Chromatin Looping and Cell Type-Specific Regulation of Sox2 Expression

By

Harry Yu Zhou

A thesis submitted in conformity with the requirements for the degree of Master of Science
Department of Cell and Systems Biology
University of Toronto

© Copyright by Harry Zhou 2013
Chromatin Looping and Cell Type-Specific Regulation of Sox2 Expression

Harry Zhou

Master of Science

Department of Cell and Systems Biology

University of Toronto

2013

Abstract

The Sox2 gene encodes a HMG box transcription factor that plays a critical role in maintaining the self-renewal of embryonic stem (ES) cells and regulating neuroectodermal differentiation. Although Sox2 is expressed in ES cells and embryonic brain, cell-type specific regulation of Sox2 transcription remains unclear. In this thesis, I found that the chromatin architecture surrounding the Sox2 locus is highly cell type specific. The Sox2 promoter and proximal enhancers interacted with a cluster of putative enhancer candidates located 80-120 kilobases (kb) downstream to Sox2 in ES cells but not in embryonic brain, whereas in embryonic brain the Sox2 proximal region adopted a more compact chromatin conformation than ES cells. Also, the distal enhancer cluster showed higher enhancer activities than the Sox2 proximal enhancers in ES cells. Together, these results suggest that the novel enhancer cluster downstream to Sox2 plays an important role in ES cell specific regulation of Sox2 transcription.
Acknowledgements

First and foremost, I would like to thank my supervisor, Dr. Jennifer Mitchell, for allowing me to conduct a Master’s thesis in her lab. The past two years have been very challenging yet extremely fruitful and rewarding for me. I would not have been able to successfully complete the thesis without Dr. Mitchell’s guidance and expertise. I would like to thank her also for giving me the opportunities to present my results at two international conferences: the 2012 Cold Spring Harbor Laboratory Meeting on the Dynamic Organization of Nuclear Function and the 2013 Annual Meeting for International Society for Stem Cell Research. These meetings greatly broadened my understanding of my research fields and enriched my graduate school experience.

I would like to thank my supervisory committee members, Dr. Janet Rossant and Dr. Ashley Bruce, for their inputs on my project and their kind support in my career development.

I would like to thank Felicia Collura, an undergraduate student in the lab, for her tremendous help with cloning the Sox2 enhancer candidates into the luciferase vector. I could not have made all of the constructs used in this thesis without her dedication and hard work.

I would like to thank Dr. Vince Tropepe for teaching me the techniques for dissecting embryonic brain tissues. I would like to thank Daniel Rivero at the BioScience Support Facilities for providing me with the experimental mice. And I would like to thank Malgosia Kownacka at the SLRI ES Cell Facility for supplying me with the ES culture reagents.

I am very grateful for everyone in Mitchell lab, who have taught me many different lab techniques and made my time in the lab very enjoyable. I want to thank Anandi Bhattacharya for patiently teaching me the nitigrities of the 3C, which was a very challenging technique to do. I want to thank Mike Schwartz for showing me how to culture mouse ES cells and for his helpful advice regarding graduate studies.

Last but not least, I would like to thank my family and close friends for their continued support and encouragement no matter what situation I found myself in during the past two years. It has truly been a transformative process for me.
Table of Content

List of Abbreviations ........................................................................................................ vi
List of Figures .................................................................................................................... vii
List of Appendices ............................................................................................................ viii
Chapter 1 Introduction ...................................................................................................... 1
  1.1 Overview of eukaryotic transcription ........................................................................ 2
  1.1.1 Players in transcriptional regulation ............................................................... 2
  1.1.2 The role of enhancers in transcriptional regulation ........................................ 5
  1.1.3 Proteins that mediate enhancer-promoter communication ............................. 6
  1.1.3 Chromatin features of enhancers ..................................................................... 8
  1.2 Transcription factor SOX2 .................................................................................... 10
  1.2.1 Sox2 expression pattern .................................................................................. 11
  1.2.2 The role of Sox2 in pluripotent stem cells ..................................................... 11
  1.2.3 The role of Sox2 in neural development ....................................................... 13
  1.2.4 Transcriptional regulation of Sox2 ................................................................. 14
  1.2.5 Novel predicted enhancer candidates near Sox2 locus ................................. 15
  1.3 Objective ................................................................................................................. 18
Chapter 2 Methods ........................................................................................................... 19
  2.1 Cell culture ............................................................................................................... 19
  2.2 Chromosome Conformation Capture (3C) .......................................................... 19
  2.3 Luciferase Reporter Constructs ............................................................................ 25
  2.4 Dual Luciferase Reporter Assay ............................................................................ 27
Chapter 3 Results .............................................................................................................. 28
  3.1 Long-range chromatin looping at the Sox2 locus ................................................ 28
  3.2 Enhancer activity of novel enhancer candidates .................................................. 34
Chapter 4 Discussion ......................................................................................................... 37
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>4.1</td>
<td>Summary of Results</td>
<td>37</td>
</tr>
<tr>
<td>4.2</td>
<td>ES cell specific chromatin architecture at the Sox2 locus</td>
<td>38</td>
</tr>
<tr>
<td>4.3</td>
<td>Factors that mediate chromatin looping at the Sox2 locus</td>
<td>40</td>
</tr>
<tr>
<td>4.4</td>
<td>ES cell-specific regulation of Sox2 expression</td>
<td>44</td>
</tr>
<tr>
<td>4.5</td>
<td>Accuracy of <em>in silico</em> prediction of active enhancers</td>
<td>46</td>
</tr>
<tr>
<td>4.6</td>
<td>Future directions</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td><strong>Chapter 5 Summary</strong></td>
<td>50</td>
</tr>
<tr>
<td></td>
<td><strong>References</strong></td>
<td>51</td>
</tr>
<tr>
<td></td>
<td><strong>Appendices</strong></td>
<td>64</td>
</tr>
</tbody>
</table>
# List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>3C</td>
<td>Chromosome conformation capture</td>
</tr>
<tr>
<td>5C</td>
<td>Chromosome conformation capture carbon copy</td>
</tr>
<tr>
<td>bp</td>
<td>Base pair</td>
</tr>
<tr>
<td>BMP</td>
<td>Bone morphogenetic protein</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>Chromatin immunoprecipitation coupled with high throughput sequencing</td>
</tr>
<tr>
<td>CTD</td>
<td>C-terminal domain</td>
</tr>
<tr>
<td>E11.5</td>
<td>Embryonic day 11.5</td>
</tr>
<tr>
<td>Enh</td>
<td>Putative enhancer</td>
</tr>
<tr>
<td>ES cell</td>
<td>Embryonic stem cell</td>
</tr>
<tr>
<td>GTF</td>
<td>General transcription factor</td>
</tr>
<tr>
<td>kb</td>
<td>Kilobases</td>
</tr>
<tr>
<td>LCR</td>
<td>Locus control region</td>
</tr>
<tr>
<td>LIF</td>
<td>Leukemia inhibitory factor</td>
</tr>
<tr>
<td>MEFs</td>
<td>Mouse embryonic fibroblasts</td>
</tr>
<tr>
<td>NPCs</td>
<td>Neural precursor cells</td>
</tr>
<tr>
<td>O/S/N</td>
<td>OCT4, SOX2, NANOG co-bound sites</td>
</tr>
<tr>
<td>PIC</td>
<td>Pre-initiation complex</td>
</tr>
<tr>
<td>RNAPII</td>
<td>RNA polymerase II</td>
</tr>
<tr>
<td>SRR</td>
<td>Sox regulatory region</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription factor</td>
</tr>
<tr>
<td>TSS</td>
<td>Transcription start site</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1: Novel enhancer candidates surrounding the Sox2 locus are bound by ES cell-specific TFs ......................16

Figure 2: Cell-type specific patterns of H3K4me1 and p300 binding surrounding the Sox2 locus .............................17

Figure 3: qPCR primers for 3C quantification and digestion efficiency check .........................................................24

Figure 4: Cloning enhancer candidates into luciferase reporter plasmid .................................................................26

Figure 5: HindIII digestion efficiency is uniform across restriction sites and cell types ........................................29

Figure 6: Chromatin interaction profile of the Sox2 locus with Sox2 promoter as the anchor .................................32

Figure 7: Chromatin interaction profile of the Sox2 locus with the distal enhancer cluster as the anchor ...............33

Figure 8: Luciferase activity of enhancer candidates surrounding the Sox2 locus in mouse ES cells and MEFs ........36

Figure 9: Cell-type specific chromatin architecture surrounding the Sox2 locus ..................................................42

Figure 10: Model for transcriptional regulation of Sox2 in mouse ES cells and neural precursor cells ..................43
List of Appendices

Appendix 1: Schematic for chromosome conformation capture (3C) technology .............................................................. 64
Appendix 2: 3C qPCR primers for murine Sox2 locus ........................................................................................................... 65
Appendix 3: Optimization of fixation condition for 3C analysis .......................................................................................... 66
Appendix 4: Gel-electrophoresis of qPCR products for Sox2 promoter 3C profile .............................................................. 67
Appendix 5: Gel-electrophoresis of qPCR products for distal enhancer cluster 3C profile ................................................ 68
Appendix 6: QPCR primers for checking HindIII digestion efficiency .................................................................................... 69
Appendix 7: Gel-electrophoresis of qPCR products for HindIII digestion efficiency check .................................................. 70
Appendix 8: Schematic for cloning enhancer candidates into luciferase reporter vector ....................................................... 71
Appendix 9: Schematic for dual luciferase reporter assay ..................................................................................................... 72
Appendix 10: Primers for cloning enhancer candidates surrounding the Sox2 locus ............................................................ 73
Appendix 11: Plasmids containing enhancer candidates ..................................................................................................... 74
Appendix 12: Diagnostic digest of luciferase plasmids containing enhancer candidates ...................................................... 75
Chapter 1 Introduction

The human genome encodes an amazing diversity of cellular morphology and function. All of the cells within the human body contain the same genome, because they are all derived from a single fertilized egg. Yet, different cell types such as cardiomyocytes and neurons have dramatically different morphologies and functions. This diversity is attributed to different gene expression programs generated by the same genome in different cell types. Therefore, understanding how the genome functions in different cellular contexts is an important area of genomic research in the post-sequencing era. Sequencing of the human genome has revealed that only approximately 1.5% of the DNA sequence actually encode proteins. Over the past decade, it has become increasingly clear that the rest of the genome is not “junk DNA”, as traditionally thought, but contains regulatory elements that control the temporal and tissue-specific patterns of gene expression (Consortium et al., 2012). Regulatory elements can control different steps of transcription and can be regulated by epigenetic mechanisms in a cell-type specific manner (Heintzman and Ren, 2009; Maston et al., 2006; Ong and Corces, 2011).

Identification of regulatory elements has been facilitated by the development of high throughput genomic technologies as well as better understanding of the epigenetic landscape and nuclear organization of the genome. Major efforts now are aimed to understand how regulatory elements control complex biological processes, such as embryonic development, and contribute to human diseases (ENCODE Project Consortium et al., 2012).
1.1 Overview of eukaryotic transcription

Transcription of eukaryotic protein-coding genes is carried out by RNA polymerase II (RNAPII). Transcription involves a series of steps that can be subjected to regulation (Fuda et al., 2009). First, RNAPII needs to be recruited to the promoter, where it gets assembled into the pre-initiation complex (PIC) with general transcription factors (TFIIA, TFIIB, TFIID, TFIIE, TFIIF, and TFIIH) (Ranish and Hahn, 1996). Second, the PIC unwinds DNA and activates RNAPII transcriptional activity by phosphorylating its C-terminal domain (CTD) on Ser5 and Ser7 (Akhtar et al., 2009; Glover-Cutter et al., 2009). Once RNAPII exits the core promoter, it pauses at the proximal-promoter region. Phosphorylation of CTD Ser2 is required to release RNAPII from the pause region and allow it to enter productive elongation phase (Peterlin and Price, 2006). Eventually, transcription is terminated and RNAPII is released from DNA and recycled through dephosphorylation of the CTD (Cho et al., 1999). Experimental evidence suggests the recruitment of RNAPII to promoters and the release of paused RNAPII are major rate-limiting steps of transcription (Fuda et al., 2009). This multi-step process of transcriptional regulation involves a large number of factors that integrate a variety of cellular signals.

1.1.1 Players in transcriptional regulation

Transcriptional regulation involves cis-acting regulatory DNA elements and trans-acting DNA-binding factors (Maston et al., 2006). There are two general classes of cis-acting regulatory DNA elements: promoters and distal regulatory elements. Promoters consist of a core promoter, where the PIC is assembled, and a proximal promoter located a few hundred base pairs (bp) upstream of the core promoter (Juven-Gershon and Kadonaga, 2010). The core promoter contains a variety of motifs, such as the TATA-box, BRE (TFIIB recognition element), DPE
(downstream core promoter element), MTE (motif ten element), and Inr (initiator), which encompasses the transcriptional start site (TSS) (Juven-Gershon and Kadonaga, 2010). Distal regulatory elements include enhancers, silencers, insulators, and locus control regions (Heintzman and Ren, 2009; Maston et al., 2006). These elements can regulate genes that are located hundreds of kilobases (kb) away. Enhancers and locus control regions (LCRs) activate transcription (Banerji et al., 1981; Grosveld et al., 1987), whereas silencers repress transcription (Brand et al., 1985). Insulators set up barriers that prevent enhancer-promoter communication (Kellum and Schedl, 1991).

