sox1* and *Olig2*, respectively. Data are presented as mean ± _s.e.m. (n = 3). T0, before induction.

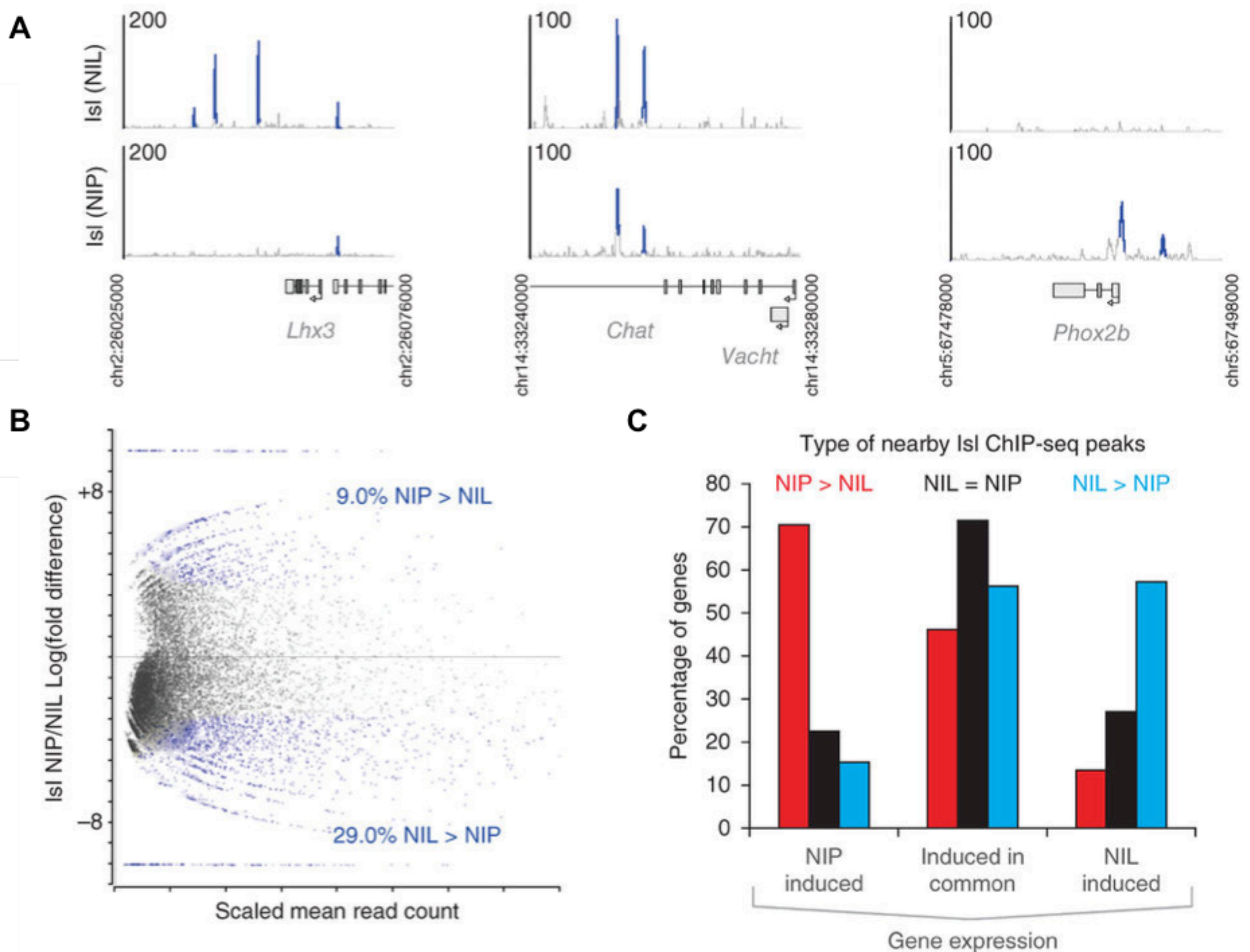
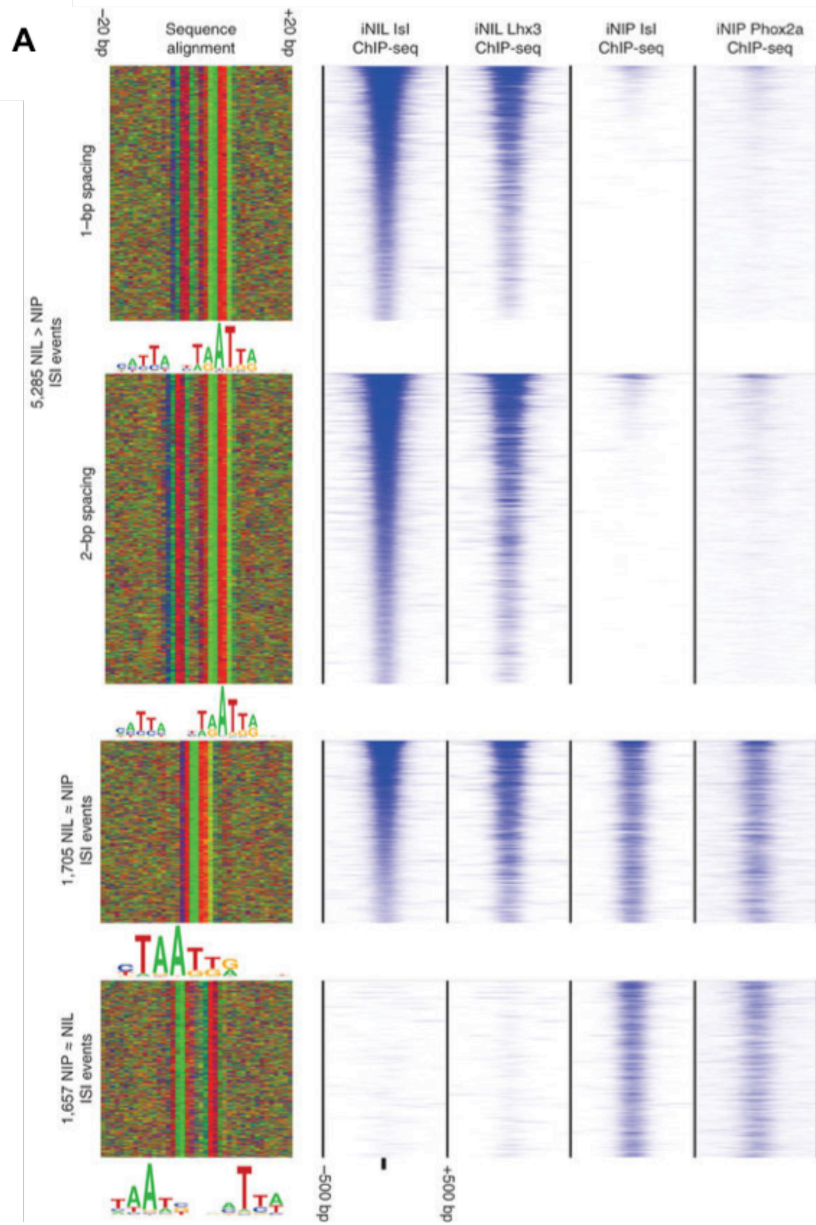


Figure 2.6. Context-specific Isl1 genome association correlates with differential gene expression. (A) Isl binding at developmentally regulated genes in iNIL and iNIP cells treated with Dox for 48 h. Isl ChIP-seq signals over *Lhx3*, *Chat* and *Phox2b* are shown. Blue peaks represent significant ($P < 0.01$) read enrichment over control. Genomic loci coordinates are shown next to the *x* axis. **(B)** Isl genome association was NIL and NIP specific. Shown is a comparison of Isl read enrichment from iNIL and iNIP cells at all detected peaks. Blue represents peaks that were significantly differentially enriched in one experiment over the other ($\log_2$; Online Methods). **(C)** Condition-specific Isl binding was associated with condition-specific gene expression. Differentially expressed genes were divided into ones selectively induced in iNIL and iNIP cells and ones induced in common in both cell types. The bar graph represents the percentage of genes containing a proximal Isl peak that was condition specific (Figure 2.4B) for each group.


A
−20 bp
Sequence alignment
+20 bp
iNIL Isl ChIP-seq
iNIL Lhx3 ChIP-seq
iNIP Isl ChIP-seq
iNIP Phox2a ChIP-seq
1−bp spacing
2−bp spacing
5,285 NIL > NIP ISl events
1,705 NIL ≈ NIP ISl events
1,657 NIP ≈ NIL ISl events
−500 bp
+500 bp

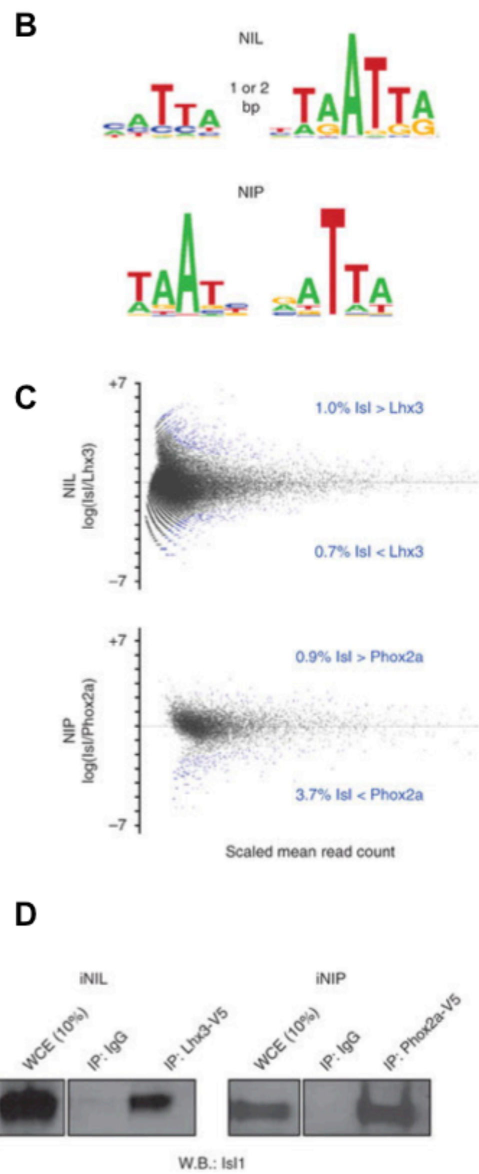

B
NIL
1 or 2 bp
NIP
C
+7
−7
NIL log(Isl/Lhx3)
1.0% Isl > Lhx3
0.7% Isl < Lhx3
NIP log(Isl/Phox2a)
0.9% Isl > Phox2a
3.7% Isl < Phox2a
Scaled mean read count
D
iNIL
iNIP
WCE (10%)
IP: IgG
IP: Lhx3-V5
IP: Phox2a-V5
W.B.: Isl1

Figure 2.7. Sequence motifs occupied by Isl1 in iNIL and iNIP cells. (A) Condition-specific Isl enrichment is explained by condition-specific DNA sequence motifs. All Isl binding sites determined to be significantly differentially enriched in either iNIL or iNIP cells were plotted, alongside a set of sites that were similarly enriched in both conditions. The sites are centered on the closest match to the condition-specific or condition-independent motifs within 25 bp of the predicted binding position. Isl ChIP-seq peaks that were similarly enriched or significantly enriched in either iNIL or iNIP cells were ordered on the basis of Isl read enrichment levels. Each column in the figure plots either a representation of the sequence alignment in a 40-bp window around the motif match or ChIP-seq read enrichment in a 1-kbp window around the predicted binding site. **(B)** Primary DNA motifs over-represented under enriched peaks obtained from Isl ChIP-seq experiments in iNIL and iNIP cells treated for 48 h with Dox. **(C)** Lhx3 and Phox2a colocalized with Isl genomic binding sites in iNIL and iNIP cells, respectively. Shown is a comparison of read enrichment from Isl with either Lhx3 or Phox2a at all detected peaks. Blue represents peaks that were significantly differentially enriched in one experiment over the other. **(D)** Isl1 associated with Lhx3 and Phox2a. V5 epitope–tagged Lhx3 or Phox2a transcription factor containing protein complexes were immunoprecipitated with antibody to V5 from Dox-treated iNIL or iNIP cells. The presence of Isl1 was examined by western blot analysis. WCE, whole cell extract; IP: IgG, negative control to test nonspecific binding; IP, V5 immunoprecipitation of Lhx3 from iNIL cells and Phox2a from iNIP cells. Shown is a representative blot (n = 3).

Chapter 3

Neuronal gene expression is controlled by constellations of enhancers distributed across large regulatory domains

Michael Closser, Yuchun Guo, Rachel Kopunova, Tulsi Patel, Yujin Ruan, David K Gifford, Hynek Wichterle

The data presented in this chapter make up the majority of thesis work and a manuscript that is in preparation. The concepts, analysis and majority of experiments of this study were conceived and performed by me with input and assistance from all collaborators. In particular, Yuchun Guo, a post doc in David Gifford's lab at MIT, led the development of novel computational methods for motif discovery, transcription factor mapping, and high-resolution mapping of chromatin interactions. Rachel Kopunova assisted with the CRISPR/Cas9 genome engineering of mouse ESCs to generate enhancer mutant cell lines. Tulsi Patel assisted with isolation of primary embryonic motor neurons. Yujin Ruan's lab was instrumental in generating ChIA-PET libraries. Hynek Wichterle and David Gifford advised and guided me through the project.

Introduction

Transcription factors are the principal drivers of cell specification in multicellular organisms, a property that has been successfully exploited in cell reprogramming experiments in multiple organisms (Mazzoni et al., 2013b; Takahashi and Yamanaka, 2006; Tursun et al., 2011). However, the mechanisms by which programming factors drive global changes in gene expression required for the successful acquisition of appropriate biochemical, physiological, and morphological properties remains poorly understood. Recent advances in biochemical analysis of protein-DNA interaction have facilitated the generation of genome-wide maps of transcription factor binding sites and begun to elucidate the locations of putative cis-regulatory regions throughout the genome. Despite substantial advances in mapping regulatory regions, our understanding of how recruitment of a transcription factor throughout the genome impacts patterns of gene expression during embryonic development and cell differentiation remains rudimentary.

The interpretation of transcription factor binding data during cellular specification is complicated by the fact that the number of transcription factor binding sites in the genome often significantly exceed the number of genes that are differentially regulated, the binding sites are often found far from transcription start sites and transcription of immediately proximal genes often does not correlate with transcription factor binding (Mazzoni et al., 2013b; Wapinski et al., 2013). Together, these issues confound the ability to infer regulatory mechanisms by transcription factors during specification and raise important questions with regard to developmental control of gene expression. To what degree does widespread transcription factor binding result in functional control of gene expression? Are all binding sites engaged with target genes to control the precise patterns of gene expression?

The importance of precise specification of cellular identity is perhaps best illustrated by the complexity of the nervous system where the proper assembly and function of neural circuits requires generation of an immensely diverse repertoire of neuronal cell types. The process relies on robust gene regulatory programs activating batteries of genes imparting distinct nerve cells with unique molecular and functional phenotypes including axon targeting, electrophysiological properties, and neurotransmitter

identity (Fishell and Heintz, 2013; Hobert, 2011). Interestingly, cell-specific gene batteries are often overlapping and highly complex throughout the nervous system. Individual genes are expressed in many different neurons raising the question how specificity is maintained to regulate the same genes in a cell-specific manner.

To this end, substantial progress has been made towards identifying transcription factors controlling cell fate specification and regulation of anatomical and developmental properties in many different neural structures including the retina, cortex and spinal cord (Jessell, 2000b; Livesey and Cepko, 2001; Lodato and Arlotta, 2015; Molyneaux et al., 2007). Despite the plethora of genetic information, the molecular mechanisms by which expression of specific transcription factors control neuron identity remain enigmatic. The difficulty in studying transcriptional mechanisms controlling cell identity is exaggerated in the developing mammalian nervous system. In addition, to difficulties in assigning function to complex patterns of transcription factor binding within a single cell type, there are difficult technical challenges introduced in studying neurons First, there is poor experimental access to a sufficient cell numbers necessary for the genome-wide interrogation of transcription factor binding, enhancer activation, and long distance chromatin interactions. Second, there is an inherent lack of purity due to the immense diversity within the primary tissue that impedes our ability to precisely dissect the regulatory mechanisms in a single neuronal cell type. Together, these issues make it essentially impossible to dissect the regulatory control by transcription factors in a pure cell population using only primary tissue. Based on this I reasoned we would need to devise a way to study the specification of a distinct neuronal cell type in isolation.

To deepen our understanding of the molecular mechanisms specifying cell identity in the nervous system I reasoned that a cell type that is well characterized would provide the best model of specification. Motor neurons in the spinal cord are particularly well characterized based on extensive studies of transcription factors controlling distinct aspects of specification and diversification conducted in recent years. Spinal motor neurons provide an excellent entry point to studying the biochemical processes controlling cell identity in the nervous system as they can be efficiently generated from stem cells by directed differentiation using embryonic patterning cues, or directly programmed with transcription factors (Mazzoni et al., 2013b; Peljto et al., 2010; Wichterle et al., 2002). These *in vitro* methods allow for robust

generation of a nearly unlimited supply of motor neurons, providing a unique opportunity to study their specification in molecular detail. As I described in Chapter 2, misexpression of three transcription factors, Ngn2, Isl1, and Lhx3 (NIL), is sufficient to activate an expression program that generates nascent postmitotic motor neurons that exhibit the general hallmarks of their in vivo counterparts, including ventral axonal projections, expression of axon guidance molecules (Robo1/2, Slit2, Nrp1/2), expression of cholinergic identity components (ChAT, Vacht) and finally, the transcription factor code associated with generic motor neuron identity (Lhx3/4+, Isl1/2+, Hb9+) (Jessell, 2000a; Lee and Pfaff, 2001). Importantly, using this programming paradigm we can generate almost completely pure populations of motor neurons. Together, these data suggest that induced motor neurons (iMNs), phenotypically model general features of *in vivo* motor neuron identity (Mazzoni et al., 2013b). Based on these general phenotypic similarities, I reasoned that I could use iMNs to model the specification of motor neuron identity *in vivo*.

Using an integrated approach, I generated genome-wide datasets from *in vitro* and *in vivo* motor neurons to uncover general principles of neuronal gene expression. I show that during development the motor neuron programming transcription factors Isl1 and Lhx3 drive dynamic changes in the distal regulatory landscape to establish new sets active enhancers from previously inaccessible DNA. Enhancers bound by these transcription factors control expression of the core features of motor neuron identity by directly engaging promoters via chromatin interactions. Chromatin interaction analysis revealed a surprising organization of regulatory elements controlling genes induced in motor neurons with interactions spanning significantly larger distances compared to regulatory interactions described in other cell types. These constellations of enhancers around neuronal genes are distributed over large regulatory domains to concomitantly control expression of motor neuron genes. Finally, meta-analysis of other neuronal datasets suggests that distributed dynamic enhancers organized across large regulatory domains are not a unique property of spinal motor neurons, but are broadly employed in the regulation of gene expression within nervous system. Together, our findings support the view that cellular diversity of the nervous system associated with highly complex patterns of gene expression relies on comparably complex regulatory elements distributed across large territories of non-coding DNA.

Results

An integrated approach to map in vivo relevant regulatory regions in motor neurons

Regulation of gene expression during development is a dynamic process under the control of distant acting enhancers. Genome-wide profiles from DNA accessibility and histone modifications from multiple tissues and cell types have uncovered a general principle that distal regulatory regions controlling gene expression are highly dynamic and cell type specific (Heintzman et al., 2009; Nord et al., 2013; Shen et al., 2012). Therefore, to begin to understand the processes that control motor neuron identity, I focused on mapping regulatory elements at the time of specification. I performed an assay for transposon accessible chromatin (ATAC-seq) in iMNs to identify putative regulatory regions that exhibit local regions of accessibility (Figure 3.1) (Buenrostro et al., 2013). Following peak calling and bioinformatic analysis we generated genome-wide maps of regions of enriched accessibility in motor neurons and identified 36316 peaks of distal accessibility (Zhang et al., 2008). In Chapter 2, we observed that iMNs had global phenotypic properties that mirrored in vivo motor neurons. I reasoned that these phenotypic similarities implied a similar overall mechanism of regulatory control. To further validate this model of specification I performed ATAC-seq on primary motor neurons from Hb9::GFP-positive mouse embryonic spinal cords. I FACS purified from Hb9::GFP positive motor neurons at e10.5 when the peak of motor neuron specification is in progress. With maps of *in vivo* and *in vitro* profiled MNs in hand I then asked to what extent is distal chromatin accessibility conserved between the two sets of motor neurons. To investigate this, I used a statistical model, EdgeR, that tests differential enrichment at a given peak in a genome-wide manner. This analysis revealed a high degree of overlap between putative distal regulatory regions with approximately eighty-seven percent of all distal accessible regions unchanged between *in vivo* and *in vitro* profiled MNs (31516/36316; LogFC<2, p<.01, Figure 3.1B, C, D). These initial analyses confirm the robustness and accuracy of this integrated system for studying gene regulatory mechanisms in the nervous system. Importantly, this technique can likely be extended to any cell type as an efficient model for studying specification. Together these experiments identify the total set of putative regulatory elements in motor neurons, yet the identity and function of these elements is not clear.

Dynamic reorganization of the active enhancer landscape in motor neurons

Distal accessible sites can be organized functionally into diverse subsets, including enhancers, insulators and silencers (Bulger and Groudine, 2011; Gorkin et al., 2014; Levine, 2010). Enhancers are the elements, which are primarily responsible for activation of gene expression and tend to be highly dynamic and cell type, even within a specific tissue. A group of studies over the past few years have uncovered specific chromatin signatures associated with active enhancers in mammals, which facilitate their genome-wide identification (Heintzman et al., 2009; Nord et al., 2013). For example, enhancers can be differentiated from promoters based on enrichment for specific chromatin modifications (high H3K4me1 and low H3K4me3 with promoters displaying the opposite characteristics), while both active enhancers and promoters are highly enriched for H3K27 acetylation (H3K27ac). Thus, based on the chromatin signature associated with a specific genomic region one can predict it's regulatory function (Creyghton et al., 2010; Heintzman et al., 2007, 2009; Rada-Iglesias et al., 2011).

Following this logic, I set out to map the distal enhancers in motor neurons responsible for expression of motor neuron identity genes. I performed chromatin immunoprecipitation followed by deep sequencing (ChIP-seq) to generate genome-wide maps of active distal enhancers, marked by the enhancer associated chromatin signature, H3K4me1 and H3K27ac (Figure 3.2A,B). As distal enhancers regulating cell identity are thought to be largely cell type specific I reasoned that functional specification of motor neuron identity would be accompanied by a reorganization of activated distal regulatory elements and those specific to motor neurons are likely relevant to the motor neuron gene expression program (Heintzman et al., 2009; Nord et al., 2013). To investigate this I compared the active enhancer landscape in motor neurons and mouse embryonic stem cells (Fig 3.2A,B) (Whyte et al., 2013). Bioinformatic analysis of peaks in each cell type identified 13176 active distal enhancers in ESCs and 11470 active enhancers in induced motor neurons. Interestingly, upon comparing the locations of active enhancers in ESCs and iMNs I observed that the majority are highly dynamic during motor neuron specification. More than eighty percent (10792/13176 or 81.9%) of enhancers active in ESCs lose their activity and accordingly, more than ninety percent (10521/11470 or 91.7%) of enhancers active in iMNs gained the activity during motor neuron specification (Figure 3.2B). Together these data demonstrate the global

transformation in cell identity is accompanied by a similar transformation of the global regulatory landscape.

I next sought determine the mechanisms responsible for the establishment of these enhancers prior to motor neuron specification. Mouse embryonic stem cells have been proposed to exhibit a bivalent chromatin signature at both promoters and enhancers. Indeed, the Isl1 promoter was shown to be poised for activation in ESCs, but to what extent this is the case at distal enhancers remains poorly understood (Bernstein et al., 2006; Rada-Iglesias et al., 2011). A specific combination of H3K4me1 and H3K27me3 with the absence of the activating histone mark, H3K27ac, is posited to prime lineage specific enhancers for rapid and effective activation during early stages of neural differentiation (Rada-Iglesias et al., 2011). Together these studies suggest priming of motor neuron enhancers as a potential mechanism for robust activation during specification. To investigate whether newly established motor neuron enhancers were primed during development I investigated the enrichment of H3K4me1 at enhancers. Remarkably, newly activated motor neuron enhancers appear to be almost entirely void of H3K4me1 chromatin modifications suggesting they are not pre-established as primed enhancers in ESCs (96% of active enhancers lack H3K4me1 in ESCs) (Figure 3.3A, B, C).

Despite the lack of H3K4me1 chromatin modifications it is possible that motor neuron enhancers are primed in ESCs by another mechanism. One possibility is that enhancers exist in an accessible state to facilitate transcription factor binding during differentiation. This mechanism is used during T-cell specification when the transcription factor Foxp3 binds preferentially in pre-existing sites to promote T-reg differentiation (Samstein et al., 2012). To exclude the possibility that motor neuron enhancers are kept in an accessible, poised state by other mechanisms, I investigated the chromatin accessibility based on DNase I hypersensitivity in ESCs (DHS). Again, I observed enhancers to be almost entirely void of priming-associated pre-accessibility. The large majority (84%) of motor neuron enhancers lack DNA accessibility in ESCs. Taken together, these data indicate that in order to program motor neuron identity the enhancer landscape must be established de novo.

De novo establishment of enhancers during specification is likely under the control of programming transcription factors (Wapinski et al., 2013). Isl1/2 and Lhx3 co-occupy thousands of sites

throughout the genome during specification of motor neuron identity implicating them as potential regulators of the enhancer landscape (Lee et al., 2012; Mazzoni et al., 2013b). To determine to what extent the changes in the enhancer landscape might be attributed to the activity of Isl/Lhx3 binding I performed ChIP-seq on Isl1/2 and Lhx3 in MNs and assessed the coincidence of IL binding and newly established distal regulatory elements. Interestingly, I observed that redistribution of distal H3K27 acetylated regions throughout the genome is highly coincident with IL binding. IL heterodimers were found in approximately sixty percent of all newly activated sites, suggesting that they play a principal role in the establishment of motor neuron regulatory landscape (Figure 3.2A,B,C,E). Accordingly, co-occupancy of Isl/Lhx3 is highly correlated with recruitment of the histone acetyltransferase p300 ($r2 = .78$). When superimposed on enhancers active in ESCs or MNs, IL binding is almost exclusively (>97%) restricted to newly activated enhancers in iMNs (Figure 3.2A,B,C). Together, these data support a global regulatory function of Isl and Lhx3 heterodimers as transcriptional activators during motor neuron specification that, either on their own, or in conjunction with additional factors establish a new enhancer landscape from previously inaccessible chromatin.

Mapping high-resolution enhancer-promoter interactions in motor neurons

The distal nature of *cis*-regulatory elements and their numerical complexity in the mammalian genome, often acting up to a megabase away, makes functional assignment to their target genes, and by extension their role in regulating gene expression, extremely difficult (Bulger and Groudine, 2011; Gorkin et al., 2014). Functional regulation of cell type specific gene expression is posited to involve chromatin interaction between distal enhancers, often bound by programming transcription factors, and the promoters of their target genes (de Laat and Duboule, 2013; Tolhuis et al., 2002; de Wit et al., 2013). Recent technological advances, especially the advent of chromosome conformation capture techniques and their genome-wide derivatives, have begun to make functional assignment of distal enhancers to their target genes possible (Dekker, 2002; Fullwood et al., 2009; Kieffer-Kwon et al., 2013; Lieberman-aiden et al., 2009). However, assays to study chromosome confirmation are dependent on large cell numbers (on the order of tens to hundreds of millions), which makes studying *in vivo* derived neurons nearly impossible. However, taking advantage of the purity of iMN programming and its global similarity to

in vivo motor neuron accessibility allowed me to identity the functional targets of enhancer in motor neurons (Figure 3.1).

I reasoned that distal regulatory regions actively involved in regulating transcription of genes underlying motor neuron identity are likely engaged with active promoters. In order to identify which distal regulatory regions control the motor neuron gene expression program we mapped promoter-tethered interactions genome-wide using a modified chromosome confirmation capture technique, namely chromatin interaction analysis with paired end tagging (ChIA-PET) (Fullwood et al., 2009; Kieffer-Kwon et al., 2013; Li et al., 2012). ChIA-PET is particularly advantageous over alternative chromatin interactions methods, including Hi-C, because it facilitates relatively high-resolution mapping of genome-wide chromatin interactions that are associated with a protein of interest, providing precise binding and chromatin interaction information from a single experiment. Moreover, based on the protein of interest, from a single experiment, one can identify specific types of chromatin interactions (i.e. enhancer-promoter or insulator-insulator, etc.). RNA polymerase II is known to occupy both promoters and distal enhancers of actively transcribed genes (Kagey et al., 2010a; Kieffer-Kwon et al., 2013; Kim et al., 2010b). I reasoned that by using promoters as an anchor we could then search the genome for distal enhancers that are engaged in high-confidence interactions and thus, likely involved in regulation.

In order to map high-confidence promoter-tethered interactions we developed several novel computational methods to map distal interaction anchors at high spatial resolution. Existing computational methods detect interactions by first calling peaks (anchor regions) and then quantifying interaction frequencies between the anchors. However, the peak-calling step ignores paired-end linkage information, and thus these methods fail to detect interactions that have weak enrichment of reads in one or both anchors. This is particularly problematic for mapping interactions associated with proteins that have diffuse patterns of binding and broad regions of enrichment (RNA PolII or histone modifications). To detect interactions more sensitively and reliably, we developed a method called Chromatin Interaction Discovery (CID) that uses unbiased density-based clustering to predict interactions directly from paired-end reads (Guo Y., Closser M., in preparation; (Reeder et al., 2015); see methods for detailed description). Using CID, we cluster local densities of single end reads into anchor regions across the

linear genome. Then we map their distal anchors using paired-end information to identify long distance interactions.

We applied CID to map RNA PolII-anchored interactions in motor neurons and identified 5153 unique distal enhancer elements engaged in high-confidence (FDR<.01) interactions with promoters (Figure 3.4 A, B). Consistent with previous studies, the distal enhancers we identified are highly enriched for components of structural protein complexes including Med1, of the mediator complex, and Smc1, of the cohesin complex that have established roles in dynamic enhancer utilization during development (Dowen et al., 2014; Kagey et al., 2010a; Phillips-Cremins et al., 2013). Additionally, the high-confidence enhancer-promoter interactions were marked by the chromatin signatures previously described to differentiate enhancers and promoters, with TSS anchors highly enriched for H3K4me3 and depleted for H3K4me1 and enhancers highly enriched for H3K4me1 and depleted for H3K4me3 (Figure 3.4B). Together, these data provide the first genome-wide map of enhancer-promoter interactions in a single neuron type allowing for the precise dissection of the regulatory control of motor neuron identity.

