Function of Sox2 as a transcriptional repressor

By

Yu-Ru Liu, M. Sc

Thesis submitted to the University of Nottingham for the degree of Doctor of Philosophy

September 2011

Abstract

Sox2 is one of the earliest known transcription factors to be expressed during development of the nervous system (Rex *et al.,* 1997; Silvia Brunelli, 2003; Wang et al., 2006b; Dee et al., 2008). Ectodermal cells expressing Sox2 have the potential to differentiate into nerve cells. Cells expressing Sox2 are specified to a neural fate during neural induction. Sox2 belongs to the SoxB1 family, comprising Sox1, Sox2 and Sox3, which are generally considered to activate specific target genes, whereas, the SoxB2 group, Sox14 and Sox21, act as transcriptional repressors (Uchikawa, Kamachi, & Kondoh, 1999). However, Sox2 has also been demonstrated to act as a repressor (Kopp et al., 2008) which implies that Sox2 could have a dual-function *in vivo*. Previous studies indicated that the HMG box-containing protein, Tcf/Lef, interacts with the transcriptional co-repressor, Groucho (Helen Brantjes, 2001). We therefore set out to determine if interaction with the Groucho co-repressor could also explain the repressor ability of Sox2.

In this study, we have examined the interaction between Sox2 and Groucho using nuclear translocation, yeast-two-hybrid and co-immunoprecipitation assays. The data suggest that Sox2 interacts with Groucho through a C-terminal, engrailed-like motif. The effect of Groucho on Sox2 function was measured using a luciferase reporter assay. The transcriptional activation activity of Sox2 was repressed after co-expressing with Groucho. To address the biological function of Sox2-Groucho interaction, a loss-of-repressor-function mutant of Sox2 was created by point mutating the essential engrailed-like motif. Analysis in Zebrafish embryos indicated that the of loss-of-repressor-function mutant of Sox2 ($Sox2^{LQY/AAA}$) lost the ability to repress the expression of *chordin*. In

human neural stem cells, Affymetrix arrays revealed that 676 genes were activated and 786 genes were repressed by Sox2 overexpression. Within the genes that were repressed by Sox2, approximatly 7% were less repressed by $Sox2^{LQY/AAA}$. Together, these data suggest that Sox2 functions not only as a transcriptional activator but also as a repressor through interacting with the co-repressor Groucho. However, because only 7% of the repressed genes were affected by the $Sox2^{LQY/AAA}$ mutant, this suggests that there are other mechanisms involved in Sox2 transcriptional repressor function.

Publications

Yu-Huan Shih, Cheng-Liang Kuo, Caroline S. Hirst, Chris T. Dee, **Yu-Ru Liu**, Zulfiqar Ali Laghari and Paul J. Scotting (2010). SoxB1 transcription factors restrict organizer gene expression by repressing multiple events downstream of Wnt signalling. Development. 137, 2671-2681.

Acknowledgments

First of all, I would like to thank my supervisor Dr. Paul J. Scotting for giving me the opportunity to work in his lab and for his valuable guidance and optimistic attitude throughout my studies. Thanks also go to Professor Jane Hewitt for her useful suggestions and encouragement.

Big thank-you goes to Zulfiquar for his endless support and friendship. To Dr. Jennie N. Jeyapalan, thanks for her time in patiently listening and encouraging me to pursue a career in science. I would also like to thank Ivan, Peggy, Sarah and Thomas for soothing me from homesickness. Thanks to Elaine and Magda for their technical support. To Chris Tan, Chris Dee and Dylan for the interesting scientific chats. Caz, Dzul, Michelle and all the lab members, thanks for sharing enthusiasm and team spirit. Caro, Tom, Paul Goodwin and all the master students worked with me, thanks for your help and friendship.

A special thank goes to Remy and my families for balancing my life outside work. Belinda, Yi-Fang, Zora and all my friends, thanks for the accompaniment and support.

I would like to dedicate this thesis to my mother in heaven. I could not fulfil my dream without her.

List of Contents

Table of Figures

List of Tables

Chapter 1 Introduction

1.1 SoxB1 in early neural development

1.1.1 Neural induction

During embryonic development, neural induction is one of the earliest events. After fertilization is completed, a series of cell divisions occurs, this process is called cleavage. Later, a fluid-filled cavity is formed and the embryo is now called a blastula (Figure 1.1.1 A). Gastrulation is the stage of embryonic development which initiates the establishment of the three germ layers – ectoderm, endoderm and mesoderm (Figure 1.1.1 B, C). The nerve system is derived from the dorsal ectoderm called the "neuroectoderm", whereas the ventral ectoderm will generate epidermis. This decision to become neural or epidermal is through the process of neural induction - formation of the neural plate, and neurulation (establishment of the neural tube) (Figure 1.1.2). Signals that originate from the midline mesoderm cause thickening of adjacent ectoderm to form the neural plate. The lateral edges keep rolling to form neural folds that fuse dorsally to form the neural tube (Figure 1.1.2). During the formation of the neural tube, neural crest cells migrate between the epidermis and the neural tube. The rostral region of the neural tube progresses to become the brain whereas the caudal region gives rise to the spinal cord. The proliferation rate is not uniform along the neural tube, resulting in the formation of three brain vesicles: the forebrain (prosencephalon), the midbrain (mesencephalon), and the hindbrain (rhombencephalon).

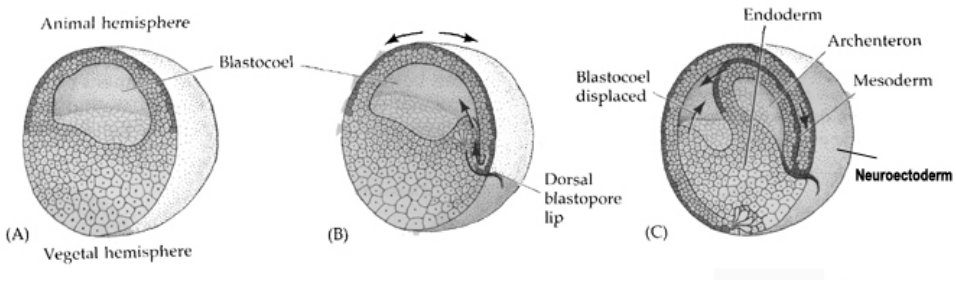


Figure 1.1.1 Gastrulation of *Xenopus* embryo

(A) Blastula. (B) Gastrulation start from the dorsal lip of the blastopore. Cell movements are indicated with arrows. (C) The formation of the three germ layers (From Gilbert, 2000).

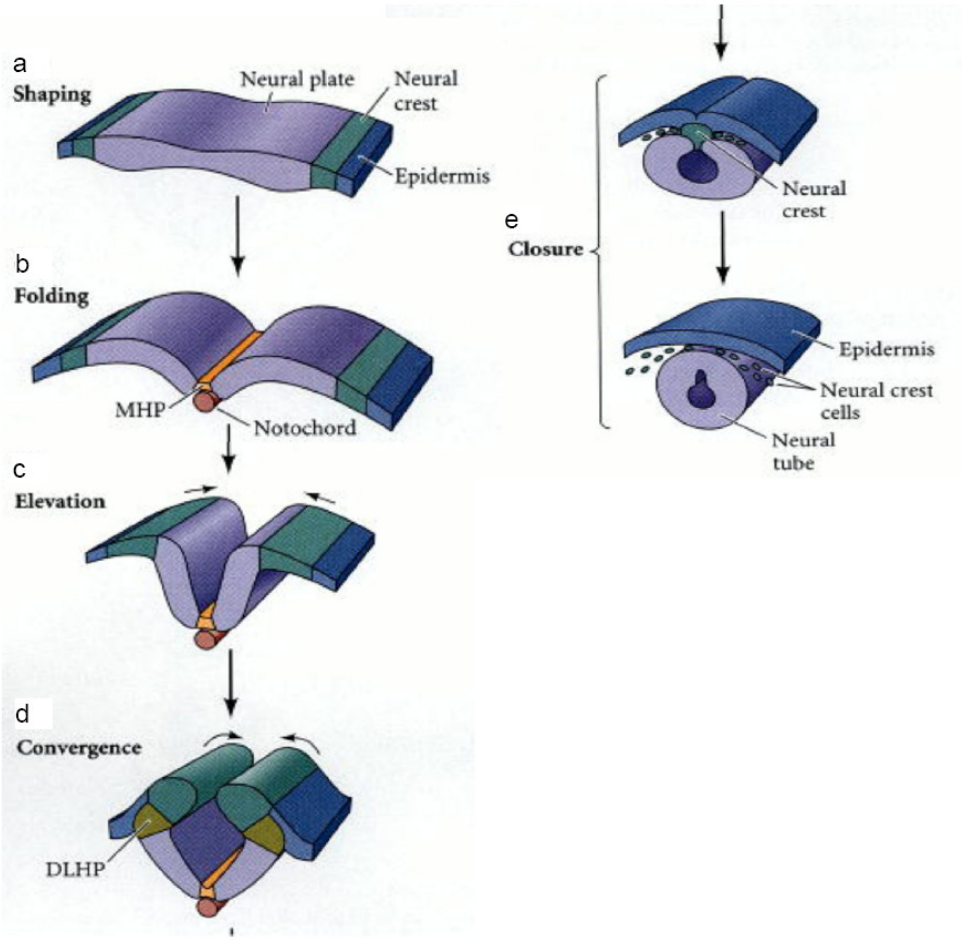


Figure 1.1.2 Illustration of neural tube formation.

(a) Neural plate can be distinguished in the midline of the dorsal ectoderm. (b) When the neural ectoderm starts folding, the medial neural hinge point (MHP) cells associate with the notochord whilst the adjacent ectoderm rolls up. (c) The epidermis keep moving toward the dorsal midline causing elevation of the neural fold. (d) The forming of wedge-shaped dorsolateral hinge point (DLHP) cells results in the convergence of the neural fold. (e) The neural tube is formed after closure of neural fold. (From Gilbert, 2000)

Neural induction has been studied for many decades, and a number of molecules have been identified that play central roles in neural fate specification and commitment. Bone morphogenetic proteins (BMPs), which belong to the TGFβ superfamily, are the most thoroughly studied of these signalling molecules (Figure 1.1.3). The ventral expression of BMPs induces epidermal (non-neural) fate, whereas the expression of BMP antagonists, such as noggin (Smith and Harland, 1992), chordin (Sasai et al., 1994) and follistatin (Hemmati-Brivanlou et al., 1994), has been shown to be involved in the initiation of neural induction. The mechanism behind the contrary functions of these molecules is that the antagonists of BMPs interfere with BMP signalling by directly binding to BMPs or blocking the phosphorylation of the Smad complex (Massague and Chen, 2000; Onichtchouk et al., 1999). Hence, these BMP antagonists are defined as neural inducers.

There are two types of BMP receptors (Type I and Type II), both are transmembrane proteins. When BMP/typeI and typeII receptors form a ternary complex, signal transduction occurs via the serine/threonine kinase activity of the receptors (Figure 1.1.3) (Munoz-Sanjuan and Brivanlou, 2002). BMP signalling result in phosphorylation of the intracellular signal transducer, Smad complex, which translocates into the nucleus after activation and triggers the transcription of downstream genes by cooperating with other transcription factors (Bubnoff and Cho, 2001; Massague and Chen, 2000). BMP4 inhibits neural differentiation and induces epidermis (Wilson and Hemmati-Brivanlou, 1995). The inhibition of BMP signalling triggers the expression of a specific set of transcription factors, including Sox family proteins, and subsequently induces neural fate.

Following neural induction, the cells of the neural ectoderm form neural progenitors later giving rise to oligodendrocytes and astrocytes through gliogenesis or neurons through neurogenesis.

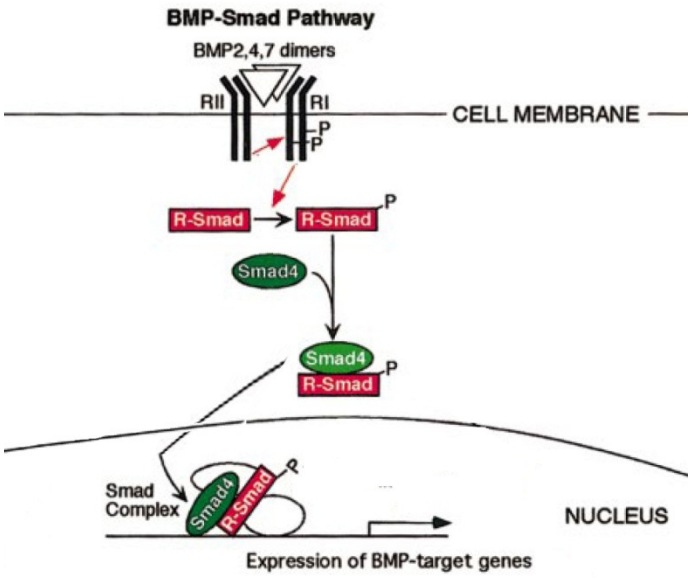


Figure 1.1.3 BMP signalling

Binding of the ligand (BMP2, 4, 7) with the typeI (RI) and typeII (RII) receptors leads to a series of phosphorylation events on the type I receptor and appropriate R-Smads (Smad1, 5, 8). This results in the formation of a Smad complex with co-Smad (Smad4) and causes the complex to relocate the nucleus. This allows the Smad complex to activate or repress target genes (From Bubnoff and Cho, 2001).

1.1.2 Transcription factors

The genome contains the complete information of all genes, but not all of genes are expressed in all cells. Different genes are expressed in different tissues and different development stages. Transcription factors play a key role ensuring the right genes are expressed at the right time.

The term “transcription factor” broadly refers to any protein that has the ability to regulate gene transcription in cells. Generally, transcription factors regulate gene transcription by directly binding to specific DNA regions called “regulatory elements” through their DNA-binding domains, affecting chromatin structure or recruiting cofactors to the target genes (Lee and Young, 2000). Generally, according to their different DNA-binding domains, transcription factors have been categorized into five classes: basic domains, zinc-coordinating DNA-binding domains, helix-turn-helix, beta-scaffold factors with minor groove contacts, and other transcription factors (Matys et al., 2006).

Transcription factors can also be divided into two classes according to their function on target genes: transcriptional activators and repressors. The transcriptional activators assemble enhanceosomes that contain appropriate regulators through protein-protein interaction. Chromatin-modifying complexes such as Swi/Snf can be recruited by transcriptional activators, which loosen the structure of the chromatin. On the other hand, the transcription initiation apparatus that contains RNA polymerase II, can also be recruited by transcriptional activators. The mechanisms of transcriptional repression are less well understood. To date, two possible hypotheses of transcriptional repression have been proposed. One is that a repressor interacts

with an activator and interferes with the binding of the activator on target promoters. The other is that repressors bind specific DNA sites and recruit other factors to block transcription (Hanna-Rose and Hansen, 1996; Johnson, 1995). The balance between transcriptional activation and repression is important to regulate gene expression *in vivo* (Lee and Young, 2000). During neural induction, transcription factors play an important role in responding to the differential signals such as BMPs and turn on/off the transcription of appropriate genes for neural fate commitment.

1.1.3 Sox family

Sox family members are highly conserved transcription factors in both vertebrates and invertebrates. Almost 20 years ago, the Sox (Sry-related high-mobility-group box) gene family was first identified. To date, the large family of Sox factors has more than 30 members, which share a similar DNA-binding domain – the HMG (high-mobility-group) domain. The large Sox family has been categorized into 10 subgroups according to similarity both within and outside of the HMG box domain (Figure 1.1.4). The HMG domain is composed of 79 amino acids and exhibits a preference for binding to the variant linear DNA sequence ($^{A}/_{T}$ $^{A}/_{T}$ CAA$^{A}/_{T}$G) in the minor groove (Laudet et al., 1993). The binding of the L-shaped HMG domain forces the DNA to bend significantly (Lefebvre et al., 2007). The Sox family proteins contain a non-canonical HMG domain, which evolved from the canonical HMG domain found in SRY (the sex determining gene on the Y chromosome). Although the identity of the HMG domain of the Sox family and SRY can be as low as 50%, the ability to alter DNA conformation is conserved (Murphy et al., 1999).

Within the Sox family proteins, sequences are quite variable except within the HMG domain.

