From DEPARTMENT OF CELL AND MOLECULAR BIOLOGY
Karolinska Institutet, Stockholm, Sweden

SOX TRANSCRIPTION FACTORS: MULTIFACETED REGULATORS OF CENTRAL NERVOUS SYSTEM DEVELOPMENT

Cécile Zaouter

Karolinska Institutet

Stockholm 2017

On the cover: Sculpture of a human brain made of laboratory gloves created by Cécile Zaouter and Gaëlle Combeau.

Published by Karolinska Institutet.
Printed by E-Print AB 2017

ISBN 978-91-7676-568-5

Sox transcription factors: multifaceted regulators of central nervous system development

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Cécile Zaouter

Principal Supervisor:
Jonas Muhr
Karolinska Institutet
Department of Cell and Molecular Biology

Co-supervisor(s):
Ola Hermanson
Karolinska Institutet
Department of Neuroscience

Maria Bergsland
Karolinska Institutet
Department of Cell and Molecular Biology

Opponent:
Aixa V Morales
Instituto Cajal
Department of Cellular and Developmental Neurobiology

Examination Board:
Mattias Mannervik
Stockholm Universitet
Department of Molecular Bioscience

Eva Hedlund
Karolinska Institutet
Department of Neuroscience

Anna Falk
Karolinska Institutet
Department of Neuroscience

À ma Mamie

« Patience et longueur de temps
Font plus que force ni que rage. »

– Jean de La Fontaine, *Le lion et le rat* (1668)

ABSTRACT

The central nervous system (CNS) is composed of three major cell types, namely neurons, astrocytes and oligodendrocytes that are mainly generated during embryonic stages by multipotent neural progenitor cells (NPCs). The decision of a NPC to remain in a self-renewing, undifferentiated state, or to commit to neuronal or glial differentiation is guided by the combined actions of transcription factors on their downstream target genes. SOX transcription factors have been shown to have key roles during neural lineage formation. In this thesis, we used several next-generation sequencing approaches to study the molecular mechanisms by which SOX transcription factors regulate the maintenance and differentiation of NPCs during early neuronal and glial lineage development.

In **Paper I**, we investigated how SOX2, SOX3 and SOX11 proteins achieve their distinct functions during neuronal lineage development by characterizing their genome-wide binding profiles in embryonic stem cells (ESCs), NPCs and neurons. We propose a model of sequentially acting SOX proteins during neural lineage formation, whereby SOX proteins pre-bind large sets of poised silent genes in ESCs and NPCs that are subsequently activated by alternative SOX proteins at later stages of neuronal differentiation.

In **Paper II**, we examined how chromatin accessibility and transcription factor binding interact to regulate the establishment of different gene expression profiles in NPCs originating from the developing mouse cortex and spinal cord. We found that despite being ubiquitously expressed in all NPCs of the developing CNS, SOX2 regulates the establishment of spatially distinct gene expression programs by interacting with region-specific partner factors on a permissive chromatin landscape.

In **Paper III**, we characterized the genome-wide binding profile of SOX3 in SOX9 in glial progenitor cells in order to determine their regulatory roles during the development of astrocytes and oligodendrocytes. We show that glial gene expression, similar to neuronal gene expression, is regulated by sequentially acting pre-binding SOX proteins.

Altogether, the results presented in this thesis provide new molecular insights into the mechanisms by which SOX proteins achieve their distinct functional roles during neuronal and glial lineage development.

LIST OF SCIENTIFIC PAPERS

This thesis is based on the following papers, which will be referred to in the text by their roman numbers.

I. Maria Bergsland*, Daniel Ramsköld*, **Cécile Zaouter**, Susanne Klum, Rickard Sandberg and Jonas Muhr (2011).
Sequentially acting Sox transcription factors in neural lineage development.
Genes & Development 25, 2453-64.

II. Daniel W. Hagey*, **Cécile Zaouter***, Gaëlle Combeau, Monika Andersson Lendahl, Olov Andersson, Mikael Huss and Jonas Muhr (2016).
Distinct transcription factor complexes act on a permissive chromatin landscape to establish regionalized gene expression in CNS stem cells.
Genome Research 26, 908-17.

III. **Cécile Zaouter***, Susanne Klum*, Zjanna Alekseenko, Åsa K. Bjöklund, Daniel W. Hagey, Johan Ericson, Jonas Muhr* and Maria Bergsland*.
scRNA-seq and ChIP-seq analyses in glial progenitors reveal extensive SOX9 pre-binding to astrocyte and oligodendrocyte specific genes.
Manuscript.

*These authors contributed equally to this work.

CONTENTS

LIST OF ABBREVIATIONS

A/P	Antero-posterior
bHLH	Basic helix-loop-helix
BMP	Bone morphogenetic protein
ChIP	Chromatin immunoprecipitation
CNS	Central nervous system
CX	Cortex
DNA	Deoxyribonucleic acid
ESC	Embryonic stem cell
hESC	Human embryonic stem cell
mESC	Mouse embryonic stem cell
FACS	Fluorescent-activated cell sorting
GPC	Glial progenitor cell
HMG	High-mobility group
ICM	Inner cell mass
iPSC	Induced pluripotent stem cell
kb	Kilo base
MBP	Myelin basic protein
NFI	Nuclear factor I
NPC	Neural progenitor cell
OPC	Oligodendrocyte progenitor cell
PCR	Polymerase chain reaction
pPCR	Quantitative PCR
PLP	Proteolipid protein
RNA	Ribonucleic acid
scRNA	Single cell RNA
siRNA	Small interfering RNA
SC	Spinal cord
SHH	Sonic hedgehog
SOX	Sry-related HMG-box
TF	Transcription factor
SRY	Sex-determining region Y
TE	Trophoectoderm
TSS	Transcription start site
VZ	Ventricular zone

1 INTRODUCTION

The central nervous system (CNS) gathers and processes information from the entire body, and consists of the brain and the spinal cord. The brain is a remarkable organ that serves as the control center of the body. Indeed, as already described by Hippocrates in the 4th century B.C, "From the brain alone, arise our pleasures, joys, laughter and jests, as well as our sorrows, pains, griefs and tears. Through it, in particular, we think, see, hear and distinguish the ugly from the beautiful, the bad from the good, the pleasant from the unpleasant" (Pandya 2011). The spinal cord, on the other hand, carries motor information from the brain to the diverse muscles and glands of the body, while also relaying sensory information from the peripheral nervous system to the brain.

The mature CNS contains about 100 billion neurons that mediate the transmission of information. Despite their essential function, neurons are outnumbered by glial cells that constitute approximately 90% of the cells of CNS. Glial cells, which include oligodendrocytes and astrocytes, also perform key functions that are vital for the normal functioning of the CNS. Oligodendrocytes form the myelin sheaths that insulate axons and enhance signal transduction, whereas astrocytes provide structural support for neurons, participate in the formation of the blood-brain barrier, and regulate synaptic transmission (Rowitch 2004; Kessaris et al. 2008).

The generation of the myriad of differentiated cells that compose the mature CNS mainly occurs during embryonic development, and starts with the decision of pluripotent embryonic stems cells (ESCs) to become multipotent neural progenitor cells (NPCs) that will in turn differentiate into neurons or glial cells. The proper control of gene expression is essential to the process of cell differentiation, and largely relies on the activity of transcription factors. Members of the SOX transcription factor family have been identified to be key regulators of neural lineage formation, and dysregulation of their activities has been implicated in numerous diseases, including developmental disorders and cancer.

In this thesis, we took advantage of the recent advances in next-generation sequencing technologies to better understand the molecular mechanisms by which SOX proteins regulate the maintenance and differentiation of NPCs during the formation of the CNS. Before discussing our findings, I will highlight the importance of the SOX protein family in the transcriptional regulation of gene expression. I will also describe their functions during early embryogenesis and CNS development, with an emphasis on SOXB, SOXC and SOXE proteins.

2 THE SOX FAMILY

2.1 DISCOVERY OF SOX GENES

In 1990, the *Sry* gene (Sex-determining region Y) was discovered in both mouse and human as being responsible for male sex determination (Gubbay et al. 1990; Sinclair et al. 1990). Shortly after this discovery, a number of genes sharing sequence similarities with *Sry* were identified. *Sry* was therefore the first-identified member of a large gene family that was later named the *Sox* gene family (Sry-related HMG-box) (Gubbay et al. 1990; Denny et al. 1992; Wright et al. 1993).

2.2 CLASSIFICATION OF SOX PROTEINS

The hallmark of the SOX protein family is the presence of a high-mobility group (HMG) box that mediates DNA binding and thereby allows them to act as transcription factors. Members of the SOX family exhibit approximately 50% or higher amino acid sequence similarity to the HMG box of the founding member of the family, SRY (Bowles et al. 2000).

SOX proteins are present throughout the animal kingdom, in both invertebrate and vertebrate organisms, from *Trichoplax adhaerens* to human. 20 different SOX proteins have been identified in both mouse and human, 27 in zebrafish (*Danio rerio*), 8 in *Drosophila* and 5 in *Caenorhabditis elegans* (Bowles et al. 2000).

SOX proteins have been classified into clearly defined groups, termed A to H in mammals, based on the phylogenetic analysis of their HMG domains. The SoxB group was further divided into two subgroups, SoxB1 and SoxB2, according to their full-length sequences and functional characteristics (Bowles et al. 2000) (**Table 1**). Members of a same group share at least 80% sequence identity within their HMG domain (Wright et al. 1993; Wegner 1999). Moreover, the comparison of full-length proteins further revealed that sequence similarities extend outside the HMG domain within each group, but not between groups (Wegner 1999; Bowles et al. 2000).

Table 1. Sox groups, their members and their respective transcriptional regulation domain in Mus musculus (Kiefer 2007; Guth & Wegner 2008).

Group	Protein	Transregulation domain
SoxA	SRY	activation
SoxB1	SOX1, SOX2, SOX3	activation
SoxB2	SOX 14, SOX21	repression
SoxC	SOX4, SOX11, SOX12	activation
SoxD	SOX5, SOX6, SOX13	none
SoxE	SOX 8, SOX9, SOX10	activation
SoxF	SOX7, SOX17, SOX18	activation
SoxG	SOX15	activation
SoxH	SOX30	activation

In addition to amino acid sequence similarities, members within a group share similar biochemical properties and have consequently been described to have similar biological functions (Bowles et al. 2000; Wegner 2010). For instance, the members of the SoxB1 group, SOX1-3, have been shown to exert almost identical biological activities in neural stem cells

during embryonic development (Wood & Episkopou 1999; Bylund et al. 2003; Graham et al. 2003). In contrast, members from different groups exhibit distinct functions.

Furthermore, members of a same group have similar genomic organizations, with members of the SoxA, SoxB and SoxC groups containing a single exon (Bowles et al. 2000). However, *Sox* genes are localized randomly throughout the mammalian genome and are not grouped in gene clusters (Wegner 1999).

2.3 EVOLUTIONARY CONSERVATION OF SOX PROTEINS

Members of the SOX family are highly conserved across species, especially among vertebrates where orthologous proteins show a high degree of conservation both within and outside their HMG domain. An example is human SOX2 that shares 98% similarity with the mouse protein, and 88% with its putative orthologue Dichaete in *Drosophila* (Bowles et al. 2000; Pevny & Lovell-Badge 1997). SRY is however an exception to the rest of the SOX family as it is poorly conserved outside its HMG domain, even across closely related species, and is restricted to mammals (Soullier et al. 1999). Additionally, amino acid sequence conservation correlates with conservation of biochemical properties. Ectopic expression of the mouse SOX2 in transgenic flies has been shown to rescue the CNS phenotype of Dichaete mutants suggesting a functional conservation during evolution (Soriano & Russell 1998).

2.4 SOX PROTEIN STRUCTURE

SOX proteins belong to the HMG box-containing protein superfamily that can be divided into two subfamilies based on the number of HMG domains they contain. Members of SOX/TCF/MATA subfamily have a single HMG domain that binds the DNA in a sequence specific manner whereas members of the HMG/UBF subfamily possess multiple HMG domains and bind DNA non-specifically (Soullier et al. 1999).

The HMG domain of SOX proteins contains 79 amino acid residues and consists of three α-helices and a N-terminal β strand that are arranged in a twisted L-shape (Weiss 2001; Pevny & Lovell-Badge 1997). This domain carries out several important functions. It allows for precise DNA recognition and binding, bends the DNA, mediates interactions with partner proteins, and contains nuclear import or export signals (Wilson & Koopman 2002; Lefebvre et al. 2007).

SOX proteins feature other functional domains. The sequences of most SOX proteins are comprised of three main domains: a N-terminal region followed by the HMG domain and a C-terminal region (Kamachi & Kondoh 2013; Kamachi et al. 2000). Sequence truncations experiments and transactivation assays have shown that the majority of the SOX proteins possess a transcriptional regulation domain located in their C-terminal region (Kiefer 2007). For most SOX proteins it corresponds to a transactivation domain (Kamachi et al. 1995; Südbeck et al. 1996; Kamachi et al. 1998), with the exceptions of SOXB2 proteins, which carry a transrepression domain (Uchikawa et al. 1999; Bylund et al. 2003), and of SOXD proteins, which reportedly lack transregulation domain (Roose et al. 1998) (**Table 1**).

Some SOX proteins additionally contain dimerization domains. Members of the SoxD group bear a coiled-coil domain that enables dimerization with members of this group, and SOXE proteins have a self-dimerization domain (Kamachi & Kondoh 2013).

3 TRANSCRIPTIONAL REGULATION OF GENE EXPRESSION BY SOX PROTEINS

Multicellular organisms contain a wide range of different cell types with distinct functions. This great cell diversity is achieved by the establishment of a unique gene expression program in each cell type. Gene expression is defined as the complex process "through which a cell converts the nucleotide sequence of a gene first into the nucleotide sequence of an RNA molecule, and then into the amino acid sequence of a protein". The regulation of gene expression can occur at any step of this process, including transcription, post-transcription (RNA splicing), translation, and post-translation (modification of a protein), with transcription being the primary mode of regulation (Alberts et al. 2008).

The transcriptional regulation of gene expression is the result of two main mechanisms. The first one involves the combined actions of transcription factors (TFs) and their interacting factors on their downstream target gene regulatory regions. The second involves the degree of DNA compaction into chromatin, which determines whether a gene regulatory region will be accessible or not to TFs and other proteins (Alberts et al. 2008).

SOX proteins have been suggested to display properties of both classical TFs and architectural components of chromatin due to their ability to bend the DNA upon binding (Pevny & Lovell-Badge 1997). In this chapter, I will describe how SOX proteins bind DNA and act as transcriptional regulators of gene expression, on one hand by interacting with partner proteins in order to modulate their target gene selection and their activity, and on the other hand by influencing the chromatin structure. I will mainly focus on the best characterized SOX protein, SOX2, during the development of the CNS.

3.1 DNA BINDING OF SOX PROTEINS

Transcription factors are regulatory proteins whose function is to regulate gene expression by controlling the transcription of a given gene into a molecule of messenger RNA (mRNA). The hallmark of TFs is their DNA binding domain that mediates binding to specific DNA sequences within gene regulatory regions, such as promoters and enhancers, in order to activate or repress the expression of a gene (Alberts et al. 2008).

As mentioned previously, SOX proteins contain a HMG domain that mediates binding to the DNA, and thus they function as TFs. The optimal DNA-binding motif for SOX TFs was identified *in vitro* using random oligonucleotide selection assays and was defined as 5´-(A/T)(A/T)CAA(T/A)G -3´ (Harley et al. 1994). More recently, experiments using chromatin-immunoprecipitation followed by deep sequencing (ChIP-seq) for different SOX TFs have further confirmed that this sequence corresponds to the SOX consensus binding motif (Chen et al. 2008; Scharer et al. 2009; Hagey & Muhr 2014). In addition, SOX TFs have individual preferences for nucleotides flanking this consensus sequence (Mertin et al. 1999). These preferences may to some extent explain how SOX TFs achieve their target gene specificity despite binding to the same consensus site.

However, even though all SOX TFs are able to bind a common consensus motif *in vitro*, analysis of SOX bound enhancers *in vivo* revealed that SOX TFs also bind sequences that only partially match the consensus site. During chondrocyte development, SOX5, SOX6 and SOX9

were found to bind the SOX consensus motif with one or two mismatches in the examined *Col2a1* enhancer (Lefebvre et al. 1997; Lefebvre et al. 1998). Therefore, SOX TFs bind DNA with little sequence specificity, indicating that other mechanisms may be involved in directing SOX TFs to their target genes *in vivo*. Moreover, it is noteworthy that SOX TFs bind DNA with relatively low affinity compared to other TFs, suggesting that their binding alone might not be stable enough to induce transcriptional regulation *in vivo* (Kamachi et al. 2000; Lefebvre et al. 2007).