IL mediated enhancer-promoter connectivity directly control MN expression program

To gain a global understanding of the regulatory mechanisms controlling specification of motor neuron identity, I focused my attention to genes that are highly induced at the time of post mitotic motor neuron specification with the idea that these genes would encompass the core features of motor neuron identity. I first asked if these genes were targets of distal enhancer interactions. To investigate this I performed RNA-sequencing experiments followed by differential expression analysis and integrated the set of induced genes with my enhancer interaction maps. This integrated analysis identified 626 highly induced genes interacting with 1629 distal enhancers (Figure 3.5A, B, C). Next I asked whether known motor neuron effector genes were targeted by these distal interactions. I observed highly specific interactions between distal enhancers and promoters of MN effector genes, including genes associated with neurotransmitter identity, axon guidance and transcriptional regulation (Chat, Slc18a3, Slit2, Robo1/2, Nrp1, Dcc, Mnx1, Lhx3, Lhx4, Isl1, Isl2) suggesting the newly established enhancer interactions drive expression of the motor neuron program (Figure 3.4A, 3.5A,B,).

My previous analysis of Isl/Lhx3 programming factor binding suggests they are principle drivers of the active enhancer landscape in motor neurons, but it isn't clear to what extent they contribute to induction of the motor neuron gene expression program. Previously, we observed a few examples of proximal binding to motor neuron genes, including ChAT and Hb9, suggesting potential regulation, but at the time very few binding sites were easily assigned to target genes and able to predict expression (Mazzoni et al., 2013b). To investigate the extent to which Isl/Lhx3 contribute to enhancer-promoter communication I mapped Isl/Lhx3 binding onto enhancers engaged with highly induced genes and asked what proportion were occupied by Isl/Lhx3. Altogether, I was able to identify 299 (or 48 % of all engaged genes) highly induced genes engaged with 834 IL bound enhancers (Figure 3.5.A, B, C, D). These data strongly support Isl/Lhx3 heterodimers as principle regulators of the motor neuron expression program. In further support of this argument, effector genes encompassing multiple aspects of motor neuron biology, including axon guidance, neurotransmitter identity and transcriptional control were all engaged with IL enhancers. These results uncovered an interesting finding that there are a disproportionate number of enhancers engaged with motor neuron identity genes. We observed that 52% of genes with distal IL bound enhancer interactions were engaged with multiple distinct enhancers and 35% are engaged with three or more (Figure 3.4A, 3.5A, B, C, D). Additionally, those genes associated with the core motor neuron program are associated with multiple distinct IL bound enhancers (Isl1, 2; Isl2, 4; Chat, 3; Slit2, 3; Dcc, 4; Mnx1(Hb9), 3;). Together these data solidify the role of Isl1 and Lhx3 as master regulators responsible for induction and regulation of the motor neuron expression program during specification.

Constellations of distributed enhancers control the expression of motor neuron genes

Our observations preferentially linking IL enhancers to the core identity genes unveiled a complex regulatory mechanism in motor neurons where multiple distinct enhancers interact with the same target genes to control expression of the postmitotic motor neuron identity program. Recent evidence has suggested that both local and long distance modes of genomic organization with respect to regulatory elements play an important role in maintaining specific regulation of cell identity genes. In many cell types, in both mouse and human, cell identity genes are regulated by spatially clustered sets of enhancers called super-enhancers. These super-enhancers reside within larger regulatory domains

called insulated neighborhoods that restrict enhancer activity to genes within the neighborhood (Dowen et al. 2014; Hnisz et al. 2013; Whyte et al. 2013).

Based on this I next analyzed the genomic organization of the multimeric enhancers around the core identity genes with established roles in motor neuron biology including axon guidance, neurotransmitter identity, morphology and transcriptional regulation. Presumably, if the clustered organization of enhancers was somehow important of regulation of cell identity genes then we expect that the distal enhancers interacting with promoters of motor neuron genes should be spatially clustered. Surprisingly, we observed that instead, the majority of motor neuron genes are engaged with multiple distinct enhancers distributed over large distances in the genome (Figure 3.4A, 3.6A,B). Enhancers connected to motor neuron genes are separated by distances spanning from 8kb to 1.5mb with a median distance of 120 kb of genomic space (Figure 3.6C). Importantly, genes associated with all aspects of motor neuron identity are associated with distantly distributed constellations of enhancers bound by Isl and Lhx3. For example, two genes that are responsible for motor axon guidance, Dcc and Slit2, are both regulated by multiple distinct enhancers separated by more than 100kb in the genome (Figure 3.6B). Interestingly, both of these genes, as well as the majority of motor neuron identity genes, are expressed throughout the nervous system suggesting there might be a unique mode of regulatory organization around neuronal genes.

To rule out the fact that super-enhancers might play a role in motor neuron biology we performed an analysis of super-enhancers with respect to gene expression. We mapped super-enhancers throughout the genome using previously published methods and identified 235 in the motor neuron genome (Hnisz et al., 2013; Whyte et al., 2013). Surprisingly, we observed that very few highly induced genes important for MN identity have any association with SEs. We found that only 3% of Isl/Lhx3 bound enhancers are involved in such clustered organization and only 8% of the induced target genes these enhancers are associated by such clustered enhancer structure. Additionally, only, 28% of super-enhancers are associated with induced genes upon motor neuron specification. Together these data verify that the primary regulatory mechanism during neuronal programming is not through super-enhancers. In stark contrast to recent studies in many cell types and many different tissues, describing major regulatory control of key cell identity genes by spatially cluster super-enhancer elements, the

regulatory organization around motor neuron genes is fundamentally different with their expression regulated by multiple distinct enhancers distributed over large regions of genomic space.

Enhancer constellations are distributed across large regulatory domains

I next investigated what could explain this distinct regulatory organization in motor neurons. Interestingly, analysis of the enhancers around motor neuron genes revealed another interesting organizational difference. I found that enhancers for the core motor neuron genes tend to be more distantly located than enhancers in ESCs (Figure 3.6A,B). To investigate whether this was a general principle of MN gene regulation I compared the distance between enhancer-promoter pairs genome-wide around genes expressed specifically in motor neurons or ES cells (Dowen et al., 2014). Indeed, motor neuron enhancer-promoter interactions are significantly larger and span ~2.5 times larger genomic distances than ES genes (MN =~120kb; ES = ~50kb, $p<.0001$; Figure 3.6C). Together the longer interaction distance and distributed organization of motor neuron enhancers suggest an interesting idea that the regulatory space around motor neuron genes might be different. Presumably, if the interactions are both long distance and spatially distributed then there would need to be more genomic space to facilitate such a phenomenon. Indeed, my initial analysis of the Isl1, Dcc and Slit2 loci support this idea. The genes are surrounded by remarkably large intergenic regulatory domains with individual interactions each spanning hundreds of kilobases to target the gene suggesting that the local enhancer organization around genes might be linked to the larger regulatory space within which the gene is located (Figure 3.6B,C). To investigate this concept more deeply, I compared the intragenic distance surrounding genes enriched in motor neurons or ES cells. Remarkably, similar to the comparison of enhancer-promoter distances, motor neuron genes are surrounded by intergenic regions that are approximately twice the size of ES genes (~230kb and ~110kb respectively, $p<.0001$; Figure 3.6D). Together these analyses uncover a striking principle of enhancer-promoter organization in motor neurons compared to what has been previously described in other cell types. Motor neuron identity genes dependent on regulation from multiple long-distance interactions from enhancers scattered throughout large genomic regions where ESC genes are regulated by distinct clustered enhancers over relatively small genomic distances from their target genes (Figure 3.6B, C, D).

In addition to relatively local regulatory organization of enhancers, recent studies have uncovered the importance of higher order chromosome organization and the mechanisms, which restrict enhancer control of gene expression within topologically associating domains, (TADs) anchored by insulated neighborhoods. These regions are anchored by CTCF and cohesion complexes and are generally stable across cell types forming the regulatory domains responsible for specificity in gene expression (Dixon et al., 2012; Dowen et al., 2014; Guo et al., 2015; Nora et al., 2012). Based on these studies I next asked whether the distantly distributed enhancers of motor neuron identity genes were restricted to TADs. Upon comparison of enhancer-promoter pairs with respect to topological domain boundaries I found that the majority of enhancer constellations are indeed restricted within the same TAD as the genes they regulate. Ninety percent of all enhancers and ninety two percent of enhancers bound by Isl/Lhx3 are in the same TAD as their target genes. Together, these results suggest that the regulatory control of motor neuron genes comes from a distributed organization of enhancer constellations within large TAD restricted regulatory domains to impart cell type specific control of gene expression.

Despite the fact that TAD boundaries tend to be stable in most cell types tested to date their sizes are quite variable and range from hundreds of kb up to multiple megabases (Dixon et al., 2012; Gorkin et al., 2014). Based on with my previous results showing larger intergenic regulatory domains surround motor neuron genes. I reasoned that perhaps motor neuron genes were also located in larger TADs to facilitate the distantly distributed enhancer organization. Surprisingly, unlike the intergenic regions, I observed no significant differences in the size of the TAD that motor neuron genes are located in when compared to ES cell genes, stably expressed genes, or all genes (Figure 3.6.E). Since the intergenic region surrounding motor neuron genes is significantly larger and the total genomic space per TAD is the same, I reasoned that there must be more potential intergenic regulatory space allotted to motor neuron genes. Indeed, analysis of the Dcc regulatory landscape supports this idea with the distributed interactions located throughout a large gene poor topological domain. To investigate this idea more globally I computed a “normalized regulatory domain size” by normalizing the TAD sizes around motor neuron genes to the number of genes within the domain. This value roughly approximates the amount of potential regulatory space in each TAD per gene. After doing this, I observed a significant increase in the size of regulatory space around motor neuron (Figure 3.6.F). These results uncover a surprising

organization of enhancers around motor neuron genes whereby multiple distinct enhancers are distributed over large patches of genomic distance surrounding their same target gene. The target genes are located in large gene poor regulatory domains that are restricted by TAD boundaries, which likely maintain specificity of enhancer-promoter communication. Together, these results uncover a unique principle of regulatory organization in motor neurons, which might be a general feature of neurons as the majority of motor neuron identity genes are broadly expressed in the nervous system.

Independent regulation from enhancer constellations

I next investigated the contribution of individual distributed enhancers, as combinatorial control of gene expression by groups of developmental enhancers remains poorly understood. One model for super-enhancers is that spatial clustering of enhancer elements somehow facilitates efficient recruitment of transcription factor and co-activator complexes to promote enhancer activation and induction of gene expression (Hnisz et al., 2013; Whyte et al., 2013). It's not clear whether there is synergism or combinatorial requirements of enhancers. To investigate the role of individual enhancers on gene expression levels during motor neuron specification I turned my focus to the master regulator of motor neuron identity, Isl1 where a striking example of distributed organization was observed. Analysis of TF binding and chromatin interactions identified two high-confidence interactions between distally located IL bound regulatory elements spanning a genomic region of ~650kb (~220(E1) and ~650kb(E2)) downstream of the Isl1 TSS. Both enhancers are highly enriched for factors associated with an active enhancer chromatin signature including K3K27ac, p300, mediator, and cohesin complexes and show significant interaction with the promoter region of Isl1 suggesting a strong regulatory function (Figure 3.6A, 3.7A).

I first investigated the regulatory function of individual enhancers by performing enhancer reporter assays in developing embryonic spinal cords of chick embryos, which have homologous progression of motor neuron specification as mice (Jessell 2000). I cloned ~1kb DNA fragments encompassing the active enhancer regions for each individual Isl1 enhancer directly upstream of a minimal promoter sequence driving a destabilized GFP reporter. This construct allowed me to test the precise

spatiotemporal specificity of motor neuron enhancers *in vivo* as the destabilized GFP has a half-life of two hours and is therefore essentially visible only when being actively transcribed. Cloned enhancers were then electroplated into the neural tube of developing chick embryos at HH stage 14, (cite HH chick paper) prior to motor neuron birth and reporter expression was assessed ~36 hours later after at the peak of motor neuron generation (Figure 3.7B,C). Analysis of the electroporated embryonic spinal cords revealed robust and specific expression of each individual construct in MNs suggesting each enhancer is sufficient to drive gene expression in isolation. Interestingly, I did not observe any reporter expression in more dorsal cells where a small population of Isl1 expressing interneurons (dI3) and Lhx3-positve V2 interneurons reside suggesting this enhancer is likely motor neuron specific (Figure 3.7B,C).

Having verified the independent sufficiency of individual Isl1 enhancers, I next investigated the contribution of each enhancer on expression of the endogenous Isl1 gene by generating loss of function mutants. I performed CRISPR/cas9-mediated genome engineering to selectively delete each enhancer from the endogenous genomic context. Deletion of each enhancer in isolation resulted in a significant decrease in expression of Isl1 suggesting that both enhancers are necessary for endogenous expression of Isl1. E1 deletion led to a sixty percent reduction of endogenous gene expression and E2 deletion led to an eighty percent reduction. Interestingly, neither enhancer completely abolished Isl1 gene expression in isolation further supporting the findings from the electroporation experiments that each enhancer is independently sufficient to drive target gene expression. I next tested the effect of compound deletions of both high confidence enhancers on Isl1 gene expression. Double mutant cells showed an even greater decrease with over 90% reduction in expression in Isl1 (Figure 3.7.D). Together, these data suggest a quantitative, probabilistic and independent regulatory input from each individual enhancer. Each enhancer is both necessary and sufficient to drive expression of Isl1 expression. Additionally, the distributed independently acting enhancers, at least for Isl1, seem to be probabilistic in nature. Each enhancer activates gene expression to a different extent and activation in compound mutants is essentially the product of the probabilities of each individual enhancer. This result is consistent with recently performed live imaging experiments in drosophila which showed that enhancer control of gene expression is dynamic in nature and target genes exhibit “bursting” frequencies that are dependent on their enhancer (Hong et al., 2008).

General regulatory organization in the nervous system

Having established a unique regulatory organization around motor neuron genes, I next investigated whether this phenomenon was specific to motor neurons or if it might be a general principle of regulatory organization in the nervous system. To investigate this I performed a similar set of analyses on a set of whole tissue expression data from the mouse ENCODE project, including lung, liver, heart, kidney and brain. First, I generated a set of tissue-enriched genes by performing a global differential expression analysis. With these gene sets I compared the size of intergenic regions surrounding genes preferentially expressed in neurons compared to the other tissues. Remarkably, I observed a median increase of around ~200kb ($p < 1e\text{-}10$) compared to other tissue specific genes suggesting that neurons, in general, are surrounded by larger intergenic regions (Figure 3.8.A). As in motor neurons, I mapped genes to their respective topologically associated domains and observed that the overall distribution of TAD sizes is approximately the same for genes expressed specifically in motor neurons compared to other tissues (Figure 3.8.B). This again, supports the idea that disproportionately large intergenic regulatory domains surround neuronal genes, as there is more intergenic space yet the size of the regulatory domain is not changing. Indeed, as with motor neuron genes, when I normalized for gene density within a TAD I observed that neuronal genes tend to be located in TADs with much larger intergenic regions and fewer genes (Figure 3.8.C). It is worth noting that the ENCODE data and motor neuron data are from adult and embryonic ages respectively suggesting this is likely a general feature of neuronal gene regulatory organization despite independent of developmental age.

Having established this unique regulatory organization around neuronal genes, I next sought to investigate what might explain this difference. One possibility is that a larger intergenic regulatory domain might contain a larger number of regulatory elements and therefore have an increased regulatory capacity. However, while increased sizes in intergenic regions are suggestive for an increased regulatory potential they are merely suggestive. It's entirely possible that the increased intergenic space around these genes is non-functional or made up of repetitive elements (Singer et al., 2010). I reasoned that in order to facilitate the increased requirement of regulatory complexity there would actually need to be an increase in the number regulatory regions surrounding neuronal genes. To this end, I took advantage of recent studies from the ENCODE project that identified enhancers in many mouse tissues and their

general characteristics of conversation (Shen et al., 2012). We mapped conserved intergenic regions in 500bp windows throughout the genome to identify putative regulatory elements and asked how they were distributed with respect to regulatory domains containing tissue specific genes. Interestingly, I observed no clear differences in the frequency and density of putative regulatory regions with respect to tissue suggesting that there is approximately the same frequency of regulatory region throughout the genome (Figure 3.8.D). This data supports the idea that if a larger regulatory domain surrounds a gene it would also have more regulatory elements. Remarkably, when I computed the frequency of putative regulatory regions with respect to gene density I observed a specific increase around genes expressed in the nervous system (Figure 3.8.E). Together, these data reveal that disproportionately large intergenic regulatory domains with the potential to facilitate more regulatory potential surround genes expressed in the nervous system.

I next investigated potential explanations for this increase in regulatory control of neuronal gene expression. I reasoned that neuronal genes might require larger regulatory domains to allow for specific enhancer regulation within the complexity of the nervous system. For example, many genes are expressed broadly across the nervous system and my studies on Isl and Lhx3 control of motor neuron identity suggests that the master regulatory factors within a given neuron are likely to control their expression. Based on this idea, I hypothesized that these large domains would allow for cell-type specific utilization of regulatory elements. To investigate this idea I compared DNA accessibility (DHS and ATAC) within multiple distinct neuronal cell types from different ages and regions of the nervous system. As a start I first analyzed the local DNA accessibility patterns from embryonic spinal motor neurons, adult cerebellum, and three distinct cortical cell types, excitatory cells (Camk2a), VIP interneurons and PV interneurons in the vicinity of the neuronal splicing factor Nova1 (Jensen et al., 2000; Mo et al., 2015)(Figure 3.8.F). Nova 1, expressed broadly across the nervous system to control neuronal alternative splicing programs, is located in a remarkably large regulatory domain spanning more than four megabases in the genome implying it might require a complex regulatory organization to specifically control it’s expression (Jensen et al. 2000; Ule et al. 2003). Upon manual investigation, I identified multiple unique regions of DNA accessibility supporting the model of unique regulatory utilization in the

nervous system (Figure 3.8G). This initial hint towards the regulatory control was somewhat biased as I was manually scanning regions to look for differences in accessibility.

To investigate this idea in a more unbiased manner I analyzed the accessibility patterns in regulatory domains around genes with fairly broad expression patterns throughout the nervous system including, neuronal splicing factors, Nova1, Rbfox3 and Ptbp2, neurotransmitter receptors, Grin3a, Gria3 and Gabrb1, synaptic proteins, Syt1, Nlgn1, and Cntnap5a/b, and finally, cell adhesion molecules including, Nrcam, Ncam1, Ncam2 and Pcdh genes (Figure 3.9.B). Most, but not all of these genes are located in extremely large regulatory domains up to 5MB which in theory should provide immense regulatory complexity. Based on results from this study and others, regulatory regions are restricted within insulated TAD domains to specifically control their target genes (Dixon et al., 2012, 2015; Dowen et al., 2014; Guo et al., 2015). Using this assumption we assigned all distal peaks of DNA accessibility to their respective regulatory domains around this subset of broadly expressed neuronal genes and then mapped regions of unique DNA accessibility. Remarkably, we observed broad patterns of unique accessibility in both age and cell type specific fashion with all genes analyzed displaying multiple unique peaks of distal accessibility (Figure 3.8H). Taken together these data support a model where neuronal genes reside in remarkably large, gene poor regulatory domains surrounded by an increase in putative regulatory elements. This increase in regulatory elements might facilitate unique modes of control in a neuron specific manner. All together, these results imply that the cellular complexity in the nervous system is matched by a similarly, if not more complex regulatory complexity in non-coding DNA.

Discussion

In this study, I use an integrated approach, generating maps of distal regulatory elements in both *in vivo* and *in vitro* derived motor neurons to study the chromatin level mechanisms controlling expression of postmitotic motor neuron identity genes. Using this complementary set of data, this study provides the first genome-wide map of high-resolution enhancer-promoter interactions in a single mammalian neuron during development. I identified a set of enhancers that are dynamically established during development from inaccessible DNA to directly control the expression of motor neuron identity genes via direct

enhancer-promoter interaction. A large majority of these enhancers are bound by the motor neuron programming factors Isl1 and Lhx3, which broadly control the expression of the core motor neuron identity program. These core identity genes are controlled by constellations of enhancers distributed across remarkably large regulatory domains. Finally, I show that this distributed organization within large regulatory domains is a general principle of enhancer organization within the nervous system and could explain for the mechanisms controlling specify of gene expression within the diversity of the nervous system.

Programming factors control the postmitotic expression program in motor neurons

Dynamic changes in the expression program during cellular specification are largely attributed to a correspondingly dynamic change in the distal enhancer landscape associated with the dynamic binding of programming transcription factors. However, causal linkage of transcription factor binding to transformation in gene expression programs has remained elusive. Here I utilized an *in vitro* programming paradigm to study the role of transcriptional specification in the nervous system. Using this highly purified system I was able to uncover the general principles of regulatory control by the programming transcription factors Isl1 and Lhx3. The transcription factors synergistically regulate the core identity of postmitotic motor neurons controlling both aspects that are highly cell type specific and those that are more generic in the nervous system. For example, Hb9, a transcription that is highly specific to motor neurons in the nervous system is regulated by multiple Isl/Lhx3 bound enhancers. Additionally, Isl/Lhx3 regulate genes that are more ubiquitously expressed in the nervous system including axon guidance molecules and structural components associated with neuronal morphology (Dcc, Robo1/2, Nefm/Nefm). Together, these results reveal a common theme of neuronal gene regulation, namely co-regulation. The model of co-regulation has been well appreciated in simpler nervous systems like *C. elegans* where the genetic dissection of regulatory elements is more straightforward given their proximal nature (Deneris and Hobert, 2014; Flames and Hobert, 2009; Hobert, 2011; Kratsios et al., 2012). Additionally, co-regulation was observed in a rare case in mammalian neurons where a transcription factor binds primarily to proximal promoters, which allowed for more direct association with target genes (Lodato et al., 2014). However, causal assignment of distal enhancers to their targets and the scale of global co-regulation remained

unclear. Together, combined with the previous examples, this study suggests a general model of gene regulation in the nervous system, conserved from worms to mammals, where master regulator transcription factors (or terminal selectors) regulate the majority of cell identity genes in a cell specific manner using specific combinatorial codes of transcription factor activity. These studies suggest the requirement for a potentially complex regulatory mechanism where genes expressed in the nervous system will require a regulatory environment to facilitate regulation of expression programs in a cell type specific manner.

Distributed enhancer constellations in motor neurons

In this study, I identified the mechanisms controlling robust and specific expression of cell identity genes in motor neurons. Multiple distinct enhancers that are organized in “constellations” regulate motor neuron genes. These constellations are distantly distributed over large swatches of genomic DNA with interactions spanning up to megabases. This finding was initially surprising given the recent studies describing spatially clustered super-enhancer organization around cell identity genes whose specificity is controlled by their location within insulated regulatory domains (Dowen et al., 2014; Hnisz et al., 2013, 2016; Whyte et al., 2013). Remarkably, despite the long distance organization of motor neuron enhancers their interactions still tend to be restricted within the same regulatory domain as their target gene suggesting a similar overall model of specificity despite the increase in size of regulatory regions.

Interestingly super-enhancers were previously described to regulate cell identity genes in the brain of mice, but these results were not without caveats (Hnisz et al., 2013). The data analyzed in these studies come from whole brain tissue instead of a purified single cell type. Conversely, in my studies using pure populations of motor neurons we do not see clusters of enhancers, but distantly distributed enhancers. Based on my results it is likely that previous described super-enhancers in brain come from the mixed population of cell types in the tissue. What then explains the difference in clustered versus distributed regulation of gene expression?

One possible explanation to reconcile the difference in regulatory organization around cell identity genes could be the regulatory requirements to maintain spatiotemporal expression of a gene during

development. For example, if a gene were expressed in a fairly restricted manner it would likely need much less regulatory complexity to control its expression in a cell type specific manner. My data suggests that both genes that are more specific for motor neurons and genes that are expressed more ubiquitously in the nervous system are under the control of the programming transcription factors Isl1 and Lhx3. If we assume that each cell identity gene is under the control of a unique set of programming factors in each cell type then a gene would need at least one unique distal enhancer per cell type and by extension more regulatory space around that gene. Accordingly, Oct4 and Nanog, core regulators of the pluripotency program in mESCs, are expressed in a fairly restricted manner while many of the motor neuron identity genes (Dcc, Slit2, Nrp1, Nefl, Isl1) are expressed broadly across the nervous system and even in other tissues. Together, this study provides initial insight into how complex patterns of gene expression in the nervous system might translate into correspondingly complex regulatory organization in the genome.

Despite the inherent difference in the spatial organization of enhancers in motor neurons compared to other cell types. The mode of regulation by the enhancers is likely shared. Whether the enhancers are spatially clustered or distantly distributed cell identity genes are controlled by multiple distinct enhancers that are acting concomitantly to control expression of their target gene within insulated regulatory domains. Mechanistic understanding of regulation of gene expression by multiple enhancers is still very poorly understood with multiple proposed models including additive, synergistic and combinatorial regulation. The dissection of Isl1 gene expression tested in this study suggests that each individual enhancer acts in an independent and probabilistic manner. That is, they each consistently activate expression of Isl1 at a specific level and their compound deletion results in a gene expression change that is the product of their individual contributions. This dynamic probabilistic model is in agreement with recent live imaging studies in drosophila that support a bursting model of transcription with different enhancers exhibiting different kinetics of transcription activation. Together these studies suggest a model for robust activation of gene expression by multiple independently acting enhancers.

We posit that this system is essential for the establishment and maintenance of cell identity in developing organisms. Cell identity genes are likely regulated by multiple enhancers to ensure stable and robust expression of both key effector genes and regulators of cell identity. Such a system would protect gene expression programs and thus cell identity from mutational perturbation. Indeed, the shadow

enhancer model in drosophila supports this idea where multiple genes are regulated by two distinct enhancers, which allows for robust maintenance of gene expression in the face of environmental perturbations (Hong et al., 2008; Spitz and Furlong, 2012). One question, which remains open for investigation, is the how multiple enhancers work within the dynamic nuclear environment. Are both enhancers active at the same time in the same cells? Are the enhancers engaged with target genes at the same time? Is there synergism or mutual exclusivity between enhancers? Live imaging experiments would likely provide the best approach to gain insight into the mechanisms of enhancer action that are not easily investigated in the population studies performed in this study.

Regulatory complexity in the nervous system

My studies dissecting the regulatory logic of motor neuron identity uncovered a unique property of enhancer organization in the nervous system. Based on my findings that Isl1 and Lhx3 co-regulate the core identity program, I hypothesized that neuronal genes would require an increased regulatory complexity for master transcription factors to control gene expression in a cell type specific manner. Presumably the identity of each type of neuron will be under the control of a unique combination of programming transcription factors. These programming factors will each bind to distinct enhancer elements to control expression of their target genes. Depending on the definition of a cell type, the nervous system is composed of arguably hundreds to many thousands of distinct classes of neurons suggesting there could be thousands of different regulatory elements for a given gene that is expressed across the nervous system. Accordingly, my preliminary analysis of distinct components of neuronal identity including synapse formation, ion channel expression, neurotransmitter identity and neuronal RNA splicing supports this model. Analysis of DNA accessibility revealed that each gene is surrounded by many hundreds of distinct peaks of DNA accessibility suggesting a largely complex regulatory environment.

Considering that this study is the first step into the control of neuronal identity within the complexity of the mammalian nervous system, many new questions arise. As I have shown, analysis of DNA accessibility is not sufficient to predict enhancer activity. It will be important to investigate to what

extent this differential accessibility observed between different neuron types correlates with enhancer activity and regulation of gene expression. My analysis of motor neuron enhancers suggest each gene is regulated by a few distinct enhancers, but there are many additional peaks of DNA accessibility around the gene. Analysis of cortical excitatory and inhibitory neurons revealed many more peaks of accessibility than seen in motor neurons (Mo et al., 2015). It is possible, but unlikely that other classes of neurons would have many more enhancers. Instead, I would hypothesize that these studies lack the purity of cell types and thus likely represent the regulatory landscapes of many distinct subtypes of excitatory and inhibitory neurons. Indeed, this is supported by the fact that within the cortex there are distinct laminar differences in transcription factors expression and immense subtype diversity (Arlotta et al., 2005; Lodato and Arlotta, 2015; Molyneaux et al., 2007). It will be important to repeat these experiments with more specific populations to validate this model and determine how far the regulatory complexity of enhancers extends. Indeed, the nervous system is made of many thousands of cell types many of which are closely related, but specified into subtypes upon differentiation (Dasen and Jessell, 2009; Fishell and Heintz, 2013; Lodato and Arlotta, 2015; di Sanguinetto et al., 2008). How stable are patterns of enhancer activity during subtype specification? Do cells maintain expression of identity genes through stable activity of distal enhancers? We will investigate these questions in Chapter 4.