Sox proteins have been found to function in many tissues and developmental processes. For example during embryonic development, Sox9 is expressed in the embryonic gonad and has a role in sex determination (Kent et al., 1996). In addition, Sox2, together with Oct4, was demonstrated to regulate FGF4 and osteopontin which play roles in early embryogenesis (Botquin et al., 1998; Yuan et al., 1995). As for neural development, Sox2 has been defined as one of the earliest pan-neural markers and is known to maintain the multipotency of neural stem cell (Bylund et al., 2003; Uwanogho et al., 1995). On the other hand, Sox10 is expressed in the neural crest and contributes to peripheral nervous system development (Kuhlbrodt et al., 1998). In lens development, all of the SoxB1 subgroup members (Sox1-3) were shown to stimulate δ1-crystallin through binding to the DC5 enhancer (Kamachi et al., 1995). The differentiation of optic cup progenitors is also regulated by Sox2 and Pax6 (Matsushima et al., 2011). In other tissues, Sox9 mutations also cause defects of skeletal structure in human (Südbeck et al., 1996). Finally, the B-cells of Sox4 knockout mice are blocked in a pro-B-cell stage (Urbánek et al., 1994). The correct function of Sox2 is critical in early embryonic development. Mutations in the Sox2 gene cause anophthalmia, microphthalmia and anomalies in brain, pituitary, genitourinary and gastresophageal (Reis et al., 2010). Recently, Sox2 has been found to be related to several cancers as an oncogene, such as lung squamous cell carcinomas, glioblastoma, gastric carcinomas and breast cancer (Gangemi et al., 2008; Lengerke et al., 2011; Lu

et al., 2010; Zhang et al., 2010a). These examples show the importance and broad functions of Sox family proteins in vertebrates.

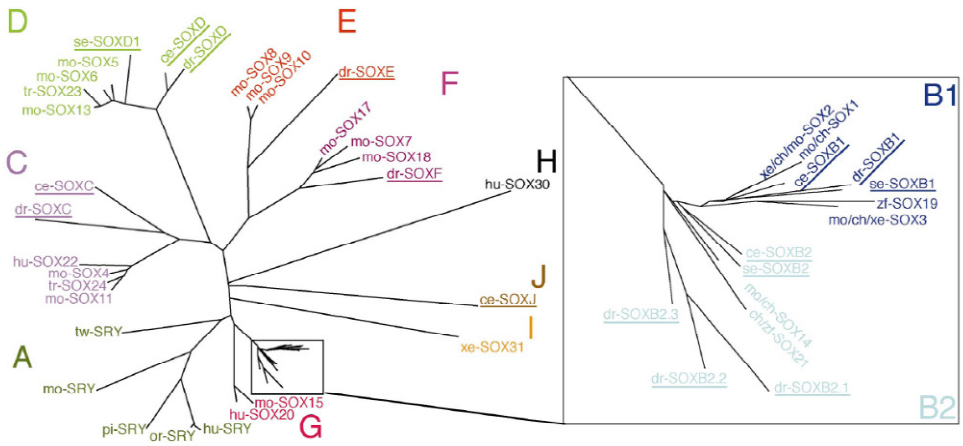


Figure 1.1.4 Phylogenical categorisation of Sox genes.

The Sox HMG domains were analysed by distance method FITCH (GCG). Abbreviations: ce, nematode, *Caenorhabditis elegans*; ch, chicken, *Gallus gallus*; dr, fruit-fly, *Drosophila melanogaster*; hu, human, *Homo sapiens*; mo, mouse, *Mus musculus*; pi, pig, *Sus scrofa*; tw, tammar wallaby (marsupial), *Macropus eugenii*; tr, rainbow trout, *Oncorhynchus mykiss*; se, sea urchin, *Strongylocentrotus purpuratus*; xe, frog, *Xenopus laevis*; zf, zebrafish, *Danio rerio*. (From Bowles et al., 2000)

1.1.4 SoxB1 subgroup

The expression and function of the SoxB subgroup proteins in neural progenitors are critical during neural induction. Research on the SoxB1 subgroup (Sox1, Sox2 and Sox3) has focused on transcriptional activation, whereas the SoxB2 subgroup factors (Sox14 and Sox 21) have been shown to exhibit transcriptional repression (Uchikawa et al., 1999). Within SoxB1 subgroup, high sequence conservation can be seen not only in the HMG domain but also in the C-terminus (Figure 1.1.5).

In addition to DNA-binding ability, the HMG domains of all Sox proteins have nuclear localization signals (NLS) that allows Sox proteins to be imported into

the nucleus to perform as transcription factors. The NLS sequences of Sox proteins have been identified in various parts of the HMG domain (Poulat et al., 1995; Sudbeck and Scherer, 1997). Importin β has been shown to be responsible for nuclear import of SRY through the NLS that is present in the carboxy-terminus of the HMG domain (Forwood et al., 2001). According to the conservation of the NLS of Sox proteins, nuclear transportation via importin β might be common to all Sox proteins.

The transcriptional activation domain of SoxB1 subgroup members has been localized broadly to the C-terminus of the protein, while the N-terminal DNA-binding domain appears to mainly function to recognize a specific binding site on the promoter of target genes (Uchikawa et al., 1999). It has been suggested that the HMG domain DNA-binding ability might not be strong enough to maintain binding. Therefore, the HMG box containing proteins are often found to act with partner factors (Kamachi et al., 2000). For example, Sox2 was demonstrated to function together with Oct4 (containing the DNA-binding domain POU) and a bipartite consensus binding sequence of HMG/POU factor has been identified (Figure 1.1.6) (Rizzino, 2009).

Transcriptional activators trigger the transcription machinery to initiate expression of their target genes. Many factors such as chromatin remodelers and co-activators are recruited to their appropriate locations by transcriptional activators. The transcriptional activation function of SoxB1 proteins might be mediated by the recruitment of such co-activators. The co-activator p300 was shown to be recruited by Sox2 to activate the transcription of FGF4 (Nowling et al., 2003). It is known that p300 activates transcription through the basal transcription machinery and chromatin remodelling (Ogryzko et al., 1996).

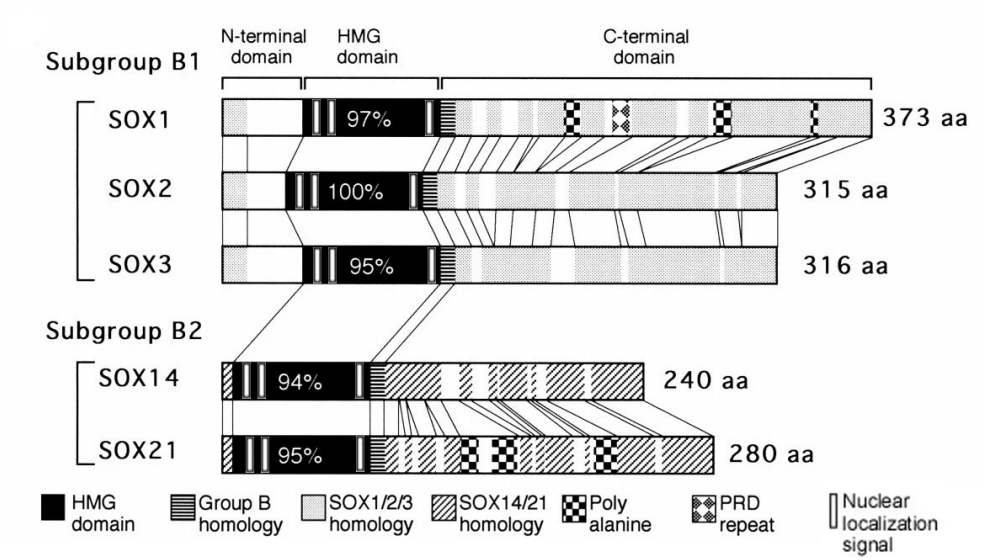


Figure 1.1.5 Conserved domains of SoxB1 subgroup factors.

The percentage of identity within the HMG domain of the chicken SoxB1 subgroup is shown in comparison with the Sox2 HMG domain. Group B homology: SoxB1 and SoxB2 conserved region. SOX 1/2/3 homology: the conserved domain among SoxB1 genes. Polyalanine, three amino acids (PRD) repeats, nuclear localization signal and SUMOylation site are also indicated (from Uchikawa et al., 1999).

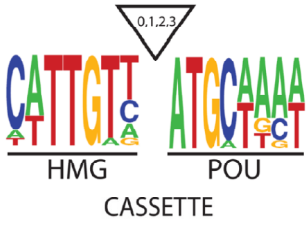


Figure 1.1.6 Consensus DNA binding sequence of HMG/POU cassette.

ChIP-chip and ChIP-PET studies has identified a consensus sequence, the HMG/POU cassette, which provides the binding sites for Sox2 and Oct4 (Rizzino, 2009). The interval between the two binding sites can be from 0 to 3 nucleotides (from Rizzino, 2009).

1.1.5 Sox B1 genes in neural development

The expression of SoxB1 genes is almost ubiquitous in neural progenitors. They function as the earliest transcriptional activators and play a central role in the determination of neural fate. However, the timing of appearance of SoxB1 group members is different in various species. In mouse embryos, Sox2 and

Sox3 are expressed pan-ectodermally initially, subsequently Sox1 expression and neural induction occur when Sox2 and Sox3 also become restricted to the neural ectoderm (Collignon et al., 1996; Wood and Episkopou, 1999) (Figure 1.1.6). In contrast, in chick embryos it is Sox3 that is expressed throughout the ectoderm at first, followed by a dramatic increase of Sox2 expression in the neural ectoderm (Okuda et al., 2006) (Figure 1.1.7). The same scenario has also been observed in zebrafish embryos (Okuda et al., 2006).

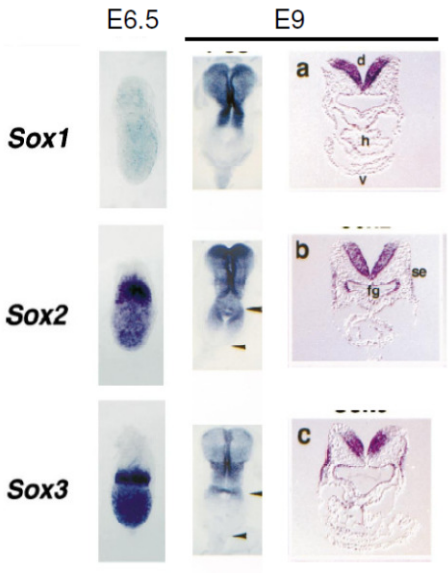


Figure 1.1.7 SoxB1 family expression in early mouse embryo.

Whole mount *in situ* hybridization of Sox1, Sox2 and Sox3 on 6.5 dpc (E6.5) and 9 dpc (E9) mouse embryos is shown. At E6.5, Sox1 was not detectable whereas Sox2 and Sox3 were expressed pan-ectodermally. At E9, all three members of SoxB1 subgroup were expressed in neural ectorderm. Key: d,dorsal; h, heart; v, ventral; fg, foregut; se, surface ectoderm (from Wood and Episkopou, 1999).

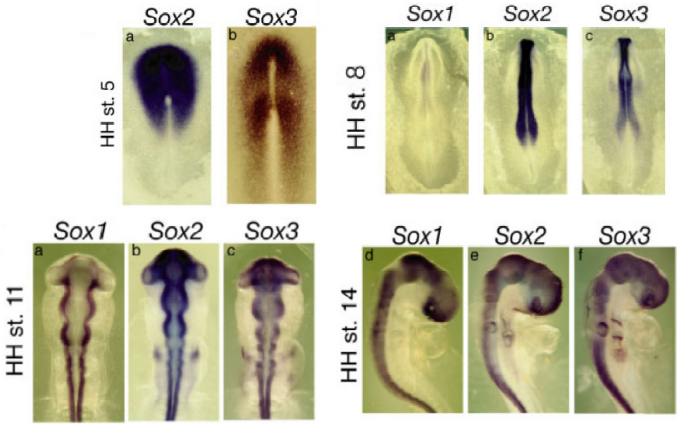


Figure 1.1.8 SoxB1 family expression in early chicken embryo.

Whole mount *in situ* hybridization of Sox1, Sox2 and Sox3 on Hamburger and Hamilton stage 5, 8, 11 and 14 (HH st.5, 8, 11 and 14) chicken embryos is shown. At HH st. 5, Sox1 was not detectable (data not shown). Sox2 is only expressed in anterial ectoderm whereas Sox3 expressed pan-ectodermally. At HH st.8, Sox1 was still not detectable whereas the expression of both Sox2 and Sox3 was detected in neural ectoderm. From HH st.11 onward, all three members of SoxB1 subgroup were expressed in neural ectorderm and CNS (from Okuda et al., 2006).

Many reports have demonstrated that SoxB1 genes play an important role in neural determination. The expression of Sox2 and FGF initiates neural differentiation of the ectoderm in *Xenopus* (Mizuseki et al., 1998). Conversely, expression of dominant negative forms of Sox2 significantly inhibits neural differentiation by affecting BMP signals (Kishi et al., 2000). Intriguingly, dominant negative forms of Sox2 also block cells from being committed to epidermal cell fate, suggesting that Sox2 might generally prevent cells from differentiation (Mizuseki et al., 1998). Overexpression of Sox1 has the opposite effects, inducing mouse embryonic carcinoma cells (P19) to undergo neuronal differentiation programme (Pevny et al., 1998). Similarly, mouse embryonic stem cells (ES) differentiate into neuroectoderm when overexpressing Sox1 or Sox2 (Zhao et al., 2004).

After neural induction, the expression of Sox B1 genes is restricted to neural precursors located in the CNS. Here, Sox B1 genes maintain neural progenitor identity and prevent further differentiation (Pevny and Placzek, 2005). Overexpression of Sox2/3 inhibits neural progenitor differentiation in chick embryos whilst, in contrast, expression of dominant negative forms of Sox2/3 results in the premature onset of neuron formation (Bylund et al., 2003; Graham et al., 2003). The exact mechanism by which SoxB1 genes keep cells from differentiation remains unclear.

1.2 Sox2 in stem cells

1.2.1 Stem cells

From the mid 1800's, people knew that some cells could generate other cells. However, not until 1963, did Canadian scientists, McCulloch and Till, identify the self-renewal ability of transplated mouse bone marrow cells (McCulloch and Till, 1960). Later, embryonic stem cells (ES) were derived from the inner cell mass (ICM) of blastocysts (Thomson et al., 1998). Also, germ cells were derived from fetal gonad tissue in the same year (Shamblott et al., 1998). Broadly defined, stem cells have the ability to self-renew and potency to differentiate into several cell types. According to the ability to differentiate, stem cells could be classified into totipotent, pluripotent and multipotent stem cells (Figure 1.2.1). Totipotent stem cells (zygote) can form all living cells and generate whole individuals whereas the pluripotent stem cells (ES cells) could differentiate into any of the three germ layers (endoderm, mesoderm and ectoderm) but not a whole individual because they lack the ability to form extraembryonic tissues. Multipotent stem cells (fetal tissue; e.g., cord blood) have limited ability to differentiate and could develop only into a closely related family of cells.

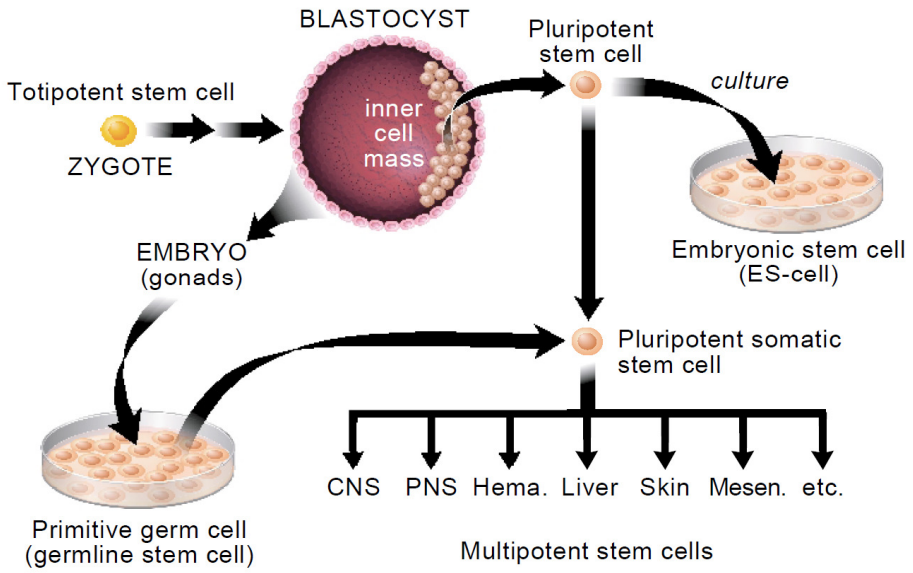


Figure 1.2.1 Different potency of stem cells.