Another feature of SOX proteins that may have important implications for gene regulation is their ability to bend the DNA. While the overall structure of SOX proteins remains unaltered upon DNA binding, SOX proteins induce a sharp bend to the DNA they bind to. This bending is established by the interaction of the HMG domain with the minor groove, which results in the widening of the minor groove and causes the DNA to bend towards the major groove with an approximate 80° angle. Thus, unlike most TFs, which bind to the major groove of the DNA, SOX proteins are unique in that they interact with the minor groove (Ferrari et al. 1992; Werner et al. 1995).

SOX proteins induce very similar DNA bending angles suggesting that there is no correlation between the bending angle and the regulatory potential of distinct SOX proteins (Palasingam et al. 2009). Nevertheless, due to their unusual intrinsic bending capacity, it has been speculated that SOX proteins may act as architectural proteins by promoting the interaction of distantly bound proteins and thereby forming a transcriptional complex (Lefebvre et al. 2007; Pevny & Lovell-Badge 1997). Such regulatory function has been described for the HMG protein LEF1 but remains to be fully uncovered for SOX proteins (Giese et al. 1992). Evidence supporting the importance of the DNA bending ability of SOX proteins come from studies showing that SOX mutants that are able to bind but not to bend the DNA result in mutant phenotypes *in vivo*. An example is a *de novo* mutation that reduces the bending activity of SRY and leads to sex-reversal in humans, pointing to the possible functional role of SOX DNA bending (Pontiggia et al. 1994; Scaffidi & Bianchi 2001).

3.2 SOX PROTEIN ACTIVITY DEPENDS ON THEIR PARTNERS

SOX TFs bind a similar DNA binding motif with low specificity and low affinity. Moreover, although most SOX TFs contain a transregulation domain, they have been shown to exert weak transcriptional activities on their own. As a result, a SOX protein alone might not to be able to specifically select its target genes, and its binding might not lead to transcriptional activation or repression of gene regulatory regions *in vivo*. Indeed, SOX TFs have been shown to critically depend on their interaction with other proteins in order to be able to exert their gene regulatory functions (Kamachi et al. 2000; Lefebvre et al. 2007; Kamachi & Kondoh 2013).

3.2.1 Partnership with transcription factors

Studies performing motif analysis in functional enhancers of known SOX target genes showed that SOX binding motifs are often in close proximity of other TF motifs. The first described example is the *Fgf4* enhancer that contains adjacent binding sites for SOX and POU domain proteins. SOX2 and OCT4 (also called POU5F1) bind and activate this enhancer in embryonic carcinoma cells and embryonic stem cells (ESCs) (Yuan et al. 1995; Ambrosetti et al. 1997). Another well characterized example is the DC5 enhancer of the δ-crystalline gene, which

features SOX and Paired domain binding sites, and is activated by SOX2 and PAX6 in lens cells (Kamachi et al. 1995; Kamachi et al. 2001). In both cases, the SOX protein and its partner factor bind cooperatively and in turn activate the enhancer. Mutating either binding site or increasing the spacing between the two sites diminishes their cooperative binding and results in the complete loss of enhancer activation. Therefore, the functional cooperation of SOX proteins with their partners is necessary to stabilize their binding and to elicit the synergistic activation of regulatory regions *in vivo*.

Furthermore, each SOX protein appears to recognize and regulate distinct sets of target genes in different cell types. The interaction of SOX proteins with cell type-specific partner factors may thus provide a mechanism of target gene selection *in vivo* (Kamachi et al. 2000; Lefebvre et al. 2007). This is illustrated by the finding that SOX2 acts cooperatively with OCT4 to activate the *Fgf4* enhancer, but is unable to do so with the related OCT1 protein. This specificity offers some insight into how selective gene activation by SOX proteins is achieved (Yuan et al. 1995). Additionally, the status of the chromatin, and thus the accessibility to the DNA, may also have an important regulatory role in defining SOX target gene selection in different cell types, as described later in this chapter.

Since these pioneer discoveries, SOX proteins have been found to interact with a variety of proteins belonging to many TF families (Kondoh & Kamachi 2010) (**Paper II**). Recently, a proteomics analysis was performed in order to identify other putative SOX2 interacting proteins in neural progenitor cells (NPCs). In this screen, a total of 50 proteins were found to physically interact with SOX2, many of which are known to be involved in neural development. Among the 50 proteins constituting the SOX2 interactome, 19 are TFs belonging to various families, including the SOX family itself. This confirms that SOX proteins cooperate with a multitude of TFs to regulate gene expression and refine their target gene selection (Engelen et al. 2011).

Interestingly, studies have shown that SOX-partner interactions are often mediated by the DNA binding HGM domain itself. These interactions can also occur via other domains, such as the coiled-coil domain or the C-terminal end of the protein (Wilson & Koopman 2002).

3.2.2 Partnership with co-regulators

In addition to TFs, several members of co-activator (TRRAP) and co-repressor (SMRT/NcoR) complexes were also identified to be strong partners of SOX2 in NPCs (Engelen et al. 2011). Moreover, SOX2 has been suggested to regulate the expression of pluripotency genes in ESCs by recruiting the co-activator p300, whose presence is known to predict enhancer activity (Chen et al. 2008). SOX2 has also been shown to interact with the repressor Groucho/TLE in NPCs to repress the *Ccnd1* promoter (Hagey & Muhr 2014). Hence, SOX TFs can modulate their transcriptional activity by interacting with either co-activators or co-repressors. This is supported by several genome-wide studies where downregulation of SOX2 followed by gene expression analysis have demonstrated the dual role of SOX2, which can function as both an activator and a repressor of downstream target genes (Engelen et al. 2011; Bergsland et al. 2011; Hagey & Muhr 2014). This is in contrast with the traditional way of characterizing TFs as being either activators or repressors, based on the transregulation domain they carry.

In summary, SOX TFs work in a combinatorial fashion and their transcriptional activity depends on their interaction with diverse types of partners.

3.3 SOX PROTEINS AND CHROMATIN

3.3.1 Chromatin structure

As mentioned earlier, the modulation of the chromatin structure is another important layer of transcriptional regulation of gene expression. Indeed, even when TFs and their partners are expressed at the right time and place, the state of chromatin might limit the access to their target sequences and thereby impair their transcriptional activity. In order to discuss the possible role of SOX TFs in altering chromatin structure, it is necessary to first describe the structure of chromatin and how it is regulated.

The chromatin is a structure that allows the efficient packaging of the entire genome into the small cell nucleus. This condensation is achieved by the wrapping of approximately 146 bp of DNA around an octamer of core histone proteins made of two copies of H2A, H2B, H3 and H4 proteins to create a nucleosome. Arrays of nucleosomes are in turn compacted to form chromatin fibers (Kornberg 1977; Venkatesh & Workman 2015). Chromatin exists in two main forms: euchromatin, which is loose and permissive for transcription, and heterochromatin, which is more condensed and transcriptionally silent (Benayoun et al. 2015). Based on its sensitivity to DNase I cleavage, chromatin can further be characterized into regions of "open" and "closed" chromatin, which correspond to transcriptionally active and silent gene regulatory regions, respectively (Gilbert & Ramsahoye 2005). Regions of open or accessible chromatin are thus referred to as DNase I hypersensitivity sites (DHSs) and can be mapped at the genome-wide level using DNase-sequencing (Natarajan et al. 2012).

3.3.2 Chromatin remodeling

Chromatin remodeling corresponds to changes in chromatin compaction that affect the accessibility of TFs to their target gene regulatory regions, and that can subsequently lead to the altering of gene expression (Venkatesh & Workman 2015). It is a dynamic process that plays an essential role during lineage specification and differentiation. Indeed, while the chromatin generally becomes more condensed as development proceeds, a substantial number of open chromatin regions are also formed *de novo* (Stergachis et al. 2013; Lara-Astiaso et al. 2014; Raposo et al. 2015). Accumulating evidence have shown that SOX TFs are engaged in some of the mechanisms involved in chromatin remodeling, which will be discussed below.

ATP-dependent chromatin remodeling enzyme complexes can rearrange nucleosome to create regions of accessible DNA for TF binding, which in turn affects gene expression (Wang et al. 2007; Benayoun et al. 2015). In an aforementioned study (Engelen et al. 2011), several proteins of the SOX2 interactome in NPCs were found to belong to chromatin remodeling complexes, including the SWI/SNF and NuRD/CHD complexes. In particular, SOX2 interacts very strongly with the major chromatin remodeling factor CHD7. This discovery suggests that SOX2 may have a role in chromatin remodeling by altering the local chromatin structure through the recruitment of components of chromatin remodeling complexes.

The post-translational modification of histone protein tails can also adjust chromatin accessibility and thus lead to transcriptional activation or repression. There are at least eight different types of covalent modifications including methylation, acetylation and phosphorylation. Some of the best studied histone modifications correspond to the modifications of lysine residues of histone H3, including H3K4me3, which marks active gene regulatory regions and H3K27me3, which is a feature of repressed regulatory regions. Thanks to the development of ChIP-sequencing technologies, the global distributions of many different chromatin modifications have been successful mapped (Kouzarides 2007; Benayoun et al. 2015).

SOX2 ChIP-seq experiments in ESCs have revealed that SOX2 binds to regulatory regions of active genes involved in pluripotency and self-renewal, and of silent genes that are not expressed in ESCs and will be expressed at later stages of development. Interestingly, the gene regulatory regions associated to active genes carry the H3K4me3 mark, whereas those associated to silent genes that will be expressed later carry the bivalent mark H3K4me3/H3K27me3 instead. This bivalent state is believed to keep genes in a permissive chromatin state for later activation (Boyer et al. 2006). Strikingly, we have described a similar situation where SOX2 in NPCs binds H3K4me3 marked regulatory regions of transcriptionally active genes, as well as pre-binds bivalently marked regulatory regions of silent genes involved in later stages of neural development (**Paper I**). It is unclear whether SOX proteins are directly involved in regulating the deposition of histone modification marks. Nevertheless, a study showed that SOX2 binding contributes to the formation of activating H3K4me2 and H3K4me3 marks at the λ5-VpreB1 enhancer in ESCs (Liber et al. 2010). In addition, we have shown that the ectopic expression of SOX3 in C2C12 mesodermal progenitor cells is followed by accumulation of the H3K4me3 mark at SOX3 bound enhancers (**Paper I**). These two examples suggest that SOX proteins may be able to induce local epigenetic changes at bound enhancer regions. This is consistent with the fact that SOX2 is a crucial factor in cell reprogramming, as chromatin remodeling is essential to this process (Takahashi & Yamanaka 2006; Nakagawa et al. 2008).

3.3.3 Pioneer factors

Pioneer factors are defined as a specific functional class of TFs that have the ability to bind regions of closed chromatin and remodel the chromatin landscape locally to increase target site accessibility and thus facilitate the binding of additional TFs. This is in strong contrast to other types of TFs that cannot directly bind closed chromatin individually and rely on cooperative interactions to access their target sites (Cirillo et al. 2002; Zaret & Carroll 2011).

In line with this definition, the binding of pioneer factors to chromatin precedes the binding of other TFs, and often occurs prior to enhancer activation and gene expression. As an example, FOXD3 binds the enhancer of the silent *Alb1* liver-specific gene in ESCs and thereby ensures the establishment of a permissive chromatin state for subsequent activation by other TFs. Indeed, as cells become committed to the endodermal lineage, FOXD3 is replaced by FOXA1, which will help activate *Alb1* gene expression (Xu et al. 2009). This pre-binding by pioneer factors is believed to prime silent enhancers for later activation when cells differentiate, as a means to confer rapid transcriptional activation of genes in response to inductive signals.

Subsequent to binding, pioneer factors recruit additional TFs that bind to adjacent binding sites and in turn contribute to the modulation of gene expression. To enable their binding, pioneer factors need to remodel the chromatin landscape locally to increase target site accessibility. One important feature of pioneer factors is their capacity to open closed chromatin. To do so, pioneer factors can evidently recruit chromatin remodelers to rearrange nucleosomes or actively contribute to the establishment of local epigenetic changes at enhancers (Magnani et al. 2011). Indeed, in addition to inducing local nucleosome remodeling, the pioneer factor PU.1 promotes the deposition of the active H3K4me1 mark at enhancers during B cell lineage differentiation (Heinz et al. 2010), and FOXA1 promotes of the deposition of the H3K4me2 mark during neuronal differentiation (Sérandour et al. 2011).

There is growing body of evidence supporting the idea that SOX2 may function as a pioneer factor. Indeed, as mentioned above, SOX2 binding in ESCs and NPCs precedes gene activation. It binds enhancers of silent genes that are in a bivalent state, and that will be activated by other TFs at a subsequent stage of the differentiation process (Boyer et al. 2006) (**Paper I**). Moreover, SOX2 has been shown to interact with components of multiple chromatin remodeling complexes, and appears to be involved in the establishment of chromatin modifications at bound enhancers (Liber et al. 2010; Engelen et al. 2011).

In addition, a study using a single-molecule imaging strategy in order to track the binding dynamics of SOX2 and OCT4 in ESCs indicates that SOX2 binds first, and subsequently recruits OCT4 to SOX/OCT binding sites. However, this assay did not investigate the influence of the chromatin state on this hierarchical binding (Chen et al. 2014). Furthermore, another study found that during the reprogramming of fibroblasts, SOX2, OCT4 and KLF4 can on their own target regions of closed chromatin that are insensitive to DNase I cleavage and lack active chromatin modification marks such as H3K4me3 and H3K27ac. Additionally, c-MYC was found to also be able to bind closed chromatin, but only when expressed in the presence of the three other factors (Soufi et al. 2012).

In conclusion, SOX TFs primarily function as classical TFs in order to regulate gene expression. With the help of their partner TFs, they bind and regulate distinct sets of genes in different cell types. They also recruit diverse types of interacting partners in order to modulate gene activity. SOX TFs also have alternative roles of transcriptional regulation, amongst which are chromatin remodeling and epigenetic functions. Finally SOX2 appears to have some features of pioneer factors in ESCs and NPCs. Hence, SOX proteins are versatile transcriptional regulators that employ different modes of action to fulfill their diverse functions.

4 SOX PROTEIN FUNCTION DURING EARLY EMBRYOGENESIS AND PLURIPOTENCY

In this chapter, I will present an overview of the roles of SOX proteins during early mouse embryogenesis and in pluripotent stem cells, with a focus on SOXB proteins, which play essential roles in these processes.

4.1 PRE-IMPLANTATION DEVELOPMENT

During mouse embryogenesis, the fertilized egg (oocyte) undergoes a series of rapid cell divisions, known as cleavages, resulting in the formation of the morula (E2.5) that further develops to form the blastocyst (E3.5). The blastocyst consists of a fluid-filled cavity (blastocoel) and two spatially separated cell populations: the inner cell mass (ICM) and the outer cells that form the trophoectoderm (TE). Cells of the ICM are pluripotent and will give rise to the complete embryo, whereas TE cells will form the extra-embryonic ectoderm and generate the placenta (Wennekamp et al. 2013; Zhang & Cui 2014). By the time of uterine implantation, the late stage blastocyst (E4.5) already contains three distinct cell lineages: the TE and two new cell lineages originating from the ICM. The first of these two new lineages is the primitive endoderm, which will develop into extra-embryonic tissues. The second lineage, termed epiblast, is pluripotent and will give rise to the embryo proper, which is all the cells in the body (Wennekamp et al. 2013; Mihajlović et al. 2015) (**Figure 1**).

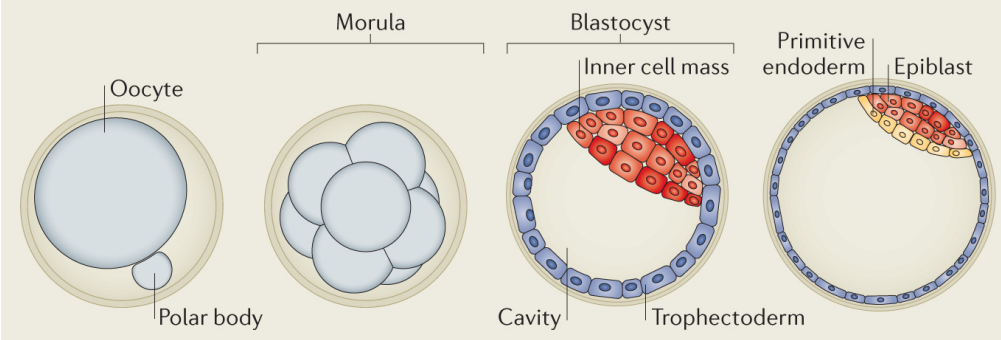

Figure 1. Overview of mouse pre-implantation development: from fertilized oocyte to blastocyst (Wennekamp et al. 2013).

