GENOME-WIDE ANALYSIS OF TRANSCRIPTION FACTORS ASCL1 AND PTF1A IN DEVELOPMENT AND CANCER

APPROVED BY SUPERVISORY COMMITTEE

Jane E. Johnson, Ph.D.

Gang. Yu, Ph.D.

Q. Richard Lu, Ph.D.

Tae-Kyung Kim, Ph.D.

DEDICATION

I dedicate this to my mother and wife, for their love and devotion, and their endless support throughout my life.

GENOME-WIDE ANALYSIS OF TRANSCRIPTION FACTORS ASCL1 AND PTF1A IN DEVELOPMENT AND CANCER

by

MARK DOMINIC BORROMEO

DISSERTATION

Presented to the Faculty of the Graduate School of Biomedical Sciences

The University of Texas Southwestern Medical Center at Dallas

In Partial Fulfillment of the Requirements

For the Degree of

DOCTOR OF PHILOSOPHY

The University of Texas Southwestern Medical Center at Dallas

Dallas, Texas

December, 2013

GENOME-WIDE ANALYSIS OF TRANSCRIPTION FACTORS ASCL1 AND PTF1A IN DEVELOPMENT AND CANCER

Mark Dominic Borromeo, Ph.D.

The University of Texas Southwestern Medical Center at Dallas, 2013

Jane E. Johnson, Ph.D.

Cell fate specification in the developing embryo relies on combinations of transcription factors to regulate tissue specific gene programs. Many of the same transcription factors can be found in multiple tissue types and are crucial for their development, and at other times these same factors can be misused in disease states. The basic helix-loop-helix (bHLH) factors Ascl1 and Ptf1a are examples of factors that give rise to and function in multiple tissues. Ascl1 and Ptf1a are essential for generating the correct number and sub-type of neurons in multiple regions of the nervous system. In addition, Ptf1a is required in the developing pancreas for both its formation and maturation, while Ascl1 is crucial for tumor

growth in malignant small cell lung carcinoma (SCLC). It is unknown if Ascl1 and Ptf1a directly regulate different genes programs in these disparate tissues. Furthermore, Ptf1a and Ascl1 are members of same transcription factor family, which recognize and bind a similar DNA sequence. How these two factors achieve specificity of DNA binding and gene target selection *in vivo* is unknown. These questions have long been unanswered due in part to the lack of known direct transcriptional targets. Thus, to understand how Ascl1 and Ptf1a function in these processes, the direct transcriptional targets were identified genome-wide in the multiple tissues using ChIP-Seq and RNA-Seq. Overwhelmingly, Ascl1 and Ptf1a directly regulate different gene programs important for each tissue. Within a given tissue, the specificity of Ascl1 and Ptf1a function is partly explained by their differences in E-box sequence preferences. Ptf1a and Ascl1 are co-expressed in a subset of cells in the dorsal neural tube, and comparative analysis of their binding sites show that they bind a common E-box. However, Ptf1a can also bind a distinct E-box, which is found enriched in binding sites unique to Ptf1a. However, analysis of Ascl1 binding in SCLC and Ptf1a binding in the developing pancreas, shows that their E-box preferences change and does not reflect the same type of E-box they bind in the neural tube. Mechanisms in addition to E-box specificity are likely in use because tissue specific binding coincides with tissue-specific chromatin accessibility and enrichment of lineage-specific transcription factor binding motifs. Thus, Ascl1 and Ptf1a make use of different tissue-specific co-factors to regulate tissue-specific genes. This study provides insights into how a single factor can regulate the transcription of different genes in different tissue types, and how two related E-box binding proteins regulate distinct genes.

ACKNOWLEDGEMENT

This dissertation would not have been possible without the support and guidance from my friends, family, and colleagues.

I would like to acknowledge my mentor Dr. Jane Johnson, for her endless support and guidance throughout my graduate school career. Because of Dr. Johnson's patience and dedication towards teaching her students, I feel prepared to take on any challenge that I may face during my next phase of my scientific career. I have learned through Dr. Johnson that there is no shortcut or substitute for good quality of work, and that good work will always earn the highest respect and appreciation from your colleagues.

I would like to thank the members of my dissertation committee: Drs. Gang Yu, Richard Lu, and Tae-Kyung Kim. Together they have provided me with both technical and biological insight that was crucial for the success of my projects.

I was privileged to work with so many talented scientists on several different projects and I would like to thank all of them for sharing their knowledge and assisting me at the lab bench. I would specially like to thank members of the Ray MacDonald, John Minna, and Francois Guillemot Labs.

Finally, I would like to give special thanks to all the past and present members of the Jane Johnson lab. I was blessed to be surrounded by so many gifted scientists that were willing to pass their knowledge on to me. Without their training and helpful discussion I would not have succeeded.

TABLE OF CONTENTS

PRIOR PUBLICATIONS

Meredith DM*, **Borromeo MD***, Deering TG, Casey B, Savage TK, Mayer PR, Hoang C, Tung KC, Kumar M, Shen C, Swift GH, Macdonald RJ, Johnson JE. 2013 Program specificity for Ptf1a in Pancreas versus Neural Tube Development correlates with distinct collaborating cofactors and chromatin accessibility. Mol Cell Biol. PMCID: PMC23754747
*authors contributed equally

Chang JC, Meredith DM, Mayer PR, **Borromeo MD**, Lai HC, Ou YH, Johnson JE. 2013 Prdm13 mediates the balance of inhibitory and excitatory neurons in somatosensory circuits. Dev Cell. 25:182-95. PMCID: PMC3644180

Henke RM, Meredith DM, **Borromeo MD**, Savage TK, Johnson JE. 2009 Ascl1 and Neurog2 form novel complexes and regulate Delta-like3 (Dll3) expression in the neural tube. Dev. Biol. 328(2):529-40. PMCID: PMC2698949

Adkins DL, Campos P, Quach D, **Borromeo M**, Schallert K, Jones TA. 2006 Epidural cortical stimulation enhances motor function after sensorimotor cortical infarcts in rats. Exp Neurol. 200(2):356-70. PMCID: 16678818

LIST OF FIGURES

LIST OF APPENDICES

LIST OF ABBREVIATIONS

Ascl1	achaete-scute like 1
Atoh1	atonal homolog 1
Bcl2	B-cell CLL/lymphoma 2
bHLH	basic helix-loop-helix
Brn2a	brain-specific homeobox/POU domain protein 2A
ChIP	chromatin immunoprecipitation
CNS	central nervous system
dI	dorsal interneuron
dIL^{A}	dorsal interneuron late A
dIL^{B}	dorsal interneuron late B
Dll1	deltalike1
Dll1	deltalike3
Dlx	distal-less
DNA	deoxyribonucleic acid
E	embryonic day
FACS	fluorescence-activated cell sorting
Fox	forkhead box
Gbx	Gastrulation brain homeobox
GABA	γ-Aminobutyric acid
GAD	glutamic acid decarboxylase
GFP	green fluorescent protein
Gsx	GS homeobox
HD	homeodomain
IP	immunoprecipitation
Isl	insulin gene enhancer protein
Kirrel	Kin of irregular chiasm-like protein
Lbx1	Ladybird homeobox 1
Lhx	LIM/homeobox protein

LIM	Lin11, Isl-1 and Mec-3
Lmx1b	LIM homeobox transcription factor 1-beta
Mash1	mammalian homolog of achaete-scute complex 1
mRNA	messenger Ribo Nucleic Acid
MZ	mantle zone
NeuroD	neurogenic differentiation factor 1
Neurog1/2	neurogenin1/2
Nphs	Nephrin
Olig	oligodendrocyte transcription factor
Pax2	paired box gene
PBS	phosphate buffered saline
PNS	peripheral nervous system
POU	Pituitary specific, Octamer, Unc transcription factor
Prdm13	PRDI-BF1 and RIZ homology domain containing 13I
Ptf1a	pancreatic transcription factor 1a
PTF1-J	trimeric complex of Ptf1a, E-protein and Rbpj
Ret	ret proto-oncogene
Rbpj	recombining binding protein suppressor of hairless
SCLC	small cell lung carcinoma
Seq	sequencing
Sox	SRY-box containing gene
Tlx	T-cell leukemia homeobox protein
vGlut2	vesicular glutamate transporters 2
VZ	ventricular zone
ZF	zinc finger

LIST OF KEYWORDS

Keyword 1; cell fate specification

Keyword 2; neuronal specification

Keyword 3; neuronal development

Keyword 4; dorsal spinal cord development

Keyword 5; bHLH transcription factors

Keyword 6; inhibitory and excitatory neurons

Keyword 7; Pancreas Development

Keyword 8; Small Cell Lung Carcinoma

Keyword 9; Ascl1

Keyword 10; Ptf1a

Keyword 11; Neurod1

Keyword 12; Rbpj

Keyword 13; ChIP-Seq

Keyword 14; Rna-Seq

CHAPTER ONE

INTRODUCTION

The genomic DNA for any multicellular organism encodes the information necessary to specify each cell-type in its body. During development, the proper differentiation of various cell types requires the precise spatial and temporal expression of specific genes. Essential to gene regulation are transcription factors that work together to orchestrate the events necessary to activate or repress the correct gene programs. In the following sections I will describe some key principles of tissue specific gene regulation.

Tissue-specific gene regulation through cis-regulatory elements

Decades of research have established that tissue-specific transcription factors regulate gene targets by binding to short specific DNA sequences, normally located in non-coding regions of DNA. DNA regulatory elements that contain the necessary transcription factor binding sites, known as cis-regulatory elements or enhancers, have characteristics of being highly conserved through evolution and are located distal (up to hundreds of Kb away) from a gene's transcriptional start site (TSS) (for review see Visel et al., 2009b). Moreover, enhancers can activate transcription independent of their location, distance or orientation with respect to the TSS of a gene (for review see Ong and Corces, 2011). Because of this latter property, researchers for the past two decades have been able to take cis-regulatory elements and place them into reporter constructs to determine what tissue the enhancer is

transcriptionally active in and what DNA sequences are required for activity (for review see Farnham, 2009).

Before the recent advent of genome-wide binding assays, identifying functional cis-regulatory regions was difficult and yielded informative, but limited results. This is due in part to the extensive amount of time it takes researchers to find enhancer regions around a single gene of interest. Scientists relied on testing random regions around TSS or on bioinformatic approaches to find candidates of highly evolutionary conserved DNA regions around a gene of interest. As a result, large kilobase (Kb) segments of DNA, classified as candidate regions, would be tested in transcriptional reporter assays; followed by several experiments that generated DNA deletions or mutations to the candidate regions to narrow down the sequences required for transcriptional activity. This approach is tedious, but has been used extensively to identify important regulatory regions. For example, Meredith and colleagues identified several highly conserved regions of DNA that spanned 15.6 kb upstream of the gene, *Ptf1a* (Meredith et al., 2009, Masui et al., 2008). The entire 15.6 kb region was sufficient to activate a lacZ reporter in the proper domains of neural tube of which Ptf1a is expressed. Ultimately, the 15.6 kb region was narrowed down to a smaller 2.3 kb region that was still able to recapitulate Ptf1a's expression pattern. This work demonstrated that a non-coding region of DNA, located far away from the genes promoter is capable driving a tissue-specific pattern in the developing nervous system. The limitation of this approach illustrated by this example is that only one regulatory element was identified for a single gene. Thus, finding the entire repertoire tissue-specific enhancers used to regulate this single gene is a huge task.

The significant advancements of Chromatin Immunopreciptation (ChIP) and next-generation sequencing over the past five years has made it possible to identify tissue-specific enhancers more readily. Scientists are now able to biochemically purify a transcription factor and the DNA the regions it occupies to identify where in the entire genome a transcription factor binds. In a hallmark paper, Johnson *et al.* were the first to demonstrate *in* vivo, the genome-wide binding for a sequence-specific factor, NRSF (Johnson et al., 2007). Following this study, Visel and colleagues adapted the use of the technology to identify relatively small DNA cis-regulatory elements bound by the general transcriptional activator P300, a non-DNA binding protein. P300 is recruited by transcription factor complexes bound to enhancers to activate transcription (for review see Visel et al., 2009b). By testing and comparing the DNA enhancers bound specifically by P300 in the developing limb versus the brain, Visel and colleagues demonstrated the tissue-specific activity of hundreds of DNA elements *in vivo* (Visel et al., 2009a). The P300 experiments clearly demonstrated that many non-coding DNA elements can regulate tissue-specific expression; however, the P300 and classical approaches of finding enhancers still leaves the question of what tissue-specific transcription factors bind to these cis-regulatory elements.

Epigenetic regulation of tissue specific gene expression

There are multiple mechanisms for tissue specific gene regulation, including the chromatin accessibility of the cis-regulatory elements. Genomic DNA is packaged around nucleosomes which can create a physical obstacle for transcription factors to bind DNA targets. The process of determining what DNA elements are nucleosome free or “open chromatin” is

highly regulated by the cell. For several years now, it has been appreciated that the chromatin landscape for pluripotent cells changes as it becomes more restricted to a specific cell-type (for review see Ong and Corces, 2011); and once again, lineage-specific transcription factors play a central role determining the accessibility (Magnani et al., 2011, Zaret and Carroll, 2011). Some lineage-specific sequence specific transcription factors have been classified as "pioneering factors" that are capable of binding closed chromatin and interacting with protein complexes to alter the accessibility. (Zaret and Carroll, 2011). The role of pioneering factors was demonstrated during the process of macrophage and B-cell differentiation (Heinz et al., 2010). The pioneering factor PU.1 was demonstrated to bind an area of closed chromatin and shortly after, the nucleosome in that region was repositioned; thus, opening the chromatin region for other lineage determining factors to bind. Because PU.1 and other pioneering factors such Foxa1 and Foxa2 are sequence-specific DNA binding proteins (Heinz et al., 2010, Lee et al., 2005a), their DNA binding sites should be detected at the tissue specific enhancers they establish. Thus, as more lineage-specific enhancers or open chromatin regions become identified in different tissues, the lineage determining factors that give rise to these tissues could potentially be identified by their DNA binding sequence.

Sequence specific binding of transcription factors

Transcription factors are key players that interpret the information encoded in DNA to give rise to multicellular organism. Thus, the interaction between a transcription factor and their DNA binding site are essential for the tissue-specific regulation of gene expression. It has long been appreciated that members of the same transcription factor family can bind similar

DNA sequences, making it unclear how a transcription factor can achieve tissue-specific function. Large scale studies using *in vitro* binding assays have shown that transcription factors of the same family do indeed bind a similar degenerate sequence, but each member of that family may prefer a specific variation of that DNA sequence (Badis et al., 2009). Thus, subtle differences in a DNA binding site may be important in gene regulation by recruiting a specific member of a transcription factor family that is expressed among other members in the same cell. To determine if small variations in DNA binding affect tissue specific regulation, binding sites of structurally related factors need to be identified *in vivo.*

Examination of functional enhancers show the presence of several transcription factor binding sites (for review see Farnham, 2009). This led to the hypothesis that open chromatin regions harbor DNA binding sites for several prospective transcription factors, and through the interactions with DNA and other transcription factors, stabilization of the DNA region will occur, resulting in the recruitment of the necessary components to activate transcription (for review see Heinz and Glass, 2012). This hypothesis could explain some tissue-specific expression patterns, given that different cell-types express different classes of transcription factors. Thus, a single factor expressed in multiple cell-types can regulate different genes by interacting with different tissue specific co-factors. These transcription factor interactions are likely to be uncovered with the identification of additional binding sites for a specific factor in multiple tissues.

Class II bHLH transcription factors bind a degenerate E-box sequence

The focus of this thesis is to elucidate how structurally related transcription factors with highly similar DNA binding motifs are able to achieve specificity of binding and activation of select target genes. While these questions are likely to apply to most tissue-specific transcription factors, I decided examine members of the basic helix-loop-helix (bHLH) transcription factor family. The bHLH family comprises a large group of transcriptional regulators that function in the development of several organ systems (Massari and Murre, 2000). The bHLH family can be subdivided into seven classes based on differences in structure, tissue expression, DNA-binding, and protein dimerization (Murre et al., 1994). Here, we focus our on work on the Class II bHLH factors, particularly Ascl1 (previously Mash1) and Ptf1a, which display a tissue-restricted expression pattern and function as heterodimers with the more ubiquitously expressed class I bHLH factors known as E-proteins (Fig 1.1). Ptf1a and Ascl1 give rise to multiple tissues in the developing embryo and a detailed description of their function is described below. Other examples of Class II bHLH factors include the neural-specific factors Atoh1, and Neurogenin1, and the muscle-specific factors MyoD and myogenin.

The bHLH heterocomplex binds to a canonical DNA sequence known as the E-box (CANNTG) (Massari and Murre, 2000). For decades, researchers have tried to understand how bHLH factors can select and regulate their specific gene targets given that they bind the same degenerate E-box sequence *in vitro* (for review see Bertrand et al., 2002). Part of the explanation for their specific target selection could be that each bHLH binds to a more specific E-box sequence *in vivo*, requires additional collaborating factors to select their

target, or chromatin accessibility regulates which E-boxes they can bind. These questions remain largely unanswered to due to lack of direct targets of the bHLH factors *in vivo.*

Tissue-specific bHLH and HD factors specify somatosensory neurons in the dorsal spinal cord

The developing neural tube is an excellent *in vivo* model for studying the function of multiple tissue-specific transcription factors, such as proneural bHLH, homeodomain (HD) and other families of factors (for review see Alaynick et al., 2011). The neural tube is a simple structure that gives rise to different neuronal populations, and the cell fate specification of these populations relies on combinations of transcription factors to activate or repress specific neurogenic programs. The bHLH and HD families of transcription factors are particularly important players in generating the correct number and sub-types of neurons in the dorsal spinal cord (Fig 1. 1) (Gowan et al., 2001, Gross et al., 2002, Glasgow et al., 2005, Helms et al., 2005, Wildner et al., 2006, Mizuguchi et al., 2006, Cheng et al., 2004, Cheng et al., 2005, Muller et al., 2005). Throughout the developing nervous system, bHLH transcription factors have a prominent role in establishing progenitor domains and coordinating the cellular events for the transition from a progenitor cell to a differentiated neuron (Castro et al., 2006, Gohlke et al., 2008, Bertrand et al., 2002, Nakada et al., 2004). In the dorsal neural tube, multiple progenitor domains can be identified by the neural bHLH factors Atoh1 (previously Math1), Neurog1 (previously Ngn1), Ascl1 (previously Mash1), and Ptf1a (Fig. 1.2) (Helms et al., 2005, Gowan et al., 2001, Glasgow et al., 2005). The neurons that arise from these progenitor domains are defined by the stage they become

postmitotic and the combination of HD factors they express (Fig. 1.2) (for reviews see Helms and Johnson, 2003, Alaynick et al., 2011, Hori and Hoshino, 2012). Many of the HD factors have been shown to be essential for continued specification of the neuronal subtypes such as the dorsal horn glutamatergic and GABAergic neurons (Cheng et al., 2004, Cheng et al., 2005, Pillai et al., 2007, Gross et al., 2002, Batista and Lewis, 2008, Huang et al., 2008). Given the temporal and genetic relationship between bHLH and HD factors, HD factors are proposed to be direct transcriptional targets of the bHLH factors in the dorsal spinal cord. The spatial restrictions of the bHLH factors and the fact that they give rise to specific populations, highlights that each of proneural bHLH factor has a specific function. It is unclear how they achieve this specificity given that they all can bind the same degenerate E-box sequence. Moreover, the bHLH factors Ascl1 and Ptf1a are responsible for giving rise to distinct neuronal populations, and genetic analysis shows that they have opposing functions (described below), but these factors are co-expressed in a subset of neural progenitors in the dorsal spinal cord (Fig 1.2). It is unknown how Ascl1 and Ptf1a each achieve their specificity in binding and selection of target genes, while repressing the other factors function.

Ascl1 and Ptf1a have opposing roles in neuronal subtype specification in the dorsal spinal cord

The focus of this thesis will be on two bHLH factors Ascl1 and Ptf1a. As described above it is well established that bHLH factors such as Ascl1 and Ptf1a influence the expression of several HD factors that mark specific neuronal populations and function in the maturation of those neurons in the dorsal neural tube (Glasgow et al., 2005, Gowan et al., 2001, Helms et

al., 2005, Batista and Lewis, 2008, Cheng et al., 2004, Huang et al., 2008, Pillai et al., 2007, Cheng et al., 2005, Gross et al., 2002, Muller et al., 2002). To summarize prior studies that revealed the genetic network regulating neuronal subtype specification in the E11.5 dorsal neural tube, and to introduce the populations used in the current study for the genome-wide binding and gene expression experiments, I show here the expression patterns of the bHLH factors Ascl1 and Ptf1a, and some of the HD factors that specify the excitatory (Tlx1/3) and inhibitory (Pax2, Lhx1/5) neurons in wild type and mutants of *Ascl1* and *Ptf1a* (Fig. 1.3). Ascl1 and Ptf1a are largely restricted to the ventricular zone of the dorsal neural tube although Ascl1 is also present in some ventrally located cells (Fig. 1.3A, D). Their expression is transient, seen as the lack of expression in regions lateral to the ventricular zone where cells are postmitotic and have initiated expression of neuronal specific genes. Ascl1 is in both mitotically active cells and cells that have just exited the cell cycle, whereas Ptf1a co-localizes with Ascl1 in a subset of the post-mitotic cells (Hori et al., 2008, Glasgow et al., 2005). Thus, Ascl1 is present at an earlier stage than Ptf1a in these dorsal neural tube progenitor cells.

The opposing functions of Ascl1 and Ptf1a in specifying neuronal subtype in the dorsal neural tube is illustrated by the changes in HD factor expression in *Ascl1* and *Ptf1a* mutant mouse embryos. The excitatory populations in the dorsal neural tube (dI3, dI5, and dIL^{B}), marked by the HD factors Tlx1 and Tlx3 (Fig. 1.3G), are dramatically reduced at E11.5 in the *Ascl1* null (Fig. 1.3H), whereas in the *Ptf1a* null they are markedly increased (Fig. 1.3I). In contrast, inhibitory populations (dI4 and dIL^{A}) marked by Pax2, Lhx1 and Lhx5 (Fig. 1.3J,M), are lost in the *Ptf1a* null (Fig. 1.3L,O). These inhibitory markers are also

diminished in the *Ascl1* null (Fig. 1.3K,N), a phenotype likely secondary to the dependence of *Ptf1a* expression on Ascl1 at this stage (Fig. 1.2E) (Mizuguchi et al., 2006). There are additional complexities and feedback loops not shown here in the transcriptional network including repressive interactions of Tlx1/3 on Pax2 levels (Cheng et al., 2005), and involvement of other transcription factors such as Prdm13 (Chang et al., 2013) (see diagram, Fig. 1.3P). In summary, Ascl1 and Ptf1a are at the head of a transcription factor network that is critical in generating GABAergic and glutamatergic neuronal populations in the dorsal spinal cord.

Ptf1a and Ascl1 give rise to multiple tissue types outside the neural tube

One of the fundamental questions of developmental biology is how a single transcription factor can give rise to different tissues. The challenge of answering this question is that it has been difficult to identify the genome-wide transcriptional targets for any transcription factor. With the advancement of ChIP and next-generation sequencing, identifying transcriptional targets has become significantly easier. Ascl1 and Ptf1a are examples of transcription factors that function in multiple tissue types. A part of this thesis will focus on identifying the transcriptional targets of these two factors in different tissues. A detailed description of their function and the tissue types they give rise to are described below.

Ptf1a specifies GABAergic neurons in multiple regions of the nervous system and is also require for pancreas development

Beyond the dorsal neural tube, Ptf1a can be found in specific regions of the developing cerebellum, hindbrain, and retina. Similar to Ptf1a's function in the dorsal neural tube, Ptf1a is required for the proper specification of the inhibitory neurons in these additional populations. (Glasgow et al., 2005, Hoshino et al., 2005, Fujitani et al., 2006, Pascual et al., 2007). The absence of Ptf1a in the dorsal neural tube not only results in failure to specify inhibitory neurons but also creates a significant increase in the excitatory populations. A similar phenotype can be found in the cerebellum and retina. Specifically in retinal development, the loss of Ptf1a results in the reduction of both amacrine and horizontal inhibitory cells, and a substantial increase in the number of supernumerary excitatory retinal ganglion cells (Fujitani et al., 2006). In the cerebellum, the inhibitory Purkinje cells and interneurons fail to form, and an increase in the number precursor cells with an excitatory identity is observed (Pascual et al., 2007).

Ptf1a is not only important for the developing nervous system but also for formation and maturation of the pancreas (Fig 1.4). Loss-of-function studies have demonstrated the necessity of Ptf1a for the commitment to the pancreatic lineage from endoderm and the specification and maintenance of the acinar pancreas (Krapp et al., 1998). In the absence of Ptf1a the pancreas fails to form, and Ptf1a-lineage cells instead adopt either a duodenal or biliary identity (Krapp et al., 1998, Masui et al., 2007, Kawaguchi et al., 2002). It is unknown if the direct transcriptional targets of Ptf1a are the same in these different tissues. If the targets are different, how does Ptf1a regulate these different genes?

Ptf1a forms a trimeric complex with E-protein and Rbpj to activate transcription

Like other class II bHLH factors, Ptf1a forms a heterodimer with E-proteins; however, it requires a third component, Rbpj, to form a trimeric complex (PTF1) that activates transcription of target genes (Fig. 1.1) (Beres et al., 2006, Masui et al., 2007, Hori et al., 2008). In both the pancreas and nervous system, Rbpj is required for the proper Ptf1a function (Masui et al., 2007). For example, in the nervous system, the absence of Rbpj or the disruption of the physical interaction between Rbpj and Ptf1a results in the incorrect specification of the GABAergic neurons (Hori et al., 2008). During acinar pancreatic development, the PTF1 complex activates the expression of the *Rbpjl*, which then replaces Rbpj in the PTF1 complex to complete acinar maturation and maintain exocrine gene expression into adulthood (Masui et al., 2010, Masui et al., 2007). Irrespective of whether Rbpj or Rbpjl is present in the PTF1 complex, in the small number of downstream targets identified, the complex binds a bipartite motif containing an E-box (the bHLH consensus site) and a TC-box (the Rbpj/Rbpjl consensus site) separated by one, two, or three helical turns of DNA (Cockell et al., 1989, Beres et al., 2006, Masui et al., 2008, Henke et al., 2009b, Meredith et al., 2009, Thompson et al., 2012). However, how Ptf1a in the PTF1 complex regulates different target genes in developing pancreas versus the neural tube remains unknown.

Direct Transcriptional Targets of Ptf1a

In the early stages of this thesis research, eight genes were known to be direct Ptf1a targets during development (Fig. 1.1). During neurogenesis, Ptf1a directly regulates *Neurog2, Ptf1a, Kirrel2 and Nphs1* (Henke et al., 2009b, Nishida et al., 2010, Meredith et al., 2009). As for the pancreas, *Rbpjl, Ptf1a, Pdx1,* and the notch ligand *Dll1*, are known to be PTF1 targets during development (Masui et al., 2010, Ahnfelt-Ronne et al., 2012, Masui et al., 2007, Masui et al., 2008). Even with this limited number Ptf1a targets identified, it seemed that Ptf1a might regulate distinct genes in each tissue. However, a genome-wide approach to identify Ptf1a in these disparate tissues would clarify this issue and provide further insights into the how this transcription factor functions in these two distinct lineages.

Ascl1 has neurogenic, oncogenic and gliogenic functions in the nervous system

Functional analysis of Ascl1 has demonstrated that Ascl1 is required for the generation of specific neuronal populations throughout the central and peripheral nervous systems. Mice deficient for Ascl1 have defects in the development of olfactory epithelium, telencephalon, hindbrain, retina, spinal cord, and ganglia of the autonomic nervous system (Akagi et al., 2004, Blaugrund et al., 1996, Casarosa et al., 1999, Cau et al., 1997, Guillemot et al., 1993, Helms et al., 2005, Hirsch et al., 1998, Pattyn et al., 2004). Ectopic expression of Ascl1 in the developing neural tube induced neuronal differentiation and increased specification of discrete neuronal cell-type populations (dI3, dI5, and dILa) (Helms et al., 2005, Mizuguchi et al., 2006, Nakada et al., 2004). Ascl1 also has an essential role in oligodendrogenesis; mice lacking Ascl1 have a deficiency in the differentiation of oligodendrocytes in the spinal cord,

telencephalon, cerebellum, and olfactory bulb (Sugimori et al., 2008, Parras et al., 2007, Grimaldi et al., 2009, Parras et al., 2004). Accumulated work over the past two decades has identified Ascl1 as an important transcription factor for the differentiation and specification of multifarious neural cells in the nervous system; however, the direct transcriptional targets of Ascl1 that lead to neural cell diversity still remain largely unknown.

In addition, neuroendocrine cells located in the thyroid and lung (Fig. 1.4) are absent in Ascl1 deficient mice (Borges et al., 1997, Lanigan et al., 1998). Interestingly, Ascl1 expression is aberrantly up-regulated in small cell lung carcinoma (SCLC), which is thought to originate from lung neuroendocrine cells (Park et al., 2011). In the clinic, Ascl1 is used as a biomarker for SCLC and other neuroendocrine tumors. In pre-clinical models, ASCL1 is found to be highly expressed in both human derived SCLC cell lines (Jiang et al., 2009, Borges et al., 1997, Ball et al., 1993) and mouse models of SCLC (Meuwissen et al., 2003, Schaffer et al., 2010). ASCL1 is necessary for tumor-initiating capacity in SCLC cell lines (Jiang et al., 2009, Osada et al., 2005). Similar to the situation of Ptf1a in the neural tube and pancreas, it is not clear how many of Ascl1 transcriptional targets are the same in neural and neuroendocrine tissues. Moreover, by comparing what genes are regulated by Ascl1 in normal neural and the abnormal tissue might give insight in to mechanisms of how Ascl1 identifies targets in different tissues. Most importantly, identifying the targets of Ascl1 in SCLC may uncover genes required for growth and survival of the tumor, which could be later targeted pharmacologically.

Transcriptional Targets of Ascl1

Several direct targets of Ascl1 have been identified in different tissues (Fig 1.1). ChIP-chip analysis, which only searches for binding events near a gene's transcriptional start site, for Ascl1 in the developing telencephalon revealed hundreds of direct targets and together these targets show that Ascl1 has multiple functions such as regulating genes involved in cell-cycle exit, neuronal migration, and differentiation. Moreover, in a subset of cells found in the telencephalon, Ascl1 regulates genes that promote progenitor proliferation (Castro et al., 2011). In SCLC tumor cells, Ascl1 targets the cancer stem cell markers *CD133* and *ALDH1A1* which are essential for tumorigenicity in some SCLC. Neither *CD133* nor *ALDH1A1* are targets of ASCL1 in telencephalon, highlighting the cell context dependency of Ascl1 function. However, some targets do appear to be the same across different tissues, in developing telencephalon, neural tube, and SCLC tumors, the notch ligand *Dll3* appears to be a common target of Ascl1 (Castro et al., 2006, Castro et al., 2011, Henke et al., 2009a, Ball et al., 1993). How much of Ascl1's function is the same or different in each tissues type is unknown. Thus, identification of the full repertoire of Ascl1 targets in each tissue type will provide insight into how much of Ascl1 function is different in each cellular environment.

THESIS RATIONALE AND GOALS

For several decades, investigators have attempted to elucidate how members of transcription factor families with highly similar DNA binding motifs are able to achieve specificity of binding and activation of select target genes. Equally puzzling is how a factor may be utilized in multiple organ systems or in a disease state to direct exceedingly disparate genetic

programs. These questions largely remain unanswered due to the lack on known transcription transcriptional targets for any given transcription factor. The work presented here focuses on the proneural bHLH transcription factor family, particularly Ascl1 and Ptf1a. In the dorsal neural tube, these two bHLH factors are co-expressed in a subset of neural progenitors, although they have distinct activity in neuronal specification, they function in a similar cellular environment. Thus, Chapter 2 addresses the basic concepts in transcriptional control of cell fate determination by probing how these related transcription factors select and regulate gene expression *in vivo* in the developing mouse neural tube. Chapter 3 and 4 address the fundamental question of how a transcription factor can regulate different sets of genes in different tissues. Specifically in Chapter 3, I identify and compare Ptf1a-chromatin interactions in the developing neural tube versus the pancreas, and in Chapter 4, I identify Ascl1 target genes in the neuroendocrine lineage tissue of small cell lung carcinoma.

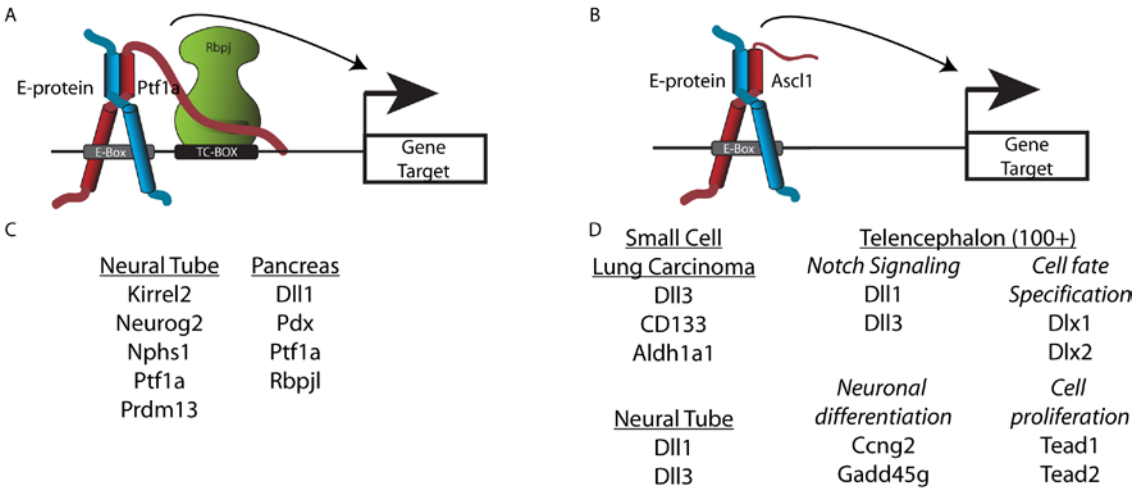


Figure 1.1 Direct transcriptional targets of Ptf1a and Ascl1. (A-B) Class II bHLH factors Ptf1a and Ascl1 form a heterodimer complex with class I E-proteins, and bind the E-box sequence. (A) The Ptf1a forms a trimeric complex with Rbpj (PTF1) to activate its genes targets (Beres et al., 2006, Hori et al., 2008). (C-D) Direct targets of Ptf1a and Ascl1 in different tissues (Henke et al., 2009b, Henke et al., 2009a, Jiang et al., 2009, Castro et al., 2011, Castro et al., 2006, Masui et al., 2010, Ahnfelt-Ronne et al., 2012, Wiebe et al., 2007, Nishida et al., 2010, Meredith et al., 2009)

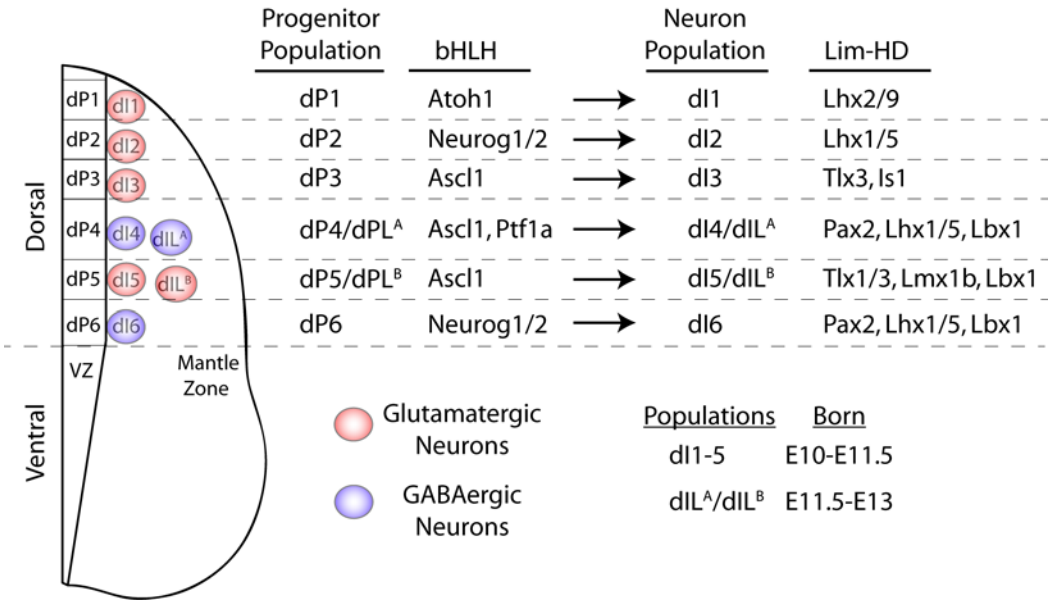


Figure 1.2 Transcription factor network specifying the excitatory and inhibitory neurons in the dorsal spinal cord. The specification of the dorsal spinal cord neurons directed by the bHLH and HD factors (for review see ((Hori and Hoshino, 2012)). The bHLH transcription factors are patterned in the ventricular zone (VZ) and the Lim-HD transcription factors label different populations of interneurons in the mantle zone (MZ). Two temporal waves of neurogenesis occur, an early wave starting at E10 creating dI1-6 populations and later wave starting at E11.5 that gives rise to dIL^{A}/dIL^{B} populations. dIL^{A} and dIL^{B} share markers with dI4 and dI5, respectively. The dI4/dILA neurons will further develop to inhibitory GABAergic neurons (blue), and the dI1-3 and dI5/dILB will become excitatory glutamatergic neurons (red).

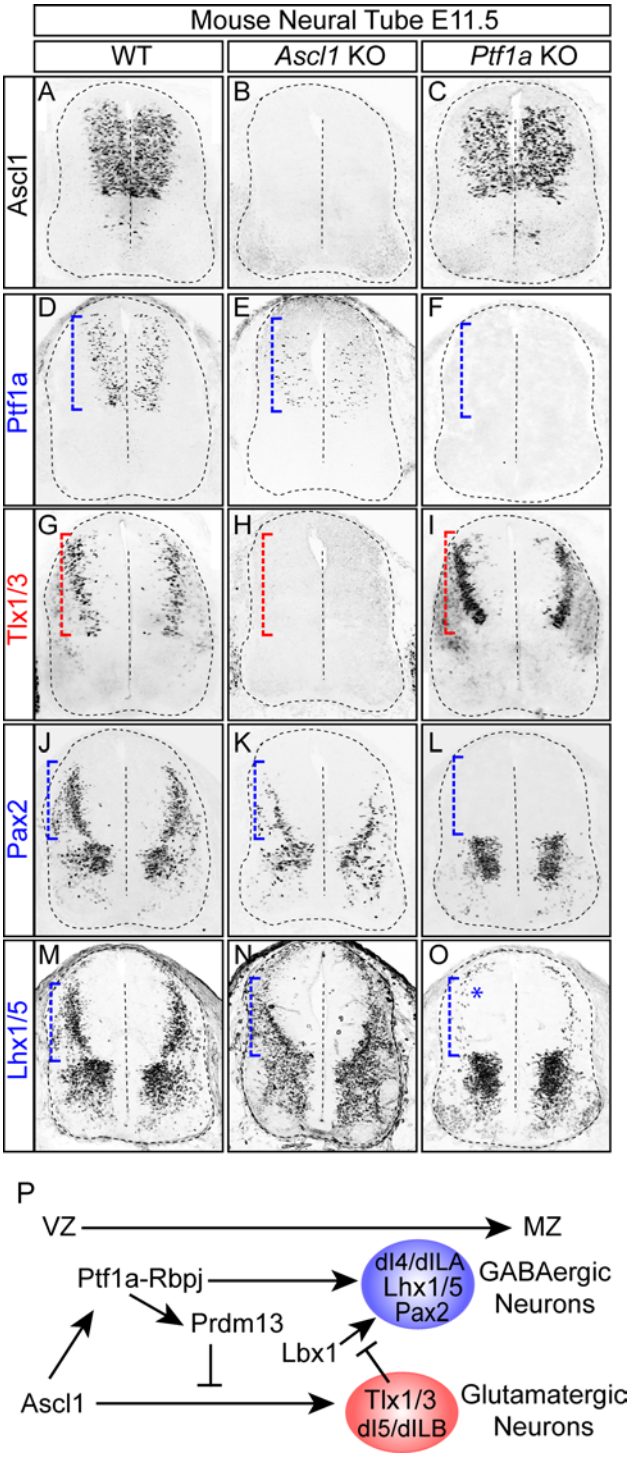


Figure 1.3 bHLH factors Ascl1 and Ptf1a have opposite actions in neuronal subtype specification in the developing spinal cord. (A-O) Immunofluorescence for Ascl1, Ptf1a, Tlx1/3, Pax2, and Lhx1/5 on cross sections of mouse E11.5 neural tube from wild type and *Ascl1 or Ptf1a* null embryos (Kawaguchi et al., 2002, Guillemot et al., 1993). Dashed brackets indicate the dorsal neural tube where Ascl1 and Ptf1a are expressed and phenotypes are detected in the mutants. Red indicates the transcription factor is a marker of excitatory neurons, and blue indicates those marking inhibitory neurons or their precursors. Blue asterisk in (O) indicates a Ptf1a-independent Lhx1/5+ population (dI2) that is unaffected in the *Ptf1a* null. (P) A diagram summarizing the known transcription factor network involved in generating excitatory and inhibitory populations in the dorsal spinal cord. Data generated by Joshua Chang of the Jane Johnson Lab.

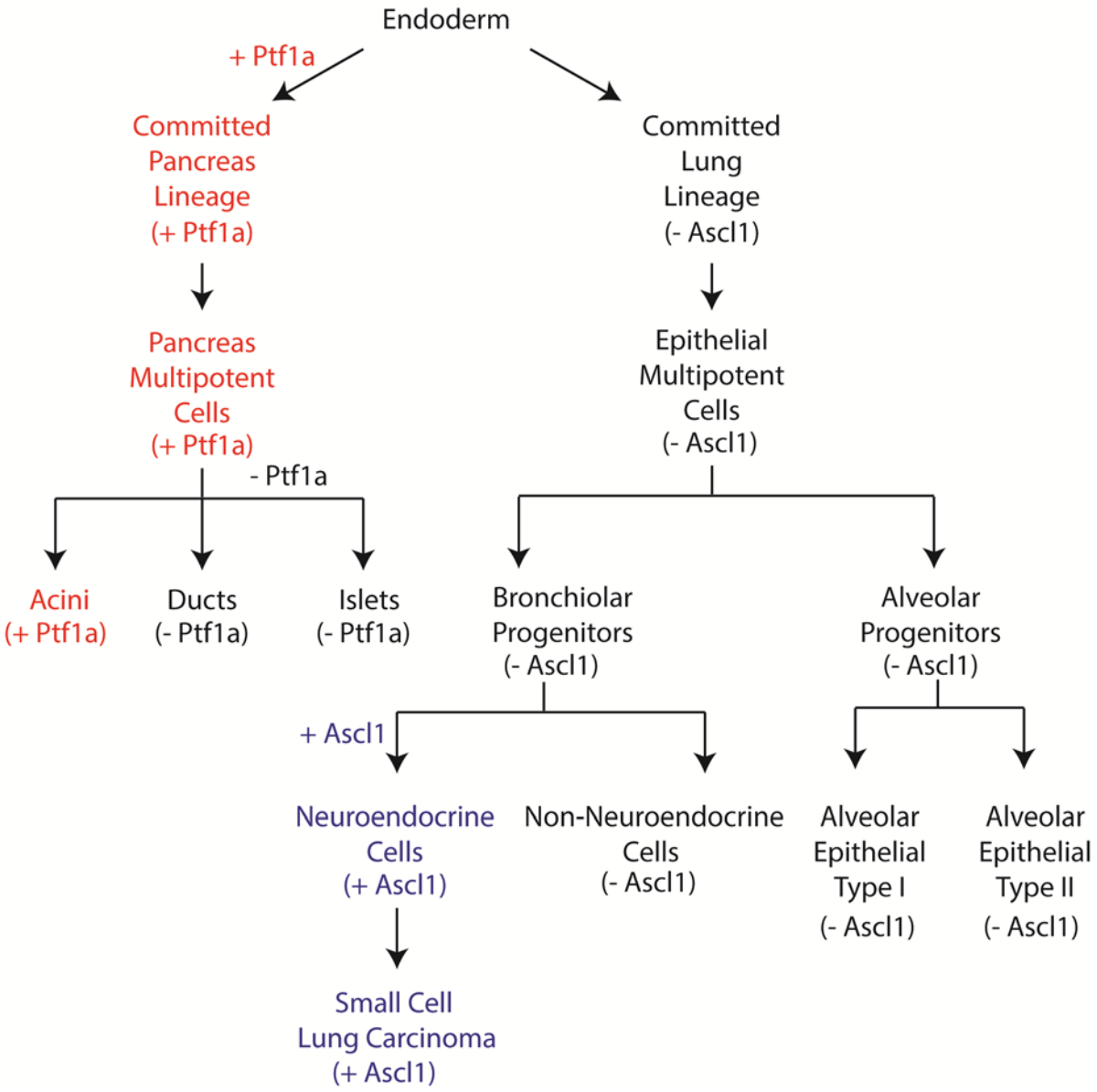


Figure 1.4.Ptf1a and Ascl1 function in discrete tissues outside the CNS. Schematic diagram shows where Ptf1a and Ascl1 are found during the formation of the pancreas and lung. Ptf1a is critical for determining the pancreatic lineage, and the specification and maturation of the Acinar cells (for review see MacDonald et al., 2010). Ascl1 is required for the formation of pulmonary neuroendocrine cells and survival and proliferation of SCLC (for review see Rock and Hogan, 2011). Note, the cells that expressed Ptf1a (Red) or Ascl1 (Blue) are marked.

CHAPTER TWO

Transcription Factor Network Specifying Inhibitory versus Excitatory Neurons in the Dorsal Spinal Cord

The work presented in this chapter was performed in collaboration with David M. Meredith, Joshua C. Chang, and Kuang C. Tung of the Jane Johnson lab, Axel Visel of the Len Pennacchio Lab and Diogo S. Castro of the Francois Guillemot lab.

INTRODUCTION

The neurons within the dorsal spinal cord provide the initial integration for somatosensory information originating in the periphery. These neurons relay the sensory information to local spinal cord neurons and higher brain centers to modulate and coordinate the appropriate physiological response to environmental stimuli (for reviews see Ross, 2011, Liu and Ma, 2011). The proper processing of somatosensory information requires the correct balance of excitatory and inhibitory neurons within the dorsal spinal cord; disruption of this balance may cause neurological disorders such as hyperalgesia and allodynia (for reviews see Tavares and Lima, 2007, Fitzgerald, 2005). Revealing the genetic programs that give rise to these different classes of neurons will provide insight into these disorders as well as address

fundamental concepts in transcriptional control of cell fate determination particularly in neuronal subtype specification.