Zygote (totipotent stem cell) divided into blastocyst. The inner cell mass could be isolated and cultured as pluripotent stem cells. The multipotent stem cells have limited ability to differentiate (From Anderson et al., 2001).

In differentiated tissues, not all cells are lineage restricted. Adult stem cells were found to have plasticity, able to give rise to many cell types within one particular lineage (Ferrari et al. 1998). The physical functions of adult stem cells include regeneration of injured tissues or replacement of cells lost in organs with a high turnover rate. There are, hence, limitations to adult stem cell differentiation which allowed them to be defined as multipotent stem cells. To date, adult stem cells had been found in many organs such as blood, skin, gut, testis, the respiratory tract, brain, skeleton and muscle (Martin Raff, 2003).

1.2.2 Sox factors in neural stem cells

Neural stem cells (NSCs) or the later neural progenitor cells (NPCs) with more restricted potency can be isolated from the CNS and cultured as primary cells using mitogens such as epidermal growth factor (EGF) and basic fibroblast

growth factor (bFGF) (Reynolds and Weiss, 1992; Richards *et al*., 1992). Musashi 1 (Kaneko *et al.,* 2000), Nestin (Lendahl *et al.,* 1990), and Sox1 (Peveny *et al*., 1998) have been used as markers of NSCs. NSCs are present in the developing embryonic neuroectoderm and adult brain subventricular zone of the lateral ventricles and the subgranular layer of hippocampal dentate gyrus (Figure 1.2.2). The process by while NSCs differentiate into neurons is called neurogenesis.

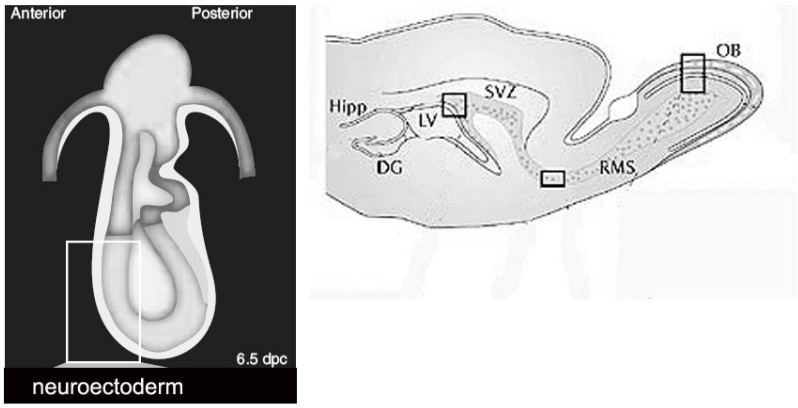


Figure 1.2.2 Illustration of the origins of NSCs

Embryonic NSCs were isolated from neuroectoderm. (A) shows the position of neuroectoderm in mouse embryo. (B) shows the origin of adult NSCs in rat brain. Adult NSCs were isolated from subventricular zone (SVZ) of the lateral ventricles (LV) and the subgranular layer of hippocampal dentate gyrus (DG). Hippocampus (Hipp); olfactory bulb (OB); rostral migratory stream (RMS). (From Lledo et al., 2006)

During neurogenesis, cells begin to express differentiation markers instead of progenitor markers and migrate from the ventricular zone to the marginal zone. On the other hand, proneural basic helix-loop-helix (bHLH) transcription factors, such as Ngn1, Ngn2 and Mash1, trigger cell cycle exit and the differentiation processes that lead to the expression of pan-neuronal markers (Kiefer, 2005).

Many experiments provide evidence that SoxB1 genes play key roles in maintaining the stem cell-like state of NSCs. Firstly, NSCs were inhibited from

differentiation due to the overexpression of Sox1-3 (Bylund *et al*., 2003; Graham *et al.*, 2003). The expression of Sox1 and Sox2 does not affect the propagation of undifferentiated ES cells but promotes ES cell differentiation into neuroectoderm after release from self-renewal (Zhao, *et al*., 2004). Furthermore, the overexpression of the HMG domain of Sox2 or Sox3 fused with the repression domain of the *Drosophila* protein Engrailed (EnR) (regarded as a dominant negative form of Sox2), resulted in premature exit from the cell cycle and expression of differentiation markers. In addition, the other Sox family member, Sox10, maintains neural crest stem cells as multipotent cells in the peripheral nervous system (Kim et al., 2003) (Figure 1.2.3). In adult neural stem cells, Sox genes are also expressed and are critical for neurogenesis (Ferr *et al*., 2004).

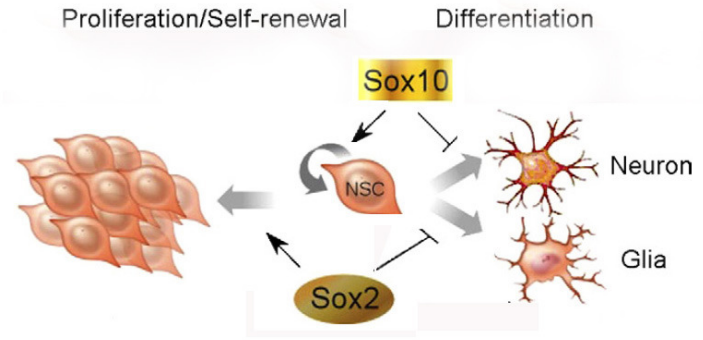


Figure 1.2.3 Sox factors in neurogenesis.

Sox2 is necessary for maintaining the proliferation and self-renewal in NSCs. Sox10, a member of SoxE subgroup, maintains the multipotency of neural crest stem cells. (From Shi et al., 2008)

The balance of SoxB1 and SoxB2 proteins is critical in maintaining the character of stem cells in the CNS. The expression of Sox1, Sox2 or Sox3 keeps stem cells from differentiating whereas inhibition of their activity leads to premature neuronal differentiation (Avilion *et al*., 2003). On the other hand, SoxB2 subgroup proteins (Sox14 and Sox21) are coexpressed with SoxB1 subgroup proteins throughout the CNS (Pevny et al., 1998). Due to the high similarity of HMG domains in SoxB1 and SoxB2 proteins, they have similar DNA-binding specificity and can target the same genes. Therefore, SoxB2 subgroup proteins could antagonise SoxB1 subgroup proteins through competing for the DNA-binding sites on target genes and thus blocking SoxB1 subgroup protein function (Wegner and Stolt, 2005).

1.2.3 Sox2 in the core transcriptional network of the ES cell

The homeodoman transcription factors, OCT4 and NANOG, have been reported to be essential for ES cells pluripotency maintenance (Mitsui et al.,

2003; Nichols et al., 1998). Later, Sox2 was found to couple with Oct4 as a heterodimer and upregulate the expression of Oct4, Sox2 and Nanog (Chew et al., 2005; Kuroda et al., 2005; Rodda et al., 2005). According to genome-wide binding site analysis, these three factors co-occupy the regulatory regions of several genes that function in the self-renewal and pluripotency of ES cells (Boyer et al., 2005; Loh et al., 2006). Therefore, they have been called the "master regulators" of ES cells (Zhou et al., 2007). Intriguingly, these three master regulators not only regulate a set of target genes but also autoregulate themselves by binding at their own promoters (Loh et al., 2006). This feedback circuit allows them to be kept in balance and stabilize the pluripotent state of ES cells (Jaenisch and Young, 2008) (Figure 1.2.4). In addition to the three master regulators, many other factors have been reported to be crucial in maintaining pluripotency, such as Sal1, Esrrb, Dax1 and Rif1 (Loh et al., 2006; Wang et al., 2006a) (Figure 1.2.4).

Although the three master regulators were defined as transcriptional activators, array data suggest that they repress almost half of their target genes in ES cells (Boyer et al., 2005; Loh et al., 2006); implying that this finding as transcriptional repressor is important to their biological roles. Moreover, they could also regulate gene expression indirectly through affecting chromatin structure, DNA methylation, microRNA expression and X chromosome inactivation. Our understanding of the whole regulation network is getting more and more complicated as much research effort is put into this area. However, only a few transcription factors represent central nodes of the complicated network. This finding should make it possible to manipulate the pluripotency in the future.

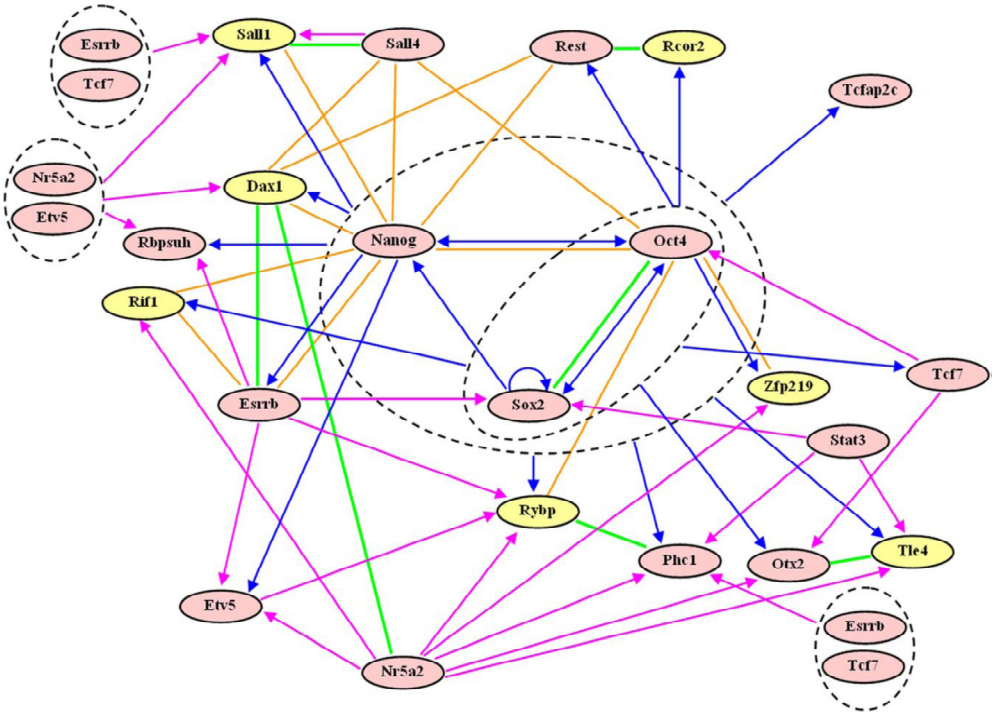


Figure 1.2.4 A pluripotency regulatory network in mouse ES cells.

The interactions between factors are showed in lines and the regulatory interaction are showed in arrows (From Zhou et al., 2007). The interaction among the core regulators (Pink), core regulators' protein-interaction partners (yellow), regulatory interactions inferred by anchor sites and by sites of coregulators within 150 bp of the anchor sites (Blue and Pink arrows), protein interactions (Orange).

1.2.4 Sox2 in induced pluripotent stem (ips) cells

During development, the totipotent zygote undergoes epigenetic changes and loss of differentiation capacity to form various cell lineages. Recent research has provided evidence of the reversibility of differentiation. Many strategies have been used to drive unipotent somatic cells to pluripotent cells. Nuclear transplantation reprogrammes somatic nuclei by transfer into enucleated oocytes (Wilmut et al., 1997). Fusion of somatic cells and embryonic stem cells can also confer pluripotency on somatic cells (Cowan et al., 2005; Tada et al., 2001). Under the culture conditions of germline stem cells, some unipotent cells were reprogrammed and became ES-like cells (Kanatsu-Shinohara et al.,

2004). On the other hand, expression of permeabilized somatic *Xenopus* cells to egg extracts results in chromatin remodelling that reverted the cells to a dedifferentiated status (Kikyo et al., 2000).

More recently, Takahashi and Yamanaka successfully reprogrammed somatic mouse cells to pluripotent cells by viral-mediated overexpression of four transcription factors; Oct4, Klf4, c-Myc and Sox2 and named these ES-like cells, "induced pluripotent stem" (iPS) cell (Takahashi and Yamanaka, 2006). Later, Thomson and colleagues reprogrammed human somatic cells to pluripotent stem cells with OCT4, SOX2, NANOG and LIN28 (Yu et al., 2007). In mouse neural stem cells, which express endogenous Sox2 and c-Myc, only Oct4 and Klf4 or Oct4 alone was enough to trigger reprogramming (Eminli et al., 2008; Kim et al., 2009a; Kim et al., 2008b). These results indicate that Sox2 is one of the key transcription factors regulating stem cell pluripotency.

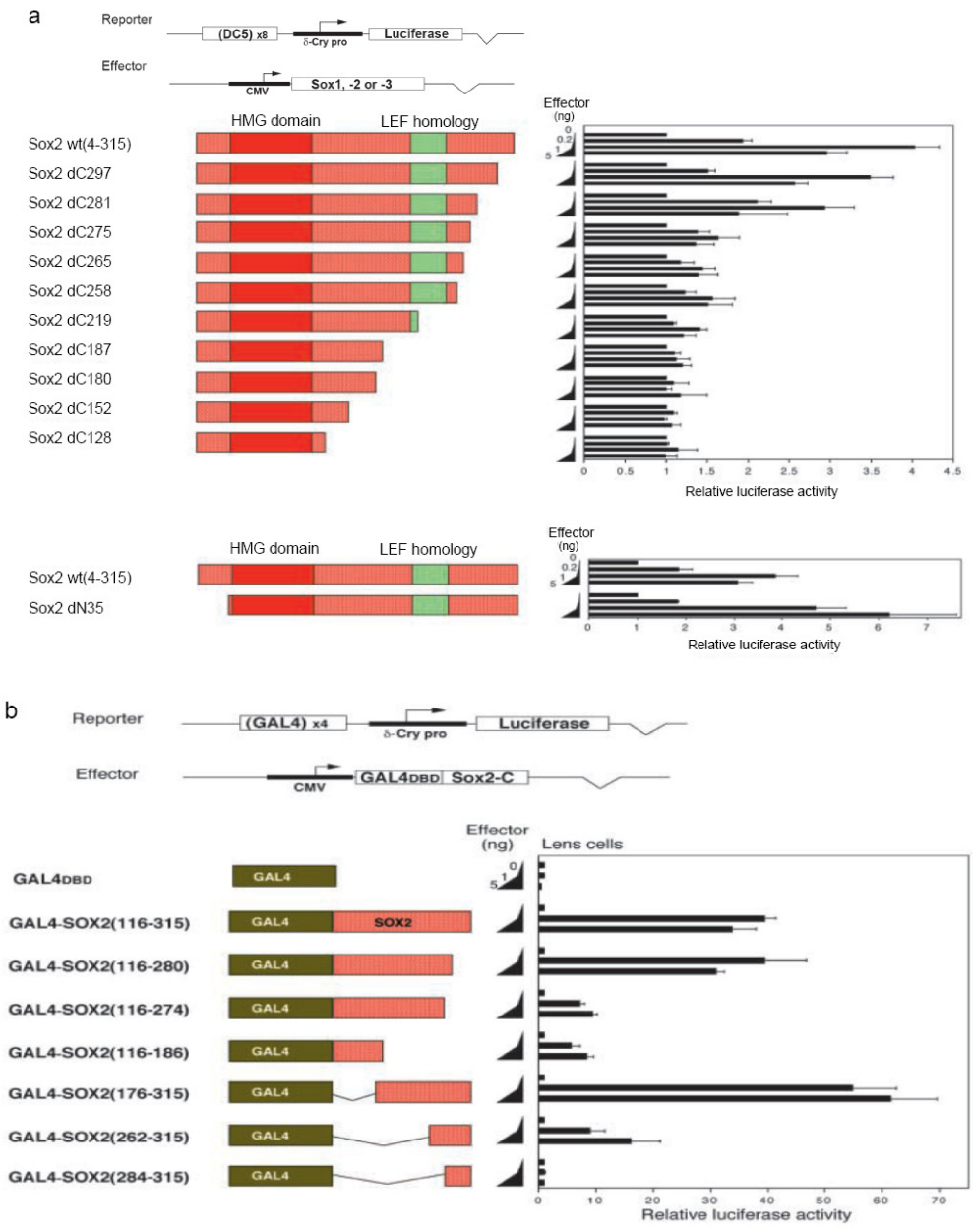
1.3 Sox2 as a transcription factor

1.3.1 The transcriptional activation activity domain of the SoxB1 subgroup

Sox proteins function as transcription factors through the binding of the HMG domain to a specific DNA sequence and subsequent activation of the transcription of target genes (Wilson and Koopman, 2002). Several studies have attempted to map the transcriptional regulation domains of the SoxB1 family. Kamachi *et. al.* showed that transcriptional activation activity locates to the C-terminus in chicken Sox2 (Kamachi et al., 1999). They inserted the Sox2 endogenous target sequence, DC5, in front of a luciferase reporter and a luciferase assay was carried out, co-expressing a series of C-terminal truncated versions of Sox2. The removal of 41 amino acids from the C-terminus had a dramatic effect on Sox2 transcriptional activation activity (Figure 1.3.1 a). In contrast, the deletion of the Sox2 N-terminus was dispensable for Sox2 transcriptional activation activity (Figure 1.3.1 a). In another study, several deletions of the Sox2 C-terminus were fused with the Gal4 DNA-binding domain to determine the key domains of Sox2 transcriptional activation activity. A significant change was observed between GAL4-SOX2 (116-280) and GAL4-SOX2 (116-274) which differ in length has only six amino acids difference. These results imply the importance of these six amino acids in transcriptional activation activity (Figure 1.3.1 b). Likewise, Nowling *et. al.* (2000) also identified the transcriptional activation activity domain of mouse Sox2 using C-terminal deletions in the GAL4 luciferase assay system (Figure 1.3.1 c). The results suggested that the transcriptional activation activity domain of Sox2 is located in the C-terminus and contains a serine-rich region

(Figure 1.3.1 c). However, the key amino acids or small motifs in the C-terminus of Sox2 that are required for Sox2 transcriptional function remain unclear.