Cis-acting regulatory DNA elements contain recognition sites for trans-acting DNA-binding proteins, which can be classified into three groups: general transcription factors (GTF), transcription activators or repressors, and coactivators (Lee and Young, 2000; Maston et al., 2006). GTFs can recognize specific DNA elements within the core promoter, such as TATA box and Initiator element (Mahowald et al., 2009), and recruit RNAPII to promoter (Ranish and Hahn, 1996). Although the assembled PIC confers a basal level of transcription, its transcriptional activity can be stimulated by transcription activators, which can directly associate with DNA through their DNA binding domains (Pabo and Sauer, 1992). Activators are important players in transcriptional regulation, because they are often expressed in specific cell types and activate specific sets of genes that possess the appropriate recognition sites (Pabo and Sauer, 1992). In addition, activators can form homo- or heterodimers that recognize combinatorial TF binding sites; they can also work synergistically by forming large complexes called enhanceosomes (Thanos and Maniatis, 1995). The advantage of enhanceosomes is that they can integrate information from multiple regulatory pathways to control the output of a
single promoter (Thanos and Maniatis, 1995). Activators can stimulate transcription by recruiting chromatin modifiers that remodel the chromatin structure. For example, activators can recruit ATP-dependent nucleosome modifier SWI/SNF complex, which generates nucleosome-free chromatin at promoters (Yudkovsky et al., 1999). Activators can also recruit histone modifiers such as histone acetyltransferases (HATs) that covalently modify histone tails (Merika et al., 1998). As a result of these modifications, chromatin becomes more open and accessible to components of the transcriptional machinery, facilitating the recruitment of RNAPII to promoters. In addition, activators can directly recruit the transcriptional machinery to promoters or indirectly stimulate its activity via coactivators (Maston et al., 2006).

Unlike transcription activators, coactivators cannot directly associate with DNA but can interact with DNA-bound activators and transcriptional machinery (Malik and Roeder, 2005). One of the most studied coactivators is the evolutionarily conserved Mediator complex, which contains 26 subunits and is 1.2 MDa in size. Because of its large size and surface area, Mediator complex can interact with a large number of DNA-bound factors and thereby plays multifaceted roles in transcriptional regulation (Taatjes, 2010). For example, Mediator can serve as a scaffold that promotes the assembly and stabilization of PIC (Baek et al., 2006; Cantin et al., 2003). Interactions with activators can trigger structural shifts within the Mediator complex that allow it to stimulate RNAPII transcriptional activity (Balamotis et al., 2009; Meyer et al., 2010). Mediator component CDK8 can phosphorylate TFs such as SMADs and STATs to either activate TF activity or target them to proteasome degradation (Alarcon et al., 2009; Bancerek et al., 2013) CDK8 can also stimulate transcriptional elongation by regulating the kinase activity of
positive transcription elongation factor b (P-TEFb), which phosphorylates RNAPII on Ser2 of CTD (Donner et al., 2010).

1.1.2 The role of enhancers in transcriptional regulation

Enhancers are traditionally defined as clusters of DNA sequences that recruit TFs and activate transcription independent of distance or orientation relative to gene promoters (Ong and Corces, 2011). They are often found in intergenic regions of the genome and can regulate genes located hundreds of kb away (Consortium et al., 2012). Several models have been proposed to explain the mechanism of enhancer-mediated transcriptional regulation (Dean, 2006). The looping model (Choi and Engel, 1988), which proposes that enhancers physically associate with distally located promoters by looping out the intervening DNA sequences, gained substantial support from recent studies using chromosome conformation capture (3C) (Krivega and Dean, 2012). Indeed, chromatin loops have been identified at many gene loci, including α-globin (Drissen et al., 2004), β-globin (Palstra et al., 2003), T helper type 2 (T\(_{H2}\)) cytokine (Spilianakis and Flavell, 2004), Ifng (Eivazova and Aune, 2004), MHC class II (Kumar et al., 2007), and IgH (Sayegh et al., 2005). These chromatin loops are established in specific cell types and are associated with onset of transcriptional activation. For example, the murine β-globin locus consists of four genes (ε\(_y\), βH1, βmaj, and βmin) that are expressed at different stages of erythroid development (Kim and Dean, 2012). The ε\(_y\) and βH1 globin genes are expressed in primitive erythroid cells from embryonic yolk-sac, whereas βmaj and βmin are predominantly expressed in definitive erythroid cells from fetal liver and adult spleen (Trimborn et al., 1999; Whitelaw et al., 1990). Transcription of the β-globin genes is regulated by the β-globin LCR located 25 kb upstream of the ε\(_y\) gene. In definitive erythroid cells, the actively expressed βmaj
and βmin genes associate with the LCR, whereas the inactive εy and βH1 genes are looped out (Palstra et al., 2003; Tolhuis et al., 2002). In addition, the LCR interacts with DNase I hypersensitive sites located upstream of the gene cluster (5’HS60/62) and downstream of the LCR (3’HS1) to form an active chromatin hub (ACH) (Palstra et al., 2003). Recently, it was found that artificial looping between LCR and βmaj promoter was sufficient to activate βmaj expression, suggesting that chromatin looping causes, rather than follows, transcriptional activation (Deng et al., 2012).

Once enhancers associate with promoters via chromatin loops, they can activate transcription through several different mechanisms. First, enhancer-bound TFs or coactivators can directly facilitate the recruitment of RNAPII and the assembly of PIC. Second, enhancers can recruit histone modifiers that generate various histone marks, which serve as docking sites for other TFs or coactivators (Ong and Corces, 2011). For example, FOSL1 enhancer recruits histone acetyltransferase MOF, which generates H3K16ac marks that recruit the bromodomain-containing protein BRD4. BRD4 then recruits P-TEFb, which releases RNAPII from proximal-promoter pause and triggers RNAPII elongation (Karam et al., 2010).

### 1.1.3 Proteins that mediate enhancer-promoter communication

Chromatin looping is often established by cell-type specific TFs that bind enhancers and promoters. For example, β-globin genes are regulated by erythroid-specific transcription factors that include KLF-1, GATA-1, and NF-E2 (Cantor and Orkin, 2002; Dean, 2006). These factors localize to LCR and β-globin promoters and are required for physical associations of LCR and promoters (Vakoc et al., 2005). GATA-1 also recruits co-factors such as NLI/Ldb1 that form
multi-component complexes on LCR and promoters. These co-factors are also required for chromatin looping (Deng et al., 2012; Song et al., 2007; Tripic et al., 2009).

The cohesin complex also plays a role in stabilizing chromatin loops (Dorsett and Merkenschlager, 2013; Merkenschlager and Odom, 2013). Cohesin is known to ensure proper chromosome segregation and DNA repair by tethering sister chromatids (Michaelis et al., 1997). Recently, it was found that cohesin co-localizes with Mediator and cohesin loading factor nipped B-like protein (NIPBL) at promoters and enhancers in embryonic stem cells (Kagey et al., 2010). Knockdown of cohesin, Mediator, or NIPBL can disrupt the pluripotency-specific gene expression program and induce differentiation. These phenotypes are due to the disruption of pluripotency-specific chromatin organization, which requires Mediator and cohesin (Apostolou et al., 2013). Moreover, the genomic occupancy of cohesin and Mediator varies between cell types, suggesting that they also contribute to the regulation of cell-type specific gene expression program (Kagey et al., 2010).

The CCCTC-binding factor CTCF is another important factor that mediates long-range chromatin interactions (Phillips and Corces, 2009). CTCF contains an evolutionarily conserved zinc-finger DNA binding domain that recognizes a wide range of derivative sequences from an 11-15 bp core consensus sequence (Chen et al., 2008; Kim et al., 2007). Traditionally, CTCF is thought to bind insulators and play a dual role of blocking enhancer-promoter interaction (Hark et al., 2000) and demarcating boundaries between heterochromatin and euchromatin (Burgess-Beusse et al., 2002). CTCF-bound insulators have been identified at the murine H19/lgf2 locus, where interactions between these insulators at the maternal H19/lgf2 locus form a chromatin loop that prevents lgf2 within the loop from being activated by enhancers outside the loop,
resulting in paternal imprinting of Igf2 (Kurukuti et al., 2006; Murrell et al., 2004). However, recent studies found that CTCF can positively regulate gene expression through facilitating looping interaction between enhancers and promoters (Majumder et al., 2008; Seitan et al., 2011; Splinter et al., 2006). At the MHC class II locus, CTCF mediates the interactions between XL9, an intergenic enhancer element, and promoters of HLA-DRB1 and HLA-DQA1 genes (Majumder et al., 2008). Knockdown of CTCF by siRNA disrupted not only these looping interactions but also the expression of HLA-DRB1 and HLA-DQA1 genes. Genome wide studies showed that CTCF-mediated chromatin interactions form active or repressive domains that regulate gene expression within the loops (Handoko et al., 2011). CTCF also frequently colocalizes with cohesin and the two factors can work together to establish chromatin interactions (Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008).

1.1.3 Chromatin features of enhancers

Enhancers are associated with a number of chromatin features that have been identified by chromatin immunoprecipitation coupled with massively parallel DNA sequencing (ChIP-seq). These features include sequence conservation, nucleosome composition, histone modifications, and TF and coactivator binding (Barski et al., 2007; Consortium et al., 2012; Ernst et al., 2011; Jin and Felsenfeld, 2007; Visel et al., 2009). Identification of these chromatin features has facilitated the prediction of cell-type specific enhancers (Heintzman & Ren, 2009).

The nucleosome is the fundamental unit of chromatin that consists of approximately 147 base pairs of DNA wrapped around a histone octamer, which normally contains two copies each of the core histones H2A, H2B, H3, and H4. For a long time, enhancers were thought to be devoid of nucleosomes due to their sensitivity to DNasel. However, recent studies mapping
genome-wide nucleosome occupancy indicate that enhancers are enriched with unstable nucleosomes containing histone variants H3.3 and H2A.Z (Barski et al., 2007; Jin and Felsenfeld, 2007). Thus, nucleosome instability allows TFs and transcriptional machinery to bind enhancers and promoters (Ong and Corces, 2011).

Histone proteins can be subjected to a wide range of post-translational modifications at the N-terminal tails, such as lysine acetylation, lysine and arginine methylation, and serine and threonine phosphorylation (Cosgrove et al., 2004). Interestingly, the presence of different histone modifications on enhancers is correlated with different levels of enhancer activity. Active enhancers are associated with H3K4me1, H3K4me2, H3K27ac and H3K9ac, whereas weak enhancers are associated with only H3K4me1 (Creyghton et al., 2010; Ernst et al., 2011; Rada-Iglesias et al., 2011). Another class of enhancers called poised enhancers has been identified in pluripotent and multipotent stem cells (Bernstein et al., 2006; Cui et al., 2009). These enhancers are associated with “bivalent histone marks” – presence of both H3K4me3 and H3K27me3. These poised enhancers can be activated upon cell differentiation when they lose H3K27me3 and acquire H3K27ac (Creyghton et al., 2010; Rada-Iglesias et al., 2011). Therefore, enhancer marks can be established long before gene induction during the course of cell differentiation (Ong and Corces, 2011).

Finally, enhancers are also associated with multiple transcription factor-binding loci (MTL) and coactivator p300 binding (Chen et al., 2008; Visel et al., 2009). For example, the gene expression program in ESCs is mainly regulated by three TFs: OCT4, SOX2, and NANOG (Loh and Lim, 2011; Silva and Smith, 2008; Young, 2011). These TFs frequently colocalize to gene promoters and enhancer elements (Boyer et al., 2005; Chen et al., 2008). Putative enhancers in
ES cells have been identified based on colocalization of OCT4/SOX2/ NANOG and a subset of these enhancers has been validated by functional assays (Chen et al., 2008). In addition, coactivator p300 has been shown to be associated with enhancers in a cell-type specific manner (Heintzman and Ren, 2009). Pennacchio and colleagues have identified p300 binding sites in forebrain, mid brain, and limb of embryonic day 11.5 mouse embryos (Visel et al., 2009). They tested a subset of these p300 binding sites in mouse transgenic reporter assays and found that over 80% of the tested elements show tissue-specific enhancer activity.

Although combinations of chromatin features have been used to predict enhancers with various degree of accuracy, they cannot link putative enhancers to target genes. This is because enhancers can be located hundreds of kb away from their target genes and there can be other genes in the intervening region. Recently, Dekker and colleagues used 3C carbon copy (5C) technology to identify long-range chromatin interactions in 1% of the human genome in three cell types (Sanyal et al., 2012). They found that only 7% of the looping interactions involved distal elements and their closest gene. Furthermore, the same enhancers were found to contact different genes in different cellular contexts. Therefore, it is important to experimentally characterize putative enhancers using methods such as 3C and luciferase reporter assays.