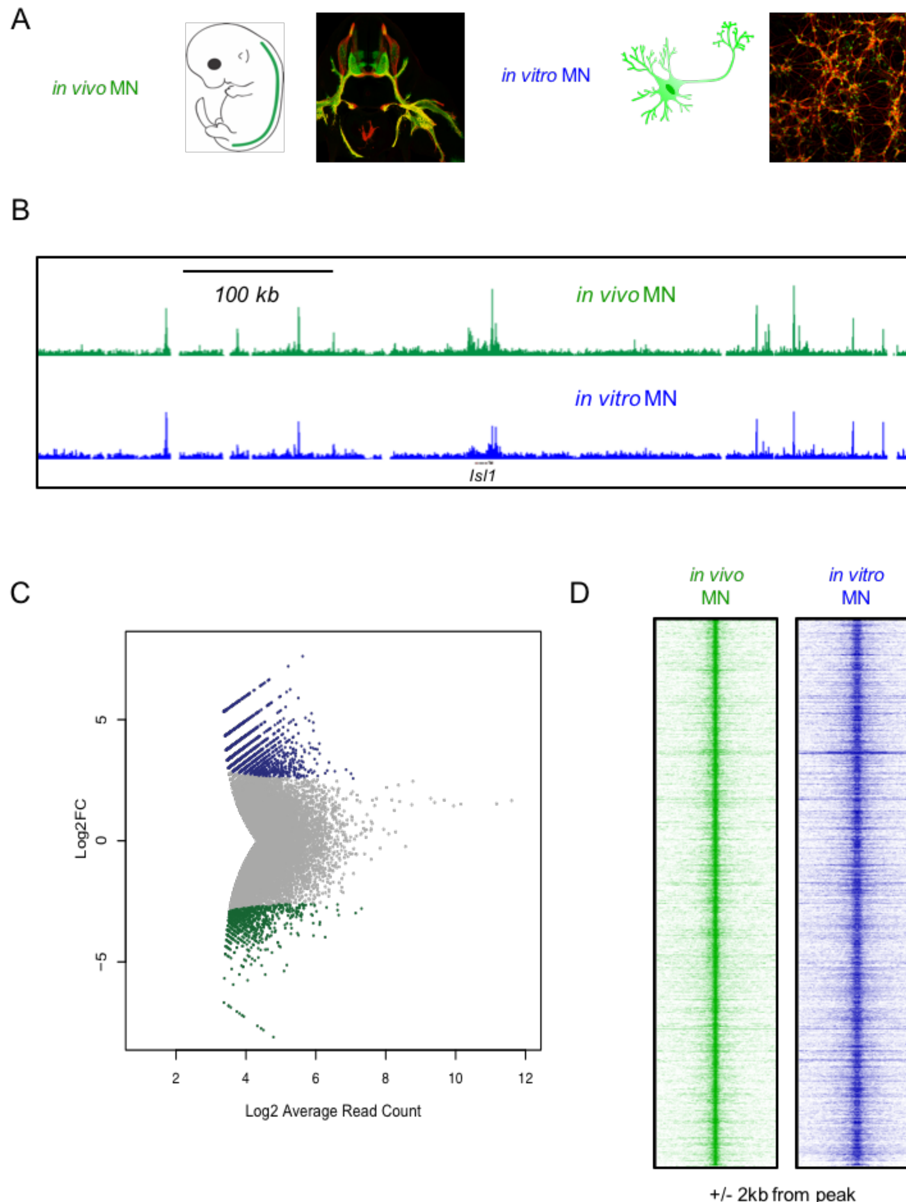


Figure 3.1. A comparison of ATAC-seq based chromatin accessibility in *in vitro* and *in vivo* motor neurons. (A) Schematic representation of *in vivo* and *in vitro* sources of motor neurons for ATAC-seq assays. GFP-positive primary motor neurons were purified from Hb9::GFP expressing mice. GFP-positive induced motor neurons were FACS purified from NIL-programmed motor neurons. **(B)** Genome browser view of peaks of DNA accessibility in an approximately 600kb region surrounding the Isl1 locus in *in vivo* (green) and *in vitro* (blue) derived motor neurons. The patterns of accessibility are larger similar around the Isl1 gene. **(C)** A global quantitative comparison of distal ATAC-seq peaks between *in vivo* and *in vitro* derived motor neurons. More than 90% of peaks are not significantly changed between the two samples (Log FC<2, p<.01; shared(gray), in vivo enriched (green), in vitro enriched (blue)) (**D)** Line plots showing the global binding pattern of shared peaks of DNA accessibility between *in vivo (green)* and *in vitro* (blue) derived motor neurons. Raw sequencing tags are plotted in 2kb windows surround peaks coordinates.

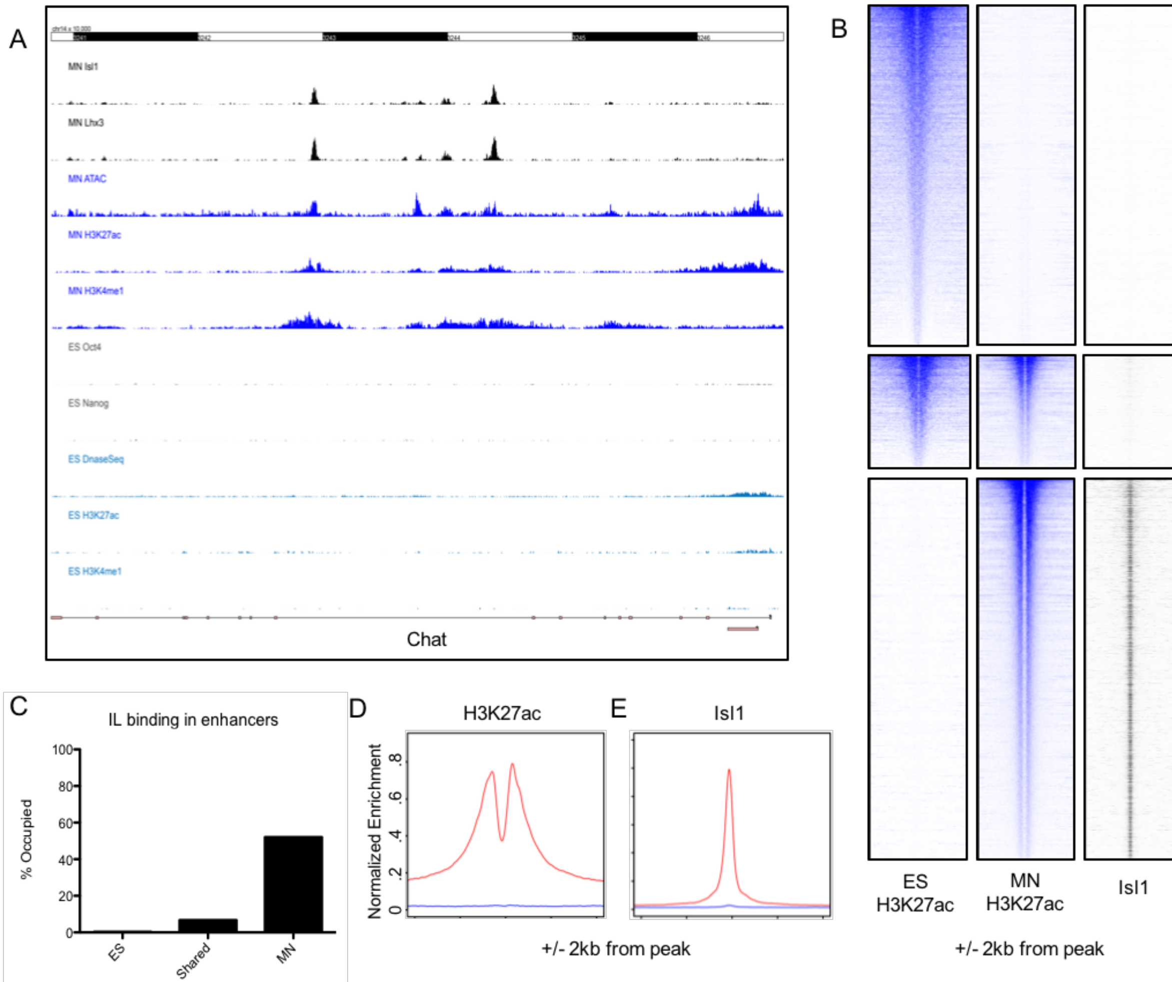


Figure 3.2. Isl/Lhx3 binding drives dynamic reorganization of active enhancer landscape in motor neurons. (A) Genome browser view of Isl and Lhx3 (black tracks) binding to two intronic enhancers of the choline acetyl transferase (Chat) gene in motor neurons. IL binding is coincident with new established regions of enrichment for H3K27ac, H3K4me1 and DNA accessibility in motor neurons, which is completely absent in mouse embryonic stem cells. **(B)** Genome-wide maps of active enhancers in motor neurons and mESCs. Enhancers are highly dynamic with the majority new established during specification. Newly established enhancers are highly coincident with Isl1 binding (black). **(C)** Quantification of IL binding in enhancers from mESCs, Shared, and motor neurons. **(D and E)** Composite plots of normalized enrichment of H3K27ac (D) and Isl1(red)/Oct4(blue) (E) at Isl1 binding sites in motor neurons (red) and ESCs (blue).

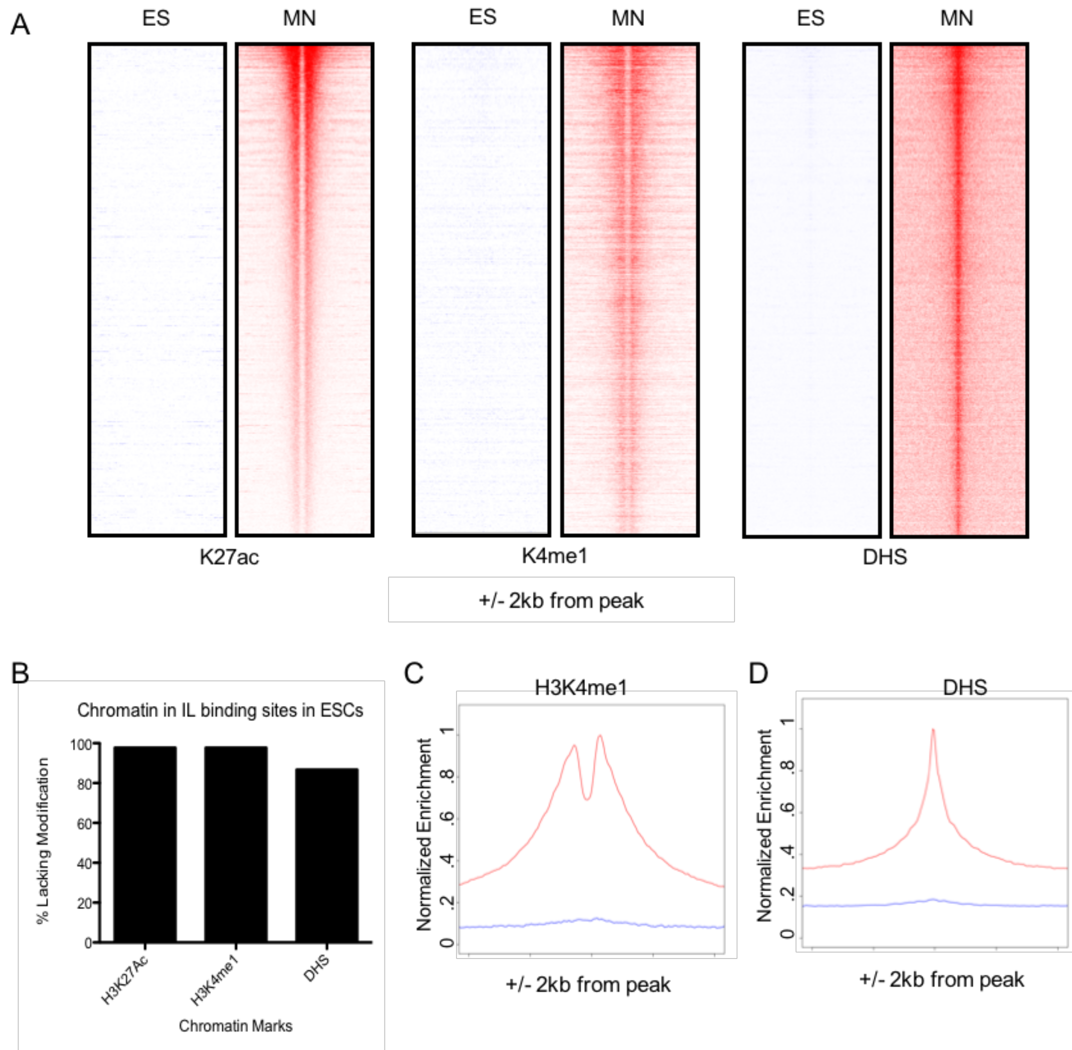


Figure 3.3. Isl/Lhx3 bound enhancers are pioneered during motor neuron specification. (A) Line plots showing genome-wide enrichment for signatures of primed enhancers. Newly established enhancers in motor neurons (red) lack evidence of any form of priming in ESCs (blue). Enhancers are highly dynamic with the majority new established during specification. **(B)** Quantification of newly established enhancers that lack priming signatures in mESCs. **(C and D)** Composite plots of normalized enrichment of H3K4me1 (D) and DNase1 hypersensitivity (DHS) (D) corresponding to line plots in (A) showing all regions.

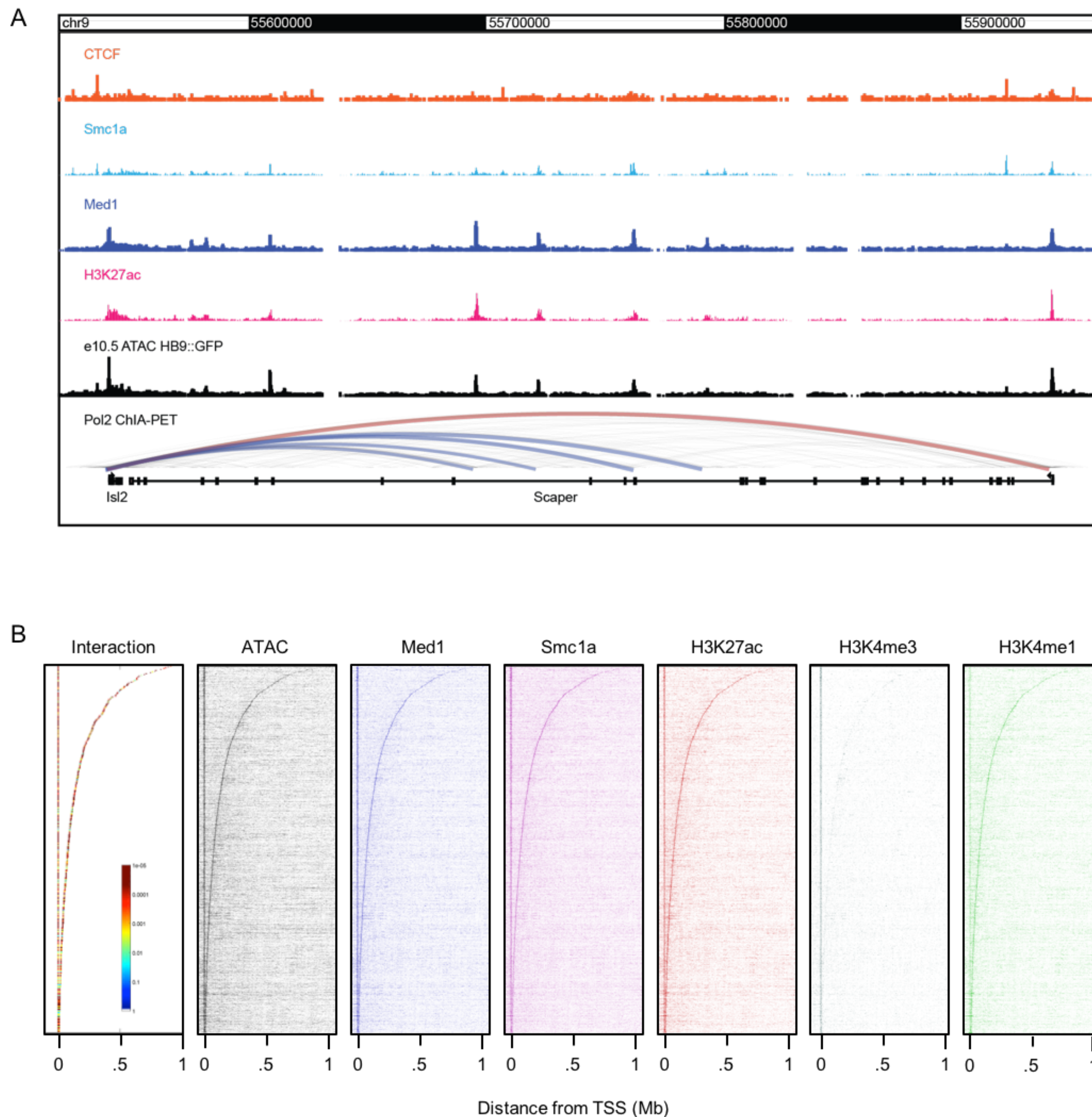


Figure 3.4. Isl/Lhx3 binding drives dynamic reorganization of active enhancer landscape in motor neurons. (A) Genome browser view of high-confidence interactions mapped by CID at the Isl2 locus in motor neurons. Distal enhancers are highly occupied for previously identified enhancer associated co-factors and histone modifications. Enhancer-promoter interactions are in blue and Promoter-Promoter interactions are in red. **(B)** Genome wide maps of all enhancer-promoter pairs identified by CID. Heatmap (left) represent the signficiance of interaction (FDR) based on CID. Enhancers and promoters are highly enriched for local DNA accessibility (black), structural protein complexes associated with enhancer-promoter interactions (blue and magenta), histone modifications associated specifically with enhancers (red and green) and promoters (red and cyan).

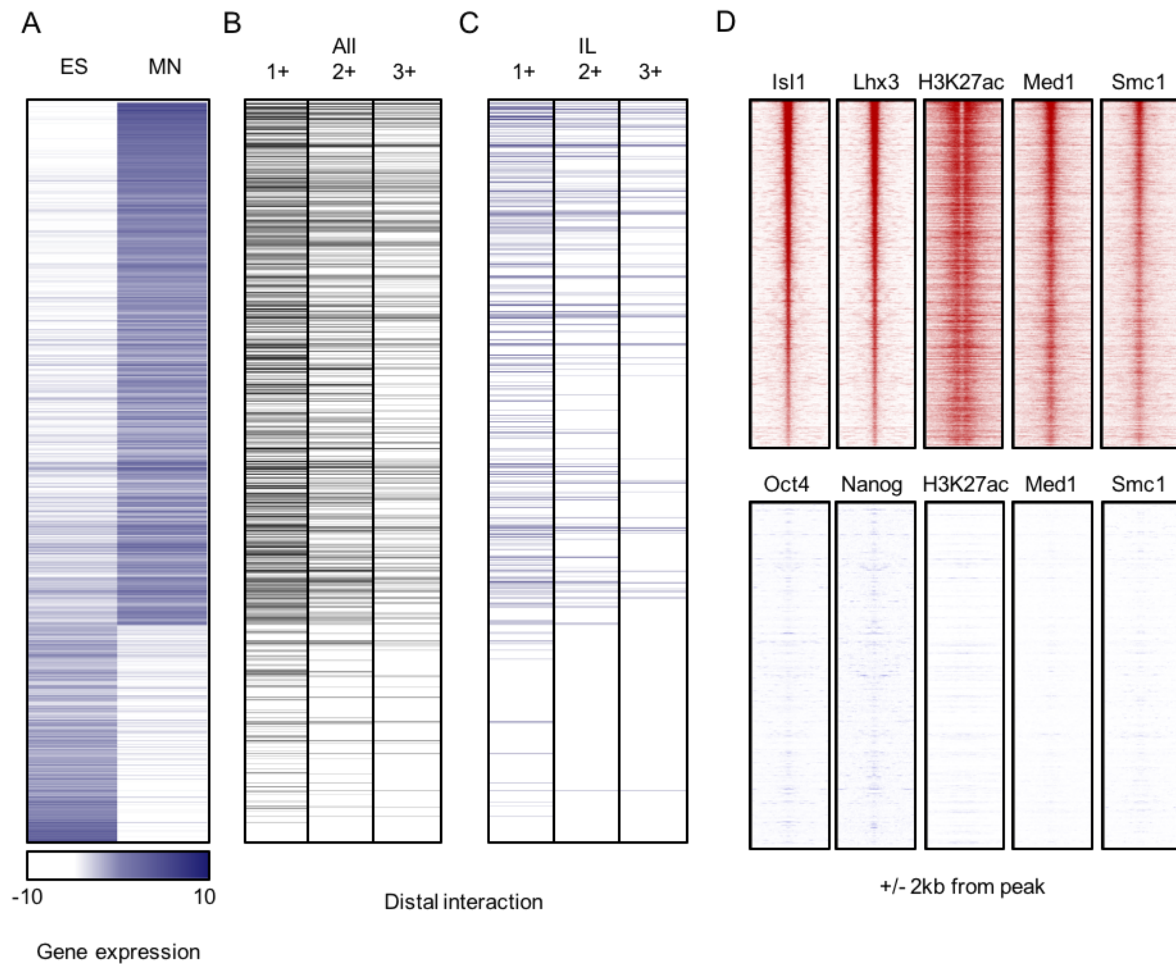
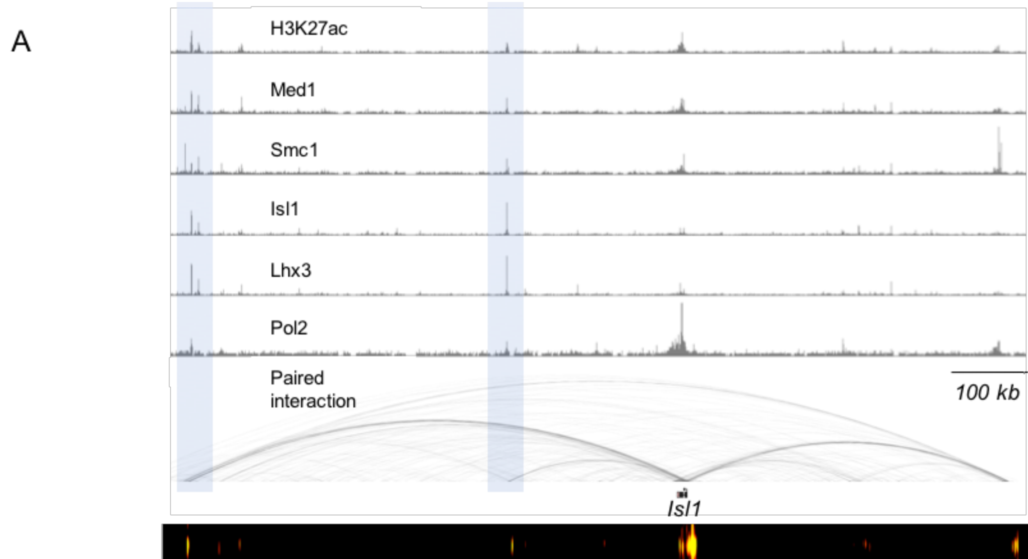
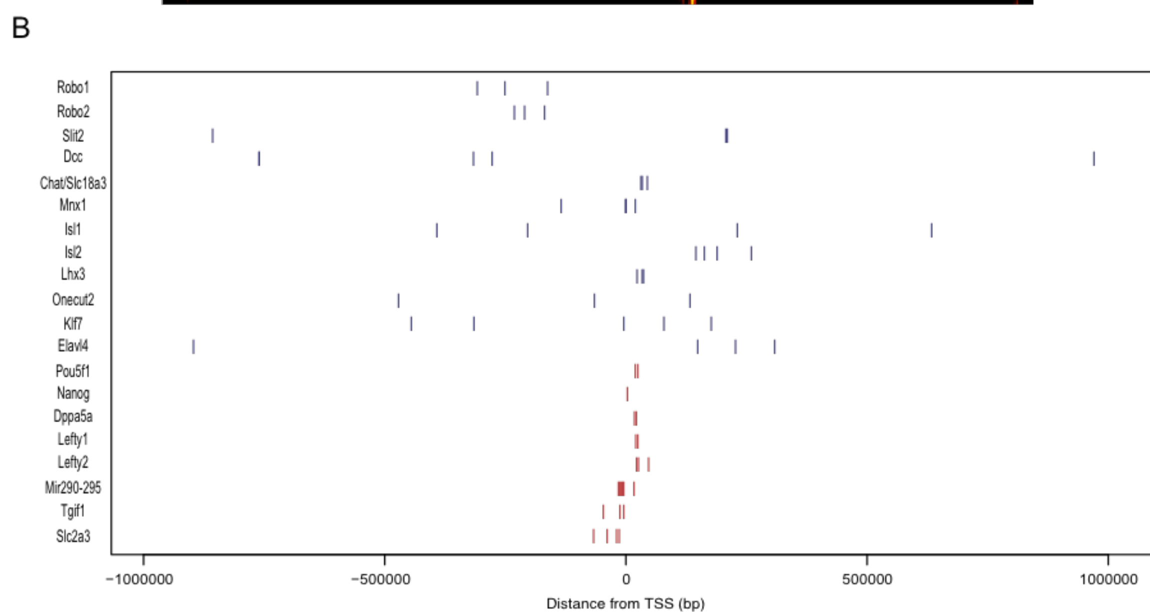


Figure 3.5. IL specifically induce MN gene expression program through E-P interaction. (A) Heatmap of genes differentially expressed between ES and iMNs. Heatmap shows Log2 FPKM values for one gene per row. **(B and C)** Interactions between differentially expressed genes with distal enhancers. (All enhancer (B) (black ticks), IL bound enhancers (C) (blue ticks)). Each row of the plot corresponds to one gene (same gene as heatmap for expression). Each column in the heatmap corresponds to the number of interactions for that gene (1+, 2+, 3+ interactions). Interactions are largely restricted to induced genes. **(D)** Line plots showing enrichment for Isl1 and Lhx3 transcription factor binding, cofactor co-occupancy (mediator (Med1) and cohesin (Smc1)) and H3K27ac activating histone modifications at enhancers engaged with motor neuron identity genes. Enhancers are plotted as one 2kb region per line anchored on Isl1 binding sites.

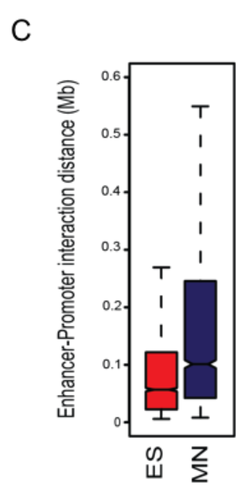
A



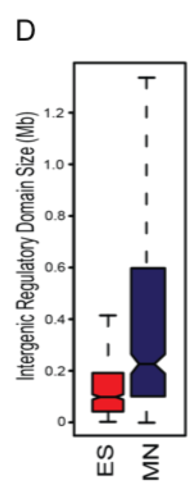
B



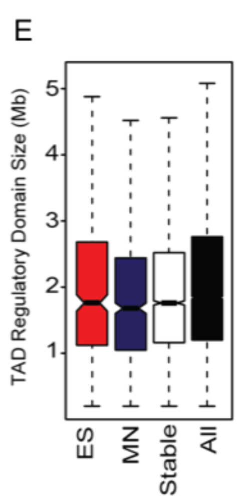
C



D



E



F

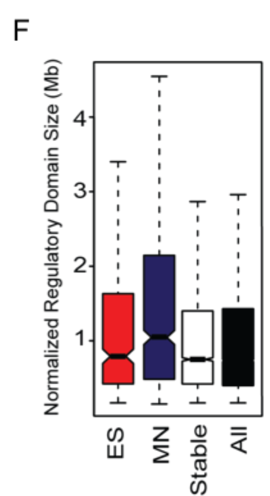


Figure 3.6. Regulation of MN expression program from distantly distributed enhancer constellations. (A) Genome browser view of the Isl1 locus and its distal enhancer interactions. Isl1 is regulated by two high-confidence interactions with Isl and Lhx3 bound enhancers located ~220 and ~600kb downstream of the TSS (blue highlights). **(B)** Distributed long distance enhancer constellations around motor neuron identity genes (blue ticks) and spatially clustered proximally located enhancers around ESC identity genes. Genes regulating all aspects of motor neuron identity have multiple distributed enhancer interactions. **(C)** Comparison of enhancer-promoter interaction distances in mESCs (red) and motor neurons (blue) **(D)** Comparison of intergenic regulatory domain sizes of genes uniquely expressed in mESCs (red) and motor neurons (blue) **(E)** Comparison of topologically associated domain sizes associated with genes uniquely expressed in mESCs (red) and motor neurons (blue), stably expressed between the two cell types (white) or all genes (black). **(F)** Comparison of regulatory region size normalized to the gene density within a topologically associated domain associated with genes uniquely expressed in mESCs (red) and motor neurons (blue), stably expressed between the two cell types (white) or all genes (black).