As for Sox1, significant reduction in transcriptional activation activity was found when 51 amino acids were removed from the C-terminus of Sox1 (ΔC323) (Figure 1.3.1 d) (Kamachi et al., 1999). Therefore, the transcriptional activation activity domain of Sox1 might also locate to its C-terminus.



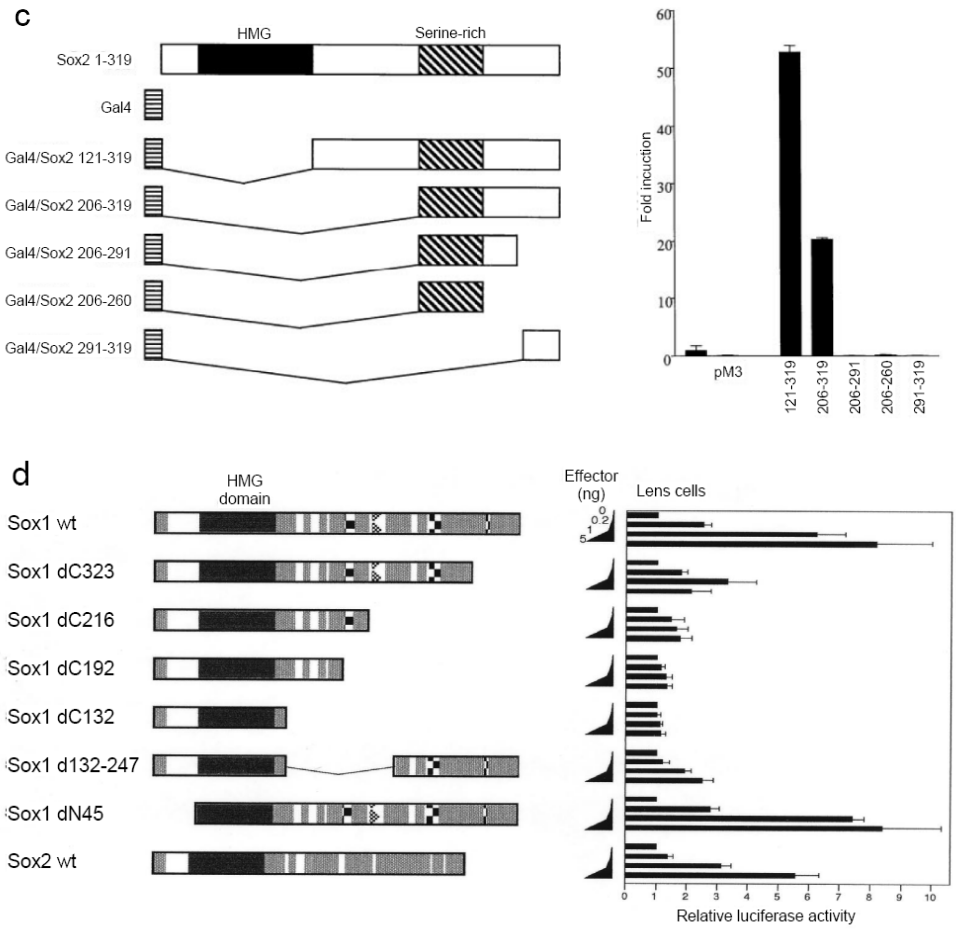


Figure 1.3.1 Transcriptional activation activity domain mapping within SoxB1 subgroup.

(a) Sox2 C-terminal and N-terminal deletion schema on the left. Chicken lens cells were co-transfected with various amounts of pcDNAI/Amp-based SOX2 expression vectors (0,0.2,1 and 5 ng) and luciferase reporter plasmid containing the octamerized DC5 fragment (illustrated at top). The relative luciferase activity of all the deletions is shown on the right. (Kamachi *et. al.,* 1998) (b) The activation potential of Gal4 DNA-binding domain and Sox2 C-terminus deletion fusion proteins. The luciferase reporter plasmid contains tetramerized Gal4 binding sequences and the *δ1- crystallin* minimal promoter (-51 to +57; *δ*-Cry pro). Varying amounts (0, 1 and 5 ng) of deletion Gal4-Sox2 fusion construct were co-transfected with reporter plasmid into the chicken lens cells. The relative luciferase activity of all the deletions is shown on the right (Kamachi et al., 1998). (c) Mouse Sox2 transactivation domain mapping. Various mouse Sox2 C-terminal deletions fused with Gal4 DNA-binding domain as illustrated on the left. The luciferase activity was measured after co-transfection of various deletions of Gal4-Sox2 (1µg) with the pG5EC (2µg) reporter plasmid into HeLa cells (Nowling et al., 2000). (d) Mapping of Sox1 transactivation domain. Various C-terminal deletion of Sox1 were expressed with reporter plasmid as described in (b) (From Kamachi et al., 1999).

1.3.2 Partners of SoxB1 transcriptional activation activity

The DNA binding dissociation constant of the Sox family HMG domain is 10^{-8}-10^{-9} $_{M}$ *in vitro*, which is not alone sufficiently stable to activate transcriptional regulation (10^{-11} $_{M}$ is the standard DNA binding affinity for a transcription factor) (Mertin et al., 1999). Sox proteins therefore need partners to enhance their transcriptional activation ability (Kamachi et al., 2000). This is an important characteristic of Sox proteins that allows them to distinguish the correct regulatory target (Figure 1.3.2). A good example is cooperation of Sox2 with the POU domain factor Oct3/4 to enhance the expression of fibroblast growth factor 4 (Fgf4) in embryonic stem cells and embryonal carcinoma cells (Yuan et al., 1995). The effects of Sox2 and Oct3/4 as a heterodimer are synergistic, affecting the expression of many genes in embryos, such as undifferentiated embryonic cell transcription factor 1 (UTF1), osteopontin, and the Sox2 gene itself (Botquin et al., 1998; Nishimoto et al., 1999). Also, PAX6 and/or δEF3 form a complex with Sox2 to activate lens-specific *δ–crystallin* expression through the DC5 enhancer (Kamachi et al., 2001) and the interaction of Sox2 and Oct1 increases the transcriptional activity of the Hox/Pbx1 complex (Di Rocco et al., 2001). All the interactions of Sox proteins and their partners so far analysed are through the HMG domain (Kamachi et al., 1999; Kamachi et al., 1998). The Sox2 target genes were activated through the activation domain close to Sox2 C-terminus. A GAL4 DNA-binding domain fused with the Sox2 C-terminal activation domain alone trigger target gene expression just as well as full-length Sox2 gene (Kamachi et al., 1999). On the other hand, when the Sox2 C-terminus was replaced by the highly potent viral activation domain VP16, the chimera still required a partner protein to activate

down-stream gene expression (Kamachi et al., 1999). In summary, Sox proteins function as transcription factors by cooperating with specific partners which restrict Sox proteins to particular targets.

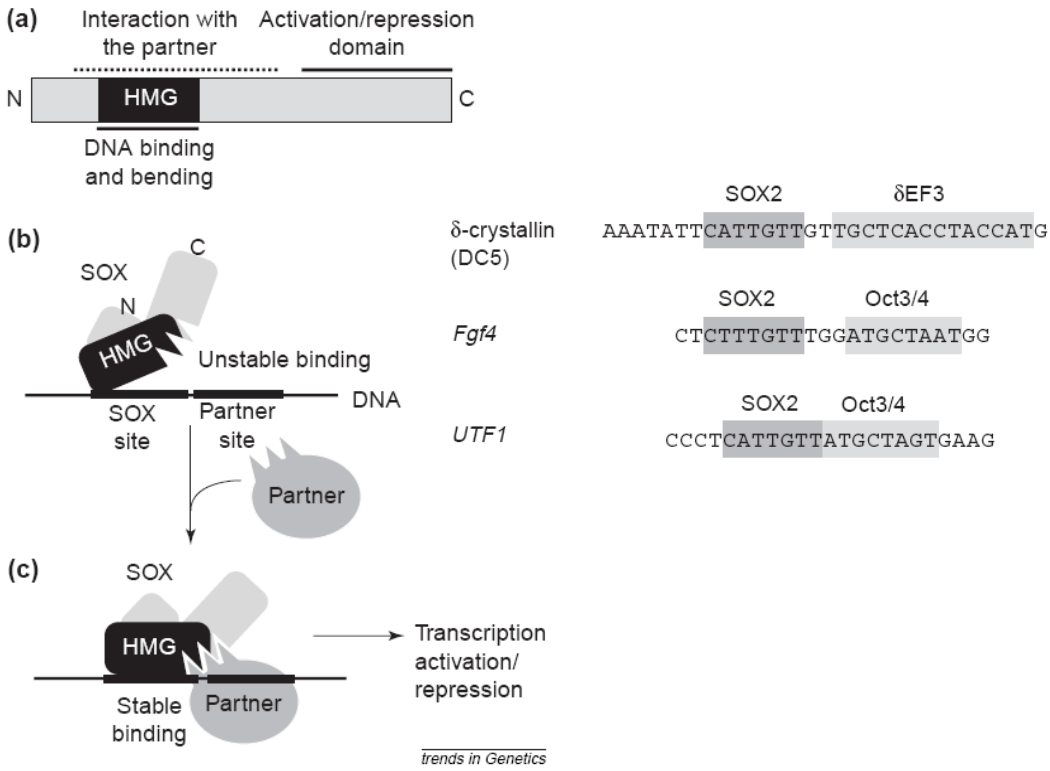


Figure 1.3.2 Sox and partner interaction model and binding sites.

(a) Diagram illustrating the three major functional domains of Sox proteins. (b). When a partner factor binds to a DNA site near a Sox protein, a ternary interaction will be stabilized and transcription activation/repression of the down stream gene achieved. The binding sites of Sox2 are showed on the right. Fgf4: fibroblast growth factor 4; UTF1: undifferentiated embryonic cell transcription factor 1 (From Kamachi et al., 2000)

1.3.3 Sox B1 genes as transcriptional repressors

In addition to a transcriptional activation function, recent studies have shown that SoxB1 proteins can also function as transcriptional repressors. Kan *et al*. (2004) carried out a luciferase reporter assay in HEK293T cells and P19 cells using the Hes1-luc reporter construct, which has the Hes1 (helix-loop-helix transcription factor) promoter in front of a luciferase coding sequence. The results showed that transcription of the reporter gene was suppressed when Sox1 was overexpressed. As for Sox2, it has been reported that a decrease in Sox2 expression is closely related to the increasing expression of the L1 elements in rat neuronal stem cells (Muotri et al., 2005). This implies the negative regulation of the LINE1(L1) element by Sox2, although the effect could be indirect. On the other hand, the non-coding RNA *Xist* was reported to be bound by Sox2 in intron 1 and was down regulated by Sox2 in undifferentiated embryonic stem cells (Navarro et al., 2008). This result suggests the direct down-regulation of a target gene by Sox2. In addition, overexpression of Sox2 can repress the expression of a glial fibrillary acidic protein (GFAP) luciferase reporter (Figure 1.3.3 A). Chromatin-immunoprecipitaion (Chip) and electrophoretic mobility shift assays (EMSA) were used to test the direct binding of Sox2 on the GFAP promoter region (Figure 1.3.3 B, C). These results indicate that Sox2 directly represses GFAP expression in mouse neural stem cells (Cavallaro et al., 2008). As for Sox3, in *Xenopus*, Zhang *et al*. (2003) found that *XSox3* can bind to the promoter region of *Xnr5* which is a Nodal-related gene. After knockdown of the endogenous *XSox3* by morpholino injection in embryos, the RNA level of *Xnr5* increased significantly (Zhang et al., 2003a).

The mechanism of SoxB1 transcriptional repression remains unclear. One possible hypothesis is that Sox proteins interact with co-repressors to suppress the transcription of target gene. For example, Groucho, a co-repressor (Courey and Jia, 2001), has been shown to interact with the transcription factor, TCF, which also contains an HMG domain (Brantjes et al., 2001; Cavallo et al., 1998). Since Sox proteins also have an HMG domain, they might interact with co-repressors in similar way.

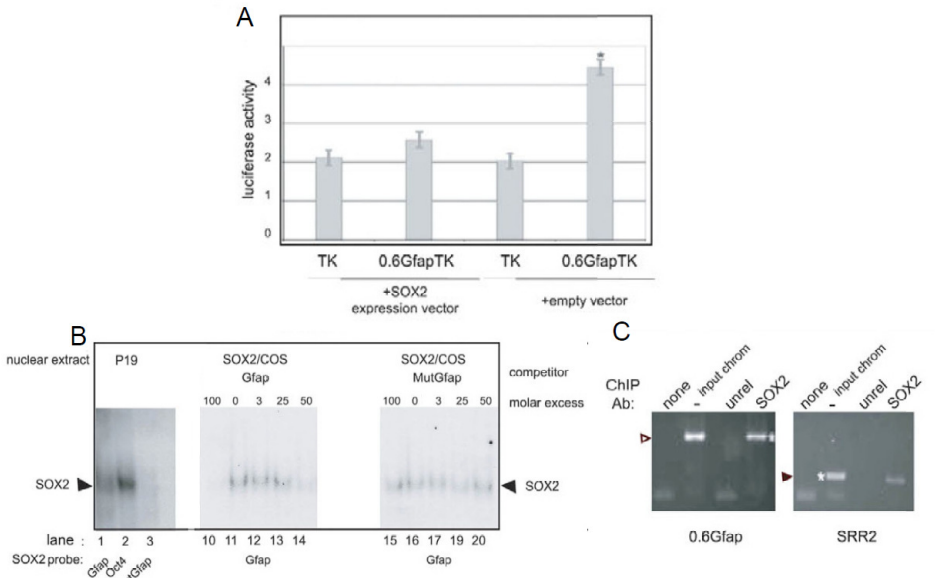


Figure 1.3.3 Sox2 directly represses GFAP expression.