1.2 Transcription factor SOX2

SOX2 belongs to the Sry-related high mobility group (HMG) box (SOX) family of transcription factors. The SOX family consists of 20 members that are categorized into eight groups (A-H) based on sequence homology of the HMG DNA binding domain (Sarkar and Hochedlinger, 2013). SOX proteins play important roles in stem cell maintenance and cell fate specification in
a variety of epithelial tissues throughout embryonic development. In particular, SOXB group
members regulate the self-renewal and differentiation of neural stem and progenitor cells
(Pevny and Placzek, 2005). The SOXB1 subgroup consists of SOX1, SOX2, and SOX3 that function
as transcriptional activators due to the presence of a C-terminal transactivation domain. The
SOXB2 subgroup includes SOX14 and SOX21 that have a C-terminal transrepression domain and
thus function as repressors. The SoxB1 and SoxB2 genes have similar expression patterns in the
neuroectodermal lineage (Pevny and Placzek, 2005).

1.2.1 Sox2 expression pattern

The expression pattern of Sox2 in the developing embryo has been extensively characterized
(Avilion et al., 2003a). Sox2 transcription is detected in morula at embryonic day 2.5 (E2.5) and
in both inner cell mass (ICM) and trophectoderm (TE) of E3.5 blastocysts. By E6.5, Sox2 is
expressed in epiblast and extraembryonic ectoderm (ExE). By E7.5, Sox2 expression becomes
restricted to the anterior region of the presumptive neuroectoderm (Avilion et al., 2003a). Sox2
is continually expressed in fetal and adult epithelial tissues in brain, retina, inner ear, trachea,
tongue, lungs, foregut, skin, and gonads (Arnold et al., 2011). Genetic lineage tracing
experiments demonstrated that Sox2-expressing cells in adult tissues are capable of self-
renewal and can differentiate into multiple mature cell types in each tissue, suggesting that
Sox2 may be a marker for multipotent stem and progenitor cells (Arnold et al., 2011).

1.2.2 The role of Sox2 in pluripotent stem cells

The hallmarks of pluripotent stem cells are the ability to differentiate into derivatives of all
three germ layers (pluripotency) and the ability to replicate indefinitely without undergoing
differentiation (self-renewal). The pluripotency and self-renewal capabilities of ES cells are maintained by a transcriptional regulatory network that controls the ES cell-specific gene expression program (Jaenisch and Young, 2008; Loh and Lim, 2011; Silva and Smith, 2008). The key components of this network are transcription factors OCT4, SOX2, and NANOG. SOX2 dimerizes with OCT4, a member of the POU family of homeodomain proteins, and binds to the OCT4/SOX2 consensus sequence, which is enriched in promoters and enhancers (Masui et al., 2007). NANOG frequently colocalizes with OCT4/SOX2 dimers and coassociation of O/S/N results in transcriptional activation of pluripotency genes and repression of lineage-specifying genes (Boyer et al., 2005; Chen et al., 2008; Loh et al., 2006; Marson et al., 2008). O/S/N upregulate each other’s expression through an autoregulatory feedback loop (Young, 2011). In addition, O/S/N recruits transcription factors involved in self-renewal (ESRRB, ZFX, KLF4), extrinsic growth signaling (STAT3, SMAD1), and cell cycle progression (E2F1) as well as coactivators (p300, Mediator) (Chen et al., 2008). These multiple transcription factor-binding loci (MTL) are enriched in intergenic regions of the genome and have ES cell specific enhancer activity (Chen et al., 2008). The requirement of Sox2 in ES cells has been demonstrated by functional studies in mice. Homozygous deletion of Sox2 causes embryonic lethality due to a failure to form pluripotent epiblast (Avilion et al., 2003a). Sox2-null ES cells show downregulation of Oct4 and Nanog and undergo differentiation into trophectoderm (Masui et al., 2007). Interestingly, Sox2-null phenotype can be rescued by forced expression of Oct4, suggesting that the main role of Sox2 in ES cells is to regulate Oct4 expression (Masui et al., 2007). On the other hand, overexpression of Sox2 can induce differentiation with a bias toward the neuroectodermal lineage (Kopp et al., 2008). Therefore, Sox2 functions in a dose-
dependent manner in ES cells and precise regulation of Sox2 expression is necessary to maintain the pluripotent state (Sarkar and Hochedlinger, 2013).

1.2.3 The role of Sox2 in neural development

Sox2 plays important roles in regulating the development of the central nervous system (CNS) (Pevny and Nicolis, 2010; Sarkar and Hochedlinger, 2013; Wegner, 2011). Indeed, Sox2 expression in the CNS is evolutionarily conserved from zebrafish to chick to human (Collignon et al., 1996). SOX2 promotes neuroectodermal differentiation of ES cells by directly suppressing the key regulators of meso- and endodermal lineages such as brachyury (Thomson et al., 2011). SOX2 also acts as a pioneering factor that binds promoters and enhancers of neural genes, which are enriched with bivalent histone marks, and preselects these genes for future activation by SOX2 in neural stem cells (Bylund et al., 2003a). Sox2 is expressed in neural precursor cells (NPCs) located in the neurogenic regions of embryonic and adult brain, which include the ventricular zone (VZ) of telencephalon and subgranular zone (SGZ) of hippocampus (Arnold et al., 2011; Ellis et al., 2004; Zappone et al., 2000). Functional studies have shown that Sox2 is required for maintaining the self-renewal of NPCs (Wegner and Stolt, 2005). For example, hypomorphic Sox2 mutations or conditional deletion of Sox2 in embryonic brain resulted in the loss of GFAP/nestin-expressing neural stem cells in the VZ and the SGZ (Favaro et al., 2009; Ferri et al., 2004b). Overexpression of Sox2 in chicken neural tube prevented neuronal differentiation of NPCs, whereas overexpression of a dominant-negative Sox2 allele induced cell-cycle exit and neuronal differentiation (Bylund et al., 2003b; Graham et al., 2003). However, Sox2 has also been shown to be required for proper neurogenesis. NPCs with Sox2 hypomorph mutations showed defects in neuronal differentiation (Cavallaro et al., 2008;
Taranova et al., 2006). Therefore, the function of Sox2 in NPCs is influenced by its expression level, similar to the situation in ES cells (Pevny and Nicolis, 2010; Sarkar and Hochedlinger, 2013; Wegner, 2011).

1.2.4 Transcriptional regulation of Sox2

Although the expression patterns of Sox2 in fetal and adult tissues have been extensively studied, the transcriptional regulation of Sox2 in different cell types remains largely unclear. Previous studies identified two enhancer elements, referred to as Sox2 Regulatory Region 1 (SRR1) and SRR2, which are located approximately 4 kb upstream and 4 kb downstream from the Sox2 transcriptional start site (TSS) (Miyagi et al., 2004; Tomioka et al., 2002; Zappone et al., 2000). These enhancers have been shown to drive transgene expression in undifferentiated ES cells and multipotent NPCs in VZ of embryonic brain (Miyagi et al., 2004; Tomioka et al., 2002; Zappone et al., 2000). These enhancers do not function in mature neurons or trophoblast stem cells, which express Sox2 (Miyagi et al., 2004). Although Sox2 is haplosufficient in mice (Ferri et al., 2004a), knocking out one Sox2 allele and the SRR1 enhancer of the other Sox2 allele in mice reduced the level of SOX2 proteins and caused significant neurological defects and loss of NPC population (Ferri et al., 2004b). This suggested that SRR1 is required for regulating Sox2 expression in NPCs. SRR1 contains two POU binding sites that are required for enhancer activity (Catena et al., 2004). SRR2 contains a POU/SOX2 composite binding site that is required for enhancer activity of SRR2 in ES cells and NPCs (Miyagi et al., 2004; Tomioka et al., 2002). This composite site can recognize members of the POU family of TFs, such as OCT4 and OCT6 in ES cells and OCT6, BRN1, BRN2, and BRN4 in NPCs (Miyagi et al., 2004; Tomioka et al., 2002).
1.2.5 Novel predicted enhancer candidates near Sox2 locus

Although Sox2 is known to be regulated by SRR1 and SRR2 in both ES cells and NPCs, it is not known whether Sox2 can be regulated by other enhancers in ES cells. Recently, Dr. Mitchell’s lab used a computational modeling approach to predict enhancer regions in mouse ES cells (Chen et al., 2012). We developed a multinomial logistic regression model that classified 1 kb genomic bins as either enhancer-like or promoter-like regions based on chromatin features, including histone modifications, p300, Mediator, and cohesin binding (Chen et al., 2012). The model was trained to classify OCT4/SOX2/NANOG (O/S/N) co-bound sites as putative enhancer regions and classify MYC/MYCN co-bound sites as putative promoter-like regions. The model predicted a large number of previously known enhancers in ES cells with high probability scores (prob>0.8) (Chen et al., 2012). Interestingly, a number of these enhancer candidates were located within a 150 kb region surrounding the Sox2 locus. In the proximal region to Sox2, the model predicted five enhancer candidates located -4 kb, +4 kb, +10 kb and +18 kb relative to Sox2 TSS. The enhancer candidates located at -4 kb and +4 kb overlapped with SRR1 and SRR2. In addition, the model identified a cluster of high probability enhancer candidates 80-120 kb downstream to Sox2. This distal enhancer cluster is bound by a large set of TFs involved in regulating the self-renewal of ES cells (Figure 1). These TFs include OCT4, SOX2, NANOG, KLF4, ESRRB, STAT3, SMAD1, and ZFX (Xiao et al., 2012). Furthermore, this distal enhancer cluster is enriched with active enhancer marks (H3K27ac, H3K4me1, and p300 binding) only in ES cells (Figure 2). These marks are lost as ES cells differentiate toward the mesendodermal lineage (Xiao et al., 2012). This suggested that this distal enhancer cluster might contain ES cell-specific enhancers.
Figure 1: Novel enhancer candidates surrounding the Sox2 locus are bound by ES cell-specific TFs. The non-coding Sox2 overlapping transcript (Sox2ot) is shown as a blue line. The Sox2 transcript including untranslated regions is shown as a blue box. The high probability enhancer candidates identified by Chen et al. (2012) are shown in red bars with heights corresponding to the probability score. The previously identified SRR1 and SRR2 enhancers are shown in green bars. Transcription factor ChIP-seq peaks identified using SISSRs algorithms are shown in black boxes and numbers indicate peak intensity (Adapted from (Chen et al., 2012)).
Figure 2: Cell-type specific patterns of H3K4me1 and p300 binding surrounding the Sox2 locus. The high probability enhancer candidates are shown in red bars with heights corresponding to the probability score. The previously identified SRR1 and SRR2 enhancers are shown in green bars. The p300 ChIP-seq data in E11.5 embryonic forebrain are obtained from Visel et al. (2009) and are shown in black boxes. The mouse ENCODE ChIP-seq data for p300, H3K4me1, H3K27ac, and CTCF in E14.5 embryonic brain, mouse ES cell, and mouse embryonic fibroblasts (MEFs) are displayed in UCSC browser (Consortium et al., 2012).
1.3 Objective

The goal of my Master’s thesis was to determine whether the high probability enhancer candidates predicted by Chen et al. (2012) are ES cell-specific enhancers for Sox2. To investigate this question, I examined three tissue types that differentially express Sox2: E14TG2a (E14) mouse ES cell line, embryonic brains isolated from E11.5 mouse embryos, and mouse embryonic fibroblasts (MEFs). At E11.5, the embryonic brain contains a large population of Sox2-positive NPCs that start to undergo neurogenesis (Bani-Yaghoub et al., 2006; Zappone et al., 2000). As both ES cells and embryonic brain express Sox2, comparing these two tissue types would allow me to identify cell-type specific enhancers for Sox2. MEFs were used as the negative control cell type, because they do not express Sox2. I adopted a two-pronged approach to evaluate enhancer candidates. First, I used chromosome conformation capture (3C) technology to detect long-range chromatin interactions near the Sox2 locus. I hypothesized that the distal enhancer cluster downstream to Sox2 would interact with the Sox2 promoter in mouse ES cells but not in embryonic brains or MEFs. Second, I cloned the enhancer candidates individually into luciferase reporter plasmids and measured enhancer activity in mouse ES cells and MEFs. The ES cell-specific enhancers were expected to generate luciferase activity in ES cells but not MEFs. The results from this project would provide important new insights into cell-type specific regulation of Sox2 transcription.
Chapter 2 Methods

2.1 Cell culture

E14 mouse ES cells were cultured on gelatin-coated plates in DMEM containing 15% FBS, 0.1 mM MEM non-essential amino-acids, 1 mM sodium pyruvate, 2 mM Gluta-MAX™, 0.1 mM 2-mercaptoethanol, and 1000 U/ml LIF. The medium was further supplemented with 3 μM CHIR99021 (GSK3β inhibitor) and 1 μM PD0325901 (MEK inhibitor), which have been demonstrated to maintain ES cells in a pluripotent state in the absence of a feeder layer (Ying et al., 2008). MEFs were isolated from E12.5 C57Bl/6 mouse embryos and cultured in DMEM containing 10% FBS and 2 mM Gluta-MAX™.