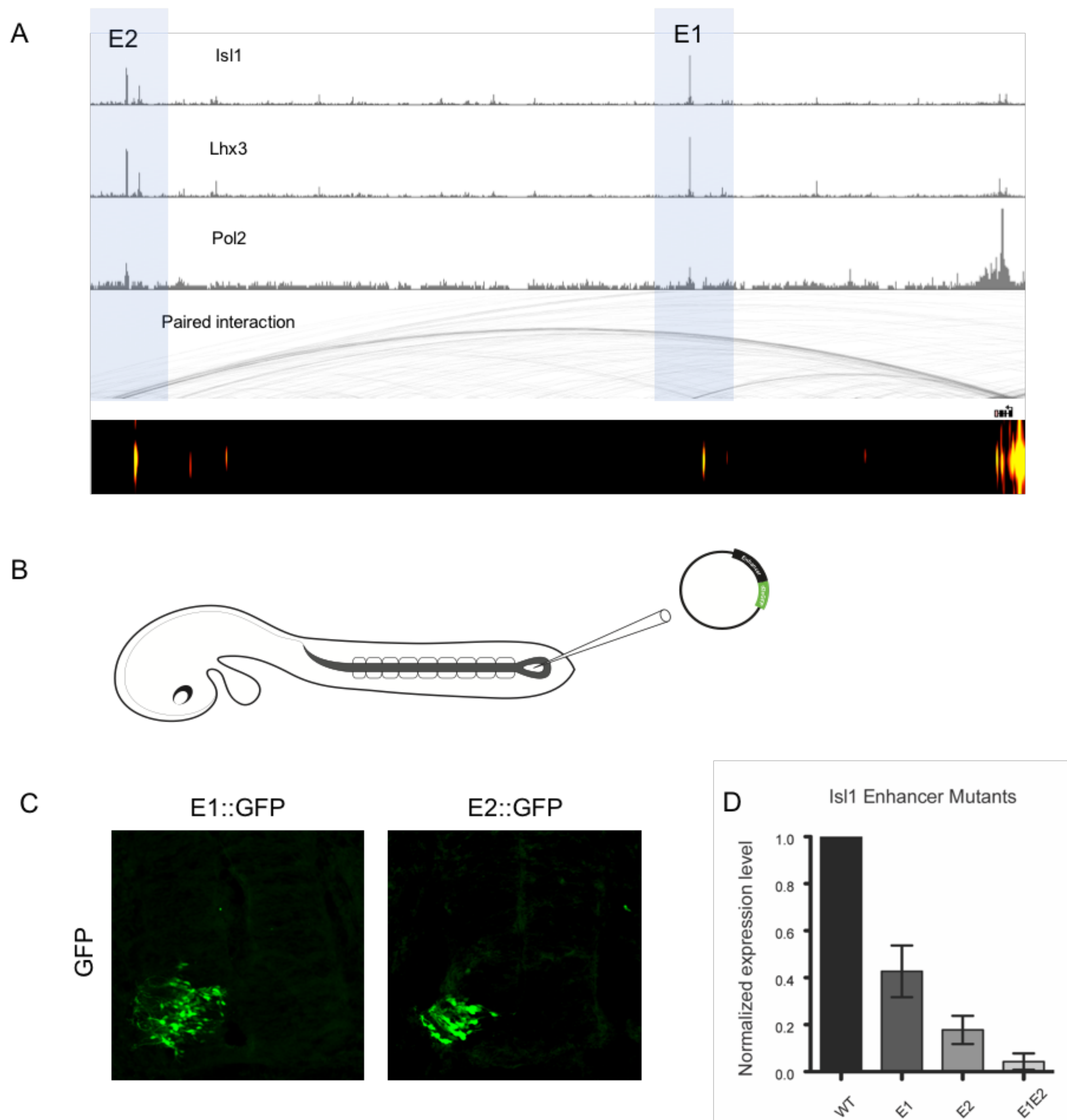
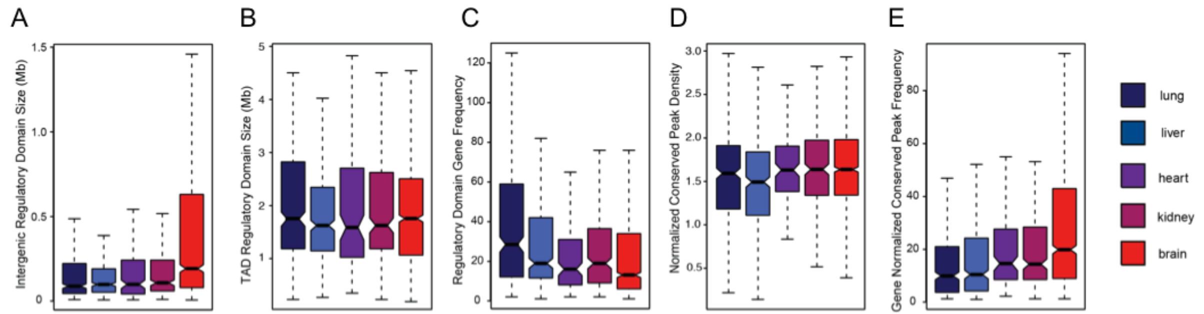
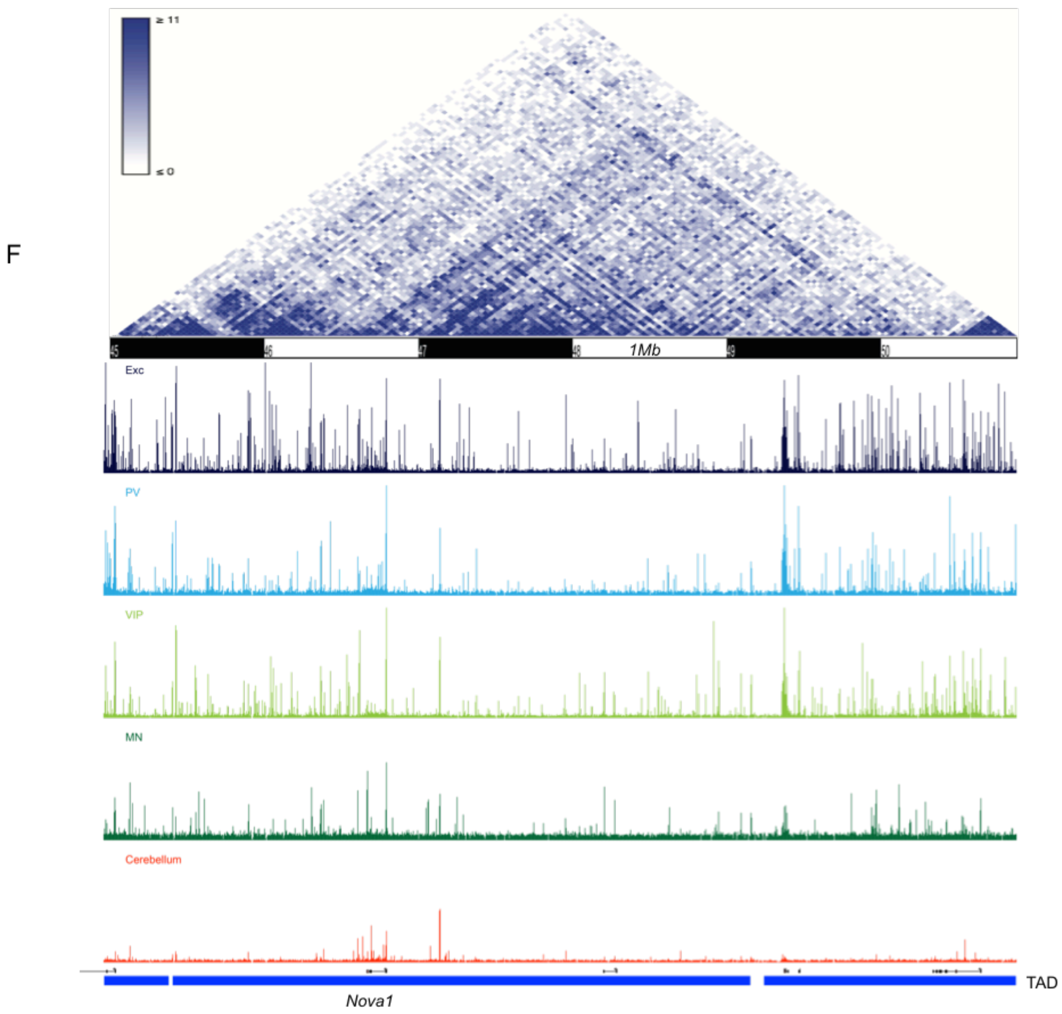


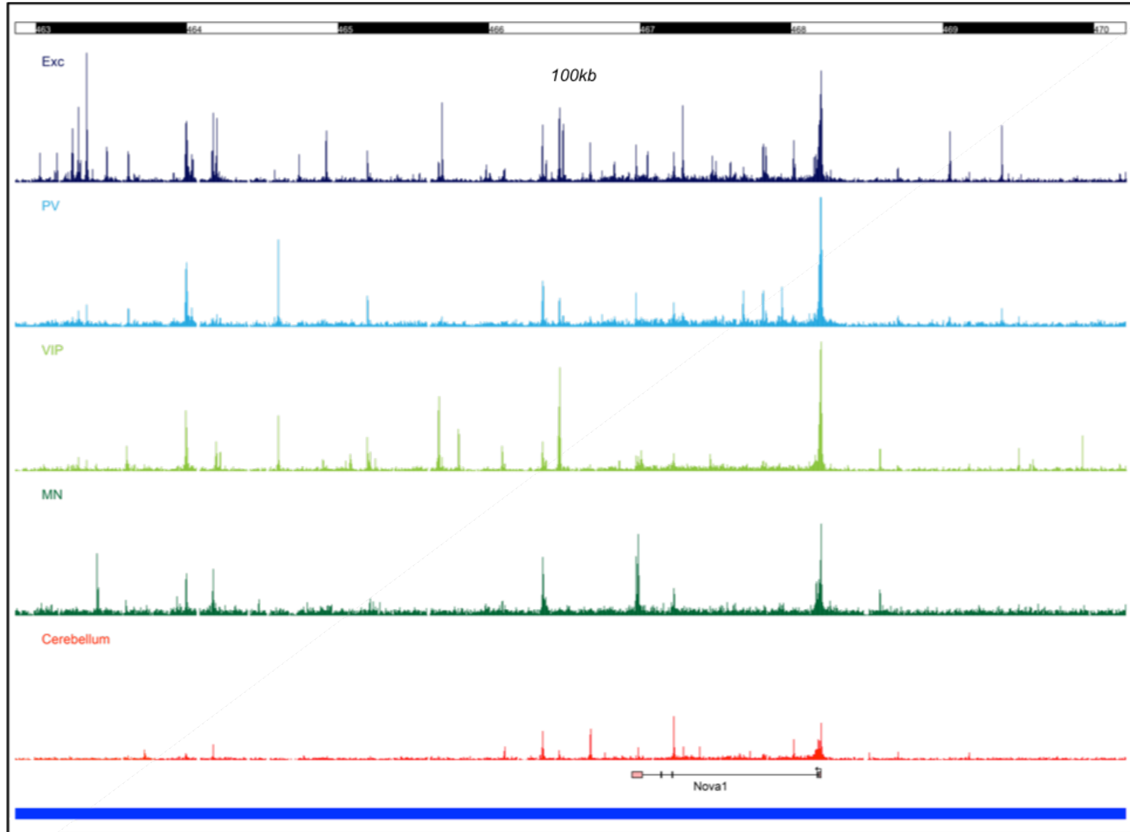
Figure 3.7. Functional dissection of distributed enhancers in motor neurons. **(A)** High-confidence interactions surrounding the master regulator of MN identity Isl1. Two high-confidence enhancers (highlighted in blue) were detected as interacting with the Isl1 gene at ~220kb (E1) and ~600kb (E2) downstream of the TSS. **(B)** Schematic representation of experimental set up for *in ovo* electroporation. ~1kb enhancer regions drive destabilized GFP in a reporter construct electroporated into the neural tube of developing chick embryos. **(C)** Electroporation of enhancer reporters in developing embryonic spinal cords in chick. E1 and E2 enhancer reporters drive d2EGFP expression specifically in the ventral domain of the spinal cord where motor neurons reside. (D) Expression analysis of enhancer mutant cell lines after CRISPR/Cas9 mediated genome engineering. Enhancer mutation leads to significant loss of Isl1 expression levels. Single deletion is insufficient to block Isl1 expression suggesting individual input from each enhancer.



F



G



H

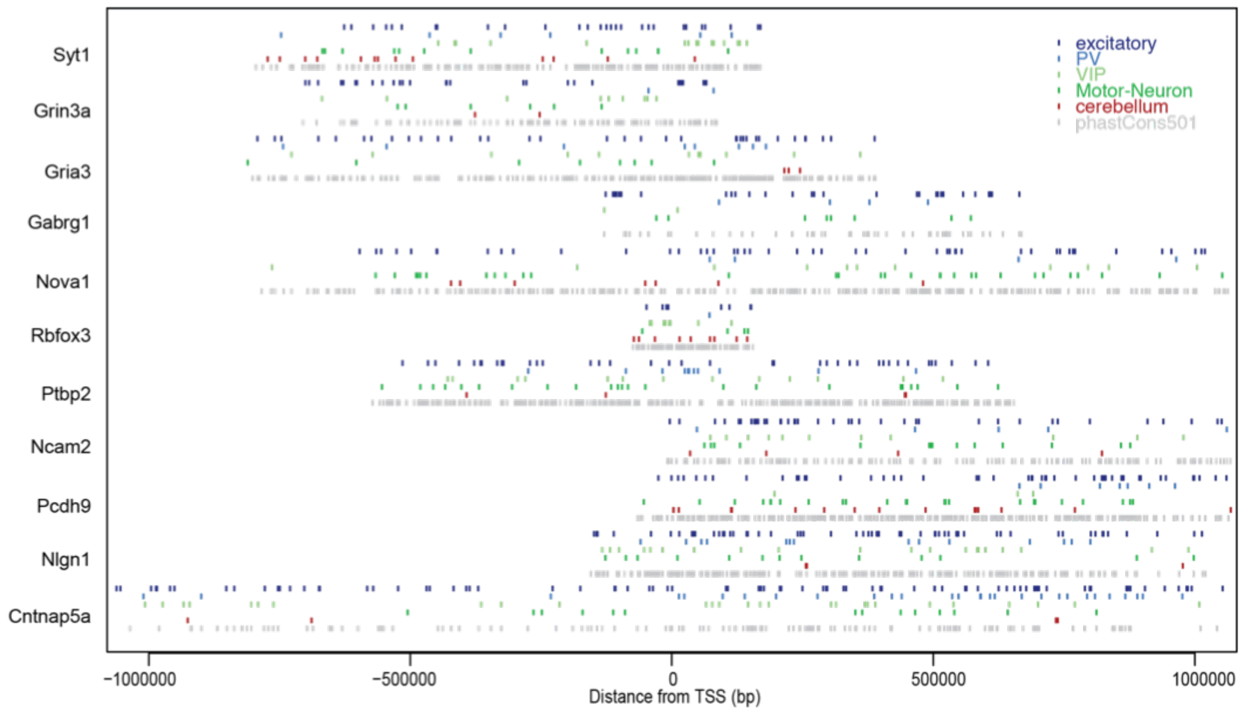


Figure 3.8. Neuronal genes are associated with an increased regulatory complexity in large gene poor regulatory domains. (A) Comparison of intergenic regulatory domain sizes of genes specifically expressed in different tissues at 8 weeks in adult mice. Lung(navy), Liver(blue), heart (purple), kidney (magenta), brain (red). **(B)** Comparison of topologically associated domain sizes associated with genes uniquely expressed in each tissue **(C)** Comparison of gene density within a topologically associated domain associated with genes uniquely expressed in each tissue. **(D)** Comparison of conserved peak density (proxy for putative regulatory elements) each tissue. **(E)** Comparison of conserved peaks per gene in each tissue (i.e. regulatory capacity around each gene). **(F)** Genome browser view displaying the regulatory complexity around neuronal genes. Approximately 5 megabase region surrounding the Nova1 locus. Nova1 is located in a TAD (blue bar on bottom of plot) that spans more than 3 megabases in the genome. DNA accessibility from 5 different neuronal cell types: Cortical excitatory (Navy), Cortical PV-positive interneurons (cyan), Cortical VIP-positive interneurons (dark green), Spinal motor neurons (light green), and Cerebellum (red). DNA accessibility shows hundreds of unique peaks spanning across the entire regulatory domain suggesting immense regulatory complexity. **(G)** Genome browser view of the same region zoomed in to ~700kb window around the Nov1 gene. Cell types are same as previous plot. Each cell type displays complex patterns of accessibility that are composed of both unique and overlapping regions in different cell types. **(H)** Plots showing unique regions of accessibility in a cell type specific manner in a 2 megabase window broadly expressed neuronal genes. Each tick mark corresponds to a unique region of accessibility in a specific cell type. Each cell type is colored the same as genome browser plots and on its own line in the plot.

Chapter 4

Expression of terminal effector genes in mammalian neurons is maintained by a dynamic relay of transient enhancers

Ho Sung Rhee, Michael Closser, Yuchun Guo, Elizaveta V. Bashkirova, G. Christopher Tan, David K. Gifford, Hynek Wichterle

The aim of this study was to uncover the dynamic regulatory mechanisms that establish and maintain cell identity during the specification and maturation of motor neurons. The data in this chapter are published in Rhee et. al Neuron 2017. The text is adapted from the published manuscript, which I co-wrote.

Author contributions:

H.S.R., M.C. and G.C.T. performed ChIP-seq experiments. H.S.R. performed ChIP-exo experiments. G.C.T. generated the inducible Ngn2 cell line. M.C. and E.V.B. performed ATAC-seq and DNase-seq experiments. H.S.R. and M.C. performed RNA-seq experiments. H.S.R. and M.C. performed *in ovo* electroporation experiments. H.S.R. and M.C. performed CRISPR mediated enhancer deletions. Y.G. performed computational analyses of deep sequencing data. H.S.R., M.C., D.K.G., and H.W. conceived the experiments and analyses. H.S.R., M.C., and H.W. co-wrote the manuscript.

Introduction

The vertebrate central nervous system is a complex organ composed of many thousands of neuronal subtypes organized in a highly stereotypical pattern. Construction of such a sophisticated system relies on the precise and reproducible establishment of cell type-specific gene expression programs that define the morphology, connectivity, and functionality of individual neuronal subtypes.

The acquisition of neuronal identity is driven by a developmental transcriptional program executed in neural progenitors that culminates in the activation of a unique combination of transcription factors (TFs) in nascent postmitotic neurons (Bertrand et al., 2002; Shirasaki and Pfaff, 2002). These TFs are often referred to as master regulators or programming factors, as they are both necessary and sufficient to activate expression of cell type-specific combinations of effector genes that define neuronal identity and functionality(Hobert, 2008; Vierbuchen et al., 2010). How programming TFs selectively recognize their genomic binding sites and which of their bound regions will get activated and regulate target genes have been the topic of intense scrutiny, yet our understanding of the process remains rather limited. Furthermore, many programming TFs are expressed only transiently in the developing mammalian nervous system and are rapidly downregulated in maturing postmitotic neurons, raising the question how expression of the cell type-specific effector genes is maintained in maturing neurons (Son et al., 2011; Wapinski et al., 2013).

Recent studies established that enhancers in progenitor cells transitioning from one developmental state to another are highly dynamic (Hong et al., 2013; Johnson et al., 2011). In contrast to dividing cells, many genes induced in postmitotic neurons (axon guidance receptors, neurotransmitters and receptors, ion channels, etc.) operate as terminal effector genes and remain stably expressed, despite the downregulation of TFs that initiated their expression. Whether enhancers established in nascent postmitotic neurons become stabilized by the recruitment of secondary TFs, or whether the continuity of expression of postmitotic neuronal effector genes depends on a transcriptional cascade engaging independent stage-specific enhancers in maturing motor neurons, remains unknown.

Generic vertebrate spinal motor neuron identity is established by three TFs - proneural basic helix-loop-helix (bHLH) TF Neurogenin 2 (Ngn2), and two LIM homeodomain (HD) TFs Lhx3 and Islet 1

(Isl1) (Briscoe et al., 2000; Jessell, 2000; Novitch et al., 2001; Pfaff et al, 1996; Sharma et al., 1998). The three factors function as programing TFs, sufficient to reprogram embryonic stem cells (ESCs) or neural progenitors into spinal motor neurons (Hester et al., 2011; Lee and Pfaff, 2003; Lee et al., 2012; Mazzoni et al., 2013b). Ngn2 is expressed only transiently in motor neuron progenitors as they initiate their terminal differentiation. Isl1 and Lhx3 are initially co-expressed in nascent postmitotic motor neurons, but Lhx3 is rapidly downregulated in all hypaxial and limb muscle innervating motor neurons (Dasen and Jessell, 2009; Sharma et al., 1998). Mapping TF-binding sites during transcriptional reprogramming of ESCs to motor neurons revealed that Isl1 and Lhx3 co-occupy more than 10,000 genomic sites, many of which engage in direct enhancer-promoter interactions to control the expression of motor neuron identity genes (see Chapters 2 and 3; Mazzoni et al., 2013). However, the precise syntax of cis- regulatory elements underlying cooperative recruitment of Isl1 and Lhx3 to their preferred binding sites remained occluded by the poor spatial resolution of chromatin immunoprecipitation- sequencing (ChIP-seq) assays.

ESC differentiation into spinal hypaxial motor neurons undergoes a progressive transition from Ngn2+ progenitors into Isl1+/Lhx3+ nascent postmitotic motor neurons, and ultimately into Isl1+/Lhx3- maturing motor neurons (Tan et al., 2016). The temporal downregulation of two of the three motor neuron programming TFs (Ngn2 and Lhx3) under this differentiation protocol provides an opportunity to study gene regulatory mechanisms contributing to the establishment and stable maintenance expression of motor neuron effector genes in the context of a dynamic transcriptional environment. Our study demonstrates that the downregulation of Lhx3 results in the release of Isl1 from the Isl1/Lhx3 co-bound sites followed by the decommissioning of nascent motor neuron enhancers. Isl1 is then recruited to a new set of enhancers activated in maturing motor neurons. By using high-resolution mapping of TF-binding sites, we discovered that the dynamic behavior of Isl1 can be in part explained by its recruitment to enhancers through protein-protein interactions with clusters of Lhx3 and Onecut1 TFs in nascent and maturing motor neurons, respectively (Rhee and Pugh, 2011). Gene expression analysis revealed that the majority of motor neuron effector genes expressed in postmitotic neurons are associated with transiently active stage-specific enhancers. Together, these results indicate that outwardly stable expression of effector genes in terminally differentiated cells is controlled by transient enhancers established and activated by stage-specific combinations of TFs.

Results

Dynamic changes in the genomic regulatory landscape in postmitotic motor neurons

The progressive differentiation of pluripotent cells into their terminal identity is controlled by developmentally regulated TFs that dynamically establish, activate, silence, and decommission enhancers regulating lineage-specific gene expression programs (Dixon et al., 2015; Shlyueva et al., 2014; Whyte et al., 2012). Mouse ESCs induced on day 2 of differentiation with patterning factors, retinoic acid (RA) and smoothened agonist (SAG) acquire motor neuron progenitor identity characterized by expression of Olig2 on day 4 of differentiation (Wichterle et al., 2002). Treatment of motor neuron progenitors with ɣ-secretase inhibitor (DAPT) transiently activates expression of Ngn2, followed by cell cycle exit and terminal differentiation of progenitors into postmitotic motor neurons by day 5 (Figure 4.1A) (Tan et al., 2016). Under this protocol, there are virtually no dividing progenitors remaining on day 5 of differentiation with ~80% of cells becoming postmitotic motor neurons. Motor neuron maturation under this protocol is temporally well synchronized, as 63% of day 5 cells co-express nascent motor neuron markers Isl1 and Lhx3, but one day later 91% of maturing motor neurons downregulate Lhx3 and acquire an expression profile of hypaxial motor neurons (Figure 4.1B).

Developmentally active enhancers are commonly associated with enriched histone H3 lysine 27 acetylation (H3K27ac) and increased chromatin accessibility (Calo and Wysocka, 2013; Heintzman and Ren, 2009). To examine genome-wide changes in active enhancer signatures in differentiating motor neurons, we performed ChIP-seq of H3K27ac and an assay for transposase-accessible chromatin with sequencing (ATAC-seq) in mouse ESCs (day 0), motor neuron progenitors (day 4), nascent postmitotic motor neurons (day 5), and maturing motor neurons (day 6) (Buenrostro et al., 2013). We identified 30,648 unique distal genomic regions (referred to as 'active enhancers'), having both significant increase of H3K27ac and ATAC-seq intensity, across the four stages of motor neuron differentiation. We observed that a significant fraction of active enhancers is not developmentally stable (Figure 4.1C). Approximately 66% of active enhancers are reorganized during a 24-hour period when cells undergo a transition from dividing motor neuron progenitors to nascent postmitotic motor neurons (10,404 of 15,802 identified

active enhancers). Surprisingly, comparably dynamic enhancer behavior is detected in maturing postmitotic motor neurons. More than 67% of enhancers are temporally reorganized (13,628 of 20,237 active enhancers) during the first 24 hours of motor neuron maturation. The highly dynamic enhancer turnover in postmitotic motor neurons contrasts with the relative stabilization in the total number of developmentally regulated genes (Figure 4.1D, E). RNA-seq analysis revealed that 39% of all expressed genes (4,256/11,049) significantly change their expression between day 4 progenitors and day 5 nascent postmitotic motor neurons (>2 fold), while only 24% of all expressed genes (2,607/11,049) change expression between nascent and maturing motor neurons. These results suggest that the genomic organization of active enhancers remains highly dynamic even in postmitotic neurons exhibiting fewer changes in gene expression.

Divergent binding of Ngn2 and Isl1 to stage-specific enhancers

Considering the dynamic behavior of active enhancers in postmitotic motor neurons, we examined whether programming TFs controlling terminal effector gene expression preferentially bind to the small subset of stably maintained enhancers in postmitotic motor neurons. Previously, it has been shown that during neuronal reprograming of fibroblasts, the proneural bHLH TF Ascl1 establishes accessible chromatin regions to which the programming TF Brn2 is recruited (Wapinski et al., 2013). To examine whether the proneural bHLH TF Ngn2 functions similarly to recruit the programming TF Isl1 in motor neuron differentiation, we performed a series of ChIP-seq experiments to identify Ngn2-bound enhancers (1 kb region around Ngn2-bound site) in day 4 motor neuron progenitors and Isl1-bound enhancers in day 5 nascent motor neurons. Surprisingly, we found that only 4% of bound enhancers (262/5,982) were co-occupied by both TFs (Figure 4.2A), suggesting that Ngn2 expressed in progenitors and Isl1 expressed in postmitotic neurons control gene expression programs through distinct sets of stage-specific enhancers.

Next, we asked whether Ngn2 and Isl1 are preferentially recruited to existing accessible chromatin regions. Mapping of chromatin accessibility by ATAC-seq revealed that neither Ngn2- nor Isl1-bound enhancers were accessible in primitive ectodermal cells on day 2 of differentiation (Figures 4.2B). Ngn2-bound enhancers became accessible in progenitors, coincident with the onset of Ngn2 expression,

but rapidly lost their chromatin accessibility one day later when Ngn2 is downregulated and cells differentiate into postmitotic motor neurons. Similarly, we found that Isl1-bound enhancers were largely inaccessible in day 4 progenitors prior to Isl1 expression, but exhibited increased chromatin accessibility in day 5 nascent postmitotic motor neurons. These findings indicate that Ngn2 and Isl1 contribute to the highly dynamic genomic regulatory landscape by establishing largely non-overlapping sets of stage-specific accessible enhancers during the transition from progenitors to nascent postmitotic motor neurons. Interestingly, these results are in disagreement with the previous model of programming factor action at the Hb9 enhancer during motor neuron specification where Ngn2 was found to bind within the same enhancer as Isl1 and Lhx3 heterodimers. I next investigated what might explain this difference. One possibility is that a small subset of sites might be bound in this highly specialized manner to induce a set of genes, which are crucially important for motor neuron identity.

We next investigated this by analyzing Ngn2 and Isl1 binding around the key motor neuron identity genes Isl1 and Lhx3 (Pfaff et al., 1996; Sharma et al., 1998). Interestingly, distinct distal enhancers in the vicinity of both genes were bound by Ngn2 in differentiating progenitors and Isl1/Lhx3 heterodimers in nascent motor neurons providing further support that the general mechanism of programming factor regulation is through distinct stage-specific enhancers (Figure 4.2C). I next tested the functional role of these independent stage-specific enhancers in regulation of Lhx3 gene expression. Are these enhancers functionally independent from Isl/Lhx3 bound enhancer? Are they required for the Lhx3 expression? To investigate these questions I performed CRISPR/Cas9 genome-engineering experiments to selectively delete the two Ngn2 bound enhancers in the proximity of Lhx3. Enhancer deletion resulted in marked decrease in the expression of Lhx3, but only in a transient manner (Figure 4.2D). On day 5 of differentiation as Ngn2 drives cell cycle exit and specification of nascent motor neuron identity I observed a seventy percent decrease in Lhx3 expression. However, as generic motor neuron identity is consolidated the cells recovered expression of Lhx3 to wild type levels. Together these experiments support a model for dynamic utilization of stage specific enhancers bound by distinct programming factor complexes including Ngn2 in progenitors, which are exiting the cell cycle to become postmitotic motor neurons and Isl/Lhx3 in nascent motor neurons.

Isl1 rapidly relocates to Onecut1-bound sites following Lhx3 downregulation

During motor neuron specification, Isl1 and Lhx3 form a heterodimer complex that binds distal enhancers of many effector genes expressed in nascent postmitotic motor neurons. As I showed in Chapter 2, these enhancers directly engage the promoters of motor neuron identity genes to regulate their expression (see Chapter 3; Lee et al., 2012; Mazzoni et al., 2013; Thaler et al., 2002). However, the fate of the Isl1/Lhx3-bound enhancers following Lhx3 downregulation in maturing postmitotic motor neurons has not been previously examined. To investigate this we performed high resolution ChIP-exo mapping of Isl1 binding before and after Lhx3 downregulation in nascent and maturing HMC motor neurons. We identified 4,128 genomic sites (22-28 bp protected regions, referred to as Isl1- bound sites) in nascent (Lhx3-positive) motor neurons and 3,153 Isl1-bound sites in maturing (Lhx3-negative) motor neurons. Remarkably, only 6% of Isl1-bound sites in nascent neurons (250/4,128) were occupied by Isl1 in maturing motor neurons. Analysis of Isl1-bound enhancers (1 kb region around Isl1-bound site) revealed that 80% of enhancers bound by Isl1 in nascent motor neurons (2,446/3,060) exhibited a significant depletion of Isl1 binding in maturing motor neurons and vice versa, 74% of mature Isl1-bound enhancers lacked Isl1 binding in nascent motor neurons (1,704/2,318) (Figure 4.3A). Most Isl1-bound sites were far from annotated transcription start sites (TSSs) and the median distance between closest nascent and maturing motor neuron Isl1-bound sites was 109 kb, demonstrating that Isl1 does not simply move to a new site within the same genomic territory.

Considering that Isl1 tends to form heterodimers with other TFs, we examined DNA sequences at the new Isl1-bound sites in maturing motor neurons using motif enrichment analysis (MEME) (Bailey et al., 2015). We identified an unexpected primary motif, AATCAATA, annotated as the Onecut/HNF6 DNA-binding motif (Figure 4.3A). Onecut1 and Onecut2 members of the Cut- HD TF family are induced 6 to 12 fold in postmitotic motor neurons relative to motor neuron progenitors. Loss of function analysis established that Onecut TFs are necessary for motor neuron maturation, specification of motor neuron subtype identity, and formation of functional neuromuscular synapses (Audouard et al., 2012; Francius and Clotman, 2010; Roy et al., 2012). To determine whether Onecut TFs co-occupy genomic regions bound by Isl1 in maturing motor neurons, we performed ChIP-exo analysis for Onecut1. We observed that more than 60% of all Isl1-bound sites (1,918/3,153) in maturing motor neurons were located within 50

bp of Onecut1- bound sites (Figure 4.3B). Interestingly, ChIP-exo and ATAC-seq analyses revealed that many Onecut/Isl1-bound enhancers are already accessible and bound by Onecut1 in nascent motor neurons, prior to Isl1 recruitment indicating that Isl1 presence is not a prerequisite for the binding of Onecut1 to maturing motor neuron enhancers (Figure 4.3B, C).

As Isl1 does not appear to be necessary for the establishment of Onecut1-bound sites, we wondered whether it might be involved in functional regulation of these enhancers. ChIP-seq profiling of histone modifications revealed a correlation between Isl1 binding and accumulation of H3K27ac (Figure 4.3A-D). In nascent motor neurons, Isl1-bound enhancers exhibited high levels of H3K27ac, but histone acetylation was significantly decreased in maturing motor neurons, suggesting that Isl1-bound enhancers become rapidly deactivated following the displacement of Isl1. In contrast, Onecut1-bound enhancers in nascent neurons exhibited low levels of H3K27ac, but their acetylation levels significantly increased following Isl1 recruitment in maturing motor neurons. The close correlation between Isl1 binding and H3K27ac enrichment raises the possibility that Isl1 is instrumental in the productive recruitment of a histone acetyltransferase (HAT) and the activation of motor neuron enhancers.

Stage-specific enhancers maintain expression of postmitotic motor neuron genes

Neuronal progenitors undergo a major transition from highly proliferative cells with a simple bipolar morphology to postmitotic motor neurons extending synaptic processes, acquiring mature electrophysiological properties, and synthesizing appropriate neurotransmitters. This transition is accompanied by the activation of a battery of postmitotic effector genes that are expressed throughout the lifespan of motor neurons. Our observation that motor neuron programming transcription factors bind to transient stage-specific enhancers raises the important question how expression of terminal effector genes is maintained in postmitotic neurons.

To correlate enhancers with changes in gene expression, we assigned distal enhancers to the nearest TSS (Creyghton et al., 2010; Rada-Iglesias et al., 2011; Wang et al., 2015; Whyte et al., 2013). We defined three classes of Isl1-bound enhancers (Figure 4.4A,B): (I) enhancers bound by Isl1 in both nascent and maturing motor neurons (referred to as ‘stable’ Isl1 enhancers), (II) enhancers differentially bound by Isl1 in nascent or maturing motor neurons, but co-associated with the same gene (referred to as

‘transient’ Isl1 enhancers), and (III) enhancers associated with genes only in nascent or maturing motor neurons (day 5 or day 6 only). Approximately 62% of all Isl1-bound enhancers (2,932/4,764) belonged to the class I or II. RNA-seq analysis revealed that day 5 and day 6 Isl1 co-associated genes (class I or II) were significantly induced in postmitotic motor neurons and many of the genes were previously established as postmitotic motor neuron effector genes (e.g. Hb9, Isl1, Slit2, Lhx3, Chat, and Nrp1). Moreover, these genes were highly expressed in both nascent and maturing hypaxial motor neurons and over 85% (754/885) are associated with distinct transient Isl1 enhancers (Figure 3C). If we integrate our previous results with respect to dynamic association of Ngn2 bound enhancers with motor neuron effector genes, these results uncover a general model of motor neuron gene regulation. Effector genes playing an important role in motor neuron identity are associated with transient enhancers occupied by stage-specific combinations of programming TFs from late progenitor to postmitotic stages.