Sox2 transcriptional activation activity was measured by luciferase assay (A). When Sox2 expression was absent, the luciferase activity of the GFAP reporter was increased. (B) Electrophoretic mobility shift assay (EMSA) showed that Sox2 could directly bind on GFAP promoter probes. (C) ChIP showed Sox2 could directly bind on GFAP promoter endogenously in p19 cells. (From Cavallaro et al., 2008)

1.4 Co-repressors of transcription

Correct gene expression is controlled by the dynamics of co-activator and co-repressor transcriptional complexes (Figure 1.4.1) (Perissi et al., 2010). However, the mode of transcriptional repression is less well understood than activation. According to their abilities, repressors are categorized into two classes, long-range and short-range repressors (Courey and Jia, 2001). Most transcriptional repressors are long-range; they inhibit all of a gene's enhancers which regulate gene expression, even if those enhancers are far distant from the repressor binding site. Long-range repression could cause chromosomal locus inactivation through post-translational modifications of the lysine in histone tails or bring the locus close to regions of heterochromatin to silence gene expression. As for short-range repressors, they only block the function of neighbouring activators (Gray and Levine, 1996). One mechanism of transcriptional repression is that the transcriptional repressor binds to the promoter region of a target gene and represses its expression by recruiting a "co-repressor". For example, a co-repressor could interact with histone deacetylase (HDAC) and alter the chromatin structure through post-translationally modifying the N-terminal tails of core histones (H2A, H2B, H3, H4).

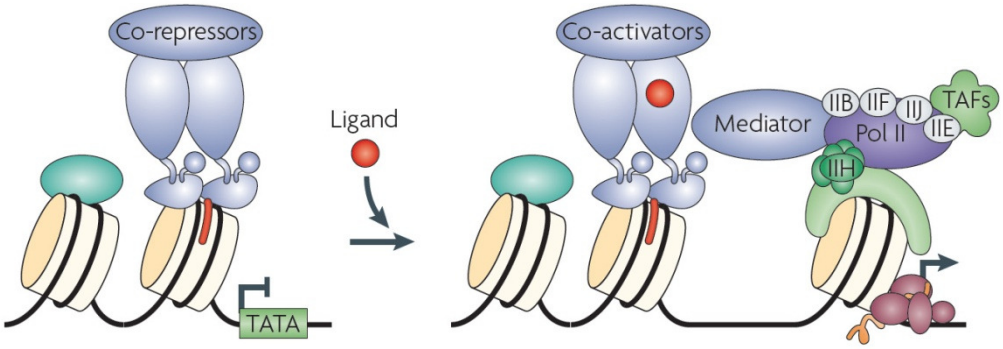


Figure 1.4.1 Model of transcription regulation.

The ligand signal-dependent switch of co-repressors and co-activators is illustrated. A co-repressor binds to on the promoter region of target gene. When the ligand transmits the signal into the nucleus, the co-repressor is replaced by a co-activator to trigger the transcription of down-stream gene (From Perissi et al., 2010).

1.4.1 Groucho

One well-studied co-repressor in long-range repression is Groucho (Fisher and Caudy, 1998). Groucho was first identified in *Drosophila* (Lindsley and Grell, 1968). It was named according to the dense bristles above the eyes of Groucho mutant fly, reminiscent of the eyebrows of the Groucho Marx (Figure 1.4.2).

Figure 1.4.2 Groucho mutant *Drosophila.*

Picture of the wild type (A) and the Groucho mutant *Drosophila* (B) are shown. Mutant has clumps of extra bristles above eyes and extra bristles on humerus. The pictures were generated by UCL cancer institute (**http://www.ucl.ac.uk/cancer/research-groups/transcriptional-regulation/index.htm**). (C) Picture of Groucho Marx (http://www.quotecollection.com/image-view.php?img=groucho-marx-3.jpg).

Groucho family members share two highly conserved domains; the Q domain and the WD-repeat (Figure 1.4.3). There are also a loose conserved CcN motif, containing a nuclear localization sequence (NLS) and two phosphorylation sites (cdc2 kinase and casein kinase II phosphorylation sites (Stifani et al., 1992)). The CcN motif is flanked by glycine/proline rich (GP) and serine/proline rich (SP) domains. The WD-repeat consists of four to eight repeats, however, only a few residues are well conserved ($X_{6\text{-}94}$---[GH---$X_{23\text{-}41}$--WD]). Groucho proteins have been found to form homotetramers. The Q domain, which contains two putative amphipathic α-helical motifs (AH1, AH2), might contribute to the oligomerization of Grouchos. The transcriptional repression function of Groucho has been shown to require its homotetramerization (Song et al., 2004). The WD-repeat domain in Groucho is involved in interacting with DNA-bound transcriptional repressors.

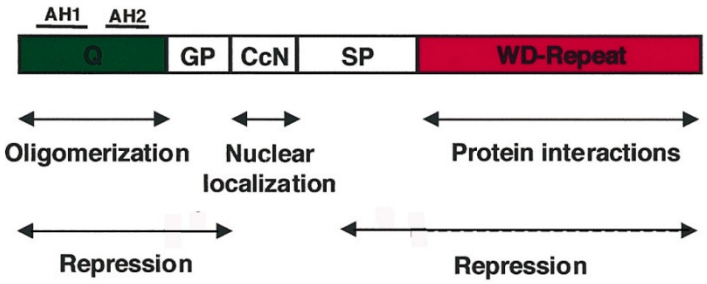


Figure 1.4.3 Illustration of Groucho structure.

The highly conserved Q domain and WD-repreat are shaded. The repression activity has been reported to be localized in either N-terminal or C-terminal. Oligomerization is believed to be dependent on the Q domain which has two putative amphipathic α-helical motifs (AH1, AH2). Protein interaction occurs through the WD-repeats. The nuclear localization sequence was identified within the CcN domain. The relative locations are labelled beneath. (From Chen and Courey, 2000)

Groucho proteins are abundant and broadly expressed across species. 19 Gro-homologous proteins have been identified to date (Chen and Courey, 2000). Both vertebrates and invertebrates have Gro-like proteins, but, no Gro-like proteins are found in plants, fungi, or protozoans. According to sequence conservation, there are two subgroups of the Groucho family (Figure 1.4.4). The GRO subgroup includes *Drosophila* Groucho and human Groucho orthologs, also called transducin-like enhancer-of-split 1-3 (TLE1-3). GRO subgroup members have the structure described above. Members of the other subgroup, mouse amino Enhancer of split (AES), only have the domains equivalent to the N-terminal Q, GP and partial CcN domain of the long form Groucho. Groucho homologues in human include TLE1-4 and the truncated version hAES, whereas in mouse they are named *Grg*1-4 and the truncated version *AES* (Fisher and Caudy, 1998).

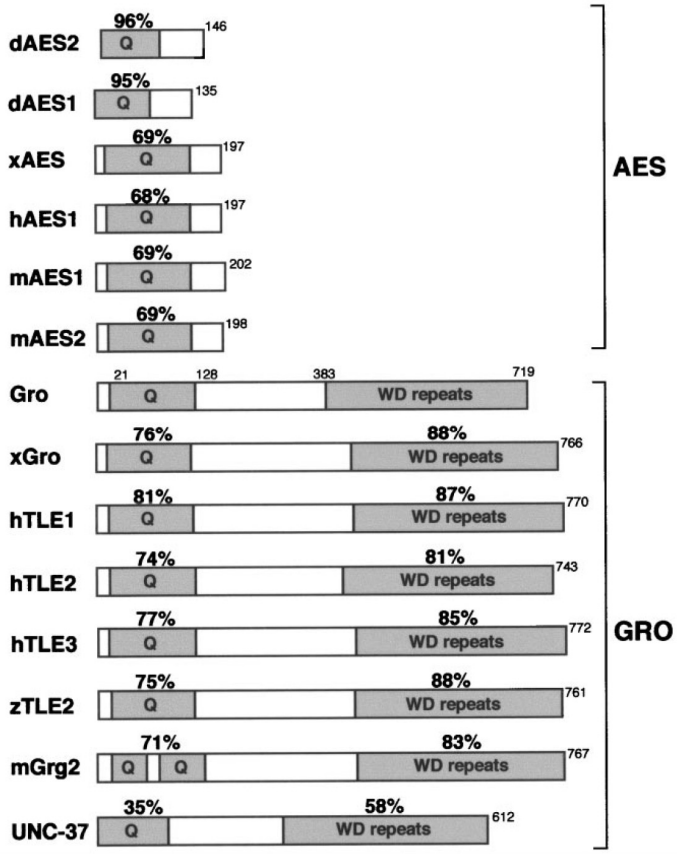


Figure 1.4.4 Sequence conservation of Gro/TLE family.

The highly conserved Q domain and WD-repeat are shaded and the similarity is marked above the box. d: *Drosophila*; x: *Xenopus*; h: Human; m: mouse; z: zebrafish (From Beagle and Johnson, 2010; Chen and Courey, 2000)

1.4.2 Groucho interaction motifs in transcription factors

Grouchos have been reported to interact with a number of transcription factors. Some are transcriptional repressors and their repression function depends on the interaction with Groucho. There include Hairy (Fisher et al., 1996), HES (Grbavec and Stifani, 1996), and Blimp-1 (Ren et al., 1999). On the other hand, some transcriptional activators could be converted into repressors when interacting with Groucho-related proteins. These include Tcf (Roose et al., 1998), Runt (Aronson et al., 1997) and Dorsal (Dubnicoff et al., 1997).

The specific Groucho interaction sequence of some transcription factors has been identified as a short peptide motif. Based on loose similarity of the Groucho interaction motifs, they can be classified into two motifs, WRPW/Y and eh-like. The tetrapeptide motifs WRPW and WRPY are conserved and necessary for the interaction between Groucho and Hairy-related and Runt domain proteins, respectively (Aronson et al., 1997; Paroush et al., 1994). However, many other similar motifs were found to interact with Groucho. Huckebein proteins were reported to interact with Groucho through an FRPW motif (Goldstein et al., 1999) whereas Brinker proteins interact with Groucho through an FKPY motif (Zhang et al., 2001). Moreover, eyeless proteins interact with Groucho via another variant, the YSPW motif (Choi et al., 2005). Taken together, in many variants, the third residue of the Groucho interaction motif, proline, is never changed. Also, the first and fourth position are occupied by the aromatic amino acids, Phenylalanine (F), Tyrosine (Y) and Tryptophan (W). These could be considered as the principle amino acids of WRPW/Y motif (Figure 1.4.5 A).

The eh-like motif was originally found in the Engrailed protein, which is a transcriptional repressor. Engrailed has two Groucho interaction motifs named eh1 and eh2 (engrailed homology 1 and 2) respectively. The amino acid sequence of the eh-like motifs can be summarized as FxIxxI/L (Figure 1.4.5 B) (Goldstein et al., 2005). However, many variants have been found that have loose similarity to the eh-like motif. Pax proteins contain an eh-like motif in which the phenylalanine is replaced by tyrosine (YSISGILG) (Eberhard et al., 2000). Otx2 has tryptophan replacing the phenylalanine (WSPASISP) (Heimbucher et al., 2007). Previous studies in our lab also identified a Groucho

interaction motif on Sox3 which has loose similarity with eh-like motif (YDMPGL) (PhD thesis, Zulfiqar Laghari, 2010). These data indicate that the sequence of the eh-like motif is quite flexible except for a few key residues such as the first aromatic amino acid, the second serine or proline and the final leucine or isoleucine.

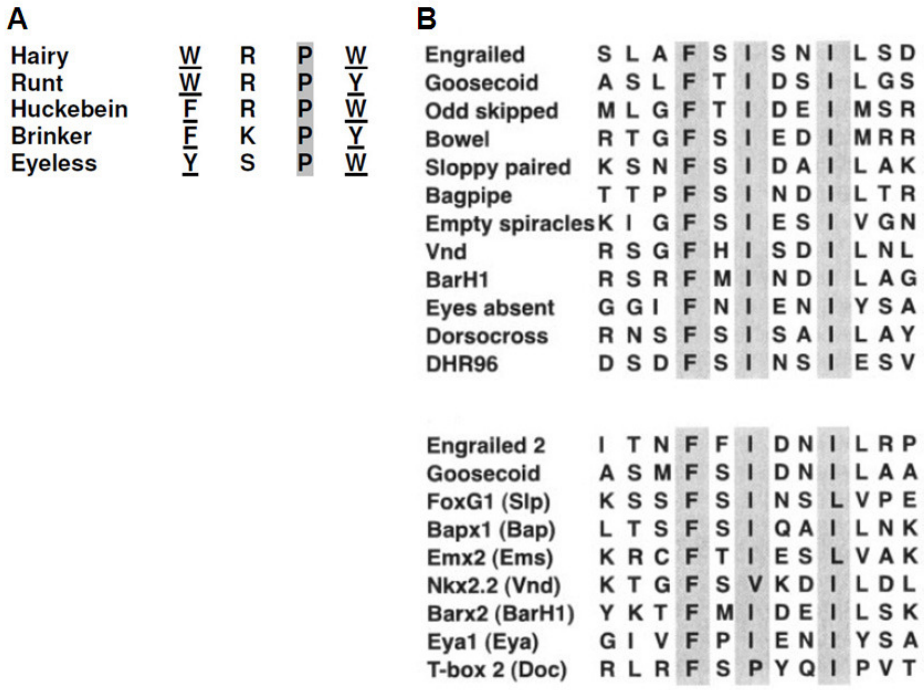


Figure 1.4.5 Similarity of Groucho interaction motifs.

WRPW/Y motifs of various genes are showed in (A). The conserved proline is shadowed in gray box whereas the aromatic amino acids are underlined. The FxIxxI/L motifs of various genes are showed in (B). The conserved phenylalanine and isoleucine are shadowed in gray boxes (From Goldstein et al., 2005).

1.4.3 Long form and short form of Groucho

Groucho proteins can be classified into long forms (GRO subgroup) and short forms (AES subgroup) according to the length of amino acid sequence (Figure 1.4.3). The short form Grouchos have only Q and GP domains which are conserved with the long form Grouchos (Figure 1.4.4) (Beagle and Johnson, 2010). However, the short forms of Groucho lose potential protein interactions through the WD-repeat, such as the interaction with HDACs (Zhang et al., 2008). It has been considered that the short form Grouchos, which retain the Q domain, therefore might replace the long form Grouchos in the Groucho tetrameric repressor complex or compete for recruitment on promoters with the long form Groucho (Gasperowicz and Otto, 2005; Sekiya and Zaret, 2007). Therefore, short form Grouchos have been considered as dominant-negative family members (Bajoghli et al., 2007). For example, *Runx2* interacts with the both long form and short form Grouchos, but, the short form Groucho (*Grg5*) enhances *Runx2* transcriptional activation activity whereas the long form Groucho (*Grg3*) inhibits it (Wang et al., 2004).

Recently, more and more exceptions were found which indicate that the short forms Groucho are not always antagonists of long forms Groucho. For example, both long form and short form Groucho can interact with and repress NF-κB mediated transcriptional activity (Tetsuka et al., 2000). Also, the short form Groucho functions as a co-repressor of the androgen receptor (AR) (Zhang et al., 2010b). Although the short form Groucho does not interact with HDACs, it has been reported that it could interact with TFIIE and so negatively regulate basal transcription (Yu et al., 2001). These data suggest that the short form Grouchos are capable of directly modulating transcription.

1.4.5 Functions of Groucho in development

Groucho proteins are involved in many developmental processes. Human Grg2 (TLE2) was reported to down regulate the transcriptional activation ability of Runx2, which is an important factor in the process of skeletal development (Thirunavukkarasu et al., 1998). Pax5 is a bifunctional transcription factor and plays a critical role in B-cell lineage commitment (Nutt et al., 1999; Rolink et al., 1999). Mouse Grg4 was demonstrated to interact with Pax5 and mediate its transcriptional repression function (Eberhard et al., 2000). Grouchos are also involved in myogenesis; TLE1 can interact with HES6 which mediated repression and promotes myoblast differentiation (Gao et al., 2001; Sasai et al., 1992). Both long form and short form Grouchos interact with Six3 and Six6, which induce eye development (Bernier et al., 2000; Lagutin et al., 2001).

Moreover, Groucho proteins have been studied in most detail in neurogenesis. HES1 (Hairy/Enhancer of split), a basic helix-loop-helix (bHLH) family transcription repressor, negatively regulates neuronal differentiation (Knust et al., 1992). This negative regulation of Notch signalling, which triggers neurogenic gene expression, requires both Groucho and HES in *Drosophila* (Paroush et al., 1994). Similarly, human Grg1 (TLE1) forms a repressor complex with HES1 and PARP1 to regulate neuronal differentiation through control of MASH1 expression (Ju et al., 2004).

1.4.6 Groucho and HMG transcription factors

A well-studied Groucho co-repressor function is in the canonical Wnt signalling pathway (Kormish et al., 2010) (Figure 1.4.7). In the absence of Wnt signalling, gene targets are silenced by TCF/Lef via the Groucho co-repressors (Brantjes et al., 2001).