The neural basic helix-loop-helix transcription factors, Ascl1 and Ptf1a, are essential for generating the correct number and sub-type of neurons in multiple regions of the nervous system. In the dorsal neural tube, Ptf1a and Ascl1 are co-expressed in a subset of neural progenitors, thus, although they have distinct activity in neuronal specification, they function in a similar cellular environment. In this chapter, I address basic concepts in transcriptional control of cell fate determination by investigating how two bHLH transcription factors select and regulate gene expression *in vivo*.

Over a decade of research has gone into exploring Ascl1 and Ptf1a's function in neural development; however, only few direct targets of Ptf1a and Ascl1 have been identified in the developing spinal cord. Therefore, my goal was to identify the direct downstream targets of Ptf1a and Ascl1 that are involved in specifying inhibitory and excitatory neurons in the dorsal spinal cord. In Chapter 1, I highlighted the point that the HD factors are genetically downstream of the bHLH factors, thus I hypothesized that they are direct transcriptional targets of Ascl1 and Ptf1a in the dorsal spinal cord. In the following sections, I provide evidence supporting this hypothesis as well as identify a cohort of genes directly regulated by Ascl1 and Ptf1a that allow these factors to direct neuronal differentiation and neuronal subtype specification.

Genes directly regulated by Ascl1 and Ptf1a were identified by performing ChIP-Seq and RNA-Seq experiments using mouse neural tube tissue. Ascl1 activates a glutamatergic

specification and maturation program by directly regulating genes encoding multiple HD factors including *Tlx1, Tlx3, Lmx1b, Gsx2, Isl1 and Pou4f1 (Brn3a)*. In contrast, Ptf1a directly activates GABAergic specification and maturation programs that include the genes for HD factors such as *Lhx1, Lhx5, Gbx1, Gbx2, and Pax2*. Hundreds of additional targets for Ascl1 consistent with its essential role in regulating several aspects of neurogenesis, while targets for Ptf1a illustrates its role in specifically regulating genes encoding components of GABA biosynthesis, GABA and glycine transport pathways, and genes involved in inhibitory synaptic function and formation. These distinct targets for Ascl1 and Ptf1a reflect differences in the timing of their function with Ascl1 being expressed earlier during neurogenesis than Ptf1a. The specificity of Ascl1 and Ptf1a function results, at least in part, through differences in E-box sequence preferences and co-factors. Notably, Ptf1a antagonizes Ascl1 in neuronal subtype specification through direct and indirect repression of several Ascl1-activated targets.

MATERIALS AND METHODS

Mouse Strains

Ptf1a^{Cre} (*p48*^{Cre}), where the *Ptf1a* coding region is replaced by that coding for Cre recombinase was used as the *Ptf1a* null (Kawaguchi et al., 2002). *12.4Ptf1a::mCherry* transgenic mice, where a 12.4 kb regulatory sequence from 3' of the *Ptf1a* gene drives expression of mCherry (Meredith et al., 2009), were used for fluorescence activated cell sorting (FACS) of *Ptf1a* lineage cells from E11.5 wild-type or *Ptf1a* null neural tubes. *Ascl1*^{GFP} (*Ascl1*^{tm1Reed}/J), where the *Ascl1* coding region is replaced by that coding for GFP

(Leung et al., 2007) was used for (FACS) of *Ascl1* lineage cells from E11.5 control or *Ascl1* null (Guillemot et al., 1993) neural tubes. PCR genotyping was performed as previously described (Meredith et al., 2009, Glasgow et al., 2005, Kim et al., 2007). All procedures on animals follow NIH Guidelines and were approved by the UT Southwestern Institutional Animal Care and Use Committee.

Chromatin Immunoprecipitation (ChIP) and Sequencing Library Preparation

Detailed descriptions of Ptf1a, Rbpj (Meredith et al., 2013), and Ascl1(Castro et al., 2011) E12.5 NT ChIP protocols have previously been published. Descriptions of the sequencing library preparations for Ptf1a, Rbpj and control samples have also been previously published (Meredith et al., 2013); in short, Ascl1, Ptf1a and Rbpj with respective control libraries were made according to Illumina's ChIP-seq DNA sample prep protocol. Single-end Sequencing runs were performed on an Illumina GAIIx Sequencer for Ptf1a and Rbpj, and an Illumina GA Sequencer for Ascl1.

ChIP-Seq Samples, Peak Calling, Intersections, and Quantification

All ChIP-seq samples are from the E12.5 neural tube. Ptf1a and Rbpj ChIP-seq samples are currently available on the GEO database. For the Ptf1a ChIP-Seq sample (GSM1150324), we used the telencephalon Ptf1a ChIP-Seq sample (obtaining) as the control since the telencephalon is a neural tissue of similar developmental stage that does not express Ptf1a, thus controlling for non-specific binding of the antibody (Glasgow et al., 2005). Both Rbpj (GSM1150327) and Ascl1 (obtaining) ChIP-Seq samples were compared to the E12.5 neural

tube input (GSM1150340 and (obtaining), respectively). For Ascl1 and its control, sequence reads were obtained on multiple lanes and combined prior to alignment.

Sequence reads for each sample were mapped to the mm9 assembly of the mouse genome with Bowtie (Trapnell et al., 2009). Duplicate reads were removed, and the remaining unique reads were normalized to 10 million reads. Peak calling was performed by HOMER (Heinz et al., 2010) using an FDR cutoff of 0.001. We used an additional cutoff of a cumulative Poisson p-value of <0.0001 and required a 4-fold enrichment of normalized sequenced reads in the treatment sample over the control/input sample. We defined a common binding site between two samples when the peak summits of each sample were found within 150bp of each other.

Quantification the ChIP-Seq samples were performed by HOMER (annotatePeaks.pl –size 5000 –hist 10 -ghist) (Heinz et al., 2010). Normalized sequence reads around each peak were counted in 10bp bins, and the results were then loaded into Matlab® to generate the heatmap. Using the normalized data, the histogram plot was generated by first calculating the Ascl1 ChIP-Seq fragment size, and then extending the sequence reads to that estimated size (ChIP-Fragment Coverage). The fragments were then tabulated in 10bp windows around the peaks and normalized to the number peaks used in each group (i.e. Ptf1a peaks subdivided by E-box type).

mRNA Isolation and Sequencing Library Preparation

Individual mouse neural tubes from the *12.4kbPtf1a::mCherry;Ptf1a*$^{Cre/+}$, *12.4kbPtf1a::mCherry;Ptf1a*$^{Cre/Cre}$, *Ascl1*$^{GFP/+}$, or *Ascl1*$^{GFP/-}$ lines were dissected into

DMEM/F12 on ice and dissociated in 0.25% trypsin for 15 minutes at 37°C. Trypsin activity was quenched with 2% fetal bovine serum, and GFP/mCherry positive cells were purified from the resulting single cell suspension by fluorescence activated cell sorting. Cells from the same genotype were pooled (~1 million) and total RNA from the sorted cells were extracted and purified with Zymo's Mini RNA Isolation Kit. 1µg of total RNA was used for mRNA (polyA) isolation and sequencing library preparation using Illumina's mRNA-Seq kit. Two independent libraries were made for each genotype (except for *12.4kbPtf1a::mCherry;Ptf1a$^{Cre/+}$).*

RNA-Seq Read Alignment, Expression Level Estimation and Significance

Sequence reads were aligned to the mm9 build of the mouse genome using TopHat v2.0.9 (Trapnell et al., 2009). All Default settings were used with the following exceptions: -G option (instructs TopHat to initially map reads onto a supplied reference transcriptome) and --no-novel-juncs to ignore putative splice junctions occurring outside of known genes in the reference. If the sequence reads were 36bp long, then –segment-length 18 was used. If a biological replicate was available, then it was specified and used to build an expression level model determined by the FPKM method of Cuffdiff v2.1.1 (Trapnell et al., 2010, Trapnell et al., 2013). The options used were multiple read correction (-u) to better distribute reads mapping to multiple genomic locations and the bias correction (-b), to correct any sequence-specific bias introduced during the library preparation option. All other settings were left at default values. A gene was considered to be expressed if it had FPKM greater than 1. For a

gene to be called as differentially expressed, it required a q-value<0.05. Scatter plots and expression bar plots were created by cummerbund (Trapnell et al., 2012).

GO Classification and ChIP-Seq Peak Gene Annotation

Distance to gene and gene annotations for ChIP-Seq peaks were obtained using GREAT v1.82 (McLean et al., 2010). GREAT assigns a gene to a binding region if the region falls within 5 kb 5' or 1 kb 3' of the transcription start site (basal region) with a maximum extension of 1,000 kb in either direction. If the binding region falls within the basal region of multiple genes, then more than one assignment is made. All parameters were left at their default settings. Gene names obtained from Great were then converted to Refseq names for any downstream analysis. The software Webgestalt (Wang et al., 2013) was used for Gene Ontology and KEGG pathway analysis. All settings were left at their default settings.

Motif Discovery and Density Plots

For motif discovery, all tests were conducted with the HOMER package v4.2 using the following settings: -size 150 –S 10 –bits.(Heinz et al., 2010). We limited our motif analysis by only using 150bp DNA region around each peak summit. For statistical analysis, random background sequences with similar GC content to the test sample was generated for comparison, except for when we searched for Ascl1 or Ptf1a specific motifs. To find the factor specific motifs, either all of the Ascl1 or Ptf1a peaks were used as the treatment, while the peak regions not being tested were specified as the background.

The E-box and Rbpj motif density plots were generated in HOMER (annotatePeaks.pl –size 1000 –hist 10). The program identified all sequences that matched to the Rbpj (Ptf1a specific) or primary E-box HOMER generated motif matrices within 1 kb of the surrounding peak regions. Density plots were then generated by Matlab®.

RESULTS

Ascl1 and Ptf1a bind largely distinct sites within neural tube chromatin and have distinct E-box sequence preferences

In order to uncover mechanisms by which two neural class II bHLH factors regulate different sets of gene programs that give rise to distinct subtypes of neurons in dorsal neural tube, I compared and contrasted the genome-wide binding sites of Ascl1 and Ptf1a by ChIP-Seq in E12.5 mouse neural tubes. ChIP-Seq data sets for Ascl1 and Ptf1a have been recently published (Meredith et al., 2013, Sun et al., 2013) but were re-evaluated here and compared using the peak calling software Homer (Heinz et al., 2010). Using the parameters of an FDR cutoff of 0.001, a 4-fold enrichment of sequence tags in the target experiment over control, and a cumulative Poisson p-value threshold of 0.0001, Ascl1 was found to bind 4,082 sites, and Ptf1a was found at 7,749 sites, with 1,588 of those sites bound by both factors (Fig. 2.1A). Heat maps of the sequence reads +/- 2.5 kb of the Ascl1 and Ptf1a binding site summits, show the binding profiles of the two factors (Fig. 2.1A). The stringent criteria for peak calling discards many low affinity Ptf1a and Ascl1 binding events, and visual inspection of the heat maps (Fig. 2.1A) suggests that the 1,588 overlapping sites may be an under estimate. The location of the Ascl1 and Ptf1a bound sites with respect to gene transcriptional

starts sites (McLean et al., 2010) shows that Ascl1 and Ptf1a preferentially bind distal DNA elements (>5 to 500kb) rather than proximal promoters although they can be found at proximal promoters as well (Fig. 2.1B). Genes with Ascl1 or Ptf1a bound sites within 5 kb of their transcription start sites, and those with multiple Ascl1 or Ptf1a sites were expressed at higher mean levels (p-values < 0.05), although the difference in levels is not dramatic (Fig. 2.2A-B).

De novo motif analysis (Heinz et al., 2010) of the Ascl1 and Ptf1a called peaks returned the canonical E-box (CANNTG) (Fig. 2.1C), the known class II bHLH consensus binding site (Murre et al., 1989). I found that 99% of Ascl1 and 89% of Ptf1a bound sites contained a generic CANNTG E-box within 75 bp of the peak center (data not shown). The specific primary E-box motifs identified in the analysis are in agreement with previous findings (Meredith et al., 2013, Sun et al., 2013), and show that while the CAGCTG E-box is enriched in both Ascl1 and Ptf1a bound sites, Ptf1a peaks are also enriched with the CATCTG/CAGATG E-box (Fig. 2.1C). The primary motif sequences of Ascl1 and Ptf1a are commonly found near the peak centers (Fig. 2.1C). Because on average there are two or more E-boxes found in each Ascl1 and Ptf1a ChIP-Seq peak (Fig. 2.2C), I counted the type of E-box closest to the peak summit (Fig. 2.1D). This analysis confirms that the GC core is the most common E-box found near center of the peaks, while Ptf1a sites also show enrichment for the TC/GA core E-box. An example of a Ptf1a bound site that is not shared with Ascl1 can be seen within the inhibitory neuronal specification gene, *Pax2* (Fig. 2.1E). The DNA sequences under the summit of the Ptf1a peak within *Pax2* shows a sequence highly conserved between vertebrates that contains a TC/GA core E-box and no GC core E-box.

Moreover, I divided the Ptf1a peaks into 3 categories (peaks with TC/GA only, GC only, or both E-boxes), based on the E-box found within 75 bp of the Ptf1a peak centers, and determined the average Ascl1 ChIP-Seq coverage at these peaks. Since the TC/GA E-box is specific to Ptf1a binding, I predicted the average coverage of Ascl1 to be low at those peaks. Indeed, the average Ascl1 read occurrences is lowest at Ptf1a peaks that contain only a TC/GA core E-box compared to Ptf1a peaks that have a GC core E-box (Fig. 2.1F). A total of 1,332 Ptf1a peaks were classified as peaks with a TC/GA E-box only; and among those sites, only 145 overlap with Ascl1 peaks. These results suggest that *in vivo* Ptf1a can bind DNA with GC or TC/GA E-boxes; however, Ascl1 preferentially binds to regions with the GC E-box. From these data it cannot be determined if Ascl1 and Ptf1a bind these common E-boxes at the same time, or even in the same cells, or if they compete for occupancy at these sites.

Enrichment of non-E-box transcription factor motifs within Ascl1 and Ptf1a bound genomic regions

The specific E-box sequence may influence binding site selection or affinity of binding, but co-operation from additional transcription factors may also modulate Ascl1 and/or Ptf1a binding. Therefore, I searched for additional sequence motifs that might be enriched within the regions bound by both Ptf1a and Ascl1 (Fig. 2.1G), and for motifs enriched in those peaks specific to Ascl1 or Ptf1a (Fig. 2.1H). I found that sites shared by Ptf1a and Ascl1 are enriched with the GC E-box, Sox, Homeodomain, Rfx and Pou motifs (Fig. 2.1G). In fact, regardless of how the data are binned, shared sites, Ascl1 only, or Ptf1a only sites, are all enriched with Sox, Homeodomain, Rfx and Pou motifs (data not shown). Moreover, it is

known that Ptf1a is a component of a trimeric complex that includes Rbpj in addition to the heterodimeric E-protein partner, and this complex is required for Ptf1a function (Beres et al., 2006, Hori et al., 2008, Masui et al., 2008). With over 1,500 sites shared by Ptf1a and Ascl1, I anticipated an enrichment of the Rbpj binding site. Indeed, embedded within the Rfx motif is the canonical Rbpj binding site known as the TC-box (Fig. 2.1G, underline, YTYYC sequence). Thus, the Rfx motifs could be bound by Rfx factors or Rbpj, or both.

In order to find motifs enriched specifically in the Ptf1a bound sites, *de novo* motif analysis using all of the Ptf1a peaks was performed, but Ascl1 bound sites were used as the background sequences instead of using randomly generated sequences; this approach masks out common motifs and identifies motifs specifically found with Ptf1a. This strategy again revealed the TC/GA core E-box is more specific to Ptf1a sites than to Ascl1 sites. In addition, one other motif enriched within the Ptf1a bound sites was found, the Rbpj motif (Fig. 2.1H). This motif contains a stronger consensus sequence for the Rbpj binding site than that found within peaks in common between Ascl1 and Ptf1a (Fig. 2.1G), and it is no longer embedded inside the Rfx motif (Fig. 2.1H). If the Rbpj motif is specific to Ptf1a binding sites, there should not be enrichment of the Rbpj motif in the vicinity of Ascl1 only sites. Indeed, the frequency of the Rbpj motif can be found highest near the center of Ptf1a bound sites (Ptf1a only and shared), while Ascl1 only sites display a low frequency of the motif (Fig. 2.1J). ChIP-Seq studies have shown that Rbpj binding sites overlap with Ptf1a in neural tube and that the binding sites of these two factors can be found with a fixed DNA spacing of 1, 2, or 3 DNA helical turns (detailed description in Chapter 2 (Meredith et al., 2013)). In the reciprocal *de novo* motif analysis of the Rbpj binding sites from Rbpj ChIP-Seq analysis of

E12.5 neural tube (Meredith et al., 2013), the Rbpj motif is enriched along with an E-box that is similar to the Ptf1a primary motif (Fig. 2.1I). This is consistent with a role for Rbpj in influencing the selection of Ptf1a binding.

I also searched for Ascl1 specific co-factors using all the Ascl1 peaks found in the neural tube and using Ptf1a sites as the background sequence. Unexpectedly, no other transcription factor motifs were found enriched beyond the preferential binding of Ascl1 to the GC core E-box (Fig. 1H). This suggests that Ascl1 specific co-factor binding sites, if present, occur at a low frequency and are not detected by our genome-wide approach or that the co-factor shares a redundant motif with Ptf1a sites. For both Ptf1a and Ascl1, no significant occurrence of a fixed DNA spacing for any of the common motifs with the E-box was detected (data not shown). It was previously shown that Ascl1 binds DNA regions with single base pair spacing between an E-box and a Pou motif (Castro et al., 2006). I also find Ascl1 binding at these regions such as the proximal promoters of *Dll1* and *Dll3*; however, because the Pou domain occurs frequently within Ascl1 peaks and at a random distribution around the E-box, I could not detect a significant enrichment of the single base pair spacing using the *de novo* motif analysis (data not shown).

Ascl1 and Ptf1a have opposite actions in neuronal subtype specification

In order to identify the transcriptomes downstream of Ascl1 and Ptf1a, I took advantage of engineered mouse strains that mark Ascl1 and Ptf1a lineage cells with fluorescent proteins and crossed these strains with null mutants for these genes. Fluorescence-activated cell sorting (FACS) from E11.5 neural tubes was used to isolate cells for RNA-Seq. For the

Ascl1-lineage cells I used FACS isolated cells from *Ascl1*$^{GFP/+}$ knockin embryos (Leung et al., 2007) compared to *Ascl1*$^{GFP/null}$ embryos that completely lack Ascl1 protein. Because the GFP in the locus is more stable than Ascl1, Ascl1-expressing progenitor cells and their immediate progeny were isolated in this paradigm (Fig. 2.3A). For the Ptf1a-lineage cells, we isolated cells by FACS from mCherry expressing transgenic embryos where the mCherry is driven by a 12.4 kb genomic region normally 3' of the *Ptf1a* coding region that directs expression to the dorsal neural tube overlapping but not restricted to Ptf1a-expressing progenitors (Fig. 2.3B) (Meredith et al., 2009). The 12.4kbPtf1a::mCherry line was crossed to the *Ptf1a*^{Cre} knockin mouse (Kawaguchi et al., 2002), allowing us to isolate mCherry positive cells from *Ptf1a* heterozygous (*Ptf1a-/+*) and *Ptf1a* mutant neural tubes (*Ptf1a-/-).* From these isolated cell populations RNA was purified and mRNA-Seq performed. Alignment of sequenced reads to the genome show that the mutant samples lack signal over the part of the Ascl1 or Ptf1a coding regions that were deleted in the respective mutants (Fig. 2.3C-D, see brackets), confirming the identity and purity of the samples sequenced.

Analysis of the RNA-Seq data from the *Ascl1* heterozygotes compared to the *Ascl1* mutants, identify 1,173 genes with a significant difference in gene expression (q-Value<0.05) (449 Ascl1-activated; higher in heterozygotes versus the nulls) (724 Ascl1-repressed; lower in the heterozygotes versus the nulls) (Fig. 2.3E, red dots top panel). There were fewer genes that had altered expression in the *Ptf1a* heterozygous cells compared to the *Ptf1a* null cells with 361 genes showing a significant change in expression (132 Ptf1a-activated; higher in the heterozygotes versus the nulls) (229 Ptf1a-repressed; lower in the heterozygotes versus the nulls) (Fig. 2.3E, red dots lower panel). As expected (described in Chapter 1), the HD

transcription factors Tlx1, Tlx3 (Fig. 2.3E, top panel), and Lmx1b—which mark the excitatory neurons in the dorsal spinal cord—require Ascl1 for expression since there is a significant decrease in their expression in the *Ascl1* mutant versus *Ascl1* heterozygous populations (4.9, 6.6, and 9.7 fold change, respectively) (Fig. 2.3E, 2.4B). Conversely, Tlx1, Tlx3 (Fig. 2.3E, bottom panel), and Lmx1b expression significantly increase in the absence of Ptf1a (56.2, 24.7, and 33.3 fold change, respectively) (Fig. 2.3E, 2.4B). Among the genes that require Ptf1a for expression are those encoding the inhibitory neuronal markers Pax2, Lhx1 and Lhx5 (145.1, 69.6, and 48.1 fold change, respectively) (Fig. 2E, 2.4B). These inhibitory neuronal markers also require Ascl1 as seen in the more subtle decrease in Pax2, Lhx1 and Lhx5 expression in *Ascl1* mutants (1.73, 1.49, and 1.63 fold change, respectively) (Fig. 2E, 2.4B). The mRNA-Seq analysis is consistent with the known expression profile of these HD factors in the Ascl1 and Ptf1a wild-type and mutant tissue as discussed Chapter one (Fig. 1.2A-O).

In order to gain insight into the gene network interactions dependent on Ascl1 and Ptf1a, the genes that displayed significant change in expression in each lineage in the mutants versus controls were intersected (Fig. 2.3F). The majority of genes that change when Ascl1 or Ptf1a are mutated are non-overlapping, suggesting distinct functions for these bHLH transcription factors in the neural tube. However, there is a subset of the Ptf1a-activated genes that are also activated by Ascl1 (32 genes, App. A), and this subset includes the HD factors that specify the GABAergic lineage Pax2 and Lhx1/5 (Figs 1.2J-O). Thus, Ptf1a and Ascl1 activated genes distinctly define dI4/dIL^{A} Inhibitory neurons (App. A). Among these genes are those encoding other classes of proteins such as cell adhesion molecules Kirrel2

and Nphs1. Co-regulation by Ptf1a and Ascl1 of a set of genes expressed in dI4/dIL^{A} neurons may reflect the dependence of some Ptf1a expression on Ascl1.

More strikingly, 63 of the Ptf1a-repressed genes are activated by Ascl1 (Fig. 2.3F, App. A). This group of genes specifically marks the dI3/5, dIL^{B} neurons, and includes the HD factors Tlx1/3 (Figs 1.2G-I and 2.3B). This group also contains genes that code for factors such as Cbln1 and Cbln2 that are involved in forming connections that promote synapse formation in glutamatergic neurons (Ito-Ishida et al., 2012). In the dorsal horn of the spinal cord, Cbln1/2 expression is restricted to the excitatory neurons (Cagle and Honig, 2013). Thus, the genes activated by Ascl1 but repressed by Ptf1a define a subset of excitatory neurons in the dorsal spinal cord and is consistent with the opposing phenotypes of *Ptf1a* and *Ascl1* mutants in specifying dorsal interneurons in the first wave of neurogenesis (Helms et al., 2005, Glasgow et al., 2005).

Homeodomain neuronal specification factors are direct downstream targets of Ascl1 and Ptf1a

Identifying direct downstream targets of bHLH factors has been a challenge in the past due to the lack of specificity of the binding consensus for bHLH factors (i. e. the E-box CANNTG) and the technology to map bound DNA regions at distal enhancers. ChIP-Seq overcomes these challenges and allowed us to identify potential direct downstream targets of Ascl1 and Ptf1a, particularly those that influence neuronal sub-type specification. I define direct downstream targets here as genes that 1) show a significant change of expression between controls and mutants, and 2) have an Ascl1 or Ptf1a binding site within the gene's regulatory

region as identified by the GREAT algorithm (McLean et al., 2010). In GREAT, every gene is assigned a regulatory domain and if the transcription factor bound region falls within the regulatory region of multiple genes, then more than one gene assignment is made. As mentioned earlier, Ascl1 and Ptf1a bind regions distal to the TSS, thus binding often occurs in the extended regulatory regions. By intersecting the genes called in the ChIP-Seq data with those genes shown in the RNA-Seq data to require Ascl1 or Ptf1a for normal levels of expression, genes most likely to be direct downstream targets of these bHLH factors were identified. Using this strategy, I identified 449 putative targets regulated directly by Ascl1. Of these, 224 are activated and 225 are repressed by Ascl1. In contrast, 207 putative targets of Ptf1a were identified. Of these, 101 are activated and 106 are repressed by Ptf1a (Fig. 2.4A, App. B).

A major function of Ascl1 and Ptf1a is in specification of neuronal sub-type as seen in the dramatic changes in expression of HD factors in gain- and loss-of-function experiments (Fig. 1.2) (Glasgow et al., 2005, Nakada et al., 2004, Chang et al., 2013). I looked specifically for HD factor genes as direct targets of the bHLH factors. I found Ptf1a directly regulates known specification genes for inhibitory neurons such as *Pax2* and *Lhx1*, and Ascl1 directly regulates known specification genes for excitatory neurons such as *Tlx3* (Fig. 2.4B-D). Genomic regions containing the Ptf1a sites near *Lhx1* and *Pax2*, and the Ascl1 binding site near *Tlx3* (Fig. 2.4C,D, beige highlighted peaks) are reported to drive reporter activity in the dorsal neural tube at E12.5 (Chang et al., 2013, Meredith et al., 2013). In addition to *Pax2* and *Lhx1*, genes encoding other HD factors involved in development of the dorsal horn spinal cord GABAergic neurons, such as *Pax8*, *Lhx5*, *Gbx1*, and *Gbx2* (Pillai et

al., 2007, Luu et al., 2011, John et al., 2005) are also direct targets of Ptf1a (Fig. 2.4C). As was seen with *Pax2*, these genes also require Ascl1 to reach normal levels of expression (Fig. 2.4B). However, in most cases this is likely indirect through Ascl1 regulation of *Ptf1a* (Fig. 1.2D, E) since Ascl1 is not found bound near these genes. The exception is *Lhx1* and *Gbx1*, both genes have an Ascl1 binding site (overlaps with the Ptf1a sites) (Fig. 2.4C) and their expression significantly decreases in the *Ascl1* null (Fig. 2.4B). Ptf1a binds at two sites near *Lhx1*. The site not co-bound by Ascl1 is sufficient to drive dorsal neural tube restricted expression in transgenic mice (Fig. 2.4C, beige highlighted peak) (Meredith et al., 2013).

In contrast to the HD factor genes regulated by Ptf1a for specification of the GABAergic neurons, direct targets of Ascl1 comprise a different set of HD factor genes known to contribute to the proper development of dorsal horn spinal cord glutamatergic neurons. These include *Tlx3*, *Tlx1*, *Lmx1b*, *Isl1*, *Gsx2*, and *Pou4f1* (*Brn2a*) (Fig. 2.4D) (Ding et al., 2004, Mizuguchi et al., 2006, Avraham et al., 2010, Zou et al., 2012). Genes encoding the dorsal neural tube HD factors Pou3f1 (Oct6) and Uncx are also targets of Ascl1 (Saito et al., 1996, Gray et al., 2004), but their function in glutamatergic neurons in the dorsal horn has not been examined. It is striking that the HD factor genes activated by Ascl1 are repressed by Ptf1a (Fig. 2.4B). The repression by Ptf1a likely involves both direct and indirect mechanisms because there is no binding of Ptf1a near genes such as *Tlx1*, *Pou4f1*, and *Gsx2*; but genes such as *Tlx3*, *Isl1*, *Uncx* and *Pou3f1* have Ptf1a bound sites independent of those bound by Ascl1. Only *Lmx1b* has an overlapping Ptf1a and Ascl1 site, but it also has independent Ascl1 and Ptf1a sites. Thus, no single model explains the mechanism of Ptf1a repression of this set of genes.

In summary, I demonstrate a direct transcriptional regulatory relationship between the bHLH factors and genes coding for HD factors involved in neuronal subtypes specification in the dorsal spinal cord. Ptf1a directly activates six HD factor genes that function in the GABAergic neuronal lineages (Fig. 2.4C), while Ascl1 directly activates eight HD factor genes that function in the glutamatergic neuronal lineages (Fig. 2.4D). This places the bHLH factors at the head of a transcription factor network controlling the generation of these major classes of neurons in the dorsal spinal cord.

Ascl1 directly regulates genes involved in multiple processes of neurogenesis

Like other proneural bHLH factors, Ascl1 coordinates the transition from a neural progenitor cell to a differentiated neuron, and this is reflected in the Gene Ontology analysis of Ascl1 targets (Wang et al., 2013) (Fig. 2.5A). This cellular transition requires several biological processes to occur within a short window of time, such as cell-cycle exit, cellular migration, and cell-type specific gene expression. Therefore, I anticipated Ascl1 neural tube targets to be involved in several biological processes, similar to known functions of Ascl1 in the developing telencephalon (Castro et al., 2011, Castro et al., 2006). Indeed, Ascl1 directly regulates differentiation and specification not only through the HD factors, but though a larger complement of transcription factors (31 out 224 activated target genes are classified as encoding proteins with transcription factor activity; Fig. 2.5C). Ascl1 also targets components of the Notch signaling pathway, such as *Dll1*, *Dll3*, *Mfng*, *Numbl*, and *Hes5*. Additional aspects of neurogenesis regulated by Ascl1 include genes involved in neuronal projections and axon guidance, such as *Nfasc*, *Epha2*, *Ephb3*, *Sema7a*, *Sema6b*, *Dcc*, *Plxna2*,

Pak3, *Rgs3*, and *Slit*. Moreover, Ascl1 regulates several genes found in synaptic terminals that regulate neurotransmitter release, such as *Snap25* and *Syt6*. Notably, Ascl1 does not directly regulate genes involved in generating or transporting the neurotransmitter glutamate, and unlike the reported function for Ascl1 in the developing telencephalon (Castro et al., 2011), Ascl1 in the caudal neural tube does not activate many genes associated with proliferation.

Not only do genes need to be activated for cellular differentiation, but certain genes also need to be shut down. Indeed, our data suggest that Ascl1 also promotes differentiation by repressing s neural stem maintenance genes *Sox2* and *Pax3* (Pevny and Nicolis, 2010, Nakazaki et al., 2008). Ascl1 also represses expression of several genes involved in cell differentiation, many of which are not involved specifically in neuronal differentiation. Genes repressed by Ascl1 that have known neuronal function are involved in early dorsal-ventral pattering of the spinal cord, such as the Wnt signaling components, notably *Wnt1* (Augustine et al., 1995, Lee and Jessell, 1999), and sonic hedgehog components such as *Gli3* (Persson et al., 2002).

Ptf1a directly represses the glutamatergic fate and upregulates components of the GABAergic machinery

Similar to Ascl1, Ptf1a turns on a cascade of transcription factors that function in neuronal differentiation and specification (Fig. 2.6A, E). Approximately a quarter (24 out 101) of Ptf1a activated targets have transcriptional activity. In contrast to Ascl1, Ptf1a directly activates neural genes that are specific to the GABAergic program, including genes encoding

the transcription factors Prdm13, Lhx1, Lhx5, and Pax2, which are important for maturation of these neurons (Pillai et al., 2007, Chang et al., 2013). Nine of the target genes encode synaptic proteins such as Sv2c (Fig. 2.6E), Sez6, and Iqsec3, which have all been shown to localize specifically at inhibitory synapses (Gronborg et al., 2010, Gunnersen et al., 2009, Fukaya et al., 2011). In addition, Ptf1a activates genes involved in GABA biosynthesis and transport pathways, such as Gad1 (Gad67), Abat (GABA transaminase), Slc32a1 (Viaat), and Slc6a5 (Glyt2) (Fig. 2.6D-F). Genes encoding both Slc32a1 and Slc6a5 transporters have Ptf1a sites nearby with no Ascl1 sites present (Fig. 2.6C-E), highlighting Ptf1a's unique role in regulating genes specific to inhibitory neurons.

Whether the other Ptf1a regulated genes are also involved in GABAergic neuronal development is less clear. Several of these genes contribute to the extracellular matrix and cell adhesion such as *Adamts4*/*20* and *Adamtsl1*, *Nrxn1*, *Vcan*, *Gpc3/4*, *Ccbe1*, *Nphs* and *Kirrel2*. Genes encoding three subunits of the voltage-gated calcium channel (Cacna2d2, Cacna2d3, and Cacna1g), which play a role in MAPK signaling pathway are also found to be regulated by Ptf1a.

Ptf1a not only activates components necessary for the GABAergic lineage but also represses the glutamatergic fate. Approximately a third (33 out 106) of the genes repressed by Ptf1a are directly activated by Ascl1, notably HD factors Tlx3 and Lmx1b (Fig. 3A, App B). In addition, Ptf1a has been shown to repress the glutamatergic fate through the activation of a target gene Prdm13 (Chang et al., 2013). Prdm13 directly interacts with Ascl1 to block its transcriptional activity. All of the inhibition could potentially be explained by Prdm13's

function; however, the ChIP-Seq analysis revealed that Ascl1 and Ptf1a commonly occupy the same genomic regions around 22 of 33 Ascl1-activated; Ptf1a-repressed genes. A total of 42 Ascl1 and Ptf1a shared binding sites are found around these 22 genes, suggesting Ptf1a repression of the glutamatergic fate may occur through alternative mechanisms involving direct binding by Ptf1a.

Ascl1 and Ptf1a bind active neural enhancers but are not required for their activity

Our ChIP-Seq assay shows that both Ascl1 and Ptf1a bind thousands of sites throughout genome, but only a fraction of the genes associated with sites require either factor for expression. I expected more genes to require Ascl1 or Ptf1a for expression because analysis of all the Ascl1 and Ptf1a binding sites using GREAT (McLean et al., 2010), shows both factors preferentially occupy regions near genes with known nervous system development functions (binominal p-value= 7.23e-19 and 1.03e-44 respectively). With that said, it is unclear if these binding sites are active neural enhancers or if they are just poised or competent to be activated. One indication of an active enhancer is the presence of the transcriptional co-activator P300. In a large project from the Pennachio laboratory, enhancers identified by P300 ChIP-Seq from neural tissues were tested for activity in transgenic mice. Most of these were examined as whole mount E11.5 embryos, and data are posted in the VISTA Enhancer Browser database (Visel et al., 2009a, Visel et al., 2007). When identified 10 enhancers that were tested in this study that encompass Ascl1 and Ptf1a ChIP-Seq peaks. Axel Visel and Len Pennachio provided the whole mount embryos so I could cross section them and determine if expression of the reporter gene was restricted to the Ascl1 and/or Ptf1a domain in the dorsal ventral axis of the neural tube. All 10 DNA regions contained

overlapping Ptf1a and Ascl1 sites, but the neighboring genes did not require Ascl1 or Ptf1a for expression and transgene expression was not restricted to the dorsal neural tube as would be expected if these enhancers were Ascl1 and Ptf1a specific (Fig. 2.7A-C and data not shown). Thus, I find that Ascl1 and Ptf1a bind to genomic regions in neural tissue where they are not necessarily required for enhancer activity, yet the regions bound are active enhancers in neural tissue.

DISCUSSION

Over two decades of research have established the genetic requirement for neural bHLH transcription factors in generating the proper number and composition of neurons in the central nervous system. However, a mechanistic understanding of bHLH factor function is only just emerging as transcriptional targets are beginning to be identified through genome-wide strategies utilizing ChIP-Seq. In this study, I identify numerous direct transcriptional targets of two of these bHLH factors in the mouse dorsal neural tube using a combination of ChIP-Seq and RNA-Seq. Ascl1 and Ptf1a provide an informative model pair of factors for uncovering how two related transcription factors function to specify distinct cell fates from a common progenitor domain. Ascl1 defines neural progenitors that will give rise to most of the interneurons in the dorsal horn of the spinal cord. Ptf1a comes on later in a subset of these neural progenitors and directs the cells to an inhibitory neuronal fate. In the Ascl1 population lacking Ptf1a, the progenitors give rise to excitatory interneurons. I show here that Ascl1 and Ptf1a directly activate sets of genes encoding HD factors that are known specifiers of the glutamatergic or GABAergic neuronal fates, respectively. In addition, Ptf1a represses

many of the Ascl1-activated HD genes, ensuring that the glutamatergic phenotype is repressed in GABAergic neurons. Furthermore, Ascl1 targets genes early in the differentiation process whereas Ptf1a targets genes required for the activity of GABAergic neurons, reflecting the temporal difference in their expression and fundamental differences in their function.

Ascl1 and Ptf1a directly regulate distinct neuronal programs in the dorsal neural tube

Many of the Ascl1 and Ptf1a targets have been shown to play a key role in neural tube development. Not only does Ascl1 specify the glutamatergic population through the activation of key transcription factors, but also contributes to other components of the neurogenic program. Historically, Ascl1 has been known as transcriptional activator in the dorsal spinal cord (Nakada et al., 2004). However, our data suggest that Ascl1 could potentially repress several genes too. In support, of the repression model, Ascl1 has been shown directly interact with a putative repressor, Prdm13 (Chang et al., 2013). Moreover, I find that Ascl1 and Ptf1a bind a common E-box, thus Ascl1's repression could simply be a passive event, where Ascl1 blocks E-box binding sites from other bHLH factors. Nonetheless, I observe that Ascl1 drives neuronal differentiation in two ways, by directly repressing several genes involved in cell-cycle progression such as *Cdk1* and *Chek1*, (Chen et al., 2012, Tsunekawa et al., 2012), and promoting cell-cycle exit genes such as *Gadd45g* and *Cdkn1c* (Huang et al., 2010, Mairet-Coello et al., 2012). Moreover, Ascl1 helps in defining boundaries of its progenitor population through suppression of different morphogenic signaling factors (e.g., Wnt and Gli3 (Persson et al., 2002, Lee and Jessell, 1999)), and

guides axon terminals through a large network of genes such as *Dscam*, *Epha2/3*, *Dcc*, *Sema7a*, *Sema6b*, *Plxna2*, *Pak3*, *Rgs3*, and *Slit1*. Interestingly, in the developing telencephalon, where Ptf1a is not expressed (Glasgow et al., 2005), Ascl1 specifies the GABAergic cell population. Both mRNA expression and ChIP-chip analysis revealed that Ascl1 can directly specify the telencephalon GABAergic population through regulating several transcription factor genes including the HD factor *Dlx2*, and activate *Gad2*, encoding an enzyme needed in synthesizing the neurotransmitter GABA (Castro et al., 2011). *Dlx2* and *Gad2* are not Ascl1 targets in the neural tube. Moreover, it is the role of Ptf1a to specify the GABAergic population in the neural tube. With that said, several gene targets of Ascl1 do remain the same between different tissues such as the Notch ligands *Dll1* and *Dll3*. Thus, to fully understand how Ascl1 functions, one most carefully examine its role in each tissue-type of interest.

In contrast, Ptf1a function is not as general as Ascl1 in the dorsal neural tube. Ptf1a does not suppress cell-cycle genes or promote cell-cycle exit genes like Ascl1. For the most part, the genes regulated directly by Ptf1a are focused on ensuring proper specification and maturation of inhibitory neurons. These include activation of HD genes important in this lineage (i.e. *Pax2*, *Gbx1*, *Gbx*2, *Lhx1*, and *Lhx2*), and activation of genes encoding cellular components specifically needed for GABAergic neuronal function such as the machinery needed to make GABA (i.e. *Gad1* and *Abat*). In addition, Ptf1a appears to directly suppress Ascl1 target HD genes important in the glutamatergic lineage. The list of target genes for Ptf1a and Ascl1 in the developing neural tube provided from this study will serve as a rich source of data for further probing the functions of these essential transcription factors.

Multiple mechanisms for cross repression of Ascl1 and Ptf1a in neuronal subtype specification

Through identifying HD factor genes as direct targets of Ascl1 and Ptf1a, I are gaining mechanistic insight into fundamental processes of cell fate determination, specifically those generating the balance of excitatory and inhibitory neurons in the dorsal spinal cord. Function assays for Ascl1 or Ptf1a have shown that each will induce one cell fate while suppressing the other (Helms et al., 2005, Glasgow et al., 2005, Hori et al., 2008, Chang et al., 2013). Because Ascl1 and Ptf1a are known as transcriptional activators, identifying HD genes specifying the relevant neurons as targets directly activated by these bHLH factors provides a simple model for generation of the excitatory and inhibitory neurons, respectively. However, it is much less clear how the alternative cell fates are repressed. Our results provide multiple insights into mechanisms for this cross repression.

Understanding how Ascl1 can repress the Pax2 defined GABAergic lineages comes from our data showing Ascl1 directly activates *Tlx1* and *Tlx3* defined glutamatergic lineages and the current model for cross repression between the Tlx factors and Pax2 (Cheng et al., 2005). Cheng et al. demonstrated that Tlx1 or Tlx3 suppress the GABAergic fate by antagonizing the ability of the HD factor Lbx1 to induce Pax2 expression (see diagram Fig. 1.2P). Thus, the higher the levels of Tlx1 and Tlx3 driven by Ascl1, the lower the levels of Pax2. Consistent with this indirect mechanism for Ascl1 suppression of the Pax2 GABAergic lineage, Ascl1 did not directly suppress Ptf1a activated target genes as determined by ChIP-Seq. Rather, I found that several of the Ptf1a activated genes were also activated by Ascl1.

Taken together, it appears that Ascl1 activates the glutamatergic lineage through direct regulation of HD factors such as the *Tlx1/3* genes, but only indirectly represses the opposing lineage.

In contrast, Ptf1a uses both direct and indirect mechanisms for suppressing the glutamatergic neuronal fate. I recently identified *Prdm13* as a direct downstream target of Ptf1a that provides an indirect mechanism for Ptf1a suppression of the glutamatergic fate (Chang et al., 2013). In this case Prdm13, induced by Ptf1a, directly suppresses Ascl1 activity in inducing its targets such as *Tlx1* and *Tlx3* (see diagram Fig. 1.2P). In addition, I show here that Ptf1a can suppress several Ascl1 targets directly, although the specific mechanism used is not clear. In several cases, I show that Ptf1a may suppress Ascl1 activated genes through binding sites that do not overlap with Ascl1 sites (i. e. *Tlx3, Uncx* and *Pou3f1)*. This would suggest a mechanism by which Ptf1a recruits a transcriptional repressor and blocks Ascl1 activity from a distant site; a capability that has never been shown for Ptf1a. In other cases, such as binding around *Lmx1b*, I show that Ascl1 and Ptf1a can locate to the same genomic site. This opens the possibility that some type of interaction or competition may occur between Ptf1a and Ascl1. In support of this mechanism, in a transcriptional assay with an E-box reporter, the ability of Ascl1 to activate the reporter was suppressed by Ptf1a (Obata et al., 2001). Given that Ptf1a and Ascl1 can bind a common CAGCTG E-box *in vivo* as shown here, Ptf1a could compete with Ascl1 for these E-boxes and passively block Ascl1 activity. On the other hand, *in vitro* studies have shown that Ascl1 can form complexes with another class II bHLH factor Neurog2 (Henke et al., 2009a, Gradwohl et al., 1996), thus an alternative mechanism could involve Ptf1a forming a novel complex with Ascl1 that is

transcriptionally inactive. I cannot distinguish between these models with our *in vivo* data because the ChIP-Seq resolution is limited to around 150 base pairs. I find Ascl1 and Ptf1a localized to common regions throughout the genome, but an examination of those regions reveal the presence of multiple E-boxes in close proximity with each other; specifically the GC-core E-box and at times, the Ptf1a specific TC-core E-box. In addition, the ChIP-Seq data cannot determine whether Ascl1 and Ptf1a are binding the shared sites at the same time or even within the same cell in the population. Regardless of these limitations, the results support the possibility that Ptf1a has the potential to directly block the activation of Ascl1 target genes. Thus, Ptf1a appears to deploy several mechanisms to ensure the proper repression of a subset of Ascl1 target genes to ensure repression of glutamatergic lineage genes in the GABAergic neurons.

Ascl1 and Ptf1a bind neuronal enhancers enriched with neuronal lineage factor motifs

Our *in vivo* analysis of Ptf1a and Ascl1 binding sites found enrichment of other transcription factor family motifs such as Sox, Pou, Homeodomain, and Rfx. It remains unclear which specific factor or factors bind to these sites and if they directly co-regulate transcription with Ascl1 or Ptf1a. It has been shown that the combination of lineage-specific transcription factors can establish the cell-type identity by priming cis-regulatory elements around the genes needed for that cell-type (Heinz et al., 2010). As pluripotent cells become more restricted to a certain fate, changes in open chromatin and the motifs found in these regions provide information about that cell's fate and its lineage (Stergachis et al., 2013). In neural development, Sox and Pou factors such as Sox2 and Brn2 are lineage-specific factors that collaborate in order to transition an embryonic stem cell to a neural progenitor cell (NPC)

(Bergsland et al., 2011, Lodato et al., 2013); Sox3 and Sox11 can sequentially replace Sox2 to coordinate the later stages of neural development (Bergsland et al., 2011). Point being, once the lineage-specific factors establish the chromatin, other transcription factors can help coordinate the remaining tissue-specific events through the same cis-regulatory region (Heinz et al., 2010). Consistent with this model, motif analysis of the Ascl1 and Ptf1a ChIP-Seq data found similar motifs with Sox2 and Brn2 shared binding sites in NPC (Sox, Pou, Homeodomain, Rfx, NfI and E12 (E-box)) (Lodato et al., 2013), suggesting that these families of factors are important throughout the development of the nervous system. The neural landscape is established prior to Ascl1 and Ptf1a expression so these neural transcription factors could be maintaining the landscape and/or co-regulating gene expression with Ascl1 or Ptf1a. Indeed, Brn2a, a Pou factor, which helps establish the neural competency with Sox2 (Lodato et al., 2013), can also work with Ascl1 to co-regulate genes such as those encoding Notch ligands Dll1 and Dll3 (Castro et al., 2006). Our RNA-Seq analysis shows that numerous members Pou/Homeodomain, Rfx, and NFI families are co-expressed in Ascl1 and Ptf1a lineage cells. However, with so many of these factors being expressed in the dorsal neural tube and the limitation of our motif analysis, which does not determine which specific factor or factors co-bind with Ascl1 or Ptf1a, makes it difficult to identify specific collaborations. Nonetheless, determining which specific factor and what combinatorial interactions that occur between these neural factors and with Ptf1a or Ascl1 will be essential for understanding the specific neural programs they regulate.