2.2 Chromosome Conformation Capture (3C)

The 3C protocol was adopted from (Dekker et al., 2002) with minor modifications (Appendix 1).

Preparation of single-cell suspension

E14 mouse ES cells and MEFs were detached from culture plates by a brief incubation with trypsin. Cells were transferred into 50 ml falcon tubes and centrifuged at 1300 rpm for 4 min. The cell pellets were resuspended in 36 ml of DMEM/FBS at room temperature. Embryonic brains isolated from E11.5 C57Bl/6 mouse embryos were strained through a 70 μm filter to achieve a single cell suspension in 36 ml DMEM/FBS.

Fixation and restriction digestion of nuclei

Cells were fixed in DMEM/FBS containing 1% (v/v) formaldehyde for 10 min at room temperature. Fixation reaction was quenched by incubation with 2.85 ml of 2M glycine on ice for 10 min. Cell pellets were collected by centrifugation at 1300 rpm for 8 min and washed with cold PBS. Then, cells were resuspended in 20 ml of cold lysis buffer (10 mM Tris-HCl, pH 8, 10
mM NaCl, 0.2% NP-40, and complete protease inhibitors) and incubated on ice for 30 min. During this time, cell membranes were solubilised and nuclei were released. After centrifugation at 1800 rpm for 5 min, pelleted nuclei were resuspended in 1-2 ml of cold lysis buffer and counted using Cell Counter. 1x10^7 nuclei were resuspended in 500 μl of 1.2x NEB Buffer2 containing 20% SDS and incubated on shaker for 1 hr at 37°C, 950 rpm. Then, SDS was quenched by adding 50 μl of 20% Triton-X100 to the reaction and incubating for 1 hr at 37°C, 950 rpm. A small aliquot was taken from each sample and transferred into a tube labeled as undigested genomic DNA (UND) and stored at -20°C. Then, 1500 U HindIII restriction enzyme (NEB) was added to the reaction and incubated overnight at 37°C, 950 rpm.

**Ligation of chromatin fragments**

After overnight digestion with HindIII, another small aliquot was taken from each sample and transferred into a tube labelled digested genomic DNA (D) and stored at -20°C. Restriction enzyme was deactivated by adding 40 μl of 20% SDS and incubating at 65°C for 25 min. Then, the reaction was diluted in 7 ml of 1.1x T4 DNA ligase reaction buffer (NEB). 375 μl of 20% Triton-X100 was added and incubated at 37°C for 1 hr to quench SDS. Digested chromatin was re-ligated by adding 800 U of T4 DNA ligase (NEB) to the reaction and incubating at 16°C for 4 hrs. After ligation, formaldehyde crosslinks were reversed by adding 900 μg of Proteinase K and incubating overnight at 65°C. Contaminating RNA was degraded by adding 30 μg of RNase A and incubating at 37°C for 1 hr.

**DNA purification**

3C libraries were purified by Phenol-Chloroform extraction and precipitated by adding 2.5 vol of 100% ethanol and 0.1 vol of 2 M sodium acetate and incubating at -20°C overnight. Precipitated
DNA was collected by centrifugation at 10,000 rpm for 30 min and washed in 70% ethanol. After evaporating the ethanol, DNA was resuspended in DNase-, RNase-free water (Sigma). DNA was cleaned up using DNeasy Blood & Tissue Kit (Qiagen). The UND and D aliquots were treated with Proteinase K and RNase A and purified in the same way as 3C libraries.

**3C validation**

To validate the 3C libraries, I compared the efficiency of *HindIII* digestion at 14 restriction sites in mouse ES cells, embryonic brains, and MEFs. Large differences in digestion efficiency between cut sites or cell types would result in different availability of restriction ends for ligation, creating artifacts in the 3C analysis. Previous studies suggested that digestion efficiency should be above 80% across all cut sites in all biological systems for accurate comparison. Thus, I performed real-time quantitative PCR (qPCR) on undigested and digested aliquots obtained from 3C samples before and after adding *HindIII*. The digestion efficiency primer pairs amplified 14 cut sites surrounding the *Sox2* locus (Figure 3). To correct for the loading differences of template DNA in the qPCR reactions, I used a control primer pair that amplified a DNA sequence within a restriction fragment. The formula used to calculate digestion efficiency at a restriction site is shown below (Hagege et al., 2007). 

\[
\text{\% restriction} = 100 - 100/2^{((C_{\text{R}} - C_{\text{C}})_D - (C_{\text{R}} - C_{\text{C}})_{\text{UND}})}
\]

where \(C_{\text{R}}\) and \(C_{\text{C}}\) represent the cycle threshold (Ct) values for the digestion efficiency and control primer pairs, respectively.
Optimization of formaldehyde fixation

3C involves fixing cells with formaldehyde to cross-link chromatin. The extent of fixation can affect the efficiency of restriction digestion and ligation (Dekker, 2006). Therefore, the formaldehyde concentration and time of incubation were optimized. Four fixation conditions were tested on E14 ES cells: 2% formaldehyde for 10 min, 2% for 5 min, 1.5% for 10 min, and 1% for 10 min. After fixation, cells were lysed and chromatin was digested overnight with HindIII. Undigested and digested chromatin were collected and DNA was purified by Phenol-Chloroform extraction. The efficiency of HindIII digestion at 14 restriction sites was evaluated by qPCR. The results are shown in Appendix 3. The fixation condition that yielded the highest digestion efficiency was fixing cells in 1% formaldehyde for 10 min at room temperature. Therefore, this fixation condition was used for all 3C experiments.

3C quantification

The ligation products in 3C libraries were quantified by qPCR using primers that annealed 50-80bp away from 3’ end of restriction fragments (Figure 3). To eliminate biases in the quantification due to different amplification efficiencies of primer pairs, I prepared serial dilutions of a 3C control library that contained all-possible ligations of the chromatin fragments surrounding the Sox2 locus. The 3C control library was prepared by digesting a BAC clone (RP23-274P9) that covers a 223 kb region containing the Sox2 locus on mouse chromosome 3 with HindIII restriction enzyme and ligating the fragments with T4 DNA ligase. Another BAC clone (RP23-2N15) that covers a 217 kb region containing the Alpha Aortic Actin-2 (a-A2) locus on mouse chromosome 19 was digested and ligated in the same fashion. The ligation products from Sox2 BAC and Acta2 BAC were mixed at equimolar ratio.
Serial dilutions of the 3C control library were used to generate standard curves for each 3C primer pair. An appropriate amount of 3C library that fell within the linear range of amplification (usually 50 ng of DNA) was used in qPCR. The starting quantity of a ligation product in the 3C libraries represents the frequency of interaction between two chromatin fragments. The interaction frequency of fragments in the Sox2 locus was normalized to that in the Alpha aortic actin locus (\(a\)-A2), which is expressed at similar levels in mouse ES cells, embryonic brains, and MEFs and presumably adopts similar chromatin conformation in all three cell types (Bhattacharya et al., 2012). Normalizing 3C data to a control locus, such as Acta2, accounts for the effects of internal environments of different cell types on cross-linking, digestion, and ligation of chromatin (Gavrilov et al., 2009). All primers used in 3C analysis are listed in Appendix 2.

**Statistical analysis**

The 3C data were analyzed by two-way ANOVA using Sigma Plot12. Post tests were performed using the Holm-Sidak method to assess significant differences in interaction frequency between ES cells, MEFs, and E11.5 embryonic brain at specific genomic loci.
Figure 3: qPCR primers for 3C quantification and digestion efficiency check. The high probability enhancer candidates are shown in red bars with heights corresponding to the probability score. The previously identified SRR1 and SRR2 enhancers are shown in green bars. The HindIII restriction sites are shown in black lines. The qPCR primers for 3C quantification are shown in red lines and qPCR primers for digestion efficiency check are shown in brown lines. The 3C primers were located 20-100 bp upstream to HindIII restriction sites. The digestion efficiency primers were located 50-80 bp downstream to HindIII restriction sites. The ENCODE ChIP-seq data for p300 binding in mouse ES cell (Bruce 4) is displayed in UCSC browser (Consortium et al., 2012)
2.3 Luciferase Reporter Constructs

The pipeline for cloning enhancer candidates into luciferase reporter plasmids is illustrated in Appendix 8. SRR1, SRR2, and nine enhancer candidates (prob>0.8) were amplified from Sox2 BAC (RP23-274P9) by PCR (Figure 4). The PCR products were 900-1,600 bp in size and enhancer primers are listed in Appendix 10. Amplified enhancer candidates were cloned into the pJET1.2/blunt cloning vector (Fermentas). The insert orientation and sequence were confirmed by gel electrophoresis and DNA sequencing done by CAGEF. From pJET1.2 vectors, enhancer candidates were amplified by primers containing 15 bp 5’overhangs that were homologous to the sequences flanking the NotI restriction site in pGL4.23 luciferase reporter vector. The PCR products were cloned into the NotI site of pGL4.23 vector by In-Fusion® cloning reaction (Clontech). By exchanging the overhangs on the forward and reverse primers, enhancer candidates were cloned into pGL4.23 in either forward or reverse orientation. Once again, insert orientation and sequence were confirmed by gel electrophoresis and DNA sequencing. The pGL4.23 vector contains firefly luciferase gene luc2 regulated by a TATA-box minimal promoter. It also contains a synthetic poly (A) signal/transcriptional pause site upstream to the minimal promoter that reduces transcriptional background. The NotI site is located 1344 bp downstream to luc2. High concentrations of pGL4.23-enhancer plasmids and pGL4.75 Renilla luciferase plasmid, which contains a Renilla luciferase gene hRluc regulated by CMV promoter, were purified from bacterial culture using PureLink® HiPure Plasmid Midiprep Kit (Invitrogen). Each pGL4.23-enhancer plasmid was mixed with pGL4.75 at 50:1 molar ratio. The concentration of pGL4.23-enhancer plasmid in the mixture was 0.1 µg/µl. The pGL4.75 plasmid was used as an internal control in performing the dual-luciferase reporter assays (Appendix 9).
Figure 4: Cloning enhancer candidates into luciferase reporter plasmid. The high probability enhancer candidates are shown in red bars with heights corresponding to the probability score. The previously identified SRR1 and SRR2 enhancers are shown in green bars. The ENCODE ChIP-seq data for p300 binding in mouse ES cell (Bruce 4) is displayed in UCSC browser (Consortium et al., 2012). Enhancer candidates that have been cloned into pGL4.23 luciferase reporter plasmid in both forward and reverse orientations are shown in red boxes.
2.4 Dual Luciferase Reporter Assay

E14 mouse ES cells were seeded onto gelatin-coated 96-well plates at $1.0 \times 10^4$ cells/well density and grown overnight in 37°C incubator. Luciferase plasmids were diluted 5X in ES medium; Lipofectamine 2000 reagent was also diluted in ES medium at 4:46 (v/v) ratio. Before transfection, 100 μl of fresh ES medium was added to each well on the 96-well plates. Diluted luciferase plasmids and diluted Lipofectamine reagent were mixed at 1:1 (v/v) ratio and incubated at room temperature for 30min. Then, 20 μl of DNA-Lipofectamine mixture was added to each well on 96-well plates. The medium was changed after 24 hrs of incubation. Transfection with each enhancer plasmid was done in triplicate. Dual-luciferase reporter assays (Promega) were performed 48 hrs after transfection. Cells were lysed in Passive Lysis Buffer and 20 μl of lysate from each well were transferred into a white 96-well plate for measurement. Dual luciferase reporter assays were performed on Fluoroskan Ascent FL plate reader using the DLR 1sec protocol. The ratio of firefly/Renilla luciferase activity was calculated for each tested enhancer candidate and was normalized to the empty vector. The luciferase data were analyzed by two-way ANOVA using Sigma Plot12. Post tests were performed using the Holm-Sidak method to assess significant differences in luciferase activity for each enhancer candidate between ES cells and MEFs.
Chapter 3 Results

3.1 Long-range chromatin looping at the Sox2 locus

Although Sox2 is expressed in pluripotent ES cells and multipotent stem and progenitor cells of fetal and adult tissues (Arnold et al., 2011; Avilion et al., 2003b; Marson et al., 2008), the transcriptional regulation of Sox2 in different cell types is not clearly understood. The SRR1 and SRR2 proximal enhancers surrounding Sox2 have been shown to be active in ES cells and NPCs, but it is not known whether they are necessary or sufficient to regulate Sox2 expression in ES cells. Also, it is not known whether there might be ES cell-specific enhancers for Sox2. However, a recent study by Dr. Mitchell’s lab predicted a number of high probability enhancer candidates surrounding the Sox2 locus (Chen et al., 2012). In particular, a cluster of enhancer candidates located 80-120kb downstream to Sox2 shows high levels of active enhancer marks (H3K4me1, H3K27ac, MED12, and p300) in ES cells but not embryonic brain tissues (Figure 2). Therefore, I hypothesized that this distal enhancer cluster might contain ES cell-specific enhancers for Sox2. To test this hypothesis, I used the 3C technology to profile the three-dimensional chromatin architecture of a 150kb region surrounding the Sox2 locus in mouse ES cells, E11.5 embryonic brains, and MEFS. I hypothesized that the distal enhancer cluster would interact with Sox2 promoter only in ES cells.