Similar to TCF/Lef, Sox proteins also have HMG DNA-binding domains. Therefore, the machinery of switching from transcriptional activator to repressor might be shared. Previous studies in our lab have demonstrated the interaction of Sox3 with a Groucho co-repressor (PhD theses, Caroline Hirst, 2009; Zulfiqar Laghari, 2010). Given the high similarity of Sox2 and Sox3, the same scenario might also be true for Sox2.

1.4 Research objectives

The early expression in embryos and the ability to reprogram fully differentiated cells, suggest an essential role of Sox2 in both embryo development and pluripotency. Previous studies have focused on the transcriptional activation activity of Sox2 in embryonic stem cells. However, genome-wide analysis showed almost the same number of genes were repressed as were activated by Sox2 (Boyer et al., 2005). Although this repression has been interpreted as an indirect effect of Sox2, there are a number of lines of evidence suggesting that Sox2 acts directly as a transcriptional repressor.

The main aims of this thesis are to demonstrate the transcription repressor function of Sox2 and identify the possible mechanism of it. Also, I aimed to

map the functional domain on Sox2 in order to investigate the biological role of Sox2 transcriptional repression.

Data in this thesis show that Sox2 indeed acts as a transcriptional repressor through interaction with Groucho co-repressor. The key domain on Sox2 which relates to the repressor function was dissected and a loss of repressor form of Sox2 was generated. In order to identify the possible target genes of Sox2 repressor function, genome-wide analysis was carried out to compare the gene expression profile when either wild-type Sox2 or the loss of repressor function form of Sox2 was overexpressed in human neural stem cells. Wild-type and mutant Sox2 were also overexpressed in early zebrafish embryos in consideration of the effect on organizer genes that were demonstrated to be repressed by Sox3 previously (Shih et al., 2010).

Chapter 2 Materials and methods

2.1 Construction of Sox2 mutants expression plasmid

A single bacterial colony of DH5α E. coli carrying the required plasmid was inoculated into 5 ml of LB broth containing appropriate antibiotic. Incubation at 37 ºC for at least 16 hours, with shaking at 230rpm was carried out. High quality plasmid DNA was then purified using a QIAprep® Spin Miniprep Kit (Qiagen, UK) following the manufacturer's instructions. At the final spin, the purified DNA was eluted in 50 µl of sterile distilled water (SDW).

2.1.2 Measuring DNA concentrations

The concentration of purified DNA was measured by a NanoDrop 1000 Spectrophotometer (Thermo Scientific) by applying 1 µl of DNA sample onto the lower measurement pedestal after calibrating the blank curve using pure water. The software calculates the concentration and the purity (the ratio of the reading under 260/280 wave length) of nucleic acid.

2.1.3 Nucleic acid restriction digests

In this study, restriction enzymes were purchased from Invitrogen or New England Biolabs. Restriction digests were carried out following the manufacturer's instructions. The total reaction volume ranged from 10-50 µl and with incubations at 37 ºC for 1-3 hours. Digested DNA was then checked by gel electrophoresis.

2.1.4 Agarose gel electrophoresis

Gel electorophoresis was carried out using 1-1.5% agarose gels, made with electrophoresis grade agarose (Invitrogen). 1x TAE buffer (0.04 M Tris, 0.02 M Acetic acid, 0.001M EDTA at pH 8.0) and ethidium bromide (0.1 µg/ml). DNA samples were loaded with 1x DNA loading buffer (New England Biolabs). 1Kb or 100bp DNA ladders (New England Biolabs) were loaded alongside the samples. Electrophoresis was carried out at 5-10 Volt/cm for 15-20 minutes. The gels were imaged using the AlphaImager 1220 Documentation & Analysis System (Alpha Innotech Corporation).

2.1.5 DNA purification from agarose gel

After gel electrophoresis, clearly separated DNA fragments were cut out with a scalpel blade under UV light. The extraction of DNA from each gel slice was carried out using a GeneCleanII kit (QBIOgene) according to the manufacturer's instructions.

2.1.6 DNA ligation and bacterial transformation

Desired DNA inserts were ligated into mammalian expression vectors, eg. pcDNA3 (Invitrogen), using T4 DNA ligase (NEB), 1x ligase buffer in 10 µl reactions. 5 µl of ligation reaction was then added to DH5α competent cells (One Shot® TOP 10F Invitrogen) incubated on ice for 30 minutes. The samples were incubated at 42 ºC for 30 seconds then placed on ice immediately for 5 minutes. Recovery of cells was carried out by adding in 50 µl of SOC medium and incubation at 37 ºC for 1 hour with shaking. The cells were then plated on

LB agar plates which contained the appropriate antibiotic and incubated at 37 ºC for 16 - 20 hours.

2.1.7 Glycerol stock of bacteria

Bacteria were cultured in appropriate selective medium for at least 8 hours. The cultures were then mixed with 80% glycerol in 1:1 ratio. The stock solution was vortexed before storage at -80 ºC.

2.1.8 Polymerase chain reaction (PCR)

Primer sequences are listed in Table 2.1.1. Oligonucleotide primers were synthesized by MWG (MWG-Biotech, UK). PCRs were carried out using Phusion® High-Fidelity DNA Polymerase (Finnzymes) in order to reach the high denaturation temperature for the GC-rich Sox2 and Sox3 sequences. For 20 µl reactions, 40 ng of template DNA, 0.5 µmole of each primer, 200 µmole of dNTP, 1x GC buffer, 0.2 µl (2U/µl) of polymerase and 10% DMSO were mixed thoroughly. PCR was carried out using a G-storm thermocycler PCR following the Phusion® polymerase manufacturer's instructions; 98 ºC for 30 seconds as initial denaturation, 30 cycles at 98 ºC for 10 seconds denaturation, annealing under appropriate temperature as primers required for 20 seconds, and 72 ºC for 30 seconds per kb. The final extension was 72 ºC for 5 minutes. PCR products were stored at 4 ºC.

Primer name	Sequence 5′ – 3′
mSox3	
mSox3 SU PM up	CGATCGCACCACGGAGCCCAT
mSox3 SU PM down	CGATCGGAGCCCAGCTCTCCG
mSox3 SUMO KR F	GGCTCCGTGGTGCGATCGGAGCCCAGC
mSox3 SUMO KR R	GCTGGGCTCCGATCGCACCACGGAGCC
mSox31M-F	GAATTCATGCGACCAGCTCGAGAGAA
mSox3 d90 F	GAA TTC ATGTACAGCCTGCTGGAGACT
mSox3 d427 R	CTCGAGCAGCCGGCCGCCTGGGAGCGG
mSox3 d221 R	AAGCTTGCGGCGGGGCCGGTACTT
mSox3 d278	AAGCTTGTGCAGGAGCAGCTGGGC
mSox3ups	GGATCCGGCAAGCTGCGAATGCGACCA
mSox3delYLP	GAATTCTTAGTCGCGCAGGTCGCCGAG
mSox3endYLP	GAATTCTTATGGCAGGTACATGCTGATCAT
mSox3 dNLS F	AACGACCAGGACCGGGTGAAG
mSox3 dNLS R	CCCGGGGCCAGGAGTCCCCCTGG
mSox3 R	
mSox2	
mSox2 F1	GGATCCGCCCGCATGTATAACATGATGGAG
mSox2 R1	GAATTCAGCGGGCTCCGGCACCTC
mSox2 R2	GAATTCCATGTCCCGGAGGTCCCC
mSox2 SUUP F	GGGAAGCTTCGATGTATAACATGATGGAGACG
reverse share with mSox3 SUUP R	
mSox2 SUDN F	GGGCGATCGGAGGCCAGCTCCAGC
reverse share with mSox2 WT R (stop)	
mSox2 d1	GGGGGATCCTCAGTTGAGGCCCGGGTGCTGCGG
mSox2 d2	GGGGGATCCTCACGGTTGCATCTGTGCCGCGCC
mSox2 d3	GGGGGATCCTCAGACGTAGCGGTGCATCGGTTG
mSox2 d4	GGGGGATCCTCAGTACTGCAGGGCGCTGACGAC
mSox2 d5	GGGGGATCCTCAGGTCATGGAGTTGTACTGCAG
mSox2 d6	GGGGGATCCTCACTGCGAGCTGGTCATGGAGTT
mSox2 WT F	GGGGAATTCATGTATAACATGATGGAGACG
mSox2 WT R (stop)	CCCGGATCCTCACATGTGCGACAGGGGCAG
Sox2hmg F	GGGTCTAGAATGGACCGCGTCAAGAGGCCCATG
Sox2hmg R	GGGCTCGAGAAGCGTGTACTTATCCTTCTT
mSox2 dgrgUP	CGATCGGTGCATCGGTTGCATCTGTGC
mSox2 dgrgDN	CGATCGATGACCAGCTCGCAGACCTAC
mSox2 LQY F	CGCTACGTCGTCAGCGCCGCGGCGGCCAACTCCATGACCAGCTCG
mSox2 LQY R	CGAGCTGGTCATGGAGTTGGCCGCCGCGGCGCTGACGACGTAGCG
SUMO	
SU1 F BamHI	GCGGGATCCATGTCTGACCAGGAGGCC
SU1 R EcoRI	GCGGAATTCACCCCCCGTTTGTTCCTG
S2 KpnI R	CCCGGTACCCATGTGCGACAGGGGCAG
SU2 F BamHI	CCCGAATTCACCTCCCTGCTGCTGTTGGAACAC
SU2 R EcoRI	GAATTCACCTCCCTGCTGCTGTTG
SU3 R EcoRI	GCGGGATCCATGGCCGACGAA
SU1 HindIII F	AAGCTTATGTCTGACCAGGAGGCCAAA
SU1 HindIII R	AAGCTTACCCCCCGTTTGTTCCTG
SU2 HindIII F	AAGCTTATGTCCGAGGAGAAGCCC
SU2 HindIII R	AAGCTTACCTCCCTGCTGCTGTTG
SU3 HindIII F	AAGCTTATGGCCGACGAAAAGCCC
Promoters	
GFAP F	AAACTCGAGCCTGTGTTAGTCAGGGTTCTCTAG
GFAP R	AAAAAGCTTTACAGTGAATGGGTAATAAAAATA

Primer name	Sequence 5′ – 3′
Rexp1500	CTCGAGCCTTGCCACAGCCTCACCCTGATAG
Rexp500	CTCGAGTGGAAAAAGTTCAGGCAACTAGTGTAC
Rex R	AAGCTTCTCCTTGGACCCCTCCCTTTTTAGATG
Nanog Pro F	CTCGAGTAAAGTGAAATGAGGTAAAGCC
Nanog Pro R	AAGCTTGGAAAGATCATAGAAAGAAGAG

Table 2.1.1 Primer sequences

2.1.9 TA cloning

Before subcloning into an expression vector, the PCR products were cloned into a pJET vector using a CloneJET^{TM} PCR Cloning Kit (Fermentas). Cloning into the pJET vector allowed DNA sequencing, to confirm insert sequence. The reactions were carried out following manufacturer's instructions. Purified PCR products mixed with T4 DNA ligase (5 units), 1x ligase buffer, 50 ng vector and 150 ng PCR product and water to bring the final reaction volume to 20 µl. The mixture was incubated at room temperature for 5 - 30 minutes before transformation (see section 2.1.6).

2.1.10 Site directed mutagenesis

To produce Sox2 mutants, amino acid changes were created using the QuickChangeII Site-Directed Mutagenesis Kit (Stratagene). Primers were designed to cover 40 nucleotides with the mismatched bases placed in the middle of the primer. pcDNASox2 plasmid was used as template DNA. PCR reactions were carried out following the manufacturer's instruction in the presence of 5% DMSO, for amplification of highly GC-rich DNA. The PCR program was set following the manufacturer's instructions. After digestion of parental DNA by *DpnI*, the DNA products were then transformed into competent cells.

2.1.11 DNA sequencing

Constructs were sequenced using the BigDye Terminator V3.1 kit (Applied Biosystems). Sequencing reactions contained 100 ng DNA, 2 µl BigDye Terminator V3.1, 2 µl sequencing buffer (1x), primer (1.6 pmol) and an appropriate volume of SDW to bring the final reaction volume to 10 µl. The PCR thermal cycler program was set according to the manufacturer's instruction; 96 ºC for 2 minutes as initial denaturation, 30 cycles of 96 ºC for 20 seconds denaturation, annealing under 50 ºC for 20 seconds, and 60 ºC for 1 minute as extension. The sequencing reactions were then precipitated using 3M sodium acetate (pH5.2), 1 µl of glycogen (10 mg/ml) and 100% ethanol. The DNA pellets were dried at room temperature then sent for sequencing (GeneService DNA sequencing, Nottingham).

2.2 Cell culture

2.2.1 COS-7 and P19 cell maintenance

COS-7 cells were maintained in Duldecco's Modified Eagle's Medium (D-MEM, GIBCO 31885) with 10% Fetal Bovine Serum (FBS, Sigma F7524), 0.5% penicillin100 Unit/ml /streptomycin 100 µg/ml (P/S, Sigma T3924), 1% L-Glutamate 200mM (GIBCO 250300) and 0.5% MEM non-essential amino acids 100x (GIBCO 11140). P19 cells were maintained in Alpha MEM (Lonza BE02-002F) supplied with 10% FBS and 0.5% P/S. Cells were cultured in a humidified 5% CO_2 incubator at 37°C.

2.2.2 hNSC cell maintenance

ReNcell^{TM}VM (Millipore) neural stem cells (hNSC) were maintained in 50% DMEM/F12 (GIBCO 21331), 50% NEUROBASAL^{TM} Medium(GIBCO 12348) supplied with 1% N2 supplement (Invitrogen), 2% B27 NeuroMix (Invitrogen), 20 mg/ml basic fibroblast growth factor 2 (bFGF-2, Invitrogen), 200 mg/ml epidermal growth factor (EGF, Invitrogen) and 0.5% P/S. Cells were seeded onto a laminin-coated flask to grow as adherent cells. To coat the flask, 15 µg/ml recombinant murine laminin (Sigma) in DMEM/F12 was applied to cover the surface of the flask and incubated at 37°C for 4 hours. Before use, the laminin solution was removed and the flask rinsed with phosphate buffered saline (PBS).

2.2.3 Storage and revival of cells in frozen stocks

To detach COS-7 and P19 cells, 0.05% Trypsin-EDTA (GIBCO 25300-062) was used, while for hNSCs, Accutase^{TM} (500 unit/ml) (Innovative Cell

Technologies) was used. A pellet of approximate $1x10^6$ cells were obtained by centrifugation at 16.1 g and was resuspended in 1ml culture medium with 10% of DMSO (Sigma D2650). The mixture was then transferred into Cryotubes and stored at -80°C.

To thaw and revive cells from frozen stocks, cells were defrosted quickly in a 37°C water bath and transferred into 5ml of warm medium. After centrifugation at 16.1 g, cells were resuspended in fresh medium and seeded onto culture flasks.

2.2.4 Cell transfection by electroporation

COS-7 cells were cultured to 90% confluence and detached using Tripsine-EDTA. For Co-immunoprecipitation (Co-IP) experiments, $6x10^6$ cells were suspended in 700 µl of medium with up to 20 µg of DNA. For immunostaining, $1x10^6$ cells were suspended in 500 µl of medium with up to 1 µg of DNA. The mixture of cells and DNA were then transferred to 4 mm-gap, 22 mm electrode height electroporation cuvettes (Yorkshire Bioscience) and 210 volts of current was applied for 50 ms using a BTX, ECM 830 electroporator. Cells were then seeded in warm medium in flasks (for Co-IP) or on to poly-D-lysine coated coverslips (for immunostaining).

To prepare the poly-D-lysine coated coverslips, the coverslips were cleaned using Trigene and rinsed with water. Coverslips were soaked in 38% HCl for 30 minutes and rinsed with water. 95% EtOH was used to sterilise the coverslips which were then dried in the cell culture hood. The coverslips were covered by 500 µl of poly-D-lysine hydro-bromide (100 µg/ml, Sigma) overnight. After removing the poly-D-lysine, the coverslips were washed in

water and 95% EtOH. The coated coverslips could be stored in room temperature after being dried.

2.2.5 Cell transfection by liposome reagent

For P19 cells, DharmaFECT®3 transfection reagent (Thermo Scientific) was used following the manufacturer's instruction. For hNSCs, Mouse Neural Stem Cell Nucleofector® (Lonza) was used following the manufacturer's instructions.