Previous work has shown that enhancers containing sites independently bound by Ascl1 or Ptf1a can recapitulate the expression pattern of its target gene, such as an enhancer

bound by Ascl1 near *Tlx3* (Chang et al., 2013), or enhancers bound by Pft1a near *Pax2*, *Lhx1*, and *Tcfap2b* (see chapter 3 (Meredith et al., 2013)). I also found that genomic regions bound by both Ascl1 and Ptf1a are also bound by the co-activator P300, and are active neural specific enhancers when tested in transgenic mice. However, the genes identified near these P300 marked enhancers do not require Ascl1 or Ptf1a for normal expression, at least when assayed by RNA-Seq. Indeed, a majority of the genes called near the Ascl1 and Ptf1a binding sites did not change when assessed in the Ascl1 or Ptf1a mutants. This highlights a major question in the field, what determines when a transcription factor actively alters gene expression? Functional redundancy is one likely possibility for some of the sites, context specific co-factors is another. As for the latter model, I isolated the peak regions around the genes that showed a significant change in gene expression from Ascl1 and Ptf1a wild-type compared to mutant tissues. I searched for known motifs within those sites, but did not find any unique motif combination that could explain the changes in expression. In fact, like the majority of the Ascl1 and Ptf1a bound sites, I found E-boxes and enrichment of the neural factors previously mentioned (data not shown). Moreover, I expected to find that *Ascl1* directly regulates genes such as Dll1, Dll3, and Fbxw7 found to be co-regulated by Ascl1 and Brn2a in the developing telencephalon (Castro et al., 2006). This was expected because Ascl1 and Brn2a are also co-expressed in the neural tube and Ascl1 binds the same regions in neural tube that were identified by Castro *et al*. in the telencephalon. However, although *Dll1* and *Dll3* were significantly decreased in expression in the *Ascl1* mutant neural tubes relative to controls (data not shown), *Fbxw7* levels were unaltered (Fig. 2.7C). Thus, not all genes predicted to require Ascl1 did so, highlighting a gap in understanding of the determinants for

transcriptional activity and reinforces the importance of cell context on transcription factor function.

Ptf1a binds a distinct E-box to regulate its specification program.

Ascl1 and Ptf1a ChIP-Seq data alone have provided valuable insight into how two different bHLH factors can select their DNA binding site. Motif analysis shows that Ptf1a binds a common E-box with Ascl1 (CAGCTG), but also has a preference for a distinct E-Box (CATCTG/CAGATG). Other genome-wide studies have shown bHLH factors, such as Atoh1 and Neurod2, have an E-box preference similar to Ptf1a (GC and TC/GA cores) (Klisch et al., 2011, Fong et al., 2012, Lai et al., 2011). In contrast, the myogenic factor, Myod, binds CAGCTG and CAGGTG/CACCTG E-boxes (Fong et al., 2012). Interestingly, E-box preferences for a specific bHLH factor can change depending on the cell context; for example, Ptf1a in the developing pancreas preferentially binds E-boxes similar to Myod (GC, and GG/CC cores) not the TC/GA core preference found in the neural tube (Meredith et al., 2013). Several studies have suggested that E-box binding by bHLH factors is heavily dictated by the availability of the site, thus the chromatin landscape of each tissue-type strongly influences a transcription factor's target selection (Meredith et al., 2013, Fong et al., 2012). In this study, both Ascl1 and Ptf1a are functioning in a similar chromatin landscape, yet I was still able to detect that some of the specificity can still be accounted for by the sequence preferences. Teasing out the consequences of this sequence preference is complicated by the fact that most Ptf1a bound regions contain the GC-core E-box or sites with both the GC and TC/GA core E-boxes. Nonetheless, examination of the genes near Ptf1a binding sites that

contain just the specific GA/TC core E-box alone revealed this was the case in 41 out of the 101 Ptf1a activated genes. Included in this subset of Ptf1a targets are some the known specification factors such as *Pax2* (Fig 2.1E), *Pax8*, *Lhx*5, *Gbx2*, and *Prdm13*. Thus, Ptf1a preference for a specific E-box, distinct from Ascl1, may be a particularly important mechanism in activating genes for the GABAergic neuronal identity and function.

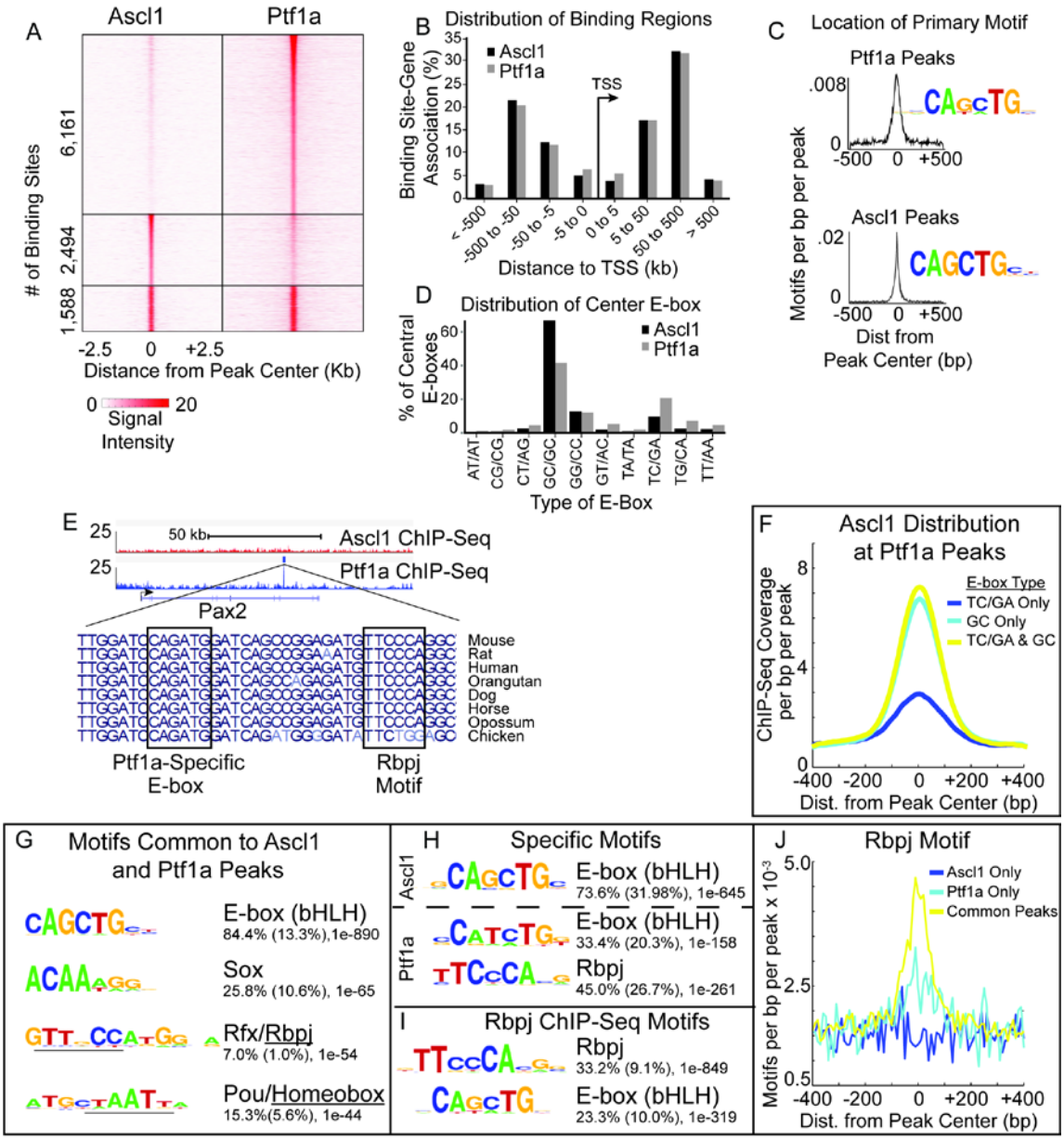


Figure 2.1. Global characterization of Ptf1a and Ascl1 binding sites in the developing neural tube reveals distinct and overlapping preferences for specific E-boxes, co-factor motifs, and binding sites. (A) Heat map summarizes the ChIP-Seq signal intensity +/- 2.5 kb around the identified peak centers bound by Ascl1, Ptf1a, or both in E12.5 mouse neural tube. (B) Ascl1 and Ptf1a preferentially bind far away from promoters shown by the distribution of Ascl1 and Ptf1a bound sites with respect to gene transcriptional start sites (TSS). (C) *De novo* motif analysis of Ascl1 and Ptf1a bound sites show both are enriched with a common GC core E-box, but Ptf1a sites are also enriched with GC/TA core E-boxes. Density plots show that the primary motif is enriched near the peak center. (D) The bar graph

displays the distribution of the E-box core type found nearest to the peak centers. (E) Example of a Ptf1a bound site, not shared by Ascl1, located within the Pax2 gene that contains a highly conserved GA/TC core E-box and Rbpj motif. (F) Coverage of Ascl1 ChIP-Seq fragments at Ptf1a peaks that contain a GA/TC E-box only, a GC E-box only or both GA/TC and GC E-boxes. (G) Additional *de novo* motif analysis shows the GC core E-box, Sox, Rfx/Rbpj, and Pou/Homeobox motifs are found enriched in sites shared by Ptf1a and Ascl1. The percent of motifs found in the test sequences, (background sequences), and calculated statistical p-values are provided. (H) Motifs enriched in Ascl1 sites over Ptf1a, or Ptf1a sites over Ascl1 are shown. (I) Motif analysis of Rbpj ChIP-Seq sites (from E12.5 neural tube) identified the known Rbpj motif and a Ptf1a-like E-box motif allowing for the GA/TC core. (J) The frequency of finding the Rbpj motif in Ascl1 only, Ptf1a only, or Ascl1/Ptf1a shared sites.

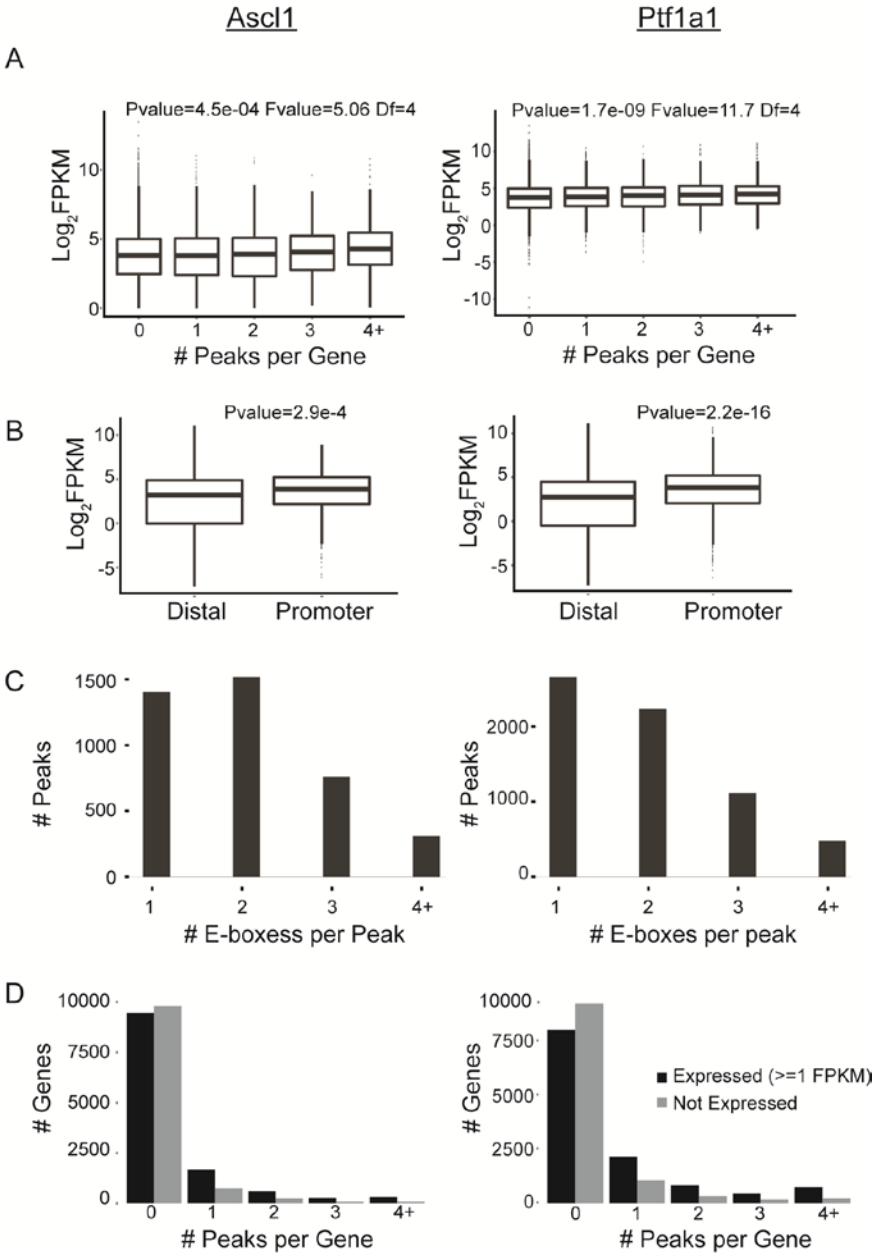


Figure 2.2. Correlation analysis of Ascl1 and Ptf1a binding with gene expression and the number of E-boxes found per peak. (A) Boxplot of the genes expressed (FPKM>1) and the number of Ascl1 or Ptf1a binding sites found for each gene. Genes with a higher number peaks were expressed at higher FPKM levels (Linear regression analysis, p-value<0.05) (B) Boxplot showing expression for all genes that have an Ascl1 or Ptf1a binding site in the promoter (+/-500 bp of TSS) or outside the promoter (distal). Genes with an Ascl1 or Ptf1a binding site within promoter region were expressed significantly higher than genes with a distal site only (T-test, p-value<0.05)**.** (C) The number of E-boxes (CANNTG) found per peak. (D) The number of binding sites found per gene that are expressed (FPKM>1) or not expressed (FPKM<1).

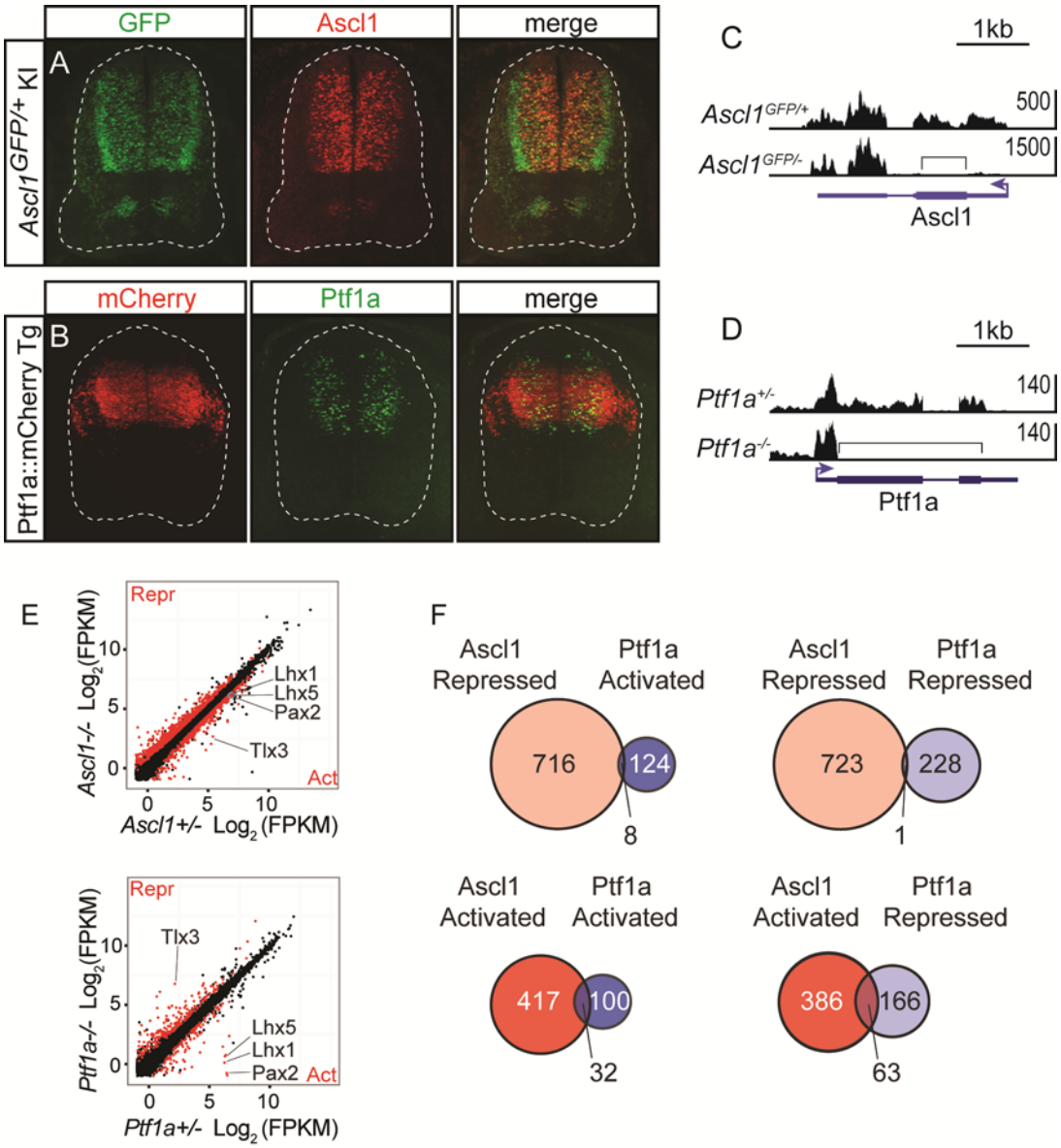


Figure 2.3. Ascl1 and Ptf1a lineage cells isolated for RNA-Seq demonstrates their opposing roles in neural tube specification. (A-B) Neural tube cross sections at E11.5 of the *Ascl1*$^{GFP/+}$ knockin and 12.4kbPtf1a::mCherry transgenic mice used to isolate Ascl1 and Ptf1a lineage cells by FACS. (A) Ascl1 antibody was used to show that GFP co-localizes with Ascl1 expressing progenitors in the ventricular zone. In addition, because GFP is more stable than Ascl1, the FACS isolation of GFP cells will also isolate Ascl1-lineage cells that have down regulated Ascl1. (B) Ptf1a antibody was used to show that mCherry is expressed in the dorsal neural tube overlapping but not restricted to Ptf1a-expressing progenitors. (C) Sequence reads (RNA-Seq) mapping to the Ascl1 locus from GFP isolated cells from *Ascl1*$^{GFP/+}$ or *Ascl1*$^{GFP/-}$ E11.5 mouse neural tubes. The brackets indicate the genomic regions replaced by GFP in the *Ascl1*^{GFP} knockin mice. . The increase in the number of sequence

reads over the *Ascl1* 3' UTR in the *Ascl1* mutant is likely due to auto-regulation at the *Ascl1* locus (note scale) (Meredith and Johnson, 2000). (D) Sequence reads (RNA-Seq) mapping to the Ptf1a locus from mcherry isolated cells from *Ptf1a::mCherry;Ptf1a*$^{Cre/+}$ or *Ptf1a::mCherry;Ptf1a*$^{Cre/Cre}$ E11.5 mouse neural tubes. The brackets indicate the genomic regions replaced by Cre in the *Ptf1a*^{Cre} knockin mice. (E) Scatterplots of RNA-Seq data shows expression levels (FPKM) in Ascl1-lineage cells from *Ascl1* control or null embryos (top plot), and Ptf1a-lineage cells from *Ptf1a* control or null embryos (bottom plot). Genes with a significant change in expression in the mutants versus controls are marked in red (q-value < 0.05). Genes above the diagonal are repressed (Repr) by Ascl1 or Ptf1a, and genes below the diagonal are activated (Act). The expression levels for the HD factors Tlx3, Pax2 and Lhx1/5 are marked in each condition. (F) Venn diagrams show the number of genes activated or repressed by Ptf1a or Ascl1, and how many genes overlap in each condition.

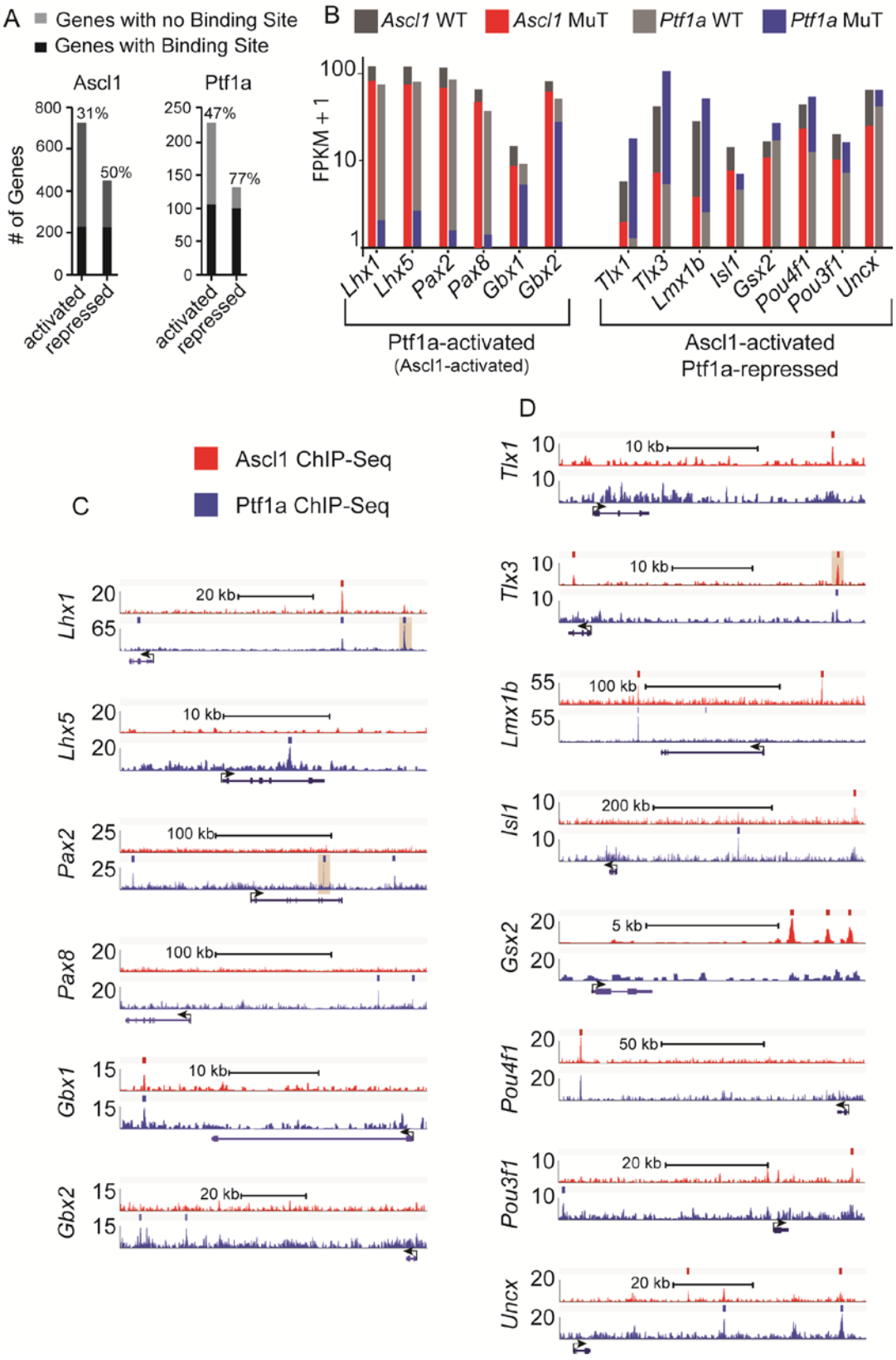

A
Genes with no Binding Site
Genes with Binding Site
Ascl1
Ptf1a
# of Genes
31%
50%
47%
77%
activated
repressed
B
Ascl1 WT
Ascl1 MuT
Ptf1a WT
Ptf1a MuT
FPKM + 1
Lhx1
Lhx5
Pax2
Pax8
Gbx1
Gbx2
Tlx1
Tlx3
Lmx1b
Isl1
Gsx2
Pou4f1
Pou3f1
Uncx
Ptf1a-activated
(Ascl1-activated)
Ascl1-activated
Ptf1a-repressed
C
Ascl1 ChIP-Seq
Ptf1a ChIP-Seq
D

Figure 2.4. Ascl1 and Ptf1a directly regulate the Homeodomain factors involved in neuronal specification. (A) Bar graph shows the total number of genes activated or repressed by Ascl1 or Ptf1a determined by RNA-Seq, and how many of those genes contain an Ascl1 or Ptf1a binding site nearby determined by ChIP-Seq (black fill). Thus, the black fill represents direct target genes of Ascl1 or Ptf1a as defined by having a significant change in expression between control and null tissues, and having a ChIP-Seq binding site assigned by the software GREAT (McLean et al., 2010). (B) Expression levels (FPKM +1) in dorsal neural tubes of *Ascl1* or *Ptf1a* control and mutant embryos of the HD factor genes that are directly regulated by Ptf1a (left panel) or by Ascl1 (right panel). Y-axis is in log10 scale. Directly activated targets of Ptf1a partially depend on Ascl1 expression, while direct targets of Ascl1 are repressed by Ptf1a. All changes in gene expression are significant (q-value < 0.5), except *Pax8* & *Gbx2* in the *Ascl1* null, and *Isl1* & *Gsx2* in the *Ptf1a* null. (C-D) Genomic regions around Ptf1a targets (C) or Ascl1 targets (D) displaying Ascl1 (red track) and Ptf1a (blue track) ChIP-Seq data. Note that the Ascl1 and Ptf1a sites in *Tlx3* and *Uncx* are not within 150 bp, and thus, are near each other but not overlapping (D). Target gene names are indicated to the left as is the scale indicating normalized number of sequence reads. Binding sites called by Homer are indicated by tick marks above each track. Beige box marks DNA elements previously shown to drive proper dorsal neural tube expression (Meredith et al., 2013, Chang et al., 2013).

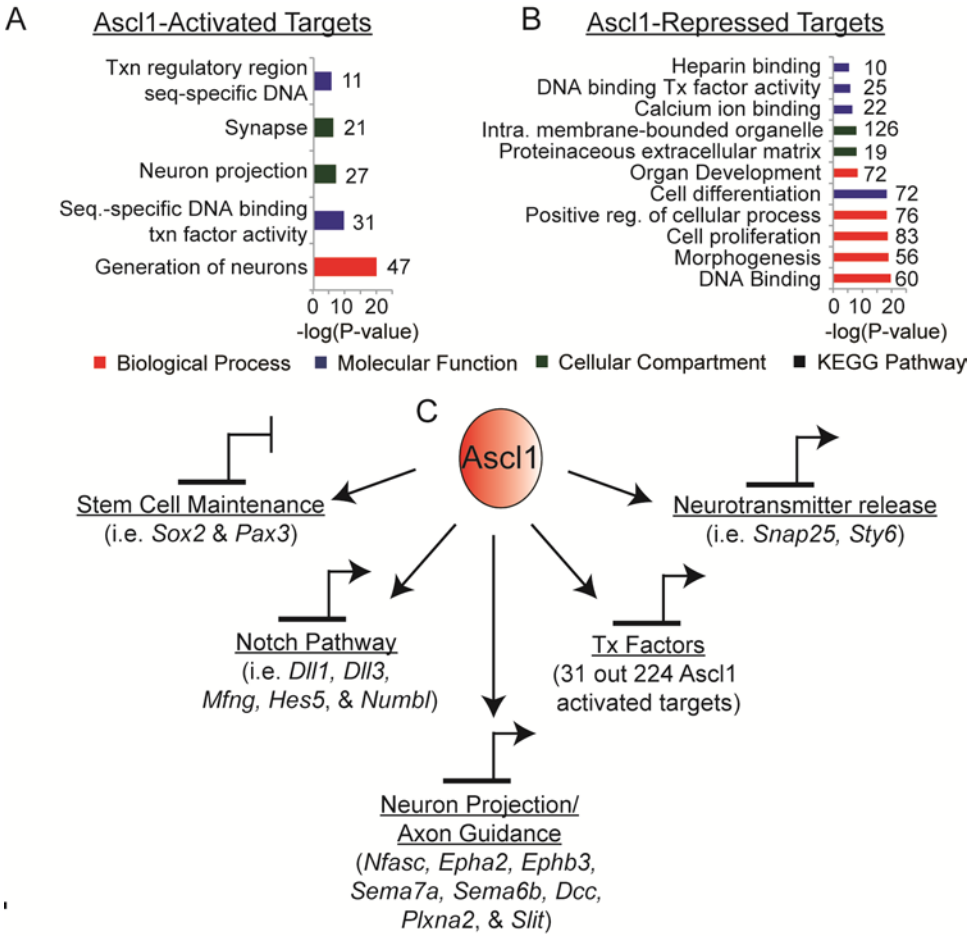


Figure 2.5. Ascl1 directly regulates several components of the neurogenic program in the developing spinal cord. (A-B) Gene ontology/KEGG analysis of direct Ascl1 activated (A) or repressed (B) targets. The color denotes the GO category or KEGG pathway the term comes from, and adjacent number indicates the number of target genes in that category. (C) Diagram summarizing Ascl1 gene targets that are involved in different steps the neurogenic program.

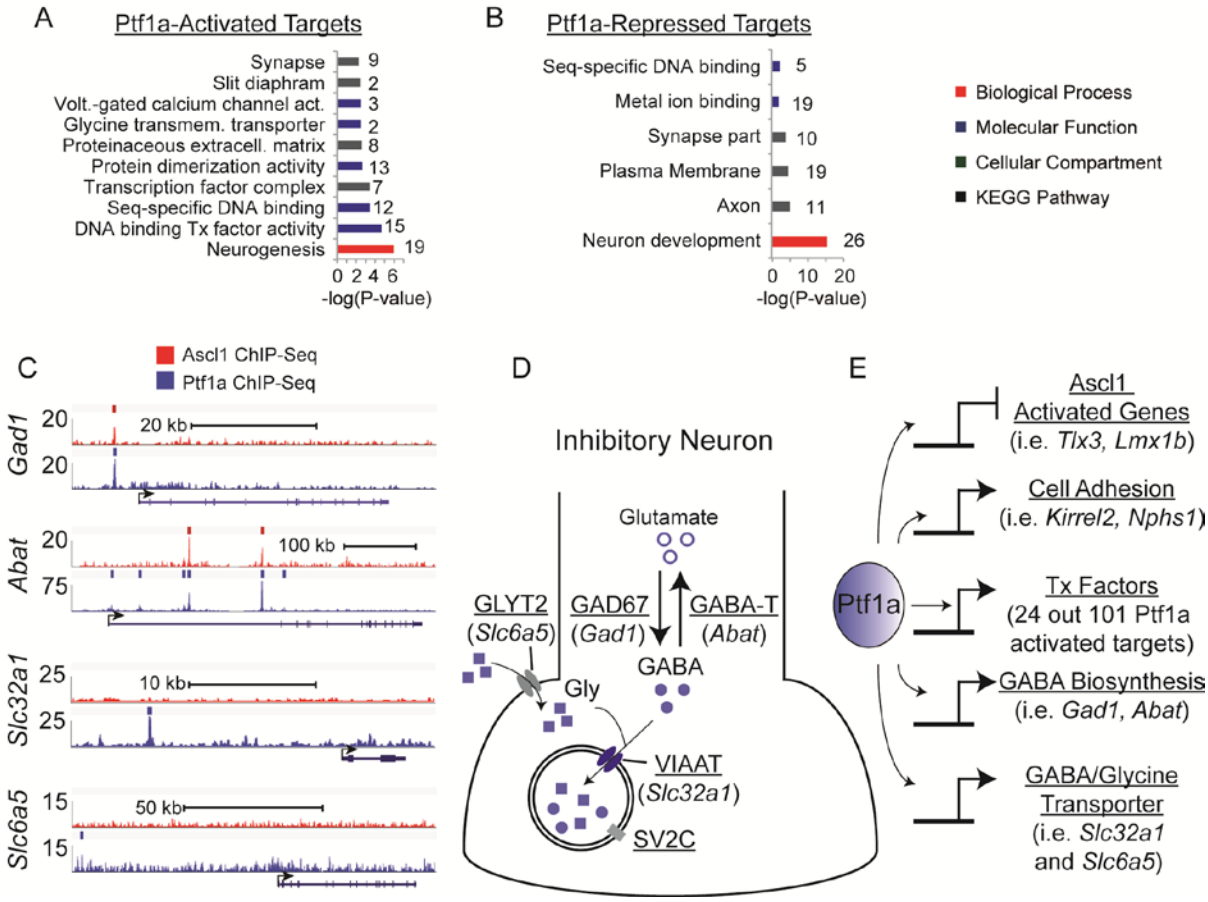


Figure 2.6. Ptf1a directly opposes Ascl1 target genes and activates several components of the GABAergic machinery. (A-B) Gene ontology/KEGG analysis of Ptf1a activated (A) or repressed (B) targets. The color denotes the GO category or KEGG pathway the term comes from, and adjacent number indicates the number of target genes in that category. (C) Genomic regions around Ptf1a target genes involved in GABA biosynthesis and GABA/glycine transport, along with Ascl1 (red track) and Ptf1a (blue track) ChIP-Seq data. Binding sites called by Homer are indicated by tick marks above each track. (D) Diagram of Ptf1a target genes that are specifically found in inhibitory presynaptic terminals (in parentheses). (E) Diagram summarizing Ptf1a gene targets that are involved in different steps of inhibitory neuron specification and include repressing Ascl1 target genes and activating genes encoding components of the GABAergic machinery.

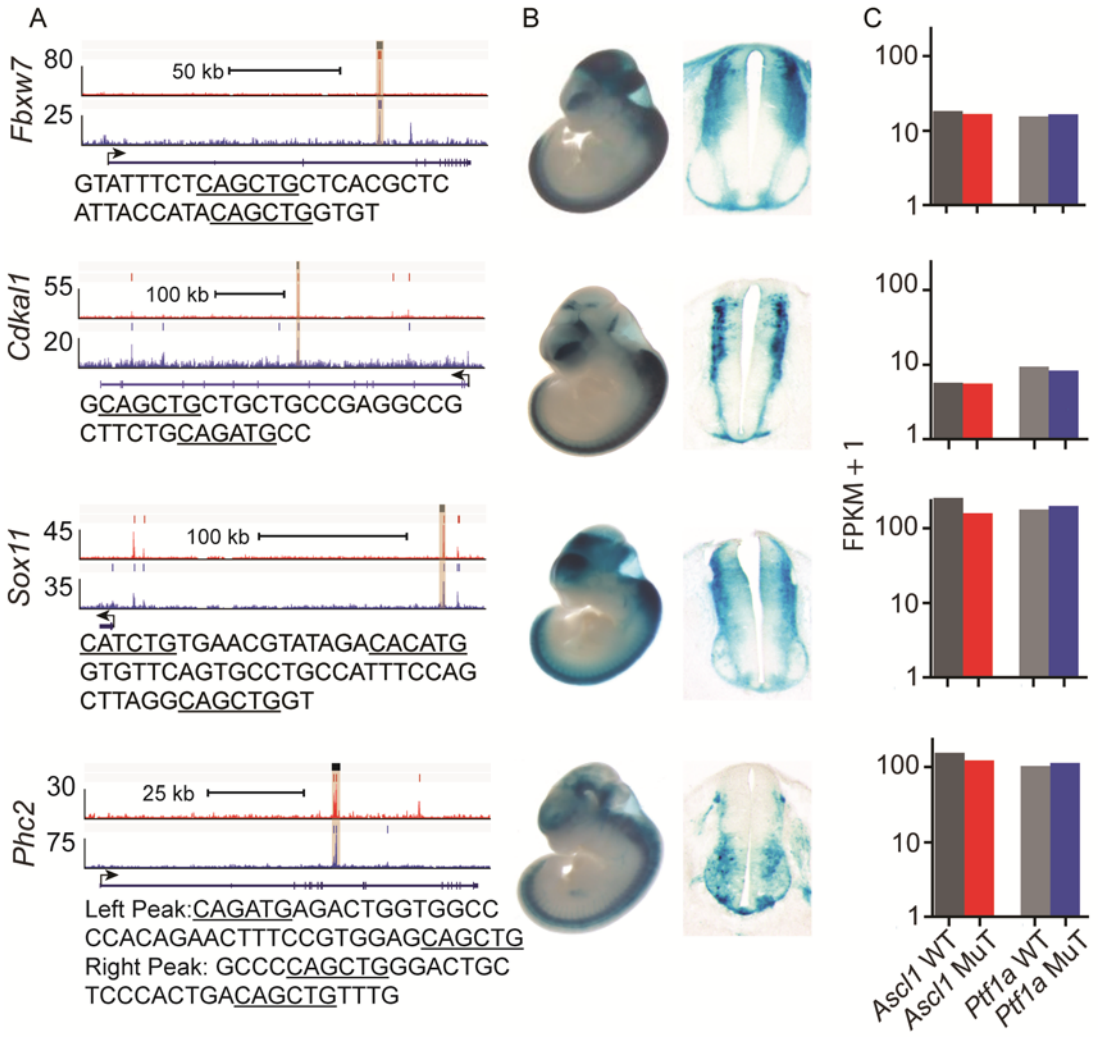


Figure 2.7. Ascl1 and Ptf1a bound genomic regions overlap with known neural enhancers. (A) Genomic regions near four genes showing Ascl1 (red track) and Ptf1a (blue track) ChIP-Seq data from E12.5 neural tube. Tick marks above each signal file indicate a called binding site determined by the program Homer. The beige box marks the DNA region tested in transgenic mice, a region identified by binding of the transcriptional activator p300 (Visel et al., 2009a, Visel et al., 2007) in neural tissue. DNA sequence near the Ascl1 and Ptf1a peak summits is shown and all E-boxes are underlined. (B) E11.5 transgenic mouse embryos testing the enhancer::LacZ transgenes are shown in whole mount (from VISTA Enhancer Browser) (Visel et al., 2007) and in cross-section through upper limb level neural tube. Although LacZ expression is neural specific it is not restricted to Ascl1 and/or Ptf1a domains. (C) Expression levels from RNA-Seq (FPKM) for the genes near these enhancers from RNA isolated from Ascl1 or Ptf1a-lineage cells in control or mutant neural tubes indicate these genes do not require Ascl1 or Ptf1a for expression.

CHAPTER THREE

Program specificity for Ptf1a in Pancreas *versus* Neural Tube Development correlates with distinct collaborating cofactors and chromatin accessibility

The work presented in this chapter was performed in collaboration with David M. Meredith, Bradford B. Casey, Trisha K. Savage, Paul R. Mayer, Kuang C. Tung, of the Jane Johnson lab, and Tye G. Deering, Chinh Hoang, Manonmani Kumar, and Galvin H. Swift of the Raymond MacDonald lab. The data presented is published in the following article:

Meredith DM*, **Borromeo MD***, Deering TG, Casey B, Savage TK, Mayer PR, Hoang C, Tung KC, Kumar M, Shen C, Swift GH, Macdonald RJ, Johnson JE. 2013 Program specificity for Ptf1a in Pancreas versus Neural Tube Development correlates with distinct collaborating cofactors and chromatin accessibility. Mol Cell Biol. PMCID: PMC23754747
*authors contributed equally

INTRODUCTION

In the first chapter I described how loss-of-function studies have demonstrated the necessity of Ptf1a both for the commitment to the pancreatic lineage from endoderm and the maintenance of the acinar pancreas, as well as for the specification of inhibitory neuron

populations in the dorsal spinal cord, cerebellum, and retina (Krapp et al., 1998, Glasgow et al., 2005, Hoshino et al., 2005, Fujitani et al., 2006, Pascual et al., 2007). In short, the absence of Ptf1a the pancreas fails to form, and Ptf1a-lineage cells instead adopt either a duodenal or biliary identity (Krapp et al., 1998, Masui et al., 2007, Kawaguchi et al., 2002). Similarly, in multiple regions of the nervous system inhibitory neurons fail to form in *Ptf1a* mutants, and Ptf1a-lineage cells switch fates to acquire an excitatory neuronal identity (Glasgow et al., 2005, Hoshino et al., 2005, Fujitani et al., 2006, Pascual et al., 2007, Storm et al., 2009). These functions of Ptf1a are dependent on a Ptf1a:Rbpj transcription complex (Masui et al., 2007, Hori et al., 2008). However, how Ptf1a in the PTF1 complex regulates different target genes in developing pancreas versus the neural tube remains unknown.

In this chapter, we again use a combination of genome-wide deep sequencing technologies to address possible mechanisms by which Ptf1a can give rise to such different tissues as the pancreas and neural tube. The majority of Ptf1a binding regions do not overlap between these tissues, and those that do tend to be near genes for non-tissue-specific developmental signaling processes. Results from experiments to distinguish between the role of lineage-specific chromatin accessibility and lineage-specific co-factors support the involvement of both mechanisms for the tissue-specific activities of Ptf1a. In particular, the forkhead factor Foxa2 is localized with Ptf1a in the pancreas-specific genomic regions, whereas Sox motifs are found near Ptf1a-bound sites in the neural tube. Further experiments testing direct interactions between Foxa2 and Ptf1a showed no such complex formation and suggest alternative mechanisms for their collaboration in pancreas development. Taken together, our results support a model whereby Ptf1a directs distinct gene expression

programs that are established by the interplay of chromatin accessibility, collaborating transcription factors, and the DNA motifs within tissue-specific regulatory enhancers.

MATERIALS AND METHODS

Chromatin Immunoprecipitation and Sequencing Library Preparation

Neural tube, telencephalon, and limb tissues from E12.5 mouse embryos, and pancreata from E17.5 embryos were dissected and placed in Buffer A (15 mM HEPES pH 7.6, 60 mM KCl, 15 mM NaCl, 0.2 mM EDTA, 0.5 mM EGTA, 0.34 M sucrose) on ice. Nuclei were liberated by Dounce homogenization and purified by centrifugation through a sucrose gradient (15 mM HEPES pH 7.6, 60 mM KCl, 15 mM NaCl, 0.1 mM EDTA, 0.25 mM EGTA, 1.25 M sucrose). Nuclei were then fixed in 1% formaldehyde for 10 minutes at 30°C, and fixation was terminated by adding glycine to a final concentration of 0.125 M. After centrifuging through another sucrose gradient, fixed nuclei were lysed in sonication buffer (1% Triton, 0.1% sodium deoxycholate, 150 mM NaCl, 50 mM Tris, 5 mM EDTA). Chromatin was sheared using a Diagenode Bioruptor for 30 minutes on high power with 30 second on:off cycles. 100 µg (pancreas) or 250 µg (neural tube, telencephalon/limb) chromatin was immunoprecipitated with either 2.4 µg affinity purified rabbit anti-Ptf1a antibody (Rose et al., 2001), 5 µl of rabbit Rbpjl antiserum (Beres et al., 2006), 5 µl of rabbit Rbpj antisera (laboratory generated), or 4 µg of goat anti-Foxa2 antibody (Santa Cruz; Hnf3b M-20, cat no. sc-6554) and Protein A/G agarose beads (Santa Cruz). Rabbit polyclonal Rbpj antibodies were generated (Covance) using the peptide Sequence (CKKK)NSSQVPSNESNTNSE, an epitope near the C-terminus of Rbpj that has no similarity to the related Rbpjl. Captured

bead:antibody complexes were washed twice with sonication buffer, three times with a high salt buffer (1% Triton, 0.1% sodium deoxycholate, 500 mM NaCl, 50 mM Tris, 5 mM EDTA), twice with LiCl (0.5% NP-40, 0.5% sodium deoxycholate, 250 mM LiCl, 10 mM Tris, 1 mM EDTA), and once with TE (10 mM Tris pH 8.0, 1 mM EDTA). Two 15 minute elutions were performed with 1% SDS, 0.1 M $NaHCO_3$, 10 mM Tris at room temperature. The immunoprecipitated DNA was then purified by QIAGEN PCR Purification kit.

Independent sequencing libraries were prepared for two E12.5 neural tube and one E12.5 telencephalon for Ptf1a and Rbpj ChIPs, and for two E17.5 pancreas for Ptf1a, Rbpjl, and Foxa2 ChIPs. In addition, an input sample from E17.5 pancreas and E12.5 neural tube were also prepared for sequencing. All libraries were made according to Illumina's ChIP-seq DNA sample preparation protocol. Amplified libraries were prepared using Invitrogen's Platinum *Pfx* polymerase for 16 cycles of 94^{o}C for 15 seconds, 62^{o}C for 30 seconds, 72^{o}C for seconds. Single-end Sequencing of 36 bp (Ptf1a, Rbpjl) or 40 bp (Rbpj, Foxa2) was conducted for all samples on an Illumina GAIIx Sequencer. Data generated by David Meredith(NT Ptf1a), Rbpj (NT Mark Borromeo), and Tye Deering (Pancreas Ptf1a/Rbpjl/Foxa2).

FAIRE and Sequencing Library Preparation

Tissue samples from E12.5 neural tube and E17.5 pancreas were dissected in cold PBS, homogenized, and crosslinked with 1% formaldehyde for 8 minutes at room temperature and fixation was terminated by adding glycine to a final concentration of 0.125 M. Samples were sonicated using a Diagenode Bioruptor for 36 minutes with 30 second on:off cycle times.