Zygotic SOX2 is first expressed in the ICM and becomes restricted to the epiblast as development proceeds (Guth & Wegner 2008; Avilion et al. 2003). The ICM and the epiblast are both pluripotent cell populations, which suggests that SOX2 may be involved in the formation of early pluripotent embryonic cells. Indeed, zygotic deletion of *Sox2* is embryonically lethal shortly after implantation, and is the result of a failure to develop an epiblast (Avilion et al. 2003). Interestingly, maternal SOX2 expression starts in the oocyte and continues throughout pre-implantation development. Therefore, the expression of maternal SOX2 could mask an earlier phenotype in zygotic SOX2 mutants (Avilion et al. 2003; Keramari et al. 2010). A subsequent study used small interfering RNA (siRNA) to deplete both maternal and zygotic *Sox2* transcripts, which caused developmental arrest at the morula stage and thereby a failure to form a blastocyst (Keramari et al. 2010). These findings demonstrate that SOX2 plays critical roles in the formation of the ICM and the epiblast and therefore is an essential factor of pluripotency during pre-implantation development.

Similarly to SOX2, OCT4 and the homeodomain protein NANOG are also expressed in the ICM and epiblast of the developing embryo. Null mutations of either gene lead to embryonic lethality due a failure to form the ICM or the epiblast, which indicates that these proteins also play important roles in maintaining pluripotency in the early embryo (Nichols et al. 1998; Mitsui et al. 2003).

4.2 EMBRYONIC STEM CELLS

The first derivation of ESCs from the pre-implantation mouse embryo was reported in 1981 (Martin 1981; Evans & Kaufman 1981). During the following two decades, ESCs lines were derived from a number of mammalian species, including mouse and humans ESCs (mESCs, hESCs) (Thomson et al. 1998). ESCs can be derived from the pluripotent ICM or epiblast of a blastocyst-stage embryo. Similarly to ICM and epiblast cells, ESCs are pluripotent and can give rise to cells from all three embryonic germ layers, namely ectoderm, mesoderm, and endoderm. However, while pluripotency is a transient *in vivo* state that is limited to the blastocyst, ESCs grown in culture possess the unique ability to self-renew indefinitely (Plusa & Hadjantonakis 2014). Due to these properties, ESCs have been valuable tools to study pluripotency and early lineage specification. Importantly, hESCs are believed to hold great promise for regenerative medicine and cell-based therapies. However, a better understanding of how ESCs identity and differentiation are regulated on the transcriptional level is needed in order to realize their full therapeutic potential (Czechanski et al. 2014; Boyer et al. 2005).

Three TFs, SOX2, OCT4 and NANOG, are highly expressed in ESCs. mESCs cannot be derived from the ICM of *Sox2*-deficient embryos (Avilion et al. 2003). Moreover, deletion of *Sox2* in already established mESCs severely affects their pluripotency and self-renewal capacities as shown by the loss of pluripotency markers and by their differentiation into TE-like cells. Hence, these findings suggest that SOX2 is required both for the establishment and the maintenance of the pluripotent state in mESCs. Interestingly, OCT4 overexpression can rescue this phenotype. One possible function of SOX2 in mESCs may thus be to maintain a sufficient level of OCT4 expression (Masui et al. 2007).

Surprisingly, overexpression of SOX2 in mESCs does not lead to a reinforced pluripotent state, but instead leads to a loss of self-renewal and triggers their differentiation into multiple cell types (Kopp et al. 2008). Indeed, forced expression of SOX2 leads to the downregulation of endogenous SOX2 expression. It also downregulates the expression of important embryonic target genes of SOX2 such as *Nanog* (Kopp et al. 2008; Boer et al. 2007). Furthermore, modifying OCT4 expression levels in mESCs largely resembles the phenotypes observed with SOX2, with a loss of pluripotency and the differentiation of these cells, although towards different lineages (Nichols et al. 1998). Conversely, overexpression of NANOG in mESCs maintains their self-renewal capacities (Chambers et al. 2003). All together, these results demonstrate that the precise control of SOX2, OCT4 and NANOG expression levels is critical to maintaining pluripotency and self-renewal in ESCs.

A number of genome-wide binding studies have been performed in order to identify the downstream target genes of SOX2, OCT4 and NANOG as a means to better understand how they achieve their regulatory functions in ESCs. ChIP experiments combined with microarray (ChIP-on-Chip) or ChIP-seq experiments revealed that these factors bind to both promoters

and distal enhancers of thousands of genes in ESCs. It also showed that they co-bind a large number of target genes that they functionally regulate. Indeed, about half of the promoters bound by OCT4 are also bound by SOX2 in hESCs, and almost all of these promoters are additionally bound by NANOG. Many of their shared target genes are active in hESCs and encode TFs and components of signaling pathways known to be involved in pluripotency and self-renewal. Notably, they bind to the promoter of their own gene, forming auto regulatory loops in order to stabilize their own expression. Interestingly, these three factors also occupy genes that are inactive in hESCs and that are enriched for TFs implicated in various developmental processes such as PAX6, HOXB1 and LHX5 (Boyer et al. 2005; Chen et al. 2008; J. Kim et al. 2008). Thus, SOX2, OCT4 and NANOG bind and activate the expression of pluripotency genes in order to keep ESCs undifferentiated, while also occupying key developmental genes.

Collectively, functional experiments as well as genome-wide binding data in ESCs demonstrate the existence of a core transcriptional regulatory network in ESCs that is necessary to sustain the ESC state and consists of three TFs: SOX2, OCT4 and NANOG.

4.3 INDUCED PLURIPOTENT STEM CELLS

The importance of SOX2 in establishing and maintaining pluripotency was further confirmed with the fascinating discovery that somatic cells can be reprogrammed into ESC-like cells (Takahashi & Yamanaka 2006). Only five years after the publication of this groundbreaking work, Shinya Yamanaka was awarded the Nobel Prize in Physiology or Medicine.

The so-called induced pluripotent stem cells (iPSCs) were initially generated by the transduction of a cocktail of four TF genes that are important for pluripotency in ESCs, namely *Sox2*, *Oct4*, *Klf4* and *c-Myc* (Takahashi & Yamanaka 2006; Takahashi et al. 2007). iPSCs are similar to ESCs in that they are pluripotent cells capable of long-term self-renewal. They have been generated from various somatic cell types including mouse embryonic fibroblasts, adult mouse and human fibroblasts, and adult mouse neural stem cells (Takahashi & Yamanaka 2006; Takahashi et al. 2007; J. B. Kim et al. 2008).

Further studies have demonstrated that the other members of the SoxB1 group, SOX1 and SOX3, can substitute for SOX2 to generate iPSCs, but that other members of the SOX family, like SOX7 and SOX15, cannot (Nakagawa et al. 2008). Moreover, while SOX2 and OCT4 are sufficient for the reprogramming process, KLF4 and C-MYC are dispensable factors that can nevertheless be used to improve reprogramming efficiency (Huangfu et al. 2008; Nakagawa et al. 2008). SOX2 is thus essential for the efficacious reprogramming of somatic cells into iPSCs, and its physical interaction with OCT4 may have central to this process.

5 SOX PROTEIN FUNCTION DURING CENTRAL NERVOUS SYSTEM DEVELOPMENT

How a common pool of cells can generate all the differentiated cell types that compose the CNS is a fundamental question in developmental biology. In this chapter, I will highlight the functions of the SOX proteins involved in controlling cell proliferation and their differentiation into neurons and glial cells, in particular those of the SoxB, SoxC and SoxE groups, which are of relevance to the work presented in this thesis. I will mainly focus on the functions they exert in cells of the developing vertebrate spinal cord.

5.1 NEURAL TUBE DEVELOPMENT

Shortly after implantation of the blastula in the uterus, a complex and coordinated cell reorganization occurs, whereby cells of the epiblast migrate through a structure called the primitive streak, and move toward the inside of the blastula. This process, known as gastrulation (E6.5), gives rise to all three primary germ layers and establishes the major body axes of the embryo. Through a developmental process called organogenesis, each germ layer of the embryo will later differentiate and develop into specific tissues and organs. The endoderm will thus form the lining of the digestive and respiratory systems, while the mesoderm will form the bones, cartilages, muscles as well as some internal organs. The ectoderm becomes, upon neural induction, subdivided into the neuroectoderm, which will later give rise to the nervous system, and the non-neural ectoderm, which will form to the epidermis (Sanes et al. 2011; Wolpert 2002).

Soon after the formation of the neuroectoderm begins another fundamental embryonic event called neurulation, which leads to the formation of the neural tube. Induced by the mesoderm, neurulation corresponds to a sequence of steps that start with the thickening of the neuroectoderm and the elevation of the neural plate borders, causing the neural plate to fold onto itself and close, thereby creating the neural tube (E8.5-E10.5).

SOX2 and SOX3, which are both expressed in the epiblast of the blastula, become restricted to cells of the ectoderm during gastrulation (Wood & Episkopou 1999; Avilion et al. 2003). By E7.5, their expression is further confined to the neuroectoderm, where the third member of the SoxB1 group, SOX1, is additionally turned on (Wood & Episkopou 1999; Avilion et al. 2003). The continuous restriction of SOXB1 expression in cells fated to become part of the CNS points to their potential role in the neural commitment of the embryonic ectoderm. This is evidenced by studies in *Xenopus laevis* where the ectopic expression of *Sox2* or the injection of dominant negative forms of *Sox2* showed that *Sox2* is required for the differentiation of the ectoderm into the neuroectoderm (Kishi et al. 2000). Similarly, overexpression of either SOX1 or SOX2 in ESCs under culture conditions that do not support self-renewal promotes their differentiation into the neuroectodermal lineage at the expense of the mesodermal and endodermal fates (Zhao et al. 2004). Furthermore, recent studies have shown that SOX2, OCT4 and NANOG, which are crucial to maintaining pluripotency, also act as lineage specifiers, with SOX2 promoting ESC differentiation towards neuroectoderm, and OCT4 and Nanog towards mesendoderm (Loh & Lim 2011; Thomson et al. 2011). Taken together, these results highlight the role of SOXB1 proteins in early neural lineage commitment in the embryo and in ESCs.

All three SOXB1 proteins are expressed in a highly overlapping and conserved manner in the developing neuroectoderm of many species, such as in chicken, mouse, and human (Wood & Episkopou 1999; Collignon et al. 1996). Given the aforementioned results and their co-expression pattern, it is likely that SOXB1 proteins might function redundantly. This assumption is supported by the phenotypes of *Sox1* and *Sox3* knockout mice as well as *Sox2* hypomorphic mouse mutants (whose SOX2 expression level is lower), which surprisingly display relatively normal neuroectoderm formation and development, and therefore suggest that SOXB1 proteins might compensate for the loss of each other (Nishiguchi et al. 1998; Rizzoti et al. 2004; Ferri et al. 2004). However, genetic studies in which members of the SoxB1 group have been deleted in a combinatorial fashion have not yet been reported.

As development proceeds, the neural tube will evolve into the central nervous system, with the anterior part of the neural tube giving rise to the three primary brain vesicles (forebrain, midbrain and hindbrain), and the posterior part generating the spinal cord (Sanes et al. 2011; Wolpert 2002). The developing neural tube consists of neuroepithelial cells which are self-renewing multipotent NPCs lining the brain ventricles and the spinal central canal. These progenitor cells generate all the differentiated neural cell types that will constitute the mature CNS, by giving rise to neurons first and glial cells later (Reiprich & Wegner 2015). In the developing spinal cord, the transition from a proliferative progenitor cell to a postmitotic neural cell can be visualized. Indeed, the different cellular stages of this multiple step process are spatially organized into three distinct cell layers that are characterized by the expression of specific marker genes. The dividing NPCs are located in the innermost layer of the developing spinal cord called the ventricular zone (VZ). Upon differentiation, NPCs exit the cell cycle and migrate laterally to the intermediate zone. Finally, differentiating cells settle in the outer marginal zone. Each step is tightly controlled by TFs that coordinate the expression of specific sets genes in order to regulate neural cell fate specification and differentiation (Diez del Corral & Storey 2001; Kintner 2002) (**Figure 2**).

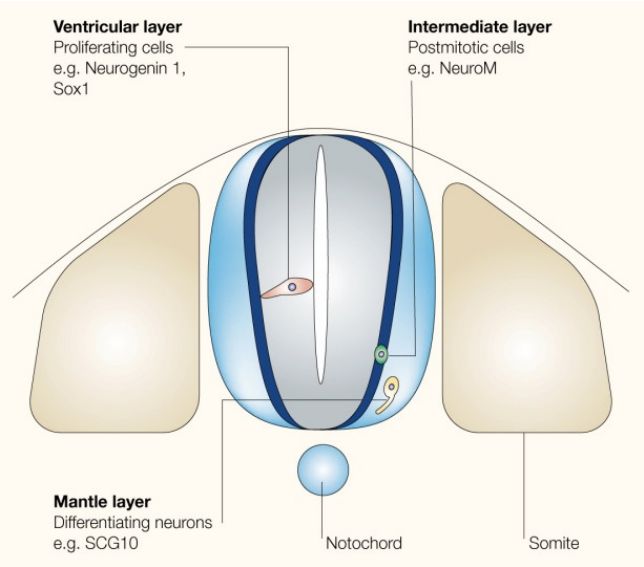

Figure 2. Cross-section of the developing spinal cord illustrating a layered organization characterized by the expression of specific markers genes (Diez del Corral & Storey 2001).

5.2 NEURONAL DEVELOPMENT

5.2.1 Neuronal progenitor cells maintenance

After neural tube formation, SOXB1 proteins continue to be expressed along the antero-posterior (A/P) axis of the developing neural tube (Collignon et al. 1996). With the onset of neurogenesis, the process by which a NPC differentiates to form a mature neuron, their expressing becomes restricted to the proliferating NPCs residing in the VZ of the developing CNS. Indeed, as cells differentiate, the expression of SOXB1 proteins is rapidly downregulated (Bylund et al. 2003; Graham et al. 2003).

Key misexpression experiments in the chicken neural tube have provided a better understanding of SOXB1 functions during early CNS development. Forced expression of either SOXB1 proteins using *in ovo* electroporation effectively keeps NPCs in a proliferative state and inhibits neuronal differentiation. Conversely, inhibition of their activity hinders self-renewal and proliferation, promotes premature exit from the cell cycle and induces neuronal differentiation, as evidenced by the precocious expression of neural differentiation markers (Bylund et al. 2003; Graham et al. 2003). Moreover, the finding that the three SOXB1 proteins produce the same effect indicates that they function redundantly in NPCs.

The differentiation of NPCs into neurons is dependent on the activity of proneural basic Helix-Loop-Helix (bHLH) proteins, such as NGN2. In NPCs, the expression of proneural bHLH proteins is maintained at a low level through active Notch signalling (Bertrand et al. 2002; Nakada et al. 2004). Interestingly, although Notch signalling can block NPCs differentiation, the maintenance of NPCs in an undifferentiated state still requires the activity of SOXB1 proteins (Holmberg et al. 2008). Indeed, forced expression of SOXB1 proteins block the activity of proneural bHLH proteins in NPCs. Conversely, the ability of proneural bHLH proteins to promote neural differentiation is partly dependent on its ability to downregulate SOXB1 expression in NPCs (Bylund et al. 2003). However, how SOXB1 and proneural bHLH proteins regulate each other's activity and expression is not fully understood.

Because deletion of *Sox2* is embryonically lethal, several studies have used conditional knock-out approaches to investigate the role of SOX2 in the developing mouse CNS. In a first study, the conditional neural-specific deletion of *Sox2* using a *Nestin*-Cre transgene is lethal at birth and only causes minor brain defects such as enlarged ventricles. Interestingly, a reduced number of progenitors was observed at E14.5, suggesting that SOX2 might be involved in the maintenance of NPCs in the developing brain (Miyagi et al. 2008). A second study also used a *Nestin*-Cre transgene to effectively delete *Sox2* from E12.5. Although only small brain defects can be detected, a complete loss of NPCs and neurogenesis is observed in the hippocampus of mutant mice one week after birth. This study additionally assessed the effect of the loss of *Sox2* *in vitro*, in NPCs cultured as neurospheres. In the absence of *Sox2*, the sphere formation capacity is progressively lost, cell death is increased, cell proliferation is decreased, and the neuronal marker TUJ1 is concomitantly expressed (Favaro et al. 2009). Consequently, *Sox2* plays a pivotal role in the maintenance of NPCs, both *in vivo* and *in vitro*.