2.3 Protein analysis

2.3.1 Cell preparation and Co-immunoprecipitation

Transfected cells were cultured for 24 hours before being used for to Co-IP experiments. The culture medium were removed from the flask and cells rinsed with PBS. Cells were detached by incubating with harvest solution (40 mM Tris-HCl pH7.4, 1 mM EDTA, 150 mM NaCl) for 10 minutes followed by scraping. The whole cell/harvest solution mixture was then transferred into eppendorf and centrifuged at 1844 g for 5 minutes. 300 µl of lysis buffer (0.1% NP40/PBS) and protease inhibitor cocktail (chymotrypsin, 1.5 µg/ml; thermolysin, 0.8 µg/ml; papain, 1 mg/ml; pronase, 1.5 µg/ml; pancreatic extract, 15 µg/ml; trypsin, 0.2 µg/ml; Roche 1 836 170) was added to lyse the cell pellet for 30 minutes at 4°C on a low speed rotating platform. The supernatant was obtained by centrifugation at 10625 g for 10 minutes at 4°C. 30 µl of the supernatant was retained as whole cell lysate (WCE) and the rest was mixed with 25 µl Myc-agarose beads suspension (Sigma) and incubated at 4°C overnight on a low speed rotating platform. The mixture of the cell lysate and beads were centrifuged very gently for 5 seconds and washed with lysis buffer 5-8 times. An equal volume of SDS-PAGE loading buffer was added to WCE and beads. The samples were heated to 100°C for 5 minutes before loading onto SDS-PAGE gels.

2.3.2 Crosslinker treated cell lysate preparation

For the Co-IP of Myc-Groucho and endogenous Sox2 in hNSC, Dithiobis[succinimidylpropionate] (DSP) was used to crosslink proteins. DSP

can form a β-mercaptoethanol breakable crosslink between molecules at a distance of 12Å (Fujita and Wade, 2004; Lomant and Fairbanks, 1976).

Before harvesting the cells, 1 ml of 2 mM DSP/PBS was added onto PBS washed cells and incubated at room temperature for 30 minutes. The crosslinking was stopped by "stop solution" (20 mM Tris-HCl, pH7.5) and the cells were rinsed in PBS twice. To lyse the cells, RIPA lysis buffer (25 mM Tris-HCl, pH7.5, 150 mM NaCl, 1% NP40, 1% Deoxycholate, 0.1% SDS) with cocktail protease inhibitor was used and incubated with cells for 30 minutes. The mixture was transferred into an eppendorf tube and centrifuged at 10625 g for 10 minutes. The cleared cell lysate was used for Co-IP as previously described.

2.3.3 SDS-PAGE electrophoresis

The ingredients of SDS-polyacrylamide gel, running buffer and western blot transfer buffer were made according to Molecular Cloning (Sambrook et al. 1989). SDS-PAGE and western blot equipment were purchased from BioRad. Protein samples were prepared in sample buffer, (laemmli 2x concentrate; Sigma) and run through 4% stacking gel {4% acrylamide/bis-acrylamid, 0.125 M Tris-HCl (pH6.8), 0.1% SDS mixture was polymerized by 0.025 % tetramethyl-ethylenediamine (TEMED) and ammonium persulphate (APS)} and 12% resolving gel {12% acrylamide/Bis-acrylamide (30%/0.8% w/v), 0.375 M Tris-HCl (pH 8.8), 0.1% SDS mixture was polymerized by TEMED and APS in the same way as stacking gel} along side with SeeBlue® Plus2 prestained standard (Invitrogen) ladder. The voltage applied for the stacking gel and resolving gel was 30v and 60v respectively (Laemmli, 1970).

2.3.4 Western blot

The SDS-PAGE gel was placed on a Hybond-C extra nitrocellulose membrane (Amersham) and transferred overnight at 30v in cold transfer buffer (25 mM Tris-base; 192 mM glycine and 20% methanol). The membrane was then blocked in PBST (PBS with 0.3% Tween 20) with 5% skimmed milk (Marvel) for 1 hour. Table 2.2.1 lists all the antibodies used and their working concentration. Primary antibody was diluted in PBST with 3% skimmed milk. After incubation in primary antibody for 1-2 hours at room temperature, the membrane was washed 3x 10 minutes in PBST and then incubated in secondary antibody solution for 1 hour. The membrane was washed 3x 10 minutes before detecting the signal. For all of phosphatase conjugated secondary antibodies, the ECL-plus kit (Amersham) was used following the manufacturer's instructions. Koda medical X ray film (general purpose blue; MOL7010) was used to detect the emitted signal on the membrane. The film was then developed by machine. For Li-cor secondary antibodies, the membrane was scanned by Odyssey infrared imaging system (Li-cor) and the images were processed by the Odyssey 2.1 software.

1st Antibody	Host Species	Dilution	Manufacturer
Anti-myc (9E10)	Mouse	1:20 ICH 1:1000 WB	Hybridoma bank
Anti- Human/Mouse Sox2 (MAB2018)	Mouse	1:200 ICH 1:3000 WB	R & D
Anti-HA (ab9110)	Rabbit	1:4000 WB	Abcam
Anti-pan-TLE	Rat	1:200 WB	Stefani et al 1992
2ed Antibody	Host Species	Dilution	Manufacturer
IRDye 800 CW anti-mouse IgG (Green)	Donkey	1:15000 WB	Li-cor
IRDye 800 CW anti-rabbit IgG (Red)	Donkey	1:15000 WB	Li-cor
IRDye 800 CW anti-goat IgG (Red)	Donkey	1:15000 WB	Li-cor
Anti-mouse IgG, Flurescein-conjugated	Horse	1:200 ICH	Vector
Anti-mouse IgG, Texas Red-conjugated	Hores	1:200 ICH	Vector
Anti-rabbit IgG, Fluorescein-conjugated	Goat	1:200 ICH	Vector
Anti-rat IgG, peroxidise-conjugated	Rabbit	1:1000 WB	Vector

* ICH: Immunocytochemistry; WB: Western Blotting

Table 2.3.1 Antibodies list

2.3.5 Cell preparation and immunostaining

Cells were seeded on coverslips and cultured for 24 hours. After removing the culture medium, cells were rinsed by PBS. The cells were fixed by 4% paraformaldehyde (PFA) in PBS for 10 minutes and permeabilized with 0.2% Triton X-100 (Sigma) in PBS for 20 minutes. Blocking solution (10% skimmed milk, 0.1% Triton X-100/PBS) was added for 30 minutes. Primary antibodies were diluted in PBS according to Table 2.3.1 and incubation was 1 hour at room temperature or overnight at 4°C. After washing three times in wash buffer (0.1% Triton X-100/PBS), cells were soaked in secondary antibodies for 1 hour at room temperature or overnight at 4°C. Cells were washed three times in wash buffer and mounted in DAPI (4,6 diamidino-2-phynylinodole) mounting medium (Vector). Fluorescent signals were detected using a Leica DMRB light microscope connected with a Nikon Digital Sight DS-SM camera by NIS Elements (F2.30, SP2, Build332) image capture software.

2.3.6 Cell preparation and Luciferase assay

Luciferase assays were carried out following the manufacturer's instructions (Dual-Luciferase Reporter 1000 Assay System, Promega), briefly described below. Cells were seeded in a 24 well plate 16 hours before transfection reagent treatment. Every transfection reagent/DNA mixture was prepared for 3-5 wells for a set of repeats. 24 hours after transfection, cells were lysed by 50 µl of passive buffer for 10 minutes at room temperature. The lysates were then transferred into eppendorf tubes and centrifuged at 1844 g for 5 minutes to get a cleared supernatant. 20 ul of the supernatant was transferred into a white

background flat bottom 96 well plate and the luciferase activity was measured by GloMax 96 microplate Luminometer (Promega).

2.4 Yeast manipulations

2.4.1 Yeast maintenance

The Matchmaker3^{TM} two-hybrid system (Clontech) was used to test protein interactions. The yeast strain, AH109 (Clontech), contains ADE2, HIS3, lacZ, and MEL1 reporter constructs that are activated only when GAL4-based protein interactions take place. Yeast were grown on YPD agar plates (pepton 20 g/L; yeast extract 10 g/L; D-glucose 20 g/L; agar 20 g/L pH 5.8) at 30°C until colonies reached 2-3 mm diameter. All the yeast plates could be stored at 4°C for at least one month. For long term storage, the mixture of yeast culture medium with 25% glycerol were kept in cryotubes and stored at -80°C.

2.4.2 Preparation of yeast competent cells

To make competent cells, AH109 colonies were cultured in 10 ml of YPD media containing 0.003% of adenine hemisulphate at 30°C with shaking at 230rpm overnight. The overnight culture was diluted in 50 ml YPDA media (OD_{600} = 0.2) and cultured for another 3 hours until they reached an OD_{600} = 0.5. Cells were collected by centrifugation at 1000g for 5 minutes and washed with 40 ml 1x TE buffer (0.01M Tris-HCl; 1 mM EDTA pH7.5). 1.5 ml of freshly made 1x TE/LiAc (1x TE buffer with 100 mM LiAc) was used to resuspend the cells which were left standing for 10 minutes.

2.4.3 Lithium acetate mediated yeast transformation

The yeast expression vectors (pGAD-T7 containing GAL4 activation domain and pGBK-T7 containing GAL4 DNA binding domain) were used to construct the GAL4 fusion proteins. Here, LiAc transformation was used to introduce the

plasmids into AH109 yeasts. The mixture of transformation solution contained 1 µg of plasmid DNA, 100 µg of denatured herring testes carrier DNA (Clontech), 100 µl of competent yeast cells and 700 µl of freshly made 1x LiAc/40% PEG-3350/1x TE. The mixture was vortexed vigorously and incubated at 30°C with 230rpm shaking for 30 minutes. 88 µl of DMSO (Sigma) was added and tubes gently inverted 2-3 times. Heat shock was carried out by placing the mixture at 42°C for 15 minutes and immediately setting on ice for 2 minutes. Cells were collected by centrifugation at 14000rpm for 5 seconds and washed in 1 ml of 1x TE. Cells were pelleted by centrifugation and resuspended in 100 µl of 1x TE and plated on proper selective plates. The plates were incubated at 30°C for 4-5 days until colonies were visible.

2.4.4 Protein detection in transformed yeast

To test if the transformed yeast express the GAL4 fusion protein, a large colony was picked from the selective plates and inoculated in to 5 ml of selection media. After overnight incubation at 30°C while shaking at 230rpm, the cells were pelleted at 12470 g for 1 minute. To break the cells, half the weight of the pellet of Glass beads (Sigma) and 60 µl of Sample buffer laemmli (Sigma) were added and vortexed twice for 30 seconds with a 30 second interval on ice. The cell lysate was boiled for 10 minutes and centrifuged at 664 g for 1 minute. The supernatant was resolved on SDS-PAGE gel.

2.4.5 Detection of protein interaction by yeast-two-hybridization

When the fusion proteins interacted with each other, the GAL4 DNA-binding domain (BD) and transcriptional activation domain (AD) were coopted onto the GAL4 responsive promoter to trigger the expression of the reporter genes. The interaction of the fusion proteins was therefore detected by growth of the co-transformed yeast on synthetic "drop out" (SD)-leu/-trp/-his/-ade plates. The co-transformed (pGAD-D7/pGBK-T7) yeast expressing *LEU2* and *TRP1* respectively were grown on (SD)-leu/-trp plates. To detect the interaction of the fusion proteins, a single colony was picked and resuspended in 100 µl of sterile distilled water (SDW). 10 µl of the mixture was plated on a SD-leu/-trp/-his/-ade plate and incubated at 30°C for 7 days.

2.5 hNSC mRNA manipulation

2.5.1 hNSC transfection and harvest

hNSCs were transfected using a Mouse Neural Stem Cell Nucleofector Kit (Lonza) according to the manufacturer's instruction. Approximately $5x10^6$ early passage (10-20 passage) hNSCs were mixed with 100 µl of "Nucleofector solution" and the appropriate amount of required plasmid DNA. The cells were transferred into a cuvette and electroporated using a NucleofectorTM II (Amaxa). The transfected cells were seeded on to a laminin coated 75T flask and cultured for 14 hours. All the cells were collected after 5 minutes of AccutaseTM treatment by centrifugation at 1000 g for 5 minutes. The pellets were resuspended in 1 ml of fresh medium.

2.5.2 Fluorescence Activated Cell Sorting (FACS)

Freshly collected hNSCs were kept on ice before sorting. Around $5x10^6$ GFP expressing cells were sorted into fresh medium using a Coulter Altra Flow Cytometer (Beckman). The cells were centrifuged at 1000 g for 5 minutes and the pellets were lysed in 1 ml of TRI reagent (Sigma) immediately.

2.5.3 Total RNA extraction

Extraction of total RNA from transfected hNSCs was carried out according to manufacturer's instructions of TRI reagent, as briefly described below. 0.2 ml of chloroform was added into the mixture of TRI reagent and cell lysate (see section 2.5.2) and vortexed vigorously for 15 seconds. The samples were allowed to stand at room temperature for 15 minutes before being centrifuged at 12,000 g for 15 minutes at 4°C. For RNA extraction, the top clear aqueous

phase was transferred into a new eppendorf tubes and mixed with 0.5 ml of isopropanol by gentle inversions. The samples were allowed to stand for 15 minutes at room temperature or overnight at -80°C before being centrifuged at 12,000 g for 10 minutes at 4°C. The pellet was washed with 75% ethanol and centrifuged at 12,000 g for 5 minutes at 4°C. The pellets were then briefly air dried for 5-10 minutes and dissolve in 100 µl of DEPC treated water (Ambion).

2.5.4 RNA purification

Total RNA samples were purified using the RNeasy® Mini Kit (Qiagen) according to the manufacturer's instructions. 100 µl of RNA sample was mixed with 350 µl of Buffer RLT and 250 ul 100% ethanol and centrifuged through the spin column at 8,000 g for 15 seconds. The spin column was washed with 500 µl Buffer RPE twice. The purified RNA was eluted with 30 µl of RNase-free water and centrifuged at 8,000 g for 1 minute. The total RNA samples were stored in -80°C for further use.

2.5.5 Total RNA quality test

The quality of total RNA samples was tested before use for microarray assay. The quality of purified total RNA was tested using a 2100 BioAnalizer with RNA 6000 Nano kit (Agilent) according to the manufacturer's instructions. The RNA integrity number (RIN) was analysed by 2100 Expert software (Agilent) using the eukaryote total RNA Nano setting.

2.5.6 GeneChip® Affymetrix HG-U133 plus 2 analysis

Purified total RNA of hNSCs was sent to Nottingham Arabidopsis Stock Centre (NASC) for the Affymetrix HG-U133 plus 2 analysis. The brief

principle of Affymetrix HG-U133 plus 2 is described in this section. First, total RNA (1 µg) was used as a template to synthesize double-stranded cDNA using the T7-Oligo(dT) Promoter Primer in the first-strand cDNA synthesis reaction. RNaseH-mediated second-strand cDNA synthesis was carried out on the samples. Double-stranded cDNA was purified and used as a template to synthesize cRNA by *in vitro* transcription (IVT) reaction using T7 RNA Polymerase and a biotinylated nucleotide analog/ribonucleotide mix. The labelled cRNA was cleaned up and fragmented before performing hybridization on microarray chips. A streptavidin phycoerythrin conjugated antibody was used to detect the hybridization on the probe chip. After proper washing and staining, the probe chips were scanned by a GeneArray® Scanner (Affymetrix).

2.5.7 Data analysis

The raw data of Affymetrix array were analysed by GeneSpringGX10 (Agilent) and further manipulation was done using Microsoft Excel. The experimental noise cut-off level was set as a raw fluorescent reading of 100.