Samples were centrifuged at 14,000 rpm to pellet debris. Two rounds of phenol:chloroform extraction were performed yielding 500 µl aqueous phase, to which 8.25 µl 5M NaCl solution was added. Samples were incubated overnight at 65 ^{o}C to reverse crosslinks. Samples were ethanol precipitated and resuspended in 50 µl 10mM Tris-HCl pH 7.5, digested for 4 hours using 5 µl 10 mg/ml Proteinase K. 1 µl of 10 mg/ml RNAse A was added, and incubated an additional 30 minutes at 37 ^{o}C. Samples were purified using a Qiagen PCR Cleanup kit, and prepared for sequencing using NEBNext ChIP-Seq Library Prep Set Master Mix for Illumina and sequenced on the Illumina HiSeq2000. Data generated by Chinh Hoang.

mRNA Isolation and Sequencing Library Preparation

Mouse neural tubes were dissected from E12.5 *2.3Ptf1a-GFP* embryos (Meredith et al., 2009) into DMEM/F12 on ice and dissociated in 0.25% trypsin for 15 minutes at 37°C. Trypsin activity was quenched with 2% fetal bovine serum, and GFP positive cells were purified from the resulting single cell suspension by fluorescence activated cell sorting. Total RNA from the sorted cells was extracted and purified with Zymo's Mini RNA Isolation Kit. mRNA was isolated, reverse transcribed and amplified for sequencing with Illumina's mRNA-Seq kit. Two independent libraries were made from RNA isolated from 6 and 9 embryos, respectively. For the pancreas and transgenic embryos, total RNA was isolated using Trizol (Invitrogen). mRNA was isolated, reverse transcribed and amplified to prepare libraries for Sequencing with Illumina's Tru-Seq kit. Whole pancreata from E15.5 *2.3Ptf1a::Sox1* and wild-type littermates, and neural tubes from E12.5 *R7Ptf1a::Foxa2*

embryos and wild-type littermates were used. Two Foxa2 transgenic E12.5 neural tube libraries and three Sox1 transgenic E15.5 pancreas libraries were made from RNA isolated from two embryos that were confirmed to express each transgene. One wild-type E12.5 neural tube library was made from RNA isolated from three embryos, and two wild-type E15.5 pancreas libraries were made from RNA isolated from three pancreata. Data generated by David Meredith (NT), and Tye Deering (Pancreas)

Bioinformatics

Peak Calling, Quantification, and Intersections

ChIP-Seq and FAIRE-Seq sequence reads were mapped to the mm9 assembly of the mouse genome with Bowtie (Trapnell et al., 2009). Duplicate reads were removed, and the remaining unique reads were normalized to a 10 million read count for peak calling analysis. Peak calling was performed by HOMER (Heinz et al., 2010) using an FDR cutoff of 0.001. Putative peaks were required to have 4-fold enrichment over the control/input sample and a cumulative Poisson p-value of <0.0001. All pancreas ChIP-Seq samples were compared to the pancreas input, while the neural tube RbpJ samples were compared to the neural tube input. Neural tube Ptf1a ChIP-Seq samples were compared to a Ptf1a ChIP-Seq sample performed in the telencephalon (a non-Ptf1a expressing neural tissue of similar developmental stage to control for non-specific background from the antibody). This control was more stringent than using neural tube input chromatin but similar results were found with both methods. For downstream analysis, only ChIP-Seq peaks that were called in both biological replicates were used. Common binding sites were required to be within 150 bp of each peak center. Quantification of the data at peak regions was performed by HOMER and

loaded into Multiple Experiment Viewer (MeV) v4.8.1 (Saeed et al., 2003) to generate heat maps. All sequencing datasets have been submitted to GEO (Accession #'s in process).

RNA-Seq Read Alignment and Expression Level Estimation

Sequence reads were aligned to the mm9 build of the mouse genome using TopHat v2.0.4 (Trapnell et al., 2009). Default settings were used with the following exceptions: -G option (instructs TopHat to initially map reads onto a supplied reference transcriptome) and –no-novel-juncs option, which ignores putative splice junctions occurring outside of known gene models. If a biological replicate was available, then it was specified and used to build an expression level model determined by the FPKM method of Cuffdiff v2.0.2 (Trapnell et al., 2010). Upper quartile normalization (-N option) was performed to more accurately estimate levels of low-abundance transcripts, and multiple read correction (-u option) was selected to better distribute reads mapping to multiple genomic locations. To correct any sequence-specific bias introduced during the library preparation option –b was used. All other settings were left at default values.

GO Classification and ChIP-Seq Peak Gene Annotation

Gene ontologies and gene annotations for ChIP-Seq peaks were obtained using GREAT v1.82 (McLean et al., 2010). GREAT assigns a gene to a binding region if the region falls within 5 kb 5' or 1 kb 3' of the transcription start site (basal region) with a maximum extension of 1,000 kb in either direction. If the binding region falls within the basal region of

multiple genes, then more than one assignment is made. All parameters were left at their default settings.

Motif Discovery and Occurrence Probability

Peak regions were first trimmed to 150 bp centered about the peak summit to increase the likelihood of finding motifs associated with binding and to decrease computational burden. Motif discovery was conducted with the HOMER package v3.8.1 using the following settings: -size 150 –S 10 –len 8,10,12 –bits. Motif density plots were generated in HOMER (annotatePeaks.pl) by identifying all matches to HOMER generated motif matrices with 1 kb surrounding peak regions.

Half Site Spacing Analysis

Locations and numbers of enriched motifs were obtained by HOMER (annotatePeaks.pl). Motif spacing was calculated from midpoint to midpoint between any E-box and any TC-box, Sox, Hox, Fox, or GATA motif found in the same peak. Histogram frequencies were then normalized to the total number of spacings found within each ChIP-Seq dataset for better comparison.

Transgenic Mice and Tissue Processing

LacZ-reporter transgenes were generated by cloning PCR fragments (~1 kb) from mouse genomic DNA for the regions listed in Supplementary Table 3 5’ of the basal promoter of a *lacZ* reporter plasmid. The *lacZ* reporter contains the basal promoter from the rat *Ela1* gene, an SV40 nuclear localization signal, the *lacZ* coding sequence, and the 3’ untranslated region

plus polyA addition signal from bovine growth hormone, as well as an insulator sequence from the *Gallus gallus* HS4 gene (Masui et al., 2008). Transgenic constructs for misexpressing Foxa2 or Sox1 were generated using regulatory sequences from the *Ptf1a* locus. *R7Ptf1a::Foxa2* transgene contains a 1.2 kb enhancer 3' of the *Ptf1a* gene called R7 that has the same activity in transgenic mice as 12.4 kb *Ptf1a* enhancer previously described (Meredith et al., 2009) (and unpublished data). R7 (chr2:19380202-19381456 from mm9) was cloned upstream of the *β-globin* basal promoter driving mouse *Foxa2* cDNA followed by a bovine growth hormone polyA signal sequence. *2.3Ptf1a::Sox1* transgene contains the 2.3 kb autoregulatory enhancer region (chr2:19351717-19354014 from mm9) (Masui et al., 2008) cloned upstream of the *β-globin* basal promoter driving mouse *Sox1* cDNA followed by a bovine growth hormone polyA signal sequence.

Transgenes were prepared for microinjection using Elutip-d columns (Schleicher & Schuell). Transgenic embryos were generated by pronuclear injection of linear transgene DNA into mouse eggs by the UT Southwestern Transgenic Core Facility. All housing, husbandry, and procedures were performed in accordance with NIH guidelines, under protocols approved by the Institutional Animal Care and Use Committee at UT Southwestern Medical Center. Data was generated by Bradford Casey, Galvin Swift, and Yanjie Chang

For *lacZ* reporter gene expression, embryos were harvested at E12.5 or E15.5, and yolk sacs PCR genotyped using primers directed against the *lacZ* gene. Primers used were 5'-TGCTGCTGGTGTTTTGCTTCC – 3' and 5' – CGATATTATTTGCCCGATGTA – 3'. E12.5 embryos were immersion fixed in 4% formaldehyde/PBS for 30 minutes, then washed in PBS before transferring to an X-gal staining solution overnight. E15.5 embryos were

dissected to remove the head and the contents of the peritoneum, and fixed for 1 hour, washed in PBS before transferring to an X-gal staining solution overnight. X-gal staining solution consists of 5 mM potassium ferrocyanide trihydrate, 5 mM potassium ferricyanide, 2 mM $MgCl_2$, 1 mg/ml X-gal in PBS. Embryos/tissues were washed in PBS, post-fixed in 4% formaldehyde/PBS at 4^oC before embedding for cryosection.

Cell transfection and transcription reporter assays

HEK293 or pancreatic 266-6 cells were transfected with plasmid DNA using FuGene 6 (Promega) as described previously (Masui et al., 2008). Transfection results reflect at least four independent experiments assayed in duplicate. Beta-galactosidase levels expressed by reporter genes were normalized according to the levels of Renilla Luciferase from the co-transfected pRL-CMV (Promega). To permit transfections performed at different times to be analyzed together in the HEK293 experiments, final results were normalized according to the level of the positive control transfections of pCMVbeta (Clontech). Results are presented as normalized relative light units (nRLU). Transcription factor expression vectors used the CMV promoter to direct expression of mouse Ptf1a (in pcDNA3.1) (Masui et al., 2007) and mouse Foxa2 (in pCIG) (Megason and McMahon, 2002). For reporter constructs, one to three repeats in tandem of the tested motifs (sequences in Fig. 7C) were cloned adjacent to the promoter of the *lacZ* reporter plasmid described above for mouse transgenes. Empty pcDNA3.1 vector was used to keep the total amount of expression vector and CMV promoter constant in each transfection. Data was generated by Galvin Swift.

Electrophoretic Mobility Shift Assay

Proteins were synthesized *in vitro* (IVTT) using the TnT Coupled Reticulocyte System (Promega). DNA probes were radiolabeled with (γ-^{32}P)-ATP using T4 polynucleotide kinase (New England Biolabs). Binding reactions (20 µl) were incubated for 30 minutes at 30°C and setup as follows: 5 µl 4X binding buffer (20% glycerol, 40 mM HEPES-NaOH pH 7.9, 16 mM Tris-HCl pH 8.0, 320 mM NaCl and 4 mM EDTA pH 8.0), 1 µl 100 mM DTT, 2 µl 800 ng/µl poly-dI/dC, 1-5 µl IVTT protein extract(s) and 1 µl labeled probe (20,000 CPM). Binding reactions were separated on 4.5% polyacrylamide gel by applying a constant current of 1.5 mA/cm for 2 hours at 4°C; gels were prepared using 1X running buffer: 50 mM Tris-HCl, 380 mM glycine and 2.8 mM EDTA, and were pre-run for 15 minutes at a constant current of 0.75 mA/cm at 4°C. Data was generated by Paul Mayer.

Co-immunoprecipitation

HEK293T cells were seeded at a concentration of 8×10^5 cells/6 cm dish. The following morning, cells were co-transfected as indicated using FuGene 6 (Roche Life Sciences). The cells were washed with ice-cold PBS and collected 48 hours after transfection. Cell pellets were dissolved with 5 pellet volumes of lysis buffer (50 mM HEPES-NaOH pH 7.4, 5 mM $MgCl_2$, 100 mM KCl, 1 mM DTT, 0.2% Triton X-100, 1 mM PMSF and EDTA-free protease inhibitor cocktail (Roche Life Sciences)) and homogenized by passing through a 25*g* needle 5 times. After resting on ice for 20 minutes, lysates were cleared by centrifugation at 20,000 *x* g for 20 minutes at 4°C. Anti-Ptf1a antibody (Hori et al., 2008) was conjugated to protein A/G beads (Santa Cruz) in lysis buffer by rotating for 2 hours at

4°C. 100 µl of the resulting antibody/bead mixture was added to cleared lysate and rotated overnight at 4°C. Beads were washed 4 times with lysis buffer and eluted with a small volume of 2X SDS-PAGE sample buffer (30 mM Tris-HCl pH 6.8, 1% SDS, 5% glycerol, 0.0125% bromophenol blue and 100 mM DTT). Co-immunoprecipitation was assessed by western blot using anti-myc (A-14, Santa Cruz, 1:2000) and anti-Foxa2 (#3143, Cell Signaling Technology, 1:1000) antibodies. Note that the heavy-chain fragment (approximately 50 kDa) from the anti-Ptf1a antibody used for immunoprecipitation co-migrates with Foxa2 and weakly cross-reacts with the HRP conjugated secondary antibody used for detection. Data was generated by Paul Mayer.

RESULTS

Ptf1a is bound largely to non-overlapping genomic sites in the pancreas and neural tube

An intuitive explanation for the divergent functions of Ptf1a in pancreatic and neural development is that it binds to distinct enhancers and, thus, regulates tissue-specific gene programs. To test this, we performed genome-wide chromatin immunoprecipitation sequencing (ChIP-Seq) to identify chromatin sites bound by Ptf1a in the E12.5 neural tube and E17.5 pancreas. Using ChIP-Seq data from two replicates each from neural tube and pancreas we identified genomic regions with enriched Ptf1a binding (peaks) signifying potential Ptf1a-regulated enhancers. Taking the peaks present in both replicates in a given tissue, our analysis revealed 2,241 Ptf1a-bound sites in the neural tube and 9,985 in the pancreas, of which 501 were common to both tissues (Fig. 3.1A,B). Our approach for

defining a Ptf1a binding site in this chapter is more conservative than the approach we took in the previous chapter; thus, the number of Ptf1a binding sites in the neural tube is less. In addition, the neural tube ChIP had a lower signal to noise probably due to the heterogeneity in the tissue used relative to the pancreas sample, and thus is likely an underestimate of the number of Ptf1a-bound sites in neural tissue. Nevertheless, a majority of genomic sites bound by Ptf1a are distinct between neural tube and pancreas. This conclusion holds true even when reducing the stringency of the peak calling.

To gain more insight into the characteristics of Ptf1a binding, we assigned peaks to genes using the algorithm GREAT developed for ChIP-Seq data (McLean et al., 2010). Notably, many genes have more than one associated Ptf1a-bound region (Fig. 3.1C), and the overwhelming majority of Ptf1a-bound regions fall well distal to their assigned genes (i.e. 5-500 kb either upstream or downstream from the transcription start sites) (Fig. 3.1D) (Thompson et al., 2012). These findings are consistent with other recent studies for site-specific binding factors (McLean et al., 2010). Gene ontology analysis from GREAT showed that genes assigned to neural tube restricted binding sites are significantly enriched for neural processes such as the development of spinal cord and eye, and axon guidance. Similarly, genes assigned to pancreas-specific sites are heavily enriched for processes of glandular and exocrine development and differentiation, including enhancements for protein synthesis, endoplasmic reticulum functions, intracellular protein trafficking, and the response pathways for unfolded secretory proteins. Target genes common to both tissues are involved in more general functions, including cell proliferation and regulation of signal transduction (Fig. 3.1E).

Ptf1a binds similar consensus motifs in the neural tube and pancreas

Ptf1a exists in a trimeric complex with an E-protein and Rbpj or Rbpjl, and it is these complexes that are required for transcription activation and the development of the neural tube and pancreas (Hori et al., 2008, Beres et al., 2006, Masui et al., 2007). From previous studies largely using select gene targets in the pancreas, the trimeric form of Ptf1a (PTF1) was found to bind to sequences containing an E-box (CANNTG) and a TC-box (YTYYCA) separated (midpoint to midpoint) by one or two helical turns (Beres et al., 2006).

One possible explanation for the tissue-specific binding preferences of Ptf1a is that the PTF1 complex binds distinct DNA motifs in the neural tube and pancreas. To address this, we analyzed 150 bp surrounding the apex of the Ptf1a peaks for motifs using HOMER (Heinz et al., 2010). For both tissues, we identified an E-box as the most enriched motif, with only subtle sequence preferences in the two tissues (Fig. 3.2A; also see Fig. 3.5). Both tissues have a strong prevalence of the palindromic CAGCTG E-box; however, the E-box CATCTG/CAGATG is enriched in neural tube binding regions, while CACCTG/CAGGTG E-boxes are enriched in the pancreas (Fig. 3.2A and App. C). The second most enriched motif was the cognate Rbpj/Rbpjl binding sequence, or TC-box, which also exhibited only slight differences in each tissue. The neural tube TC-box, YTSYCA, shows greater degeneracy at the third position than does the pancreas TC-box, YTCYCA (Fig. 3.2A). The TC-box from Ptf1a-bound regions shared between the two tissues, TTCYCA, shows even less degeneracy. Because the differences represent preponderance rather than the absolute range of variations, which is the same among the three groups, it is unclear whether these

subtle tissue-specific preferences affect the specificity of in vivo binding or function of PTF1.

To investigate whether the previously proposed spacing constraint for 1 or 2 helical turns between the E-box and TC-box applied to all Ptf1a-bound sites, we determined the spacing of all E-box and TC-box motifs in each peak measuring the distance midpoint to midpoint. The most frequent spacing in both tissues occurs around 11 and 22 bp, consistent with previous studies (Fig. 3.2A) (Cockell et al., 1989, Beres et al., 2006, Masui et al., 2008, Henke et al., 2009b, Meredith et al., 2009). The Ptf1a peaks in pancreas demonstrated enrichment for a 3 helical turn spacing as well, consistent with a recent study using the acinar cell line 266-6 (Thompson et al., 2012). A majority of Ptf1a peaks in each group (pancreas only 67%, neural tube only 53%, and common 81%) have at least one E-box:TC-box with the PTF1 constrained spacing (Fig. 3.2A, pie charts). It is likely that this analysis underestimates bona fide PTF1 sites given PTF1's tolerance of motif degeneracy (Beres et al., 2006). Taken together, the genome-wide analysis of Ptf1a binding alone in the two tissues confirms the E-box/TC-box with spacing constraints as previously predicted for Ptf1a in the pancreas. No novel spacing preference between E-box and TC-box in either tissue was revealed.

Many genomic regions bound by Ptf1a are also bound by Rbpj or Rbpjl

The motif analysis of Ptf1a-bound sequences implies Rbpj and Rbpjl binding at many of these same sites, which is also expected given the apparent requirement of Rbpj or Rbpjl for the transcriptional activity of Ptf1a (Beres et al., 2006, Masui et al., 2007, Hori et al., 2008).

To assess this genome-wide, we performed ChIP-Seq for Rbpj and Rbpjl with chromatin from the neural tube and pancreas, respectively. Rbpjl was used for pancreas rather than Rbpj since it is the dominant binding partner for Ptf1a at E17.5 in the pancreas (Masui et al., 2007), and ChIP-Seq for Rbpj at this stage identified fewer than 1,000 peaks making comparative analysis more difficult. Using similar analyses as above, we identified 11,678 Rbpjl binding sites in the pancreas at E17.5 and 3,035 significantly enriched Rbpj binding sites in the neural tube at E12.5 (Fig. 3.2B,C). As predicted, most (71%) Ptf1a-bound regions in the pancreas displayed co-localization with Rbpjl (Fig. 3.2B). Significant co-localization between Ptf1a and Rbpj occurred for the neural tube although with decreased frequency (25% of Ptf1a peaks) (Fig. 3.2C), consistent with the lower frequency of E-box/TC-boxes with constrained helical turn spacing found by the motif analysis. The stringent criteria for peak calling discards many low affinity Ptf1a and Rbpj/Rbpjl binding events, and visual inspection of the heat maps (Fig. 3.2B,C) suggests that the overlap is likely greater than 25% for neural tube and may approach 100% for pancreas. Together these results assessing Ptf1a and Rbpj/Rbpjl binding genome-wide support the broad existence of the PTF1 complex in both pancreas and neural tube, although the decreased frequency in the neural tube suggests a possible Rbpj-independent role for Ptf1a in the nervous system.

Tissue-specific Ptf1a binding correlates with differences in tissue-specific chromatin accessibility

Chromatin accessibility as defined by the epigenetic landscape is known to change over developmental time from stem cell stages to differentiated cell types in mature tissues (Ong

and Corces, 2011, Ladewig et al., 2013). Chromatin accessibility is a likely candidate for explaining the distinct Ptf1a binding in the pancreas and neural tube chromatin given the different developmental histories for these tissues. To determine if chromatin accessibility differences correlate with the ability of Ptf1a to bind its recognition motif in the two tissues, we performed formaldehyde-assisted identification of regulatory elements followed by deep sequencing (FAIRE-Seq) (Giresi et al., 2007). In this procedure, open chromatin regions are preferentially isolated and subsequently sequenced. We performed FAIRE-Seq with E12.5 neural tube and E17.5 pancreas chromatin (Fig. 3.3). A heat map of Ptf1a-bound sites designated as neural tube only, pancreas only, or in common in the two tissues is shown next to the FAIRE-Seq signals at these sites in the two tissues. Sites bound by Ptf1a in each tissue show concordant FAIRE enrichment, as the chromatin accessibility model predicts. In the pancreas-only Ptf1a-bound regions, FAIRE-Seq indicated 53% of these were ‘open’ (>5 sequence tags) in pancreas and only 10% were ‘open’ in the neural tube. Conversely, in the neural tube-only Ptf1a-bound regions, FAIRE-Seq indicated 75% were ‘open’ in the neural tube compared to 15% in the pancreas. Similarly, Ptf1a-bound sites in common between the two tissues are at regions with high FAIRE in both tissues with 61% of the sites ‘open’ in both tissues. This is illustrated both globally (Fig. 3.3A) and for specific examples around *Klf7* (neural tube only), *Cela3b* (pancreas only), and *Fbxl19* (common target) (Fig. 3.3B, also see Fig. 3.4). Thus, chromatin accessibility is at least part of the mechanism of tissue-specific Ptf1a binding in the genome.

The Ptf1a-bound sites with flanking sequences retain tissue-specific activity

To determine whether genomic regions bound by Ptf1a and Rbpj/Rbpjl represent active enhancers that retain their tissue-specificity out of the normal genomic context, we tested the ChIP-Seq identified regions in *lacZ* reporter constructs for activity in transgenic mouse embryos. One such region located between *Kirrel2* and *Nphs1* is occupied by PTF1 complexes in both the neural tube and pancreas (Fig. 3.4A). This region contains two consensus PTF1 motifs; both are required for reporter expression in vitro (Nishida et al., 2010, Mizuhara et al., 2010). At E12.5, the *Kirrel2/Nphs1* enhancer faithfully recapitulates endogenous *Ptf1a* expression, as well as that of *Kirrel2* and *Nphs1* (Volker et al., 2012, Mizuhara et al., 2010, Nishida et al., 2010, Minaki et al., 2005, Putaala et al., 2001) in both dorsal neural tube and in pancreas (Fig. 3.4B-C, App. D). Similarly, reporter expression is detected in the correct *Ptf1a*-expression domains in the developing cerebellum and medulla, and in the pancreas, at E15.5 (Fig. 3.4D-D'', App. D).

We next examined whether regions with tissue-specific Ptf1a binding in vivo were capable of directing tissue-restricted expression in transgenic animals despite the similarity in the consensus PTF1 motifs found in these regions. In this fashion, we hoped to address whether the intrinsic sequence, and hence the binding of additional tissue specific factors contributes to the distinct Ptf1a binding in the two tissues. Regions chosen for testing had strong Ptf1a binding in the neural tube and no significantly enriched binding in the pancreas and vice versa (App. D). Additionally, genes near each tested region had expression only within the tissue in which there was Ptf1a binding, as determined by RNA-Seq analyses from

each tissue (App. D). Preference was also given to regions with significant evolutionary conservation and the presence of a PTF1-binding motif.

Of the six neural tube-specific regions tested in this study, three had reproducible activity in the dorsal neural tube at E12.5 (Fig. 3.4E-G, App. D-E). Histologic analysis of these embryos revealed reporter staining in the correct Ptf1a domain in the ventricular zone, as well as in more mature neurons of this lineage in the mantle zone (Fig. 3.4F-G', App. E). No pancreatic expression was detected in any of these embryos (Fig. 3.4F''-G'', App. E). Similarly, of the eight pancreas-specific regions tested, five efficiently drove reporter expression within the pancreas at both E12.5 (Fig. 3.4H-J', App. E) and E15.5 (Fig.3. 4I''-J'', App. D-E). Only one region, that upstream from *Dlk1*, also directed expression in Ptf1a domains in the nervous system at E12.5 (App. E). The tissue-specific activity of Ptf1a-bound regions in vivo, therefore, is intrinsic to the DNA sequence and is not due to the imposition of chromatin architecture by the flanking regions of the endogenous loci.

Lineage-specific factor-binding motifs co-localize with Ptf1a-bound chromatin in pancreas and neural tube

The intrinsic tissue-specific activity of Ptf1a-bound regions strongly suggests that other tissue-restricted factors are needed for pancreas versus neural tube specific PTF1 functions. To investigate this possibility, we searched for other enriched motifs within Ptf1a-bound regions. Indeed, besides the E-box and TC-box, which were found with the highest frequency, HOMER identified additional enriched motifs that are distinct for the two tissues (Fig. 3.5).

For the neural tube, the consensus Sox family binding site ACAAWG (van Beest et al., 2000, Maruyama et al., 2005) was present in 22% of Ptf1a-bound regions, and the consensus Hox family binding site RTTAATY (Pellerin et al., 1994) occurred in 19% (Fig. 3.5A). These consensus sites are not enriched in the pancreatic Ptf1a-bound regions. To determine if these motifs are indicative of additional or alternative binding partners for Ptf1a, we examined whether they occurred at stereotyped intervals from the E-box. Unlike the E-box and the TC-box, which are centered around the apex of the Ptf1a peak and are enriched in the 1 and 2 helical spacing (Fig. 3.2A, 3.5B), the Sox and Hox motifs are randomly spread across the peak region and have no apparent spacing constraints relative to the E-box (Fig. 3.5B,C). These results suggest Sox and Hox factors are not forming DNA-bound complexes with Ptf1a and Rbpj that have strict stereochemical constraints, but rather may be serving a more general role to stabilize the region in a neural-specific open chromatin state.

In contrast, in the Ptf1a-bound regions specific to the pancreas, the next most enriched motifs after the E-box and TC-box were a consensus forkhead motif (Fox) RYMAAYA (Benayoun et al., 2008) found in 16% of the peaks, and a GATA family motif GATAA in 12% of the peaks (Fig. 3.5D). Like the Sox and Hox motifs in the neural tube Ptf1a peaks, the GATA motif was randomly spread across the peak region and had no spacing constraint relative to the E-box. However, the Fox motif was also enriched at the center of the peak with the E-box and TC-box (Fig. 3.5E). Notably, a compound motif with the E-box:Fox with a single base-pair separation 3’ of the E-box was identified (Fig. 3.5F). Neither this compound motif nor the Fox motif alone occur at any particular interval with respect to the TC-box (data not shown), suggesting that Ptf1a, and not Rbpjl or the PTF1

complex, may form a novel stereochemically constrained complex with a pancreas-specific forkhead protein.

Foxa2 binds to the same genomic regions as Ptf1a without forming a stable complex with PTF1

The presence of the E-box:Fox motifs with the constrained spacing in 430 (4.3%) of the pancreatic Ptf1a-bound regions led us to hypothesize that a forkhead factor might bind cooperatively with Ptf1a on such sites. Among forkhead family members present in the pancreas at E15.5 (App. F), we chose Foxa2 as the best candidate for interaction with Ptf1a based on its high level of expression and its role in pancreatic development (Gao et al., 2008, Lee et al., 2005b). To determine if Foxa2 localizes to the same genomic regions as Ptf1a in vivo, we performed ChIP-Seq for Foxa2 with chromatin from E17.5 pancreas. Surprisingly, Foxa2 localizes to at least 31% of all Ptf1a localized regions, and 38% of PTF1 sites defined by Ptf1a and Rbpjl overlapping peaks (Fig. 3.6A,B). An example of a site with co-localized Foxa2, Ptf1a, and Rbpjl is shown for the *Gata4* locus (Fig. 3.6C). GO analysis reveals that genes near regions bound by Ptf1a, Rbpjl, and Foxa2 function generally in organogenesis, such as stem cell development, branching morphogenesis, and Notch and RhoA signaling (Fig. 3.6D).

The ChIP-Seq results demonstrate that Ptf1a, Rbpjl, and Foxa2 localize to many of the same genomic regions, but they cannot determine if these factors are present at these sites at the same time. To test whether Foxa2 binds cooperatively with Ptf1a to the compound E-box:Fox motifs, we performed EMSAs with several sequences present in pancreatic Ptf1a-

and Foxa2- bound regions. Foxa2 and PTF1 (Ptf1a:E12:Rbpj) bound the motifs separately, but no evidence of a higher order complex between Foxa2 and the PTF1 trimer or Ptf1a:E12 dimer was detected indicating PTF1 and Foxa2 bind DNA independent of one another (Fig. 3.7A,B, and data not shown). Furthermore, co-immunoprecipitation studies from transfected cells failed to demonstrate an association of Foxa2 and Ptf1a (Fig. 3.7D). Thus, there is no evidence for a stable complex containing Foxa2 and Ptf1a in vitro.

Three compound sites that bind Ptf1a, Rbpjl and Foxa2 in close proximity were selected from genomic control regions that direct pancreas-specific expression (*Clps*, *Dlk1*, and *Rbpjl* regulatory regions; see App. D) and examined for their ability to activate transcription in response to Ptf1a and Foxa2 in cell transfection assays (Fig. 3.7C,E,F). These experiments provided insight into how Ptf1a and Foxa2 work together to activate pancreas-specific target genes. First, in mouse pancreatic 266-6 cells, which express endogenous PTF1 and Foxa2, all three PTF1-Fox compound sites could activate reporter gene transcription (Fig. 3.7E). Activity of the Fox-PTF1 site from the *Clps* control region in particular required all three binding motifs (E-box, TC-box, and Fox).

We then used HEK293 cells, which do not contain endogenous Ptf1a or Foxa2 but contain ample Rbpj and E-proteins (Sathira et al., 2010), to test the individual versus joint contribution of these factors. Each compound site responded to Ptf1a and Foxa2 differently. For the Fox-PTF1 site from the *Clps* regulatory region, Ptf1a and Foxa2 had strong synergistic activity (Fig. 3.7F). Importantly, this activity required the presence of each individual motif, just as for 266-6 cells. In contrast, the addition of Foxa2 had no effect on the Ptf1a induced activity of the Fox-PTF1 site from the *Dlk1* regulatory region. And finally,

Foxa2 inhibited the Ptf1a activation of the Fox-PTF1 site from the *Rbpjl* regulatory region (Fig. 3.7F). Thus, Ptf1a and Foxa2 can work together in different ways to provide distinct regulatory outcomes, depending on the nucleotide sequence of compound sites.

Our results demonstrate that Foxa2 and Ptf1a bind in close apposition in many genomic regions in vivo, including hundreds that contain a compound E-box:Fox motif. It appears, however, that no stable physical interaction between these factors occurs either in solution or on DNA, indicating independent rather than cooperative binding and the possibility of sequential or alternating binding. In addition, the effect of Foxa2 on Ptf1a-mediated transcription is variable and may depend on the relative spacing or precise sequences of their binding motifs in compound sites.

Foxa2 and Sox1 are not sufficient in combination with Ptf1a to activate tissue-specific gene expression programs

We tested whether Foxa2 or Sox1 might collaborate with Ptf1a to control pancreas and neural programs, respectively, given the enrichment of motifs for these factors in Ptf1a-bound regions. Transgenic embryos were generated that misexpressed Foxa2 in Ptf1a-expressing cells in the dorsal neural tube, or Sox1 in Ptf1a-expressing cells in the developing pancreas. In this paradigm, the ectopic expression of Foxa2 or Sox1 occurs simultaneously with the endogenous Ptf1a rather than prior to Ptf1a as would happen during normal development. Changes in gene expression as a consequence of misexpressing Foxa2 or Sox1 were assayed by RNA-Seq. We found that the combination of Foxa2 and Ptf1a in neural tissue, or Sox1 and Ptf1a in the pancreas used in this paradigm is not sufficient in vivo to

fully reprogram neural tissue to pancreas, or pancreas to neural tissue, respectively (data not shown). The effects are most supportive of a model whereby the lineage-specific factors Foxa2 and Sox1, normally expressed prior to Ptf1a, generate lineage-specific chromatin accessibility that allows Ptf1a to bind when it appears at later developmental stages (see model Fig. 3.8 and Discussion).

DISCUSSION

Distinct binding of Ptf1a within chromatin of neural tube versus developing pancreatic tissue

The bHLH transcription factor Ptf1a is critical to the development of both the pancreas and nervous system. In this study, we address a fundamental question in developmental biology of how one transcription factor can control diverse gene expression programs by assessing in vivo Ptf1a binding genome-wide in the two developing tissues. We demonstrate that genomic occupancy of Ptf1a in the pancreas and neural tube are largely distinct, yet there is little difference in the most common binding motifs in these regions. These motifs contain an E-box and a TC-box separated by 1, 2, or 3 helical turn spacing as previously predicted by a select set of targets identified in the pancreas, and in pancreatic 266-6 cells (Thompson et al., 2012, Beres et al., 2006). Thus, the broad existence of this compound E-box:TC-box motif was shown in vivo in pancreas and in neural tissue. Despite the use of an alternate binding partner in the pancreas (Rbpjl, which recognizes the identical TC-box consensus sequence), the motifs bound by the PTF1 trimer in the pancreas and neural tube are nearly identical. Indeed, Rbpj can compensate for the absence of Rbpjl in E17.5 and adult pancreatic acinar

cells (Masui et al., 2010). Although we detected subtle differences in E-box and TC-box preferences between neural tube and pancreas, it seems unlikely that such subtleties alone could dictate the observed binding profiles of Ptf1a. Indeed, several of the first validated PTF1 sites near exocrine pancreatic genes possess widely degenerate sequences (Beres et al., 2006).

Although the PTF1 motif was found in a majority of Ptf1a-bound sites, there were still many where HOMER did not detect a TC-box, particularly in the neural tube. While visual inspection often identified potential TC-boxes not detected by HOMER, there remain numerous examples where no clear PTF1 motif or TC-box is evident. This suggests that Ptf1a may serve another function independent of the PTF1 complex at least in neural tissue. Since the Ptf1a/E-protein heterodimer does not activate transcription (Beres et al., 2006), it is possible that Ptf1a competes at enhancers used by other E-box binding proteins near genes involved in opposing programs.

Ptf1a-bound regions define tissue-specific enhancers

Ptf1a-bound sites restricted to either tissue define tissue-specific enhancers in transgenic mice. This result was not necessarily predicted because transgenes integrate in non-homologous loci where the lineage-specific chromatin landscape is presumably absent. The alternative was that the enhancers would be inactive or their specificity lost, and thus be active in both neural tube and developing pancreas where there is endogenous Ptf1a. The fact that this did not occur indicates that information present in the transgene, including the Ptf1a-bound site and surrounding sequence, is sufficient to retain the tissue specificity. Motif

analysis with HOMER identified multiple tissue-specific consensus motifs and suggested that additional collaborating factors may provide such specificity. Fox and GATA motifs were enriched specifically in the Ptf1a peaks restricted to the pancreas, whereas Sox and Hox motifs were enriched specifically in the Ptf1a peaks restricted to the neural tube. Factors in these families, such as Foxa2 and Gata4/6, or the SoxB1 family, represent known lineage-defining factors for pancreas and neural tissue, respectively (Xuan et al., 2012, Bergsland et al., 2006, Lodato et al., 2013). The simplest model currently is that factors such as Foxa2 and Gata4 combine to define accessible sites within the chromatin allowing Ptf1a in the PTF1 complex to bind and activate transcription of its targets in pancreas (Xu et al., 2011). Conversely, SoxB1 and Hox family factors combine to define sites accessible to PTF1 in neural tissue (Bailey et al., 2006, Lodato et al., 2013) (see model Fig. 3.8). The transgenes chosen originally based on Ptf1a binding also possess consensus sites for these collaborating factors and potentially confer tissue-specificity on these enhancers. That artificial co-expression of Foxa2 or Sox1 with Ptf1a was not sufficient to direct Ptf1a to regulate pancreas versus neural targets in transgenic mice suggests that additional factors are also required, or that the correct temporal relationship of expression must also be respected.

Not all Ptf1a regions tested for enhancer activity were sufficient to direct expression in transgenic mice. Since Ptf1a often binds multiple regions near genes it regulates, it may require more than one of these enhancers for detectable activity. Similar observations have been made with Crx in the adult mouse retina, in which Crx binding regions cluster around genes important for photoreceptor identity (Corbo et al., 2010). When tested individually, many of these regions were not able to direct reporter expression; however, several such

bound regions in the *Rho* locus dramatically increased the activity of a functioning enhancer when tested in combination. The use of alternate enhancers to regulate critical development genes has been well described in *Drosophila* (Hong et al., 2008, Frankel et al., 2010), implying that the combinatorial action of cis-regulatory elements in modulating target levels is a widespread and highly conserved phenomenon. Perhaps in more complicated genomes there exist many partial enhancers necessary for proper gene regulation that are insufficient on their own to induce gene transcription. If this is true, then it could account for the fact that not all Ptf1a-bound regions we tested functioned individually as enhancers.

Lineage-specific chromatin landscapes correlate with lineage-specific binding of Ptf1a

It has been appreciated for many years that the chromatin landscape changes as development proceeds and cells become restricted to specific lineages (Ladewig et al., 2013, Ong and Corces, 2011). We assessed differences in chromatin accessibility between embryonic pancreas and neural tube tissue using FAIRE-Seq. This analysis illustrated the differences in these two tissues and demonstrated that Ptf1a often binds to chromatin defined as open by this assay, consistent with a model whereby the chromatin landscape defines which PTF1 sites are available for binding. Although from these data alone we cannot distinguish whether Ptf1a binding is a consequence of the availability of its motif or whether Ptf1a can direct the opening of the chromatin, the fact that Ptf1a binding is tissue-specific suggests the former. Furthermore, based on previous studies with Foxa2 and Sox factors, it is more likely that these factors prepare specific regions of chromatin in early developmental progenitor cells

with broad potency in anticipation of binding of more restricting lineage factors – such as Ptf1a – later (Fig. 3.8) (Xu et al., 2011, Bailey et al., 2006).

Seminal work by Zaret and colleagues has elucidated the pioneering function of Foxa2 in liver development, in which Foxa2 acts as a place holder to preserve access to critical regulatory regions for specification factors (summarized in (Zaret and Carroll, 2011)). Our observations are consistent with a similar role for Foxa2 in pancreatic development. Although Foxa2 and Ptf1a bind many of the same regions in vivo, and there is a compound motif with a constrained spacing of the E-box and Fox motif, we were unable to detect any physical interaction between Foxa2 and Ptf1a either in solution or on DNA. These findings favor a model of mutually independent binding, perhaps with Foxa2 protecting PTF1 binding regions for future use. Given the disparate combined activities of Ptf1a and Foxa2 in our cell-transfection experiments, it is tempting to speculate that Foxa2 buffers PTF1 activity to achieve optimal levels of target transcription. For example, Foxa2 augments Ptf1a function on the *Clps* element, which itself possesses a low affinity PTF1 binding site, while it decreases Ptf1a mediated transcription on the high affinity PTF1 site in the *Rbpjl* element. Nonetheless, we cannot discount that additional co-factors or post-translational modifications may be needed to correctly assemble a higher order complex including Foxa2 and Ptf1a.

In summary, the tissue specific transcription factor Ptf1a utilizes multiple strategies to activate transcription of lineage-specific genes. In previous work, Ptf1a was shown to form a transcriptional activator complex with its E-protein partner and Rbpj or Rbpjl. These factors cooperatively bind a bipartite DNA motif resulting in increased sequence specificity and efficient transcriptional activation, a mechanism shared in pancreas and neuronal

development. From results in this study, we show that the distinct activity of Ptf1a in pancreas versus neural tube utilizes additional mechanisms that likely combine collaboration with other tissue-specific factors, either directly or indirectly, with lineage-specific chromatin accessibility to regulate target genes. Furthermore, we identify Foxa2 and GATA factors, or Sox and Hox factors, as the most likely candidates for the collaborating factors that influence Ptf1a specificity in the pancreas and neural tube, respectively.

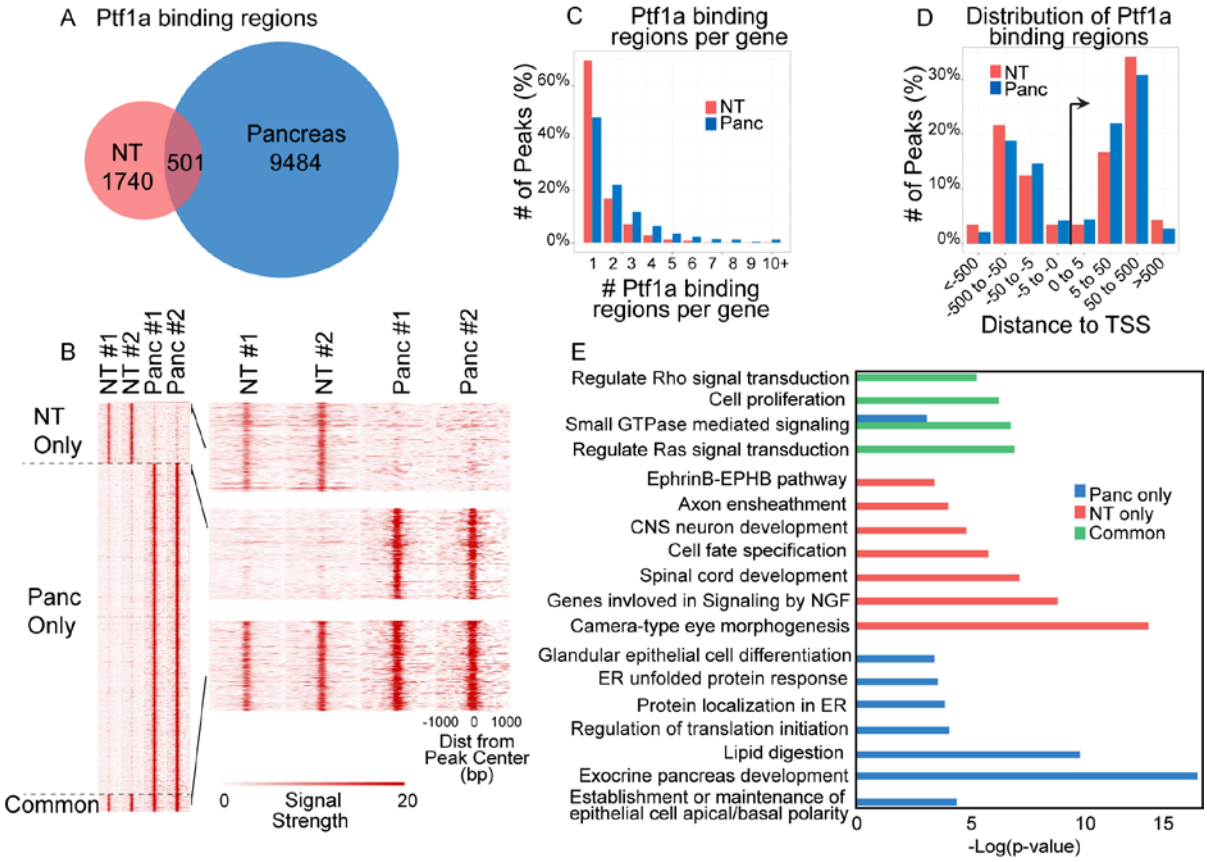


Figure 3.1. Ptf1a-bound sites in embryonic neural and pancreatic chromatins show distinct in vivo binding. (A) Comparison of genomic sites bound by Ptf1a in vivo for mouse E12.5 neural tube (NT) and E17.5 pancreas (Panc). Venn diagram shows Ptf1a binds largely dissimilar genomic regions. (B) *Left*: Heat map signatures of all called Ptf1a binding sites for two neural tube and two pancreas ChIP-Seq experiments. *Right*: Enlarged image of Ptf1a binding signatures at NT only, Panc only, and Common sites. Heat signatures show 1000-bp left and right of each peak center. (C) ~35% of NT and ~55% of pancreas genes associated with Ptf1a ChIP-Seq peaks identified by GREAT have more than one region with a Ptf1a-bound site. (D) Ptf1a-bound regions tend to fall distal (5-500 kb) to the transcription start site (TSS) of genes. (E) Select GO terms for the genes near NT-specific, panc-specific, or common Ptf1a-bound sites.

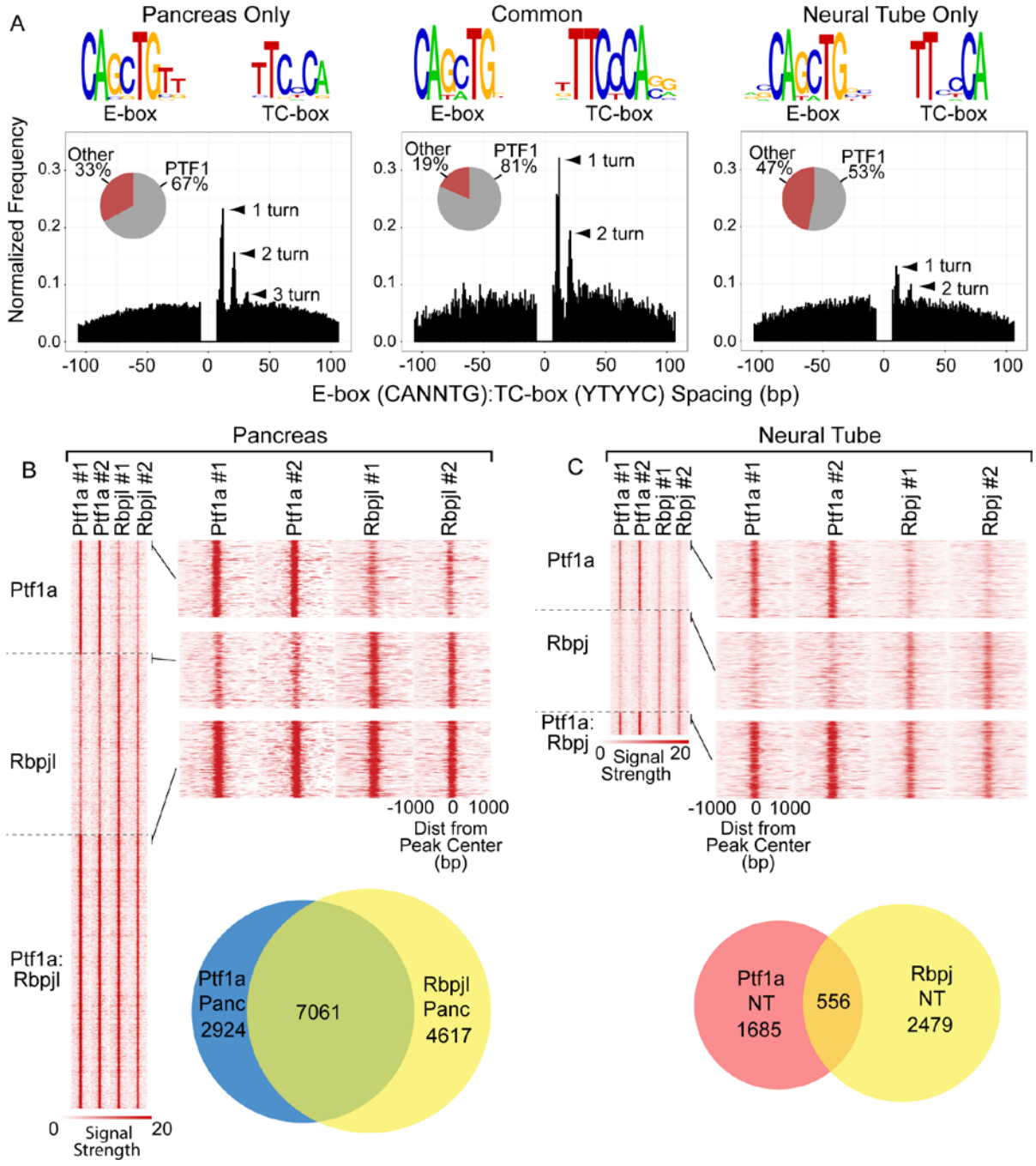


Figure 3.2. Ptf1a co-localizes with co-factor RbpJ or RbpJL and defines the E-box:TC-box compound Motif. (A) Primary and secondary motifs (E-box and TC-box, respectively) enriched in the Ptf1a-bound regions restricted to pancreas, neural tube, or those common to both tissues. Histogram of distances (in bp) between any E-box (CANNTG) and TC-box (YTYYCA), N=A/G/C/T, Y= C/T within each ChIP-Seq peak. Frequencies were normalized to total number of binding sites in each condition to facilitate comparison. Preferential

spacing of 1, 2, or 3 integral helical turns of DNA separating the E-box and TC-box is consistent with that proposed for the PTF1 binding consensus (Beres et al., 2006). The pie charts show the percentage of peaks that contain 1, 2, or 3 helical turn spacing defined as a PTF1 site. (B, C) Heat map signatures and Venn diagrams of Ptf1a and Rbpj binding sites from ChIP-Seq analyses of biological replicates for the pancreas (B) and the neural tube (C). *Right*: Zoomed images of regions indicated. Heat signatures were obtained for 1000 bp left and right of each peak center. The Venn diagrams show the extent of Rbpj/Rbpjl co-localization with Ptf1a. E17.5 pancreas has a 71% co-localization of Ptf1a with Rbpjl, and E12.5 neural tube a 25% co-localization with Rbpj.