In addition to being expressed from the earliest stages of the developing CNS, SOXB1 expression persists in NPCs of neurogenic niches in the adult brain, that is the subventricular zone of the lateral ventricles and the subgranular zone of the dentate gyrus of the hippocampus

(Ellis et al. 2004; Ferri et al. 2004; Wang et al. 2006). Reduction of *Sox2* expression in the adult mouse brain of *Sox2* hypomorphic mutants leads to a significant decrease in NPCs proliferation and an impaired adult neurogenesis. Therefore, in analogy with SOX2 function in embryonic NPCs, SOX2 maintains cells as proliferative progenitors in the adult brain (Ferri et al. 2004). Interestingly, the finding that adult *Sox2*-deficient mice exhibit defects in neural differentiation suggests that apart from maintaining NPCs undifferentiated, *Sox2* might also be required for the generation of subsets of neurons (Ferri et al. 2004).

Moreover, ectopic expression of SOX2 alone or together with other TFs was shown to directly reprogram fibroblasts into multipotent self-renewing induced NPCs without passing through a pluripotent state, which further supports the importance of SOX2 in NPCs (Han et al. 2012; Ring et al. 2012).

The global binding pattern of SOXB1 proteins, in particular that of SOX2, has been extensively studied in NPCs, both derived from mESCs and isolated from mouse embryonic tissues (Engelen et al. 2011; Lodato et al. 2013; Hagey & Muhr 2014) (**Paper I**, **Paper II**). These ChIP-seq experiments have revealed that SOX2 binds thousands of genes, many of which are expressed in NPCs and are known to be important for NPCs maintenance. However, SOX2 also binds genes that are not yet expressed in NPCs and that will be expressed at later stages of neural development, which will be discussed later in this thesis.

In summary, SOXB1 proteins are essential for the maintenance of NPC properties in the developing and adult CNS.

5.2.2 Neuronal differentiation

SOX21, a member of the SoxB2 subgroup exhibiting trans-repression activity, is co-expressed with SOXB1 proteins in NPCs of the developing and adult CNS (Rex et al. 1997; Uchikawa et al. 1999; Matsuda et al. 2012). However, SOX21 has the opposite function compared to SOXB1 proteins and promotes neural differentiation. Indeed, ectopic expression of SOX21 in the chicken neural tube induces NPCs to downregulate progenitor makers, exit the cell cycle and express neuronal makers. It appears that the precise balance between SOX21 and SOXB1 proteins influences whether cells will remain as progenitors or undergo neurogenesis, and it is thought that they might compete for binding to the same target genes. Interestingly, the ability of proneural proteins to repress SOXB1 proteins in NPCs, and thereby to promote neuronal differentiation, depends on their capacity to influence this balance by upregulating SOX21 expression (Sandberg et al. 2005). Furthermore, SOX21 was reported to regulated hippocampal adult neurogenesis by repressing the Notch-responsive gene *Hes5* (Matsuda et al. 2012).

The differentiation of NPCs into neurons is characterized by the switch from SOXB1 and SOXB2 proteins to the expression of SOXC proteins, which consist of SOX4, SOX11 and SOX12. Thus, as NPCs exit the cell cycle and migrate away from the VZ, SOXB1 and SOXB2 are downregulated and SOXC proteins start to be expressed. SOXC proteins show overlapping expression in postmitotic neural cells committed to neuronal differentiation, where they are responsible for the establishment of pan-neuronal characteristics (Bergsland et al. 2006). Indeed, ectopic expression of SOX4 or SOX11 in NPCs of the chicken neural tube induces the precocious expression of pan-neuronal genes such as *Tubb3* and *Map2*. Unexpectedly, these cells remain in the cell cycle and continue to express SOXB1 proteins, and therefore do not

differentiate per se. Importantly, the ability of proneural bHLH proteins to induce neuronal differentiation relies on the activity of SOXC proteins, their downstream effector (Bergsland et al. 2006). Hence, the activation of pan-neuronal genes by SOXC proteins is an important step towards the formation of a differentiated neuron, and appears to be independent of cell cycle exit.

In addition to the CNS, SOXC proteins are expressed in numerous other tissues during mouse development. Consequently, mice deficient for *Sox4* or *Sox11* display multiple organ defects, which hints to their role in organogenesis (Schilham et al. 1996; Sock et al. 2004; Bhattaram et al. 2010). *Sox12*-deficient mice however develop normally despite its widespread expression (Hoser et al. 2008). Nevertheless, while *Sox4* and *Sox11* double knockout embryos die at E10.5, the additional inactivation of *Sox12* worsens this phenotype and embryos die earlier, at E8.5, indicating that these three genes function in redundancy in multiple developmental processes (Bhattaram et al. 2010). With regard to the CNS, the neural-specific deletion of both *Sox4* and *Sox11* using a *Brn4*-Cre transgene results in a strong increase in cell death, demonstrating the requirement of *Sox4* and *Sox11* for cell survival in the developing mouse neural tube (Thein et al. 2010; Bhattaram et al. 2010).

SOXC proteins are transiently expressed in neuronal cells and are downregulated as development proceeds. However, SOXC proteins continue to be highly expressed in neurogenic niches of the adult mouse brain. In the dentate gyrus of the hippocampus, the expression of SOX4 sand SOX11 is initiated in committed neuronal cells and coincides with the downregulation of SOX2 and the onset of neuronal markers expression. Their expression is maintained in immature neurons but is eventually downregulated in mature neurons (Haslinger et al. 2009; Mu et al. 2012). Overexpression of either SOX4 or SOX11 in adult hippocampus-derived NPCs is sufficient to induce the expression of neuronal markers such as DCX and TUJ1, despite culture conditions promoting proliferation. Conversely, NPCs in which *Sox4* and *Sox11* were ablated exhibit a reduced expression of the aforementioned neuron-specific proteins. Consistent with these *in vitro* experiments, NPCs in the hippocampus of *Sox4* and *Sox11* double conditional knockout mice have a decreased expression of neuronal markers, and instead maintain SOX2 expression. However, in contrast to the phenotype observed during embryonic development, deletion *Sox4* and Sox11 does not appear to affect cell survival in the adult brain. Together, these gain and loss-of-function experiments demonstrate that SOX4 and SOX11 are essential regulators of neuronal fate commitment in the adult hippocampus (Mu et al. 2012), which is reminiscent of their role in the developing chicken neural tube (Bergsland et al. 2006).

Strikingly, ChIP-seq experiments have shown that despite having opposite functions during neural development, the binding patterns of SOX2 and SOX3 in NPCs and of SOX11 in early neurons overlap greatly (**Paper I**).

Interestingly, in addition to regulating the early steps of neurogenesis, SOXB1 proteins also function in subsets of postmitotic neurons in the adult mouse brain. Indeed, while SOXB1 proteins are normally downregulated upon cell cycle exit, they are maintained or re-expressed in specific and mostly non overlapping populations of differentiated neurons where they exert distinct functions, including cell differentiation and migration (Malas et al. 2003; Rizzoti et al. 2004; Ferri et al. 2004; Ekonomou et al. 2005).

In conclusion, these studies demonstrate that neuronal development relies on multiple SOX proteins that have distinct functions, with SOXB1 maintaining NPCs undifferentiated, and SOX21 and SOXC proteins promoting neuronal differentiation.

5.3 GLIAL DEVELOPMENT

5.3.1 Spinal cord patterning and glial specification

As mentioned previously, the vast majority of neuronal and glial cell types that constitute the mature CNS are produced during embryonic stages by multipotent NPCs residing in the VZ of the developing neural tube. Along the dorso-ventral axis of the embryonic spinal cord, the VZ is actually divided into distinct regionally-restricted progenitor domains that will each give rise to a specific neuronal cell type. For instance, in the ventral spinal cord, NPCs of the pMN domain will generate motor neurons, while NPCs of the p0-p3 domains will produce different subtypes of interneurons (Rowitch 2004; Kessaris et al. 2008). The patterning of the developing spinal cord into distinct progenitor domains is mediated ventrally by the concentration gradient of sonic hedgehog (SHH) and dorsally by the gradient of bone morphogenetic proteins (BMPs), which are secreted by the floor plate and the roof plate, respectively. These signaling molecules control the expression of sets of TFs that are spatially restricted to a particular progenitor domain. These TFs in turn regulate distinct downstream target genes in order to establish unique NPCs identities and consequently generate specific neuronal cell types (Rowitch 2004; Kessaris et al. 2008) (**Figure 3**).

During development, multipotent NPCs first give rise to neurons, followed by astrocytes and oligodendrocytes. They do so by transitioning from a neurogenic to a gliogenic potential (Kiefer 2007). The so-called neuron to glia switch predominantly takes place in two of the ventral progenitor domains of the developing spinal cord, the pMN and the p2 domains, wherein the vast majority of glial cells are generated (Richardson et al. 2006). NPCs of the p2 domain first generate V2 interneurons and later switch competence to give rise to astrocytes. By contrast, NPCs of the pMN domain, which are characterized by the expression of the transcription factor OLIG2, produce motor neurons first and then oligodendrocytes (Kessaris et al. 2001; Zhou & Anderson 2002).

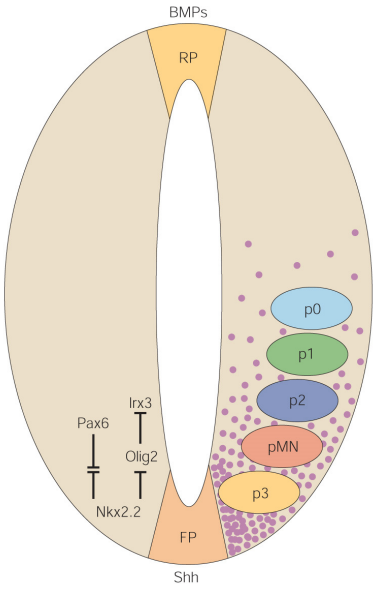

Figure 3. Patterning of the ventral embryonic spinal cord into distinct progenitor domains. Pink dots: SHH concentration gradient; RP: roof plate; FP: floor plate (Rowitch 2004)

It is well established that SOX9, a member of the SoxE group, is a major regulator of this competence switch. In NPCs of the mouse spinal cord, SOX9 starts to be expressed ventrally at E9.5 and is then expressed throughout the VZ by E10.5 where its expression strongly overlaps that of SOXB1 proteins (Scott et al. 2010). SOX9 expression continues in proliferating, migrating oligodendrocyte progenitor cells (OPCs) and is downregulated in differentiating, myelin-forming oligodendrocytes. On the other hand, SOX9 expression is maintained in astrocytes until adulthood. Conditional deletion of *Sox9* using a Nestin-Cre

transgene is characterized by the dramatic reduction in oligodendrocyte and astrocyte progenitor cells, coupled with an increased number of motorneurons and V2 interneurons. The extended period of neurogenesis together with the delayed gliogenesis point to a neuron to glia switch defect (Stolt et al. 2003). Hence, SOX9 is required for both oligodendrocyte and astrocyte fate specification in the developing mouse spinal cord.

Two members of the nuclear factor I (NFI) family of transcription factors, NFIA and NFIB, start to be expressed in the developing mouse spinal cord at E11.5, at the time when gliogenesis is initiated. These factors have been shown to be both necessary and sufficient to promote glial fate specification *in vivo*, and are thus key regulators of the neuron to glia switch (Deneen et al. 2006). Importantly, a later study has demonstrated that the expression of NFIA is directly regulated by SOX9 (Kang et al. 2012). Hence, SOX9 appears to initiate the neuron to glia switch by activating NFIA which in turn induces gliogenesis.

Although severely affected, the specification of oligodendrocytes cells is not completely lost in the absence of *Sox9* (Stolt et al. 2003). Interestingly, SOX8, another member of the SoxE group, is induced in ventral NPCs shortly after SOX9 (E11), is expressed along with SOX9 in proliferating OPCs, and in contrast to SOX9 is maintained in mature, terminally differentiated oligodendrocytes (Stolt et al. 2004; Stolt et al. 2005). Even though deletion of *Sox8* alone does not impair oligodendrocyte specification by itself (Sock et al. 2001), the combined loss of *Sox8* and *Sox9* leads to a more severe specification defect defined by the complete absence of oligodendrocytes (Stolt et al. 2005). These results suggest that SOX8 and SOX9 function redundantly to control glial specification.

5.3.2 Oligodendrocyte development

The generation of neurons and glial cells are multi-step processes that include specification, migration and differentiation. Unlike neurons, which become post-mitotic and start migrating away from the VZ shortly after being specified, glial cells do not immediately exit the cell cycle after their specification, and instead migrate throughout the CNS as proliferative progenitor cells before settling to become post-mitotic, differentiated cells (Kessaris et al. 2008). A number of SOX proteins have been identified to be expressed at different phases of oligodendrocyte development in the mouse spinal cord, and their respective roles have been uncovered thanks to the use of genetic studies, as presented below.

The third member of the SoxE group, SOX10, is selectively expressed in oligodendrocytes and is thus one of the key markers of the oligodendrocyte lineage (Stolt et al. 2002). Its expression is initiated upon glial specification (E11.5) in proliferating OPCs in the pMN domain of the ventral VZ. SOX10 is co-expressed with SOX8 and SOX9 in migrating, proliferating OPCs (Stolt et al. 2002; Stolt et al. 2003; Stolt et al. 2005). In analogy with SOX8, SOX10 expression persists in oligodendrocytes as they settle in the marginal zone of the developing spinal cord and undergo terminal differentiation (E18.5) (Stolt et al. 2002; Zhou & Anderson 2002; Stolt et al. 2003; Stolt et al. 2004; Stolt et al. 2005).

Deletion of *Sox10* does not affect the number, appearance or distribution of OPCs in the mouse embryonic spinal cord (Stolt et al. 2002). In a similar way, the selective ablation of *Sox9* in already specified OPCs using a Sox10-Cre transgene does not affect oligodendrocyte development (Finzsch et al. 2008). Considering that all three members of the SoxE group are

co-expressed once OPCs are specified, the observed lack of OPCs defect is likely due to the fact that SOXE proteins compensate for the loss of one another in proliferating OPCs. This assumption is supported by the combined deletion of *Sox9* and *Sox10* following OPCs specification. In this study, the number of OPCs is significantly reduced due to an increased apoptosis, and their migration pattern is altered. These results demonstrate that SOX9 and SOX10 function in a redundant manner during the time window when they are co-expressed, and are required for OPCs survival and migration during oligodendrocyte lineage progression (Finzsch et al. 2008).

Unlike SOX9, whose expression ends when OPCs start to differentiate, SOX8 and SOX10 are maintained throughout the maturation phase of oligodendrocyte development. SOX10 expression is even increased with the onset of terminal differentiation (Stolt et al. 2002; Stolt et al. 2003; Stolt et al. 2005). It is therefore not surprising that SOX10 plays a crucial role in the final step of oligodendrocytes development. Indeed, while proliferating OPCs are not affected by the ablation of *Sox10* alone, the terminal differentiation of oligodendrocytes is in contrast profoundly impaired. OPCs fail to undergo terminal differentiation as illustrated by the absence of myelin gene expression such as myelin basic protein (*Mbp*) and myelin proteolipid protein (*Plp*). As a consequence, *Sox10*-deficient mice display severe myelination defects and die around the time of birth (Stolt et al. 2002). These results therefore indicate that SOX10 is required for terminal differentiation of oligodendrocytes. Nevertheless, the expression of myelin genes is not completely abolished in the absence of *Sox10* (Stolt et al. 2002). The lack of complete penetrance may be due to the overlapping expression of SOX8 during the phase of terminal differentiation. Indeed, SOX8 has been shown to partially compensate for the loss of SOX10 during oligodendrocyte differentiation (Stolt et al. 2004; Kellerer et al. 2006).

ChIP experiments followed by polymerase chain reaction (PCR) have demonstrated that SOX10 binds to the *Mbp* promoter, suggesting that SOX10 directly controls the expression of myelin genes (Stolt et al. 2006). Accordingly, a recent study investigating the genomic occupancy of SOX10 in oligodendrocytes from postnatal rat spinal cords demonstrated that SOX10 binds many regulatory regions of genes critical for myelination (Lopez-Anido et al. 2015).