Because the 3C technology relies on restriction digestion and ligation of cross-linked chromatin, large variations in digestion efficiency among different restriction sites would create artifacts in the 3C results. Therefore, I performed qPCR to evaluate the efficiency of HindIII digestion at 14 restriction sites near the Sox2 locus in ES cells, embryonic brains, and MEFS. The
results showed that *HindIII* digestion efficiency was approximately 90% at all tested restriction sites in all three cell types (Figure 5). This experiment established the validity of my 3C analysis.

![HindIII Digestion Efficiency](image)

**Figure 5: HindIII digestion efficiency is uniform across restriction sites and cell types.** A small aliquot of chromatin was collected from each 3C sample before (undigested) and after (digested) *HindIII* restriction digestion. The aliquots were treated with proteinase K to reverse formaldehyde cross-linking and with RNase A to remove contaminating RNA. DNA was purified by phenol/chloroform extraction. Primers were used to amplify 14 *HindIII* restriction sites near the Sox2 locus in undigested and digested samples. The qPCR amplification of each restriction site was performed in triplicates. The Ct values were used to calculate digestion efficiency. Values are an average of three or four independent experiments, each performed in triplicate. Error bars represent standard deviation.
Next, I profiled the interactions between the Sox2 promoter and surrounding chromatin fragments using qPCR. I used a primer located within the Sox2 promoter fragment as the anchor primer and I paired it with target primers located in surrounding chromatin fragments. The results produced a number of interesting observations (Figure 6). First, the interaction profile of MEFs showed that Sox2 promoter fragment interacted with proximal fragments at high frequency and that this interaction frequency decreased dramatically as the genomic separation between the Sox2 promoter and the target fragment increased. This inverse relationship between interaction frequency and genomic separation was commonly observed in previous 3C studies (Dekker, 2006). The frequent interactions between nearby fragments are mainly due to random collisions of flexible chromatin ends, but the frequency of random collisions decreases as the genomic separation increases (Dekker, 2006). Observation of this inverse relationship in the negative control cell type provided a validation for my 3C analysis.

In the proximal region within 20 kb to Sox2 TSS in either direction, the interaction frequency in embryonic brain was significantly higher than that in ES cells or MEFs (P<0.05) (Figure 6). This was unexpected because this proximal region displays similar patterns of chromatin modifications in ES cells and embryonic brain (Figure 2). The SRR1 and SRR2 enhancers are located on fragments immediately adjacent to the Sox2 promoter fragment. Therefore, their interactions with the Sox2 promoter were masked by random collisions due to proximity. Interestingly, the Sox2 promoter interacted with the distal enhancer cluster in ES cells, as evidenced by a local peak in interaction frequency, but not in MEFs or embryonic brains (Figure 6). The frequency of this long-range interaction in ES cells was approximately 9-fold and 3.5-fold higher than that in MEFs and embryonic brain, respectively. These results suggested
that the Sox2 promoter interacted with the distal enhancer cluster through chromatin looping in ES cells but not E11.5 embryonic brains or MEFs.

To validate the looping interaction between Sox2 promoter and the distal enhancer cluster, I used a fragment within the enhancer cluster as the anchor fragment and generated 3C interaction profiles for ES cells, embryonic brain, and MEFs. The anchor fragment contained four enhancer candidates: Enh106, Enh107, Enh109, and Enh111 (See Figure 4 on page 25). The interaction profile showed that this anchor fragment interacted with the proximal region surrounding the Sox2 locus at significantly higher frequency in ES cells than in embryonic brains or MEFs ($P<0.05$) (Figure 7). Importantly, this proximal region included SRR1 and SRR2, as well as enhancer candidate Enh18 (See Figure 4 on page 25). This suggested that the distal enhancer cluster interacted with not only the Sox2 promoter but also other proximal enhancers surrounding Sox2, resulting in the formation of a chromatin hub. In embryonic brain, the distal enhancer fragment interacted with a fragment located 12.5-22.5 kb downstream to Sox2 at significantly higher frequency than MEFs ($P<0.05$). But embryonic brain did not show interaction between the distal enhancer cluster and Sox2 promoter. Therefore, my 3C analyses revealed cell-type specific chromatin architecture surrounding the Sox2 locus in mouse ES cells, embryonic brains, and MEFs.
Figure 6: Chromatin interaction profile of the Sox2 locus with Sox2 promoter as the anchor. Chromosome conformation capture (3C) was performed on mouse ES cells, E11.5 embryonic brains, and MEFs. The frequency of interaction between the Sox2 promoter (anchor) and the surrounding chromatin was normalized to that between adjacent fragments at the Alpha aortic actin (α-Act) locus. Black bar represents the anchor fragment and shaded bars represent the interacting fragments. Values are an average of three independent experiments, with each experiment performed in triplicate. Error bars represent standard deviation. A significant difference between ES cells or E11.5 brain MEFs is indicated by *(P<0.5), **(P<0.01) or ***(P<0.001). A significant differences between ES cells and E11.5 brain is indicated by Δ (P<0.05), ΔΔ (P<0.01).
Figure 7: Chromatin interaction profile of the Sox2 locus with the distal enhancer cluster as the anchor. Chromosome conformation capture (3C) was performed on mouse ES cells, E11.5 embryonic brains, and MEFs. The frequency of interaction between the distal enhancer cluster (anchor fragment) and the surrounding regions is normalized to that between adjacent fragments at the Alpha aortic actin (α-Act) locus. Black bar represents the anchor fragment and grey bars represent the interacting fragments. Values are an average of three independent experiments, with each experiment performed in triplicate. Error bars represent standard deviation. A significant difference between ES cells and MEFs is indicated by *(P<0.5), ***(P<0.01) or ****(P<0.001). A significant differences between ES cells and E11.5 brain is indicated by Δ (P<0.05), ΔΔ (P<0.01).
3.2 Enhancer activity of novel enhancer candidates

Although my 3C results demonstrated looping interactions between the distal enhancer cluster and the Sox2 promoter, these interactions do not necessarily play a role in regulating Sox2 expression. Therefore, I tested the ability of the enhancer candidates to drive reporter gene expression in mouse ES cells and MEFs. I cloned SRR1, SRR2, and eight enhancer candidates into pGL4.23 luciferase reporter vector in both forward and reverse orientations relative to the direction of Sox2 transcription. As a negative control, I cloned a 1.2 kb region (called desert) located 55 kb downstream to Sox2 TSS that did not show active enhancer marks or interact with the Sox2 promoter in my 3C results. The pGL4.23 vector contained the firefly luciferase gene luc2 regulated by a minimal promoter (Appendix 9). The pGL4.23-enhancer plasmids were co-transfected with a Renilla luciferase plasmid (pGL4.75) at 50:1 molar ratio into mouse ES cells or MEFs cultured on 96-well plates. The Renilla luciferase gene hRluc was regulated by a constitutive CMV promoter and Renilla luciferase activity was used to normalize the transfection efficiencies of different cell types. The transfection efficiencies of ES cells and MEFs were approximately 70% and 60%.

The luciferase activity was measured by the Dual-Luciferase Reporter Assay System (Promega) at 48 hrs post-transfection. The ratios of firefly/Renilla activity for cloned fragments were normalized to the empty vector (Figure 8). The desert region showed similar luciferase activity as the empty vector in both ES cells and MEFs. As expected, the luciferase activities of SRR1 and SRR2 in ES cells were significantly higher than those in MEFs (P<0.05 for SRR1, P<0.001 for SRR2). In ES cells, SRR1 and SRR2 had approximately 3- and 10-fold higher luciferase activity than the empty vector, respectively. Although SRR2 appeared to have higher
luciferase activity than SRR1 in ES cells, the difference was not significant ($P=0.091$). A previous study by Tomioka et al. (2002) showed that SRR2 had stronger enhancer activity than SRR1 in ES cells and embryonic carcinoma cells. Also, SRR1 and SRR2 in the reverse orientation had similar luciferase activities as in the forward orientation.

Four of the eight enhancer candidates (Enh18, Enh95, Enh107, and Enh111) showed significantly higher luciferase activity in ES cells than in MEFs ($P<0.001$). Enh107 and Enh111 in forward orientation had 40- and 35-fold higher luciferase activity than the empty vector in ES cells. In both forward and reverse orientations, Enh18, Enh107 and Enh111 had significantly higher luciferase activity than SRR1 ($P<0.001$). Enh107 also had significantly higher activity than SRR2 in both orientations ($P<0.05$). However, Enh111 had significantly higher activity than SRR2 only in forward orientation ($P<0.05$). Enh18 and Enh95 did not have significantly higher luciferase activity than SRR2. Interestingly, the reverse orientation of Enh107 and Enh111 appeared to have lower luciferase activity than the forward orientation, but the differences were not significant ($P=0.247$ for Enh107, $P=0.998$ for Enh111). Unexpectedly, Enh85 and Enh114 in forward orientation had significantly higher luciferase activity in MEFs than in ES cells ($P<0.01$).
Figure 8: Luciferase activity of enhancer candidates surrounding the Sox2 locus in mouse ES cells and MEFs. The ENCODE ChIP-seq data for p300 binding in mouse ES cell (Bruce 4) is displayed in UCSC browser (Consortium et al., 2012). SRR1, SRR2, and eight enhancer candidates were individually cloned into pGL4.23 luciferase plasmids downstream to the firefly luciferase gene luc2, which was regulated by a minimal promoter. A desert region was included as a negative control. The pGL4.23-enhancer plasmids were co-transfected with pGL4.75 Renilla luciferase plasmid at 50:1 molar ratio into mouse ES cells or MEFs for 48 hrs before performing the dual-luciferase reporter assays. The ratios of firefly/Renilla luciferase activity for enhancers were normalized to the empty pGL4.23 vector. Values are an average of three independent experiments, with each experiment performed in triplicate. Error bars represent SEM. A significantly higher activity over the empty vector is indicated by **(P<0.01) or ***(P<0.001). Significant differences between ES cells and MEFs for each enhancer are indicated by Δ (P<0.05), ΔΔ (P<0.01), or ΔΔΔ (P<0.001).
Chapter 4 Discussion

4.1 Summary of Results

Sox2 is known to play critical roles in regulating the gene expression program of ES cells and the expression level of Sox2 must be tightly regulated to maintain the self-renewal of ES cells (Avilion et al., 2003b; Kopp et al., 2008; Masui et al., 2007). However, the mechanism that regulates Sox2 expression in ES cells is not clearly understood. Previous studies showed that SOX2, OCT4, and NANOG upregulate each other’s expression by colocalizing to each other’s promoter (Boyer et al., 2005; Chen et al., 2008). In the proximal region to Sox2, two enhancers SRR1 and SRR2 have been shown to drive transgene expression in ES cells as well as NPCs (Catena et al., 2004; Miyagi et al., 2004; Tomioka et al., 2002; Zappone et al., 2000). SRR1 and SRR2 are targeted by POU family TFs and POU/Sox2 heterodimers in ES cells as well as NPCs (Catena et al., 2004; Miyagi et al., 2004; Tomioka et al., 2002). Mutations of OCT4 binding sites in SRR1 and SRR2 disrupted enhancer activity in ES cells and NPCs (Miyagi et al., 2004; Tomioka et al., 2002). However, it is not known whether SRR1 and SRR2 are necessary or sufficient to regulate Sox2 expression in ES cells.

In this thesis, I evaluated the functions of a set of high probability enhancer candidates surrounding the Sox2 locus in mouse ES cells, E11.5 embryonic brain, and MEFs. I used the 3C technique to profile chromatin interactions in the 150 kb region containing the Sox2 locus and showed that the chromatin architecture of this region is highly cell-type specific. The distal enhancer cluster located 80-120 kb downstream to Sox2 interacted with the Sox2 promoter-proximal region in ES cells but not embryonic brain or MEFs. On the other hand, embryonic brain displayed a more compact chromatin conformation in the Sox2 promoter-proximal region.
than ES cells or MEFs. MEFs did not show any looping interaction in the region, suggesting that the locus adopted a random conformation. To further validate the functions of the enhancer candidates, I cloned SRR1, SRR2, and eight enhancer candidates into firefly luciferase reporter plasmids and measured luciferase activity in ES cells and MEFs. The results showed enhancer candidate Enh18, Enh107 and Enh111 generated ES cell specific enhancer activity. All three enhancer candidates had significantly higher enhancer activity than SRR1 (\( P < 0.001 \)). Reversing the orientation of the enhancer sequence did not affect the activity of SRR1, SRR2, or Enh18 but decreased the activity of Enh107 and Enh111, although the differences were not significant (\( P > 0.05 \)).