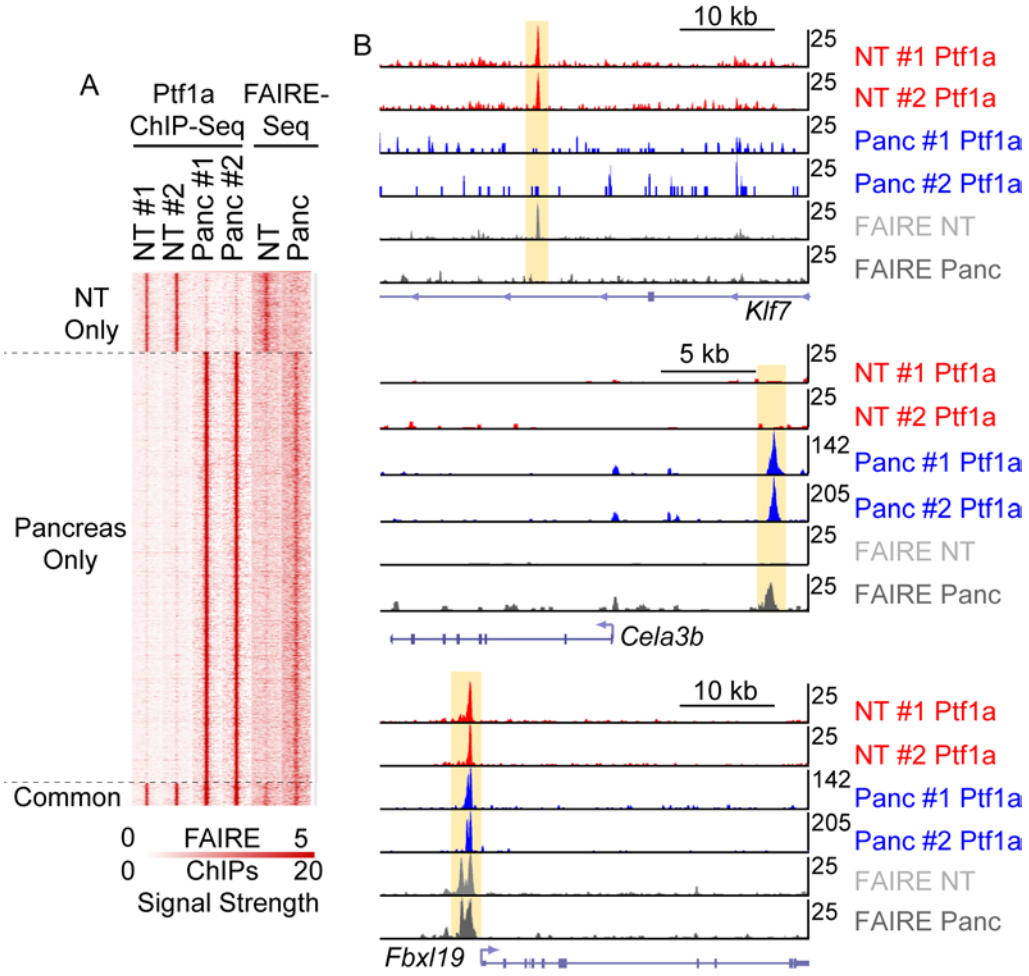


Figure 3.3. Chromatin accessibility is a major determinant of tissue-specific Ptf1a binding. (A) FAIRE-Seq (open chromatin) heat map signature at ChIP-Seq-determined Ptf1a-binding sites in the neural tube (NT), pancreas (Panc), and sites in common in both tissues (Common) (from Fig. 1) indicates tissue-specific chromatin accessibility at these sites. Heat signatures were obtained for 1000 bp left and right of each peak center. (B) Representative genomic regions showing ChIP-Seq data for Ptf1a binding at these sites in NT (red) and Pancreas (blue). This is compared to FAIRE-Seq data that indicates open-chromatin (NT, light grey; Panc, dark grey). Normalized read counts are shown on the right of each track. The *Klf7* locus represents a NT-specific Ptf1a binding site, the *Cela3b* locus represents a Panc-specific Ptf1a binding site, and the *Fbxl19* represents a locus with a Ptf1a-bound site in common in both tissues.

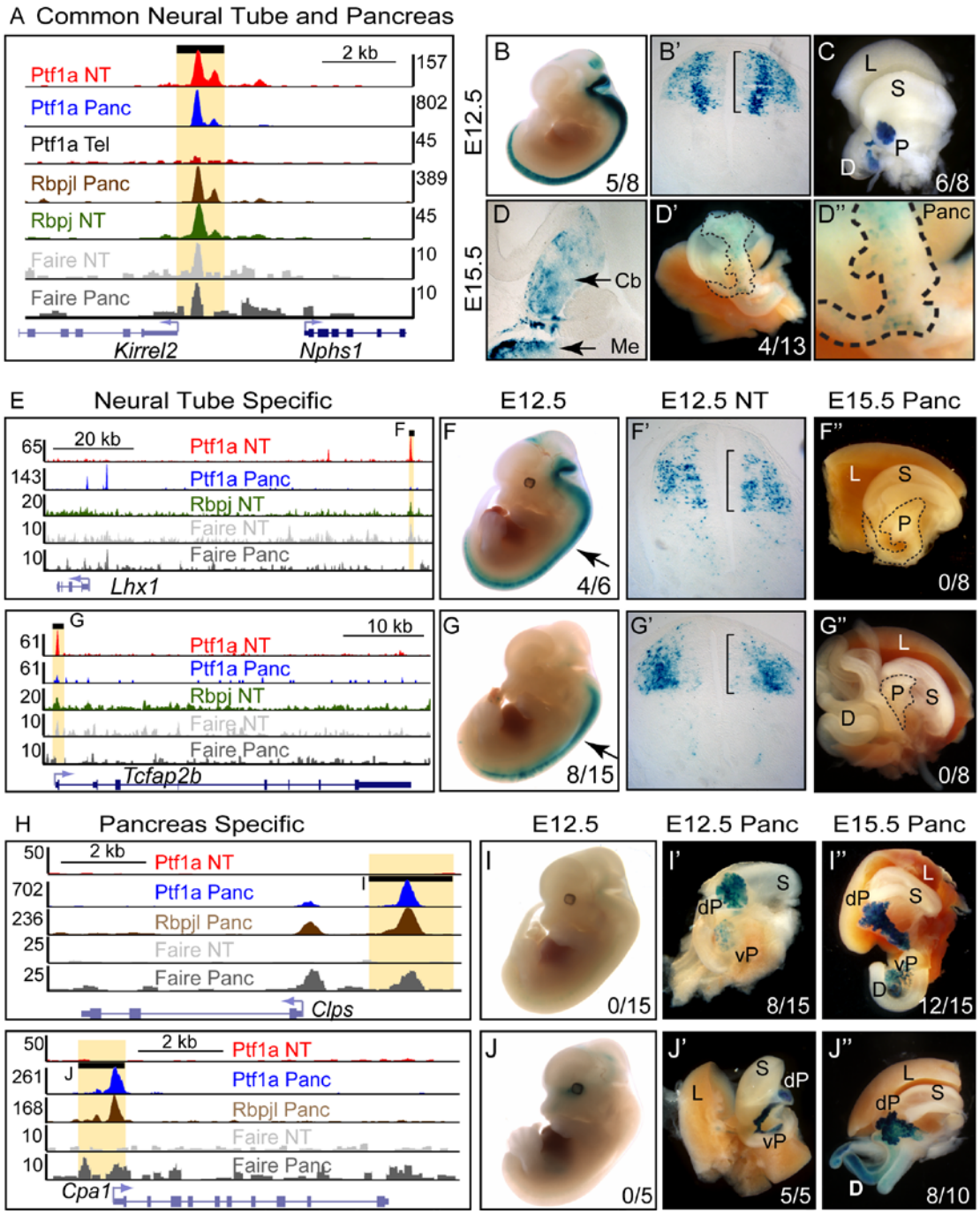


Figure 3.4. Tissue specificity of Ptf1a-bound regions is retained in transgenic embryos
(A) ChIP-Seq and FAIRE-Seq at the *Kirrel2/Nphs1* enhancer. Ptf1a/Rbpj both localize in vivo to this region at E12.5 in the neural tube (NT); similarly, Ptf1a/Rbpjl both bind at E17.5 in the pancreas (Panc). Ptf1a binding data for the telencephalon (Tel, Ptf1a-negative tissue) at E12.5 is displayed for comparison. FAIRE-Seq for each tissue is shown. Normalized read counts are at the right of each track. The black bar at the top designates the ~1 kb region used

for transgenic analysis (see Table S3 for coordinates for all transgenes). (B-D) The *Kirrel2/Nphs1* enhancer directs tissue-appropriate reporter expression in both the dorsal neural tube and pancreas. (B) Whole mount E12.5 embryo stained for β-gal showing prominent activity in the neural tube. (B') Cross-section of neural tube at forelimb level showing restricted expression in the dorsal region (bracket). (C) Isolated gut organs from the E12.5 embryo showing strong pancreatic expression. (D-D'') E15.5 transgenic embryos showing reporter expression in a sagittal section of cerebellum (Cb) and medulla (Me) (D), and in pancreas (dashed outline) (D'); expression mirrors native Ptf1a expression at this stage. (D'') Higher magnification of outlined region in (D'). (E) Ptf1a (red) and Rbpj (green) binding near the neural-specific genes *Lhx1* and *Tcfap2b* in the E12.5 neural tube. Black bars labeled 'F' or 'G' represent regions chosen for transgenic analysis (see Table S3). Lack of Ptf1a binding to these sites in the pancreas is shown for comparison (blue). FAIRE-seq for each tissue is also shown. (F-G) Representative E12.5 whole mount embryos stained for β-gal activity possessing enhancer-reporter transgenes derived from Ptf1a-bound regions in the E12.5 neural tube (Table S3). (F'-G') Cross-section of embryos shown in (F-G) at forelimb level. Brackets indicate the region of endogenous Ptf1a expression. (F''-G'') Gut organs from E15.5 transgenic embryos showing lack of pancreatic expression (dashed outlines). (H) Summary of ChIP-Seq and FAIRE-Seq near pancreas-specific genes *Clps* and *Cpa1* in E17.5 pancreas. Lack of Ptf1a binding in the neural tube (red) at E12.5 is shown for comparison. Black bars labeled 'I' or 'J' represent regions chosen for transgenic analysis (App D). (I-J) Representative E12.5 whole mount embryos stained for β-gal possessing enhancer-reporter transgenes derived from Ptf1a-bound regions in the E17.5 pancreas. Note the lack of reporter expression in the dorsal neural tube. (I'-J") Gut tissue from E12.5 and E15.5 transgenic embryos showing prominent pancreatic expression. The number of embryos exhibiting Ptf1a-like expression/number of embryos possessing transgene is shown in the bottom right of each image panel. D, duodenum; dP, dorsal pancreas; L, liver; S, stomach; vP, ventral pancreas. Transgenic animals were generated by Bradford Casey, Galvin Swift, and Yanjie Chang.

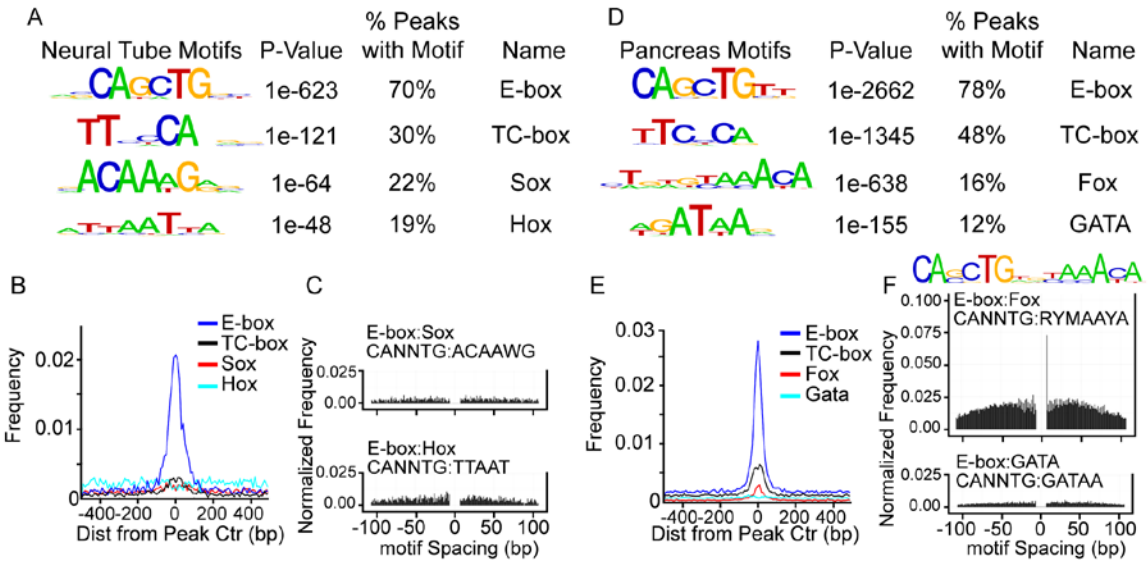


Figure 3.5. Tissue-specific motifs co-localize with Ptf1a binding regions. HOMER was used to search 150 bp regions bound by Ptf1a in neural tube (A) or pancreas (D) for enrichment of sequence motifs. In each case, the E-box and TC-box were the most highly enriched. However, secondary motifs were distinct in the two tissues. (A-C) In neural tube-specific Ptf1a peaks, consensus Sox and Hox family DNA binding motifs were identified as enriched. (B) A density plot of the motif frequencies relative to Ptf1a peak centers is shown. While E-boxes and TC-boxes cluster near the middle of peak regions, Sox and Hox motifs are more irregularly distributed. (C) Histogram of distances (in bp) between E-boxes and the Sox/Hox sites shows no preferred spacing or orientation of Sox motifs relative to E-boxes. (D-F) In contrast, consensus Fox and GATA family DNA binding motifs were identified in pancreas-specific Ptf1a peaks. (E) A density plot of motif frequencies relative to Ptf1a peak centers shows Fox motifs cluster near peak centers along with E-boxes and TC-boxes, while the GATA motif is more irregularly distributed. (F) Spacing analysis between E-box and Fox motifs reveals enrichment of a novel compound motif with a single bp separating the consensus motifs. The HOMER derived matrix of this E-box:Fox motif is shown.

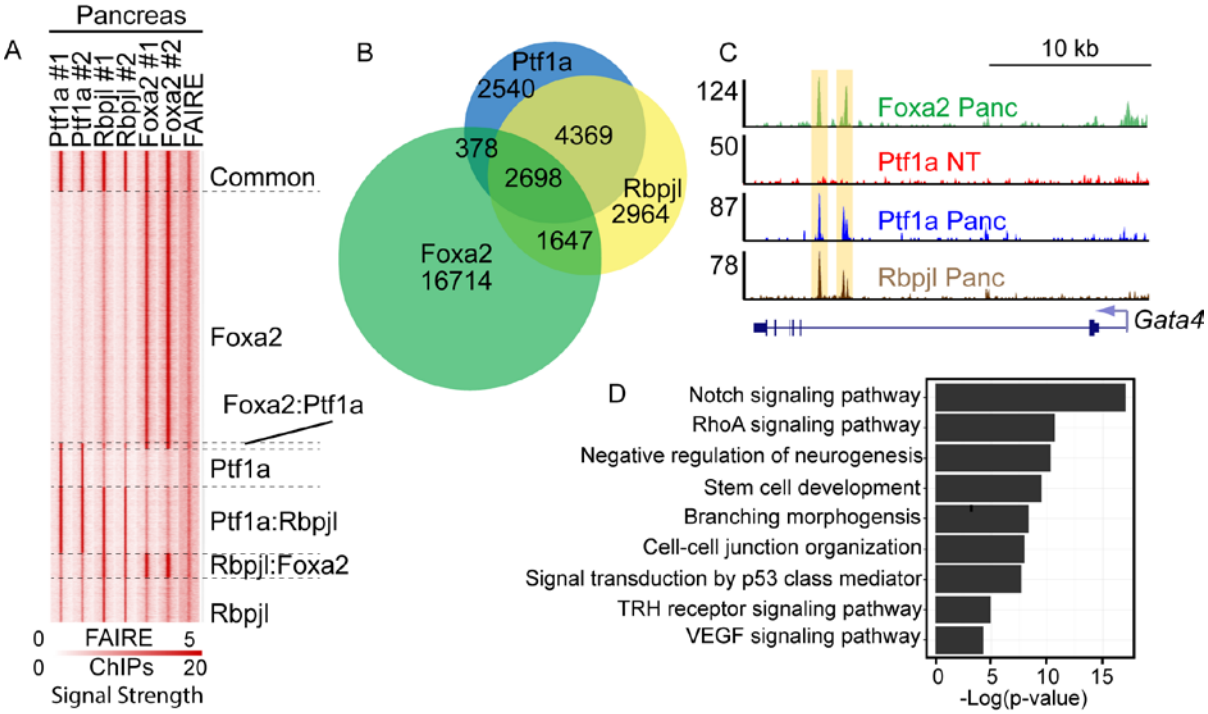


Figure 3.6. Foxa2 co-localizes with the PTF1-L complex in pancreatic chromatin. (A, B) Heat map signature and Venn diagram of ChIP-Seq data shows many regions bound by Ptf1a and Rbpjl in the pancreas at E17.5 are also bound by Foxa2. (C) Diagram of the *Gata4* locus illustrating the ChIP-Seq data in this locus and shows several intronic regions bound by Ptf1a (blue), Rbpjl (brown), and Foxa2 (green). Lack of Ptf1a binding in the neural tube at E12.5 (red) is displayed for comparison. Normalized read counts are shown on the left of each track. (D) GO classification of genes near regions occupied in common by Ptf1a, Rbpjl, and Foxa2.

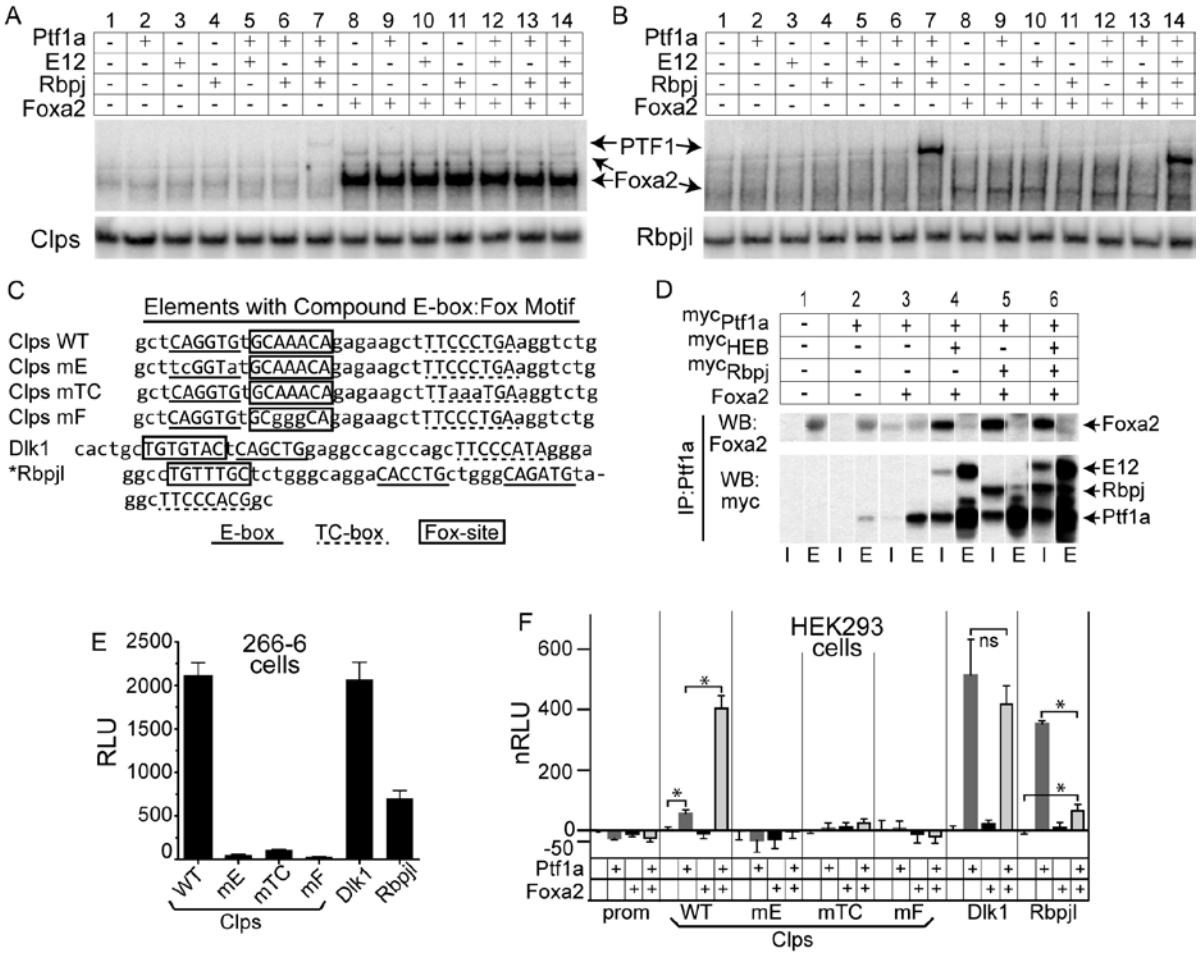


Figure 3.7. Foxa2 does not physically interact with Ptf1a but can act synergistically to activate transcription. EMSA with the *Clps* element (A) or *Rbpjl* element (B) shows that the PTF1 complex (lanes 7 and 14) and Foxa2 (lanes 8-14) bind DNA independent of one another. Free radiolabeled probe is shown as loading control (bottom panel). (C) DNA sequences of elements near *Clps*, *Dlk1*, and *Rbpjl* that contain PTF1 and Fox consensus motifs and used in the reporter constructs and EMSA. * The *Rbpjl* element does not contain the compound E-box:Fox motif as do the *Clps* and *Dlk1* elements. (D) Co-immunoprecipitations with Ptf1a antibodies using lysates from HEK293 cells transfected with combinations of expression vectors for ^{myc}Ptf1a, ^{myc}HEB, ^{myc}Rbpj, and Foxa2. Western blot with antibodies to Foxa2 or the myc epitope. No interaction between Foxa2 and Ptf1a is detected (compare elution (E) lanes to input (I)). (E) Luciferase reporter constructs containing elements from Ptf1a/Rbpjl/Foxa2 peaks with PTF1 and Fox motifs (shown in (C)) were transfected into pancreatic 266-6 cells that express these factors endogenously. Activity of the *Clps* element is abrogated when the binding site for Ptf1a (mE), Rbpjl (mTC), or Foxa2 (mF) is mutated. (F) The luciferase reporter constructs were transfected into HEK293 cells along with expression vectors for Ptf1a, Foxa2, or both. Error bars represent +/-SEM. *

p-value <0.01. RLU, relative luciferase units; nRLU, normalized RLU (see Methods). Data was generated by Paul Mayer and Galvin Swift.

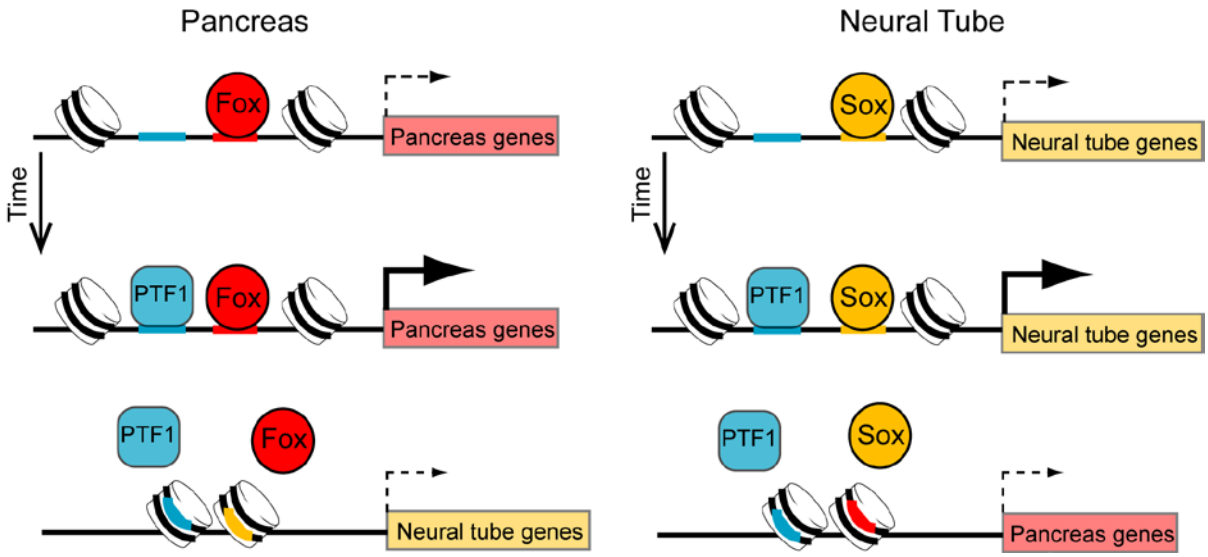


Figure 3.8. Model for generating lineage specific Ptf1a binding. Ptf1a binds to pancreas-specific regions and neural-specific regions within the genome. Lineage-specific transcription factors of the Fox and Sox families present prior to Ptf1a may function to maintain open chromatin at pancreatic and neural enhancers, respectively. When Ptf1a is expressed, it forms the PTF1 complex with its co-factors Rbpj/Rbpjl and E-protein and binds at PTF1 sequence motifs in these accessible sites. Since the relevant Fox factors such as Foxa2 are not co-expressed with Ptf1a in the neural tube, the pancreas-specific enhancers are not available for interaction with the PTF1 complex. And vice versa, Sox factors such as Sox1-3 are not expressed in the developing pancreas so the neural-specific enhancers are not available to be bound by PTF1 in this tissue even though Ptf1a is present. Thus, the lineage-specific cofactors may determine the accessibility of PTF1 sites.

CHAPTER FOUR

Genome-wide analysis of ASCL1 and NEUROD1 Targets in Small Cell Lung Carcinoma

INTRODUCTION

In the previous chapter Ptf1a binding activity was compared in two distinct tissue types during normal development. In this chapter, I examine how a bHLH factor, ASCL1, functions in a disease condition such as small cell lung carcinoma (SCLC). Ptf1a is not expressed in SCLC, but another proneural bHLH factor, NEUROD1, can be identified certain subtypes of SCLC tumors. Both Ascl1 and Neurod1 functions will be described below.

Small cell lung cancer (SCLC) accounts for 10%-15% of all lung cancers (Siegel et al., 2013). Lung cancer (SCLC and non-SCLC) in general is by far the leading cause of cancer death among men and women. In fact, more people die of lung cancer compared to colon, breast, and prostate cancers combined (Siegel et al., 2013). According to the National Institute of Cancer, SCLC patients without treatment have a median survival of 2-4 months after diagnosis. Patients that do receive treatment will typically see the return of disease within two years (90% of total SCLC population) and the overall survival at 5 years of any lung cancer patient is only 5-10%. The high reoccurrence of the tumor and the unacceptable mortality rate of the lung cancer are due in part to the lack of effective therapies. Thus,

understanding the molecular underpinnings of the SCLC and finding “druggable” targets are active areas of research.

Genetically, SCLC can harbor several different types of mutations, but the majority of SCLC tumors find a way to delete or inactivate the TP53 and Rb1 genes (D'Angelo and Pietanza, 2010). Moreover, it is also common to find transcription factors to be genetically amplified, including at least one member of the MYC Family (*MYC*, *MYCN*, or *MYCL1*), and at times Nuclear Factor I B-type (*NFIB*) (Dooley et al., 2011, Kim et al., 2006). Clinically, SCLC is presented as a highly aggressive malignant lung tumor that displays poorly differentiated neuroendocrine features. Gastrin-releasing protein, chromograinins, synaptophysin, and neural cell adhesion molecules are commonly expressed in neuroendocrine tumors and used as specific biomarkers for diagnosis (Stovold et al., 2012). However, tumors classified as SCLC histologically do not always express all of the neuroendocrine markers (Guinee et al., 1994), thus proper classification of SCLC remains a challenge, due in part to a lack of consensus in defining the subtypes neuroendocrine tumors (Moran et al., 2009).

Ascl1 is a biomarker readily used in the clinic for SCLC and other neuroendocrine tumors (Stovold et al., 2012), even though genetic alternations around the ASCL1 locus have not been reported. ASCL1 is found to be highly expressed in both human SCLC (Jiang et al., 2009, Borges et al., 1997, Ball et al., 1993) and mouse models of SCLC (Meuwissen et al., 2003, Schaffer et al., 2010). Molecular studies have shown that ASCL1 is necessary for tumor-initiating capacity in SCLC cell lines (Jiang et al., 2009, Osada et al., 2005). In addition, ASCL1 is sufficient to activate neuroendocrine markers not normally found in a

non-SCLC cell line (Osada et al., 2008). Consistent with these reports, the ectopic expression of ASCL1 with SV40 large T antigen in lung Clara cells generated aggressive adenocarcinoma with neuroendocrine phenotype (Linnoila et al., 2000).

SCLC is thought to originate from pulmonary neuroendocrine cells (PNECs) (Fig. 1.3) (Ball, 2004, Park et al., 2011). During normal development, ASCL1 is critical for the proper development of several neuronal populations throughout the nervous system, including PNECs. In mice, the ablation of ASCL1 expression results in the complete loss of PNECs (Borges et al., 1997, Ito et al., 2000). Thus, ASCL1 plays a critical role in both the normal development of the pulmonary neuroendocrine cells, and survival of SCLC cell lines.

Another proneural bHLH factor NEUROD1, is found to be expressed in a subset of SCLC cell lines (Osborne et al., 2013, Westerman et al., 2007, Hiroshima et al., 2006, Yazawa et al., 2009). Similar to ASCL1, NEUROD1 is important for the survival of SCLC cells in culture, and it also regulates the migratory capabilities of the cells (Osborne et al., 2013). NEUROD1 was also shown to be important in development of neuroendocrine cells in the lung. In contrast to the loss of ASCL1 that resulted in a complete ablation of PNECs, PNECS are still generated in the absence of NEUROD1, but the morphology and distribution of the cells are altered. The loss of NEUROD1 decreases the number of solitary PNECs and increases the number found in neuroendocrine bodies (NEBs) (Neptune et al., 2008). This suggests NEUROD1 may function genetically downstream of ASCL1. Similar to ASCL1, over-expression of NEUROD1 in non-endocrine non-small cell lung cancers cell lines is sufficient to increase cell proliferation and activate a neuroendocrine program (Neptune et al., 2008). Although ASCL1 and NEUROD1 are both capable of increasing cell proliferation

while activating neuroendocrine genes, it is unclear where these two bHLH factors bind in the genome, what genes they regulate, and if these related factors share a common function in SCLC.

Throughout the dissertation I have described that transcription factor binding and gene target regulation are influenced by many factors, including the DNA sequence the transcription factor can bind, the chromatin accessibility to the DNA binding sites and availability of co-regulatory factors. Given, that several human SCLC cells lines have been shown to harbor different genetic mutations, and display different expression profiles, it is unclear how ASCL1 and NEUROD1 function in these different cellular conditions.

In this chapter, I used two next-generation sequencing assays, ChIP-Seq and RNA-Seq, to gain insight into the possible mechanism by which ASCL1 and NEUROD1 select their gene targets, and to identify what those gene targets are in different pre-clinical models of SCLC. The vast majority of ASCL1 and NEUROD1 sites do not overlap, and their E-box binding preferences are quite distinct. A close examination of the DNA binding motifs surrounding ASCL1 and NEUROD1 binding sites revealed potential co-regulators. Among the potential ASCL1 co-factors was FOXA2, which bound thousands of the same DNA regions as ASCL1, which included key oncogenes important for the survival of SCLC. Further analysis of all the potential ASCL1 and NEUROD1 gene targets shows that these proneural transcription factors still regulate neuronal genes, but they also activate genes important for the proliferation and survival of the tumor cells. Taken together, results here support the model of ASCL1 and NEUROD1 having distinct functions in driving the neural

phenotype in SCLC, while also promoting their own individual gene programs important for the malignant behavior of tumor.

MATERIALS AND METHODS

Cell culture conditions

Cancer lines H889, H2107,H82, and H524 were maintained in RPMI-1640 (Life Technologies Inc.) supplemented with 5% or 10% fetal calf serum (FCS) without antibiotics at 37°C in a humidified atmosphere containing 5% CO2 and 95% air**.**

mRNA Isolation and Sequencing Library Preparation

10 million lung cancer cells were prepared for RNA extraction by washing twice with cold PBS followed by trypsinization. Total RNA was extracted and purified with Zymo's Mini RNA Isolation Kit. 1µg of total RNA was used for mRNA (polyA) isolation and sequencing library preparation using SOLiD Total RNA-Seq kit. Single-end sequencing of 75 bp was conducted for all samples on the ABI SOLid sequencer.

RNA-Seq Read Alignment, Expression Level Estimation and gene clustering

Sequence reads were aligned to the Hg19 build of the human genome using TopHat v2.0.9 (Trapnell et al., 2009). All Default settings were used with the following exceptions: -G option (instructs TopHat to initially map reads onto a supplied reference transcriptome) and --no-novel-juncs to ignore putative splice junctions occurring outside of known genes in the reference. The options used were multiple read correction (-u) to better distribute reads

mapping to multiple genomic locations and the bias correction (-b), to correct any sequence-specific bias introduced during the library preparation option. Gene expression levels were determined by the FPKM method of Cuffdiff v2.1.1 (Trapnell et al., 2010, Trapnell et al., 2013).

For K-means clustering, I filtered out any gene that was expressed at a low level (<10 FPKM) in all four cell lines, and any gene that did not show more than 2-fold difference in expression between the highest expressing and lowest expressing population. Expression levels for the remaining genes were then log2 transformed before clustering, then the log2 expression levels for each gene were centered along their respective mean across all four cell lines. K-means clustering (specifying 10 groups) was performed by Cluster 3.0 (de Hoon et al., 2004). Heatmap showing gene clusters were generated by Java TreeView (Saldanha, 2004).

Chromatin immuneprecipitation (ChIP) and Sequencing Library Prep

10 million lung cancer cells were prepared for chromatin immunoprecipitation (ChIP) by washing twice with cold PBS followed by trypsinization. Nuclei were liberated from cells by dounce homogenization and then fixed in 1% formaldehyde for 10 minutes at room temperature. The fixation process was terminated by adding glycine to a final concentration of 0.125M. Chromatin was sheared by using a Diagenode Bioruptor for 30 minutes on high power with 30s:30s on:off cycles. 250 μg chromatin was immunoprecipitated with 5 μg affinity-purified mouse anti-ASCL1 antibody (BD Biosciences) followed by anti-mouse Dyna beads (Invitrogen), with 5 μg goat anti-NEUROD1 (Santa Cruz, N-19) or goat anti-

FOXA2 (Santa Cruz, sc-6554) followed by Protein G Dynabeads (Invitrogen). The immunoprecipitated chromatin was then purified with the Qiagen PCR Clean-up kit and then prepped for sequencing. ChIP-Seq libraries were prepared using the NEBNext ChIP-Seq Library kit. Indexing primers and adapters were obtained from Illumina. Single-end sequencing of 50 bp was conducted for all samples on the Illumina High-Seq 2000 sequencer.

ChIP-Seq Samples, Peak Calling, Intersections, and Quantification

ASCL1 ChIP-Seq samples from cell lines H889 and H2107 were compared to an antibody control data set, an ASCL1 ChIP-Seq data in H524 a non-ASCL1 expressing cell line. NEUROD1 ChIP-Seq samples from H82 and H524 were compared to an antibody control data set, NEUROD1 ChIP-Seq data in H889, a non-NEUROD1 expressing cell line. FOXA2 in H889 was compared input control sample.

Sequence reads for each sample were mapped to the Hg19 assembly of the human genome with Bowtie (Trapnell et al., 2009). Duplicate reads were removed, and the remaining unique reads were normalized to 10 million reads. Peak calling was performed by HOMER (Heinz et al., 2010) using an FDR cutoff of 0.001, a cumulative Poisson p-value of <0.0001 and required a 4-fold enrichment of normalized sequenced reads in the treatment sample over the control/input sample. I defined a common binding site between two samples when the peak summits of each sample were found within 150bp of each other.

Quantification the ChIP-Seq samples were performed by HOMER (annotatePeaks.pl –size 5000 –hist 10 -ghist) (Heinz et al., 2010). Normalized sequence reads around each peak

were counted in 10bp bins, and the results were then loaded into Matlab® to generate the heatmap.

GO Classification and ChIP-Seq Peak Gene Annotation

Distance to gene and gene annotations for ChIP-Seq peaks were obtained using GREAT v1.82 (McLean et al., 2010). GREAT assigns a gene to a binding region if the region falls within 5 kb 5' or 1 kb 3' of the transcription start site (basal region) with a maximum extension of 1,000 kb in either direction. If the binding region falls within the basal region of multiple genes, then more than one assignment is made. All parameters were left at their default settings. The software Webgestalt (Wang et al., 2013) was used for Gene Ontology, all settings were left at their default settings.

Motif Discovery and Density Plots

For motif discovery, all tests were conducted with the HOMER package v4.2 using the following settings: -size 150 –S 10 –bits.(Heinz et al., 2010). I limited motif analysis by only using 150 bp DNA region around each peak summit. For statistical analysis, random background sequences with similar GC content to the test samples was generated for comparison (Homer default).

The E-box and secondary motif density plots were generated in HOMER (annotatePeaks.pl –size 1000 –hist 10). The program identified all sequences that matched to the HOMER

generated motif matrices within 1 kb of the surrounding peak regions. Density plots were then generated by Matlab®.

RESULTS

ASCL1 and NEUROD1 are expressed in SCLC cell lines with different genetic and molecular profiles

For decades, human derived SCLC cell lines have been used as the model system to uncover the genetic and molecular variations that can potentially drive the malignant behavior of the tumor cells. It is known that there is variation in the genetic alterations in human SCLC cell lines, and ASCL1 and NEUROD1 can be expressed at different levels (Fig. 4.1A-B). To compare the function of ASCL1 and NEUROD1 in SCLC, I identified two SCLC lines that express high levels of ASCL1 and low levels of NEUROD1 (H889 and H2107 ($ASCL1^{Hi}$)), and two SCLC that express high levels of NEUROD1 and low levels of ASCL1 (H82 and H524, $NEUROD1^{HI}$). mRNA expression levels of ASCL1 and NEUROD1 is demonstrated by RNA-Seq from these cell lines (Fig. 4.1B). The four human SCLC cell lines used in this study have known genetic alterations such as mutations in *TP53* and *RB1* genes, and genetic amplifications around some of the oncogenic Myc genes (*MYCL1, MYC, & MYCN*) and *NFIB* (Fig. 4.1A) (Johnson et al., 1996, Kim et al., 2006, Dooley et al., 2011). Both ASCL1 expressing SCLC cell lines are characterized as having classic SCLC characteristics, while the NEUROD1 cell lines has variant features, thus these two groups have morphological differences and express different neuroendocrine markers (Carney et al., 1985). The known genetic alterations in these lines did not always reflect high gene expression, such as seen

with *MYCL1*, which is genetically amplified in both in H889 and H2107, but high levels of expression were detected only in H889 (Fig. 1B). Instead, H2107 had high levels of *MYC*, but no known amplification around that gene locus has been reported. H82 and H524 have amplifications around the *MYC* locus and show high expression levels of *MYC*. Thus, all the cell lines used for this study have at least one member of the Myc family that is highly expressed. Another oncogenic transcription factor NFIB is genetically amplified in H889 but not in either of the NEUROD1^{Hi} SCLC cells (Dooley et al., 2011). Moderate levels of *NFIB* can be detected in both ASCL1^{Hi} SCLC cells and is present at higher levels compared to the NEUROD1^{Hi} cells. Taken together, ASCL1 and NEUROD1 are differentially expressed in SCLC cell lines characterized by different types of genetic and molecular alterations.

To begin uncovering differences in the SCLC lines that could be due to transcriptional regulation by ASCL1 and NEUROD1, I used K-means clustering to identify discrete expression profiles (Fig. 4.1C). In order to find substantial differences in expression, I filtered out any gene that was expressed at a low level (<10 FPKM) in all four cell lines, and any gene that did not show more than 2-fold difference in expression between the highest expressing and lowest expressing population. 5,706 genes (data not shown) met these criteria and results show that majority of the genes can clustered into cell-specific expression profiles, where one gene expressed significantly higher in one cell line compared to the other three. Despite the heterogeneity, 263 genes were expressed higher in both ASCL1^{Hi} cells compared to NEUROD1^{Hi} cells, and 271 genes were expressed higher in the NEUROD1^{Hi} cells compared to ASCL1^{Hi} cells (Fig. 4.1C). We used these two gene clusters as pools of candidate targets specific for ASCL1 or NEUROD1, respectively. Ontology analysis of these

two gene clusters (Wang et al., 2013) revealed both groups were enriched with genes with known neural function (Fig. 4.1D). Given that the genes from this analysis are non-overlapping, it suggests that NEUROD1 and ASCL1 contribute to different aspects of neural phenotype in SCLC.

ASCL1 and NEUROD1 bind distinct sites to regulate different genes in SCLC

In order to determine if the genes found in our clusters are direct targets of ASCL1 or NEUROD1 requires us to know where these two factors bind with in each cancer's genome. Given the various genetic and molecular differences found in the SCLC cell lines, it is unclear if these conditions influence ASCL1 and NEUROD1 binding. To investigate these questions, we performed Chromatin Immunoprecipitation coupled with high throughput sequencing (ChIP-Seq) for ASCL1 in H889 and H2107, and NEUROD1 in H82 and H524. Unexpectedly, our analysis found a large degree of discrete binding sites between ASCL1 and NEUROD1 for each cell-type (Fig. 4.1E-F). For ASCL1, we discovered 29,976 and 15,077 binding sites in H889 and H2107, respectively. The majority of these ASCL1 and NEUROD1 sites are located distal from transcription start sites (App. G). Among all the ASCL1 sites found in either of the $Ascl1^{Hi}$ cell lines, 9,124 were found to overlap. For NEUROD1, 11,250 and 4,011 were found in H82 and H524, and 1,405 were found to be in common. Among all the ASCL1 and NEUROD1 bound sites, only 289 sites were found to be in common. Taken together, ASCL1 and NEUROD1 bind distinct sites between each cell line, and a small subset of sites overlap among all the cell lines.

To gain insight into the genes directly regulated by ASCL1 and NEUROD1, I intersected the common ASCL1 and NEUROD1 ChIP-Seq sites with the specific ASCL1^{Hi} and NEUROD1^{Hi} expression enriched clusters found in the RNA-Seq data, respectively. Given the large amount heterogeneity found in each cell line, analysis was restricted to the overlapping sites of ASCL1 in the ASCL1^{Hi} and the overlapping NEUROD1 sites in the NEUROD1^{Hi} cells, to find common targets specific to ASCL1 and NEUROD1. This analysis highlights the different functions ASCL1 and NEUROD1 in SCLC.

In order to assign a binding site to a gene, the algorithm defined by the Genomic Regions Enrichment of Annotations Tool (GREAT) software (McLean et al., 2010) was used. Overwhelmingly, both ASCL1 and NEUROD1 bind to DNA sites far away from transcriptional start sites (TSS) (App G), regardless of the cell line; thus, assigning the nearest gene to the binding site is not always appropriate (McLean et al., 2010). GREAT defines a basal regulatory region for a gene within 5 kb upstream or 1 kb downstream of the transcriptional start site (TSS) and defines an extended regulatory region that exists within 1000 kb both up and downstream of the TSS or until the region reaches an adjacent gene's basal region. If the binding site falls within a gene's basal regulatory region, the peak is assigned to that gene alone, but if the binding site is outside the basal region, and falls into extended regions of genes, then the peak will be assigned to both genes. Given, that ASCL1 and NEUROD1 bind sites distal from the TSS, more than one gene is commonly assigned (data not shown).

Among the 271 genes expressed higher in the NEUROD1 cell lines, only 70 had a common NEUROD1 binding site (App H). Genes of interest include oncogenes such as *MYC*

and the cell-cycle regulator *CDK6* (App. H). Among the highest expressing genes found in this group, are genes typically associated with neuronal function such as transcription factor *OTX2*, the excitatory amino acid transporter, *SLC1A7*, the axon growth regulator, *OLM1*, and even *NEUROD1* itself (Fig. 4.2A, Table App. H).

NEUROD1 binding around these genes comes in different flavors, some genes such as *SLC38A5* have one binding site, just upstream of the TSS, but in most cases multiple binding sites around a gene was found, and these were located up- or down-stream of the TSS (Fig.. 4.2B, and data not shown).

In contrast, among the 263 genes expressed higher in the ASCL1 positive cells, 200 had an ASCL1 binding site assigned to it (App H). Included among the top ASCL1 potential targets were the neuroendocrine marker GRP and ASCL1 itself (Fig. 4.2C). Also, known oncogenic genes appear to be targets such as the anti-apoptotic gene BCL2, and Ret. There were 19 ASCL1 targets that have a known function in central nervous system development (data not shown), which do not overlap with NEUROD1 targets. Genes important for lung development were also targeted by ASCL1 such as FOXA2, NKX2-1, PROX1, and WNT11; as well as genes important for endocrine function (hormone secretion) such as SEC11C, ISL1, FAM3B, GCG, PRKARIB, GCK, INHBB, and CACNA1A. Like NEUROD1, multiple ASCL1 binding sites can be found surrounding the potential target genes distal from the TSS (Fig. 4.2D).