Furthermore, SOX10 can be used together with other TFs for the direct reprogramming of mouse embryonic fibroblasts into induced OPCs capable of giving rise to myelinating oligodendrocytes, which further supports the central role of SOX10 during oligodendrocyte development (Najm et al. 2013; Yang et al. 2013).

The temporal control of oligodendrocyte development is also regulated by the SOXD proteins SOX5 and SOX6. Their expression is initiated in the VZ of the spinal cord at E10.5, shortly before the onset of oligodendrocyte specification. SOX5 and SOX6 expression pattern is similar to that of SOX9 as they are maintained in proliferating OPCs but are downregulated prior to terminal differentiation. Genetic studies have revealed that SOX5 and SOX6 control several stages of oligodendrocyte development in a redundant manner. In mice deficient for *Sox5* and *Sox6*, specification of OPCs occurs earlier and oligodendrocytes differentiate prematurely (Stolt et al. 2006). Hence, SOX5 and SOX6 are involved in repressing oligodendrocyte specification and terminal differentiation. Therefore, they exert opposite

functions to SOX9 and SOX10, which promote oligodendrocyte specification and terminal differentiation, respectively (Stolt et al. 2002; Stolt et al. 2003; Stolt et al. 2006). As for the mechanism, it appears that SOXD proteins directly counteract SOX10 function by competing for the same binding sites in myelin genes, such as the *Mbp* promoter, which likely prevents their activation by SOX10. Once SOXD proteins are downregulated, SOX10 can finally activate myelin genes and regulate terminal differentiation (Stolt et al. 2006).

In summary, SOX proteins have key roles at each step of oligodendrocyte development. How does a particular SOX protein perform distinct functions at different stages of oligodendrocytes development is not well understood. Getting a better picture of their respective downstream target genes at different stages of development would certainly help better understand how they exert their function.

5.3.3 Astrocyte development

Although astrocytes are the most abundant cell type in the CNS, they were until recently considered to mainly serve as metabolic support to neurons, and therefore have been far less studied than neurons and oligodendrocytes. Consequently, while the functions of the SOX proteins expressed in the neuronal and oligodendrocyte lineages have largely been elucidated, much less is known about the roles of SOX proteins during astrocyte development. With the growing understanding that astrocytes play active roles in CNS development and diseases, it is crucial to better understand the molecular mechanisms regulating their specification and differentiation (Molofsky & Deneen 2015).

One of the key factors of astrocyte development is SOX9, the only member of the SoxE group to be expressed in the astrocyte lineage. While SOX9 is selectively downregulated in the oligodendrocyte lineage with the onset of terminal differentiation, it is expressed in all phases of astrocyte development and continues to be expressed in the adult. This suggests that SOX9 probably has roles beyond astrocyte specification, and might be important at later stages of astrocyte development (Stolt et al. 2003).

As already discussed, an important function of SOX9 is to induce the expression of NFIA, a key factor of the neuron to glia switch. Following its induction, NFIA physically interacts with SOX9 to bind and regulate common target genes with migration and metabolic roles in astrocytes, as supported by ChIP-PCR experiments (Deneen et al. 2006; Kang et al. 2012). The importance of these factors in astrocyte development is further highlighted by the finding that the ectopic expression of SOX9, NFIA and NFIB can direct the conversion of mouse embryonic fibroblasts into astrocytes (Caiazzo et al. 2015). However, revealing the binding pattern of SOX9 through ChIP-seq experiments would greatly help gain new insights into the role of SOX9 during astrocyte development (**Paper III**).

6 AIMS

The aim of this thesis is to unravel novel insights into the mechanisms regulating the maintenance and differentiation of neural progenitor cells at the transcriptional level.

More specifically, we aim to:

I. Characterize the genome-wide binding profiles of SOXB1 and SOXC proteins to examine their sequential roles during neuronal differentiation.

II. Examine how chromatin accessibility and transcription factor binding interact to regulate the establishment of gene expression in neural progenitor cells.

III. Characterize the genome-wide binding profile of SOX3 and SOX9 in glial progenitor cells in order to determine their regulatory roles during the development of astrocytes and oligodendrocytes.

7 RESULTS

7.1 PAPER I

Sequentially acting Sox transcription factors in neural lineage development

The generation of neurons from stem cells depends on several different SOX proteins, from early neural lineage specification in ESCs to later stages of neuronal differentiation. SOX2 is essential for the establishment and maintenance of ESCs (Avilion et al. 2003), SOXB1 proteins maintain NPCs undifferentiated (Bylund et al. 2003; Graham et al. 2003), and SOXC proteins are important for the induction of neuronal characteristics (Bergsland et al. 2006). In order to understand how different groups of SOX proteins functionally achieve their distinct activities, we have used ChIP-sequencing to identify their downstream target genes during neural lineage formation. We have determined the genome-wide binding profiles of SOX2/SOX3 and SOX11 in mESCs-derived NPCs and immature neurons, respectively, and compared them to the previously identified binding profile of SOX2 in mESCs (Chen et al. 2008).

The analysis of the ChIP-seq experiments in NPCs revealed that SOX3 binds more than 9000 binding sites, which corresponds to approximately 4000 different target genes, as a given gene can be associated with several binding sites. Interestingly, consistent with the redundant functions of SOXB1 proteins in NPCs, the binding pattern of SOX2 and SOX3 largely overlapped, as 96% of the sites targeted by SOX2 are co-occupied by SOX3. Moreover, in contrast to another study, which found that SOX2 primarily binds to promoters (Engelen et al. 2011), we found that most regions bound by SOX3 (and by analogy by SOX2) are enhancers rather than promoters, which is comparable to the binding pattern of SOX2 in mESCs (Chen et al. 2008). In fact, the comparison of SOX3 binding sites in NPCs with that of the transcriptional co-activator p300, which can predict active enhancers with a high accuracy (Visel et al. 2009), revealed that SOX3 binds a large number of enhancers that are active specifically in the CNS.

Although SOX2 is expressed in NPCs, it is best known for its role in pluripotent stem cells like ESCs. By comparing the binding pattern of SOX2 in mESCs with that of SOX3 in NPCs, we found that nearly half of the genes bound by SOX2 are later bound by SOX3. Using publicly available expression profiles, we found that while the genes that are uniquely bound by SOX2 in mESCs are most significantly expressed in mESCs, the genes that are sequentially targeted by SOX2 and SOX3 are mostly expressed in NPCs. These results show that in mESCs, SOX2 pre-binds silent genes that will be expressed and bound by SOX3 at a subsequent developmental stage. However, although the binding patterns of SOX2 in mESCs and SOX2/SOX3 in NPCs overlap greatly, they mostly target their common genes through different sites. One possible explanation is that the target site selection relies on the interaction with distinct cell type-specific partner factors. Accordingly, SOX2 binding sites in mESCs, but not SOX3 (and by analogy SOX2) binding sites in NPCs, are highly enriched for OCT4 motifs.

SOX2 in ESCs has previously been shown to bind H3K4me3 marked regulatory regions of active genes involved in pluripotency, as well as bivalent (H3K4me3 and H3K27me3 marked) regulatory regions of silent genes, which will be expressed later during development. The analysis of the expression pattern of the bivalent genes bound by SOX2 in ESCs revealed that these genes become activated in the neural lineage, but not in the endodermal and mesodermal

lineages. Thus, by preselecting neural genes in mESCs, SOX2 may be involved in establishing the competence of ESCs to develop along the neural lineage.

In addition, a large proportion of the SOX3 bound genes are not pre-bound by SOX2 in mESCs, and are in fact uniquely bound by SOX3 in NPCs. These genes are most highly expressed at later stages of neural lineage development, in late populations of neurons and/or glial cells. Knowing that SOXC proteins are expressed in neurons and are important for the induction of neuronal characteristics, we performed SOX11 ChIP-seq experiments in mESC-derived early neurons, and compared the binding profile of SOX11 with that of SOX3.

Strikingly, we found that the binding profiles of SOX3 and SOX11 overlap extensively despite having different functions during neural development. Indeed, 92% of the sites bound by SOX11 in early neurons are already bound by SOX3 in NPCs. We found that the genes that are sequentially bound by SOX3 and SOX11 are most significantly expressed in NPCs and early neurons. While the binding of SOX11 to neuronal genes can be expected, its binding to progenitor genes, which are silent when SOX11 is expressed, is more surprising and suggests that SOX11 may help turn off their expression. The genes that are uniquely bound by SOX3 are mostly expressed in late populations of neurons and/or glial cells. Together, these results show that in NPCs, SOX3 preselects many silent genes that will be bound by SOX11 and expressed in developing neurons.

To examine the regulatory roles of SOX3 and SOX11 on their target genes, we performed a number of functional experiments. First, we stably overexpressed SOX3 in ESC-derived NPCs. The gene expression analysis of the misregulated genes coupled with the analysis of the SOX3 ChIP-seq data revealed that SOX3 primarily activates the NPC genes it binds to, but does activate the bound neuronal genes. Next, we investigated the role of SOX3 and SOX11 on commonly bound putative neuronal enhancers. Misexpression of SOX11 in the chicken neural tube as well as transactivation assays showed that SOX11 activates these enhancers. However, SOX3 blocks the SOX11-mediated reporter activation, and overexpression of increasing amounts of SOX3 in the chicken neural tube efficiently blocks the SOX11-mediated induction of the neuronal gene *Tubb3*. These findings suggest that SOX3 binds and activates enhancers of progenitor genes in NPCs, and also pre-binds enhancers of neuronal genes that will be targeted and activated by SOX11 in neurons. It is possible that the pre-binding of SOX3 on neuronal enhancers may prevent their precocious activation by SOX11 during the short time window when SOX3 and SOX11 are co-expressed in proliferative NPCs, until SOX3 is eventually downregulated.

Thus, in analogy with the pre-binding of SOX2 to silent progenitor genes in ESCs, and their subsequent binding and activation by SOX3 in NPCs, SOX3 also pre-binds many silent neuronal genes that are subsequently bound and activated by SOX11 in neurons.

This finding prompted us to examine the histone modification marks located at SOX3 and SOX11 bound enhancers. Using sequential ChIP-qPCR, we found that in NPCs, the SOX3-bound enhancers associated with genes expressed in NPCs carry the H3K4me3 mark only. In contrast the SOX3-bound enhancers associated with silent neuronal and glial enhancers are marked by both H3K4me3 and H3K27me3. Upon differentiation of NPCs into neurons and SOX11 binding, the H3K4me3 mark associated with the previously active progenitor

enhancers is replaced by the H3K27me3 mark, and the bivalent mark on neuronal enhancers is resolved into a monovalent H3K4me3 mark as enhancers become activated.

To test whether SOX proteins are directly involved in the deposition of histone modification marks, we ectopically expressed SOX3 in the mesodermal progenitor cell line C2C12. We observed that SOX3 promotes the accumulation of the H3K4me3 mark at its bound enhancers, indicating that SOX3 may alter the chromatin state locally.

Together, these results indicate that an important role of SOXB1 proteins in NPCs is to maintain silent genes poised for activation by the sequentially acting SOXC protein, SOX11. Consistently, in mESCs, SOX2 has been shown to bind and epigenetically prepare the locus λ5-VpreB1 by facilitating the addition of H3K4me2 and H3K4me3 marks. Upon B cell differentiation, this locus is bound and activated by the SOXC protein SOX4 (Liber et al. 2010).

In summary, these data provide evidence that SOX proteins pre-bind large sets of poised silent genes in ESCs and NPCs that are destined to be activated by alternative SOX proteins at later stages of neuronal differentiation. Here, we propose a model whereby the sequential binding of distinct SOX proteins to common sets of genes drives neurogenesis (**Figure 4**).

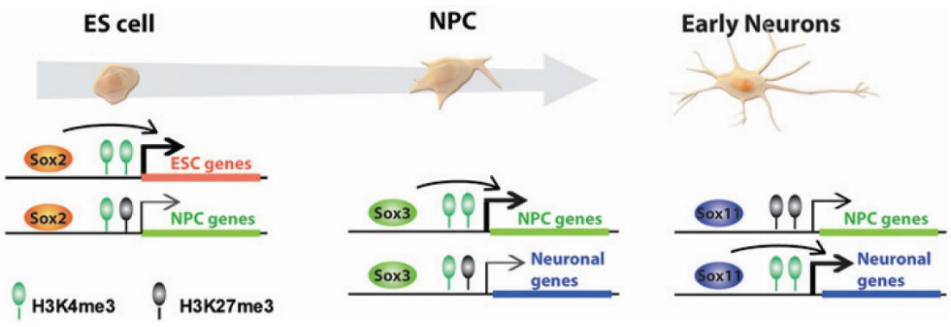

Figure 4. Sequential binding of SOX proteins during neural lineage development. Model depicting the sequential binding of distinct SOX proteins to common genes during the differentiation of ESCs towards the neuronal lineage, and highlighting the association between SOX pre-binding and bivalent histone modification marks (Bergsland et al. 2011).

7.2 PAPER II

Distinct transcription factor complexes act on a permissive chromatin landscape to establish regionalized gene expression in CNS stem cells

The CNS is a complex organ consisting of thousands of different neural cell types that are mainly produced during embryonic stages. The generation of this impressive cell diversity relies on the fact that NPCs at distinct spatial locations along the major axes of the developing embryo exhibit distinct gene expression profiles. Gene expression is largely regulated by the combined actions of TFs and their interacting partners on their downstream target genes. However, the binding of TFs is also greatly affected by the extent of chromatin accessibility and chromatin compaction. Here, we have used genome-wide approaches to investigate how these different regulatory features interact to establish specific gene expression profiles in NPCs from different regions of the developing mouse neural tube, the cortex (CX) and the spinal cord (SC). We have generated and compared DNase-seq, RNA-seq and SOX2 ChIP-seq data in NPCs from the E11.5 cortex and spinal cord.

First, we investigated how the chromatin landscape correlates with the differences in gene expression between NPCs of the CX and the SC. The stringent analysis of the DNase-seq experiments identified approximately 34,000 DHSs in each NPC population. CX and SC NPCs share a large number of open chromatin regions, as 75% of the DHSs are common to both cell types. Interestingly, the majority of these common DHSs are also present in ESCs as well as in endodermal and mesodermal progenitor cells. Conversely, most of DHSs specific to CX or SC NPCs are not shared with other stem or progenitor cells populations, and are overwhelmingly located distant from the closest transcription start site (TSS). These results suggest that cell type-specific DHSs correspond to enhancer regions that are mostly formed *de novo* during CNS development.

By comparing the RNA-seq data of CX and SC NPCs, we identified genes that are specifically expressed in CX or in SC NPCs, as well as genes expressed in both cell types. The comparison of the chromatin accessibility data with these different sets of expressed genes revealed that DHSs specific to the CX are enriched around genes that are specifically expressed in the CX, whereas DHSs specific to the SC are enriched around genes that are specifically expressed in the SC. However, common DHSs are not only enriched around commonly expressed genes, but are also enriched around genes that are specifically expressed in CX or SC NPCs. To further characterize these common DHSs, we assessed the number of DHS reads in each cell type. Interestingly, we found that common DHSs with a higher number of reads in CX NPCs compared to SC NPCs were enriched for genes that are specifically expressed in CX NPCs, and vice versa. Hence, the number of DHS reads in a particular cell type is a better predictor of gene expression than simply the occurrence of a DHS. Such observation had previously been reported in hematopoietic cells, where a high number of reads in DHSs at TSS has been correlated to a strong gene expression (Boyle et al. 2008).

Next, we investigated how differences in chromatin accessibility between CX and SC NPCs correlate with TF binding. All NPCs of the developing CNS express high levels of SOX2, which is a key regulator of NPCs. We thus performed SOX2 ChIP-seq experiments in CX and in SC NPCs, and found that the genome-wide binding profile of SOX2 differs quite extensively between the two cell types. Using our gene expression data sets and gene ontology analysis, we found that genes that are uniquely bound by SOX2 in the CX are most significantly expressed in CX NPCs, whereas genes that are exclusively bound by SOX2 in the SC are mostly expressed in SC NPCs. We then examined the overlap between SOX2 bound regions (also called peaks) and DHSs in order to study the relationship between SOX2 binding and chromatin accessibly. These comparisons revealed that common SOX2 peaks are found within regions of common DHSs. Additionally, although SOX2 specific peaks are enriched within specific DHSs, they are surprisingly often found within regions of common DHS. However, by assessing the number of DHS reads, we found a correlation between SOX2 binding and the extent of chromatin accessibility in CX and SC NPCs. These data suggest that the number of DHS reads in a particular cell type can more accurately predict TF binding than the occurrence of a DHS.