### 4.2 ES cell specific chromatin architecture at the Sox2 locus

The detection of long-range chromatin interactions between the Sox2 promoter and the distal enhancer cluster only in ES cells confirmed my initial hypothesis. The distal enhancer cluster is highly enriched with histone modifications for active enhancers such as H3K4me1 and H3K27ac (Mikkelsen et al., 2007), and factors known to mediate chromatin looping, such as Mediator, cohesin, and CTCF (Goren et al., 2010; Kagey et al., 2010), in ES cells but not in embryonic brains or MEFs. As ES cells differentiate into mesoendodermal or neuroectodermal lineages, the distal enhancer cluster loses H3K4me1 and H3K27ac marks (Xiao et al., 2012), suggesting that it is functional specifically in ES cells. My 3C results agree with recent published data from Phillips-Cremins et al. (2013). They used a high-throughput variant of 3C technology called 3C carbon copy (5C) to detect long-range chromatin interactions near the Sox2 locus in ES cells and NPCs. They found that the Sox2 promoter-proximal region interacted with the region containing the distal enhancer cluster in ES cells but not NPCs. This suggests that the chromatin
surrounding the Sox2 locus undergoes a conformational change as ES cells undergo neural differentiation. However, the resolution of the 5C interaction map was not as high as my 3C interaction map. This is because 5C relies on head-to-head ligation of forward and reverse 5C primers that bind to the antisense and sense strand at the 3’ ends of two restriction fragments. To avoid detection of self-ligated fragments, Phillips-Cremins and colleagues used alternating forward and reverse primers. This caused the 5C map to miss certain interactions. For example, the 5C map showed interactions between the Sox2 locus and the region slightly downstream to the distal enhancer cluster but not the enhancer cluster (Phillips-Cremins et al., 2013). Nevertheless, my results agree with this published data in showing ES cell-specific chromatin looping surrounding the Sox2 locus.

The interaction between the distal enhancer cluster and SRR1 and SRR2 in ES cells was intriguing, because it suggested that these enhancer regions interacted to form a chromatin hub. Previous studies have identified a similar type of chromatin architecture at the β-globin locus (Palstra et al., 2003; Tolhuis et al., 2002). The β-globin LCR interacts with the upstream 5’HS60/62 and the downstream 3’HS1 to form a chromatin complex called active chromatin hub (ACH). The β-globin genes loop in to contact the active chromatin hub only when they become transcriptionally activated. The looping interaction between LCR and 3’HS1 is required for the formation of the active chromatin hub (Hou et al., 2010). It remains to be seen whether SRR1 or SRR2 is required for looping interaction between the Sox2 promoter and the distal enhancer cluster.
4.3 Factors that mediate chromatin looping at the Sox2 locus

The 3C results in this thesis raised the question of which factors are responsible for establishing the cell-type specific chromatic architecture surrounding the Sox2 locus. Recent studies showed that CTCF, Mediator, and cohesin are important architectural proteins that mediate long-range looping interactions (Chen et al., 2008; Dixon et al., 2012; Goren et al., 2010; Handoko et al., 2011; Kagey et al., 2010; Sanyal et al., 2012). Genome-wide studies suggested that cohesin frequently colocalizes with CTCF to mediate chromatin looping (Rubio et al., 2008; Stedman et al., 2008; Wendt et al., 2008). At the T-cell receptor (TCR) α locus (Tcra), CTCF and cohesion co-occupy the TEA promoter and the Eα enhancer. Knockdown of CTCF or cohesin subunit RAD21 disrupted chromatin interaction between TEA promoter and Eα enhancer as well as Tcra transcription and rearrangement (Seitan et al., 2011). At the β-globin locus, knockdown of CTCF or cohesin subunits SMC1 and RAD21 by siRNA disrupted the interaction between HS5 and 3’HS1, suggesting that CTCF and cohesin are both required for chromatin looping (Splinter et al., 2006). Mediator and cohesin frequently co-occupy gene promoters and enhancers in ES cells (Kagey et al., 2010; Phillips-Cremins et al., 2013; Sanyal et al., 2012). Knockdown of subunits of Mediator complex (MED12) or cohesin complex (SMC1, SMC3) disrupted the expression of ES cell-specific genes including Sox2 and induced cell differentiation (Kagey et al., 2010). Indeed, Mediator and cohesin complexes were required for maintaining the chromatin interactome of the NANOG promoter in ES cells (Apostolou et al., 2013).

The distal enhancer cluster downstream to Sox2 is enriched with Mediator and cohesin co-bound sites in ES cells but not NPCs (Kagey et al., 2010; Phillips-Cremins et al., 2013; Sanyal et al., 2012). Also, a prominent CTCF peak is found within Enh109 in the distal enhancer cluster
in ES cells but not NPCs. The Mediator and cohesin co-bound sites and CTCF in the distal enhancer cluster might be responsible for interacting with the Sox2 promoter and proximal enhancers in ES cell (Figure 10). On the other hand, embryonic brain displays a cluster of CTCF peaks located 12.5-22.5 kb downstream to Sox2 (Figure 9). These CTCF sites might interact with the CTCF peaks upstream to the Sox2 locus to loop out the Sox2 locus in embryonic brain but not ES cells (Figure 9). This would explain why in embryonic brain the Sox2 promoter interacted with the proximal regions at higher frequency than in ES cells but did not interact with the distal enhancer cluster (Figure 6). Furthermore, the Sox2 locus in embryonic brain is enriched with H3K27ac only in the region flanked by the proximal CTCF peaks (Figure 2). This suggests that the CTCF peaks set up boundaries for the active chromatin domain in the Sox2 proximal region in embryonic brain.

Interestingly, my 3C results showed that the distal enhancer cluster interacted with the region 12.5-22.5 kb downstream to Sox2 at significantly higher frequency in embryonic brain than MEFs (Figure 7). This chromatin interaction might be due to the interactions between the CTCF located in the distal enhancer cluster and the CTCF located in the 12.5-22.5 kb downstream region in embryonic brain. However, in embryonic brain the distal enhancer cluster is segregated from the Sox2 locus, which is looped out due to the CTCF interaction in the proximal region (Figure 10). Therefore, CTCF play an important role in facilitating the cell-type specific chromatin architecture near the Sox2 locus in embryonic brain and ES cells.
Figure 9: Cell-type specific chromatin architecture surrounding the Sox2 locus. The high probability enhancer candidates are shown in red bars with heights corresponding to the probability score. The previously identified SRR1 and SRR2 enhancers are shown in green bars. The UCSC tracks for CTCF binding peaks in mouse ES cells (Bruce4), MEFs, and E14.5 embryonic brains are displayed. The frequency of interaction between the Sox2 promoter and the surrounding chromatin in mouse ES cells or E11.5 embryonic brains is normalized MEFs. Values are an average of three independent experiments, with each experiment performed in triplicate. Error bars represent standard deviation.
Figure 10: Model for transcriptional regulation of Sox2 in mouse ES cells and neural precursor cells. In mouse ES cells, the distal enhancer cluster 80-120kb downstream to Sox2 interacts with Sox2 and SRR1 and SRR2 enhancers via chromatin looping. The distal enhancer cluster is bound by a large number of ES cell-specific transcription factors (SOX2, OCT4, NANOG, KLF4, ESRRB), effectors of growth signaling (STAT3, SMAD1), transcription coactivators (p300, Mediator), and cohesin and CTCF. In neural precursor cells, CTCF bound sites upstream and downstream to Sox2 interact to loop out the intervening region containing Sox2, SRR1, and SRR2 enhancers. SRR1 and SRR2 can be targeted by neural specific POU transcription factors such as BRN1, BRN2, and OCT6.
4.4 ES cell-specific regulation of Sox2 expression

In this thesis, the identification of chromatin looping between the Sox2 promoter and the distal enhancer cluster sheds light on the transcriptional regulation of Sox2 in ES cells. The distal enhancer cluster is bound by a large enhanceosome consisting of intrinsic and extrinsic regulators of ES cell transcriptional network (Chen et al., 2012; Chen et al., 2008). The intrinsic regulators are TFs that regulate the self-renewal of ES cells: OCT4, SOX2, NANOG, ESRRB, ZFX, and KLF4 (Figure 1). KLF4 is a member of the Kruppel-like factor (KLF) family of zinc finger TFs. It plays an important role in regulating the transcription of the key pluripotency genes, including Nanog, Oct4, and Sox2 (Bourillot and Savatier, 2010). For example, KLF4 colocalizes with KLF2 and KLF5 to the Nanog distal enhancer and the binding of these factors are required for Nanog expression (Jiang et al., 2008). The transcription of Klf4 and Klf5 are induced by LIF/STAT3 signaling (Hall et al., 2009) and Klf4 is also regulated by NANOG and OCT4, suggesting that Klf4 is a part of the autoregulatory feedback loop in ES cells (Bourillot and Savatier, 2010). Although KLF4 has been shown to be required for Sox2 transcription in ES cells (Jiang et al., 2008), it remains unclear how KLF4 regulates Sox2 transcription. Since KLF4 binds to Enh18 as well as the Sox2 distal enhancer cluster in ES cells (Chen et al., 2012; Chen et al., 2008), it might regulate Sox2 transcription through chromatin looping between the distal enhancer cluster and the Sox2 promoter. A recent study by (Wei et al.) (2013) found that KLF4 directly regulated the recruitment of cohesin to the Oct4 distal enhancer. Depletion of KLF4 disrupted the chromatin interaction between Oct4 promoter and enhancer and reduced Oct4 expression (Wei et al., 2013). Therefore, KLF4 might also play a role in recruiting cohesin to the distal enhancer cluster downstream to Sox2. ESRRB belongs to the estrogen-related receptor subfamily of nuclear
orphan receptors. It is required for maintaining the pluripotency state of ES cells (Ivanova et al., 2006; Loh et al., 2006) and it can cooperate with OCT4 and SOX2 to reprogram MEFs into induced pluripotent stem (iPS) cells (Feng et al., 2009). Depletion of Esrrb has been shown to disrupt the expression of pluripotency genes, including Sox2 (Feng et al., 2009). A recent study showed that ESRRC can form heterodimers with SOX2 and bind Esrrb-Sox2 composite motifs (Hutchins et al., 2013) (Hutchinson et al., 1981). Similar to KLF4, ESRRB binds to Enh18 and the distal enhancer cluster (Chen et al., 2012; Chen et al., 2008).

The distal enhancer cluster is also bound by STAT3 and SMAD1, which are downstream effectors of leukemia inhibitory factor (LIF) and bone morphogenetic protein 4 (BMP4) signaling pathways, respectively (Attisano and Wrana, 2002; Matsuda et al., 1999; Niwa et al., 1998). LIF and BMP4 are both required for maintaining the self-renewal of ES cells cultured in serum-free medium, because LIF inhibits mesendodermal differentiation whereas BMP4 inhibits neuroectodermal differentiation (Ying et al., 2003). LIF signaling activates STAT3, which can activate the expression of Klf4 (Hall et al., 2009; Niwa et al., 2009). STAT3 and SMAD1 have been shown to frequently co-occupy enhancers bound by OCT4, SOX2, and NANOG in ES cells (Chen et al., 2008). NANOG directly interacts with SMAD1 and the binding of STAT3 and SMAD1 is dependent on OCT4 (Chen et al., 2008). Therefore, STAT3 and SMAD1 might be recruited to the distal enhancer cluster by OCT4, SOX2, and NANOG in response to LIF signaling. Since KLF4 is also induced by LIF signaling and might play a role in recruiting cohesin to the Sox2 distal enhancer cluster, together these evidence support the notion that Sox2 transcription is regulated by extrinsic growth signaling pathways.
Recently, Young and colleagues classified ES-specific enhancers into two groups: typical-enhancers and super-enhancers (Whyte et al., 2013). In comparison to typical-enhancers, super-enhancers are one magnitude larger in size and enriched with higher levels of H3K4me1, H3K27ac, and Mediator binding. In addition, super-enhancers have stronger enhancer activity and preferentially regulate genes that are highly cell-type specific. Therefore, Young and colleagues suggested that super-enhancers play important roles in controlling a cell-type specific gene expression programs. Young and colleagues identified a super-enhancer in the promoter-proximal region of Sox2 locus, which contains SRR1 and SRR2 enhancers, and a second super-enhancer downstream to Sox2, which overlaps with the distal enhancer cluster I characterized. In my work, I showed that the distal super-enhancer interacted with the proximal super-enhancer to regulate Sox2 expression. More importantly, I found that Enh107 and Enh111 within the distal super-enhancer had significantly higher enhancer activity than SRR1 and SRR2 in mouse ES cells.

4.5 **Accuracy of *in silico* prediction of active enhancers**

Recent studies identified a number of chromatin features associated with active enhancers. Many groups have attempted to use these chromatin features to predict cell-type specific enhancers (Heintzman and Ren, 2009). To date, between 400,000 and 1.4 million putative enhancers have been identified in mammalian genomes based on chromatin features (Consortium et al., 2012). However, relatively few of these putative enhancers have been validated by functional experiments. I tested the enhancer activity of eight high probability enhancer candidates predicted based on canonical enhancer marks such as p300, H3K4me1, MED12, NIPBL (Chen et al., 2012) in mouse ES cells and MEFs. The results showed that four out
of the eight putative enhancers showed enhancer activity in ES cells. This suggests that it is important to perform functional experiments to validate the putative enhancer candidates. However, there was a caveat to my enhancer activity results. The enhancer candidates were tested in the pGL4.23 plasmids, in which the firefly luciferase gene was regulated by a TATA-box minimal promoter. However, the Sox2 promoter lacks a TATA-box but contains a CCAAT box that binds nuclear factor-Y (NF-Y), which plays a role in bridging TFs with the basal transcription machinery (Wiebe et al., 2000). Therefore, those enhancer candidates that did not show enhancer activity in ES cells might require the proper Sox2 promoter to function. Previous studies that tested the luciferase activity of SRR1 and SRR2 in ES cells or NPCs used the thymidine kinase (tk) promoter, which also contains a CCAAT box (Miyagi et al., 2004; Tomioka et al., 2002). Therefore, the enhancer candidates surrounding Sox2 should be tested with either the Sox2 promoter or the tk promoter to determine if there is any synergistic effect between enhancers and promoters.