ASCL1 and NEUROD1 bind discrete consensus motifs in SCLC

Both ASCL1 and NEUROD1 are class II bHLH factors that bind DNA sequences containing the E-box motif (CANNTG) (for review see Bertrand et al., 2002). Given that ASCL1 and NeuroD1 are capable of binding so many distinct sites in each cell line, it is unclear if there are differences in E-box binding preferences. To gain more insight into which specific E-box motifs ASCL1 and NEUROD1 bind to in SCLC, we performed *de novo* motif analysis on all the ASCL1 and NEUROD1 sites for each cell line (Fig. 4.3A). Specifically, we analyzed the 150 bp regions surrounding the apex of each called binding site using the program Homer (Heinz et al., 2010). The primary motif returned from the analysis was the E-box motif and was found in almost all of the binding sites. With that said, ASCL1 and NEUROD1 did show differences in the type of E-box they bind. In particular, ASCL1 maintained its preference in both ASCL1^{Hi} for CASSTG E-boxes (S=G or C); specifically, CAGCTG or CAGGTG/CACCTG. In both NEUROD1^{Hi} cell lines, NEUROD1 also maintained its E-box preference for CAGMTG (M=C or A) motifs; specifically, the CAGCTG or CATCTG/CAGATG E-boxes. In order to quantitate the different E-box preferences, I determined the frequency and location for each specific E-box for all the peaks called in each cell line (Fig. 4.3B). Indeed, density plots show that the E-boxes are typically found near the center of the peak, but the occurrence of each E-box type is differentially distributed. In order to compare between cell lines, all results were normalized by the total number of peaks from each cell line. As expected, the GC core E-box is enriched in peaks found in all four cell lines, while the GG/CC-core is only enriched in the ASCL1^{Hi} cells. Conversely, the TC/GA

core E-box motif is not enriched in the ASCL1 expressing lines compared to the enrichment found in NEUROD1 cell lines NEUROD1^{Hi}.

In addition, motif analysis to binding sites unique to each cell line (App. G) and to sites found in common was performed (Fig. 4.3C). For all categories, ASCL1 and NEUROD1 maintained their E-box preferences, except for when ASCL1 and NEUROD1 where bound to the same location in all four cell lines. Moreover, the commonly shared E-box between all four cell lines appears more similar to NEUROD1's primary E-box (Fig. 4.3C, lower right). However, a closer examination of the 281 shared ASCL1 and NEUROD1 binding sites show the presence of both the GC-core and TC/GA within same binding site (data not shown), so it is unclear if ASCL1 is binding the TC/GA core E-box or just simply binding the GC E-box next to it. In summary, even though the majority of ASCL1 and NEUROD1 sites do not overlap, both factors prefer to bind to the GC-core E-box, but also have a preference for their own specific E-box, CACCTG/CAGGTG for ASCL1 or CATCTG/CAGATG for NEUROD1.

Specific transcription factor motifs co-localize with ASCL1 and NEUROD1 binding sites in SCLC

The differential binding between ASCL1 and NEUROD1 and the distinct molecular profiles found in each cell line maybe a result of cell line specific co-factors. To explore this possibility and to focus in on potential co-regulators important to both ASCL1^{HI} and NEUROD1^{Hi} tumor cells; thus, we searched for additional motifs enriched in the common

ASCL1^{HI} or the common NEUROD^{HI} sites. Indeed, we find ASCL1 and NEUROD1 sites are enriched with distinct transcription factor motifs.

Among the ASCL1^{Hi} common sites, two additional motifs matching the consensus binding sites of NF-I half site (also recognized as CTF) (Roulet et al., 2002) and forkhead factors (Badis et al., 2009) were enriched (Fig. 4.3C). Just over 40% of the ASCL1 binding sites contained the NF-I half-site motif and over 25% contained the forkhead motif. As for the NEUROD1 sites, the homeodomain motif matching to the CRX or Otx2 factors (Corbo et al., 2010, Bunt et al., 2011) was found in over 41% of the sites. In order to determine if these co-factor motifs are specific to either the ASCL1 or NEUROD1 sites, we determined the frequency of each motif with respect to the peak summits (Fig. 4.3D). For comparability, the number of occurrences was normalized to the total number of peaks in each group. Unexpectedly, the NF-I motif was enriched in both ASCL1 and NEUROD1 motifs and found near the peak centers, even though *de novo analysis* did not detect the motif in the NEUROD1 sites (Fig. 4.3D top panels). In fact, we found that 54% of the NEUROD1 sites contained the NF-I motif. The absence of the NF-I motif in the *de novo* motif analysis is likely due to the fact the NF-I site overlaps considerably with NEUROD1 primary E-box motif, thus NF-I motif is being masked by the frequent number of E-boxes found in the NEUROD1 peaks. As far as the additional motifs found in our analysis, the forkhead motif was specifically enriched with in the ASCL1 peaks, but randomly dispersed around NEUROD1 binding sites (Fig. 4.3D, middle panels). Conversely, the CRX/OTX2 motif is enriched near the NEUROD1 peak summits and is not found to occur often in the ASCL1 binding sites (Fig. 4.3D, lower panels).

The presence of the co-factor motifs suggests that one or more members of these transcription factor families are capable of binding the same chromatin regions as ASCL1 and NEUROD1. Interestingly, there are members of each transcription factor family that are specifically expressed in the ASCL1 or NEUROD1 cell lines, opening the possibility that they are cell specific factors that co-regulate genes important to the tumor with ASCL1 and NEUROD1. Factors such as FOXA2 (Fig. 4.3A) from the forkhead family, NFIB and NFIX from NF-I family are specifically expressed in the ASCL1 expressing cells (App. I). The NEUROD1 specific co-factor motif matches the known binding site of homeodomain factors CRX and OTX2 (Bunt et al., 2011). Among those two factors, only Otx2 is expressed in both NEUROD1 cell lines and neither is expressed in the ASCL1 expressing lines (App. I).

FOXA2 binds to the same DNA regions as ASCL1 in SCLC

In chapter three, it was shown that the bHLH factor Ptf1a binds to similar chromatin regions as FOXA2 in developing pancreas and that their binding sites, E-box and Fox motifs, are frequently separated by a single base pair (Meredith et al., 2013). Similarly, when we inspect the distant between the E-box and Fox motifs in the ASCL1 sites, we find a similar arrangement (data not shown). Moreover, as mentioned above, FOXA2 is specifically expressed in the ASCL1 expressing cell lines and not in the NEUROD1 lines (Fig. 4 4A, App. I). To determine if FOXA2 and ASCL1 bind the same sites in SCLC, ChIP-Seq was performed for FOXA2 in the H889 cell line. Indeed, thousands of sites overlap with ASCL1 in the SCLC.

Genome-wide analysis of FOXA2 found 14,308 binding sites and 5,190 of those sites overlapped with ASCL1 in H889 (Fig. 3.4B-C). *De novo* motif analysis of all FOXA2 sites (Fig. 4.4D) shows the primary enrichment was the Fox motif, followed by the E-box, which matches the ASCL1 primary E-box, and an NF-I motif. Additionally, a homeodomain motif was found enriched in the FOXA2 sites too. Thus, one would be able to predict that an E-box binding protein such as ASCL1 and a NF-I factor could potentially cooperate with FOXA2 in SCLC.

By re-applying the *de novo* motif algorithm to just the overlapping FOXA2 and ASCL1 binding sites in H889 shows that the enrichment of both E-box and Fox motif along with the NF-I motif once again (Fig. 4.4E). However, by altering the motif algorithm to search for longer motifs enriched in the shared sites (up to 15 bp long), it reveals that ~27% of the shared FOXA2 and ASCL1 sites have a fixed spacing of a single base pair (Fig. 4.4E). On the other hand, majority of the FOXA2 and ASCL1 shared sites did not have a DNA spacing restraint between their respective motifs. In order to determine if known oncogenes or genes essential for the survival of the lung cancer are potentially targeted by both ASCL1 and FOXA2, the ASCL1 and FOXA2 ChIP-Seq data was intersected again with the RNA-Seq. Indeed, key genes such as *GRP*, *BCL2*, and *RET* contain an overlapping binding sites of with ASCL1 (found in both $ASCL1^{HI}$ cells) and FOXA2 (Fig. 4.4 F-H). In the case of *GRP*, the fixed single base spacing between E-box and Fox motif was found (Fig. 4F), but the spacing between the two motifs go beyond the single base pair spacing for peaks assigned to *BCL2* and *RET* (Fig. 4G-H). The NF-I motif was found to be present in the peak assigned *BCL2* (Fig. 4.4G). Taken together, FOXA2 is restricted to ASCL1 expressing cells and binds

thousands of sites overlapping with ASCL1 in SCLC; and together they are found associated near key oncogenes.

DISCUSSION

Decades of research have established the developmental requirement for the bHLH factors ASCL1 and NEUROD1 throughout the nervous system. In chapters one and two, I describe and provide evidence that ASCL1 has the ability to induce a proliferating neural progenitor cell to exit the cell-cycle and differentiate; while, it well established that NEUROD1 is functionally essential for the terminal differentiation and maturation of post-mitotic neurons (Cho and Tsai, 2004). Knowing this, many researchers are left wondering why SCLC and other neuroendocrine tumors mis-regulate transcription factors such as ASCL1 and NEUROD1. Interestingly, it is actually common for other tumors, not just lung cancer, to "hijack" cell-lineage factors in order to promote its survival and malignant behavior (Garraway and Sellers, 2006). In order to explain why ASCL1 and NEUROD1 are important to the tumor, I set out to identify these factor's transcriptional targets and investigate how these factors select them. Using pre-clinical models of SCLC, I identify the genome-wide binding landscape of ASCL1 and NEUROD1 through the use of ChIP-Seq. In addition, the transcriptome profiles of SCLC models were identified through the use of RNA-Seq. Combing these two assays revealed that NEUROD1 and ASCL1 have distinct functions in SCLC, as they do in neural development. While ASCL1 and NEUROD1 appear to drive independent neural programs in SCLC, they also seem to contribute to the malignant behavior of the tumors independently. The different functions of these two bHLH factors are

reflected by their distinct binding sites throughout the genome. The difference in ASCL1 and NEUROD1 binding sites can partially be explained by each factor's specific preference for certain types of E-boxes, and their likely use of cell-specific co-factors.

ASCL1 and NEUROD1 maintains their neuronal function while regulating oncogenic genes

My investigation focused in on the functional differences between ASCL1 and NEUROD1 in SCLC. Analysis of their candidate gene targets and differential expression in each cell line highlighted the fact that both factors still maintain a neuronal function, but their functions are distinct (Fig. 4.1D, App. I). When either ASCL1 or NEUROD1 is overexpressed in non-neuroendocrine lung cells, they are capable of activating a neural program (Neptune et al., 2008, Osada et al., 2005); this phenotype is consistent with types of genes found in my study. Indeed, I find that ASCL1 targets neuronal genes such as the neural cadherin gene, *CDH2*, synaptic scaffolding protein, *SHANK3,* and components of the Notch pathway, *Dll1*, *Dll4*, and *DNER.* In addition, ASCL1 appears to directly contribute to the endocrine function of SCLC by regulating genes such as Gastrin-releasing peptide (*GRP*), Dopa decarboxylase (*DDC),* and Dopamine beta-hydroxylase (*DBH*). In contrast, NEUROD1 activates neural transcription factors such as *Neurod4 and OTX2*, and several genes found on neuronal axons and synaptic terminals such as *OLM1, SDC2, KIRREL2*, SCN3B, and *THY1*. As far as the endocrine function specific to NEUROD1, we find that NEUROD1 targets the somatostatin receptor, *SSTR2*.

The major goal in this study was to uncover why ASCL1 and NEUROD1 are important for the survival of SCLC. By identifying their target genes, it became clear why ASCL1 and NEUROD1 are essential to the tumor. For example, as mentioned above, ASCL1 regulates the neuroendocrine genes *GRP*, which has been shown to be critical for the ongoing proliferation of SCLC (Patel et al., 2006). In addition, ASCL1 and NEUROD1 specifically target well-known oncogenes. We identified that *BCL2* and *RET* are specific targets of ASCL1 (App. I). In support of this ASCL1 specificity, the application of the BCL2 inhibitor (ABT263) in mice with ASCL1 expressing H889 xenografts, showed an effective inhibition the tumor growth (Tse et al., 2008), but application of a BCL2 inhibitor (ABT-737) in mice with the NEUROD1-expressing H82 xenografts did not show changes in tumor growth (Hann et al., 2008). *RET* has not been directly associated with ASCL1 in SCLC, and the relevance of *RET* in SCLC patients remains unclear; however, *RET* expression was found to be highly correlated with ASCL1 expression in smokers with neuroendocrine lung adenocarcinoma, a non-small cell lung carcinoma (Kosari et al., 2013). Kosari *et al.* demonstrated that the reduction of ASCL1 expression also reduced RET levels, which resulted in the inhibition of growth for the tumor cells. Thus, in another neuroendocrine lung cancer, RET appears to require ASCL1. In contrast, the cell-cycle regulator *CDK6* and the transcriptional amplifier *MYC* appear to be directly regulated by NEUROD1. *CDK6* like *RET* has not been thoroughly investigated in SCLC, but in other lung cancers, the reduction of *CDK6* sends the tumor cell into cell-cycle arrest (Zhu et al., 2013). *MYC* can be found be up-regulated in subset of SCLC cell-lines (Johnson et al., 1996, Kim et al., 2006), but a correlation with NEUROD1 expression has not been reported. In summary, ASCL1 and

NEUROD specifically regulate several well-known oncogenes in pre-clinical models of SCLC.

This study revealed functional differences between ASCL1 and NEUROD1 as reflected in the minimal number of overlapping binding sites found in the tumor cell lines (Fig. 4.1F). Nonetheless, there are 289 sites that overlapped in all four cell lines. Comparing the genes associated with these sites (using GREAT) with genes expressed in all four cell lines studied here, 150 genes were identified (App H). Indeed, we found the known ASCL1 target's *Dll3* and *RGS16* (Jiang et al., 2009, Henke et al., 2009a, Castro et al., 2006) are also potential targets of NEUROD1. Interestingly, NEUROD1 has been shown to regulate lung neuroendocrine tumor migrating capabilities through its control of NCAM1 (Osborne et al., 2013). We find that NCAM1 is a target of both ASCL1 and NEUROD1. Thus, although only a small fraction of ASCL1 and NEUROD1 binding sites are found in common, some of these common gene targets may be important to the tumor.

Distinct binding profiles of ASCL1 and NEUROD1 in SCLC

This study addresses fundamental questions in the transcription factor field of how two structurally similar factors such as the bHLH proteins ASCL1 and NEUROD1 select their DNA binding targets and how their target selection is influenced by different cellular environments. The tumor models, which were clinically classified as the same tumor type of SCLC, harbored several different genetic alternations (Fig. 4.1A-B)). Moreover, the gene expression profiles of each of the four cancer cell lines show contrasting molecular differences between each other (Fig. 4.1C), an environment that could alter transcription

factor function. Coinciding with these cell-specific observations, is the finding that the majority of the ASCL1 and NEUROD1 binding sites were unique to one cell-line, even among the cells that showed similar expression levels of ASCL1 and NEUROD1. It remains unclear if these cell-line specific differences contribute to the tumor malignancy. The type of genes associated with these cell-line specific binding sites, determined by the software GREAT, shows an overwhelming an enrichment of genes with known neural function (data not shown). In attempts to explain why the cell-line specific binding differences occur, I further examined the cell-specific binding sites at the DNA level by *de novo* motif analysis; unexpectedly, there were no dramatic changes in the primary E-box binding site and no new co-factor motif appeared in any of the ASCL1 or NEUROD1 cell specific sites (App. G). Given that small differences can be detected at the DNA level for each cell, it suggests that the chromatin landscape may be the critical determent for ASCL1 and NEUROD1 binding in each cell.

ASCL1 and NEUROD1 bind distinct E-boxes

Directly comparing the ASCL1 and NEUROD1 ChIP-Seq data sets provide mechanistic insight into how these two factors select their DNA binding site, and provide additional supporting evidence that these two factors are functionally different in SCLC. Regardless of the cell line, motif analysis shows these two factors prefer distinct E-boxes, where ASCL1 specifically prefers the CAGGTG/CACCTG E-box compared to the NEUROD1 preference for the CATCTG/CAGATG E-box (Fig. 4.3A-B). Interestingly, both factors commonly bind the CAGCTG E-box in their respective cell lines; however, an

intersection of all the binding sites show very little overlap between any ASCL1 and NEUROD1 sites. The lack of binding site overlap at binding regions with a CAGCTG E-box brings into question, whether Ascl1 or Neurod1 need additional co-factors to bind these regions or if the DNA region in the other cell lines is not accessible because of chromatin structure

Identification of the ASCL1 and NEUROD1 transcription factor network in SCLC

There is a large overlap of ASCL1 sites between H889 and H2107 (9,124 sites) and NEUROD1 sites between H82 and H524 (1,405 sites). Examination of these common sites led to identifying other candidate transcription factors that may work in concert with ASCL1 and NEUROD1.

Forkhead and the NF-I motifs were found enriched within the ASCL1 bound regions. There are dozens of forkhead family members that can bind this motif and several of these factors have been shown to be mis-regulated and important in several types of cancers (Myatt and Lam, 2007). Within the ASCL1 expressing cell lines, several Fox factors are expressed (Fig.4.4A, App. I), but only FOXA2 expression was restricted to the $ASCL1^{Hi}$ cell lines and not expressed in in the $Neurod1^{Hi}$ cells. This may be due in part to FOXA2 being directly regulated by ASCL1 in these cells (Fig. 4.2C); suggesting a feedforward mechanism for maintaining gene expression. In support of this mechanism, we found thousands of FOXA2 binding sites overlapping with ASCL1 bound sites. Among the sites co-bound by FOXA2 and ASCL1 was ASCL1 itself, suggesting ASCL1 may auto-regulate its own expression (Fig. 4.2D) with the help of FOXA2. Other potential co-regulated targets known to be

important to the tumor are *GRP*, *BCL2* and *RET* (Mulligan et al., 1998, Tse et al., 2008, Patel et al., 2006); their relationship with ASCL1 will be discussed below. If and how ASCL1 and FOXA2 interact at the transcription level is unclear, the fixed spacing between the E-box and Fox motif suggest a physical interaction between the two proteins may exist. However, a similar scenario was found, with the same DNA spacing between the binding motifs, for the bHLH factor Ptf1a and FOXA2 in the developing pancreas (Meredith et al., 2013), but protein-protein interaction was never found. Nonetheless, our data suggest not only ASCL1 is directly targeting key oncogenes, but it regulates these key genes by cooperating with other factors such as FOXA2.

The NF-1 binding motif was also found to be enriched in the ASCL1 binding regions. NF-1 factors consist of 4 proteins: NFIA, NFIB, NFIC, and NFIX. Both NFIB and NFIX are expressed at higher levels in the ASCL1 expressing SCLC lines (App. I). NFIB, which is commonly over-expressed and genetically amplified in SCLC, has been shown to be an important oncogenic gene that regulates the cell viability and proliferation during the tumor formation (Dooley et al., 2011). NFIX on the other hand, has not been reported to be important for any type of tumor, but it has been studied in the nervous system. As neural stem cells enter a quiescence phase, NFIX is found to be up-regulated and required for proper induction of the quiescent state (Martynoga et al., 2013). It would be interesting to know if ASCL1 and NFIX co-regulate genes for a quiescent niche within SCLC tumors; potentially explaining why SCLC tumors are resistance to chemotherapy. Martynoga and colleagues highlight that helix-loop-helix protein ID4 is highly up-regulated and a potential direct target of NFIX in neural progenitor cells entering a quiescent phase. In the $ASCL1^{Hi}$ SCLC cell

lines, ID4 is an ASCL1 specific target (App. H) and the NFI motif is found within the ASCL1 binding sites. Thus, further investigation into the functional relationship between ASCL1 and NF-I factors is warranted.

Within the NEUROD1 bound regions there is enrichment of the binding motif for the homeodomain factor OTX2 (Bunt et al., 2011), which is expressed in the NEUROD1^{Hi} cells but not the ASCL1^{Hi}. Similar to the relationship between ASCL1 and FOXA2, NEUROD1 appears to directly regulate *OTX2* (Fig. 4.2C), suggesting another feed-forward transcriptional mechanism. During development, both NEUROD1 and OTX2 have been shown to be required for proper neuronal specification in the retina (Swaroop et al., 2010). As for a role in cancer, OTX2 is required for cell proliferation in pre-clinical models of medulloblastoma (Bunt et al., 2012). Consistent with this function in SCLC, NEUROD1 binding sites containing the OTX2 motif are found around the oncogene MYC. Altogether, the ChIP-Seq data have elucidated the binding preferences for NEUROD1 and ASCL1, but also implicated additional transcription factors that may be important collaborating factors that may function in the survival of the tumor.

ASCL1 has distinct neural functions in the developing neural tube *versus* small cell lung carcinoma

In the second chapter the genome-wide binding of Ascl1 in the mouse embryonic neural tube was reported. In both SCLC and the developing neural tube, ASCL1 occupies DNA regions near genes with known neuronal function. However, a comparison of the DNA binding sites in each tissue shows that Ascl1 has a slightly different E-box preference. In the neural tube,

the GC-box core E-box is highly preferred, but in SCLC, ASCL1 prefers the GC-core along with the GG/CC-core E-box. In addition, in each tissue type, several non-E-box motifs were enriched in Ascl1 binding regions, but those motifs were distinct between the two tissues. This suggests that tissue specific co-factors are co-regulating with Ascl1 to activate disparate neural programs. The ASCL1 binding sites in neural tube were directly compared with ASCL1 binding sites in the SCLC cancer cells by lifting over the Ascl1 binding coordinates found the mm9 mouse genome to the hg19 human genome. Indeed, only 302 of the ASCL1 neural tube binding sites were found to be homologous with ASCL1 sites in both Ascl1 expressing SCLC cell lines (appendix J). However, among these shared sites was the binding region near the notch ligand Dll3. Dll3 has been shown to be ASCL1 target in multiple tissue types (Ball et al., 1993, Castro et al., 2006, Castro et al., 2011, Henke et al., 2009a). In addition to Dll3, several other members of the notch pathway are targets of Ascl1 in the nervous system and cancer cells such as Dll1, Rbpj, Maml3, and Lfng. Altogether, analysis of the Ascl1 binding sites and potential gene targets show that Ascl1 regulates a divergent, cell-specific gene program in the neural tube and SCLC, but also a limited set of genes particularly in the notch pathway.

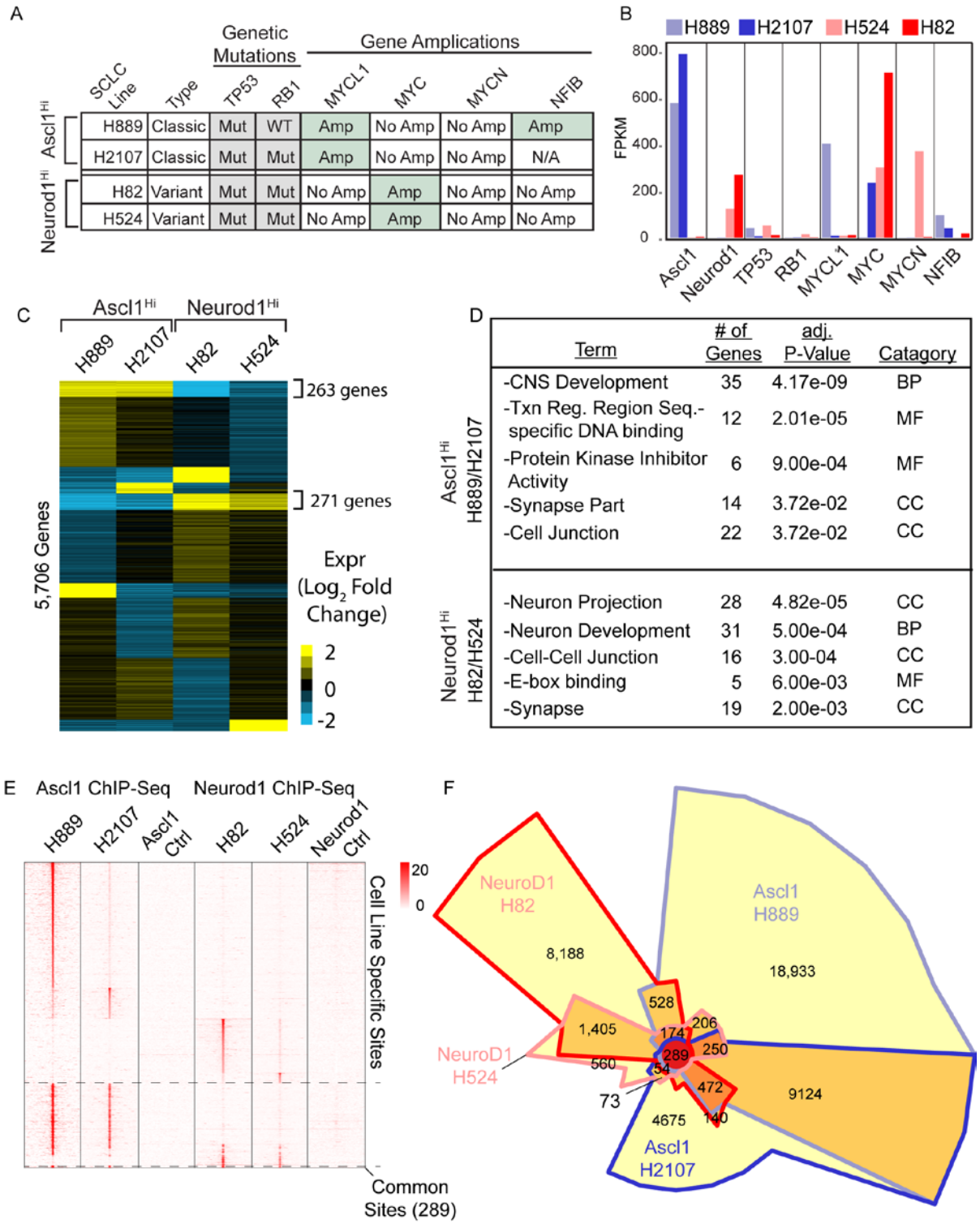


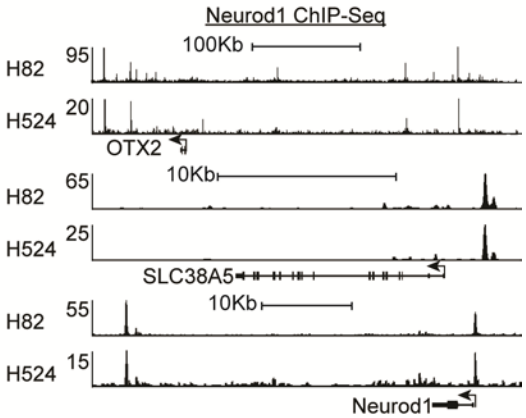
Figure 4.1. Genetic and molecular heterogeneity in ASCL1 and NEUROD1 expressing SCLC. (A-B) Summary of the known genetic alterations (Johnson et al., 1996, Kim et al., 2006, Dooley et al., 2011, Carney et al., 1985) and their expression levels determined by

RNA-Seq in ASCL1 and NEUROD1 expressing cell lines. (C) K-means clustering of differentially expressed genes between the SCLC cell lines. Note, gene expression levels were transformed to a log2 scale prior to clustering and fold changes differences are also measured in a log2 scale. (D) Gene ontology analysis (ref) of genes enriched specifically in the ASCL1 expressing cell lines (top) or NEUROD1 expressing lines (bottom). (E) Heat map summarizes the ChIP-Seq signal intensity +/- 2.5 kb around the identified peak centers bound by ASCL1 or NEUROD1 in H889, H2107, H82, and H524. Note, ASCL1 control is from ASCL1 ChIP-Seq in H82 (ASCL1 negative) and NEUROD1 control is NEUROD1 ChIP-Seq in H889 (NEUROD1 negative). (F) Chow-Ruskey diagram showing a 4-way intersection between ASCL1 and NEUROD1 ChIP-Seq binding sites found cell line. Note, overlapping sites were required to be within 150bp of each peak summit.

A

Top 10 Neurod1 Target's Expression (FPKM)				
Gene	H889	H2107	H82	H524
OTX2	0.30	0.85	670	110
SLC38A5	0.42	1.5	450	110
NEUROD1	0.35	0.79	270	120
OLFM1	0.41	0.18	100	130
RGS10	0.26	0.00	96	39
ARL4D	8.8	1.0	84	25
HPCA	0.98	0.31	72	27
TMEM54	0.18	0.00	66	26
SDC2	0.62	2.3	61	12
SLC1A7	0.00	0.05	58	16

B



C

Top 10 Ascl1 Target's Expression (FPKM)				
Gene	H889	H2107	H82	H524
GRP	1800	3200	3.7	1.4
ASCL1	580	790	5.0	2.7
TESC	350	210	3.5	0.74
IGFBP5	320	180	5.5	2.6
GPX2	180	170	0.34	0.35
HOXB2	170	180	2.2	3.5
CLDN5	160	120	0.43	0.15
HOXB5	120	110	0.08	0.50
FOXA2	110	67	0.89	0.14

D

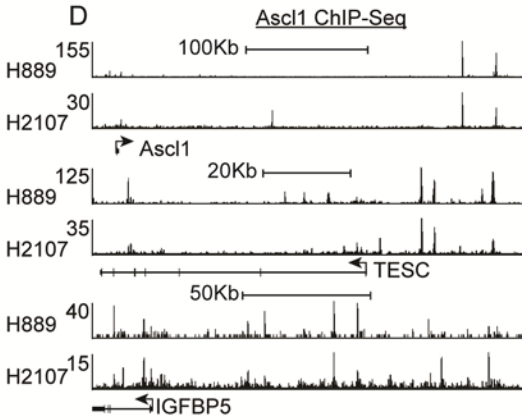


Figure 4.2. Top potential direct downstream targets of ASCL1 and NEUROD1 targets in SCLC. (A-B) The highest expressing genes determined by RNA-Seq found in both NEUROD1 positive cell lines, that show a low to no expression in both ASCL1 positive cells, (A) and vice versa (B). Note, all potential gene targets had an associated NEUROD1 or ASCL1 binding site, determined by the software GREAT (McLean et al., 2010). (C-D) Examples of the ChIP-Seq NEUROD1 and ASCL1 bound sites around their potential gene targets. The genomic scale, measured in Kilobases (Kb), is listed above each track. The highest intensity of the ChIP-Seq signal for the genomic window being shown for each sample is listed on the left of the track.

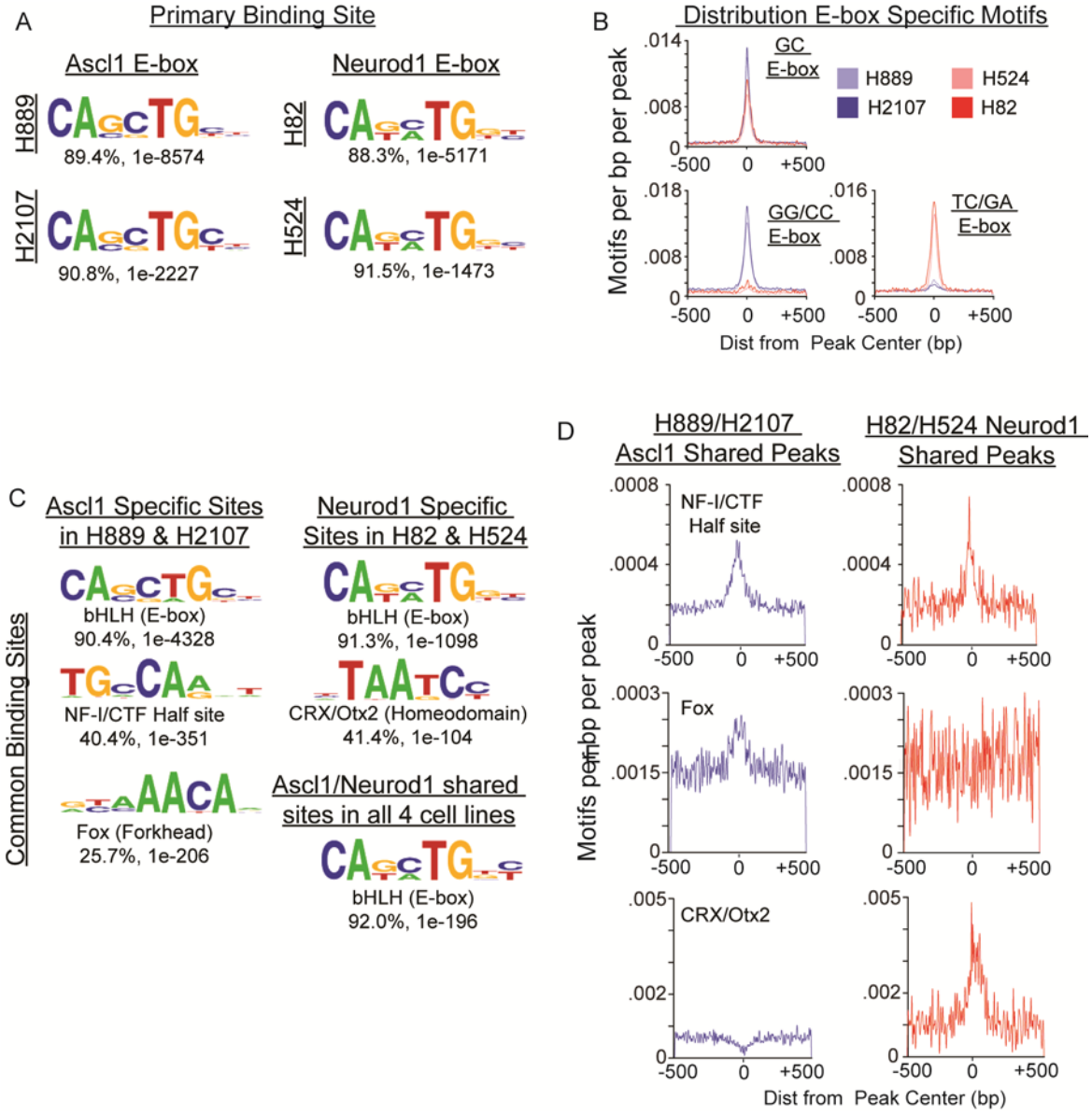


Figure 4.3. Genome-wide characterization of ASCL1 and NEUROD1 binding sites is SCLC reveals both unique and common E-box binding preferences, along with specific co-factor motifs. (A) The primary E-box motif found by *de novo* motif analysis (ref) for ASCL1 and NEUROD1 binding sites in each of their respective cell lines. (B) Density plots of specific E-box motif frequencies relative to the peak centers found in all four SCLC cell lines. The GC-core E-box is enriched in all four cell lines (top); while the GG/CC-core E-box is enriched only in the ASCL1 expressing cell lines (H889/H2107) and TC/GA-core is enriched only in the NEUROD1 expressing lines (H82/H524) (bottom). (C) The significantly enriched motifs found by *de novo* motif analysis in common ASCL1 sites (Left), NEUROD1 common sites (top right), or common NEUROD1 and ASCL1 sites in all four cell lines (bottom right). (D) Density plots of the potential co-factor motifs found in either ASCL1 or

NEUROD1 common sites. Note the percent of peaks found in each condition that contain the motif is listed below motif matrixes, along with its statistical significance. For comparison, each density plot was normalized by the total number of occurrences found in each cell line or common peak intersection.

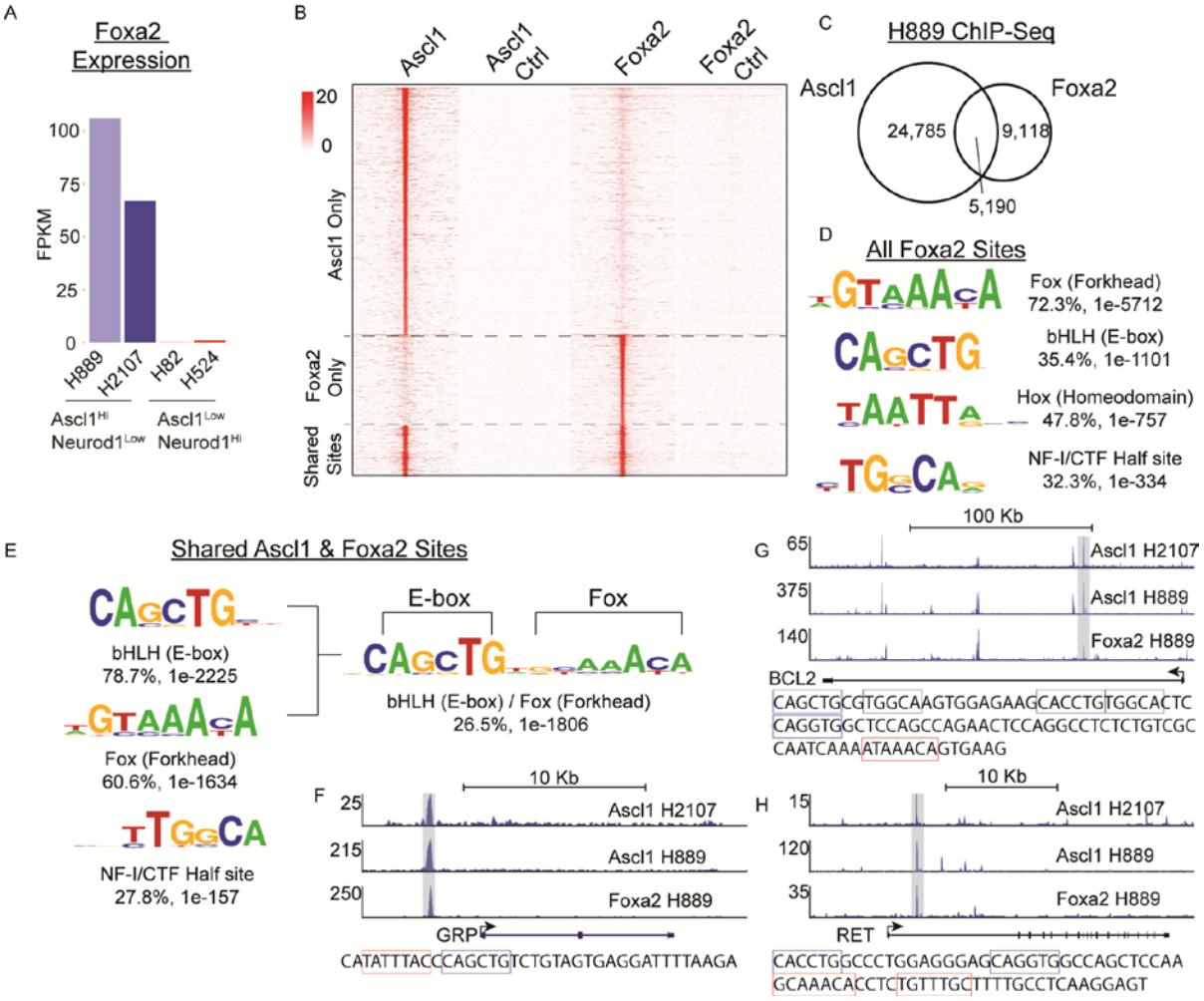


Figure 4.4. FOXA2 and ASCL1 bind to the same genomic regions in the SCLC genome and potentially co-regulate oncogenes RET and BCL2. (A) Expression levels (FPKM) of FOXA2 determined by RNA-Seq in SCLC. (B) Heat map summarizes the ChIP-Seq signal intensity +/- 2.5 kb around the identified peak centers bound by ASCL1 or FOXA2 in H889. Note, ASCL1 control is from ASCL1 ChIP-Seq in H82 (ASCL1 negative) and FOXA2 control is Input sample from H889. (C) Venn diagram showing the number overlapping ASCL1 and FOXA2 binding sites in H889. Overlapping sites were required to be within 150bp of each peak summit. (D) *De novo* motif analysis of all FOXA2 ChIP-Seq sites found in H889. (E) *De novo* motif analysis of the shared ASCL1 and FOXA2 ChIP-Seq sites found in H889 (left). Repeated *de novo* motif analysis of same common sites, but using extended motif search parameters (Motifs up to 15bp long) (right). (F-G) Examples of the ChIP-Seq FOXA2 and ASCL1 bound sites around their potential gene targets. DNA sequences around the peak summits (marked by light grey box) are shown below each track. Blue boxes marks the E-boxes, red boxes surround the Fox motifs, and grey box mark the NFI motif.

CHAPTER FIVE

Conclusions and future directions

The significance of this body of work is that for the first time the binding sites for multiple members of the bHLH family were identified *in vivo* and compared across different tissues. Using this information I was able to identify the direct target genes of these factors and gain insight into how a single factor can function in or give rise to disparate tissue types. In doing so, I expanded the understanding of how Ascl1 and Ptf1a function in neural development. By uncovering their direct target genes in the neural tube, we now know that Ascl1 and Ptf1a directly oppose each other's function through the direct activation of key transcription factors in the homeodomain family, and that Ptf1a itself may directly block Ascl1's transcriptional activity. In addition, multiple general principles in how sequence specific transcription factors function during development have been uncovered. In particular, my analysis provides three keys to target specificity for the bHLH factors: 1) selection of specific E-boxes, 2) the use of distinct co-factors in each tissue, and 3) the influence of chromatin accessibility. With this new knowledge about Ptf1a and Ascl1 function across multiple tissues, future studies should investigate which tissue-specific factors co-regulate and interact transcriptionally with Ascl1 and Ptf1a to activate the proper gene programs.

Direct Ptf1a inhibition of Asc11

In chapter two I demonstrated that Ptf1a opposes Ascl1 function by activating a downstream target such as Prdm13, or by Ptf1a itself directly inhibiting the transcription of Ascl1 target genes. How Ptf1a performs the latter is unknown. Given that a subset of Ascl1 and Ptf1a sites overlap (Fig. 2.1A), and that both factors are capable of binding a GC-core box (Fig 2.1G), one possibility is that these two factors compete for same E-box binding sites. For Ptf1a to repress Ascl1 in this scenario, the transcriptional activity of the Ptf1a heterodimer would need to be in active, or at least less active than the Ascl1 heterodimer (the heterodimers being the presumptive DNA binding complexes). Indeed, Beres et al has shown that the Ptf1a/E-protein heterodimer is not transcriptionally active in the absence of its trimeric partner, Rbpj or Rbpjl (Beres et al., 2006). Alternatively, Ascl1 and Ptf1a may form a novel heterodimer complex which renders Ascl1 inactive. Support for Ptf1a repressing transcriptional activity of Ascl1 was seen in transcription reporter assays where addition of Ptf1a decreased Ascl1 activated luciferase activity of an E-box reporter construct (Obata et al., 2001). Proving that Ptf1a and Ascl1 compete for similar E-boxes will be challenging, because this will rely on in vitro binding assays with careful protein titrations. Testing the hypothesis that Ascl1 and Ptf1a can form a complex can be done by co-immunopreciptiation assays. If Ascl1 and Ptf1a do form a novel complex, testing if they bind the same DNA region at the same time *in vivo* would require performing ChIP for one factor, followed by the sequential ChIP of the other factor (ChIP-reChip assay). However, ChIP-rechip, requires a significant amount chromatin material and efficient antibodies, making this experiment technically infeasible using neural tube tissue.

Ascl1 repression of target genes

Historically, in the neural tube, Ascl1 has been shown to be a transcriptional activator (Nakada et al., 2004), but in Chapter two, I show that several genes, particularly genes that are involved in stem cell maintenance are repressed by Ascl1. In support of this repression model, the introduction of Ascl1 into normal human astrocytes (NHA) resulted in the reduced expression levels of the negative regulator of wnt signaling, DKK1 (Rheinbay et al., 2013). Moreover, the binding of Ascl1 near DKK1 in Glioblastoma cells correlated with the a H3K4m1 histone marker, but the H3K27ac was absent, which indicates that the DNA enhancer is inactive or poised. How Ascl1 achieves this repression is unknown. One possible model is that Ascl1 recruits transcriptional repressors such as Prdm13 to these sites as was shown for Ascl1 repression of HD factor Tlx3 during specification of the neurons in the dorsal spinal cord (Chang et al., 2013). The RNA-Seq data provides a small candidate list of known transcriptional repressors that are expressed specifically in the Ascl1 lineage cells. Thus, biochemical assays such as Co-IPs and ChIP can be performed on these candidates to determine if 1) they directly interact with Ascl1 and 2) they bind the same DNA elements as Ascl1 in the neural tube. These experiments will begin show mechanistically how Ascl1 can directly repress a target gene.

Additional cooperative factors of Ascl1 and Ptf1a

In this dissertation research, I found enrichment of additional motifs that surround the primary binding sites of Ascl1 and Ptf1a. Identifying which specific factors bind to these regions is critical for our understanding of how these bHLH factors regulate specific genes. The experimental design here has provided the field with a wealth of information to identify

potential candidates. First, RNA-Seq data from each tissue type provides a short list of transcription factors expressed in the relevant cell types. Second, ChIP-Seq data have identified putative neuronal enhancers, such as the enhancers tested in Chapter three (Fig. 3.4), and a recently published Ascl1 responsive enhancer associated with Tlx3 (Fig. 2.4D (Chang et al., 2013)). This approach can be used to further examine the enhancers identified here that are bound by Ascl1 or Ptf1a. Reporter assays can be used to determine if these other motifs identified as being enriched along with Ascl1/Ptf1a's E-box are necessary and sufficient for activity and are capable of directing appropriate expression. Thus, in combination of these reporter assays and through the expression data, one can begin to uncover which specific co-factors are necessary to co-regulate an Ascl1 or Ptf1a specific target gene. These experiments will advance the field by identifying new components of the transcription factor network that work together with Ascl1 and Ptf1a in the different cellular contexts in which they function.

Requirement for Foxa2 in SCLC

One of the more interesting co-factors for ASCL1 found in the SCLC experiments was Foxa2. Outside the central nervous system, Foxa2 appears to co-regulate targets with either Ascl1 in SCLC or Ptf1a in pancreas. A closer examination of the relationship between Foxa2 and the bHLH factors is definitely needed. In the lung cancer field, several important questions about Foxa2 should be addressed. In SCLC and neuroendocrine NSCLC tumors, Ascl1 is used as a biomarker because its expression strongly correlates with most neuroendocrine tumors of the lung. Moreover, patients with Ascl1 expressing tumors show a

very poor prognosis (collaboration with John Minna Lab, data not shown). Thus, I would predict that Foxa2 is also a good biomarker for neuroendocrine lung tumors and that the presence Foxa2 would predict a poor outcome for patients. I am currently testing this hypotheses in collaboration with Dr. John Minna's lab. In the SCLC cell lines, I also predict that Foxa2 is required for the malignant behavior of the tumor cells, because many Foxa2 bound sites overlapped with Ascl1 bound sites (Fig 4.4B), and a subset of these included sites near key oncogenes (Fig 4.4F-H). This hypothesis can be tested by knocking down Foxa2 in the SCLC lines and determine if these cells phenocopy the loss of Ascl1.