SOX TF are known to interact with partner TFs not only to modulate their transcriptional regulatory activity, but also to refine their target site selection (Kondoh & Kamachi 2010). Thus, in order to better understand why DHSs that are commonly found in the CX and the SC are often bound by SOX2 in only one of these two cell types, we performed a DNA motif

analysis in CX and SC specific SOX2 peaks. This analysis revealed a strong enrichment of LHX2 motifs in SOX2 bound regions in the CX, and an enrichment of HOXA9 motifs in SOX2 bound regions in the SC. In addition, the analysis of DHSs for DNase I footprints (short DNA sequences within a DHS that are bound by regulatory proteins and are thus resistant to DNase I cleavage (Galas & Schmitz 1978)) identified neighboring SOX2 LHX2 motifs in CX NPCs, and SOX2 HOXA9 motifs in SC NPCs. These results suggest that SOX2 binds together with LHX2 and HOX proteins in NPCs of the CX and the SC, respectively.

To examine whether SOX2 bound DNA regions are involved in specifying specific gene expression in distinct NPC populations of the developing CNS, we used a GFP-reporter assay in zebrafish embryos allowing the visualization of the spatial distribution of enhancer activity *in vivo*. We found that the regions bound by SOX2 function as enhancers capable of driving reporter gene expression, and are activated in a spatially restricted fashion. Indeed, while DNA regions commonly bound by SOX2 in the CX and SC activate GFP expression along the entire A/P axis of the developing CNS, regions specifically bound by SOX2 in the CX exclusively activate reporter gene expression in the brain, and the reverse is true for regions specifically bound by SOX2 in the SC. Moreover, loss-of-function experiments, by mutating SOX motifs or injecting morpholinos against SOX2 and SOX3, indicate that SOX2 is necessary for the activation of the majority of the enhancers tested. Additionally, the mutations of LHX motifs in CX enhancers, and of HOX motifs (together with PBX and MEIS, two known partners of HOX proteins (Shanmugam et al. 1999)) in SC enhancers, result in the loss of enhancer activity. This suggests that these motifs are necessary for enhancer activation, and that the proteins binding to these motifs may function as SOX2 partners. Furthermore, interchanging HOX and LHX motifs between CX and SC enhancers is sufficient to re-specify the spatial distribution of the enhancers. Indeed, "swapped" SC enhancers are now activated in the brain and vice versa. Finally, we created enhancers in which all the nucleotides are randomized, except for SOX, LHX and HOX motifs (as well as PBX and MEIS motifs). These "synthetic" enhancers are able to activate reporter gene expression, although at lower levels compared to wild type enhancers. Taken together, these experiments demonstrate that SOX, LHX and HOX motifs (as well as PBX and MEIS motifs) are necessary and sufficient to confer cell type-specific enhancer activity. Consistent with these results, immunoprecipitation experiments as well as transactivation assays indicate that SOX2 and LHX2 can physically interact, at least *in vitro*, and cooperate to activate CX specific enhancers, whereas SOX2 and HOXB6 (together with PBX3 and MEIS1) can interact and cooperate to activate SC specific enhancers.

Our findings indicate that SOX2 often targets DNA in a cell type-specific manner although the DNA is accessible in both CX and SC NPCs. Moreover, our data further indicate that the cell type-specific binding pattern of SOX2, within a permissive chromatin landscape, can be explained by the region-specific distribution of SOX2 partner factors, LHX2 and HOX proteins. We next wanted to examine whether the misexpression of HOXB6 protein in CX NPCs could induce ectopic expression of SC genes in CX NPCs. To address this question, we performed *in utero* electroporation on E13.5 mouse cortices, followed by fluorescent-activated cell sorting (FACS) and RNA-seq. The misexpression of HOXB6 induces the upregulation of many SC expressed genes and the downregulation of many CX expressed genes. We then further analyzed these genes by characterizing their DHSs and SOX2 binding profiles, as well

as their enrichment for HOX motifs. We found that the upregulated SC genes are often associated with DHSs, bound by SOX2 in both CX and SC NPCs, and enriched in HOX motifs.

Finally, we used our DNase-seq, SOX2 ChIP-seq and RNA-seq data sets to generate a hierarchical statistical model in order to explain gene expression differences between CX and SC NPCs. This model reveals that the number of DHS reads is the best predictor of gene expression specificity in NPCs from different regions of the developing CNS. Moreover, the number of DHS reads is influenced by the relative SOX2 binding, which is in turn predicted by HOX and LHX motifs enrichment.

In this study, we have shown that despite being ubiquitously expressed in NPCs of the developing CNS, SOX2 acts on a permissive chromatin landscape to establish distinct gene expression patterns in NPCs from different regions of the developing CNS. The differences in SOX2 binding pattern between CX and SC NPCs can primarily be explained by its interaction with distinct partner factors, rather than by variations in the degree of chromatin accessibility.

7.3 PAPER III

scRNA-seq and ChIP-seq analyses in glial progenitors reveal extensive SOX9 pre-binding to astrocyte and oligodendrocyte specific genes

The vertebrate CNS is composed of three major cell types, namely neurons, astrocytes and oligodendrocytes that are generated in a temporal sequence by multipotent NPCs during development. In **Paper I**, we have shown that SOX3 activates the expression of NPCs genes and pre-binds many silent neuronal genes that are first activated in differentiated neurons when SOX3 is downregulated and replaced by SOX11. However, due to the lack of specific gene expression profiles for developing astrocytes and oligodendrocytes, it is not clear whether genes exclusively expressed in astrocytes and oligodendrocytes are pre-bound by SOX proteins prior to their activation, and whether these genes are later bound and activated by alternative lineage-specific SOX proteins. Thus, in order to investigate how neuronal and glial gene expression programs are activated in a lineage-specific and timely manner, we have used single cell RNA-sequencing (scRNA-seq) to determine the specific expression profiles of mouse spinal cord progenitor cells and their neuronal and glial progeny. Additionally, we have performed ChIP-seq in order to characterize the genome-wide binding profiles of SOX3 and SOX9 in mESCs-derived glial progenitor cells (GPCs).

We first performed scRNA-seq on NPCs isolated from the mouse embryonic spinal cord at E11.5, prior to the onset of gliogenesis, and on E15.5 cells corresponding to a mixture of progenitors, neurons and glial cells. Using advanced bioinformatics tools, the transcriptome analysis of 350 cells defined 7 distinct cell clusters. The differential expression analysis performed on the different clusters identified gene expression signatures (containing hundreds of differentially expressed genes) specific of each cluster, with the exception of one cluster that was excluded from further analysis due to the low number of differentially expressed genes. Gene ontology analysis of the distinct gene expression signatures resulted in the identification of the 6 remaining cell clusters, which correspond to NPCs, neurons, astrocytes, oligodendrocytes, as well as microglia and mesodermal cells. The gene signature of each cluster includes well-known genes characteristic of each cell type, such as *Map2* and *Stmn2* in

the neuronal cluster, *Fgfbp3* and *Slc1a3* (GLAST) in the astrocyte cluster, and *Mpb* and *Sox10* in the oligodendrocyte cluster, which further confirms their identity.

The analysis of the SOX3 ChIP-seq data in mESC-derived NPCs (from **Paper I**) coupled with the aforementioned specific gene expression profiles, revealed that in addition to neuronal genes, approximately 40% of the astrocyte and of the oligodendrocyte specific genes are already bound by SOX3 in NPCs. These results indicate that in NPCs, a cell population that is competent to give rise to neurons rather than glial cells, SOX3 pre-binds silent genes of all three neural lineages. Considering that SOX3 is expressed in both astrocyte and oligodendrocyte progenitor cells, we determined its binding profile in progenitor cells which are competent to give rise to glial cells. To this aim, we differentiated mESCs to sequentially generate NPCs and GPCs, and we performed a SOX3 ChIP-seq in GPCs. In this *in vitro* context, it is important to note the same pool of GPCs gives rise to both astrocytes and oligodendrocytes, whereas *in vivo,* astrocytes and oligodendrocytes of the developing spinal cord are generated by different populations of progenitor cells residing in distinct progenitor domains.

By comparing the binding patterns of SOX3 in NPCs and GPCs, we found that 30% of SOX3 binding sites in GPCs are already occupied in NPCs, indicating that SOX3 targets many new sites in GPCs. Hence, in a similar way to the changing binding pattern of SOX2 between ESCs and NPCs (**Paper I**), SOX3 also changes binding sites in the transition from NPCs to GPCs, which is likely due to their interaction with changing partner factors. Interestingly, while genes from all lineages are bound by SOX3 in NPCs, astrocyte specific genes are particularly enriched among the SOX3 targets in GPCs.

Next, we examined the epigenetic state of SOX3 bound enhancers in NPCs. To do this, we performed ChIP-seq experiments to identify enhancer regions associated with the H3K27ac mark in NPCs and in GPCs, a histone modification mark associated with active enhancers (Creyghton et al. 2010). The mapping of H3K27ac reads to SOX3 bound regions in NPCs defined two types of SOX3 bound enhancers in NPCs. First, the enhancers carrying the active H3K27ac mark in NPCs and that continue to carry this mark in GPCs. These enhancers are associated to genes that are primarily expressed in astrocytes. Second, the enhancers that are marked with H3K27ac in NPCs but that are no longer marked in GPCs. In contrast, these enhancers are associated to genes that are primarily expressed in NPCs, neurons, and oligodendrocytes. Taken together, these results show that genes of all lineages are associated to H3K27ac marked enhancers that are pre-bound by SOX3 in NPCs. On the other hand, astrocyte genes are associated with enhancers that maintain the H3K27ac mark and are especially enriched among the SOX3 targets in GPCs.

Another SOX protein, SOX9, which is co-expressed with SOX3, starts to be expressed in progenitor cells just before they commit to generate glial cell types, and has been demonstrated to have a key role in the neuron to glia switch (Stolt et al. 2003). SOX9 continues to be expressed in differentiating astrocytes until adulthood, whereas it is transiently expressed in differentiating oligodendrocytes. To better understand the role of SOX9 during gliogenesis, we performed ChIP-seq experiments in GPCs and compared the genome-wide binding profile of SOX9 to that of SOX3 in GPCs. We found that 25% of SOX3 binding sites are also bound by SOX9 in GPCs, and that astrocyte genes are highly enriched among the genes targeted by both

SOX3 and SOX9 in GPCs. Therefore, astrocytes genes are characterized by the continuous binding of SOX3 and the concomitant binding of SOX9 in GPCs. Additionally, we found that although oligodendrocytes genes are already bound by SOX3 in NPCs, SOX3 binding does not appear to be maintained on oligodendrocyte genes in GPCs.

In contrast to astrocyte genes, oligodendrocyte genes are preferentially bound by SOX9 only in GPCs, and not by SOX3. While SOX9 is expressed in both astrocytes and oligodendrocytes, SOX10 is specifically expressed in oligodendrocytes and is essential for their terminal termination (Stolt et al. 2002). In order to better understand how the oligodendrocyte lineage is regulated, we have compared the published SOX10 ChIP-seq data from mature oligodendrocytes (Lopez-Anido et al. 2015) to the SOX9 ChIP-seq data in GPCs. We found that the oligodendrocytes genes that are highly enriched among SOX9 targets in GPCs are later bound by SOX10 in mature oligodendrocytes. Surprisingly, we found that astrocyte genes are also enriched among genes targeted by both SOX9 and SOX10. One possibility is that in addition to regulating oligodendrocyte genes, SOX10 may help keep the expression of astrocyte genes turned off during oligodendrocyte differentiation.

In GPCs, SOX9 pre-binds, together with SOX3, astrocytes genes. SOX9 also pre-binds oligodendrocyte genes, many of which will later be bound by SOX10 in differentiated oligodendrocytes. To better understand how gene target selection is achieved in the glial two lineages, we examined SOX bound DNA regions and looked for enriched DNA motifs. In the DNA regions associated to astrocyte specific genes that are commonly bound by SOX3 and SOX9 in GPCs, we found that the NFI TF binding motif is highly enriched. In addition, transactivation assays revealed that SOX9 and NFIA can synergistically activate putative astrocyte enhancers, which is consistent with the fact that SOX9 has been shown to physically interact with NFIA during astrocyte development (Kang et al. 2012). However, NFIA is already expressed in GPCs together with SOX9, raising the question of why astrocyte genes are first activated in differentiating astrocytes. Interestingly, transactivation assays further revealed that SOX3 blocks the SOX9-mediated activation of putative astrocyte enhancers. Thus, one possibility is that the binding of SOX3 to astrocyte enhancers in GPCs prevents the premature activation of astrocyte genes in GPCs.

In summary, we have shown that SOX3 pre-binds genes of both neuronal and glial lineages in NPCs. SOX3 binding becomes more restricted towards the astrocyte lineage in GPCs, where it appears to prevent the precocious activation of astrocyte genes that are instead activated by SOX9 and NFIA. Conversely, SOX3 does not maintain its binding to oligodendrocyte genes in GPCs. Instead, oligodendrocyte genes in GPCs are preferentially bound by SOX9 alone, and many of these genes are also bound by SOX10 in mature oligodendrocytes. Additionally, SOX10 binds many astrocyte specific genes in oligodendrocytes, possibly to prevent their expression. Collectively, these results indicate that activating or repressing SOX TFs control glial gene expression programs during mouse spinal cord development.

8 CONCLUDING REMARKS AND PERSPECTIVES

The proper development of the CNS relies on the capacity of NPCs to generate neurons, astrocytes and oligodendrocytes. The decision of a NPC to self-renew or differentiate is guided by transcription factors. Many members of the SOX family of transcription factors are well known regulators of CNS development and have been shown to be involved in diverse processes, including NPC maintenance, cell specification and differentiation. Many insights into the roles of SOX TFs have come from functional studies. Owing to the advent of next-generation sequencing approaches, many studies, including ours, have started to unravel the molecular mechanisms by which SOX proteins achieve their distinct functional roles. The work presented in this thesis was aimed to provide a better understanding of the multiple mechanisms employed by SOX TFs to regulate NPCs maintenance and differentiation during embryonic development.

In **Paper I**, we have investigated the roles of SOX2, SOX3 and SOX11 during neuronal lineage progression by characterizing the activities of these distinct SOX proteins on their downstream target genes. In this study, we have provided evidence that different SOX proteins act sequentially to regulate gene expression during neurogenesis. Importantly, our findings add to the growing body of evidence suggesting that SOX proteins exhibit some features of pioneer factors. We have demonstrated that SOX binding precedes gene activation, as SOX2 and SOX3 preselect silent genes that are destined to be activated by an alternative SOX protein at a later stage of neuronal development. Additionally, we have shown that this sequential binding is accompanied by changes in chromatin modifications, and that SOX3 is able to impact the epigenetic landscape. Moreover, the identification of cell type-specific TFs that interact with the distinct SOX proteins is essential to provide a better understanding of the mechanisms governing SOX target site selection. In this study we have identified a number of DNA binding motifs that are differentially enriched within SOX2, SOX3 and SOX11 targeted DNA regions. However, further studies will be necessary to determine the identity of the SOX partner factors binding these motifs during neuronal development. Moreover, it would be interesting to investigate if gene expression in other lineages is also controlled by sequentially pre-binding and activating TFs.

Furthermore, our data raises the question of whether a sequential binding of SOX proteins also occurs in the glial lineage. In **Paper III**, we have addressed this question and found that SOX3 already pre-binds astrocyte and oligodendrocyte genes in NPCs. In GPCs, SOX3 preferentially pre-binds astrocyte genes, possibly to prevent their premature expression. Our data also suggest that SOX9 pre-binds astrocyte specific genes together with the putative partner factor NFIA. SOX9 additionally pre-binds oligodendrocyte specific genes, many of which are bound by SOX10 in mature oligodendrocytes. To further examine the role of NFIA as a potential partner factor, it would be interesting to determine its binding profile in GPCs and compare it to the biding profile of SOX9. Moreover, in this study we have used scRNA-seq in order to characterize the expression profiles of the neuronal and glial cell types in the mouse embryonic spinal cord. The determination of gene sets specific of each cell type can be very valuable for the scientific community. As an example, the identification of specific markers can be used for the generation of conditional knock out animals.