4.6 Future directions

This thesis provided evidence that the distal enhancer cluster plays an important role in regulating Sox2 transcription in ES cells. However, more experiments can be done to further characterize the function of the distal enhancer cluster. First, the activity of Enh18, Enh95, Enh107, and Enh111 could be tested in NPCs to determine whether they are active only in ES cells. In addition, the distal enhancer cluster could be deleted in ES cells and NPCs to see if it affects Sox2 transcription and self-renewal of either cell type. This thesis also revealed the important role of chromatin architecture in facilitating the cell-type specific regulation of Sox2 transcription, but it remains to be determined which factors are responsible for establishing the
chromatin architecture in ES cells or embryonic brain. I suspect that CTCF, Mediator and cohesin in the distal enhancer cluster might be responsible for facilitating the chromatin interaction between the distal enhancer cluster and Sox2 in ES cells, whereas the CTCF sites in the proximal Sox2 region might be responsible for establishing the compact chromatin conformation in embryonic brain. To test this hypothesis, the CTCF binding site in the distal enhancer cluster could be mutated or deleted to see whether it would disrupt chromatin looping in ES cells. Similarly, the CTCF sites in the proximal region surrounding the Sox2 locus could be deleted to see if it would disrupt the chromatin architecture in embryonic brain. Since Mediator and cohesion do not recognize specific DNA binding motifs, we can knockdown these proteins using shRNA or siRNA.

A recent study found that KLF4 can directly regulate the recruitment of cohesin to enhancers and promote enhancer-promoter interaction (Wei et al., 2013). Since KLF4 is enriched in the distal enhancer cluster downstream to Sox2, it might play a role in recruiting cohesin to the distal enhancer cluster and facilitating the chromatin interaction between the distal enhancer cluster and Sox2 in ES cells. To test this hypothesis, we can knockdown KLF4 by shRNA and use 3C to detect changes in chromatin architecture at the Sox2 locus. Furthermore, it would be interesting to monitor the changes in chromatin architecture of the Sox2 locus as ES cells differentiate towards the neuroectodermal lineage. During in vitro neural differentiation of ES cells, Sox2 expression has been shown to drop on day 3 but climb back up on day 5 (Abranches et al., 2009). This expression change might be due to conformational changes of chromatin surrounding the Sox2 locus that allow different enhancer elements to regulate Sox2 transcription. Finally, the distal enhancer cluster might play additional roles in regulating the
expression of other pluripotency genes. Previous studies have shown that functionally related
genes co-localized to the same transcription factory and that enhancers play important roles in
organizing these transcription hubs (Osborne et al., 2004; Schoenfelder et al., 2010). Therefore,
the distal enhancer cluster might interact with other pluripotency genes in the genome and
organize the transcriptional network in ES cells. To explore this possibility, we can use a variant
of 3C technology called enhanced 4C (e4C) (Osborne et al., 2004), which allows the
identification of all the genomic loci that interact with a bait region, which can be the distal
enhancer cluster downstream to Sox2.
Chapter 5 Summary

The work presented in this thesis highlights the important role of the three-dimensional chromatin architecture in the cell-type specific regulation of Sox2 transcription. Sox2 is expressed throughout embryonic development, first in the pluripotent ES cells of the blastocyst and later in multipotent stem or progenitor cells of all three germ layers (Sarkar and Hochedlinger, 2013). Sox2 plays a dose-dependent role in regulating the cell fate decision of stem cells, but the cell-type specific regulation of Sox2 transcription has been unclear. In this thesis, I found that Sox2 interacts with a cluster of enhancers located 80-120 kb downstream to Sox2 in mouse ES cells but not embryonic brain or MEFs. The difference in chromatin architecture between ES cells and embryonic brain suggests that the Sox2 locus undergoes a conformational change as ES cells differentiate towards the neuroectodermal lineage. The distal enhancer cluster also possesses a robust enhancer activity in ES cells but not MEFs. These results provide evidence that the distal enhancer cluster regulates Sox2 transcription specifically in ES cells. This distal enhancer cluster is the first known cell-type specific enhancer region for Sox2. Future studies could explore the role of external growth signaling in regulating the activity of the distal enhancer cluster.
References


Appendices

Appendix 1: Schematic for chromosome conformation capture (3C) technology

Cells are fixed by formaldehyde and then lysed to release the cross-linked chromatin. The chromatin is then digested with a restriction enzyme and ligated at dilute conditions to facilitate intramolecular ligation between cross-linked fragments. The cross-linking is reversed and ligation products are quantified by real-time quantitative PCR.
Appendix 2: 3C qPCR primers for murine Sox2 locus

List of qPCR primers used for 3C analyses in this thesis. All primers were used at a 3 μM concentration in qPCR reactions.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence 5'-'3'</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sox2Hind3C U1</td>
<td>TTGATTGTGATTGTGGGGTTAAA</td>
</tr>
<tr>
<td>Sox2Hind3C U3</td>
<td>AACTTAGCTTAGGTCCACTGAGTCC</td>
</tr>
<tr>
<td>Sox2Hind3C U4(2)</td>
<td>AGTAGCCTTGCTCTCTTTAGGT</td>
</tr>
<tr>
<td>Sox2Hind3C D6(2)</td>
<td>TTCTCCGTGACCTCACATAA</td>
</tr>
<tr>
<td>Sox2Hind3C D7(2)</td>
<td>AGGCCGAGCATTCACTTTG</td>
</tr>
<tr>
<td>Sox2Hind3C U8</td>
<td>AATCTGCGCTATTGCAGTAAGG</td>
</tr>
<tr>
<td>Sox2Hind3C U10</td>
<td>TCCTGAGCTCCTCTCTGATTCC</td>
</tr>
<tr>
<td>Sox2Hind3C U11(2)</td>
<td>GCCCATCTAATAGAAGCCAC</td>
</tr>
<tr>
<td>Sox2Hind3C U13(2)</td>
<td>CATGGCCTCATATGAGTTTAT</td>
</tr>
<tr>
<td>Sox2Hind3C U14(2)</td>
<td>TTTGATAGGACATTGGATTTGGAA</td>
</tr>
<tr>
<td>Sox2Hind3C U16</td>
<td>TTGCAGAAGACTCACATCTCTT</td>
</tr>
<tr>
<td>Sox2Hind3C U17(2)</td>
<td>GTGCAAGCTAGCAAGCTACATC</td>
</tr>
<tr>
<td>Sox2Hind3C U19(2)</td>
<td>TGTTTTCCACAAAGGAGTTCTCAA</td>
</tr>
<tr>
<td>Sox2Hind3C U23</td>
<td>AGAATAAACCCATGTATGCTCC</td>
</tr>
<tr>
<td>Sox2Hind3C U24</td>
<td>ACTGGCCCTCAGAACACCACCAT</td>
</tr>
<tr>
<td>Sox2Hind3C U26</td>
<td>AAAAAGGAGTGGAGGGCGTAG</td>
</tr>
<tr>
<td>Sox2Hind3C U27</td>
<td>AAGGTGTGCTCATTATCCTCTAATG</td>
</tr>
<tr>
<td>Sox2Hind3C D29(2)</td>
<td>CATCCCTCGTGAACACACCAT</td>
</tr>
<tr>
<td>Sox2Hind3C U29(3)</td>
<td>AATTCATGAAGCTACAGCTGAC</td>
</tr>
<tr>
<td>Sox2Hind3C U30</td>
<td>CTGGGTGCTTTTCTATTCAGA</td>
</tr>
<tr>
<td>Sox2Hind3C U32</td>
<td>TCCAGGGCTAGGAGCAGTTGTATAAT</td>
</tr>
<tr>
<td>Sox2Hind3C U33</td>
<td>TATAGCTCATGTTCGCTGT</td>
</tr>
<tr>
<td>Sox2Hind3C U34</td>
<td>TTTTCTTTACTCACAGGCATT</td>
</tr>
<tr>
<td>Sox2Hind3C U35</td>
<td>TCCCTGACTCACATTATCCCTTGA</td>
</tr>
<tr>
<td>Sox2Hind3C U39</td>
<td>GGGTTTCATAGGTTACTGCAC</td>
</tr>
<tr>
<td>Sox2Hind3C U40</td>
<td>GGGTCTGCAAGGTCAGTTT</td>
</tr>
</tbody>
</table>
Appendix 3: Optimization of the fixation condition for 3C analysis

E14 mouse ES cells were fixed in four different conditions that involved different formaldehyde concentrations and times of incubation. Then, the cells were lysed and the chromatin was digested overnight with *HindIII*. Undigested and digested chromatin was collected before and after *HindIII* digestion. DNA was purified by Phenol-Chloroform extraction and digestion efficiency at 14 restriction sites surrounding the *Sox2* locus were evaluated by qPCR. Data represent average of three qPCR reactions.
Appendix 4: Gel-electrophoresis of qPCR products for Sox2 promoter 3C profile

The qPCR products for Sox2 promoter 3C profile were analyzed on 2% agarose gel. The ladder used was the 100 bp DNA ladder (NEB).

List of primer pairs for top gel, lanes 1-10:
Ladder, Acta2 U2+U4, D7(2)+U1, D7(2)+U39, D7(2)+U3, D7(2)+U4(2), D7(2)+U6, D7(2)+U8, D7(2)+U10, D7(2)+U14, D7(2)+U17(2)

List of primer pairs for bottom gel, lanes 11-20:
Ladder, D7(2)+U21(2), D7(2)+U21(2), D7(2)+U24, D7(2)+U27, D7(2)+D29(2), D7(2)+U30, D7(2)+U32, D7(2)+U33, D7(2)+U34, D7(2)+U35
Appendix 5: Gel-electrophoresis of qPCR products for distal enhancer cluster 3C profile

The qPCR products for the Sox2 distal enhancer 3C profile were analyzed on 2% agarose gel. The ladder used was the 100 bp DNA ladder (NEB).

List of primer pairs for top gel, lanes 1-11:
Acta2 U2+U4, U30+U1, U30+U39, U30+U3, U30+U4(2), U30+U6, U30+D7(2), U30+U8, U30+U10, U30+U11(2), U30+U13(2)

List of primer pairs for top gel, lanes 12-21:
U30+U14(2), U30+U16, U30+U17(2), U30+U19(2), U30+U40, U30+U23, U30+U24, U30+U26, U30+U27, U30+U29(3)
**Appendix 6: QPCR primers for checking HindIII digestion efficiency**