Requirement for Ascl1 and Neurod1 in mouse models of SCLC

In chapter one I mention that in mouse models of SCLC, Ascl1 is up-regulated in the tumor cells. Genetic mouse models of cancer are powerful because one can study the initiation and progression of the cancer. Several important questions relating to Ascl1 or Neurod1 can be addressed in these mouse models of SCLC. For example, one can determine if either bHLH factor is required in tumor formation. Members of Johnson lab have begun generating the SCLC tumor models by using mice that contain loxP sites flanking the Rb, TP53, and p130 genes (Schaffer et al., 2010). These mice as adults are injected with adenoviral Cre into the trachea, making a conditional mutant and creating the SCLC tumors within a few months. In addition, these mice have been crossed to other mice containing loxP sites around Ascl1 or Neurod1 locus. These models with test whether Ascl1 or Neurod1 are required for tumor growth in vivo. Preliminary results indicate the absolute requirement for Ascl1 for SCLC formation. Surprisingly, Neurod1 is not required for tumor formation, and furthermore, in the

absence of Neurod1, the tumors are larger and cover more surface area in the lung. Further experiments with these mouse models, such as performing ChIP-Seq and RNA-Seq to compare with the human cell lines, will be used to uncover a better understanding of the lineages involved in SCLC formation and possible new pharmacological targets to stop tumor growth.

APPENDIX

App. A. Table of direct Ascl1 and Ptf1a targets genes identified in the mouse embryonic neural tube (See Chapter 2 for details).

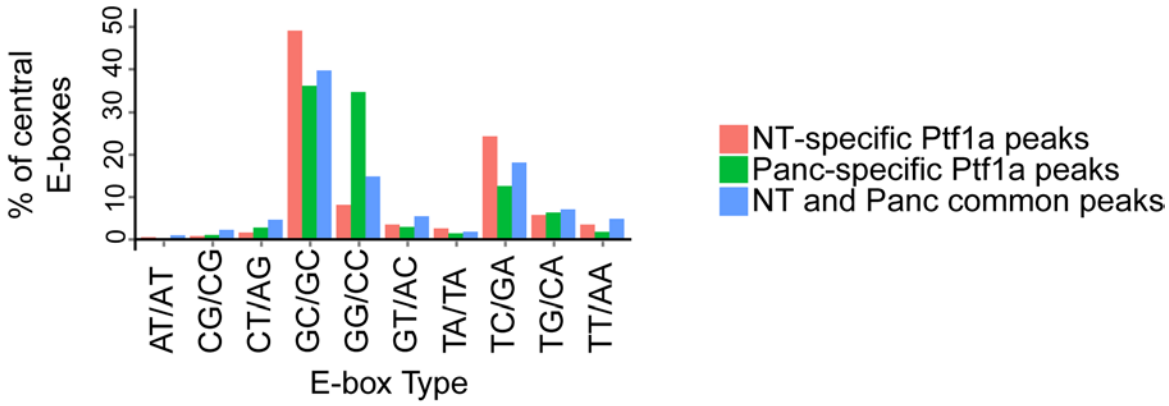
Ascl1 Activated Targets		Ascl1 Repressed Targets		Ptf1a Activated Targets	Ptf1a Repressed Targets
1810041L15Rik	Mtus1	Abcc4	Lama1	1700025G04Rik	2410004P03Rik
2300009A05Rik	Mxra7	Acsbg1	Lck	3110035E14Rik	2510009E07Rik
2510009E07Rik	Myo16	Adamts1	Lcp1	Abat	4931429L15Rik
6430548M08Rik	Myo18a	Adcy8	Lhfp	Acss1	Abcc8
Adcyap1	Myo1b	Adk	Lnx2	Adamts20	Adcyap1
Add2	Myt1	Alcam	Lrig1	Adamts4	Ajap1
Adra2a	Nav1	Angpt1	Lrp4	Adamts9	Ap1s3
Ajap1	Nbl1	Apcdd1	Lsamp	Adamtsl1	Arg1
Aldh1b1	Ndrg1	Arhgap19	Lsm2	Adra1a	Arpp21
Ank2	Necab3	Arhgef26	Ltbp1	Asap3	Atp2b2
Anks1	Nefl	Arl4a	Lypd6	B3gat1	Btbd16
Apba1	Neil1	Atp1a1	Mastl	Cacna1g	Cabp7
Arhgap20	Neurod6	Bard1	Mcm6	Cacna2d2	Calb2
Atoh8	Neurog2	Bhlhe40	Mdc1	Cacna2d3	Casz1
Bcas1	Nfasc	Bmpr1a	Mdfi	Cartpt	Cbln1
Bcl11b	Nol4	Boc	Melk	Cbs	Cbln2
Bcl7a	Nova2	Brca1	Mgst1	Ccbe1	Cd22
Brsk2	Nphs1	Brip1	Mier1	Ccnd2	Cdc42ep3
Btbd17	Ntsr2	Cachd1	Mki67	Cdh11	Chl1
C530008M17Rik	Numbl	Cad	Mms22l	Chgb	Chrna4
Cabp1	Optc	Casp8	Mpdz	Clstn2	Chrnb4
Cacna2d1	Pak3	Ccnd1	Mthfd1l	Dhrs3	Colec12
Camk2n1	Pcdh1	Ccnd2	Mycn	Dmbx1	Crabp1
Cbfa2t2	Pdzrn3	Ccne1	Nedd1	Dram2	Dlgap3
Cbfa2t3	Pdzrn4	Cdca2	Nek6	Dusp26	Dlk1
Cbln1	Pik3r1	Cdca7	Nfia	Dusp4	Dll3
Cbln2	Plekha6	Cdca8	Nfib	Eif4e3	Dmrta2
Cdc25b	Plk3	Cdh11	Nfix	Elmod2	E130309F12Rik
Cdk5r1	Plxna2	Cdk1	Notch2	Epha10	Ebf2
Cdk5r2	Pnmal1	Cdon	Nr4a3	Esrrg	Ebf3
Cdkn1a	Pnmal2	Celsr1	Ntn1	Fam102b	Edil3
Cdkn1c	Podxl2	Cep135	Ntrk2	Fndc5	Efna3
Celf4	Pou3f1	Chd1l	Oat	Foxd3	Elfn1
Chrd	Pou4f1	Chek1	Pard3b	Gad1	Elmo1
Chrna4	Ppp1r17	Clspn	Pax3	Gbx1	Ephb1
Cited2	Ppp1r9b	Cobl	Pcdh18	Gbx2	Eya2

Coro2b	Prox1	Col2a1	Pcsk6	Glcci1	Fam65b
Cotl1	Ptchd2	Crb2	Pdlim1	Gpc3	Fam78b
Ctif	Rap1gap	Cspg5	Pdpn	Gpc4	Fgd3
Cux2	Rapgef5	Cxcr4	Phactr2	Gpm6a	Fstl5
Cyth1	Rassf4	Cyr61	Phf17	Grasp	Gadd45g
Dact3	Rbfox3	Dcn	Phyhipl	Hectd2	Gap43
Dcc	Rbm24	Dctd	Plce1	Id1	Glyctk
Ddc	Rgmb	Ddit4l	Plekhg1	Id4	Hip1r
Diras2	Rgs16	Ddx11	Plk4	Ildr2	Homer2
Dll1	Rgs3	Dhrs3	Plod2	Iqsec3	Hspa12a
Dll3	Rgs8	Diap3	Plscr1	Kirrel	Kif26b
Dock4	Rhbdl3	Dock1	Pmf1	Kirrel2	L1cam
Dpysl4	Rnf112	Dsn1	Polq	Klhl14	Lars2
Dscam	Rufy3	Dtl	Ppap2b	Lhx1	Lingo1
Dusp4	Rusc1	Dtna	Prdm16	Lhx5	Lmx1b
Ebf2	Scrt2	E2f7	Prdm5	Lipg	Mapk10
Ebf3	Sdk2	E2f8	Prkcdbp	Mafb	Mbp
Efhd2	Sema6b	Ect2	Pros1	Map3k13	Megf11
Egr1	Sema7a	Ednrb	Psat1	Mybl1	Mfng
Elavl4	Setbp1	Efnb2	Ptgfrn	Myo7a	Nbl1
Epha10	Sez6l	Egln3	Ptn	Nphs1	Nckap5
Epha2	Sgip1	Eps8	Rab32	Nptx1	Necab2
Ephb3	Sgk1	Ets1	Rad18	Nrxn3	Nefl
Esyt1	Shb	Fabp7	Rbm43	Nuak1	Nell2
Eya2	Shb	Fat1	Rfx4	Pax2	Neurod4
Fam109a	Shb	Fbln2	Rreb1	Pax8	Ngfr
Fam110a	Shb	Fbxo48	Sae1	Pde9a	Nptx2
Fam171a2	Shox2	Fen1	Sall1	Pgbd5	Nrn1
Fam20c	Skil	Fgfr1	Serpine2	Pid1	Nrp2
Fbrsl1	Skor1	Fgfr2	Serpinh1	Plcxd2	Ntng2
Fbxl16	Slc2a10	Fignl1	Sfrp2	Pmp22	Ntrk3
Fcho1	Slit1	Fn1	Shroom3	Ppfibp1	Olfm1
Fgd3	Smug1	Foxm1	Slc1a3	Prdm13	Otp
Fgfr4	Snai1	Frem1	Slc22a23	Prickle2	Pde2a
Fry	Snap25	Fstl1	Slc25a13	Prkcb	Pde8b
Gadd45g	Sobp	Fut10	Slc25a18	Ptf1a	Pdgfra
Gap43	Sox10	Fut9	Slc35f1	Ptprd	Pdzrn4
Gata3	Sox4	Fzd10	Slc43a1	Rab3c	Peg10
Gatsl2	Spire2	Fzd8	Sox2	Rap1gap2	Phox2b
Gbx1	Spsb4	Galk1	Sox6	Rapgef5	Pou3f1
Glyctk	Srrm3	Gas1	Sox9	Rassf4	Pou4f1
Gpr156	Srrm4	Gdf10	Sparcl1	Rnf144b	Pou4f2
Gpr56	St8sia3	Gldc	Spata13	Runx1t1	Ppp1r12b
Grasp	Stat3	Gli3	Spc24	Samd5	Prickle1
Grik3	Stk32c	Gm5506	Spon1	Sel1l3	Rapgef4
Gse1	Stk33	Gm5506	Stim1	Sema3f	Rtn4r
Gsg1l	Stxbp1	Gm5506	Suclg2	Sez6	Scn3b
Gsx2	Syt13	Gm5506	Sulf1	Shisa2	Sez6l

Hdac5	Syt6	Gpr123	Sypl	Skor1	Shb
Hes5	Tagln3	Gpr126	Tanc1	Skor2	Shb
Hes6	Tbc1d30	Gpr83	Tcf7l2	Slc32a1	Shb
Hoxb3	Tbca	Gpr98	Tfap2c	Slc37a1	Shb
Hoxb4	Tcerg1l	Gpx7	Thbs1	Slc6a5	Shox2
Hoxc4	Tfap2b	Gramd3	Thsd4	Slco3a1	Slc17a6
Igdcc4	Tfap2e	Grb10	Thsd4	Sorcs3	Slc4a1
Igfbpl1	Tfeb	Grid2	Thsd4	Spsb1	Slitrk1
Igsf8	Th	Gsr	Thsd4	Sst	Socs2
Insm1	Tle3	Hat1	Tjp1	Sv2c	Spock1
Isl1	Tlx1	Haus4	Tmem135	Tfap2b	Syt13
Jakmip3	Tlx3	Hhip	Tns3	Tnik	Tcerg1l
Jph3	Tmcc2	Hhipl1	Tomm7	Tril	Tfap2e
Kcnab2	Tmem163	Hmcn1	Trip6	Tuba1c	Th
Kcnh2	Tmem169	Hopx	Ulk4	Vcan	Thsd7a
Kcnq2	Tmem219	Id2	Umps	Zbtb16	Tigit
Kcnq4	Trappc1	Igf2	Vamp3	Zfp804a	Tle2
Kirrel2	Trim8	Igfbp5	Vim		Tlx1
Klhl29	Trim9	Igsf11	Vldlr		Tlx3
Lbx1	Trp53i11	Igsf9b	Wls		Tshz1
Lhx1	Tspan11	Ildr2	Wnt1		Uncx
Lmx1b	Ttbk1	Impa2	Wnt4		Vstm2l
Lzts1	Tubb2a	Inhbb	Wnt5a		Wscd2
Map3k13	Uncx	Iqgap2	Wnt7a		Zfp488
Mapk10	Vav2	Irx2	Xpo4		Zfp703
Mfng	Wscd2	Itgb8	Zfp36l1		
Mgat5b	Zfhx2	Jag1	Zfp395		
Mkrn3	Zfhx3	Jun	Zic4		
Mllt11	Zswim5	Kank2	Zic5		
Mpp3		Kbtbd11	Zranb3		
		Kbtbd8	Zwilch		
		Kcnk5			

App. B. Genes from Venn diagrams in figure 3F for Ascl1 and Ptf1a regulated genes.

Ascl1 Repressed, Ptf1a Repressed	Ascl1 Repressed, Ptf1a Activated		Ascl1 Activated, Ptf1a Activated	Ascl1 Repressed, Ptf1a Repressed
Accn4	1300002K09Rik	Mapk10	A930038C07Rik	Rmrp
Cabp1	2510009E07Rik	Mbp	Ccnd2	
Chgb	9030425E11Rik	Mfng	Cdh11	
Cttnbp2	Adcyap1	Mgat4c	Dhrs3	
Dusp4	Adra2a	Nbl1	Gpc3	
Epha10	Ajap1	Necab2	Gpc4	
Gad2	C130021I20Rik	Nefl	Ildr2	
Gbx1	Cacna2d1	Nefm	Rnf144b	
Gm14204	Calb2	Ngfr		
Grasp	Cbln1	Otp		
Hsd11b2	Cbln2	Pdzrn4		
Kirrel2	Cdc42ep3	Phox2b		
Lhx1	Chrna3	Pou3f1		
Lhx5	Chrna4	Pou4f1		
Mafb	D10Bwg1379e	Pou4f2		
Map3k13	Dlgap3	Prrxl1		
March4	Dlk1	Scn3b		
Myo5b	Dll3	Sez6l		
Nphs1	Ebf2	Shb		
Nptx1	Ebf3	Shox2		
Pax2	Eya2	Slc17a6		
Ptf1a	Fam65b	Sncg		
Rap1gap2	Fgd3	Syt13		
Rapgef5	Fgd5	Tcerg1l		
Rassf4	Fstl5	Tfap2e		
Runx1t1	Gabra2	Th		
Skor1	Gadd45g	Tlx1		
Skor2	Gap43	Tlx3		
Slc30a3	Glyctk	Tmem163		
Slc32a1	Gm2694	Uncx		
Tfap2b	Lhx2	Wscd2		
Tnik	Lmx1b			



App. C. Distribution of E-box motifs of the central E-box in Ptf1a ChIP-seq peaks in the developing pancreas and neural tube.

App. D. These tables provides additional information regarding the Ptf1a ChIP-Seq regions tested as transgenes in Chapter 3

.

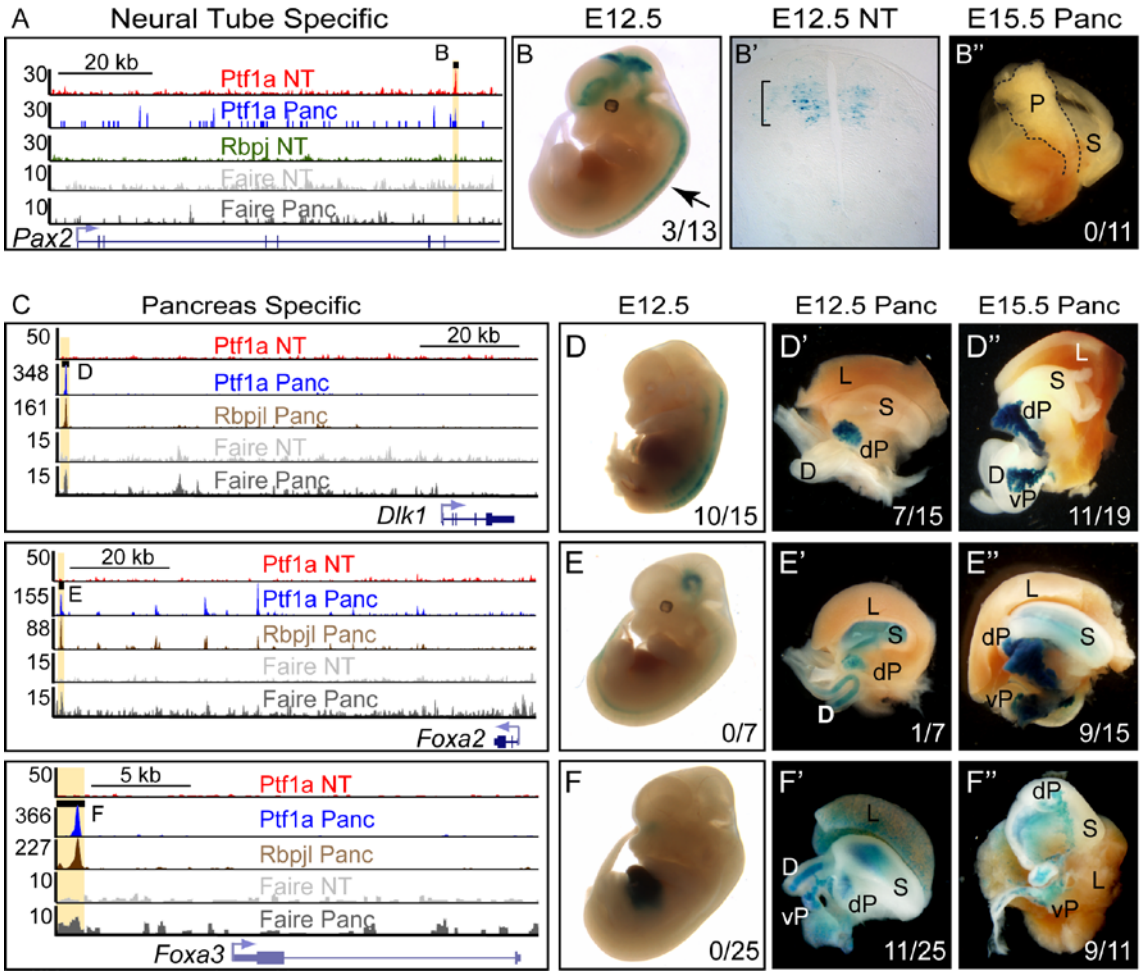
	Binding Site Parameters			Gene mRNA Levels		Activity Tests in Transgenic Embryos			
Target Gene	# PTF1 sites[&]	Peak center dist. to TSS	Position	FPKM E12.5 NT	FPKM E15.5 Panc	E12.5 NT #Exp/#Tg	E12.5 Panc #Exp/#Tg	E15.5 Panc #Exp/#Tg	Coordinates of TG regions
Lhx1	1	-60311	5'	91.1	0.35	4/6	0/6	*1/8	chr11:84,398,942-84,399,901
Tcfap2b	2	233	5' UTR	92.3	0.05	8/15	*1/15	*1/8	chr1:19,198,766-19,199,711
Prdm13	1	76	5' UTR	30.5	0.00	0/9	0/9	0/5	chr4:21,612,631-21,613,611
Rapgef5	1	172419	intron	14.6	3.91	0/7	1/7	0/11	chr12:118,927,145-118,928,069
Pax2	1	62959	intron	111.7	0.01	3/13	0/13	0/11	chr19:44,894,465-44,895,473
Nrxn3	1^	many	intron	19.0	0.37	0/9	1/9	0/9	chr12:90,387,054-90,388,512
Neurog2	2	1913	3' UTR	29.3	0.00	6/7**	ND	ND	chr3:127,337,666-127,338,713
Kirrel2	2	-714, -343	5'	47.4	21.24	5/8	6/8	4/13	chr7:31,242,523-31,243,477
Nphs1		-2200, -1829	5'	14.0	8.08				
Clps	2	-1186	5'	0	7500	0/15	8/15	12/15	chr17:28,698,416-28,699,365
Foxa3	1^	15761	3'	0	17.3	0/25	11/25	9/11	chr7:19,592,371-19,593,360
Cpa1	1	53	5' UTR	0.03	4723	0/5	5/(5)	8/10	chr6:30,588,360-30,589,398
Dlk1	1	-50914	5'	22.5	173	10/15	7/15	11/19	chr12:110,640,256-110,641,250
Foxa2	2	78286	3'	0.31	16.8	0/7	1/7	9/15	chr2:147,793,982-147,794,893
Aqp12	1	-60	5' UTR	0	19.9	0/6	0/6	0/10	chr1:94,902,374-94,903,348
Cbs	1^	5485	3'	2.3	111	0/10	*1/10	*1/20	chr17:31,768,102-31,769,173
Nupr1	1	-5582	5'	0.3	814	0/19	*1/19	0/10	chr7:133,774,257-133,774,981
Rbpjl	2	-96	5' UTR	0.03	98	ND	ND	8/22***	chr2:164,227,316-164,228,795

Neural tube only Ptf1a ChIP-seq peaks
Neural tube and Pancreas shared Ptf1a ChIP-seq peaks
Pancreas only Ptf1a ChIP-seq peaks

[&] number of PTF1 binding sites fitting the consensus CANNTG(N10-14,20-24)YTCYC.
^ The PTF1 site below the center of the peak diverged sufficiently from the consensus to be unrecognized by the search algorithm.
#Exp = the number of embryos expressing lacZ in a Ptf1a pattern in each tissue (* is just a few cells in the pancreas)
#Tg = the number of transgenic embryos obtained
**Data for E11.5 from Simmons et al. 2001
***Data for E17.5 from Masui et al. 2007
ND = not done

Gene	PTF1 motif	spacing	from peak center
Lhx1	tgagtCAGCTGtctcacTTCTCACAgacat	13	0
Tcfap2b	gcattCATGTGaggagCAGATGtctggtTTCTCAGGactcc	13, 24	1
Prdm13	agggtCACATGgcgtggTTCCCTCAgcccc	13	23
Rapgef5	gcagcCAGATGgactcagTTCCCAGTgaaag	14	54
Pax2	ggatcCAGATGgatcagccggagatgTTCCCAGGctgag	22	32
Nrxn3	none	-	-
Neurog2	gctggCATCTGctctaTTCCCATTcccat	12	18
	tttgaCAGGTGaaaTTCCCACAgcctg	10	51
Kirrel2	cagccCAGCTGggcgTTCCCACGcagcg	11	36
Nphs1	cagcaCAGCTGttcagtctgttggttTTCCCATCcctcg	22	35
Clps	tccagCAGCTGgactcTTCCCAGCctgca	12	23
	gtgctCAGGTGtGCAAACAgagaagctTTCCCTGAaggtc	23	5
Foxa3	none	-	-
Cpa1	ctctcCACCTGccttgTTCCCTGAtactt	12	5
Dlk1	TGTGTACtCAGCTGgaggccagccagcTTCCCATAgggaa	20	5
Foxa2	cgaacCACCTGtgtttgctcactagctTTCCCCAAtctgc	23	11
	cctgcCATGTGaccggctctggccattTTCTCGCAgattg	23	45
Aqp12	ggtcaCAGGTGtgggtgTTCCCAGTagcca	13	27
Cbs	none	-	-
Nupr1	ccaggCATTTGagtcTTCCCTGGaccag	11	62
Rbpjl	agggaCACCTGctgggCAGATGtaggcTTCCCACGgcagt	12, 23	32

Underlined = Fox sites

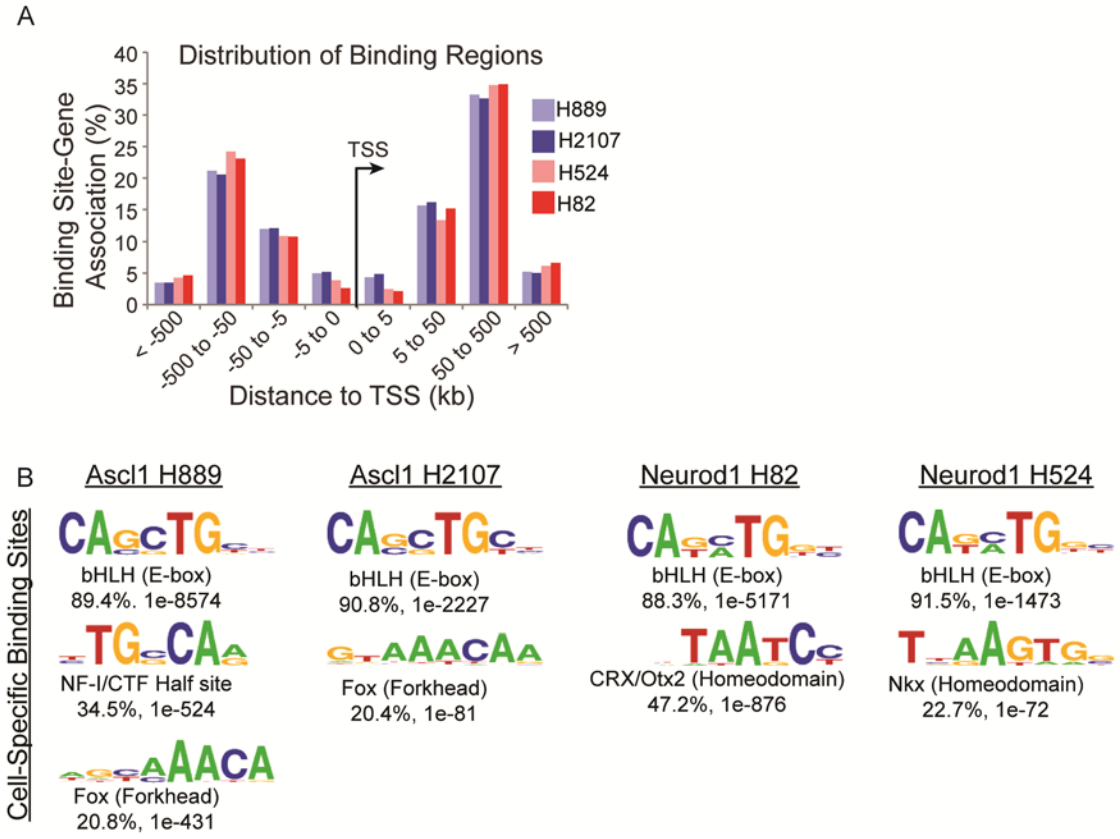


App. E. Tissue specificity of Ptf1a-bound regions is retained in transgenic embryos.

App. F. Table of expression levels of Fox and Sox family members in neural tube and pancreas.

Gene	E12.5 NT Ptf1a FPKM	E15.5 Pancreas FPKM	E17.5 Pancreas FPKM
Foxa1	0.17	0.06	0.00
Foxa2	**0.31**	16.75	24.98
Foxa3	**0.00**	17.32	35.25
Foxb1	5.25	0.02	0.00
Foxb2	1.68	0.03	0.06
Foxc1	3.73	0.31	0.81
Foxc2	3.57	0.33	0.96
Foxd1	4.42	0.03	0.00
Foxd2	0.96	0.02	0.14
Foxd3	3.14	0.29	1.38
Foxd4	0.04	0.02	0.00
Foxe1	0.00	0.00	0.00
Foxe3	0.63	0.00	0.00
Foxf1a	1.21	0.94	1.84
Foxf2	1.13	0.09	0.42
Foxg1	0.12	0.02	0.08
Foxh1	0.19	0.09	0.17
Foxi1	0.00	0.00	0.00
Foxi2	0.06	0.00	0.00
Foxi3	0.00	0.00	0.00
Foxj1	1.53	0.64	0.84
Foxj2	3.93	4.69	5.05
Foxj3	10.47	9.00	7.35
Foxk1	9.90	5.01	5.52
Foxk2	28.46	13.03	11.94
Foxl1	0.38	0.11	0.34
Foxl2	0.07	0.00	0.04
Foxl2os	0.11	0.01	0.08
Foxm1	20.48	14.53	9.28
Foxn1	0.03	0.01	0.04
Foxn2	3.60	6.91	3.31
Foxn3	10.16	2.47	1.42
Foxn4	0.16	0.06	0.00
Foxo1	1.34	6.14	5.81
Foxo3	5.62	4.04	4.25
Foxo4	9.09	19.72	19.87
Foxo6	3.84	1.04	2.04
Foxp1	8.00	7.80	6.12
Foxp2	8.18	1.72	1.72
Foxp3	0.06	0.02	0.03
Foxp4	55.39	13.33	28.98

Foxq1	0.29	0.01	0.03
Foxr1	1.01	0.73	1.42
Foxr2	0.04	0.03	0.00
Sox1	**15.66**	0.02	0.13
Sox10	3.31	2.30	4.05
Sox11	**135.83**	3.18	1.76
Sox12	55.33	12.04	18.76
Sox13	17.99	6.47	9.16
Sox14	**1.56**	0.04	0.00
Sox15	0.04	0.28	1.25
Sox17	0.42	1.25	4.33
Sox18	2.09	6.18	22.53
Sox2	**88.82**	0.09	0.04
Sox21	**12.33**	0.11	0.00
Sox2ot	5.02	0.05	0.55
Sox3	**83.17**	0.05	0.04
Sox30	0.11	0.08	0.30
Sox4	171.09	17.60	15.10
Sox5	2.09	1.09	0.69
Sox6	3.69	4.36	2.30
Sox7	0.31	3.19	3.92
Sox8	1.51	0.65	0.63
Sox9	21.17	14.07	19.50

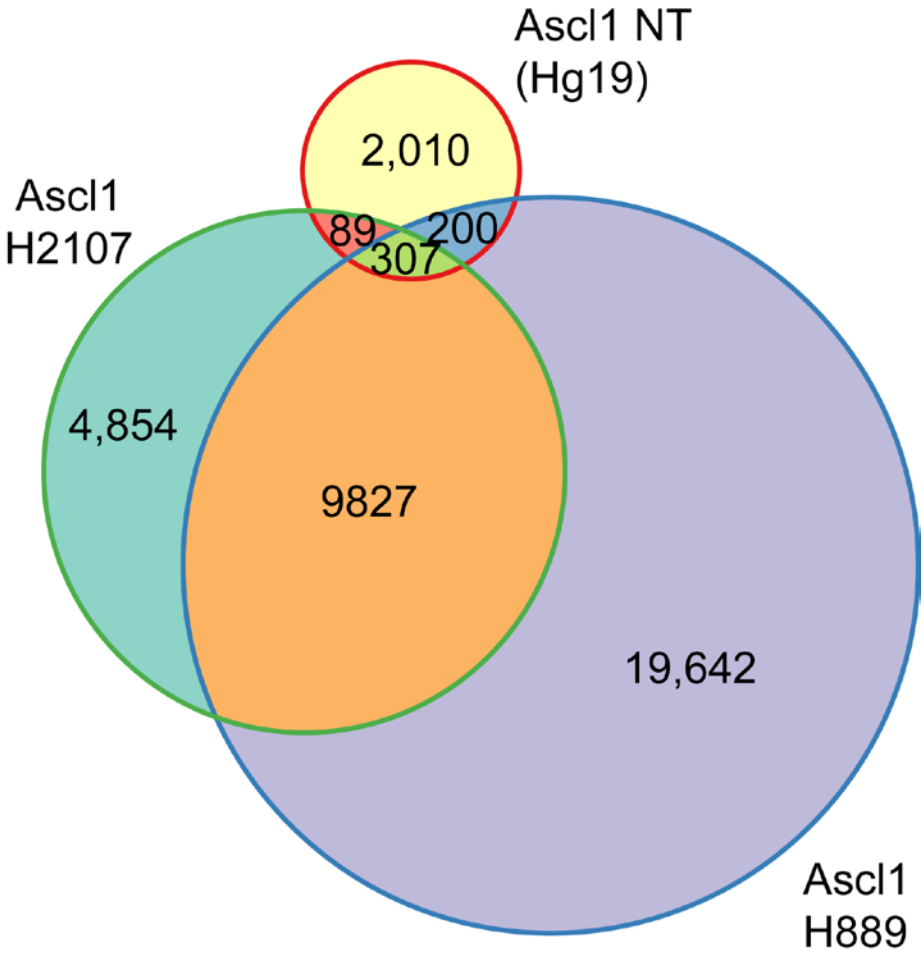


App. G. ASCL1 and NEUROD1 bind distal enhancers in SCLC and have enrichment cell-specific motifs. (A) ASCL1 and NEUROD1 binding in all SCLC cell lines with respect to its assigned genes TSS. Peaks were assigned to gene using the software GREAT (McLean et al., 2010). (B) The significantly enriched motifs found by *de novo* motif analysis in cell-specific binding sites.

App. H. List of Potential ASCL1 and NEUROD1 targets in SCLC. An ASCL1 target gene is determined by the presence of ASCL1 binding site and enrichment in gene expression in ASCL1^{Hi} over NEUROD1^{HI} cells, and vice versa for NEUROD1 targets. A shared target is defined by have a common ASCL1 and NEUROD1 site and having a gene expression greater than 10 FPKM in all four cell lines. Note, tables are separated by either ASCL1 targets in ASCL1^{Hi} cells, NEUROD1 targets in NEUROD1^{HI} cells, or ASCL1 and NEUROD1 shared targets in all four cell lines. ASCL1 targets also include sites that overlap with FOXA2 binding sites. ChIP-Seq binding sites are assigned with a unique Peak ID and how far the binding sites are away from the genes transcription start site.

Ascl1 Targets				Neurod1 Targets		Shared Targets		
ABLIM1	EFR3B	INPPL1	SDR42E1	ABCC4	PREX1	ABHD12	HES6	RBM17
ABLIM2	ELAVL3	INSM1	SEC11C	ABHD15	PTPRU	AGAP3	HIBADH	RGS16
ABLIM2	ELF3	ISL1	SETBP1	ADCY7	RGS10	ALG2	HMGA1	RNF220
AES	ELN	KCNA1	SEZ6L	ADORA2B	S100A11	ANAPC13	HOOK1	RPL19
AMACR	EPCAM	KCNF1	SGPP2	ANK1	SCN3B	ANKH	ICT1	RPL38
ASCL1	EPHX1	KCNMB2	SH3BP5	ARL4D	SDC2	ASPSCR1	IFNAR1	RPLP1
ATP2A3	ERO1LB	KIT	SH3PXD2A	ASS1	SLC1A7	ASUN	IFNAR2	SCMH1
B3GNT5	ESRP1	KITLG	SHANK3	BHLHE40	SLC35F4	ATP5G3	INSM1	SCRN1
BARX1	ETS2	KRT15	SHISA2	BIN1	SLC38A5	ATP6AP2	IRF2BP2	SEC22B
BCL2	ETV6	KRT7	SHISA8	CAMKK1	SLC39A11	ATP6V1B2	ITPK1	SEC61A1
BEST3	FAM110B	LFNG	SLC12A7	CDCA7L	SNX9	BAG5	JUND	SEPT9
BTBD11	FAM155A	LGALS3BP	SLC37A1	CDK6	SPATS2L	BCL2L1	KATNA1	SERINC1
BTBD17	FAM163B	LGALS8	SLC7A4	CERKL	SSTR2	BCL7A	KIAA0182	SH3GLB2
C19orf63	FAM19A5	LHX2	SLCO3A1	CHRDL1	STOM	BCOR	KLC1	SHKBP1
C1RL	FAM3B	LIFR	SMAD7	CLDN10	SYT6	BRD4	KLF13	SLC25A19
C9orf64	FBLN7	LMO3	SMOC2	CLSTN2	THY1	C19orf10	LHFP	SLC44A1
CA8	FGD3	MAP6D1	SMPD3	CORO2B	TMEM54	C19orf63	MAP2	SLC7A5
CACNA1A	FGF9	MAPK4	SNCAIP	DNAJC15	TRIB1	C6orf1	MAPRE3	SOX4
CAMK1D	FLI1	MBNL3	SOX1	EEF2K	ZC4H2	C8orf83	MARCKS	SRI
CAMK2N2	FOXA2	MBP	SOX2	FAM171A1	ZNF217	CADM1	MCCC1	SRSF5
CBLN2	FOXC1	MMP11	SPINT2	FHIT		CALM1	MCL1	SSBP3
CCDC3	FZD8	NCALD	ST14	GJC1		CALM2	MGAT5	STK25
CDH2	FZD9	NEURL1B	ST18	GLCE		CBFA2T2	MID1IP1	STX6
CDKN1A	GCK	NFIX	ST7	GRK5		CCDC130	MLXIP	STXBP1
CDKN2C	GFI1	NKAIN2	ST8SIA5	HPCA		CDC45	MRPL34	SYT1
CHMP4C	GJB1	NKX2-1	STMN3	ISL2		CDKAL1	MRPL37	TCF12
CLDN3	GNG8	NOL4	SYT4	KIRREL2		CENPO	MRPS23	TGFBR1
CLDN4	GPR56	NPTXR	TAGLN3	LINGO1		CHD3	MSI2	TLE3
CLDN5	GPRC5B	NR0B2	TESC	LMO1		CHGA	MUDENG	TM9SF2
CMIP	GPX2	ODZ3	TFCP2L1	LOXL1		COG2	NARS2	TMEM177
CMPK2	GRB7	OVOL2	TFF3	LRFN1		DLL3	NCAM1	TMEM66
CNKSR3	GRIP1	PAK7	TMEM132D	MAP2K6		DPYSL5	NCS1	TMUB1
COTL1	GRP	PANX2	TMEM163	MAP4K1		DSTN	NEDD4L	TMX1
CRIP2	HEPACAM2	PCP4	TMEM184A	MTHFD1L		E2F1	NFATC3	TOMM20
CSRP1	HMSD	PLEKHA2	TMEM61	MYC		E2F2	NNT	TRAF4
CXCR4	HOXB2	PLEKHG4B	TNFRSF11A	MYO10		EIF3I	OAZ2	TRAF7
CYB5A	HOXB3	POU4F1	TNS3	MYOD1		ELAVL1	ODC1	TRIM9
CYYR1	HOXB4	PRKAR1B	TOX	NEUROD1		ELOVL5	PAPOLA	TSC22D2
DBH	HOXB5	PROX1	TOX3	NEUROD4		ENY2	PEX5L	TTC35
DDC	HOXB7	PTPRN2	TRIB2	NHLH1		EPB41	PFN2	TXNDC11
DGKB	HOXB8	RAB3IP	USP18	NXN		EPT1	PGS1	UBE2Q1
DLL1	HOXD10	RASL11B	UTRN	OLFM1		ERAL1	PHC2	UMPS
DLL4	HOXD11	RBFOX2	VASH1	OTX2		EXOC7	PJA2	VRK1
DLX1	HTATIP2	RET	VPS4B	PCOLCE2		EXOSC5	POLE4	WSB1
DLX2	HYI	RGL3	WNT11	PDE2A		EZR	POR	ZC3H7A
DLX5	ICA1	RNF183	ZBTB7C	PELI2		FBXO21	PTDSS1	ZNF238
DMRTA2	ID2	S100A6	ZIC2	PITX1		G6PC3	PTRHD1	ZNF259
DNALI1	ID4	SCG2	ZIC5	PLCD3		GLRX5	PVRL1	ZNF362
DNER	IGFBP5	SCN3A	ZMAT4	PNMA2		GNAZ	PVRL2	
DSP	INHBB	SCNN1A	ZSCAN30	POU3F1		GPR107	PYGB	
						HDAC5	RAB2A	

App. I. Specific Fox, NF-I, and OTX2/CRX factors are expressed in SCLC. (A-C) Expression levels, determined by RNA-Seq, of the Forkhead, NF-I (B), and Homeodomain (CRX/OTX2) factors in our SCLC cell lines.



App. J. Venn diagram shows the overlapping ASCL1 binding sites in embryonic neural tube and SCLC. Note mouse neural tube binding sites were lifted over from mm9 to hg19 genome.

Bibliography

AHNFELT-RONNE, J., JORGENSEN, M. C., KLINCK, R., JENSEN, J. N., FUCHTBAUER, E. M., DEERING, T., MACDONALD, R. J., WRIGHT, C. V., MADSEN, O. D. & SERUP, P. 2012. Ptf1a-mediated control of Dll1 reveals an alternative to the lateral inhibition mechanism. *Development,* 139**,** 33-45.

AKAGI, T., INOUE, T., MIYOSHI, G., BESSHO, Y., TAKAHASHI, M., LEE, J. E., GUILLEMOT, F. & KAGEYAMA, R. 2004. Requirement of multiple basic helix-loop-helix genes for retinal neuronal subtype specification. *J Biol Chem,* 279**,** 28492-8.

ALAYNICK, W. A., JESSELL, T. M. & PFAFF, S. L. 2011. SnapShot: spinal cord development. *Cell,* 146**,** 178-178 e1.

AUGUSTINE, K. A., LIU, E. T. & SADLER, T. W. 1995. Interactions of Wnt-1 and Wnt-3a are essential for neural tube patterning. *Teratology,* 51**,** 107-19.

AVRAHAM, O., HADAS, Y., VALD, L., HONG, S., SONG, M. R. & KLAR, A. 2010. Motor and dorsal root ganglion axons serve as choice points for the ipsilateral turning of dI3 axons. *J Neurosci,* 30**,** 15546-57.

BADIS, G., BERGER, M. F., PHILIPPAKIS, A. A., TALUKDER, S., GEHRKE, A. R., JAEGER, S. A., CHAN, E. T., METZLER, G., VEDENKO, A., CHEN, X., KUZNETSOV, H., WANG, C. F., COBURN, D., NEWBURGER, D. E., MORRIS, Q., HUGHES, T. R. & BULYK, M. L. 2009. Diversity and complexity in DNA recognition by transcription factors. *Science,* 324**,** 1720-3.

BAILEY, P. J., KLOS, J. M., ANDERSSON, E., KARLEN, M., KALLSTROM, M., PONJAVIC, J., MUHR, J., LENHARD, B., SANDELIN, A. & ERICSON, J. 2006. A global genomic transcriptional code associated with CNS-expressed genes. *Exp Cell Res,* 312**,** 3108-19.

BALL, D. W. 2004. Achaete-scute homolog-1 and Notch in lung neuroendocrine development and cancer. *Cancer Lett,* 204**,** 159-69.

BALL, D. W., AZZOLI, C. G., BAYLIN, S. B., CHI, D., DOU, S., DONIS-KELLER, H., CUMARASWAMY, A., BORGES, M. & NELKIN, B. D. 1993. Identification of a human achaete-scute homolog highly expressed in neuroendocrine tumors. *Proc Natl Acad Sci U S A,* 90**,** 5648-52.

BATISTA, M. F. & LEWIS, K. E. 2008. Pax2/8 act redundantly to specify glycinergic and GABAergic fates of multiple spinal interneurons. *Dev Biol,* 323**,** 88-97.

BENAYOUN, B. A., CABURET, S., DIPIETROMARIA, A., BAILLY-BECHET, M., BATISTA, F., FELLOUS, M., VAIMAN, D. & VEITIA, R. A. 2008. The identification and characterization of a FOXL2 response element provides insights into the pathogenesis of mutant alleles. *Hum Mol Genet,* 17**,** 3118-27.

BERES, T. M., MASUI, T., SWIFT, G. H., SHI, L., HENKE, R. M. & MACDONALD, R. J. 2006. PTF1 is an organ-specific and Notch-independent basic helix-loop-

helix complex containing the mammalian Suppressor of Hairless (RBP-J) or its paralogue, RBP-L. *Mol Cell Biol,* 26**,** 117-30.

BERGSLAND, M., RAMSKOLD, D., ZAOUTER, C., KLUM, S., SANDBERG, R. & MUHR, J. 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes Dev,* 25**,** 2453-64.

BERGSLAND, M., WERME, M., MALEWICZ, M., PERLMANN, T. & MUHR, J. 2006. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes Dev,* 20**,** 3475-86.

BERTRAND, N., CASTRO, D. S. & GUILLEMOT, F. 2002. Proneural genes and the specification of neural cell types. *Nat Rev Neurosci,* 3**,** 517-30.

BLAUGRUND, E., PHAM, T. D., TENNYSON, V. M., LO, L., SOMMER, L., ANDERSON, D. J. & GERSHON, M. D. 1996. Distinct subpopulations of enteric neuronal progenitors defined by time of development, sympathoadrenal lineage markers and Mash-1-dependence. *Development,* 122**,** 309-20.

BORGES, M., LINNOILA, R. I., VAN DE VELDE, H. J., CHEN, H., NELKIN, B. D., MABRY, M., BAYLIN, S. B. & BALL, D. W. 1997. An achaete-scute homologue essential for neuroendocrine differentiation in the lung. *Nature,* 386**,** 852-5.

BUNT, J., HASSELT, N. E., ZWIJNENBURG, D. A., HAMDI, M., KOSTER, J., VERSTEEG, R. & KOOL, M. 2012. OTX2 directly activates cell cycle genes and inhibits differentiation in medulloblastoma cells. *Int J Cancer,* 131**,** E21-32.

BUNT, J., HASSELT, N. E., ZWIJNENBURG, D. A., KOSTER, J., VERSTEEG, R. & KOOL, M. 2011. Joint binding of OTX2 and MYC in promotor regions is associated with high gene expression in medulloblastoma. *PLoS One,* 6**,** e26058.

CAGLE, M. C. & HONIG, M. G. 2013. Parcellation of Cblns 1, 2, and 4 among different subpopulations of dorsal horn neurons in mouse spinal cord. *J Comp Neurol*.

CARNEY, D. N., GAZDAR, A. F., BEPLER, G., GUCCION, J. G., MARANGOS, P. J., MOODY, T. W., ZWEIG, M. H. & MINNA, J. D. 1985. Establishment and identification of small cell lung cancer cell lines having classic and variant features. *Cancer Res,* 45**,** 2913-23.

CASAROSA, S., FODE, C. & GUILLEMOT, F. 1999. Mash1 regulates neurogenesis in the ventral telencephalon. *Development,* 126**,** 525-34.

CASTRO, D. S., MARTYNOGA, B., PARRAS, C., RAMESH, V., PACARY, E., JOHNSTON, C., DRECHSEL, D., LEBEL-POTTER, M., GARCIA, L. G., HUNT, C., DOLLE, D., BITHELL, A., ETTWILLER, L., BUCKLEY, N. & GUILLEMOT, F. 2011. A novel function of the proneural factor Ascl1 in progenitor proliferation identified by genome-wide characterization of its targets. *Genes Dev,* 25**,** 930-45.

CASTRO, D. S., SKOWRONSKA-KRAWCZYK, D., ARMANT, O., DONALDSON, I. J., PARRAS, C., HUNT, C., CRITCHLEY, J. A., NGUYEN, L., GOSSLER, A., GOTTGENS, B., MATTER, J. M. & GUILLEMOT, F. 2006. Proneural bHLH and Brn proteins coregulate a neurogenic program through cooperative binding to a conserved DNA motif. *Dev Cell,* 11**,** 831-44.