In **Paper II**, we have examined how the interaction between chromatin accessibility and TF binding regulates the establishment of specific gene expression programs in NPCs from different regions of the developing CNS, the cortex and the spinal cord. This study reveals that despite being expressed in all NPCs of the developing CNS, SOX2 regulates the establishment of spatially distinct gene expression programs by interacting with region-specific partner factors on a permissive chromatin landscape. In addition, we have used this data to generate a model that attempts to predict gene expression differences in regionally distinct NPCs. It would be interesting to test the robustness of our statistical model by attempting to predict specific gene expression in other stem cells populations.

The generation of a complex organ such as the CNS requires the generation of the right type of cells at the right place and time. The finding that SOX2 binding differs extensively between NPCs of distinct spatial location raises the question of whether the binding of SOX2 would also differ in NPCs from different developmental stages. One interesting experiment would be to test whether our model would also be able to predict gene expression differences in NPCs over time.

Lastly, in addition to their prominent roles during the formation of the CNS, SOX proteins are essential to the development of most tissues and organs. As a result, mutations in *SOX* genes have been linked with a number of human diseases. For instance, mutations of *SOX2* cause anophtalmia, *SOX3* mutations are associated with X-linked mental retardation, and *SOX10* mutations are implicated in the Waardenburg-Hirschsprung syndrome. Furthermore, it is becoming increasingly clear that cancers can originate from adult stem cell populations, and that TFs important for the maintenance of normal stem cells are likely to play a significant role in tumor initiation and development. Many SOX proteins have indeed been found to be overexpressed in various types of tumors (Dong et al. 2004; Kiefer 2007; Chew & Gallo 2009). This highlights the importance of investigating the role of SOX proteins in relation to development and disease.

9 ACKNOWLEDGEMENTS

This is the end of a long journey, and I would probably need more words than the ones written in this thesis to express my gratitude to all the people who were a part of it. Nevertheless, I would like to thank the following people:

First of all my supervisor, **Jonas**, for giving me the opportunity to do my PhD in your laboratory. Thank you for always having your door open, for your enthusiasm, support and encouragements, especially during the last few months.

Ola, my co-supervisor, for organizing a great course in Toronto!

Maria B, my co-supervisor. I learnt so much working with you, and I am grateful for everything you have done for me, both professionally and personally.

Our collaborators: **Rickard**, **Daniel R**, **Olov**, **Monica**, **Mikael**, **Åsa**, **Johan E** and **Zjanna**, with a special thanks to Daniel R for your patience answering all my questions.

Charlotta, **Birgitta**, **Mats**, **Erika**, **Eliza**, **Soheilla** and **Jorge**. Ludwig would not run without you all!

Matti, for your help throughout the years.

Vilma and **Susie**, my friends, travel companions and partners in crimes. Thank you for the countless great moments we have spent together, from the daily laughter to the memorable weekend trips and vacation. Vilma, thank you for all the good times philosophizing about life, for being my confident and at times my shoulder to cry on. Susie, thank you for being my everyday support in the lab, for putting up sitting next to me in the office all these years, and for having the best playlists to carry us through the week.

The other members of the **Muhr lab**, past and present, for your inputs over the years, for creating a friendly and helpful work atmosphere, and for a fantastic lab retreat in Marrakech! **Gaëlle**, thank you for all the scientific and less scientific discussions, for inventing new words together, and most importantly for your support. **Daniel H**, for good discussions and collaborations. **Idha**, for being so knowledgeable. **Danny**, I learnt a lot working next to you. **Bhumica**, for your positive attitude. **Alex**, for your kindness. **Magnus**, **Alison** and **Claire**, for good company. It has undoubtedly been a great learning experience working with you all.

Thomas, **Johan H**, **Susanne**, **Jan**, **Rickard**, **Qiaolin** and their groups. Thank you everyone for making Ludwig a great place to work at! In particular, I would like to thank the following people. **Hilda**, **Linda**, **André**, **Katarina**, **Fredrick**, **Lina, Stuart**, **Julianna**, **Helena** and **Sissy**, for all the nice chats, laughs, valuable discussions, help and support. **Hugo**, for your friendship and for a crazy Easter weekend! **Konstantinos,** for being one of us. **Eva**, for the awesome company on Fridays at Boomie, which will not be the same without you.

Friends from CMB. My gorgeous **Sanja**, thank you for being such a good friend to me, for your invaluable support over the years, in good and tough times, and for all the laughs. I am so glad that you found happiness outside the brick walls of academia. **Aishe**, my dearest friend. Thank you for always taking such good care of me and for your precious advice and support,

even from across the Atlantic. I miss you! **Vanessa**, I am grateful that we got to share a room in Japan. Thank you for your support, I only wish I could have done the same for you. **Jérémie**, for many fun chats at the fish house.

I would also like to thank the other amazing friends I have made along the way, KI affiliated or not, who one way or another contributed to making these years memorable:

Karina, for your rare kindness. **Haythem**, for being so easy to talk you. **Mustafa**, for always being so positive. **Alessandro**, for your unique sense of humor. **Andrea**, for being the most genuine person I know. **Steffi**, for good times when you were in Stockholm. **Tanya**, for all the great conversations we have had. **Maria H**, for a great week in Toronto. **Vicky**, for nice chats and company. I am thankful for the many years we have spent together. Even though some of you are already off to new adventures, I hope that we will stay in touch.

All the wonderful people who made my time at Jägargatan so special, in particular: **Violaine**, for lots of fun times together. **Léa**, or should I say Léaaaaa, for your friendship and for being wise beyond your years. Mi amiga **Clarissa**, for being the strong and smart woman that you are. I cannot wait to see all the great things you will accomplish.

The CHaSE team: **Tiago**, **Su**, **Moritz**, **Michael**, **Lizan**, **Francesca**, **Thibaud, Florian**, **Emelie**, **Christian** and **Daniel** for a great experience together!

My friends from KTH: **Pierre** and **Matthieu** for many fun memories together and for being able to make me laugh through the saddest night. **Manu** and **Fred** for many good times in Stockholm.

All the French expats for the great parties, crazy boat trips and Midsommar together, with special thanks to **Guillaume**, **Cécile**, **JP** and **Arnaud**.

Je tiens aussi à remercier mes amis qui malgré la distance et les années qui passent ont toujours pris le temps de me voir lors de mes courts passages à Paris. A mes amies d'Orsay: **Babou**, **Vivi** et **Béné**, pour nos réunions annuelles et pour le weekend de mes 30 ans. A **Raph**, **Oliv'** et **l'Ancien**, pour les soirées inoubliables à Massy. A **Manu**, mon coloc préféré, et **Cyrielle**, ma Fillotte, pour tous les bons moments passés ensemble rue des Primevères. A **Joëlle** et **Lena**, pour avoir bravé le froid Suédois, pour votre soutien dans les moments difficiles, et surtout pour ces 20 ans d'amitié.

Enfin, un grand merci à ma petite famille, avec une pensée toute particulière pour **Papy** et **Mamie Jeanne**. A **Yoann** et **Blandine**, ainsi qu'à ma nièce et mon neveu adorés, **Joséphine** et **Charles.** A **Papa** et **Maman**, pour m'avoir toujours soutenue dans mon choix de partir vivre en Suède, pour prendre si bien soin de moi lors de mes retours à la maison, pour vos encouragements et votre patience tout au long de ces longues années, ainsi que pour tous vos sacrifices, pour lesquels je ne vous remercierai jamais assez. Et enfin, à ma **Mamie** que j'aime tant, c'est à toi que je dédie cette thèse.

10 REFERENCES

Alberts, B. et al., 2008. *Molecular Biology of the Cell* 5th ed., Garland Science.

Ambrosetti, D.C., Basilico, C. & Dailey, L., 1997. Synergistic activation of the fibroblast growth factor 4 enhancer by Sox2 and Oct-3 depends on protein-protein interactions facilitated by a specific spatial arrangement of factor binding sites. *Molecular and cellular biology*, 17(11), pp.6321–9.

Avilion, A.A. et al., 2003. Multipotent cell lineages in early mouse development on SOX2 function. *Genes & Development*, 17, pp.126–140.

Benayoun, B.A., Pollina, E.A. & Brunet, A., 2015. Epigenetic regulation of ageing: linking environmental inputs to genomic stability. *Nature reviews. Molecular cell biology*, 16(10), pp.593–610.

Bergsland, M. et al., 2011. Sequentially acting Sox transcription factors in neural lineage development. *Genes & Development*, 25(23), pp.2453–2464.

Bergsland, M. et al., 2006. The establishment of neuronal properties is controlled by Sox4 and Sox11. *Genes & Development*, 20(24), pp.3475–3486.

Bertrand, N., Castro, D.S. & Guillemot, F., 2002. Proneural genes and the specification of neural cell types. *Nature reviews. Neuroscience*, 3(7), pp.517–30.

Bhattaram, P. et al., 2010. Organogenesis relies on SoxC transcription factors for the survival of neural and mesenchymal progenitors. *Nature communications*, 1, p.9.

Boer, B. et al., 2007. Elevating the levels of Sox2 in embryonal carcinoma cells and embryonic stem cells inhibits the expression of Sox2:Oct-3/4 target genes. *Nucleic Acids Research*, 35(6), pp.1773–1786.

Bowles, J., Schepers, G. & Koopman, P., 2000. Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. *Developmental biology*, 227(2), pp.239–255.

Boyer, L.A. et al., 2006. Polycomb complexes repress developmental regulators in murine embryonic stem cells. *Nature*, 441(7091), pp.349–353.

Boyer, L. a L. a. et al., 2005. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*, 122(6), pp.947–956.

Boyle, A.P. et al., 2008. High-Resolution Mapping and Characterization of Open Chromatin across the Genome. *Cell*, 132(2), pp.311–322.

Bylund, M. et al., 2003. Vertebrate neurogenesis is counteracted by Sox1-3 activity. *Nature Neuroscience*, 6(11), pp.1162–1168.

Caiazzo, M. et al., 2015. Direct conversion of fibroblasts into functional astrocytes by defined transcription factors. *Stem Cell Reports*, 4(1), pp.25–36.

Chambers, I. et al., 2003. Functional expression cloning of Nanog, a pluripotency sustaining factor in embryonic stem cells. *Cell*, 113(5), pp.643–655.

Chen, J. et al., 2014. Single-molecule dynamics of enhanceosome assembly in embryonic stem cells. *Cell*, 156(6), pp.1274–1285.

Chen, X. et al., 2008. Integration of External Signaling Pathways with the Core Transcriptional Network in Embryonic Stem Cells. *Cell*, 133(6), pp.1106–1117.

Chew, L.J. & Gallo, V., 2009. The Yin and yang of Sox proteins: activation and repression in development and disease. *Journal of Neuroscience Research*, 87, pp.3277–3287.

Cirillo, L.A. et al., 2002. Opening of compacted chromatin by early developmental transcription factors HNF3 (FoxA) and GATA-4. *Molecular Cell*, 9(2), pp.279–289.

Collignon, J. et al., 1996. A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. *Development*, 122, pp.509–520.

Creyghton, M.P. et al., 2010. Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proceedings of the National Academy of Sciences of the United States of America*, 107(50), pp.21931–21936.

Czechanski, A. et al., 2014. Derivation and characterization of mouse embryonic stem cells from permissive and nonpermissive strains. *Nature protocols*, 9(3), pp.559–74.

Deneen, B. et al., 2006. The Transcription Factor NFIA Controls the Onset of Gliogenesis in the Developing Spinal Cord. *Neuron*, 52(6), pp.953–968.

Denny, P. et al., 1992. A conserved family of genes related to the testis determining gene, SRY. *Nucleic Acids Research*, 20(11), p.2887.

Diez del Corral, R. & Storey, K.G., 2001. Markers in vertebrate neurogenesis. *Nature reviews. Neuroscience*, 2(November), pp.835–839.

Dong, C., Wilhelm, D. & Koopman, P., 2004. Sox genes and cancer. *Cytogenetic and Genome Research*, 105(2–4), pp.442–447.

Ekonomou, A. et al., 2005. Neuronal migration and ventral subtype identity in the telencephalon depend on SOX1. *PLoS Biology*, 3(6), pp.1111–1122.

Ellis, P. et al., 2004. SOX2, a persistent marker for multipotential neural stem cells derived from embryonic stem cells, the embryo or the adult. *Developmental Neuroscience*, 26(2–4), pp.148–165.

Engelen, E. et al., 2011. Sox2 cooperates with Chd7 to regulate genes that are mutated in human syndromes. *Nature Genetics*, 43(6), pp.607–611.

Evans, M.J. & Kaufman, M.H., 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature*, 292, pp.154–156.

Favaro, R. et al., 2009. Hippocampal development and neural stem cell maintenance require Sox2-dependent regulation of Shh. *Nature neuroscience*, 12(10), pp.1248–1256.

Ferrari, S. et al., 1992. SRY, like HMG1, recognizes sharp angles in DNA. *EMBO Journal*, 11(12), pp.4497–4506.

Ferri, A.L. et al., 2004. Sox2 deficiency causes neurodegeneration and impaired neurogenesis in the adult mouse brain. *Development*, 131(15), pp.3805–3819.

Finzsch, M. et al., 2008. Sox9 and Sox10 influence survival and migration of oligodendrocyte precursors in the spinal cord by regulating PDGF receptor alpha expression. *Development*, 135(4), pp.637–646.

Galas, D.J. & Schmitz, a, 1978. DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic acids research*, 5(9), pp.3157–3170.

Giese, K., Cox, J. & Grosschedl, R., 1992. The HMG domain of lymphoid enhancer factor 1 bends DNA and facilitates assembly of functional nucleoprotein structures. *Cell*, 69(1), pp.185–195.

Gilbert, N. & Ramsahoye, B., 2005. The relationship between chromatin structure and transcriptional activity in mammalian genomes. *Brief Funct Genomic Proteomic*, 4(2), pp.129–142.

Graham, V. et al., 2003. SOX2 functions to maintain neural progenitor identity. *Neuron*, 39(5), pp.749–765.

Gubbay, J. et al., 1990. A gene mapping to the sex-determining region of the mouse Y chromosome is a member of a novel family of embryonically expressed genes. *Nature*, 346(7), pp.245–50.

Guth, S.I.E. & Wegner, M., 2008. Having it both ways: Sox protein function between conservation and innovation. *Cellular and Molecular Life Sciences*, 65(19), pp.3000–3018.

Hagey, D.W. & Muhr, J., 2014. Sox2 Acts in a Dose-Dependent Fashion to Regulate Proliferation of Cortical Progenitors. *Cell Reports*, 9(5), pp.1908–1921.

Han, D.W. et al., 2012. Direct reprogramming of fibroblasts into neural stem cells by defined factors. *Cell Stem Cell*, 10(4), pp.465–472.

Harley, V.R., Lovell-badge, R. & Goodfellow, P.N., 1994. Definition of a consensus DNA binding site for SRY. *Nucleic Acids Research*, 22(8), pp.1500–1501.

Haslinger, A. et al., 2009. Expression of Sox11 in adult neurogenic niches suggests a stage-specific role in adult neurogenesis. *European Journal of Neuroscience*, 29(11), pp.2103–2114.

Heinz, S. et al., 2010. Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. *Molecular Cell*, 38(4), pp.576–589.

Holmberg, J. et al., 2008. SoxB1 transcription factors and Notch signaling use distinct mechanisms to regulate proneural gene function and neural progenitor differentiation. *Development*, 135(10), pp.1843–1851.

Hoser, M. et al., 2008. Sox12 deletion in the mouse reveals nonreciprocal redundancy with the related Sox4 and Sox11 transcription factors. *Molecular and cellular biology*, 28(15), pp.4675–4687.

Huangfu, D. et al., 2008. Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2. *Nature biotechnology*, 26(11), pp.1269–1275.

Kamachi, Y. et al., 1998. Involvement of Sox1, 2 and 3 in the early and subsequent molecular events of lens induction. *Development*, 125(13), pp.2521–2532.

Kamachi, Y. et al., 1995. Involvement of SOX proteins in lens-specific activation of crystallin genes. *The EMBO journal*, 14(14), pp.3510–9.

Kamachi, Y. et al., 2001. Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. *Genes & Development*, 15(10), pp.1272–1286.

Kamachi, Y. & Kondoh, H., 2013. Sox proteins: regulators of cell fate specification and differentiation. *Development*, 140(20), pp.4129–4144.

Kamachi, Y., Uchikawa, M. & Kondoh, H., 2000. Pairing SOX off: With partners in the regulation of embryonic development. *Trends in Genetics*, 16(4), pp.182–187.