To determine whether this is a more general principle, we extended the analysis to active enhancers that are not bound by Isl1. We identified over 13,400 active enhancers in nascent and maturing motor neurons, which lacked Isl1 binding (Figure 4.4D). Both groups of putative target genes were highly induced in postmitotic motor neurons and included many important motor neuron effector genes (e.g. Robo2, Dcc, Nrcam, etc.). Altogether, these results indicate that transiently active enhancers are broadly utilized regulatory features contributing to stable expression of developmentally regulated effector genes in postmitotic motor neurons.

Transient Isl1-bound enhancers control stable motor neuron gene expression

We next sought to test the ability of transient Isl1 enhancers to control gene expression in the context of motor neuron maturation *in vivo*. We utilized the same enhancer electroporation paradigm that I described in Chapter3 to study spatiotemporal enhancer activity in developing chick embryos. We studied stage-specific enhancers associated with either nascent or maturing motor neurons. A nascent motor neuron-specific enhancer (274 bp) at a neuropilin receptor gene Nrp2 (43 kb from the TSS) and a maturing motor neuron-specific enhancer (143 bp) proximal to a glutamate receptor associated gene Gripap1 (32 kb from the TSS) were cloned upstream of a minimal promoter driving expression of a destabilized GFP reporter (Figures 4.5A,B). Electroporation of the construct carrying the nascent Isl1-

specific enhancer resulted in strong GFP expression in nascent spinal motor neurons 24 hours post electroporation, followed by rapid loss of GFP expression by 48 hours (Figure 4.5C, top). In contrast, the maturing Isl1-specific enhancer failed to activate reporter expression in nascent motor neurons at 24 hours post electroporation (Figure 4.5C, middle). However, by 48 hours we observed robust induction of the reporter in postmitotic neurons within the ventral spinal cord, recapitulating the *in vitro*-inferred temporal pattern of gene regulation. Similar results were obtained with three out of four tested nascent Isl1-bound enhancers and two out of three tested maturing Isl1-bound enhancers, indicating that transient Isl1 bound enhancers regulate gene expression with high temporal resolution. Since we found that many stably expressed genes associate with independent stage-specific enhancers, we also tested regulatory activity of an artificial enhancer that combined one nascent and one maturing Isl1-bound enhancer (Figure 4.5B). Upon electroporation of this construct, we observed robust GFP expression that persisted from nascent to maturing motor neurons (Figure 4.5C, bottom), demonstrating that a combination of stage-specific enhancers is sufficient to maintain stable reporter gene expression in the chick spinal cord *in vivo*. Interestingly, this artificial enhancer was composed of stage specific enhancers from two different genes suggesting the primary sequence and spatial activity of the enhancers is independent of genomic location supporting the model of co-regulation of motor neuron identity presented in Chapter3.

Having established the ectopic activity of stage-specific enhancers in maintaining expression of a reporter gene, we next sought investigate their role in controlling expression of the endogenous gene. To test the requirement for stage-specific enhancers during motor neuron differentiation we generated a series of ESC lines carrying homozygous deletions of nascent and/or maturing motor neuron enhancers. We used CRISPR/Cas9-mediated genome-engineering to excise ~1 kb regions harboring nascent and maturing Isl1 bound enhancers associated with motor neuron genes Hb9 (enhancers E1 and E5, respectively) and Isl1 (enhancers E4 and E3, respectively) (Figure 4.6A,B) (Ran et al., 2013). After differentiation of control and mutant ESC lines to nascent and maturing motor neurons, we examined transcription of Isl1 and Hb9 genes using quantitative reverse transcription PCR (qPCR). The deletion of transient enhancers resulted in a stage-specific decrease in Hb9 and Isl1 expression. Furthermore, double-deletion of nascent and maturing motor neuron-specific enhancers resulted in a persistent decrease in Hb9 expression (Figure 4.6A) indicating that individual nascent and maturing motor neuron

enhancers are necessary for maintained expression of associated genes. Importantly, deletion of nascent Isl1-bound enhancers did not compromise Isl1 recruitment to the maturing enhancers, demonstrating functional independence of stage-specific enhancers (Figure 4.6C). Taken together, these results demonstrate that like Ngn2 and Isl1 bound enhancers, transient stage-specific Isl1-bound enhancers in postmitotic motor neurons operate independently and that sequential activation of these enhancers in maturing neurons is sufficient and, in many instances, necessary for the sustained expression of terminal effector genes during motor neuron maturation.

Isl1 is recruited to stage-specific enhancers through interactions with Onecut1 or Lhx3

The highly dynamic changes in Isl1 binding patterns prompted us to examine mechanisms underlying the rapid relocation of Isl1 to Onecut1-bound sites upon the downregulation of Lhx3. We hypothesized that in maturing motor neurons, Isl1 might form heterodimers with Onecut1 and bind to sites containing adjacent Isl1- and Onecut1-binding motifs. To test this, we mapped Isl1- and Onecut1-binding sites with ChIP-exo at near single-base resolution (Rhee and Pugh, 2011). Indeed, at 33% of Isl1/Onecut1 co-bound regions (629/1,918) in maturing motor neurons, the two TFs occupied adjacent sites, containing appropriate Isl and Onecut binding motifs, respectively (subset ii in Figures 4.7B,C). Each of these sites also exhibited a unique 23 bp (Isl1) and 24 bp (Onecut1) ChIP-exo footprint, centered over the expected motifs. Interestingly, at two thirds of co-bound regions (1,289/1,918), Isl1 and Onecut1 bound the same genomic site (subset iii). Closer examination of the ChIP-exo footprints revealed that both Isl1 and Onecut1 binding profiles at these sites were similar and matched the footprint and motif found at sites bound by Onecut1 alone (subset iv). These observations indicate that Isl1 might be recruited to a subset of maturing motor neuron enhancers through protein-protein interactions with Onecut1. Indeed, biochemical analysis revealed that Onecut1 co-immunoprecipitates from lysates of maturing motor neurons with Isl1-containing protein complexes (Figure 4.7D).

Isl1 has been previously shown to form a complex with Lhx3 and a LIM Domain-Binding protein Ldb1, raising the possibility that even in nascent motor neurons, Isl1 might be recruited to a subset of enhancers through protein- protein interactions with Lhx3 (Lee et al., 2013; Mazzoni et al., 2013; Thaler et al., 2002). Examination of ChIP-exo profiles of Isl1 and Lhx3 in nascent motor neurons recovered

adjacent 23 bp (Isl1) and 28 bp (Lhx3) footprints at 31% of co-bound regions (subset ii in Figure 4.8A,B). Interestingly, in the remaining co-bound regions, both TFs co- occupied the same 28 bp footprint centered over a 15 bp HD motif (subset iii), which is identical to the footprint and motif recovered from sites bound by Lhx3 alone (subset iv). Our data indicate that protein-protein interactions of Isl1 with stage-specific TFs, Lhx3 and Onecut1, play an important role in the recruitment of Isl1 to transient motor neuron enhancers.

Isl1 and p300 are preferentially recruited to clusters of Onecut1 or Lhx3

Global analysis of TF-bound sites in nascent and maturing motor neurons revealed a discrepancy between the large number of Lhx3- or Onecut1-bound sites and the relatively small number of sites bound by Isl1 (Figure 4.7B and Figure 4.8A). Throughout the mouse genome, there are over 142,000 occurrences of the 8 bp Onecut consensus motif (AATCAATA). Remarkably, ChIP-exo analysis revealed that more than 75% of these sites (>108,800 sites) were occupied by Onecut1 in maturing motor neurons (Figure 4.9A). In contrast, Isl1 was present at only 5% of Onecut1- occupied sites, prompting us to investigate what additional features contribute to the productive recruitment of Isl1 to a small subset of Onecut1 sites. Notably, we found that Isl1 is preferentially recruited to closely spaced clusters of Onecut1 TFs (defined as Onecut1-binding sites within a 200-bp window). The majority of genomic regions (228/413) containing four or more Onecut1- bound sites effectively recruited Isl1 in maturing motor neurons, whereas only 1% of regions containing single Onecut1-bound site (657/61,248) recruited Isl1 (Figure 4.9B). Similarly, Isl1 was preferentially recruited to regions containing multimeric Lhx3 binding sites in nascent motor neurons.

Next, we examined whether recruitment of Isl1 to a subset of enhancers correlates with enhancer activity. Above we described a strong correlation between Isl1 recruitment and accumulation of H3K27ac at transiently bound nascent motor neuron enhancers (Figure 4.3A-C). Analysis of the histone acetyl transferase p300 and H3K27ac enrichment revealed that p300 and H3K27ac intensities correlated most closely with Isl1 binding occupancy, irrespective of whether these regions contained Lhx3 or Onecut1 binding (Figure 4.9C,D). In contrast, genomic regions that were occupied by Lhx3 alone or by Onecut1 alone exhibited little or no p300 or H3K27ac enrichment. Together these results indicate that Isl1 is

preferentially recruited to enhancers containing clusters of Lhx3 or Onecut1 and that histone acetyl transferase p300 is selectively recruited to enhancers containing Isl1 TF.

Discussion

Outwardly stable gene expression in a dynamic regulatory environment

Embryonic development can be viewed as a multi-step process during which individual cells acquire unique and cell type-specific gene expression profiles. Programming TFs play a pivotal role in this process by regulating the transition from a multipotent progenitor state to a terminally differentiated cell. The ability of programming TFs to induce expression of cell type-specific effector genes and to establish a defined cell identity has been successfully exploited in cell reprogramming experiments (Mazzoni et al., 2013; Son et al., 2011; Vierbuchen et al., 2010; Wapinski et al., 2013). Recent convergence of biochemical, genomic, and computational approaches yielded maps of genomic regions bound by programming TFs (Lodato et al., 2013; Mazzoni et al., 2013; Stampfel et al., 2015; Wapinski et al., 2013). However, these maps remain relatively coarse to reveal the fine-grain organization of cis-regulatory elements controlling the combinatorial transcriptional inputs into cell type-specific gene expression programs. Moreover, many mammalian programming TFs are expressed only transiently at the time when cell identity is being specified, leaving open the question of what ensures continuous expression of cell type-specific effector genes in the context of such a dynamic gene regulatory environment (Briscoe et al., 2000; Jessell, 2000b; Novitch et al., 2001; Sharma et al., 1998).

In theory, enhancers established by transiently expressed programming TFs might be maintained in postmitotic neurons by secondary TFs recruited to these accessible chromatin regions. Our high-resolution mapping analysis of TF-binding sites, chromatin accessibility, and histone modifications in ESC-derived hypaxial motor neurons supports an alternative model. We report that many enhancers established by Isl1 and Lhx3 TFs in nascent motor neurons are truly ephemeral, in existence for less than 24 hours, losing not only TF binding, but also histone acetylation and chromatin accessibility. At the same time, Isl1 is recruited to a distinct set of enhancers bound by clustered Onecut1 TFs that become rapidly

activated. We propose that the maintenance of motor neuron gene expression relies on a dynamic handover (relay) of regulatory control from one stage-specific enhancer to another.

Indirect recruitment of Isl1 to enhancers through protein-protein interactions

The rapidity with which Isl1 is displaced from its initial binding sites following Lhx3 downregulation and recruited to the new sites occupied by Onecut1 is notable. We speculated that this instability might be due to a strictly cooperative binding of Isl1/Lhx3 and Isl1/Onecut1 heterodimers to genomic sites containing pairs of heterodimer-specific motifs (LIM-LIM; LIM-Onecut) (Lee et al., 2012; Mazzoni et al., 2013b). Indeed, we identified such binding pattern in one third of Isl1/Lhx3 and Isl1/Onecut1 co-bound regions. Surprisingly, we discovered a significant fraction of sites at which Isl1 does not bind to its cognate motif, instead Isl1 appears to be recruited to these sites through protein-protein interactions with Lhx3 or Onecut1. We propose that the observed tendency of Isl1 to engage in protein-protein interactions contributes to the rapid displacement of Isl1 from nascent motor neuron enhancers following Lhx3 downregulation and to its relocation to a new set of Onecut1-bound enhancers in maturing motor neurons.

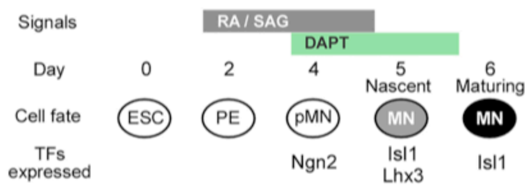
While surprising, indirect recruitment of proteins traditionally recognized as DNA-binding TFs is not without a precedent. An analogous mechanism has been reported for the recruitment of MEC-3, a LIM HD TF controlling postmitotic mechanosensory neuronal identity in C. elegans (German et al., 1992; Gordon and Hobert, 2015; Lichtsteiner and Tjian, 1995; Xue et al., 1993). Biochemical and genetic studies demonstrated that the HD DNA-binding domain of MEC-3 is dispensable for its normal function, and that MEC-3 is recruited to relevant enhancers primarily through a protein-protein interaction with a POU TF, UNC-86. Besides enhancers to which Isl1 is recruited indirectly, we identified a number of Isl1-bound sites with the preferred Isl HD-binding motif. However, since the motif is short and degenerate we favor a model in which additional, yet to be identified TFs interact with Isl1 to stabilize its binding to DNA. Ultimately, it will be interesting to systematically probe TFs with short and degenerate binding motifs to determine whether there is a larger group of TFs preferentially relying on protein-protein interactions for their recruitment to DNA.

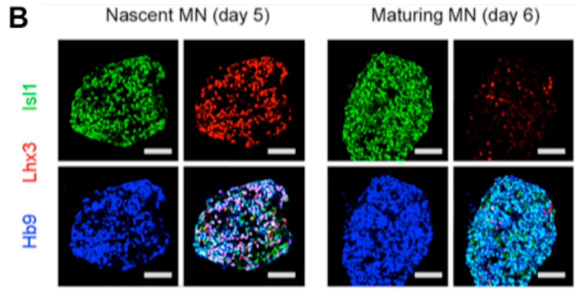
Isl1 operates as an integrator TF, functionally integrating binding patterns of anchor TFs

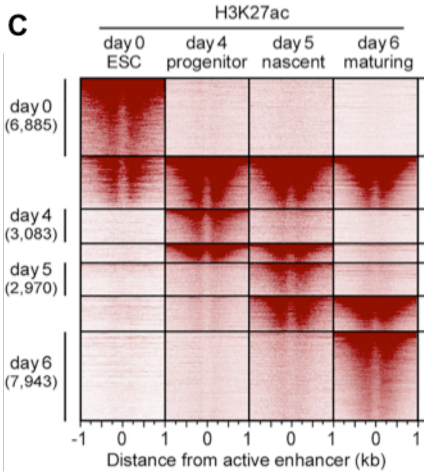
In contrast to Isl1, Onecut1 and Lhx3 TFs appear to bind many genomic sites independent of other TFs (75% of all preferred motifs in the genome are bound by Onecut1), but their recruitment is not sufficient to activate enhancers. We propose referring to such factors as anchor TFs – they strongly bind to DNA, yet possess limited inherent regulatory capacity, and their function therefore depends on combinatorial interaction with additional TFs or co-factors. We reason that Isl1 functions as a transcriptional co-activator for Lhx3 and Onecut1 that might directly or indirectly recruit HATs or other histone modifying enzymes to activate subsets of anchor TF-bound enhancers (Wang et al., 2016). Isl1 is a key TF involved in the specification of many cell types within the brain, pancreas, retina, heart, gut and spinal cord. Previous biochemical studies revealed that Isl1 interacts with several cell-type specific TFs to regulate expression of cell identity genes. Isl1 has been shown to partner with a bHLH TF Beta2 in pancreatic β-cells, a POU- HD TF Brn3b in retinal ganglion cells, and with Phox2a in cranial motor neurons (Zhang et al., 2009; Adouard et al., 2012; Mazzoni et al., 2013). Ultimately it will be interesting to determine whether Isl1 functions primarily as a co-activator of enhancers in all of these cell types and whether its partner TFs function primarily as anchor TFs.

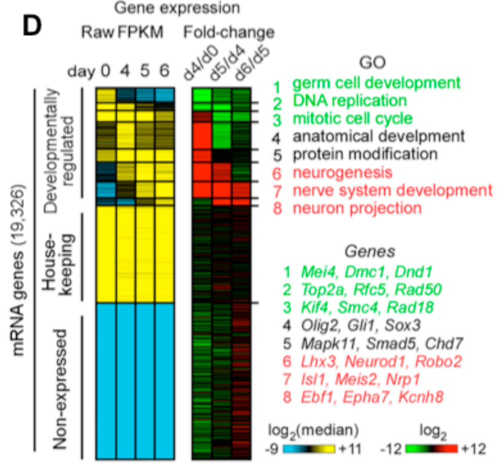
As effective Isl1 recruitment positively correlates with the number of clustered anchor TFs in postmitotic spinal motor neurons, we propose that Isl1 operates as an integrator TF, reading and translating the density of anchor TFs into enhancer activity. Integrator TFs can add a layer of complexity to transcriptional networks and contribute to higher order combinatorial regulation of gene expression. Thus, the observed "division of labor" between anchor TFs strongly interacting with their cognate DNA motifs and integrator TFs that functionally stratify anchor TF-bound regions might constitute an important logical component of transcriptional regulatory networks, contributing to specification of cell diversity in highly complex organs and organisms.

Our findings suggest that the process of neuronal maturation is controlled by cell type-specific dynamic gene regulatory networks, engaging stage specific enhancers associated with terminal effector genes. Mapping and understanding these regulatory networks might yield new strategies for the direct reprogramming of stem cells into mature or aging neurons, a critical step towards the development of more realistic models of late-onset neurodegenerative diseases.


A
Signals
RA / SAG
DAPT
Day
0
2
4
5
Nascent
6
Maturing
Cell fate
ESC
PE
pMN
MN
MN
TFs
expressed
Ngn2
Isl1
Lhx3
Isl1


B
Nascent MN (day 5)
Maturing MN (day 6)
Isl1
Lhx3
Hb9


C
H3K27ac
day 0
ESC
day 4
progenitor
day 5
nascent
day 6
maturing
day 0
(6,885)
day 4
(3,083)
day 5
(2,970)
day 6
(7,943)
a
b
c
d
e
f
g
-1 0 1 0 1 0 1 0 1
Distance from active enhancer (kb)


D
Gene expression
Raw FPKM
Fold-change
day 0 4 5 6
d4/d0
d5/d4
d6/d5
mRNA genes (19,326)
Developmentally regulated
House-keeping
Non-expressed
GO
1 germ cell development
2 DNA replication
3 mitotic cell cycle
4 anatomical develpment
5 protein modification
6 neurogenesis
7 nerve system development
8 neuron projection
Genes
1 Mei4, Dmc1, Dnd1
2 Top2a, Rfc5, Rad50
3 Kif4, Smc4, Rad18
4 Olig2, Gli1, Sox3
5 Mapk11, Smad5, Chd7
6 Lhx3, Neurod1, Robo2
7 Isl1, Meis2, Nrp1
8 Ebf1, Epha7, Kcnh8
log2(median)
-9 +11
log2
-12 +12

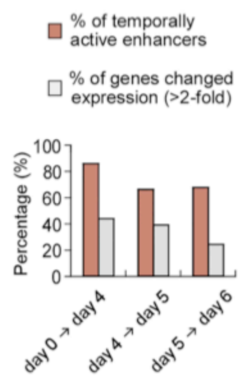

E
% of temporally active enhancers
% of genes changed expression (>2-fold)
Percentage (%)
100
80
60
40
20
0
day 0 → day 4
day 4 → day 5
day 5 → day 6

Figure 4.1. Dynamic Changes in Enhancer Organization and Gene Expression during Motor Neuron Differentiation. (A) Overview of mouse ESC-directed differentiation to postmitotic hypaxial motor neuron (MN). Differentiating cells become nascent postmitotic MNs on day 5, and maturing MNs on day 6. PE, primitive ectoderm; pMN, MN progenitor. **(B)** Expression of Isl1, Lhx3, and Hb9 in embryoid bodies. Lhx3 is transiently expressed in nascent MNs on day 5 and rapidly downregulated in maturing MNs on day 6. Scale bars represent 50 μm. **(C)** H3K27ac intensity relative to active enhancer midpoints during MN differentiation. Active enhancers were grouped and sorted by stage-specific H3K27ac intensity, demonstrating that majority of enhancers are transient during MN differentiation. **(D)** Expression levels of 19,326 annotated RefSeq genes grouped by fold-changes in expression between individual MN differentiation stages. Left, FPKM (fragments per kilobase of exon per million fragments mapped) values were median normalized, and log2 transformed. Middle: Log2 RNA fold-change, ordered as shown in the left panel. Right, representative GO terms and examples in each group of genes. **(E)** Percentage of enhancers and genes that dynamically change more than 2-fold between indicated timepoints. Enhancers remain dynamic between day 5 and day 6 MNs, while gene expression becomes more stable.

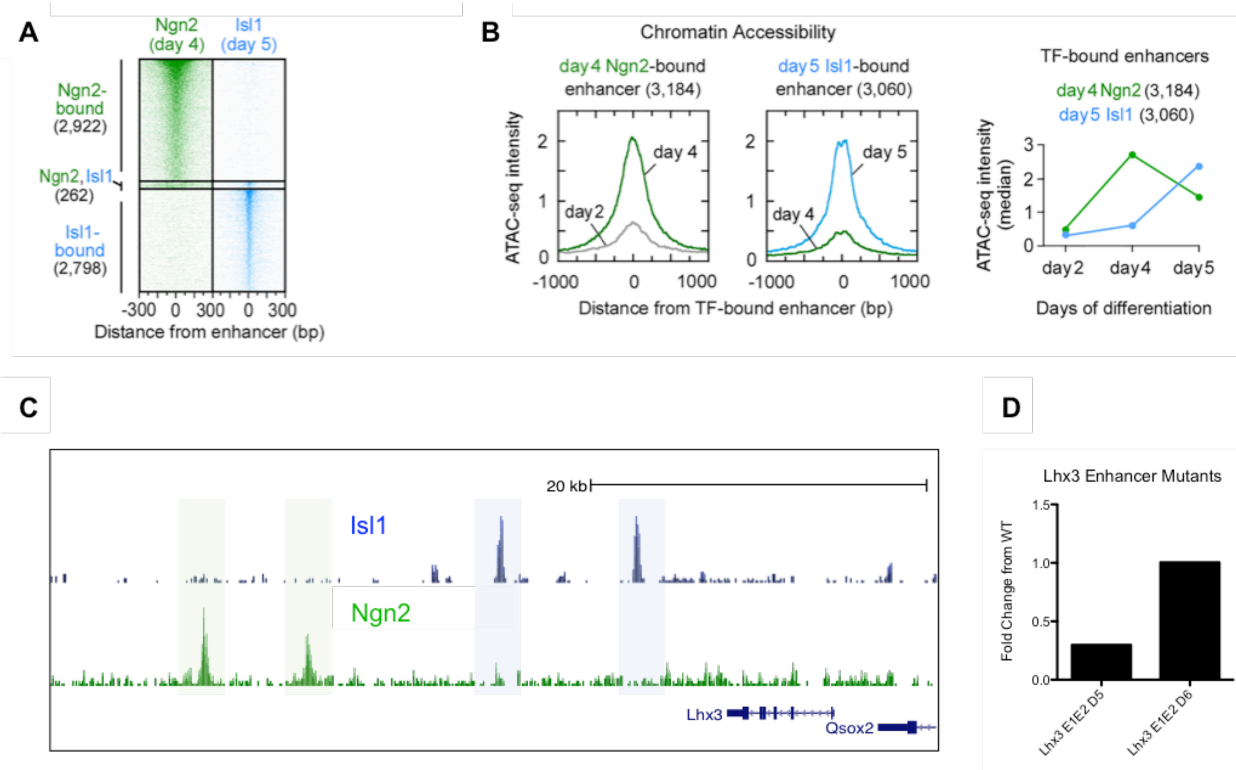


Figure 4.2. Ephemeral Nature of Enhancers Established by Motor Neuron Programming TFs. (A) ChIP mapping of Ngn2 binding enhancers in day 4 progenitors and Isl1-bound enhancers in day 5 postmitotic MNs, sorted by TF occupancy. Ngn2 and Isl1 TFs were found within 1 kb genomic region only in 262 locations. **(B)** Composite (average read counts) plots of days 2 and 4 ATAC-seq intensity for Ngn2-bound enhancers (left), and days 4 and 5 ATAC-seq intensity for Isl1-bound enhancers (middle). Right panel depicts a transient increase in median ATAC-seq intensity of Ngn2-bound enhancers on day 4 and a gain of chromatin accessibility of Isl1-bound enhancers on day 5, coincident with the time of TF expression. **(C)** Genome browser view of transient transcription factor bound enhancers in the proximity of the Lhx3 locus. Two distinct Ngn2 (green) and Isl1 (blue) bound enhancers are bound in a stage-specific manner. **(D).** Quantitative PCR to measure expression of Lhx3 after deletion of both Ngn2 bound enhancers during specification and maturation stages of motor neuron differentiation. Ngn2 bound enhancer deletion results in delayed induction of Lhx3 expression. Expression levels are rescued to WT levels upon activation of Isl bound enhancers.

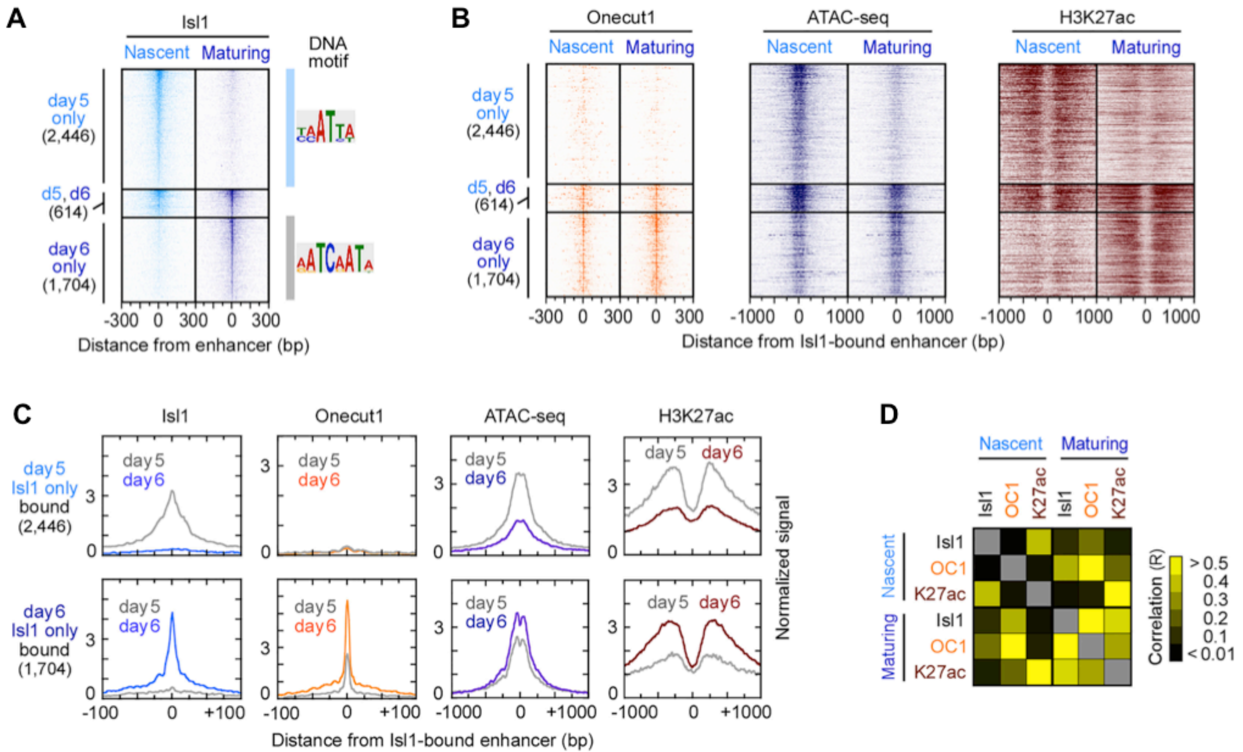


Figure 4.3. Ephemeral Nature of Isl1 bound Enhancers in Postmitotic Neurons. (A, B) Isl1, Onecut1, and H3K27ac ChIP, and ATAC-seq intensity measured in nascent (day 5) and maturing (day 6) postmitotic MNs. All data are plotted relative to the Isl1-bound enhancer midpoints, sorted and ordered by Isl1 occupancy. If the midpoints between day 5 and day 6 Isl1-bound enhancers reside within 1 kb, they were defined as maintained enhancers (n=614). DNA motif represents the most enriched sequence within ±14 bp from the midpoint of Isl1-bound site. **(C)** Composite plots of panels A and B. **(D)** A heat map of Pearson correlation coefficients (R) for pairwise combinations of TFs and H3K27ac intensity, shown in panels C and D.