List of qPCR primers used for checking HindIII digestion efficiency in this thesis. All primers were used at a 3 μM concentration in qPCR reactions.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5'→3')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gapdh-ex-in2F</td>
<td>CAAGACCTGCCTAGGCAAC</td>
</tr>
<tr>
<td>Gapdh-ex-in2R</td>
<td>GCCACGCTTAATCTCATTTTCTT</td>
</tr>
<tr>
<td>Sox2exon2F</td>
<td>CAGGGAGTTCGGCAAAGTCT</td>
</tr>
<tr>
<td>Sox2exon2R</td>
<td>TGGACATTGGATTGCCATGT</td>
</tr>
<tr>
<td>Sox2Hind3C U2</td>
<td>GCACAAAGAACGTGTATTGTATT</td>
</tr>
<tr>
<td>Sox2Hind3C D2</td>
<td>CTCTATCCCCACCGTACAGA</td>
</tr>
<tr>
<td>Sox2Hind3C U6</td>
<td>ATATCACCAGGGTTGTCC</td>
</tr>
<tr>
<td>Sox2Hind3C D6</td>
<td>TGTATGTATCTCTAGGGTCTTAATAG</td>
</tr>
<tr>
<td>Sox2Hind3C U8</td>
<td>AATCTGGGCTATTTGCCGTAAGG</td>
</tr>
<tr>
<td>Sox2Hind3C D8</td>
<td>ACGTCCCCGTAGGAAGGTAGA</td>
</tr>
<tr>
<td>Sox2Hind3C U10</td>
<td>TCCCTCAGCTCTCTTGTTTTGA</td>
</tr>
<tr>
<td>Sox2Hind3C D10</td>
<td>AGCCAGATGATAAGGGTCTCAT</td>
</tr>
<tr>
<td>Sox2Hind3C U11</td>
<td>AGAATCTTGAGCTCTACAGGGGAATG</td>
</tr>
<tr>
<td>Sox2Hind3C D11</td>
<td>GCAATAACCTGACATGAAGTTCCAA</td>
</tr>
<tr>
<td>Sox2Hind3C U13</td>
<td>CATGGCCTCAATGTGTTTAT</td>
</tr>
<tr>
<td>Sox2Hind3C D13</td>
<td>CATGAACCCAGGGTATGTAA</td>
</tr>
<tr>
<td>Sox2Hind3C U15</td>
<td>GCTTTAGCTCAGCATAAGGATGACT</td>
</tr>
<tr>
<td>Sox2Hind3C D15</td>
<td>TAGAGAGCGAGTTGCTGACAG</td>
</tr>
<tr>
<td>Sox2Hind3C U18</td>
<td>CTAGCGCTTTAGACCAGGGTAC</td>
</tr>
<tr>
<td>Sox2Hind3C D18</td>
<td>GAACTTCTGGTTTCAGCTGT</td>
</tr>
<tr>
<td>Sox2Hind3C U27</td>
<td>AAGGGTGCTCATATATCATCTAATG</td>
</tr>
<tr>
<td>Sox2Hind3C D27</td>
<td>AACTCTCACACTTCATTTCTTTTG</td>
</tr>
<tr>
<td>Sox2Hind3C U28</td>
<td>GGGAAGGTAACTTCACCCTCCAA</td>
</tr>
<tr>
<td>Sox2Hind3C D28</td>
<td>AGTATGAATCAGGGTTTCTTGATG</td>
</tr>
<tr>
<td>Sox2Hind3C U29</td>
<td>CCTCAAGTGCTCGTCAAG</td>
</tr>
<tr>
<td>Sox2Hind3C D29</td>
<td>CCTCACAATCTAAGGAGCTTC</td>
</tr>
<tr>
<td>Sox2Hind3C U30</td>
<td>CTGGCTTCTACTTTCTTGCA</td>
</tr>
<tr>
<td>Sox2Hind3C U30(2)</td>
<td>TTAAAGTGGAGGGGCGATTTG</td>
</tr>
<tr>
<td>Sox2Hind3C U33</td>
<td>TATAGCTTATGGTCTGCTGTG</td>
</tr>
<tr>
<td>Sox2Hind3C D33</td>
<td>GCACCTGAGAGATCTCTACAAGAAG</td>
</tr>
<tr>
<td>Sox2Hind3C U34</td>
<td>TTTCTTTTTACTCAAGGGCTTGA</td>
</tr>
<tr>
<td>Sox2Hind3C D34</td>
<td>GAAGTGAGCGCAACTTTAGGT</td>
</tr>
</tbody>
</table>
Appendix 7: Gel-electrophoresis of qPCR products for HindIII digestion efficiency check

The qPCR products for HindIII digestion efficiency check were analyzed on 2% agarose gel. The ladder used was the 100 bp DNA ladder (NEB).

List of primer pairs for top gel, lanes 2-17:
Step 1) Enhancer candidates are PCR amplified from BAC (RP23-274P9) and ligated into pJET-1.2/blunt vector (Fermentas). Step 2) enhancer sequence and orientation are verified by restriction digestion and DNA sequencing. Step 3) enhancers in pJET-1.2 vector are PCR amplified by primers with 15 bp overhangs that are homologous to the NotI site in pGL4.23 [luc2/minP] vector. Step 4) PCR amplified products are cloned into pGL4.23 [luc2/minP] by In-Fusion® cloning reaction (Clontech). The enhancer sequence and orientation in pGL4.23 are verified by DNA sequencing.
Appendix 9: Schematic for dual luciferase reporter assay

The Sox2 enhancer candidates were cloned into pGL4.23 [luc2/minP] plasmids. Each pGL4.23-enhancer plasmid is mixed with pGL4.75 [Rhluc/CMV] plasmid at 50:1 molar ratio. The concentration of pGL4.23-enhancer plasmid in the mixture is 10 µg/µl. ES cells and MEFs were seeded on 96-well plates at 1x10^4 cells/well and were grown overnight in a 37°C incubator. Before transfection, the plasmid mixture was diluted 5X in ES medium. The Lipofectamine 2000 reagent was diluted in ES medium at 4:46 (v/v) ratio. After 5 min incubation at room temperature, 50 µl of diluted lipofectamine was mixed with 50 µl of diluted plasmid and was incubated at room temperature for 30 min. Then 20 µl of the DNA-lipofectamine mix was added to each well of the 96-well plates. The medium was changed after 24 hrs of incubation. Transfection for each enhancer plasmid was done in triplicate. Dual-luciferase reporter assays (Promega) were performed 48 hrs after transfection. Cells were lysed in Passive Lysis Buffer and 20 µl of lysate from each well were transferred into a white 96-well plate for measurement. Dual luciferase reporter assays were performed on Fluoroskan Ascent FL plate reader using the DLR 1sec protocol.
## Appendix 10: Primers for cloning enhancer candidates surrounding the Sox2 locus

List of primers used for cloning the enhancer candidates surrounding the murine Sox2 locus. All primers were used at a 10 μM concentration in PCR reactions.

<table>
<thead>
<tr>
<th>Primer Name</th>
<th>Sequence (5'-3')</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sox2-SRR1-F</td>
<td>GCCATTCCTAAGCCTCTTAAT</td>
<td>For cloning SRR1</td>
</tr>
<tr>
<td>Sox2-SRR1-R</td>
<td>AGGCTTAACGTCCAAGCATAAC</td>
<td></td>
</tr>
<tr>
<td>Sox2-SRR2-F</td>
<td>TCCGTTCTCGGTCATCTTC</td>
<td>For cloning SRR2</td>
</tr>
<tr>
<td>Sox2-SRR2-R</td>
<td>GGGTAGCGTCGACTTTGTTAC</td>
<td></td>
</tr>
<tr>
<td>Sox2-Enh-R11-F</td>
<td>GCAGTGGTTGGAATTGGAAGT</td>
<td>For cloning Enh18</td>
</tr>
<tr>
<td>Sox2-Enh-R11-R</td>
<td>GACCAAGATTCTATGCACATTG</td>
<td></td>
</tr>
<tr>
<td>Sox2-Enh-R27(1)-F</td>
<td>TATCAGCCAACCAATTTGAGG</td>
<td>For cloning Enh85</td>
</tr>
<tr>
<td>Sox2-Enh-R27(1)-R</td>
<td>AGCCAGGCCTCAATCCTCATAC</td>
<td></td>
</tr>
<tr>
<td>Sox2-Enh-R27(2)-F</td>
<td>GATATTGGCATTGTCCCAAATG</td>
<td>For cloning Enh95</td>
</tr>
<tr>
<td>Sox2-Enh-R27(2)-R</td>
<td>TACAGCTGGACCTGGATATTGC</td>
<td></td>
</tr>
<tr>
<td>Sox2-Enh-R30(1)-F</td>
<td>GGCACCTTCAGTCCTGTA</td>
<td>For cloning Enh106</td>
</tr>
<tr>
<td>Sox2-Enh-R30(1)-R</td>
<td>TATAGCGAGGAGGGGCACCG</td>
<td></td>
</tr>
<tr>
<td>Sox2-Enh-R30(2)-F</td>
<td>AATCATCAGCTTGGCAAAGG</td>
<td>For cloning Enh107</td>
</tr>
<tr>
<td>Sox2-Enh-R30(2)-R</td>
<td>CCTGGCATGAATGGTCTTTA</td>
<td></td>
</tr>
<tr>
<td>Sox2-Enh-R30(3)-F</td>
<td>ACAGGGTTGGAGGGGAAGTG</td>
<td>For cloning Enh109</td>
</tr>
<tr>
<td>Sox2-Enh-R30(3)-R</td>
<td>TCCTGTACCCACCAAAATAA</td>
<td></td>
</tr>
<tr>
<td>Sox2-Enh-R30(4)-F</td>
<td>GTGGGGTACAAGGAGCATGGAAA</td>
<td>For cloning Enh111</td>
</tr>
<tr>
<td>Sox2-Enh-R30(4)-R</td>
<td>CCTCATCTGCAAATCTCTGG</td>
<td></td>
</tr>
<tr>
<td>Sox2-Desert-F</td>
<td>GCCAGGGATTCTAAAAAGC</td>
<td>For cloning desert</td>
</tr>
<tr>
<td>Sox2-Desert-R</td>
<td>GAGGTCCTGCTTGAATCTCTTA</td>
<td></td>
</tr>
<tr>
<td>Fusion-NotI-1F</td>
<td>GGTCTGACAGCGGCGGCACCTGTGCCTGAACACCATAC</td>
<td>For cloning enhancers from pJET1.2 into NotI site of pGL4.23 (preserving orientation in pJET1.2)</td>
</tr>
<tr>
<td>Fusion-NotI-1R</td>
<td>TTAGCATTGCGCGCGAGAGTCGATTGCGCCAAGAAAC</td>
<td></td>
</tr>
<tr>
<td>Fus-NotI-rev-1F</td>
<td>GGTCTGACAGCGGCGCGAGAGTCGATTGCGCCAAGAAAC</td>
<td>For cloning enhancers from pJET1.2 into NotI site of pGL4.23 (reversing orientation in pJET1.2)</td>
</tr>
<tr>
<td>Fus-NotI-rev-1R</td>
<td>TTAGCATTGCGCGCGACTTGTGCCTGAACACCATAC</td>
<td></td>
</tr>
<tr>
<td>pGL4.23-NotI-F1</td>
<td>AAGGGATTTTGGTCAATGAGT</td>
<td>For sequencing enhancers cloned into NotI site of pGL4.23</td>
</tr>
<tr>
<td>pGL4.23-NotI-R1</td>
<td>GCCACTATGGACGAAACGAAT</td>
<td></td>
</tr>
<tr>
<td>pGL4.23-prom F1</td>
<td>ctaactggccgtctctgag</td>
<td>For sequencing promoter in pGL4.23</td>
</tr>
<tr>
<td>pGL4.23-prom R1</td>
<td>ggccctcttaatgtttttg</td>
<td></td>
</tr>
</tbody>
</table>
Appendix 11: Plasmids containing enhancer candidates

Below is a list of plasmids constructed in this thesis. The pJet-1.2 vector is a blunt-end cloning vector (Fermentas). The pGL4.23 vector contains the firefly luciferase gene *luc2* regulated by a minimal promoter. In pGL4.23, enhancers were cloned into the *NotI* site located 1344 bp downstream to *luc2* in the pGL4.23 vector.

<table>
<thead>
<tr>
<th>Plasmid Number</th>
<th>Backbone</th>
<th>Cloning site</th>
<th>Description</th>
<th>Orientation of insert</th>
</tr>
</thead>
<tbody>
<tr>
<td>P0006</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-30(1) (Enh106)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0007</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-30(1) (Enh106)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0008</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-30(4) (Enh111)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0009</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-31 (Enh114)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0010</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-31 (Enh114)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0011</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-27(1) (Enh85)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0014</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-27(1) (Enh85)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0021</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-27(1) (Enh85)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0027</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-11 (Enh18)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0031</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 desert</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0035</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 desert</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0039</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 desert</td>
<td>Forward</td>
</tr>
<tr>
<td>P0043</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-27(1) (Enh85)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0047</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-27(1) (Enh85)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0051</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-11 (Enh18)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0055</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-11 (Enh18)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0063</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-30(1) (Enh106)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0083</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-30(4) (Enh111)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0108</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-30(3) (Enh109)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0123</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-30(4) (Enh111)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0133</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-27(2) (Enh95)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0134</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-27(2) (Enh95)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0140</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-31 (Enh114)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0162</td>
<td>pJet-1.2</td>
<td>EcoRV</td>
<td>Sox2 Enh-30(2) (Enh107)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0184</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-30(2) (Enh107)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0205</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-30(3) (Enh109)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0210</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-27(2) (Enh95)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0216</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-27(2) (Enh95)</td>
<td>Reverse</td>
</tr>
<tr>
<td>P0222</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-30(1) (Enh106)</td>
<td>Forward</td>
</tr>
<tr>
<td>P0231</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 desert</td>
<td>Forward</td>
</tr>
<tr>
<td>P0233</td>
<td>pGL4.23</td>
<td>NotI</td>
<td>Sox2 Enh-30(4) (Enh111)</td>
<td>Forward</td>
</tr>
</tbody>
</table>
Appendix 12: Diagnostic digest of luciferase plasmids containing enhancer candidates

The luciferase reporter plasmids used in this thesis were digested by NotI restriction enzyme and analyzed on a 2% agarose gel. The ladder used was the 2-Log. The top band in lane 1-22 represents the pGL4.23 backbone (4283 bp in size). The band in lane 25 represents pGL4.75 vector (4281 bp in size).

Description of inserts in the plasmids for top gel, Lane 1-13:
SRR1-Forward, SRR1-Reverse, SRR2-Forward, SRR2-Reverse, Enh18-Forward, Enh18-Reverse,
Enh85-Forward, Enh85-Reverse, Enh95-Forward, Enh95-Reverse, Enh106-Forward, Enh106-Reverse, Enh107-Forward

Description of inserts in the plasmids for lower gel, Land 14-23:
Enh109-Forward, Enh109-Reverse, Enh111-Forward, Enh111-Reverse, Enh114-Forward,
Enh114-Reverse, Desert-Forward, Desert-Reverse, empty pGL4.23, empty pGL4.75