Kang, P. et al., 2012. Sox9 and NFIA Coordinate a Transcriptional Regulatory Cascade during the Initiation of Gliogenesis. *Neuron*, 74(1), pp.79–94.

Kellerer, S. et al., 2006. Replacement of the Sox10 transcription factor by Sox8 reveals incomplete functional equivalence. *Development*, 133(15), pp.2875–86.

Keramari, M. et al., 2010. Sox2 is essential for formation of trophectoderm in the preimplantation embryo. *PLoS ONE*, 5(11).

Kessaris, N., Pringle, N. & Richardson, W.D., 2008. Specification of CNS glia from neural stem cells in the embryonic neuroepithelium. *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 363(1489), pp.71–85.

Kessaris, N., Pringle, N. & Richardson, W.D., 2001. Ventral neurogenesis and the neuron-glial switch. *Neuron*, 31(5), pp.677–680.

Kiefer, J.C., 2007. Back to basics: Sox genes. *Developmental Dynamics*, 236(8), pp.2356–2361.

Kim, J. et al., 2008. An Extended Transcriptional Network for Pluripotency of Embryonic Stem Cells. *Cell*, 132(6), pp.1049–1061.

Kim, J.B. et al., 2008. Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. *Nature*, 454(7204), pp.646–50.

Kintner, C., 2002. Neurogenesis in Embryos and in Adult Neural Stem Cells. *The Journal of Neuroscience*, 22(3), pp.639–643.

Kishi, M. et al., 2000. Requirement of Sox2-mediated signaling for differentiation of early Xenopus neuroectoderm. *Development*, 127(4), pp.791–800.

Kondoh, H. & Kamachi, Y., 2010. SOX-partner code for cell specification: Regulatory target selection and underlying molecular mechanisms. *International Journal of Biochemistry and Cell Biology*, 42(3), pp.391–399.

Kopp, J.L. et al., 2008. Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells. *Stem Cells*, 26(4), pp.903–911.

Kornberg, R.D., 1977. Structure of chromatin. *Annual review of biochemistry*, 46, pp.931–54.

Kouzarides, T., 2007. Chromatin Modifications and Their Function. *Cell*, 128(4), pp.693–705.

Lara-Astiaso, D. et al., 2014. Chromatin state dynamics during blood formation. *Science*, 345(6199), pp.943–9.

Lefebvre, V. et al., 2007. Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. *International Journal of Biochemistry and Cell Biology*, 39(12), pp.2195–2214.

Lefebvre, V. et al., 1997. SOX9 is a potent activator of the chondrocyte-specific enhancer of the pro alpha1(II) collagen gene. *Molecular and cellular biology*, 17(4), pp.2336–46.

Lefebvre, V., Li, P. & De Crombrugghe, B., 1998. A new long form of Sox5 (L-Sox5), Sox6 and Sox9 are coexpressed in chondrogenesis and cooperatively activate the type II collagen gene. *EMBO Journal*, 17(19), pp.5718–5733.

Liber, D. et al., 2010. Epigenetic priming of a Pre-B Cell-Specific enhancer through binding of Sox2 and Foxd3 at the ESC stage. *Cell Stem Cell*, 7(1), pp.114–126.

Lodato, M.A. et al., 2013. SOX2 Co-Occupies Distal Enhancer Elements with Distinct POU Factors in ESCs and NPCs to Specify Cell State. , 9(2).

Loh, K.M. & Lim, B., 2011. A precarious balance: Pluripotency factors as lineage specifiers. *Cell Stem Cell*, 8(4), pp.363–369.

Lopez-Anido, C. et al., 2015. Differential Sox10 Genomic Occupancy in Myelinating Glia Camila. *Glia*, 63(11), pp.1897–1914.

Magnani, L., Eeckhoute, J. & Lupien, M., 2011. Pioneer factors: Directing transcriptional regulators within the chromatin environment. *Trends in Genetics*, 27(11), pp.465–474.

Malas, S. et al., 2003. Sox1-deficient mice suffer from epilepsy associated with abnormal ventral forebrain development and olfactory cortex hyperexcitability. *Neuroscience*, 119(2), pp.421–432.

Martin, G.R., 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A*, 78(12), pp.7634–7638.

Masui, S. et al., 2007. Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. *Nature Cell Biology*, 9(6), pp.625-U26.

Matsuda, S. et al., 2012. Sox21 Promotes Hippocampal Adult Neurogenesis via the Transcriptional Repression of the Hes5 Gene. *Journal of Neuroscience*, 32(36), pp.12543–12557.

Mertin, S., McDowall, S.G. & Harley, V.R., 1999. The DNA-binding specificity of SOX9 and other SOX proteins. *Nucleic Acids Research*, 27(5), pp.1359–1364.

Mihajlović, A.I., Thamodaran, V. & Bruce, A.W., 2015. The first two cell-fate decisions of preimplantation mouse embryo development are not functionally independent. *Nature Publishing Group*, (April), pp.1–16.

Mitsui, K. et al., 2003. The Homeoprotein Nanog Is Required for Maintenance of Pluripotency in Mouse Epiblast and ES Cells. *Cell*, 113, pp.631–642.

Miyagi, S. et al., 2008. Consequence of the loss of Sox2 in the developing brain of the mouse. *FEBS Letters*, 582(18), pp.2811–2815.

Molofsky, A.V. & Deneen, B., 2015. Astrocyte development: A Guide for the Perplexed. *Glia*, 63(8), pp.1320–1329.

Mu, L. et al., 2012. SoxC Transcription Factors Are Required for Neuronal Differentiation in Adult Hippocampal Neurogenesis. *Journal of Neuroscience*, 32(9), pp.3067–3080.

Najm, F.J. et al., 2013. Transcription factor–mediated reprogramming of fibroblasts to expandable, myelinogenic oligodendrocyte progenitor cells. *Nature Biotechnology*, 31(5), pp.426–433.

Nakada, Y. et al., 2004. Distinct domains within Mash1 and Math1 are required for function in neuronal differentiation versus neuronal cell-type specification. *Development*, 131(6), pp.1319–1330.

Nakagawa, M. et al., 2008. Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. *Nature biotechnology*, 26(1), pp.101–6.

Natarajan, A. et al., 2012. Predicting cell-type – specific gene expression from regions of open chromatin the genome. *Genome Research*, pp.1711–1722.

Nichols, J. et al., 1998. Formation of pluripotent stem cells in the mammalian embryo dependes on the POU transcription factor Oct4. *Cell*, 95(3), pp.379–391.

Nishiguchi, S. et al., 1998. Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. *Genes & Development*, 12(6), pp.776–81.

Palasingam, P. et al., 2009. The Structure of Sox17 Bound to DNA Reveals a Conserved Bending Topology but Selective Protein Interaction Platforms. *Journal of Molecular Biology*, 388(3), pp.619–630.

Pevny, L.H. & Lovell-Badge, R., 1997. Sox genes find their feet. *Current Opinion in Genetics and Development*, 7(3), pp.338–344.

Plusa, B. & Hadjantonakis, A.-K., 2014. Embryonic stem cell identity grounded in the embryo. *Nature cell biology*, 16(6), pp.502–4.

Pontiggia, A. et al., 1994. Sex-reversing mutations affect the architecture of SRY-DNA complexes. *EMBO journal*, 13(24), pp.6115–6124.

Raposo, A.A.S.F. et al., 2015. Ascl1 Coordinately Regulates Gene Expression and the Chromatin Landscape during Neurogenesis. *Cell Reports*, 10(9), pp.1544–1556.

Reiprich, S. & Wegner, M., 2015. From CNS stem cells to neurons and glia: Sox for everyone. *Cell and Tissue Research*, 359(1), pp.111–124.

Rex, M. et al., 1997. Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue. *Developmental Dynamics*, 209(3), pp.323–332.

Richardson, W.D., Kessaris, N. & Pringle, N., 2006. Oligodendrocyte wars. *Nature reviews. Neuroscience*, 7(1), pp.11–18.

Ring, K.L. et al., 2012. Direct reprogramming of mouse and human fibroblasts into multipotent neural stem cells with a single factor. *Cell Stem Cell*, 11(1), pp.100–109.

Rizzoti, K. et al., 2004. SOX3 is required during the formation of the hypothalamo-pituitary axis. *Nature genetics*, 36(3), pp.247–55.

Roose, J. et al., 1998. High expression of the HMG box factor Sox-13 in arterial walls during embryonic development. *Nucleic Acids Research*, 26(2), pp.469–476.

Rowitch, D.H., 2004. Glial specification in the vertebrate neural tube. *Nature reviews. Neuroscience*, 5(5), pp.409–419.

Sandberg, M., Kallstrom, M. & Muhr, J., 2005. Sox21 promotes the progression of vertebrate neurogenesis. *Nature Neuroscience*, 8(8), pp.995–1001.

Sanes, D.H., Reh, T.A. & Harris, W.A., 2011. Neural induction. In *Development of the Nervous System*. Academic Press, pp. 1–22.

Scaffidi, P. & Bianchi, M.E., 2001. Spatially Precise DNA Bending Is an Essential Activity of the Sox2 Transcription Factor. *Journal of Biological Chemistry*, 276(50), pp.47296–47302.

Scharer, C.D. et al., 2009. Genome-wide promoter analysis of the SOX4 transcriptional network in prostate cancer cells. *Cancer Research*, 69(2), pp.709–717.

Schilham, M.W. et al., 1996. Defects in cardiac outflow tract formation and pro-B-lymphocyte expansion in mice lacking Sox-4. *Nature*, 380(6576), pp.711–4.

Scott, C.E. et al., 2010. SOX9 induces and maintains neural stem cells. *Nature neuroscience*, 13(10), pp.1181–1189.

Sérandour, A.A. et al., 2011. Epigenetic switch involved in activation of pioneer factor FOXA1-dependent enhancers. *Genome Research*, 21(4), pp.555–565.

Shanmugam, K. et al., 1999. PBX and MEIS as non-DNA-binding partners in trimeric complexes with HOX proteins. *Molecular and Cellular Biology*, 19(11), pp.7577–7588.

Sinclair, a H. et al., 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature*, 346(6281), pp.240–244.

Sock, E. et al., 2004. Gene Targeting Reveals a Widespread Role for the High-Mobility-Group Transcription Factor Sox11 in Tissue Remodeling Gene Targeting Reveals a Widespread Role for the High-Mobility-Group Transcription Factor Sox11 in Tissue Remodeling. *Molecular Cell Biology*, 24(15), pp.6635–6644.

Sock, E. et al., 2001. Idiopathic Weight Reduction in Mice Deficient in the High-Mobility-Group Transcription Factor Sox8 Idiopathic Weight Reduction in Mice Deficient in the High-Mobility-Group Transcription Factor Sox8. *Molecular Cell Biology*, 21(20), pp.6951–6959.

Soriano, N.S. & Russell, S., 1998. The Drosophila SOX-domain protein Dichaete is required for the development of the central nervous system midline. *Development*, 125(20), pp.3989–96.

Soufi, A., Donahue, G. & Zaret, K.S., 2012. Facilitators and impediments of the pluripotency reprogramming factors' initial engagement with the genome. *Cell*, 151(5), pp.994–1004.

Soullier, S. et al., 1999. Diversification pattern of the HMG and SOX family members during evolution. *Journal of Molecular Evolution*, 48(5), pp.517–527.

Stergachis, A.B. et al., 2013. Developmental Fate and Cellular Maturity Encoded in Human Regulatory DNA Landscapes. *Cell*, 154(4), pp.888–903.

Stolt, C.C. et al., 2005. Impact of transcription factor Sox8 on oligodendrocyte specification in the mouse embryonic spinal cord. *Developmental Biology*, 281(2), pp.309–317.

Stolt, C.C. et al., 2006. SoxD Proteins Influence Multiple Stages of Oligodendrocyte Development and Modulate SoxE Protein Function. *Developmental Cell*, 11(5), pp.697–709.

Stolt, C.C. et al., 2002. Terminal differentiation of myelin-forming oligodendrocytes depends on the transcription factor Sox10. *Genes & Development*, pp.165–170.

Stolt, C.C. et al., 2003. The Sox9 transcription factor determines glial fate choice in the developing spinal cord. *Genes & Development*, 17(13), pp.1677–1689.

Stolt, C.C. et al., 2004. Transcription factors Sox8 and Sox10 perform non-equivalent roles during oligodendrocyte development despite functional redundancy. *Development*, 131(10), pp.2349–2358.

Südbeck, P. et al., 1996. Sex reversal by loss of the C-terminal transactivation domain of human SOX9. *Nature genetics*, 13(2), pp.230–2.

Takahashi, K. et al., 2007. Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell*, 131(5), pp.861–872.

Takahashi, K. & Yamanaka, S., 2006. Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors. *Cell*, 126(4), pp.663–676.

Thein, D.C. et al., 2010. The closely related transcription factors Sox4 and Sox11 function as survival factors during spinal cord development. *Journal of Neurochemistry*, 115(1), pp.131–141.

Thomson, J.A. et al., 1998. Embryonic stem cell lines derived from human blastocysts. *Science*, 282(5391), pp.1145–7.

Thomson, M. et al., 2011. Pluripotency factors in embryonic stem cells regulate differentiation into germ layers. *Cell*, 145(6), pp.875–889.

Uchikawa, M., Kamachi, Y. & Kondoh, H., 1999. Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: Their expression during embryonic organogenesis of the chicken. *Mechanisms of Development*, 84(1–2), pp.103–120.

Venkatesh, S. & Workman, J.L., 2015. Histone exchange, chromatin structure and the regulation of transcription. *Nature Reviews. Molecular Cell Biology*, 16(3), pp.178–189.

Wang, G.G., Allis, C.D. & Chi1, P., 2007. Chromatin remodeling and cancer, part II: ATP-dependent chromatin remodeling. *Trends in Molecular Medicine*, 13(9), pp.373–380.

Wang, T.W. et al., 2006. Sox3 expression identifies neural progenitors in persistent neonatal and adult mouse forebrain germinative zones. *Journal of Comparative Neurology*, 497(1), pp.88–100.

Wegner, M., 2010. All purpose Sox: The many roles of Sox proteins in gene expression. *International Journal of Biochemistry and Cell Biology*, 42(3), pp.381–390.

Wegner, M., 1999. From head to toes: The multiple facets of Sox proteins. *Nucleic Acids Research*, 27(6), pp.1409–1420.

Weiss, M.A., 2001. Floppy SOX: mutual induced fit in hmg (high-mobility group) box-DNA recognition. *Molecular endocrinology (Baltimore, Md)*, 15(3), pp.353–362.

Wennekamp, S. et al., 2013. A self-organization framework for symmetry breaking in the mammalian embryo. *Nature Publishing Group*, 14(7), pp.452–459.

Werner, M.H. et al., 1995. Molecular basis of human 46X,Y sex reversal revealed from the three-dimensional solution structure of the human SRY-DNA complex. *Cell*, 81(5), pp.705–714.

Wilson, M. & Koopman, P., 2002. Matching SOX: Partner proteins and co-factors of the SOX family of transcriptional regulators. *Current Opinion in Genetics and Development*, 12(4), pp.441–446.

Wolpert, L., 2002. *Principles of Development* 2nd ed., Oxford University Press.

Wood, H.B. & Episkopou, V., 1999. Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. *Mechanisms of Development*, 86(1–2), pp.197–201.

Wright, E.M., Snopek, B. & Koopman, P., 1993. Seven new members of the Sox gene during mouse development family expressed during mouse development. *Nuclei Acids Research*, 21(3), p.744.

Xu, J. et al., 2009. Transcriptional competence and the active marking of tissue-specific enhancers by defined transcription factors in embryonic and induced pluripotent stem cells. *Genes & Development*, 23(24), pp.2824–2838.

Yang, N. et al., 2013. Generation of oligodendroglial cells by direct lineage conversion. *Nat Biotechnol*, 31(5), pp.434–439.

Yuan, H. et al., 1995. Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. *Genes & Development*, 9(21), pp.2635–2645.

Zaret, K.S. & Carroll, J.S., 2011. Pioneer transcription factors: Establishing competence for gene expression. *Genes & Development*, 25(21), pp.2227–2241.

Zhang, S. & Cui, W., 2014. Sox2, a key factor in the regulation of pluripotency and neural differentiation. *World journal of stem cells*, 6(3), pp.305–11.

Zhao, S. et al., 2004. SoxB transcription factors specify neuroectodermal lineage choice in ES cells. *Molecular and Cellular Neuroscience*, 27(3), pp.332–342.

Zhou, Q. & Anderson, D.J., 2002. The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification. *Cell*, 109(1), pp.61–73.