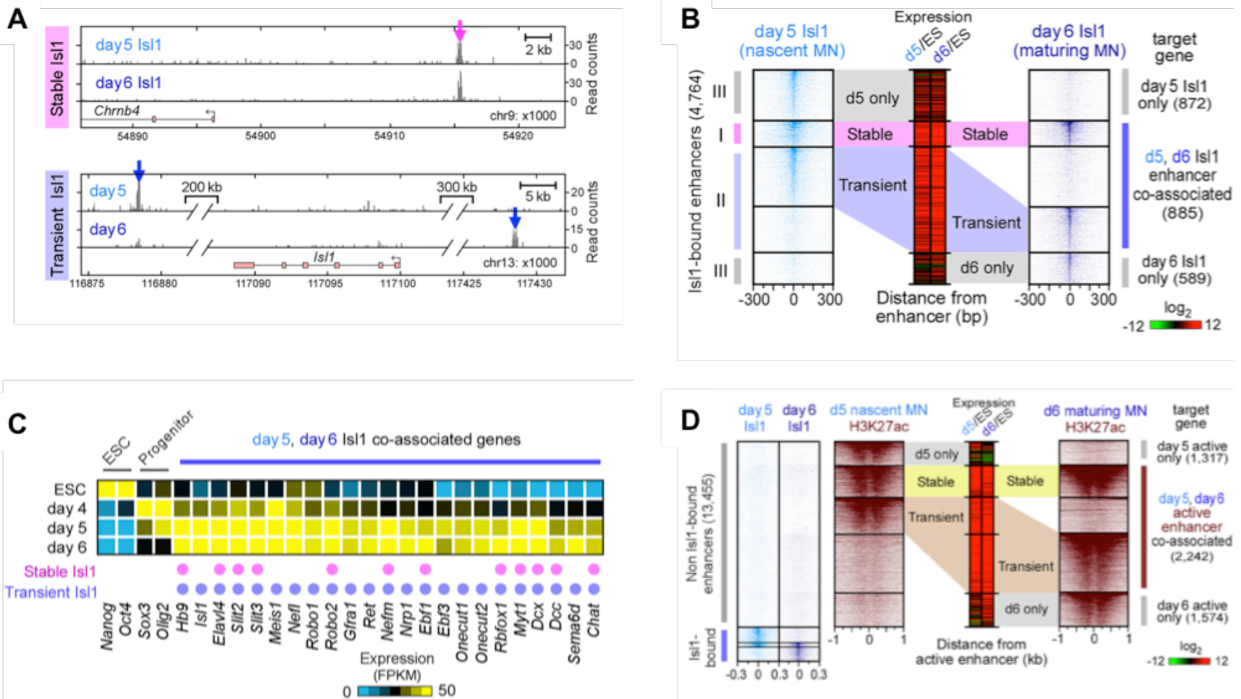


Figure 4.4. Motor Neuron Effector Genes are Associated with Stage-Specific Isl1-Bound Enhancers. (A) Examples of a stable Isl1-bound enhancer (class I, magenta) proximal to Chrnb4 and transient enhancers (class II, blue) proximal to Isl1 gene in nascent (day 5) and maturing (day 6) MNs. ChIP mapping of Isl1 is shown. **(B)** Isl1 occupancy relative to Isl1-bound enhancers, sorted and ordered by Isl1 occupancy. These enhancers were grouped by the presence of stable (I), transient (II), day 5 only (III), and day 6 only (III) Isl1-bound enhancers. Middle panel shows relative changes in associated gene expression in day 5 or day 6 MNs relative to ESCs (Log2 FPKM). **(C)** Expression profiles of selected genes developmentally regulated in ESCs, progenitors, and postmitotic MNs. Association of a gene with a stable and/or transient enhancers is marked by a pink and purple dot, respectively. **(D)** Number of genes associated only with stable, transient, or both classes of Isl1-bound enhancers, shown in panel B. **(G)** H3K27ac intensity relative to active the enhancers (n=13,455), which were enriched for both H3K27ac and ATAC-seq intensity in the absence of Isl1 binding (subset of groups d, e, f, and g in Figure 1C), sorted and ordered by H3K27ac intensity. These enhancers were grouped by the presence of stable, transient, day 5 only, day 6 only active enhancers.

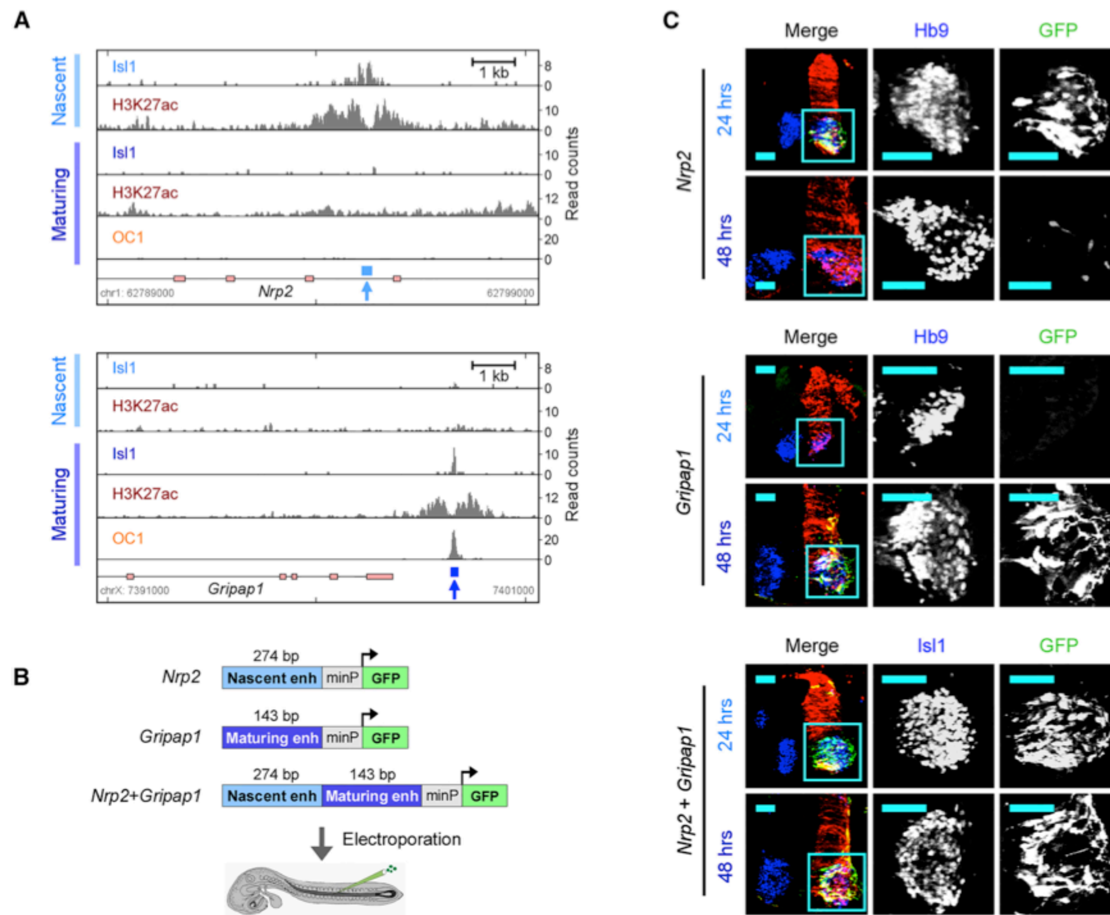


Figure 4.5. Spatiotemporal Specificity of Transient Isl1-Bound Enhancers in the Developing Neural Tube. (A) Examples of a nascent MN specific Isl1-bound enhancer (274 bp, cyan arrow) at Nrp2 and a maturing MN specific Isl1-bound enhancer (143 bp, blue arrow) proximal to Gripap1. ChIP mapping of Isl1, H3K27ac, and Onecut1 is shown. **(B)** Stage-specific enhancers were cloned individually or in combination into a destabilized GFP reporter plasmid containing a minimal promoter (minP). **(C)** Analysis of the reporter gene expression 24 and 48 hours after electroporation of the plasmids into the Hamburger Hamilton stage 13 chick neural tube. A CMV-mCherry reporter (red) was co-electroporated with GFP plasmids to assess the efficiency of electroporation. Embryos were fixed at 24 hours (stage 18) and 48 hours later (stage 23), and stained with MN marker Hb9 (blue) and GFP (green) antibodies (n=2). Scale bars represent 50 μm.

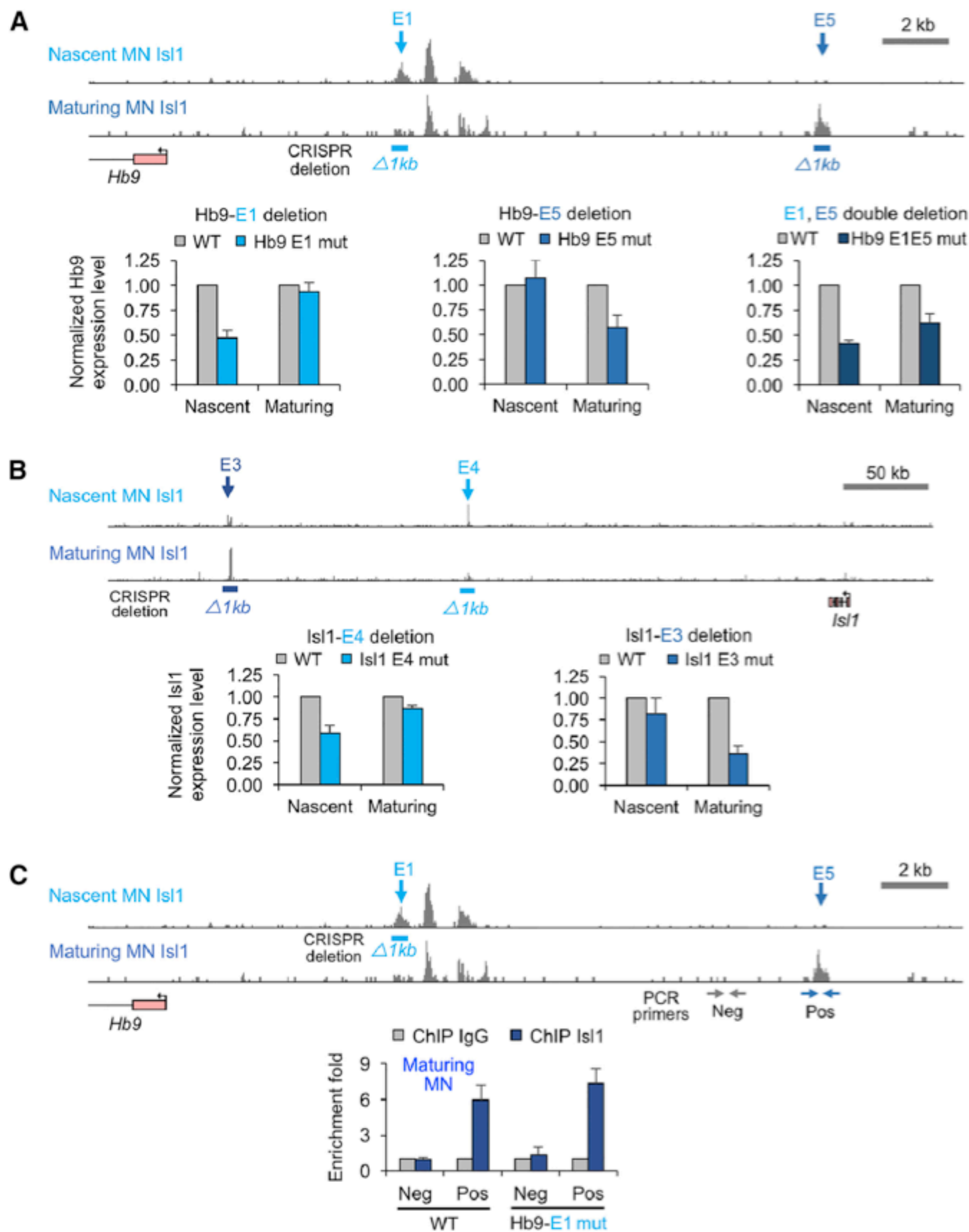


Figure 4.6. Transient Isl1-Bound Enhancers are Necessary for a Stable Expression of Motor Neuron Genes. (A) Upper: The location of Isl1-bound enhancers proximal to Hb9 gene in wild-type (WT) MNs. 629 bp genomic DNA of a nascent MN-specific Isl1-bound enhancer (Hb9-E1) and 847 bp of a maturing MN-specific Isl1-bound enhancer (Hb9-E5) were deleted using CRISPR genome editing (Ran et al., 2013). Lower: Hb9 expression levels measured by quantitative RT-PCR in nascent and maturing MNs (normalized to WT expression levels of Hb9). Error bars represent the s.d. n=4, two independent differentiations. **(B)** Same as panel A except transient enhancers proximal to Isl1 gene (1,413 bp of the maturing MN-specific enhancer (Isl1-E3) and 1,519 bp of nascent MN-specific enhancer (Isl1-E4) were deleted. n=4, two independent differentiations. **(C)** ChIP-PCR analysis of Isl1 binding to maturing MN enhancers in an ESC line containing the deletion of a nascent MN-specific enhancer (Hb9-E1). Horizontal arrows mark the position of positive (blue; Pos) and negative (gray; Neg) PCR primer sets. Error bars represent the s.d., n=3, two independent differentiations.

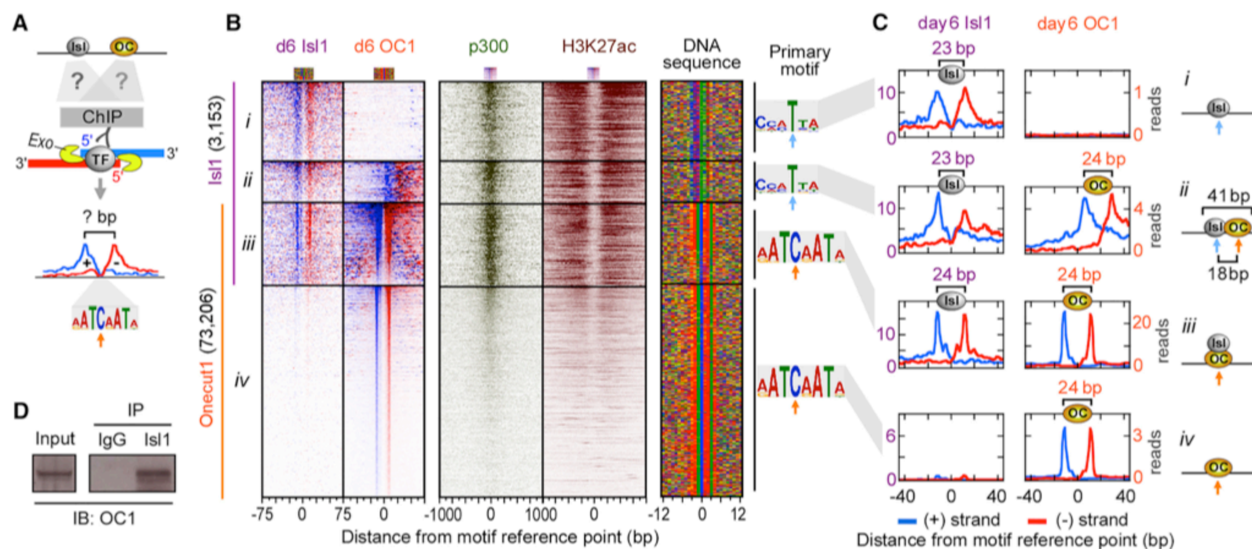


Figure 4.7. High Resolution Mapping of Isl1 and Onecut1 Binding Sites in Maturing Motor Neuron Enhancers. (A) ChIP-exo relies on the treatment of immunoprecipitated DNA with a 5' to 3' exonuclease. The 5' ends of the digested DNA on the (+) and (-) strands are enriched at a fixed distance from the TF-DNA crosslinking sites, demarcating TF footprint regions protected from the exonuclease. **(B)** ChIP-exo for Isl1 and Onecut1 and ChIP-seq for p300 and H3K27ac plotted relative to the TF motif reference point in day 6 maturing MNs, sorted and ordered by Isl1 and Onecut1 occupancy. TF occupancy on the (+) strand (blue, left border) and (-) strand (red, right border) is shown. We classified four subsets: (i) 1,235 Isl1 only bound sites, (ii) 629 Isl1-bound sites next to Onecut1 (plots were reoriented to keep Onecut1 to the right side of Isl1), (iii) 1,289 Isl1/Onecut1 co-bound sites, and (iv) 71,917 Onecut1 only bound sites (10,426 sites on chromosomes 1 and 2 are shown). Right panel shows a color chart representation of the DNA sequence located ±12 bp from the motif reference point (cyan arrow for Isl and orange arrow for Onecut motif), which is the midpoint between the left and right border of TF. **(C)** Composite plots of TF ChIP-exo profiles on the (+) and (-) strand (TF footprints) shown in panel B. Right panel shows models of TF-binding sites. The distance between adjacent Isl1 and Onecut1 motif reference points of heterodimers (subset ii) is 18 bp. **(D)** Biochemical demonstration of Isl1 and Onecut1 interactions in maturing MNs. Immunoprecipitation (IP) of day 6 MN lysates with Isl1 antibody followed by a western blot (IB) with Onecut1 antibody. The detected band is ~51 kD, consistent with Onecut1 protein. Input is 5% of whole cell extract; IgG is a negative control. Shown is a representative blot (n=2).

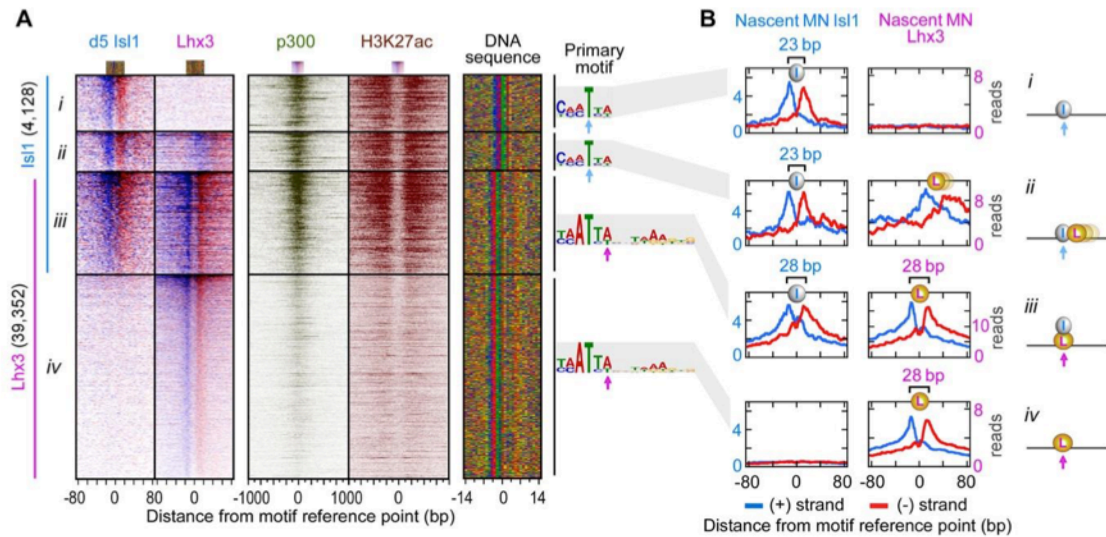


Figure 4.8. Isl1 is Recruited to DNA through Interactions with Lhx3 in Nascent Motor Neurons. (A) Isl1, Lhx3, p300, and H3K27ac intensity relative to the TF motif reference point of Isl1-bound sites in day 5 nascent motor neurons (MNs) or Lhx3-bound sites in iNIL programmed spinal MNs (Mazzoni et al., 2013), sorted and ordered by TF occupancy (Table S6). TF occupancy on the (+) strand (blue, left border) and (-) strand (red, right border) is shown. We classified four subsets: (i) 1,127 Isl1 only bound sites, (ii) 805 Isl1 next to Lhx3-bound sites (Lhx3 was reoriented to the right side of Isl1 regardless the reference Isl motif orientation), (iii) 2,205 Isl1/Lhx3 co-bound sites (iv) 37,147 Lhx3 only bound sites (10,426 sites on chromosomes 1 to 4 are shown). Right panel shows a color chart representation of the DNA sequence located ±14 bp from the motif reference point, ordered as shown in the left panels. Cyan and magenta arrows indicate the motif reference point of Isl and Lhx3 DNA-binding motif, respectively. **(B)** Composite (average read counts) plots of Isl1 and Lhx3 binding occupancy on the (+) and (-) strand shown in panel A. Right panel shows models of Isl1 and Lhx3 binding to DNA in nascent MN enhancers.

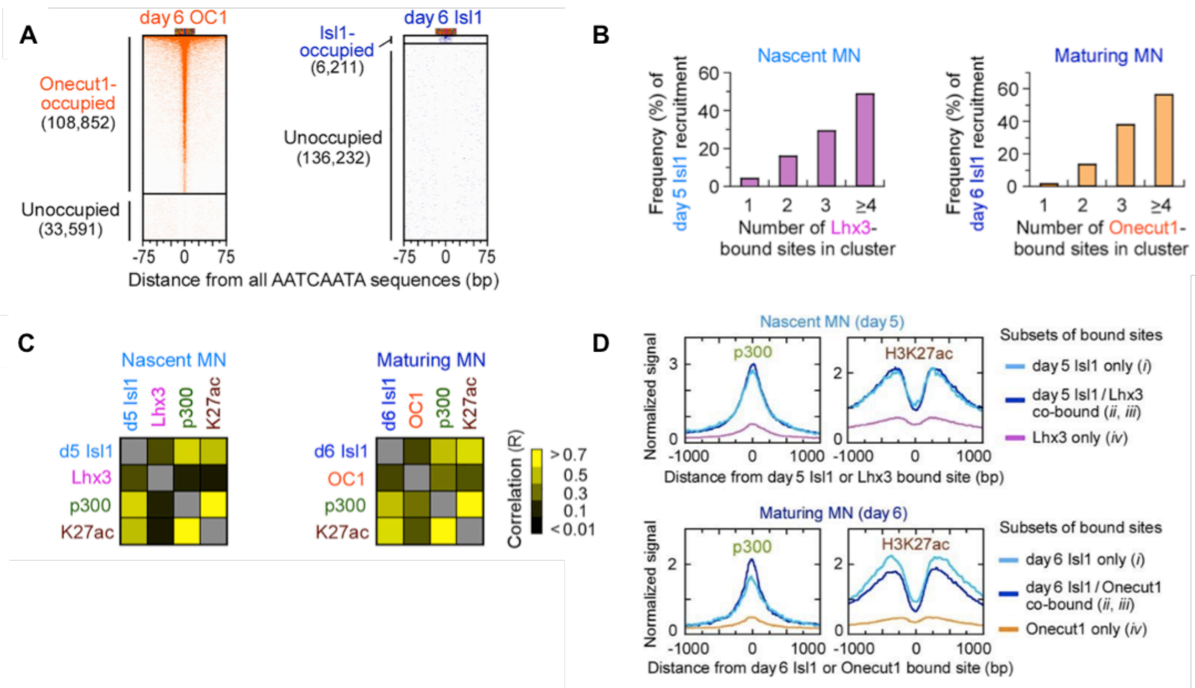


Figure 4.9. Isl1 is Preferentially Recruited to Enhancers Containing Clusters of Onecut1 or Lhx3 TFs. (A) Onecut1 and Isl1 occupancy in 142,443 genomic sites containing canonical Onecut binding motif (AATCAATA), sorted by Onecut1 (left panel) and Isl1 (right panel) occupancy. **(B)** Percentage of Lhx3 (in nascent iNIL MNs) and Onecut1 (in maturing MNs) bound enhancers that successfully recruited Isl1 in nascent and maturing MNs, respectively. The enhancers were subdivided based on the number of clustered Lhx3 or Onecut1 binding sites (number of identified biding sites within a 200 bp window). (C) Heat maps of the Pearson correlation coefficients (R) for pairwise comparisons of TFs and H3K27ac intensity in nascent MNs and in maturing MNs shown in Figure 4.7B. (D) Upper, Composite plots of p300 and H3K27ac intensity for Isl1 only (subset i), Isl1/Lhx3 (subsets ii and iii), and Lhx3 only bound sites (subset iv) in nascent motor neurons (MNs). Lower, Composite plots for Isl1 only (subset i), Isl1/Onecut1 (subsets ii and iii), and Onecut1 only bound sites (subset iv) in maturing MNs shown in Figure 4.7B.

Chapter 5

General Discussion and Future Perspectives

The work presented in this thesis was focused on understanding the molecular mechanisms that control the specification of the immense cellular diversity in the nervous system. Years of study in distinct neural structures including the spinal cord, the retina, and the cerebral cortex uncovered a remarkable amount of information regarding the general principles of neuronal diversification (Fishell & Heintz, 2013; Jessell, 2000a; Livesey & Cepko, 2001; Molyneaux, Arlotta, Menezes, & Macklis, 2007). From these studies, in neurons of distinct identities, arising from different regions along the neuraxis and with unique functional roles, arises a unifying theory which connects all aspects of cell identity and diversity in the nervous system. The combinatorial expression of specific transcription factors imparts unique features of neuronal identity on distinct cell types. However, thus far, very little was known about the molecular mechanisms and regulatory organization facilitating the generation of this immense cellular diversity. In the work presented in this thesis, we have made significant progress towards understanding the regulatory organization and molecular control of the gene expression programs underlying motor neurons identity in space and time. In this final chapter, I will discuss the interesting implications of our discoveries as well as the new questions that they raise and future experiments that will push our understanding of neural diversity onwards.

Cell- and stage-specific combinatorial binding of transcription factors

Our results have begun to elucidate the functional role of combinatorial binding to distal enhancers. We have shown that specific combinations of transcription factors control motor neuron specific as well as pan-neuronal gene expression programs. Interestingly, we have identified how different heterodimer compositions control different genes through context-dependent binding to enhancers with cell-specific co-factors.

We focused our studies to the binding of the motor neuron programming factor Isl1 in three distinct contexts of motor neuron identity (Pfaff et al., 1996). First, we compared Isl1 binding patterns in

spinal and cranial motor neurons programmed from mouse embryonic stem cells. We found that the genomic binding patters of Isl1 were context-dependent with Isl1 occupying distinct genomic locations in cranial and spinal motor neurons that was dependent on the cofactor expressed. In cranial motor neurons Isl1 formed heterodimers with Phox2a and in spinal motor neurons with Lhx3. Interestingly, Isl1 displayed both cell-specific and shared binding locations. Together, these results suggest that combinatorial binding between distinct transcription factors are able to control both specific and shared properties in related, but distinct types of neurons. These initial experiments were performed at a single stage when motor neurons are generically specified. From a single point in time we can't infer the dynamics of combinatorial binding. Are both factors recruited to DNA at the same time or is there a hierarchy in the order of recruitment? What happens when the combinatorial composition of transcription factors changes?

We next performed a similar set of experiments using motor neurons differentiated with hypaxial subtype characteristics. Initially, cells are characterized by a transcription factor code of nascent identity, co-expressing Isl1 and Lhx3. However, as cells mature they rapidly transition from Isl1/Lhx3-positive nascent identity to Isl1/Lhx3-negative mature identity. This allowed us to ask what happens to programming factor binding upon down regulation of its cofactor. Remarkably, we observed rapid relocation of Isl1 to sites bound by another transcription factor, Onecut1. Clusters of Onecut1 bound TFs act as an anchor to guide the relocation of Isl1 to a new set of enhancers. Interestingly, we found that the accumulation of activating histone modifications was correlated with the recruitment of Isl1 in a stage specific manner suggesting that Isl1 acts like a co-activator guided to specific locations in the genome based on its binding partner.

Together, these studies of Isl1 binding in three separate cellular contexts uncovered the remarkable regulatory capacity that the combinatorial binding of transcription factors can have. It will be important to test this idea more broadly. Is the dynamic relocation of Isl1 a mechanism that evolved specifically in HMC cells, which lose expression Lhx3, or is it a general principle of combinatorial action by dynamically expressed transcription factors during development. One obvious next step would be to ask what happens in LMCm motor neurons, which lose expression of Lhx3 as well (Dasen, 2009; Kania et al., 2000). Does Isl1 binding relocate to a new set of sites? What is the cofactor, which guides Isl1 binding? Does Onecut again perform the same role or is there a new factor in each subtype?

Klf factors as general activators of enhancer-promoter engagement

The maps of transcription factor binding generated in these studies present an interesting concept that is perhaps underappreciated. Within transcription factor complexes, individual factors exhibit different activities. Thus, the combinatorial binding of different TFs might impart different functions on enhancer activity. For example, we found that Lhx3 and Onecut transcription factors tend to bind a larger portion of their sites in the genome than Isl1. In the absence of Isl1 binding enhancers displayed lower levels of activating histone modifications and recruitment of the co-activator, p300. Accordingly, we found that Isl1 recruitment was highly correlated with p300 recruitment (Chapter 3 and 4). However, only a subset of Isl1 bound sites become highly activated and engage promoters via-chromatin interactions (Chapter 3). Certainly, we expect that some part of this discrepancy is due to underestimating the total enhancer-promoter interaction landscape due to technical limitations of chromatin interaction experiments. However, despite these caveats, we see only a small portion of total TF binding sites engaged in interactions suggesting that Isl1 recruitment is not sufficient to promote enhancer-promoter interaction. This suggests that there might be additional layers of regulation beyond the simple binding of cell specific heterodimers of with Isl1. What then might explain the differences in binding and enhancer-promoter engagement?

To try to address this question we performed an initial analysis to search for additional transcription factors, which promote effective looping of enhancer-promoter interactions. In order to identify candidate transcription factors that could facilitate promoter engagement we performed a differential motif enrichment analysis between Isl1/Lhx3 bound regions that either recruit high levels of mediator complexes or those that fail to recruit the complex (Figure 5.1A). Interestingly, we discovered strong enrichment for a single motif associated with the Sp/Klf family of transcription factors. We find that Klf motifs are approximately three times more likely to occur in proximity of IL bound enhancers which engage in interactions. Additionally, these motifs are six times more likely to accessible and bound based on a DNase binding algorithm (Figure 5.1B, C, D). Interestingly, Klfs were previously implicated as key players in facilitating enhancer-promoter interactions in mouse embryonic stem cells. Klf4 interacts with the cohesin complex to promote stable engagement of distal enhancers with the promoters of ES identity genes (Wei et al., 2013; Whyte et al., 2013).

If we take these preliminary results in the context of our studies of Isl1 binding with cell specific heterodimers we can add additional layers of regulation synergy between combinatorial transcription factor complexes. In this model, an anchor factor such as Lhx3 or Onecut1 would bind and create maps throughout the genome to demarcate where Isl1 is recruited. Isl1 recruitment would recruit co-activators including p300 to acetylate histones and facilitate activation of the enhancer. And finally, Klf members would bind to effectively recruit structural complexes including mediator and cohesin to promoter enhancer-promoter looping. Many important experiments are underway to test these ideas now. We have generated mouse ES lines with tagged Klf family members to map their binding locations throughout the genome in motor neurons. We are also in the process of testing the activity of enhancers in their endogenous context by mutating Klf binding motifs and assessing mediator recruitment and enhancer-promoter looping. These experiments should further our understanding of the regulatory logic of motor neuron identity by combinatorial action of transcription factors at distal enhancers. Nevertheless, a major question remains: How do enhancers find their specific target promoters?