CAU, E., GRADWOHL, G., FODE, C. & GUILLEMOT, F. 1997. Mash1 activates a cascade of bHLH regulators in olfactory neuron progenitors. *Development,* 124**,** 1611-21.

CHANG, J. C., MEREDITH, D. M., MAYER, P. R., BORROMEO, M. D., LAI, H. C., OU, Y. H. & JOHNSON, J. E. 2013. Prdm13 mediates the balance of inhibitory and excitatory neurons in somatosensory circuits. *Dev Cell,* 25**,** 182-95.

CHEN, L., CHAO, S. B., WANG, Z. B., QI, S. T., ZHU, X. L., YANG, S. W., YANG, C. R., ZHANG, Q. H., OUYANG, Y. C., HOU, Y., SCHATTEN, H. & SUN, Q. Y. 2012. Checkpoint kinase 1 is essential for meiotic cell cycle regulation in mouse oocytes. *Cell Cycle,* 11**,** 1948-55.

CHENG, L., ARATA, A., MIZUGUCHI, R., QIAN, Y., KARUNARATNE, A., GRAY, P. A., ARATA, S., SHIRASAWA, S., BOUCHARD, M., LUO, P., CHEN, C. L., BUSSLINGER, M., GOULDING, M., ONIMARU, H. & MA, Q. 2004. Tlx3 and Tlx1 are post-mitotic selector genes determining glutamatergic over GABAergic cell fates. *Nat Neurosci,* 7**,** 510-7.

CHENG, L., SAMAD, O. A., XU, Y., MIZUGUCHI, R., LUO, P., SHIRASAWA, S., GOULDING, M. & MA, Q. 2005. Lbx1 and Tlx3 are opposing switches in determining GABAergic versus glutamatergic transmitter phenotypes. *Nat Neurosci,* 8**,** 1510-5.

CHO, J. H. & TSAI, M. J. 2004. The role of BETA2/NeuroD1 in the development of the nervous system. *Mol Neurobiol,* 30**,** 35-47.

COCKELL, M., STEVENSON, B. J., STRUBIN, M., HAGENBUCHLE, O. & WELLAUER, P. K. 1989. Identification of a cell-specific DNA-binding activity that interacts with a transcriptional activator of genes expressed in the acinar pancreas. *Mol Cell Biol,* 9**,** 2464-76.

CORBO, J. C., LAWRENCE, K. A., KARLSTETTER, M., MYERS, C. A., ABDELAZIZ, M., DIRKES, W., WEIGELT, K., SEIFERT, M., BENES, V., FRITSCHE, L. G., WEBER, B. H. & LANGMANN, T. 2010. CRX ChIP-seq reveals the cis-regulatory architecture of mouse photoreceptors. *Genome Res,* 20**,** 1512-25.

D'ANGELO, S. P. & PIETANZA, M. C. 2010. The molecular pathogenesis of small cell lung cancer. *Cancer Biol Ther,* 10**,** 1-10.

DE HOON, M. J., IMOTO, S., NOLAN, J. & MIYANO, S. 2004. Open source clustering software. *Bioinformatics,* 20**,** 1453-4.

DING, Y. Q., YIN, J., KANIA, A., ZHAO, Z. Q., JOHNSON, R. L. & CHEN, Z. F. 2004. Lmx1b controls the differentiation and migration of the superficial dorsal horn neurons of the spinal cord. *Development,* 131**,** 3693-703.

DOOLEY, A. L., WINSLOW, M. M., CHIANG, D. Y., BANERJI, S., STRANSKY, N., DAYTON, T. L., SNYDER, E. L., SENNA, S., WHITTAKER, C. A., BRONSON, R. T., CROWLEY, D., BARRETINA, J., GARRAWAY, L., MEYERSON, M. & JACKS, T. 2011. Nuclear factor I/B is an oncogene in small cell lung cancer. *Genes Dev,* 25**,** 1470-5.

FARNHAM, P. J. 2009. Insights from genomic profiling of transcription factors. *Nat Rev Genet,* 10**,** 605-16.

FITZGERALD, M. 2005. The development of nociceptive circuits. *Nat Rev Neurosci,* 6**,** 507-20.
FONG, A. P., YAO, Z., ZHONG, J. W., CAO, Y., RUZZO, W. L., GENTLEMAN, R. C. & TAPSCOTT, S. J. 2012. Genetic and epigenetic determinants of neurogenesis and myogenesis. *Dev Cell,* 22**,** 721-35.
FRANKEL, N., DAVIS, G. K., VARGAS, D., WANG, S., PAYRE, F. & STERN, D. L. 2010. Phenotypic robustness conferred by apparently redundant transcriptional enhancers. *Nature,* 466**,** 490-3.
FUJITANI, Y., FUJITANI, S., LUO, H., QIU, F., BURLISON, J., LONG, Q., KAWAGUCHI, Y., EDLUND, H., MACDONALD, R. J., FURUKAWA, T., FUJIKADO, T., MAGNUSON, M. A., XIANG, M. & WRIGHT, C. V. 2006. Ptf1a determines horizontal and amacrine cell fates during mouse retinal development. *Development,* 133**,** 4439-50.
FUKAYA, M., KAMATA, A., HARA, Y., TAMAKI, H., KATSUMATA, O., ITO, N., TAKEDA, S., HATA, Y., SUZUKI, T., WATANABE, M., HARVEY, R. J. & SAKAGAMI, H. 2011. SynArfGEF is a guanine nucleotide exchange factor for Arf6 and localizes preferentially at post-synaptic specializations of inhibitory synapses. *J Neurochem,* 116**,** 1122-37.
GAO, N., LELAY, J., VATAMANIUK, M. Z., RIECK, S., FRIEDMAN, J. R. & KAESTNER, K. H. 2008. Dynamic regulation of Pdx1 enhancers by Foxa1 and Foxa2 is essential for pancreas development. *Genes Dev,* 22**,** 3435-48.
GARRAWAY, L. A. & SELLERS, W. R. 2006. Lineage dependency and lineage-survival oncogenes in human cancer. *Nat Rev Cancer,* 6**,** 593-602.
GIRESI, P. G., KIM, J., MCDANIELL, R. M., IYER, V. R. & LIEB, J. D. 2007. FAIRE (Formaldehyde-Assisted Isolation of Regulatory Elements) isolates active regulatory elements from human chromatin. *Genome Res,* 17**,** 877-85.
GLASGOW, S. M., HENKE, R. M., MACDONALD, R. J., WRIGHT, C. V. & JOHNSON, J. E. 2005. Ptf1a determines GABAergic over glutamatergic neuronal cell fate in the spinal cord dorsal horn. *Development,* 132**,** 5461-9.
GOHLKE, J. M., ARMANT, O., PARHAM, F. M., SMITH, M. V., ZIMMER, C., CASTRO, D. S., NGUYEN, L., PARKER, J. S., GRADWOHL, G., PORTIER, C. J. & GUILLEMOT, F. 2008. Characterization of the proneural gene regulatory network during mouse telencephalon development. *BMC Biol,* 6**,** 15.
GOWAN, K., HELMS, A. W., HUNSAKER, T. L., COLLISSON, T., EBERT, P. J., ODOM, R. & JOHNSON, J. E. 2001. Crossinhibitory activities of Ngn1 and Math1 allow specification of distinct dorsal interneurons. *Neuron,* 31**,** 219-32.
GRADWOHL, G., FODE, C. & GUILLEMOT, F. 1996. Restricted expression of a novel murine atonal-related bHLH protein in undifferentiated neural precursors. *Dev Biol,* 180**,** 227-41.
GRAY, P. A., FU, H., LUO, P., ZHAO, Q., YU, J., FERRARI, A., TENZEN, T., YUK, D. I., TSUNG, E. F., CAI, Z., ALBERTA, J. A., CHENG, L. P., LIU, Y., STENMAN, J. M., VALERIUS, M. T., BILLINGS, N., KIM, H. A., GREENBERG, M. E., MCMAHON, A. P., ROWITCH, D. H., STILES, C. D. &

MA, Q. 2004. Mouse brain organization revealed through direct genome-scale TF expression analysis. *Science,* 306**,** 2255-7.

GRIMALDI, P., PARRAS, C., GUILLEMOT, F., ROSSI, F. & WASSEF, M. 2009. Origins and control of the differentiation of inhibitory interneurons and glia in the cerebellum. *Dev Biol,* 328**,** 422-33.

GRONBORG, M., PAVLOS, N. J., BRUNK, I., CHUA, J. J., MUNSTER-WANDOWSKI, A., RIEDEL, D., AHNERT-HILGER, G., URLAUB, H. & JAHN, R. 2010. Quantitative comparison of glutamatergic and GABAergic synaptic vesicles unveils selectivity for few proteins including MAL2, a novel synaptic vesicle protein. *J Neurosci,* 30**,** 2-12.

GROSS, M. K., DOTTORI, M. & GOULDING, M. 2002. Lbx1 specifies somatosensory association interneurons in the dorsal spinal cord. *Neuron,* 34**,** 535-49.

GUILLEMOT, F., LO, L. C., JOHNSON, J. E., AUERBACH, A., ANDERSON, D. J. & JOYNER, A. L. 1993. Mammalian achaete-scute homolog 1 is required for the early development of olfactory and autonomic neurons. *Cell,* 75**,** 463-76.

GUINEE, D. G., JR., FISHBACK, N. F., KOSS, M. N., ABBONDANZO, S. L. & TRAVIS, W. D. 1994. The spectrum of immunohistochemical staining of small-cell lung carcinoma in specimens from transbronchial and open-lung biopsies. *Am J Clin Pathol,* 102**,** 406-14.

GUNNERSEN, J. M., KUEK, A., PHIPPS, J. A., HAMMOND, V. E., PUTHUSSERY, T., FLETCHER, E. L. & TAN, S. S. 2009. Seizure-related gene 6 (Sez-6) in amacrine cells of the rodent retina and the consequence of gene deletion. *PLoS One,* 4**,** e6546.

HANN, C. L., DANIEL, V. C., SUGAR, E. A., DOBROMILSKAYA, I., MURPHY, S. C., COPE, L., LIN, X., HIERMAN, J. S., WILBURN, D. L., WATKINS, D. N. & RUDIN, C. M. 2008. Therapeutic efficacy of ABT-737, a selective inhibitor of BCL-2, in small cell lung cancer. *Cancer Res,* 68**,** 2321-8.

HEINZ, S., BENNER, C., SPANN, N., BERTOLINO, E., LIN, Y. C., LASLO, P., CHENG, J. X., MURRE, C., SINGH, H. & GLASS, C. K. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Mol Cell,* 38**,** 576-89.

HEINZ, S. & GLASS, C. K. 2012. Roles of lineage-determining transcription factors in establishing open chromatin: lessons from high-throughput studies. *Curr Top Microbiol Immunol,* 356**,** 1-15.

HELMS, A. W., BATTISTE, J., HENKE, R. M., NAKADA, Y., SIMPLICIO, N., GUILLEMOT, F. & JOHNSON, J. E. 2005. Sequential roles for Mash1 and Ngn2 in the generation of dorsal spinal cord interneurons. *Development,* 132**,** 2709-19.

HELMS, A. W. & JOHNSON, J. E. 2003. Specification of dorsal spinal cord interneurons. *Curr Opin Neurobiol,* 13**,** 42-9.

HENKE, R. M., MEREDITH, D. M., BORROMEO, M. D., SAVAGE, T. K. & JOHNSON, J. E. 2009a. Ascl1 and Neurog2 form novel complexes and regulate Delta-like3 (Dll3) expression in the neural tube. *Dev Biol,* 328**,** 529-40.

HENKE, R. M., SAVAGE, T. K., MEREDITH, D. M., GLASGOW, S. M., HORI, K., DUMAS, J., MACDONALD, R. J. & JOHNSON, J. E. 2009b. Neurog2 is a

direct downstream target of the Ptf1a-Rbpj transcription complex in dorsal spinal cord. *Development,* 136**,** 2945-54.

HIROSHIMA, K., IYODA, A., SHIDA, T., SHIBUYA, K., IIZASA, T., KISHI, H., TANIZAWA, T., FUJISAWA, T. & NAKATANI, Y. 2006. Distinction of pulmonary large cell neuroendocrine carcinoma from small cell lung carcinoma: a morphological, immunohistochemical, and molecular analysis. *Mod Pathol,* 19**,** 1358-68.

HIRSCH, M. R., TIVERON, M. C., GUILLEMOT, F., BRUNET, J. F. & GORIDIS, C. 1998. Control of noradrenergic differentiation and Phox2a expression by MASH1 in the central and peripheral nervous system. *Development,* 125**,** 599-608.

HONG, J. W., HENDRIX, D. A. & LEVINE, M. S. 2008. Shadow enhancers as a source of evolutionary novelty. *Science,* 321**,** 1314.

HORI, K., CHOLEWA-WACLAW, J., NAKADA, Y., GLASGOW, S. M., MASUI, T., HENKE, R. M., WILDNER, H., MARTARELLI, B., BERES, T. M., EPSTEIN, J. A., MAGNUSON, M. A., MACDONALD, R. J., BIRCHMEIER, C. & JOHNSON, J. E. 2008. A nonclassical bHLH Rbpj transcription factor complex is required for specification of GABAergic neurons independent of Notch signaling. *Genes Dev,* 22**,** 166-78.

HORI, K. & HOSHINO, M. 2012. GABAergic neuron specification in the spinal cord, the cerebellum, and the cochlear nucleus. *Neural Plast,* 2012**,** 921732.

HOSHINO, M., NAKAMURA, S., MORI, K., KAWAUCHI, T., TERAO, M., NISHIMURA, Y. V., FUKUDA, A., FUSE, T., MATSUO, N., SONE, M., WATANABE, M., BITO, H., TERASHIMA, T., WRIGHT, C. V., KAWAGUCHI, Y., NAKAO, K. & NABESHIMA, Y. 2005. Ptf1a, a bHLH transcriptional gene, defines GABAergic neuronal fates in cerebellum. *Neuron,* 47**,** 201-13.

HUANG, H. S., KUBISH, G. M., REDMOND, T. M., TURNER, D. L., THOMPSON, R. C., MURPHY, G. G. & UHLER, M. D. 2010. Direct transcriptional induction of Gadd45gamma by Ascl1 during neuronal differentiation. *Mol Cell Neurosci,* 44**,** 282-96.

HUANG, M., HUANG, T., XIANG, Y., XIE, Z., CHEN, Y., YAN, R., XU, J. & CHENG, L. 2008. Ptf1a, Lbx1 and Pax2 coordinate glycinergic and peptidergic transmitter phenotypes in dorsal spinal inhibitory neurons. *Dev Biol,* 322**,** 394-405.

ITO-ISHIDA, A., MIYAZAKI, T., MIURA, E., MATSUDA, K., WATANABE, M., YUZAKI, M. & OKABE, S. 2012. Presynaptically released Cbln1 induces dynamic axonal structural changes by interacting with GluD2 during cerebellar synapse formation. *Neuron,* 76**,** 549-64.

ITO, T., UDAKA, N., YAZAWA, T., OKUDELA, K., HAYASHI, H., SUDO, T., GUILLEMOT, F., KAGEYAMA, R. & KITAMURA, H. 2000. Basic helix-loop-helix transcription factors regulate the neuroendocrine differentiation of fetal mouse pulmonary epithelium. *Development,* 127**,** 3913-21.

JIANG, T., COLLINS, B. J., JIN, N., WATKINS, D. N., BROCK, M. V., MATSUI, W., NELKIN, B. D. & BALL, D. W. 2009. Achaete-scute complex homologue 1

regulates tumor-initiating capacity in human small cell lung cancer. *Cancer Res,* 69**,** 845-54.

JOHN, A., WILDNER, H. & BRITSCH, S. 2005. The homeodomain transcription factor Gbx1 identifies a subpopulation of late-born GABAergic interneurons in the developing dorsal spinal cord. *Dev Dyn,* 234**,** 767-71.

JOHNSON, B. E., RUSSELL, E., SIMMONS, A. M., PHELPS, R., STEINBERG, S. M., IHDE, D. C. & GAZDAR, A. F. 1996. MYC family DNA amplification in 126 tumor cell lines from patients with small cell lung cancer. *J Cell Biochem Suppl,* 24**,** 210-7.

JOHNSON, D. S., MORTAZAVI, A., MYERS, R. M. & WOLD, B. 2007. Genome-wide mapping of in vivo protein-DNA interactions. *Science,* 316**,** 1497-502.

KAWAGUCHI, Y., COOPER, B., GANNON, M., RAY, M., MACDONALD, R. J. & WRIGHT, C. V. 2002. The role of the transcriptional regulator Ptf1a in converting intestinal to pancreatic progenitors. *Nat Genet,* 32**,** 128-34.

KIM, E. J., LEUNG, C. T., REED, R. R. & JOHNSON, J. E. 2007. In vivo analysis of Ascl1 defined progenitors reveals distinct developmental dynamics during adult neurogenesis and gliogenesis. *J Neurosci,* 27**,** 12764-74.

KIM, Y. H., GIRARD, L., GIACOMINI, C. P., WANG, P., HERNANDEZ-BOUSSARD, T., TIBSHIRANI, R., MINNA, J. D. & POLLACK, J. R. 2006. Combined microarray analysis of small cell lung cancer reveals altered apoptotic balance and distinct expression signatures of MYC family gene amplification. *Oncogene,* 25**,** 130-8.

KLISCH, T. J., XI, Y., FLORA, A., WANG, L., LI, W. & ZOGHBI, H. Y. 2011. In vivo Atoh1 targetome reveals how a proneural transcription factor regulates cerebellar development. *Proc Natl Acad Sci U S A,* 108**,** 3288-93.

KOSARI, F., IDA, C. M., AUBRY, M. C., YANG, L., KOVTUN, I. V., KLEIN, J. L., LI, Y., ERDOGAN, S., TOMASZEK, S. C., MURPHY, S. J., BOLETTE, L. C., KOLBERT, C. P., YANG, P., WIGLE, D. A. & VASMATZIS, G. 2013. ASCL1 and RET expression defines a clinically relevant subgroup of lung adenocarcinoma characterized by neuroendocrine differentiation. *Oncogene*.

KRAPP, A., KNOFLER, M., LEDERMANN, B., BURKI, K., BERNEY, C., ZOERKLER, N., HAGENBUCHLE, O. & WELLAUER, P. K. 1998. The bHLH protein PTF1-p48 is essential for the formation of the exocrine and the correct spatial organization of the endocrine pancreas. *Genes Dev,* 12**,** 3752-63.

LADEWIG, J., KOCH, P. & BRUSTLE, O. 2013. Leveling Waddington: the emergence of direct programming and the loss of cell fate hierarchies. *Nat Rev Mol Cell Biol,* 14**,** 225-36.

LAI, H. C., KLISCH, T. J., ROBERTS, R., ZOGHBI, H. Y. & JOHNSON, J. E. 2011. In vivo neuronal subtype-specific targets of Atoh1 (Math1) in dorsal spinal cord. *J Neurosci,* 31**,** 10859-71.

LANIGAN, T. M., DERAAD, S. K. & RUSSO, A. F. 1998. Requirement of the MASH-1 transcription factor for neuroendocrine differentiation of thyroid C cells. *J Neurobiol,* 34**,** 126-34.

LEE, C. S., FRIEDMAN, J. R., FULMER, J. T. & KAESTNER, K. H. 2005a. The initiation of liver development is dependent on Foxa transcription factors. *Nature,* 435**,** 944-7.

LEE, C. S., SUND, N. J., BEHR, R., HERRERA, P. L. & KAESTNER, K. H. 2005b. Foxa2 is required for the differentiation of pancreatic alpha-cells. *Dev Biol,* 278**,** 484-95.

LEE, K. J. & JESSELL, T. M. 1999. The specification of dorsal cell fates in the vertebrate central nervous system. *Annu Rev Neurosci,* 22**,** 261-94.

LEUNG, C. T., COULOMBE, P. A. & REED, R. R. 2007. Contribution of olfactory neural stem cells to tissue maintenance and regeneration. *Nat Neurosci,* 10**,** 720-6.

LINNOILA, R. I., ZHAO, B., DEMAYO, J. L., NELKIN, B. D., BAYLIN, S. B., DEMAYO, F. J. & BALL, D. W. 2000. Constitutive achaete-scute homologue-1 promotes airway dysplasia and lung neuroendocrine tumors in transgenic mice. *Cancer Res,* 60**,** 4005-9.

LIU, Y. & MA, Q. 2011. Generation of somatic sensory neuron diversity and implications on sensory coding. *Curr Opin Neurobiol,* 21**,** 52-60.

LODATO, M. A., NG, C. W., WAMSTAD, J. A., CHENG, A. W., THAI, K. K., FRAENKEL, E., JAENISCH, R. & 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**,** e1003288.

LUU, B., ELLISOR, D. & ZERVAS, M. 2011. The lineage contribution and role of Gbx2 in spinal cord development. *PLoS One,* 6**,** e20940.

MACDONALD, R. J., SWIFT, G. H. & REAL, F. X. 2010. Transcriptional control of acinar development and homeostasis. *Prog Mol Biol Transl Sci,* 97**,** 1-40.

MAGNANI, L., EECKHOUTE, J. & LUPIEN, M. 2011. Pioneer factors: directing transcriptional regulators within the chromatin environment. *Trends Genet,* 27**,** 465-74.

MAIRET-COELLO, G., TURY, A., VAN BUSKIRK, E., ROBINSON, K., GENESTINE, M. & DICICCO-BLOOM, E. 2012. p57(KIP2) regulates radial glia and intermediate precursor cell cycle dynamics and lower layer neurogenesis in developing cerebral cortex. *Development,* 139**,** 475-87.

MARTYNOGA, B., MATEO, J. L., ZHOU, B., ANDERSEN, J., ACHIMASTOU, A., URBAN, N., VAN DEN BERG, D., GEORGOPOULOU, D., HADJUR, S., WITTBRODT, J., ETTWILLER, L., PIPER, M., GRONOSTAJSKI, R. M. & GUILLEMOT, F. 2013. Epigenomic enhancer annotation reveals a key role for NFIX in neural stem cell quiescence. *Genes Dev,* 27**,** 1769-86.

MARUYAMA, M., ICHISAKA, T., NAKAGAWA, M. & YAMANAKA, S. 2005. Differential roles for Sox15 and Sox2 in transcriptional control in mouse embryonic stem cells. *J Biol Chem,* 280**,** 24371-9.

MASSARI, M. E. & MURRE, C. 2000. Helix-loop-helix proteins: regulators of transcription in eucaryotic organisms. *Mol Cell Biol,* 20**,** 429-40.

MASUI, T., LONG, Q., BERES, T. M., MAGNUSON, M. A. & MACDONALD, R. J. 2007. Early pancreatic development requires the vertebrate Suppressor of Hairless (RBPJ) in the PTF1 bHLH complex. *Genes Dev,* 21**,** 2629-43.

MASUI, T., SWIFT, G. H., DEERING, T., SHEN, C., COATS, W. S., LONG, Q., ELSASSER, H. P., MAGNUSON, M. A. & MACDONALD, R. J. 2010. Replacement of Rbpj with Rbpjl in the PTF1 complex controls the final maturation of pancreatic acinar cells. *Gastroenterology,* 139**,** 270-80.

MASUI, T., SWIFT, G. H., HALE, M. A., MEREDITH, D. M., JOHNSON, J. E. & MACDONALD, R. J. 2008. Transcriptional autoregulation controls pancreatic Ptf1a expression during development and adulthood. *Mol Cell Biol,* 28**,** 5458-68.

MCLEAN, C. Y., BRISTOR, D., HILLER, M., CLARKE, S. L., SCHAAR, B. T., LOWE, C. B., WENGER, A. M. & BEJERANO, G. 2010. GREAT improves functional interpretation of cis-regulatory regions. *Nat Biotechnol,* 28**,** 495-501.

MEGASON, S. G. & MCMAHON, A. P. 2002. A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. *Development,* 129**,** 2087-98.

MEREDITH, A. & JOHNSON, J. E. 2000. Negative autoregulation of Mash1 expression in CNS development. *Dev Biol,* 222**,** 336-46.

MEREDITH, D. M., BORROMEO, M. D., DEERING, T. G., CASEY, B., SAVAGE, T. K., MAYER, P. R., HOANG, C., TUNG, K. C., KUMAR, M., SHEN, C., SWIFT, G. H., MACDONALD, R. J. & JOHNSON, J. E. 2013. Program specificity for Ptf1a in Pancreas versus Neural Tube Development correlates with distinct collaborating cofactors and chromatin accessibility. *Mol Cell Biol.*

MEREDITH, D. M., MASUI, T., SWIFT, G. H., MACDONALD, R. J. & JOHNSON, J. E. 2009. Multiple transcriptional mechanisms control Ptf1a levels during neural development including autoregulation by the PTF1-J complex. *J Neurosci,* 29**,** 11139-48.

MEUWISSEN, R., LINN, S. C., LINNOILA, R. I., ZEVENHOVEN, J., MOOI, W. J. & BERNS, A. 2003. Induction of small cell lung cancer by somatic inactivation of both Trp53 and Rb1 in a conditional mouse model. *Cancer Cell,* 4**,** 181-9.

MINAKI, Y., MIZUHARA, E., MORIMOTO, K., NAKATANI, T., SAKAMOTO, Y., INOUE, Y., SATOH, K., IMAI, T., TAKAI, Y. & ONO, Y. 2005. Migrating postmitotic neural precursor cells in the ventricular zone extend apical processes and form adherens junctions near the ventricle in the developing spinal cord. *Neurosci Res,* 52**,** 250-62.

MIZUGUCHI, R., KRIKS, S., CORDES, R., GOSSLER, A., MA, Q. & GOULDING, M. 2006. Ascl1 and Gsh1/2 control inhibitory and excitatory cell fate in spinal sensory interneurons. *Nat Neurosci,* 9**,** 770-8.

MIZUHARA, E., MINAKI, Y., NAKATANI, T., KUMAI, M., INOUE, T., MUGURUMA, K., SASAI, Y. & ONO, Y. 2010. Purkinje cells originate from cerebellar ventricular zone progenitors positive for Neph3 and E-cadherin. *Dev Biol,* 338**,** 202-14.

MORAN, C. A., SUSTER, S., COPPOLA, D. & WICK, M. R. 2009. Neuroendocrine carcinomas of the lung: a critical analysis. *Am J Clin Pathol,* 131**,** 206-21.

MULLER, T., ANLAG, K., WILDNER, H., BRITSCH, S., TREIER, M. & BIRCHMEIER, C. 2005. The bHLH factor Olig3 coordinates the specification of dorsal neurons in the spinal cord. *Genes Dev,* 19**,** 733-43.

MULLER, T., BROHMANN, H., PIERANI, A., HEPPENSTALL, P. A., LEWIN, G. R., JESSELL, T. M. & BIRCHMEIER, C. 2002. The homeodomain factor lbx1 distinguishes two major programs of neuronal differentiation in the dorsal spinal cord. *Neuron,* 34**,** 551-62.

MULLIGAN, L. M., TIMMER, T., IVANCHUK, S. M., CAMPLING, B. G., YOUNG, L. C., RABBITTS, P. H., SUNDARESAN, V., HOFSTRA, R. M. & ENG, C. 1998. Investigation of the genes for RET and its ligand complex, GDNF/GFR alpha-I, in small cell lung carcinoma. *Genes Chromosomes Cancer,* 21**,** 326-32.

MURRE, C., BAIN, G., VAN DIJK, M. A., ENGEL, I., FURNARI, B. A., MASSARI, M. E., MATTHEWS, J. R., QUONG, M. W., RIVERA, R. R. & STUIVER, M. H. 1994. Structure and function of helix-loop-helix proteins. *Biochim Biophys Acta,* 1218**,** 129-35.

MURRE, C., MCCAW, P. S., VAESSIN, H., CAUDY, M., JAN, L. Y., JAN, Y. N., CABRERA, C. V., BUSKIN, J. N., HAUSCHKA, S. D., LASSAR, A. B. & ET AL. 1989. Interactions between heterologous helix-loop-helix proteins generate complexes that bind specifically to a common DNA sequence. *Cell,* 58**,** 537-44.

MYATT, S. S. & LAM, E. W. 2007. The emerging roles of forkhead box (Fox) proteins in cancer. *Nat Rev Cancer,* 7**,** 847-59.

NAKADA, Y., HUNSAKER, T. L., HENKE, R. M. & JOHNSON, J. E. 2004. Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification. *Development,* 131**,** 1319-30.

NAKAZAKI, H., REDDY, A. C., MANIA-FARNELL, B. L., SHEN, Y. W., ICHI, S., MCCABE, C., GEORGE, D., MCLONE, D. G., TOMITA, T. & MAYANIL, C. S. 2008. Key basic helix-loop-helix transcription factor genes Hes1 and Ngn2 are regulated by Pax3 during mouse embryonic development. *Dev Biol,* 316**,** 510-23.

NEPTUNE, E. R., PODOWSKI, M., CALVI, C., CHO, J. H., GARCIA, J. G., TUDER, R., LINNOILA, R. I., TSAI, M. J. & DIETZ, H. C. 2008. Targeted disruption of NeuroD, a proneural basic helix-loop-helix factor, impairs distal lung formation and neuroendocrine morphology in the neonatal lung. *J Biol Chem,* 283**,** 21160-9.

NISHIDA, K., HOSHINO, M., KAWAGUCHI, Y. & MURAKAMI, F. 2010. Ptf1a directly controls expression of immunoglobulin superfamily molecules Nephrin and Neph3 in the developing central nervous system. *J Biol Chem,* 285**,** 373-80.

OBATA, J., YANO, M., MIMURA, H., GOTO, T., NAKAYAMA, R., MIBU, Y., OKA, C. & KAWAICHI, M. 2001. p48 subunit of mouse PTF1 binds to RBP-Jkappa/CBF-1, the intracellular mediator of Notch signalling, and is expressed in the neural tube of early stage embryos. *Genes Cells,* 6**,** 345-60.

ONG, C. T. & CORCES, V. G. 2011. Enhancer function: new insights into the regulation of tissue-specific gene expression. *Nat Rev Genet,* 12**,** 283-93.

OSADA, H., TATEMATSU, Y., YATABE, Y., HORIO, Y. & TAKAHASHI, T. 2005. ASH1 gene is a specific therapeutic target for lung cancers with neuroendocrine features. *Cancer Res,* 65**,** 10680-5.

OSADA, H., TOMIDA, S., YATABE, Y., TATEMATSU, Y., TAKEUCHI, T., MURAKAMI, H., KONDO, Y., SEKIDO, Y. & TAKAHASHI, T. 2008. Roles of

achaete-scute homologue 1 in DKK1 and E-cadherin repression and neuroendocrine differentiation in lung cancer. *Cancer Res,* 68**,** 1647-55.

OSBORNE, J. K., LARSEN, J. E., SHIELDS, M. D., GONZALES, J. X., SHAMES, D. S., SATO, M., KULKARNI, A., WISTUBA, II, GIRARD, L., MINNA, J. D. & COBB, M. H. 2013. NeuroD1 regulates survival and migration of neuroendocrine lung carcinomas via signaling molecules TrkB and NCAM. *Proc Natl Acad Sci U S A,* 110**,** 6524-9.

PARK, K. S., LIANG, M. C., RAISER, D. M., ZAMPONI, R., ROACH, R. R., CURTIS, S. J., WALTON, Z., SCHAFFER, B. E., ROAKE, C. M., ZMOOS, A. F., KRIEGEL, C., WONG, K. K., SAGE, J. & KIM, C. F. 2011. Characterization of the cell of origin for small cell lung cancer. *Cell Cycle,* 10**,** 2806-15.

PARRAS, C. M., GALLI, R., BRITZ, O., SOARES, S., GALICHET, C., BATTISTE, J., JOHNSON, J. E., NAKAFUKU, M., VESCOVI, A. & GUILLEMOT, F. 2004. Mash1 specifies neurons and oligodendrocytes in the postnatal brain. *EMBO J,* 23**,** 4495-505.

PARRAS, C. M., HUNT, C., SUGIMORI, M., NAKAFUKU, M., ROWITCH, D. & GUILLEMOT, F. 2007. The proneural gene Mash1 specifies an early population of telencephalic oligodendrocytes. *J Neurosci,* 27**,** 4233-42.

PASCUAL, M., ABASOLO, I., MINGORANCE-LE MEUR, A., MARTINEZ, A., DEL RIO, J. A., WRIGHT, C. V., REAL, F. X. & SORIANO, E. 2007. Cerebellar GABAergic progenitors adopt an external granule cell-like phenotype in the absence of Ptf1a transcription factor expression. *Proc Natl Acad Sci U S A,* 104**,** 5193-8.

PATEL, O., SHULKES, A. & BALDWIN, G. S. 2006. Gastrin-releasing peptide and cancer. *Biochim Biophys Acta,* 1766**,** 23-41.

PATTYN, A., SIMPLICIO, N., VAN DOORNINCK, J. H., GORIDIS, C., GUILLEMOT, F. & BRUNET, J. F. 2004. Ascl1/Mash1 is required for the development of central serotonergic neurons. *Nat Neurosci,* 7**,** 589-95.

PELLERIN, I., SCHNABEL, C., CATRON, K. M. & ABATE, C. 1994. Hox proteins have different affinities for a consensus DNA site that correlate with the positions of their genes on the hox cluster. *Mol Cell Biol,* 14**,** 4532-45.

PERSSON, M., STAMATAKI, D., TE WELSCHER, P., ANDERSSON, E., BOSE, J., RUTHER, U., ERICSON, J. & BRISCOE, J. 2002. Dorsal-ventral patterning of the spinal cord requires Gli3 transcriptional repressor activity. *Genes Dev,* 16**,** 2865-78.

PEVNY, L. H. & NICOLIS, S. K. 2010. Sox2 roles in neural stem cells. *Int J Biochem Cell Biol,* 42**,** 421-4.

PILLAI, A., MANSOURI, A., BEHRINGER, R., WESTPHAL, H. & GOULDING, M. 2007. Lhx1 and Lhx5 maintain the inhibitory-neurotransmitter status of interneurons in the dorsal spinal cord. *Development,* 134**,** 357-66.

PUTAALA, H., SOININEN, R., KILPELAINEN, P., WARTIOVAARA, J. & TRYGGVASON, K. 2001. The murine nephrin gene is specifically expressed in kidney, brain and pancreas: inactivation of the gene leads to massive proteinuria and neonatal death. *Hum Mol Genet,* 10**,** 1-8.

RHEINBAY, E., SUVA, M. L., GILLESPIE, S. M., WAKIMOTO, H., PATEL, A. P., SHAHID, M., OKSUZ, O., RABKIN, S. D., MARTUZA, R. L., RIVERA, M. N., LOUIS, D. N., KASIF, S., CHI, A. S. & BERNSTEIN, B. E. 2013. An aberrant transcription factor network essential for Wnt signaling and stem cell maintenance in glioblastoma. *Cell Rep,* 3**,** 1567-79.

ROCK, J. R. & HOGAN, B. L. 2011. Epithelial progenitor cells in lung development, maintenance, repair, and disease. *Annu Rev Cell Dev Biol,* 27**,** 493-512.

ROSE, S. D., SWIFT, G. H., PEYTON, M. J., HAMMER, R. E. & MACDONALD, R. J. 2001. The role of PTF1-P48 in pancreatic acinar gene expression. *J Biol Chem,* 276**,** 44018-26.

ROSS, S. E. 2011. Pain and itch: insights into the neural circuits of aversive somatosensation in health and disease. *Curr Opin Neurobiol,* 21**,** 880-7.

ROULET, E., BUSSO, S., CAMARGO, A. A., SIMPSON, A. J., MERMOD, N. & BUCHER, P. 2002. High-throughput SELEX SAGE method for quantitative modeling of transcription-factor binding sites. *Nat Biotechnol,* 20**,** 831-5.

SAEED, A. I., SHAROV, V., WHITE, J., LI, J., LIANG, W., BHAGABATI, N., BRAISTED, J., KLAPA, M., CURRIER, T., THIAGARAJAN, M., STURN, A., SNUFFIN, M., REZANTSEV, A., POPOV, D., RYLTSOV, A., KOSTUKOVICH, E., BORISOVSKY, I., LIU, Z., VINSAVICH, A., TRUSH, V. & QUACKENBUSH, J. 2003. TM4: a free, open-source system for microarray data management and analysis. *Biotechniques,* 34**,** 374-8.

SAITO, T., LO, L., ANDERSON, D. J. & MIKOSHIBA, K. 1996. Identification of novel paired homeodomain protein related to C. elegans unc-4 as a potential downstream target of MASH1. *Dev Biol,* 180**,** 143-55.

SALDANHA, A. J. 2004. Java Treeview--extensible visualization of microarray data. *Bioinformatics,* 20**,** 3246-8.

SATHIRA, N., YAMASHITA, R., TANIMOTO, K., KANAI, A., ARAUCHI, T., KANEMATSU, S., NAKAI, K., SUZUKI, Y. & SUGANO, S. 2010. Characterization of transcription start sites of putative non-coding RNAs by multifaceted use of massively paralleled sequencer. *DNA Res,* 17**,** 169-83.

SCHAFFER, B. E., PARK, K. S., YIU, G., CONKLIN, J. F., LIN, C., BURKHART, D. L., KARNEZIS, A. N., SWEET-CORDERO, E. A. & SAGE, J. 2010. Loss of p130 accelerates tumor development in a mouse model for human small-cell lung carcinoma. *Cancer Res,* 70**,** 3877-83.

SIEGEL, R., NAISHADHAM, D. & JEMAL, A. 2013. Cancer statistics, 2013. *CA Cancer J Clin,* 63**,** 11-30.

STERGACHIS, A. B., NEPH, S., REYNOLDS, A., HUMBERT, R., MILLER, B., PAIGE, S. L., VERNOT, B., CHENG, J. B., THURMAN, R. E., SANDSTROM, R., HAUGEN, E., HEIMFELD, S., MURRY, C. E., AKEY, J. M. & STAMATOYANNOPOULOS, J. A. 2013. Developmental fate and cellular maturity encoded in human regulatory DNA landscapes. *Cell,* 154**,** 888-903.

STORM, R., CHOLEWA-WACLAW, J., REUTER, K., BROHL, D., SIEBER, M., TREIER, M., MULLER, T. & BIRCHMEIER, C. 2009. The bHLH transcription

factor Olig3 marks the dorsal neuroepithelium of the hindbrain and is essential for the development of brainstem nuclei. *Development,* 136**,** 295-305.

STOVOLD, R., BLACKHALL, F., MEREDITH, S., HOU, J., DIVE, C. & WHITE, A. 2012. Biomarkers for small cell lung cancer: neuroendocrine, epithelial and circulating tumour cells. *Lung Cancer,* 76**,** 263-8.

SUGIMORI, M., NAGAO, M., PARRAS, C. M., NAKATANI, H., LEBEL, M., GUILLEMOT, F. & NAKAFUKU, M. 2008. Ascl1 is required for oligodendrocyte development in the spinal cord. *Development,* 135**,** 1271-81.

SUN, W., HU, X., LIM, M. H., NG, C. K., CHOO, S. H., CASTRO, D. S., DRECHSEL, D., GUILLEMOT, F., KOLATKAR, P. R., JAUCH, R. & PRABHAKAR, S. 2013. TherMos: Estimating protein-DNA binding energies from in vivo binding profiles. *Nucleic Acids Res,* 41**,** 5555-68.

SWAROOP, A., KIM, D. & FORREST, D. 2010. Transcriptional regulation of photoreceptor development and homeostasis in the mammalian retina. *Nat Rev Neurosci,* 11**,** 563-76.

TAVARES, I. & LIMA, D. 2007. From neuroanatomy to gene therapy: searching for new ways to manipulate the supraspinal endogenous pain modulatory system. *J Anat,* 211**,** 261-8.

THOMPSON, N., GESINA, E., SCHEINERT, P., BUCHER, P. & GRAPIN-BOTTON, A. 2012. RNA profiling and chromatin immunoprecipitation-sequencing reveal that PTF1a stabilizes pancreas progenitor identity via the control of MNX1/HLXB9 and a network of other transcription factors. *Mol Cell Biol,* 32**,** 1189-99.

TRAPNELL, C., HENDRICKSON, D. G., SAUVAGEAU, M., GOFF, L., RINN, J. L. & PACHTER, L. 2013. Differential analysis of gene regulation at transcript resolution with RNA-seq. *Nat Biotechnol,* 31**,** 46-53.

TRAPNELL, C., PACHTER, L. & SALZBERG, S. L. 2009. TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics,* 25**,** 1105-11.

TRAPNELL, C., ROBERTS, A., GOFF, L., PERTEA, G., KIM, D., KELLEY, D. R., PIMENTEL, H., SALZBERG, S. L., RINN, J. L. & PACHTER, L. 2012. Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nat Protoc,* 7**,** 562-78.

TRAPNELL, C., WILLIAMS, B. A., PERTEA, G., MORTAZAVI, A., KWAN, G., VAN BAREN, M. J., SALZBERG, S. L., WOLD, B. J. & PACHTER, L. 2010. Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. *Nat Biotechnol,* 28**,** 511-5.

TSE, C., SHOEMAKER, A. R., ADICKES, J., ANDERSON, M. G., CHEN, J., JIN, S., JOHNSON, E. F., MARSH, K. C., MITTEN, M. J., NIMMER, P., ROBERTS, L., TAHIR, S. K., XIAO, Y., YANG, X., ZHANG, H., FESIK, S., ROSENBERG, S. H. & ELMORE, S. W. 2008. ABT-263: a potent and orally bioavailable Bcl-2 family inhibitor. *Cancer Res,* 68**,** 3421-8.

TSUNEKAWA, Y., BRITTO, J. M., TAKAHASHI, M., POLLEUX, F., TAN, S. S. & OSUMI, N. 2012. Cyclin D2 in the basal process of neural progenitors is linked to non-equivalent cell fates. *EMBO J,* 31**,** 1879-92.

VAN BEEST, M., DOOIJES, D., VAN DE WETERING, M., KJAERULFF, S., BONVIN, A., NIELSEN, O. & CLEVERS, H. 2000. Sequence-specific high mobility group box factors recognize 10-12-base pair minor groove motifs. *J Biol Chem,* 275**,** 27266-73.

VISEL, A., BLOW, M. J., LI, Z., ZHANG, T., AKIYAMA, J. A., HOLT, A., PLAJZER-FRICK, I., SHOUKRY, M., WRIGHT, C., CHEN, F., AFZAL, V., REN, B., RUBIN, E. M. & PENNACCHIO, L. A. 2009a. ChIP-seq accurately predicts tissue-specific activity of enhancers. *Nature,* 457**,** 854-8.

VISEL, A., MINOVITSKY, S., DUBCHAK, I. & PENNACCHIO, L. A. 2007. VISTA Enhancer Browser--a database of tissue-specific human enhancers. *Nucleic Acids Res,* 35**,** D88-92.

VISEL, A., RUBIN, E. M. & PENNACCHIO, L. A. 2009b. Genomic views of distant-acting enhancers. *Nature,* 461**,** 199-205.

VOLKER, L. A., PETRY, M., ABDELSABOUR-KHALAF, M., SCHWEIZER, H., YUSUF, F., BUSCH, T., SCHERMER, B., BENZING, T., BRAND-SABERI, B., KRETZ, O., HOHNE, M. & KISPERT, A. 2012. Comparative analysis of Neph gene expression in mouse and chicken development. *Histochem Cell Biol,* 137**,** 355-66.

WANG, J., DUNCAN, D., SHI, Z. & ZHANG, B. 2013. WEB-based GEne SeT AnaLysis Toolkit (WebGestalt): update 2013. *Nucleic Acids Res,* 41**,** W77-83.

WESTERMAN, B. A., BREUER, R. H., POUTSMA, A., CHHATTA, A., NOORDUYN, L. A., KOOLEN, M. G., POSTMUS, P. E., BLANKENSTEIN, M. A. & OUDEJANS, C. B. 2007. Basic helix-loop-helix transcription factor profiling of lung tumors shows aberrant expression of the proneural gene atonal homolog 1 (ATOH1, HATH1, MATH1) in neuroendocrine tumors. *Int J Biol Markers,* 22**,** 114-23.

WIEBE, P. O., KORMISH, J. D., ROPER, V. T., FUJITANI, Y., ALSTON, N. I., ZARET, K. S., WRIGHT, C. V., STEIN, R. W. & GANNON, M. 2007. Ptf1a binds to and activates area III, a highly conserved region of the Pdx1 promoter that mediates early pancreas-wide Pdx1 expression. *Mol Cell Biol,* 27**,** 4093-104.

WILDNER, H., MULLER, T., CHO, S. H., BROHL, D., CEPKO, C. L., GUILLEMOT, F. & BIRCHMEIER, C. 2006. dILA neurons in the dorsal spinal cord are the product of terminal and non-terminal asymmetric progenitor cell divisions, and require Mash1 for their development. *Development,* 133**,** 2105-13.

XU, C. R., COLE, P. A., MEYERS, D. J., KORMISH, J., DENT, S. & ZARET, K. S. 2011. Chromatin "prepattern" and histone modifiers in a fate choice for liver and pancreas. *Science,* 332**,** 963-6.

XUAN, S., BOROK, M. J., DECKER, K. J., BATTLE, M. A., DUNCAN, S. A., HALE, M. A., MACDONALD, R. J. & SUSSEL, L. 2012. Pancreas-specific deletion of mouse Gata4 and Gata6 causes pancreatic agenesis. *J Clin Invest,* 122**,** 3516-28.

YAZAWA, T., SATO, H., SHIMOYAMADA, H., OKUDELA, K., WOO, T., TAJIRI, M., OGURA, T., OGAWA, N., SUZUKI, T., MITSUI, H., ISHII, J., MIYATA, C., SAKAEDA, M., GOTO, K., KASHIWAGI, K., MASUDA, M., TAKAHASHI, T. & KITAMURA, H. 2009. Neuroendocrine cancer-specific up-regulating mechanism of insulin-like growth factor binding protein-2 in small cell lung cancer. *Am J Pathol,* 175**,** 976-87.

ZARET, K. S. & CARROLL, J. S. 2011. Pioneer transcription factors: establishing competence for gene expression. *Genes Dev,* 25**,** 2227-41.

ZHU, X., LI, Y., SHEN, H., LI, H., LONG, L., HUI, L. & XU, W. 2013. miR-137 inhibits the proliferation of lung cancer cells by targeting Cdc42 and Cdk6. *FEBS Lett,* 587**,** 73-81.

ZOU, M., LI, S., KLEIN, W. H. & XIANG, M. 2012. Brn3a/Pou4f1 regulates dorsal root ganglion sensory neuron specification and axonal projection into the spinal cord. *Dev Biol,* 364**,** 114-27.