Enhancer-promoter specificity

In Chapter 3, I uncovered the regulatory principles of enhancer control of motor neuron identity. I found that each motor neuron gene is regulated by multiple distinct enhancers spread across large distances that can span up to megabases of linear genomic sequence. Enhancers often do not interact with their most proximal promoter suggesting there are likely more complex regulatory mechanisms to control enhancer-promoter specificity. Together, these findings bring about interesting questions. How do enhancers find and interact with distantly located target promoters to trigger gene expression and what controls the specificity of these interactions? One mechanism has arisen in recent years to describe how such specificity might be maintained. High-throughput mapping of chromosome confirmation with Hi-C and ChIA-PET have revealed that chromosomes are organized into higher-order structures called topologically associating domains (Dixon et al., 2012, 2015; Dowen et al., 2014; Phillips-Cremins et al., 2013; Tang et al., 2015). These domains are anchored on each end by insulator elements bound by CTCF to form “insulated neighborhoods” within which enhancer-promoter interactions are contained (Dowen et al., 2014; Guo et al., 2015). Indeed, my analysis of enhancer-promoter interactions bound by

Isl/Lhx3 during motor neuron specification support this model with the majority of interactions restricted within the same topological domain as their target gene. While insulation by CTCF anchored regulatory domains might explain how enhancer interactions do not activate promoters of genes in neighboring domains, it does not explain how distal enhancers find their targets promoters within their respective domain. Interestingly, while genes within the same topological domain tend to have correlated expression it is by no means binary. That is, not all genes within a TAD are expressed and in some cases genes have divergent patterns of expression (Dixon et al., 2015; Hnisz et al., 2016). Together, this supports the idea that there are likely more complex mechanisms to control specificity in enhancer-gene interactions.

One possibility could be that there is information in the proximal promoter sequences from which specificity in enhancer-gene interactions could arise. This model would allow for the specific recruitment from distal enhancers to proximal sites. For example, sequence specific DNA binding transcription factors binding at distal enhancers and promoter-proximal sites could, in principle, recruit cofactor complexes including mediator and cohesin to guide specificity within large regulatory domains. Indeed, there have been a handful of studies, which support this idea. In studies performed in *Drosophila* embryos, a distal enhancer activates expression of the Scr gene despite another gene, Ftz, located in the intervening region. This specificity is mediated through a tethering element located upstream of the Scr promoter (Calhoun and Levine, 2003; Calhoun et al., 2002). Additionally, distal enhancers have been shown to exhibit preferences for different types of promoters. For example, promoters from housekeeping genes and those from developmentally regulated genes display different responses when combined with different enhancers. This difference is likely to be under the coordinated control of distinct DNA binding factors (Butler and Kadonaga, 2001; Ohtsuki et al., 1998; Zabidi et al., 2014).

However, it remains to be investigated how far-reaching these mechanisms are and to what extent specificity in enhancer-promoter interactions is controlled by specific factors throughout the genome. Does specificity lie in the enhancer alone or is there coordinated specificity in both and enhancer and promoter sequences? Do the same factors regulate both ends of the interactions or do distinct complexes get recruited to enhancer and promoter? These are all open questions that we are currently investigating.

A dynamic relay of enhancers initiates and maintains neuronal gene expression programs

The idea that enhancers are highly dynamic and both cell and tissue specific during development is a concept that has been appreciated for some time (Heintzman et al., 2009; Nord et al., 2013; Shen et al., 2012; Visel et al., 2009). Moreover, within a single tissue enhancers display a high degree of spatial and temporal specificity during development. However, a deep understanding of their temporal dynamics within a single cell type and their direct role with respect to the regulation of gene expression remains poorly understood.

Here we performed a precise dissection of enhancer activity at distinct stages of motor neuron specification. Interestingly, we find that enhancers are highly dynamic even within a single cell type. The core motor neuron identity genes are controlled by multiple stage-specific enhancers bound by different combinations of stage-specific transcription factor complexes. Stage-specific transcription factors are not simply controlling distinct sets of genes, but inducing and maintaining the global motor neuron gene expression program. This finding revealed a previously underappreciated idea that within postmitotic neurons the regulatory environment is highly dynamic and complex. At each stage of motor neuron specification studied, multiple distinct enhancers regulate gene expression. We tested the temporal and combinatorial function of these enhancers in motor neuron development using CRISPR-mediated genome-engineering. Our results strongly suggest that enhancers act independently of one another to control target gene expression. This idea is true in both static and dynamic contexts, as deletion of enhancers at one stage of development does not affect the activity of downstream enhancers, further supporting an independent model of enhancer regulation.

What might be the regulatory logic for the evolution of such a complex and robust enhancer system? One reasonable hypothesis is that multiple enhancers might provide a regulatory cushion to shield genes from environmental perturbation and maintain robustness in gene expression programs. (Hong et al., 2008; Levine, 2010; Spitz and Furlong, 2012). Any single mutation or perturbation to a gene regulatory network will result in only minor disruptions, occurring in narrow time windows thereby ensuring stability of gene expression programs and thus, cell identity. With this idea in mind, we propose that during cell specification, gene expression programs might function akin to attractor networks, initiating at a relatively chaotic and dynamic phase of neuronal progenitor identity, progressing to a more stable and

genetic postmitotic neuron identity, then through further refinement to specific subtype identity, and finally arriving at a stable mature neuronal identity (Hopfield 1982). The mature stage is likely the final state of specification. It will be important to understand when this final specification occurs by studying multiple stages of neuronal specification and maturation *in vivo*. This model of progressive restriction in the dynamics of gene expression networks is supported by recent experiments investigating the reprogramming competence of postmitotic neurons in the cortex in which postmitotic callosal neurons have restricted competence to be reprogrammed at late stages of development, suggesting that eventually neurons are locked into a specific identity (Rouaux and Arlotta, 2010). Taken together, our results add to the emerging idea that the outwardly stable expression program of postmitotic neuronal identity is dynamically maintained by a complex repertoire of developmental enhancers (Deneris and Hobert, 2014).

Reverse engineering cell identity: implications for development and disease

The findings that the misexpression of transcription factors is sufficient to program specific aspects of cell identity have generated an entirely new field of biomedical science (Davis et al., 1987; Takahashi and Yamanaka, 2006; Vierbuchen et al., 2010). Currently, the identification of effective programming modules relies on empiric testing of many combinations of transcription factors expressed in the target cell type. Based on the work presented here, we propose that the selection of effective programming modules for nerve cells should focus on a key set of transcription factors expressed during similar developmental windows that control global aspects of cellular identity. The combinatorial action of these key transcription factors expressed within small time windows will likely result in robust transformation of gene expression programs during programming.

In the work presented in this thesis, we took advantage of the fact that we can generate a nearly unlimited supply of motor neurons to perform chromatin profiling experiments that are not usually possible *in vivo*. From this in-depth dissection of the regulatory control of motor neuron identity, we have learned the basic principles of enhancer-promoter organization that underlie the regulation of cellular identity in the nervous system. By combining these general rules of neuronal gene regulation with the recent advent of chromatin profiling techniques for low cell numbers, we now propose that it should be possible to probe

the regulatory control of distinct neuronal identities *in vivo* in a global fashion (Buenrostro et al., 2013). By integrating the maps of putative regulatory regions with the computational discovery of regulatory motifs associated with cell identity genes we can learn the identity of programming factors in a cell specific context and more accurately inform the design of cell programming systems in an unbiased manner. This strategy should allow one to effectively reverse engineer the nervous system. Altogether, by understanding the regulatory logic and function of programming transcription factors we will not only further our understanding of the general mechanisms of development, but also substantially accelerate production of homogenous cell populations necessary for human disease modeling, cell-based drug screening and transplantation therapy.

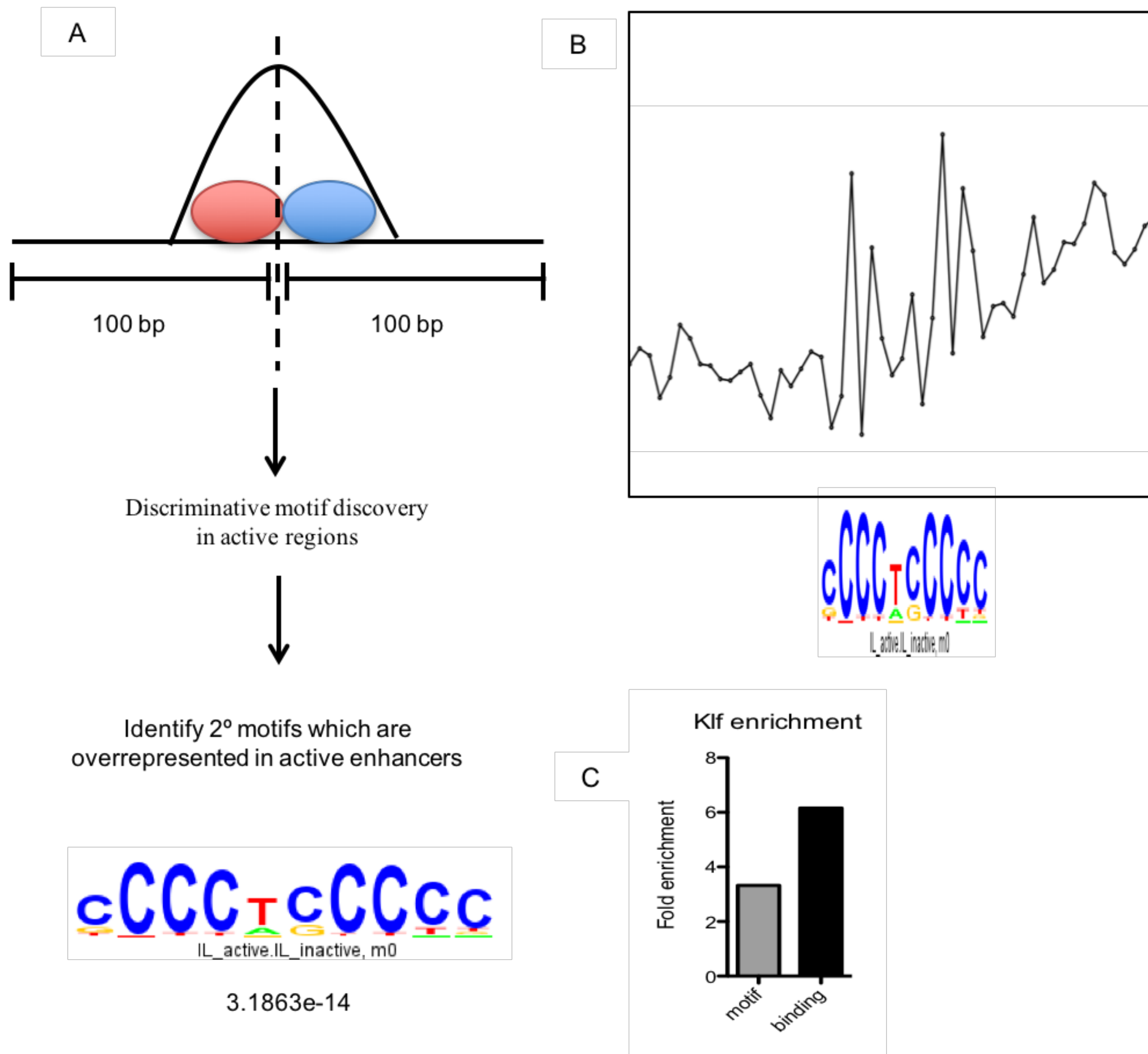


Figure 5.1. Klf factors are enriched at interacting enhancers. (A) Strategy for the de novo discovery of motifs enriched in proximity to Isl1/Lhx3 bound enhancers. **(B)** Local profile of DNaseI cleavage points at Klf motifs in enhancers. **(C)** Klf motifs are ~3 times more likely to occur in engaged enhancers and 6 times more likely to be bound.

Experimental Methods

ES cell culture

ESCs were cultured over a layer of Mitomycin-C–treated fibroblast resistant to Neomycin (Fisher) in EmbryoMax D-MEM (Fisher) supplemented with 15% ESC-grade fetal bovine serum (vol/vol, Invitrogen), L-glutamine (Gibco), 0.1 mM β-mercaptoethanol and 100 U ml^{-1} leukemia inhibitory factor.

Motor neuron differentiation and programming

Motor neuron differentiation of ESCs was performed as previously described. Briefly, ESCs were trypsinized (Invitrogen) and seeded at 5×10^5 cells per ml in ANDFK medium (Advanced DMEM/F12:Neurobasal (1:1) Medium, 10% Knockout-SR (vol/vol), Pen/Strep, 2 mM L-glutamine, and 0.1 mM 2-mercaptoethanol) to initiate formation of embryoid bodies (day 0). Medium was exchanged on days 2 and 5 of differentiation. Patterning of embryoid bodies was induced by supplementing media on day 2 with 1 µM all-trans retinoic acid (Sigma) and 0.5 µM Smo agonist of hedgehog signaling (SAG, Calbiochem). For hypaxial motor neuron differentiation DAPT (Selleckchem, S2215) was added to the culture medium at 5 µM on day 4 and day 5 of differentiation. For motor neuron programming doxycycline (Sigma) was added to the culture medium at 1-3 µg ml^{-1} on day 2 of differentiation.

Immunocytochemistry

Embryoid bodies were fixed with 4% paraformaldehyde (vol/vol) in phosphate-buffered saline, embedded in OCT (Tissue-Tek) and sectioned for staining: 24 h at 4 °C for primary antibodies and 4 h at 20–25 °C for secondary antibodies. After staining, samples were mounted with Aqua Poly Mount (Polyscience). Images were acquired with a LSM 510 or 780 Carl Zeiss confocal microscope. We used antibodies to Olig2 (MABN50, Millipore, 1:2,000), Ngn2 (SC-19233, Santa Cruz, 1:200), V5 (R960, Invitrogen, 1:500), Ki67 (550609, BD Biosciences, 1:100), Tuj1 (ab7751, Abcam, 1:500), SV2 (SV2, Developmental Studies Hybridoma Bank, 1:100), β3-Tubulin (D71G9, Cell Signaling, 1:1,000), Vacht (AB1578, Chemicon, 1:1,000); NeuN (MAB377, Chemicon, 1:500), mNCAM (1:1,000), Hb9 (1:5,000), Isl1 (1:100); Chat (1:500) (gifts from T. Jessell); and Phox2b (1:500; gift from J.F. Brunet, Institut de Biologie de l'École Normale

Supérieure). Alexa 488–(A11001, A11008), FITC-(715-095-150, 706-095-148), Cy3-(715-165-150, 711-165-152, 706-165-148) and Cy5-conjugated (715-175-150, 711-175-152) secondary antibodies were used (Invitrogen or Jackson Immunoresearch, 1:2,000). For quantification, 100 cells in random fields were scored for coexpression of transgene and motor neuron markers by at least two independent authors (n = 1). Data distribution was assumed to be normal, but this was not formally tested. No statistical methods were used to pre-determine sample sizes, but our sample sizes are similar to those generally employed in the field.

Imaging

All images were acquired using confocal laser scanning microscope (LSM Zeiss Meta 510 or 780) with either 10X or 20X objective. Images and quantifications were analyzed using Image-J or manually counted.

Electrophysiology

For whole-cell patch-clamp recordings, NIP and NIL cells were co-cultured with mouse primary cortical astrocytes. Briefly, astrocytes were prepared as previously described and plated on 25-mm diameter coverslips at a density of 50,000 cells per well in a six-well plate. 4 d following astrocyte plating, freshly dissociated NIP and NIL cells were added to the wells at a density of 100,000 cells per well. Cultures were maintained for 7 d before recording. Current-clamp recordings were performed using an Axopatch 2B amplifier. Data were digitized using a Digidata 1322A digital to analog converter and were recorded at a 10-kHz sample rate using pClamp 10 software (all equipment from Molecular Devices). Patch pipettes were fabricated using a P-97 pipette puller (Sutter Instruments). The external recording solution contained 145 mM NaCl, 5 mM KCl, 10 mM HEPES, 10 mM glucose, 2 mM $CaCl_2$ and 2 mM $MgCl_2$. The pH was adjusted to 7.3 using NaOH and the osmolality adjusted to 325 mOsm with sucrose. The pipette solution contained 130 mM CH_3KO_3S, 10 mM CH_3NaO_3S, 1 mM $CaCl_2$, 10 mM EGTA, 10 mM HEPES, 5 mM MgATP and 0.5 mM Na_2GTP (pH 7.3, 305 mOsm). Experiments were performed at 21–23 °C. During recordings, current was injected to hold the cells at −60 mV. Action potentials were evoked using incrementally increasing current steps 1 s in duration. The maximum amplitude of the current step (20–50 pA) and the size of the increment were calculated from the input resistance of the cell. No statistical

methods were used to pre-determine sample sizes, but our sample sizes are similar to those generally employed in the field.

Co-immunoprecipitation

Cells were grown, differentiated and crosslinked as described for chromatin immunoprecipitation experiments (*n* = 1). Cells were lysed in RIPA (50 mM Tris (pH 8.0), 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) with protease inhibitors (Roche, 11697498001). Extracts were sonicated four times for 20 s using a Misonix 3000 at 70%. Extracts were then centrifuged at 4 °C for 10 min at 20,000 *g*. Protein complexes were immunoprecipitated overnight at 4 °C using 5 μg of rabbit antibody to V5 (Abcam, ab1229) or rabbit IgG (NB810-56910, Novus Biologicals) bound to 50 μl Dynabeads (Life Technologies, 100-04D). Immunoprecipitates were washed three times with lysis buffer followed by 10 min of boiling in SDS running buffer. Western blots were developed using mouse antibody to Isl1 (39.4D5, 1:1,000, Developmental Studies Hybridoma Bank).

***In ovo* electroporation**

The GFP enhancer reporter plasmid containing a minimal promoter was cloned by replacing Luciferase gene in pGL4.23[luc2/minP] vector (Promega, Part# 9PIE841A) with destabilized GFP (ZsGreen1) reporter sequence from pZsGreen1-DR Vector (Clontech, Cat# 632428). Enhancer DNA fragments were subcloned into the reporter plasmid and co- electroporated with a ubiquitously expressed CMV-mCherry reporter as a positive control into the developing neural tube of the Hamburger Hamilton (HH) stage 13 chick embryos (Hamburger and Hamilton, 1951; Muhr et al., 1999), using ECM 830 Square Wave Electroporation System (BTX, 45-0002). Chick embryos were incubated in 39°C incubator, and analyzed at 24 and 48 hours after electroporation, respectively. Plasmid-injected chicks were dissected, fixed in 4% paraformaldehyde, embedded in OCT, sectioned and then processed for immunocytochemistry as described above.

CRISPR/Cas9 genome engineering

I used a plasmid generated by Martin Jacko to express the Cas9 enzyme (Cong et al. 2013; Mali et al. 2013) with a self-cleavable mCherry reporter (Cas-p2A-mCherry), in combination with a gRNA plasmid

previously established (Mali et al. 2013). The plasmids were transfected together with (or without) an ssDNA oligonucleotide (donor DNA for homologous recombination) into mouse embryonic stem cells (mESCs) using Mouse Neural Stem Cell Nucleofector Kit (setting A24). To make targeted deletions of enhancers I used pairs of gRNAs flanking the enhancer peak spanning approximately 1kb in the genome. After 24 hours of expression, cells were harvested and FACS-sorted using the mCherry reporter. mCherry+ cells were cultured at low density for 5 days and then to picked for individual clones that were expanded for genotyping and further culturing.

Sequencing and genome alignment

ChIP-seq, ATAC-Seq, ChIA-PET and ChIP-exo DNA samples were analyzed by single-end sequencing, and ATAC-seq and ChIA-PET DNA samples were analyzed using a paired-end sequencing (HiSeq 2000 Sequencing Systems, Illumina). All sequencing data sets were aligned using Bowtie (version 0.12.7)(Langmead et al., 2009) to build version NCBI38/mm10 assembly of the mouse genome.

ATAC-Seq

ATAC-seq libraries were prepared as described previously (Buenrostro et al., 2013). Briefly, approximately 50,000 cells were lysed and spun down. The nuclear pellet was resuspended in 50 µL of the transposase reaction mix (Illumina DNA Library Preparation Kit, FC-121-1030) containing 2.5 µL Tn5 transposase, 25 µL 2xTD buffer, 22.5 uL nuclease-free water, and then incubated for 30 min at 37°C. The transposase-treated DNA was PCR amplified with sequencing adaptors, purified, and sequenced by Illumina HiSeq 2000.

ChIP-seq and ChIP-exo

Approximately 30 million cells were crosslinked with 1% formaldehyde, then processed through the ChIP-exo assay as described previously (Rhee and Pugh, 2011, 2012), using Illumina adaptors. Briefly, cells were lysed, and chromatin pellets were isolated and then solubilized and fragmented by sonication. Fragmented chromatin was then subject to immunoprecipitation using magnetic beads coupled with Protein G (Life Technologies, 10004D) and antibodies (~5 µg) against protein of interest (Isl1, the

Developmental Studies Hybridoma Bank, clone number 39.3F7, gifts from T. Jessell; Onecut1, Santa Cruz, sc-13050, Lot# J04505; H3K27ac, Abcam, ab4729, Lot# GR00563-2; p300, Santa Cruz, sc-584, Lot# I2712) or V5 epitope-tagged Ngn2 (using iNgn2 cell line) and Lhx3 (using iNIL cell line) (V5, Invitrogen, R960, Lot# 1652168). After washing the beads to remove un-bound proteins and DNA, ChIP-seq samples were eluted from the magnetic beads. While ChIP-exo samples were still on the beads, the immunoprecipitated DNA fragments were polished, and ligated to an appropriate sequencing library adaptor. Samples were then subjected to lambda exonuclease digestion (NEB, Cat# M0262L), which processively removes nucleotides from the 5' end of double-stranded DNA until it encounters a protein-DNA crosslink induced by formaldehyde treatment. The resulting single-stranded DNA was eluted from the magnetic beads and converted to double- stranded DNA by primer annealing and extension. A second sequencing adaptor was ligated to exonuclease treated ends, PCR amplified, gel purified, and sequenced by Illumina HiSeq 2000.

RNA-Seq

RNA was extracted using TRIzol reagent (Thermo, 15596026). Illumina Ribo-Zero rRNA Removal Kit (MRZH116) was used with DNase-treated total RNA sample to remove ribosomal RNA. RNA-seq libraries were prepared for sequencing using standard Illumina protocols, and sequenced by Illumina HiSeq 2000. Sequencing data was aligned to the mouse genome using RNA-seq alignment algorithm in the STAR (Spliced Transcripts Alignment to a Reference) software (version2.3.1)(Dobin et al., 2013). To estimate the relative abundances of transcripts, we used the Cufflinks software (release 2.2.1)(Trapnell et al., 2013). FPKM (Fragments Per Kilobase of exon per Million fragments mapped) was calculated per individual gene. For $\log_2$ RNA fold-changes if an FPKM value less than $1.5x10^{-3}$, it was recoded to $1.5x10^{-3}$ to avoid producing denominators with a zero value and FPKM values were then $\log_2$ transformed. To determine genes that significantly

ChIA-PET

ChIA-PET experiments were performed as previously described (Fullwood et al., 2009; Kieffer-Kwon et al., 2013; Li et al., 2012). Briefly, on the appropriate day of differentiation, embryoid bodies were

dissociated in trypsin into single cell suspension. Cells were cross-linked using 2mM DSG and 1% formaldehyde. Cross-linked chromatin was fragmented by sonication to a size of approximately 300bp. Chromatin complexes were immunoprecipitated with monoclonal anti-RNAPII (Covance, 8WG16) coated protein G Dynabeads (Life Technologies). A small portion of ChIP enriched DNA was eluted from beads for quantification. To prepare ChIA-PET libraries DNA was end polished with T4 DNA polymerase (NEB). To assess the degree of intermolecular proximity ligation end polished DNA was divided into 2 aliquots and each ligated to linkers (A or B). The two samples were then joined together for proximity ligation under dilute conditions. Following ligation, samples were treated with Mme1 to release paired end tag (PET) constructs. PET constructs were amplified and submitted to sequencing on Illumina Genome Analyzer II or HiSeq2000.

ChIA-PET interaction calling (CID)

To detect interactions more sensitively and reliably, we developed a method called Chromatin Interaction Discovery (CID) that uses unbiased density-based clustering (Rodriguez and Laio, 2014) to predict interactions directly from paired-end reads. CID takes aligned paired-end reads from ChIA-PET datasets as inputs. The PETs are then segmented into independent genomic region pairs such that no PETs in different regions are within 5000bp of each other. The distance between two PETs is evaluated as the Chebyshev distance:

$$\text{Distance}(PET_1, PET_2) = \max(|read_{1,a} - read_{2,a}|, |read_{1,b} - read_{2,b}|)$$

where $read_{i,a}$ and $read_{i,b}$ are the left and right read positions of PET_i, respectively. A density-based clustering algorithm (Rodriguez and Laio, 2014) is then used to cluster the nearby PETs into candidate interactions. This clustering method recognizes clusters based on the density of the data points, automatically determines the number of clusters and automatically identify outlier data points. The only parameter for this algorithm, d_c, represents the distance within which data points are considered as neighbors to each other. CID estimate d_c based on the genomic span between the region pairs:

$d_c = 100 + span/20$

If the two region pairs overlap with each other, the span is defined as the maximum distance among all the start and end positions of the two regions. d_c is capped at 5000bp because chromatin interactions rarely contains PETs that are 5000bp apart. After clustering, singleton PETs are considered as noise and are excluded. PET clusters with count>=2 are considered candidate interactions and their statistical significance is computed using a published model MICC (He et al., 2015). Interactions with a false discovery rate 0.05 are used for downstream analysis.

References

Almer, a, Rudolph, H., Hinnen, a, and Hörz, W. (1986). Removal of positioned nucleosomes from the yeast PHO5 promoter upon PHO5 induction releases additional upstream activating DNA elements. EMBO J. *5*, 2689–2696.

Apostolou, E., and Thanos, D. (2008). Virus Infection Induces NF-??B-Dependent Interchromosomal Associations Mediating Monoallelic IFN-?? Gene Expression. Cell *134*, 85–96.

Arber, S. (2012). Motor Circuits in Action: Specification, Connectivity, and Function. Neuron *74*, 975–989.

Arber, S., Han, B., Mendelsohn, M., Smith, M., Jessell, T.M., and Sockanathan, S. (1999). Requirement for the homeobox gene Hb9 in the consolidation of motor neuron identity. Neuron *23*, 659–674.

Arlotta, P., Molyneaux, B.J., Chen, J., Inoue, J., Kominami, R., and Macklis, J.D. (2005). Neuronal Subtype-Specific Genes that Control Corticospinal Motor Neuron Development In Vivo. *45*, 207–221.

Audouard, E., Schakman, O., René, F., Huettl, R.E., Huber, A.B., Loeffler, J.P., Gailly, P., and Clotman, F. (2012). The Onecut Transcription Factor HNF-6 Regulates in Motor Neurons the Formation of the Neuromuscular Junctions. PLoS One *7*.

Bailey, T.L., Johnson, J., Grant, C.E., and Noble, W.S. (2015). The MEME Suite. Nucleic Acids Res. *43*, W39–W49.

Bell, A.C., West, A.G., and Felsenfeld, G. (1999). The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. Cell *98*, 387–396.

Berger, M.F., Badis, G., Gehrke, A.R., Talukder, S., Philippakis, A.A., Peña-Castillo, L., Alleyne, T.M., Mnaimneh, S., Botvinnik, O.B., Chan, E.T., et al. (2008). Variation in Homeodomain DNA Binding Revealed by High-Resolution Analysis of Sequence Preferences. Cell *133*, 1266–1276.

Bernstein, B.E., Mikkelsen, T.S., Xie, X., Kamal, M., Huebert, D.J., Cuff, J., Fry, B., Meissner, A., Wernig, M., Plath, K., et al. (2006). A Bivalent Chromatin Structure Marks Key Developmental Genes in Embryonic Stem Cells. Cell *125*, 315–326.

Bertrand, N., Castro, D.S., and Guillemot, F. (2002). Proneural genes and the specification of neural cell types. Nat. Rev. Neurosci. *3*, 517–530.

Boyer, L.A., Tong, I.L., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., et al. (2005). Core transcriptional regulatory circuitry in human

embryonic stem cells. Cell *122*, 947–956.

Briscoe, J., Pierani, A., Jessell, T.M., and Ericson, J. (2000). A homeodomain protein code specifies progenitor cell identity and neuronal fate in the ventral neural tube. Cell *101*, 435–445.

Buenrostro, J.D., Giresi, P.G., Zaba, L.C., Chang, H.Y., and Greenleaf, W.J. (2013). Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat. Methods *10*, 1213–1218.

Bui, T. V., Akay, T., Loubani, O., Hnasko, T.S., Jessell, T.M., and Brownstone, R.M. (2013). Circuits for Grasping: Spinal dI3 Interneurons Mediate Cutaneous Control of Motor Behavior. Neuron *78*, 191–204.

Bulger, M., and Groudine, M. (1999). Looping versus linking: Toward a model for long-distance gene activation. Genes Dev. *13*, 2465–2477.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327–339.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell *49*, 825–837.

Catela, C., Shin, M.M., and Dasen, J.S. (2015). Assembly and Function of Spinal Circuits for Motor Control. Annu. Rev. Cell Dev. Biol. *31*, 669–698.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J., et al. (2008). Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. Cell *133*, 1106–1117.

Cirillo, L.A., Lin, F.R., Cuesta, I., Friedman, D., Jarnik, M., and Zaret, K.S. (2002). Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. Mol. Cell *9*, 279–289.

Coppola, E., Pattyn, A., Guthrie, S.C., Goridis, C., and Studer, M. (2005). Reciprocal gene replacements reveal unique functions for Phox2 genes during neural differentiation. EMBO J. *24*, 4392–4403.

Creyghton, M.P., Cheng, A.W., Welstead, G.G., Kooistra, T., Carey, B.W., Steine, E.J., Hanna, J., Lodato, M.A., Frampton, G.M., Sharp, P.A., et al. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl. Acad. Sci. *107*, 21931–21936.

Dasen, J.S. (2009). Chapter 4 Transcriptional Networks in the Early Development of Sensory-Motor Circuits (Elsevier Inc.).

Dasen, J.S., and Jessell, T.M. (2009). Chapter Six Hox Networks and the Origins of Motor Neuron Diversity (Elsevier Inc.).

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

Dekker, J. (2002). Capturing Chromosome Conformation. Science (80-.). *295*, 1306–1311.

Dekker, J., Marti-Renom, M.A., and Mirny, L.A. (2013). Interpreting Chromatin Interaction Data. Nat Rev Genet *14*, 390–403.

Deneris, E.S., and Hobert, O. (2014). Maintenance of postmitotic neuronal cell identity. Nat. Neurosci. *17*, 899–907.

Dessaud, E., McMahon, A.P., and Briscoe, J. (2008). Pattern formation in the vertebrate neural tube: a sonic hedgehog morphogen-regulated transcriptional network. Development *135*, 2489–2503.

Dillon, A.K., Fujita, S.C., Matise, M.P., Jarjour, A.A., Kennedy, T.E., Kollmus, H., Arnold, H., Weiner, J.A., Sanes, J.R., and Kaprielian, Z. (2005). Molecular Control of Spinal Accessory Motor Neuron/Axon Development in the Mouse Spinal Cord. Situ *25*, 10119–10130.

Dixon, J.R., Selvaraj, S., Yue, F., Kim, A., Li, Y., Shen, Y., Hu, M., Liu, J.S., and Ren, B. (2012). Topological domains in mammalian genomes identified by analysis of chromatin interactions. Nature *485*, 376–380.

Dixon, J.R., Jung, I., Selvaraj, S., Shen, Y., Antosiewicz-Bourget, J.E., Lee, A.Y., Ye, Z., Kim, A., Rajagopal, N., Xie, W., et al. (2015). Chromatin architecture reorganization during stem cell differentiation. Nature *518*, 331–336.

Dorsett, D. (2011). Cohesin: Genomic insights into controlling gene transcription and development. Curr. Opin. Genet. Dev. *21*, 199–206.

Dowen, J.M., Fan, Z.P., Hnisz, D., Ren, G., Abraham, B.J., Zhang, L.N., Weintraub, A.S., Schuijers, J., Lee, T.I., Zhao, K., et al. (2014). Control of cell identity genes occurs in insulated neighborhoods in mammalian chromosomes. Cell *159*, 374–387.

Dubreuil, V., Hirsch, M.-R., Jouve, C., Brunet, J.-F., and Goridis, C. (2002). The role of Phox2b in synchronizing pan-neuronal and type-specific aspects of neurogenesis. Development *129*, 5241–5253.

Ericson, J., Thor, S., Edlund, T., Jessell, T.M., and Yamada, T. (1992). Early stages of motor neuron

differentiation revealved by expression of homeobox gene Islet-1. Science (80-.). *256*, 1555–1560.

Fishell, G., and Heintz, N. (2013). The neuron identity problem: Form meets function. Neuron *80*, 602–612.

Flames, N., and Hobert, O. (2009). Gene regulatory logic of dopamine neuron differentiation. Nature *458*, 885–889.

Francius, C., and Clotman, F. (2010). Dynamic expression of the Onecut transcription factors HNF-6, OC-2 and OC-3 during spinal motor neuron development. Neuroscience *165*, 116–129.

Fullwood, M.J., Liu, M.H., Pan, Y.F., Liu, J., Xu, H., Mohamed, Y. Bin, Orlov, Y.L., Velkov, S., Ho, A., Mei, P.H., et al. (2009). An oestrogen-receptor-alpha-bound human chromatin interactome. Nature *462*, 58–64.

Gao, B.X., and Ziskind-Conhaim, L. (1998). Development of ionic currents underlying changes in action potential waveforms in rat spinal motoneurons. J. Neurophysiol. *80*, 3047–3061.

Gaspard, N., Bouschet, T., Herpoel, A., Naeije, G., van den Ameele, J., and Vanderhaeghen, P. (2009). Generation of cortical neurons from mouse embryonic stem cells. Nat. Protoc. *4*, 1454–1463.

Ge, K., Guermah, M., Yuan, C.-X., Ito, M., Wallberg, A.E., Spiegelman, B.M., and Roeder, R.G. (2002). Transcription coactivator TRAP220 is required for PPARγ2-stimulated adipogenesis. Nature *417*, 563–567.

German, M.S., Wang, J., Chadwick, R.B., and Rutter, W.J. (1992). Synergistic activation of the insulin gene by a LIM-homeo domain protein and a basic helix-loop-helix protein: Building a functional insulin minienhancer complex. Genes Dev. *6*, 2165–2176.

Ghisletti, S., Barozzi, I., Mietton, F., Polletti, S., De Santa, F., Venturini, E., Gregory, L., Lonie, L., Chew, A., Wei, C.L., et al. (2010). Identification and Characterization of Enhancers Controlling the Inflammatory Gene Expression Program in Macrophages. Immunity *32*, 317–328.

Gordon, P.M., and Hobert, O. (2015). A Competition Mechanism for a Homeotic Neuron Identity Transformation in C. elegans. Dev. Cell *34*, 206–219.

Gorkin, D.U., Leung, D., and Ren, B. (2014). The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell *14*, 771–775.

Grillet, N., Dubreuil, V., Dufour, H.D.H.D., and Brunet, J.-F.J.-F. (2003). Dynamic expression of RGS4 in

the developing nervous system and regulation by the neural type-specific transcription factor Phox2b. J. Neurosci. *23*, 10613–10621.

Guo, Y., Xu, Q., Canzio, D., Shou, J., Li, J., Gorkin, D.U., Jung, I., Wu, H., Zhai, Y., Tang, Y., et al. (2015). CRISPR Inversion of CTCF Sites Alters Genome Topology and Enhancer/Promoter Function. Cell *162*, 900–910.

Hasan, K.B., Agarwala, S., and Ragsdale, C.W. (2010). PHOX2A regulation of oculomotor complex nucleogenesis. Development *137*, 1205–1213.

He, C., Zhang, M.Q., and Wang, X. (2015). MICC: an R package for identifying chromatin interactions from ChIA-PET data. Bioinformatics *31*, 3832–3834.

Heintzman, N.D., and Ren, B. (2009). Finding distal regulatory elements in the human genome. Curr. Opin. Genet. Dev. *19*, 541–549.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heintzman, N.D., Hon, G.C., Hawkins, R.D., Kheradpour, P., Stark, A., Harp, L.F., Ye, Z., Lee, L.K., Stuart, R.K., Ching, C.W., et al. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Hester, M.E., Murtha, M.J., Song, S., Rao, M., Miranda, C.J., Meyer, K., Tian, J., Boulting, G., Schaffer, D. V, Zhu, M.X., et al. (2011). Rapid and Efficient Generation of Functional Motor Neurons From Human Pluripotent Stem Cells Using Gene Delivered Transcription Factor Codes. Mol. Ther. *19*, 1905–1912.

Hirsch, M.R., Glover, J.C., Dufour, H.D., Brunet, J.F., and Goridis, C. (2007). Forced expression of Phox2 homeodomain transcription factors induces a branchio-visceromotor axonal phenotype. Dev. Biol. *303*, 687–702.

Hnisz, D., Abraham, B.J., Lee, T.I., Lau, A., Saint-André, V., Sigova, A.A., Hoke, H.A., and Young, R.A. (2013). Super-enhancers in the control of cell identity and disease. Cell *155*, 934–947.

Hnisz, D., Day, D.S., and Young, R.A. (2016). Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. Cell *167*, 1188–1200.

Hobert, O. (2008). Regulatory logic of neuronal diversity: terminal selector genes and selector motifs.

Proc. Natl. Acad. Sci. U. S. A. *105*, 20067–20071.

Hobert, O. (2011). Regulation of terminal differentiation programs in the nervous system. Annu. Rev. Cell Dev. Biol. *27*, 681–696.

Holmes, G., and Niswander, L. (2001). Expression of slit-2 and slit-3 during chick development. Dev. Dyn. *222*, 301–307.

Hong, J.-W., Hendrix, D.A., and Levine, M.S. (2008). Shadow Enhancers as a Source of Evolutionary Novelty. Science (80-.). *321*, 1314.

Hong, J.W., Park, K.W., and Levine, M.S. (2013). Temporal regulation of single-minded target genes in the ventral midline of the Drosophila central nervous system. Dev. Biol. *380*, 335–343.

Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences 79, 2554–2558.

Jensen, K.B., Dredge, B.K., Stefani, G., Zhong, R., Buckanovich, R.J., Okano, H.J., Yang, Y.Y.L., and Darnell, R.B. (2000). Nova-1 Regulates Neuron-Specific Alternative Splicing and Is Essential for Neuronal Viability. Neuron *25*, 359–371.

Jessell, T.M. (2000a). Neuronal Specification in the Spinal Cord : Inductive Signals and Transcriptional Codes. Nat. Rev. Genet. *1*, 20–29.

Jessell, T.M. (2000b). Neuronal Specification in the Spinal Cord : Inductive Signals and Transcriptional Codes. Nat. Rev. Genet. *1*, 20–29.

Johnson, S.A., Harmon, K.J., Smiley, S.G., Still, F.M., and Kavaler, J. (2011). Discrete regulatory regions control early and late expression of D-Pax2 during external sensory organ development. Dev. Dyn. *240*, 1769–1778.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010a). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kagey, M.H., Newman, J.J., Bilodeau, S., Zhan, Y., Orlando, D.A., van Berkum, N.L., Ebmeier, C.C., Goossens, J., Rahl, P.B., Levine, S.S., et al. (2010b). Mediator and cohesin connect gene expression and chromatin architecture. Nature *467*, 430–435.

Kania, A., and Jessell, T.M. (2003). Topographic motor projections in the limb imposed by LIM

homeodomain protein regulation of ephrin-A:EphA interactions. Neuron *38*, 581–596.

Kania, A., Johnson, R.L., and Jessell, T.M. (2000). Coordinate Roles for LIM Homeobox Genes in Directing the Dorsoventral Trajectory of Motor Axons in the Vertebrate Limb. Cell *102*, 161–173.

Kieffer-Kwon, K.R., Tang, Z., Mathe, E., Qian, J., Sung, M.H., Li, G., Resch, W., Baek, S., Pruett, N., Grøntved, L., et al. (2013). Interactome maps of mouse gene regulatory domains reveal basic principles of transcriptional regulation. Cell *155*, 1507–1520.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008). An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. Cell *132*, 1049–1061.

Kim, J., Woo, A.J., Chu, J., Snow, J.W., Fujiwara, Y., Kim, C.G., Cantor, A.B., and Orkin, S.H. (2010a). A Myc Network Accounts for Similarities between Embryonic Stem and Cancer Cell Transcription Programs. Cell *143*, 313–324.

Kim, T.-K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010b). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Kratsios, P., Stolfi, A., Levine, M., and Hobert, O. (2012). Coordinated regulation of cholinergic motor neuron traits through a conserved terminal selector gene. Nat. Neurosci. *15*, 205–214.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.

de Laat, W., Klous, P., Kooren, J., Noordermeer, D., Palstra, R.J., Simonis, M., Splinter, E., and Grosveld, F. (2008). Three-Dimensional Organization of Gene Expression in Erythroid Cells. Curr. Top. Dev. Biol. *82*, 117–139.

Lee, S.K., and Pfaff, S.L. (2001). Transcriptional networks regulating neuronal identity in the developing spinal cord. Nat. Neurosci. *4 Suppl*, 1183–1191.

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

Lee, S., Cuvillier, J.M., Lee, B., Shen, R., Lee, J.W., and Lee, S.-K. (2012). Fusion protein Isl1-Lhx3 specifies motor neuron fate by inducing motor neuron genes and concomitantly suppressing the interneuron programs. Proc. Natl. Acad. Sci. *109*, 3383–3388.

Lee, S.-K., Jurata, L.W., Funahashi, J., Ruiz, E.C., and Pfaff, S.L. (2004). Analysis of embryonic motoneuron gene regulation: derepression of general activators function in concert with enhancer factors. Development *131*, 3295–3306.

Levine, M. (2010). Transcriptional enhancers in animal development and evolution. Curr. Biol. *20*, R754–R763.

Li, G., Ruan, X., Auerbach, R.K., Sandhu, K.S., Zheng, M., Wang, P., Poh, H.M., Goh, Y., Lim, J., Zhang, J., et al. (2012). Extensive promoter-centered chromatin interactions provide a topological basis for transcription regulation. Cell *148*, 84–98.

Lichtsteiner, S., and Tjian, R. (1995). Synergistic activation of transcription by UNC-86 and MEC-3 in Caenorhabditis elegans embryo extracts. EMBO J. *14*, 3937–3945.

Lieberman-aiden, E., Berkum, N.L. Van, Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). of the Human Genome. Analyzer *33292*, 289–293.

Livesey, F.J., and Cepko, C.L. (2001). Vertebrate neural cell-fate determination: lessons from the retina. Nat. Rev. Neurosci. *2*, 109–118.

Lodato, S., and Arlotta, P. (2015). Generating Neuronal Diversity in the Mammalian Cerebral Cortex. Annu. Rev. Cell Dev. Biol. *31*, 699–720.

Lodato, M.A., Ng, C.W., Wamstad, J.A., Cheng, A.W., Thai, K.K., Fraenkel, E., Jaenisch, R., and Boyer, L.A. (2013). SOX2 Co-Occupies Distal Enhancer Elements with Distinct POU Factors in ESCs and NPCs to Specify Cell State. PLoS Genet. *9*.

Lodato, S., Molyneaux, B.J., Zuccaro, E., Goff, L.A., Chen, H.-H., Yuan, W., Meleski, A., Takahashi, E., Mahony, S., Rinn, J.L., et al. (2014). Gene co-regulation by Fezf2 selects neurotransmitter identity and connectivity of corticospinal neurons. Nat. Neurosci. *17*, 1046–1054.

Lomvardas, S., Barnea, G., Pisapia, D.J., Mendelsohn, M., Kirkland, J., and Axel, R. (2006). Interchromosomal Interactions and Olfactory Receptor Choice. Cell *126*, 403–413.

Luo, L., Callaway, E.M., and Svoboda, K. (2008). Genetic Dissection of Neural Circuits. Neuron *57*, 634–660.

Mahony, S., Mazzoni, E.O., McCuine, S., Young, R.A., Wichterle, H., and Gifford, D.K. (2011). Ligand-

dependent dynamics of retinoic acid receptor binding during early neurogenesis. Genome Biol. *12*, R2.

Mann, R.S., and Carroll, S.B. (2002). Molecular mechanisms of selector gene function and evolution. Curr. Opin. Genet. Dev. *12*, 592–600.

Maroof, A.M., Keros, S., Tyson, J.A., Ying, S.-W., Ganat, Y.M., Merkle, F.T., Liu, B., Goulburn, A., Stanley, E.G., Elefanty, A.G., et al. (2013). Directed Differentiation and Functional Maturation of Cortical Interneurons from Human Embryonic Stem Cells. Cell Stem Cell *12*, 559–572.

Marson, A., Levine, S.S., Cole, M.F., Frampton, G.M., Brambrink, T., Johnstone, S., Guenther, M.G., Johnston, W.K., Wernig, M., Newman, J., et al. (2008). Connecting microRNA Genes to the Core Transcriptional Regulatory Circuitry of Embryonic Stem Cells. Cell *134*, 521–533.

Mazzoni, E.O., Mahony, S., Iacovino, M., Morrison, C.A., Mountoufaris, G., Closser, M., Whyte, W.A., Young, R.A., Kyba, M., Gifford, D.K., et al. (2011). Embryonic stem cell–based mapping of developmental transcriptional programs. Nat. Methods *8*, 1056–1058.

Mazzoni, E.O., Mahony, S., Peljto, M., Patel, T., Thornton, S.R., McCuine, S., Reeder, C., Boyer, L.A., Young, R.A., Gifford, D.K., et al. (2013a). Saltatory remodeling of Hox chromatin in response to rostrocaudal patterning signals. Nat. Neurosci. *16*, 1191–1198.

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

Miles, G.B. (2004). Functional Properties of Motoneurons Derived from Mouse Embryonic Stem Cells. J. Neurosci. *24*, 7848–7858.

Mizuguchi, R., Sugimori, M., Takebayashi, H., Kosako, H., Nagao, M., Yoshida, S., Nabeshima, Y.I., Shimamura, K., and Nakafuku, M. (2001). Combinatorial roles of Olig2 and Neurogenin2 in the coordinated induction of pan-neuronal and subtype-specific properties of motoneurons. Neuron *31*, 757–771.

Mo, A., Mukamel, E.A., Davis, F.P., Luo, C., Henry, G.L., Picard, S., Urich, M.A., Nery, J.R., Sejnowski, T.J., Lister, R., et al. (2015). Epigenomic Signatures of Neuronal Diversity in the Mammalian Brain. Neuron *86*, 1369–1384.

Molyneaux, B.J., Arlotta, P., Menezes, J.R.L., and Macklis, J.D. (2007). Neuronal subtype specification in

the cerebral cortex. Nat. Rev. Neurosci. *8*, 427–437.

Montavon, T., Soshnikova, N., Mascrez, B., Joye, E., Thevenet, L., Splinter, E., De Laat, W., Spitz, F., and Duboule, D. (2011). A regulatory archipelago controls hox genes transcription in digits. Cell *147*, 1132–1145.

Moore, R.Y. (1989). Cranial motor neurons contain either galanin??? or calcitonin gene???related peptidelike immunoreactivity. J. Comp. Neurol. *282*, 512–522.

Nora, E.P., Lajoie, B.R., Schulz, E.G., Giorgetti, L., Okamoto, I., Servant, N., Piolot, T., van Berkum, N.L., Meisig, J., Sedat, J., et al. (2012). Spatial partitioning of the regulatory landscape of the X-inactivation centre. Nature *485*, 381–385.

Nord, A.S., Blow, M.J., Attanasio, C., Akiyama, J.A., Holt, A., Hosseini, R., Phouanenavong, S., Plajzer-Frick, I., Shoukry, M., Afzal, V., et al. (2013). Rapid and pervasive changes in genome-wide enhancer usage during mammalian development. Cell *155*, 1521–1531.

Novitch, B.G., Chen, A.I., and Jessell, T.M. (2001). Coordinate regulation of motor neuron subtype identity and pan-neuronal properties by the bHLH repressor Olig2. Neuron *31*, 773–789.

Panne, D., Maniatis, T., and Harrison, S.C. (2007). An Atomic Model of the Interferon-?? Enhanceosome. Cell *129*, 1111–1123.

Parelho, V., Hadjur, S., Spivakov, M., Leleu, M., Sauer, S., Gregson, H.C., Jarmuz, A., Canzonetta, C., Webster, Z., Nesterova, T., et al. (2008). Cohesins Functionally Associate with CTCF on Mammalian Chromosome Arms. Cell *132*, 422–433.

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

Pattyn, A., Morin, X., Cremer, H., Goridis, C., and Brunet, J.-F. (1997). Expression and interactions of the two closely related homeobox genes Phox2a and Phox2b during neurogenesis. Development *124*, 4065–4075.

Pearson, B.J., and Doe, C.Q. (2004). Specification of Temporal Identity in the Developing Nervous System. Annu. Rev. Cell Dev. Biol. *20*, 619–647.

Peljto, M., and Wichterle, H. (2011). Programming embryonic stem cells to neuronal subtypes. Curr. Opin.

Neurobiol. *21*, 43–51.

Peljto, M., Dasen, J.S., Mazzoni, E.O., Jessell, T.M., and Wichterle, H. (2010). Functional diversity of ESC-derived motor neuron subtypes revealed through intraspinal transplantation. Cell Stem Cell *7*, 355–366.

Pennacchio, L. a, Ahituv, N., Moses, A.M., Prabhakar, S., Nobrega, M. a, Shoukry, M., Minovitsky, S., Dubchak, I., Holt, A., Lewis, K.D., et al. (2006). In vivo enhancer analysis of human conserved non-coding sequences. Nature *444*, 499–502.

Peter, I.S., and Davidson, E.H. (2011). Evolution of gene regulatory networks controlling body plan development. Cell *144*, 970–985.

Pfaff, S.L., Mendelsohn, M., Stewart, C.L., Edlund, T., and Jessell, T.M. (1996). Requirement for LIM homeobox gene Isl1 in motor neuron generation reveals a motor neuron-dependent step in interneuron differentiation. Cell *84*, 309–320.

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

Phillips-Cremins, J.E., Sauria, M.E.G., Sanyal, A., Gerasimova, T.I., Lajoie, B.R., Bell, J.S.K., Ong, C.T., Hookway, T.A., Guo, C., Sun, Y., et al. (2013). Architectural protein subclasses shape 3D organization of genomes during lineage commitment. Cell *153*, 1281–1295.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–283.

Ran, F.A., Hsu, P.D., Wright, J., Agarwala, V., Scott, D.A., and Zhang, F. (2013). Genome engineering using the CRISPR-Cas9 system. Nat. Protoc. *8*, 2281–2308.

Reeder, C., Closser, M., Poh, H.M., Sandhu, K., Wichterle, H., and Gifford, D. (2015). High resolution mapping of enhancer-promoter interactions. PLoS One *10*, 1–22.

Rhee, H.S., and Pugh, B.F. (2011). Comprehensive genome-wide protein-DNA interactions detected at single-nucleotide resolution. Cell *147*, 1408–1419.

Rodriguez, A., and Laio, A. (2014). Clustering by fast search and find of density peaks. Science *344*, 1492–1496.

Roeder, R.G. (2005). Transcriptional regulation and the role of diverse coactivators in animal cells. FEBS Lett. *579*, 909–915.

Rouaux, C., and Arlotta, P. (2010). Fezf2 directs the differentiation of corticofugal neurons from striatal progenitors in vivo. Nat. Neurosci. *13*, 1345–1347.

Roy, a., Francius, C., Rousso, D.L., Seuntjens, E., Debruyn, J., Luxenhofer, G., Huber, a. B., Huylebroeck, D., Novitch, B.G., and Clotman, F. (2012). Onecut transcription factors act upstream of Isl1 to regulate spinal motoneuron diversification. Development *139*, 3109–3119.

Samstein, R.M., Arvey, A., Josefowicz, S.Z., Peng, X., Reynolds, A., Sandstrom, R., Neph, S., Sabo, P., Kim, J.M., Liao, W., et al. (2012). Foxp3 exploits a pre-existent enhancer landscape for regulatory T cell lineage specification. Cell *151*, 153–166.

di Sanguinetto, S.A.D.T., Dasen, J.S., and Arber, S. (2008). Transcriptional mechanisms controlling motor neuron diversity and connectivity. Curr. Opin. Neurobiol. *18*, 36–43.

Scardigli, R., Schuurmans, C., Gradwohl, G., and Guillemot, F. (2001). Crossregulation between Neurogenin2 and pathways specifying neuronal identity in the spinal cord. Neuron *31*, 203–217.

Sekiya, S., and Suzuki, A. (2011). Direct conversion of mouse fibroblasts to hepatocyte-like cells by defined factors. Nature *475*, 390–393.

Sharma, K., Sheng, H.Z., Lettieri, K., Li, H., Karavanov, A., Potter, S., Westphal, H., and Pfaff, S.L. (1998). LIM homeodomain factors Lhx3 and Lhx4 assign subtype identities for motor neurons. Cell *95*, 817–828.

Sharma, K., Leonard, A.E., Lettieri, K., and Pfaff, S.L. (2000). Genetic and epigenetic mechanisms contribute to motor neuron pathfinding. Nature *406*, 515–519.

Shen, Y., Yue, F., McCleary, D.F., Ye, Z., Edsall, L., Kuan, S., Wagner, U., Dixon, J., Lee, L., Lobanenkov, V. V., et al. (2012). A map of the cis-regulatory sequences in the mouse genome. Nature *488*, 116–120.

Shirasaki, R., and Pfaff, S.L. (2002). Transcriptional Codes and the Control of Neuronal Identity. Annu. Rev. Neurosci. *25*, 251–281.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat. Rev. Genet. *15*, 272–286.

Singer, T., McConnell, M.J., Marchetto, M.C.N., Coufal, N.G., and Gage, F.H. (2010). LINE-1 retrotransposons: Mediators of somatic variation in neuronal genomes? Trends Neurosci. *33*, 345–354.

Slattery, M., Riley, T., Liu, P., Abe, N., Gomez-Alcala, P., Dror, I., Zhou, T., Rohs, R., Honig, B., Bussemaker, H.J., et al. (2011). Cofactor binding evokes latent differences in DNA binding specificity between hox proteins. Cell *147*, 1270–1282.

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

Song, M.-R., Shirasaki, R., Cai, C.-L., Ruiz, E.C., Evans, S.M., Lee, S.-K., and Pfaff, S.L. (2006). T-Box transcription factor Tbx20 regulates a genetic program for cranial motor neuron cell body migration. Development *133*, 4945–4955.

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

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

Spitz, F., and Furlong, E.E.M. (2012). Transcription factors: From enhancer binding to developmental control. Nat. Rev. Genet. *13*, 613–626.

Stampfel, G., Kazmar, T., Frank, O., Wienerroither, S., Reiter, F., and Stark, A. (2015). Transcriptional regulators form diverse groups with context-dependent regulatory functions. Nature *528*, 147–151.

Taatjes, D.J. (2002). Structure, Function, and Activator-Induced Conformations of the CRSP Coactivator. Science (80-.). *295*, 1058–1062.

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

Tan, G.C., Mazzoni, E.O., and Wichterle, H. (2016). Iterative Role of Notch Signaling in Spinal Motor Neuron Diversification. Cell Rep. *16*, 907–916.

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

Tapscotr, S.J., Davis, R.L., Thayer, M.J., Cheng, P., Weintraub, H., and Lassar, A.B. (1988). MyoD1: A Nuclear Phosphoprotein Reqiring a Myc Homology Region to Convert Fibroblasts to Myoblasts. Sci. (New York, NY) *295*, 405–412.

Thaler, J., Harrison, K., Sharma, K., Lettieri, K., Kehrl, J., and Pfaff, S.L. (1999). Active suppression of interneuron programs within developing motor neurons revealed by analysis of homeodomain factor HB9. Neuron *23*, 675–687.

Thaler, J.P., Lee, S.K., Jurata, L.W., Gill, G.N., and Pfaff, S.L. (2002). LIM factor Lhx3 contributes to the specification of motor neuron and interneuron identity through cell-type-specific protein-protein interactions. Cell *110*, 237–249.

Thanos, D., and Maniatis, T. (1995). Virus induction of human IFN-b gene expression requires the assembly of an enhancosome. Cell *83*, 1091–1100.

Tjian, R., and Maniatis, T. (1994). Transcriptional activation: A complex puzzle with few easy pieces. Cell *77*, 5–8.

Tolhuis, B., Palstra, R.J., Splinter, E., Grosveld, F., and De Laat, W. (2002). Looping and interaction between hypersensitive sites in the active ??-globin locus. Mol. Cell *10*, 1453–1465.

Tsuchida, T., Ensini, M., Morton, S.B., Baldassare, M., Edlund, T., Jessell, T.M., and Pfaff, S.L. (1994). Topographic organization of embryonic motor neurons defined by expression of LIM homeobox genes. Cell *79*, 957–970.

Tursun, B., Patel, T., Kratsios, P., and Hobert, O. (2011). Direct Conversion of C. elegans Germ Cells into Specific Neuron Types. Science (80-.). *331*, 304–308.

Ule, J. (2003). CLIP Identifies Nova-Regulated RNA Networks in the Brain. Science (80-.). *302*, 1212–1215.

Vierbuchen, T., and Wernig, M. (2012). Review Molecular Roadblocks for Cellular Reprogramming. MOLCEL *47*, 827–838.

Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., and Wernig, M. (2010). Direct conversion of fibroblasts to functional neurons. Nature *77*, 7–8.

Visel, A., Blow, M.J., Li, Z., Zhang, T., Akiyama, J. a, Holt, A., Plajzer-Frick, I., Shoukry, M., Wright, C., Chen, F., et al. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*,

854–858.

Wang, A., Yue, F., Li, Y., Xie, R., Harper, T., Patel, N.A., Muth, K., Palmer, J., Qiu, Y., Wang, J., et al. (2015). Epigenetic priming of enhancers predicts developmental competence of hESC-derived endodermal lineage intermediates. Cell Stem Cell *16*, 386–399.

Wang, W., Shi, Q., Guo, T., Yang, Z., Jia, Z., Chen, P., and Zhou, C. (2016). PDX1 and ISL1 differentially coordinate with epigenetic modifications to regulate insulin gene expression in varied glucose concentrations. Mol. Cell. Endocrinol. *428*, 38–48.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). Hierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621–635.

Wei, Z., Gao, F., Kim, S., Yang, H., Lyu, J., An, W., Wang, K., and Lu, W. (2013). Klf4 organizes long-range chromosomal interactions with the OCT4 locus inreprogramming andpluripotency. Cell Stem Cell *13*, 36–47.

Wendt, K.S., Yoshida, K., Itoh, T., Bando, M., Koch, B., Schirghuber, E., Tsutsumi, S., Nagae, G., Ishihara, K., Mishiro, T., et al. (2008). Cohesin mediates transcriptional insulation by CCCTC-binding factor. Nature *451*, 796–801.

Wenick, A.S., and Hobert, O. (2004). Genomic cis-regulatory architecture and trans-acting regulators of a single interneuron-specific gene battery in C. elegans. Dev. Cell *6*, 757–770.

Whyte, W.A., Bilodeau, S., Orlando, D.A., Hoke, H.A., Frampton, G.M., Foster, C.T., Cowley, S.M., and Young, R.A. (2012). Enhancer decommissioning by LSD1 during embryonic stem cell differentiation. Nature *482*, 221–225.

Whyte, W.A., Orlando, D.A., Hnisz, D., Abraham, B.J., Lin, C.Y., Kagey, M.H., Rahl, P.B., Lee, T.I., and Young, R.A. (2013). Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell *153*, 307–319.

Wichterle, H., Wichterle, H., Lieberam, I., Lieberam, I., Porter, J. a, Porter, J. a, Jessell, T.M., and Jessell, T.M. (2002). Directed differentiation of embryonic stem cells into motor neurons. Cell *110*, 385–397.

Wichterle, H., Peljto, M., and Nedelec, S. (2009). Stem Cells in Regenerative Medicine. *482*, 171–183.

de Wit, E., Bouwman, B.A., Zhu, Y., Klous, P., Splinter, E., Verstegen, M.J., Krijger, P.H., Festuccia, N.,

Nora, E.P., Welling, M., et al. (2013). The pluripotent genome in three dimensions is shaped around pluripotency factors. Nature *501*, 227–231.

Xue, D., Tu, Y., and Chalfie, M. (1993). Cooperative interactions between the Caenorhabditis elegans homeoproteins UNC-86 and MEC-3. Science (80-.). *261*, 1324–1328.

Young, R.A. (2011). Review Control of the Embryonic Stem Cell State. Cell *144*, 940–954.

Zaret, K.S., and Carroll, J.S. (2011). Pioneer transcription factors : establishing competence for gene expression Parameters affecting transcription factor access to target sites in chromatin Initiating events in chromatin : pioneer factors bind first. Genes Dev. 2227–2241.

Zhang, H., Wang, W.P., Guo, T., Yang, J.C., Chen, P., Ma, K.T., Guan, Y.F., and Zhou, C.Y. (2009). The LIM-Homeodomain Protein ISL1 Activates Insulin Gene Promoter Directly through Synergy with BETA2. J. Mol. Biol. *392*, 566–577.

Zhang, X., Krutchinsky, A., Fukuda, A., Chen, W., Yamamura, S., Chait, B.T., and Roeder, R.G. (2005). MED1/TRAP220 exists predominantly in a TRAP/Mediator subpopulation enriched in RNA polymerase II and is required for ER-mediated transcription. Mol. Cell *19*, 89–100.

Zhang, Y., Liu, T., Meyer, C.A., Eeckhoute, J., Johnson, D.S., Bernstein, B.E., Nussbaum, C., Myers, R.M., Brown, M., Li, W., et al. (2008). Model-based Analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

Zhang, Y., Wong, C.-H.C.-H., Birnbaum, R.Y., Li, G., Favaro, R., Ngan, C.Y., Lim, J., Tai, E., Poh, H.M., Wong, E., et al. (2013). Chromatin connectivity maps reveal dynamic promoter-enhancer long-range associations. Nature *504*, 306–310.

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