2.5.8 Quantitative PCR (qPCR)

Purified RNA was used for the synthesis of cDNA using SuperScript™ III Reverse Transcriptase (Invitrogen) according to the manufacturer's instructions. A mixture of 2 µg of RNA sample, 200 ng of Oligo$(dT)_{18}$ Primer (Fermentase), 1 µl of 10mM dNTP Mix (Bioline) and 14 µl of RNase-free water was heated at 65°C for 5 minutes and then put on ice for 1 minute. The following solutions were added to the mixture; 4 µl 5X First-Strand Buffer, 1 µl 0.1 M DTT and 1 µl of SuperScript™ III RT. The tubes were incubated on a Techne PHC-3

thermal cycler with the programme set to single cycle at 25°C for 5 minutes, 50°C for 60 minutes and 70°C for 15 minutes. cDNA was stored at -80°C until used.

qPCR was carried out using a Rotor-Gene™ 6000 (Corbett life science) with Brilliant SYBR® Green QPCR Master Mix (Agilent). All qPCR data was analysed by Rotor-Gene™ 6000 real-time rotary analyzer version 1.7. The primers used for qPCR were designed according to the following rules. First, the length should be between 100-150 bp. Second, the melting temperature should be around 58-60°C. Third, the target region should cross an intron or be nearby the probes used in the Affymetirx microarray. The sequence details of qPCR primers are listed in Table 2.5.1.

The standard threshold of each set of primers was determined according to the qPCR results of serial template cDNA dilution (200 ng/µl, 40 ng/µl, 8 ng/µl and 1.6 ng/µl) by the auto-find function in Rotor-Gene™ 6000 real-time rotary analyzer version 1.7 software. To prepare the qPCR reaction, 7.2 µl of cDNA (50 ng/µl), 4.3 µl of each primer (20 nmole), 37.5 µl of 2X Brilliant SYBR® Green QPCR Master Mix and 26 µl of distilled water were carefully mixed in a 0.5 ml tube and aliquoted in three 0.1 ml strip tubes (Qiagen). The samples were placed in the Rotor-Gene™ 6000 and the PCR programme set at 95°C for 15 seconds, 60°C for 20 seconds and 72°C for 20 seconds. The threshold cycle (Ct) values of each sample was defined by standard threshold and the relative comparison with the housekeeping gene (Actin, GAPDH and Cyclophilin B) were calculated using the equation below (Pfaffl, 2001) :

$$\text{ratio} = \frac{(E_{\text{target}})^{\Delta \text{CP}_{\text{target}}(control-sample)}}{(E_{\text{ref}})^{\Delta \text{CP}_{\text{ref}}(control-sample)}}$$

Primer	Sequence 5′-3′
Hells QF	GGGCGCAAGAGAAGCCGTACT
Hells QR	TGCTGGATGCCTCTCTTCAGGT
IGFBP2 QF	CCACCATGCGCCTTCCGGAT
IGFBP2 QR	CCACGCTGCCCGTTCAGAGAC
IGFBP3 QF	GATGGCTTTTGCTGCGGCCC
IGFBP3 QR	TGCAGTCATCCGAAGAATTGTGCCA
NNMT QF	TCCCCACCTACTGCAGGGCG
NNMT QR	CCCAGGGGGAGGCTGGAGAA
Serpine1 QF	TCTGCCCACTCGGGTCTGCA
Serpine1 QR	CTCCGCGGTGGCAGGCAGTA
Sox8 QF	CTGTCCCCGACTGTGCCACG
Sox8 QR	ACTGGTGCCTGCGAGCCAAG
Tagln QF	TGGGCAGCTTGGCAGTGACC
Tagln QR	CCCTCCTGCAGCTGGCTCTCT
GFAP QF	CGGGTGCTCAGGGCTGACAC
GFAP QR	AAGTGGGCCCTCCCAGTCCC
Sox2 coding region F	AAAAACAGCCCGGACCGCGT
Sox2 coding region R	GCCCAGGCGCTTGCTGATCT
Sox2 non-coding region F	ACAGCGCCCCGCATGTACAACA
Sox2 non-coding region R	TTCATGGGCCGCTTGACGCG
GAPDH F	TGCACCACCAACTGCTTAGC
GAPDH R	GGCATGCACTGTGGTCATGAG
Actin F	TCTACAATGAGCTGCGTG
Actin R	ATCTCCTTCTGCATCCTGTC
Cyclophilin B F	GGCCGGGTGATCTTTGGTCTCTTC
Cyclophilin B R	CCCGGCTGTCTGTCTTGGTGCTCT

Table 2.5.1 sequence of qPCR primers

2.6 Zebrafish manipulation

2.6.1 Zebrafish (Danio rerio) maintenance and embryo collection

Standard procedures (Westerfield, 2000) were followed in maintaining zebrafish (mix strain of Tübingen and AB). Embryo stages were determined according to Kimmel et al. (1995). System water containing fungicide (methylblue) was used during embryo manipulation procedures. All embryos were incubated at 28°C.

2.6.2 *in vitro* transcription of mRNA for embryo injection

Sense RNAs were produced from plasmids (Table 2.6.1) using a mMachine® high yield capped mRNA transcription kit (Ambion) according to the manufacturer's instructions. 10 µg of plasmids was digested with an appropriate restriction enzyme (to cut 3' of the insert) in 50 µl reactions at 37°C overnight and the linearised plasmids were purified by a GENECLEAN® II kit (MP Biomedical) according to the manufacturer's instructions. The mixture of template DNA (linearised plasmids), 1x NTP/CAP, 1x reaction buffer and 2 µl enzyme were incubated at 37°C for 2 hours. The template DNA was digested by adding 1 µl DnaseI (1Unit/µl; Sigma) and incubated at 37°C for 15 minutes. The reaction was stopped by adding 15 µl ammonium acetate (5M) and 115 µl nuclease free water (Ambion). RNA was purified by mixing with an equal volume of buffer-saturated phenol/chloroform (Sigma) and vortexed for 30 seconds. The mixture was centrifuged for 3 minutes at 16162 g and the clear upper layer was transferred into a clean eppendorf tube. An equal volume of chloroform was added, vortexed for 30 seconds and centrifuged for

3 minutes at 16162 g. The clear upper layer was again transferred into clean eppendorf tube. 0.1x volumes of 3 M sodium acetate and 1x volumes of isopropanol were added and kept at -20°C for 30 minutes to precipitate the RNA. The RNA was pelleted by centrifugation at 16162 g for 30 minutes at 4°C. The pellet of RNA was resuspended in 30 µl nuclease free water and quantified by a NanoDrop 1000 Spectrophotometer. The RNA was stored at -80°C for further use.

Gene	Vector	Linearization restriction enzyme	*In vitro* Transcription primer
sox2	pBUT2-HA	EcoR I	T3
sox2$^{LQV\text{-}AAA}$	pBUT2-HA	EcoR I	T3
GFP	pCS2-nls	Not I	SP6

Table 2.6.1 Vectors for sense mRNA synthesis.

2.6.3 Zebrafish embryo microinjection

The glass needles used for microinjection were prepared by Lab technicians using a needle puller (Sutter Instrument CO). Needles were loaded with 0.5 µl of the required concentration (as described in figure legend of chapter 6) of mRNA and mounted on the micromanipulator connected to a PICOSPRITZER®III picopump (Parker Instrumentation). For the 1-2 cell stage, 0.5 nl of mRNA was injected. As for 16-32 cell stage, 0.25 nl of mRNA was injected. The volume of the drop size was adjusted to 100 µm in diameter for 0.5 nl and 50 µm for 0.25 nl. The embryos were lined up along the side of a glass microscope slide and a defined volume of mRNA was injected through the chorion and yolk into the cell.

2.6.4 Preparation of digoxygenin (Dig) labelled RNA probe for *in situ* hybridization

For anti-sense RNA probes, 10 µg of vectors was digested with an appropriate restriction enzyme (Table 2.6.2) and purified (see section 2.6.2). A mixture of 1 µg of purified linearised DNA template, 5mM DTT (Promega), 10x DIG labelling mix (Roche), 1 µl of RNasin RNase inhibitor (40Unit/ µl ; Promega), 5x transcription buffer (Roche), 2 µl of either T3/T7/SP6 polymerase (17Unit/ µl; Roche) and 7 µl RNase free water was incubated at 37°C for 3 hours. Probes were cleaned using Microspin^{TM} G50 columns (GE Health Care) and diluted 100x in hybridisation buffer I (hybeI) for -20°C long term storage.

Gene	Linearization Restriction enzyme	Transcribe primer	Source/Reference
sox31	Sal I	T7	Dee et al. 2007
bmp2	EcoR I	T3	Furthauer et al. 2004
bmp4	EcoR I	T3	Nikaido et al. 1997
n-cadherin	Xho I	T7	Lele et al. 2002
shh	BamH I	T7	Krauss et al. 1993
Krox20	Pst I	T7	Sun et al. 2007
emx1	BamH I	T7	Morita et al. 1995
goosecoid	Sma I	T7	Dr. Martin Gering
chordin	Not I	T7	Dr. Martin Gering
gata2	EcoR I	T7	Read et al. 1998
gata3	EcoR I	T7	Neave et al. 1995
p63	Sal I	T7	Bakkers et al. 2002
bozozok	BamH I	T7	Dr. Martin Gering

Table 2.6.2 Probes for *in situ* hybridisation.

2.6.5 Zebrafish embryo in situ hybridisation

The solutions used for *in situ* hybridisation are listed in Table 2.6.3. Embryos were fixed at the required stages by soaking in 4% PFA in PBS overnight at 4°C. The PFA was washed out by soaking in PBS with 0.1% Tween20 (PBSTw) 5 minutes 3 times before dechorionation. The embryos were manually dechorionated and dehydrated through a series of incubations in 25%,

50%, 75% and 100% methanol in PBSTw for 5 minutes in each step. 100% methanol was added to cover the embryos. The embryos were then stored at -20°C overnight or longer for future use.

Rehydration was carried out through serial 5 minute incubations in 75%, 50% and 25% methanol in PBSTw and 4 times 5 minutes washing in PBSTw. The embryos were equilibrated for 5 minutes in 50% hybI in PBSTw at room temperature and prehybridized for 1 hour by hybI at 65°C. 1:200 dilution of RNA probe was used to incubate with embryos at 65°C overnight.

Serial 10 minutes washes were carried out at 65°C in 100%, 75%, 50% and 25% hybridisation buffer II (hybeII) in 2x saline-sodium citrate (SSC) and incubated in 2x SSC for 10 minutes. After 10 minutes washing in 0.2% SSC, 4 times at 65°C, the embryos were taken through a series of 5 minutes incubations with 75%, 50% and 25% SSC in maleic acid buffer containing 0.1% Tween20 (MABTw) and then 100% MABTw at room temperature with gentle shaking at 5rpm. 1 hour blocking with 2% Boehringer blocking reagent in MABTw was carried out to block the non-specific antibody binding. The blocking solution was replaced with blocking reagent containing 1:5000 anti-Dig antibody and incubated at 4°C overnight.

The embryos were washed 8 times in MABTw for 15 minutes at room temperature while shaking at 5rpm and then equilibrated with BCL buffer III 3 times for 5 minutes. Colour development was performed in the dark by incubating with 50% BM Purple (Roche) in BCL buffer III while shaking gently. To stop the colour developing, embryos were washed 3 times for 5 minutes in PBSTw and incubated in 4% PFA for 20 minutes. The PFA was

rinsed out by three 5 minute PBSTw washes. The embryos were stored in 80% glycerol at 4°C.

The embryos were observed under a Nikon SMZ1500 microscope and the staining patterns were photographed by a Nikon ACT-2U 1.40 software controlled digital camera (Nikon DS-5M, DS-U1).

Solution	Ingredients
20x Saline sodium citrate (SSC)	3 M NaCl, 0.3 M sodium citrate, pH7
Maleic acid buffer (MAB)	0.1 M melic acid, 0.15 M NaCl, adjusted to pH7.5 with NaOH
Hybrisation solution I (Hyb I)	50% v/v formamide, 5% SSC, 0.5 mg/ml *E.coli* tRNA (Sigma), 5 ug/ml heparin (Sigma), 9.2 mM 1M citric acid, 0.1 Tween20
Hybrisation solution II (Hyb II)	50% v/v formamide, 5% SSC, 9.2 mM 1M citric acid, 0.1% Tween20
BCL buffer III	100 mM Tris-HCl (pH 9.5), 100 mM NaCl, 50 mM $MgCl_2$, 0.1% Tween20
4% PFA	4% paraformaldehyde in 1 x PBS
Deyolking buffer	55 mM NaCl, 1.8 mM KCl, 1.25 mM $NaHCO_3$

Table 2.6.3 Solutions for *in situ* hybridisation

2.6.6 Zebrafish embryo protein detection

To detect protein expression in the Zebrafish embryos, around 20 dechorionated embryos were collected. The embryonic yolk was dissolved by pipetting in 300 µl of deyolking buffer (Table 2.6.3) and 5 minutes rocking incubation in room temperature. After centrifugation at 73 g for 2 minutes, 20 µl of 2X Sample Buffer Laemmli (Sigma) was added and sample analysed by SDS-PAGE and western blotting with appropriate antibodies (see section 2.3.4).

Chapter 3 Sox2 interacts with Grouchos

3.1 Introduction

The Tcf family are downstream effectors of the Wnt signalling. In the absence of a Wnt signal, Tcf interacts with Groucho keeping target genes suppressed as shown in Figure 3.1.1 (Brantjes et al., 2001). Tcf is a well-studied transcriptional repressor which suppresses down-stream genes via recruited co-repressors such as Groucho and CtBP. Since the HMG domain on Tcf is highly conserved within the SoxB1 family members, the same repression mechanism is very likely to occur within the SoxB1 family.

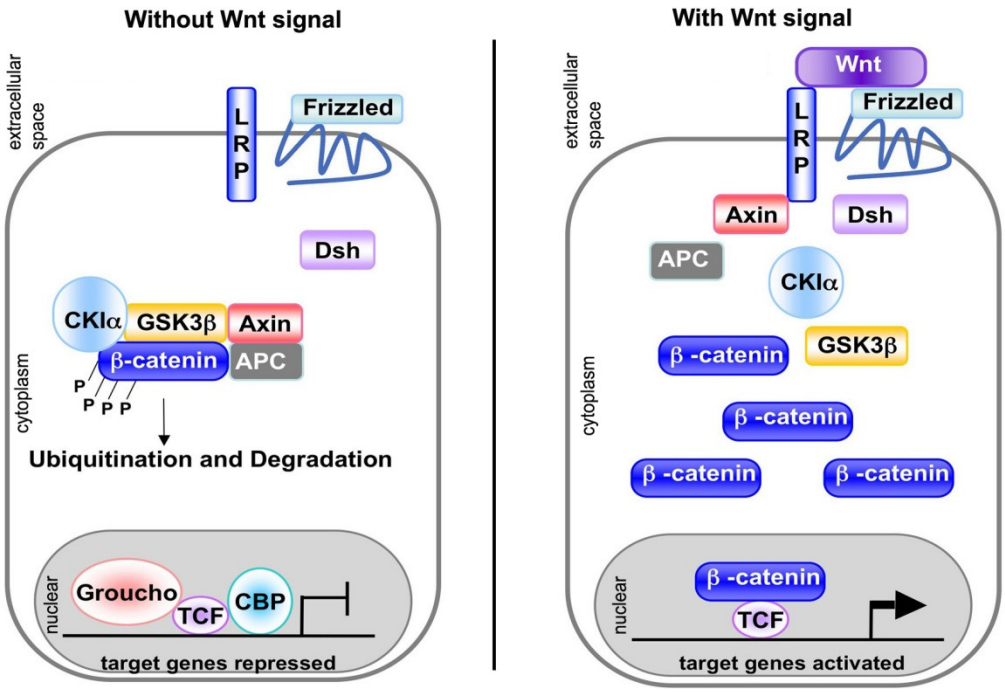


Figure 3.1.1 Schematic representation of canonical Wnt signalling cascade.

When a Wnt signal is absent, β-catenin forms a complex containing Axin, APC, CKIα and GSK3β. The phosphorylation of β-catenin by GSK3β results in ubiquitination and degradation of β-catenin and keeps the downstream genes repressed by the Tcf and Groucho complex. The Wnt signalling prevents the β-catenin complex forming which leads to the accumulation of β-catenin in the nuclear. β-catenin replaces the co-repressor in the Tcf complex. The interaction of β-catenin and Tcf triggers the downstream genes transcription.

Previous research in our lab has shown that Sox3, a SoxB1 family protein, interacts with the co-repressor Groucho (PhD thesis, Caroline Hirst, 2009). This interaction of transcription factor and co-repressor implies a repressor function for the SoxB1 family, which are generally regarded as transcriptional activators (Kamachi et al., 1999; Uchikawa et al., 1999). A repressor function for Sox3 in various aspects of early embryo developmental has also been demonstrated in *xenopus*, zebrafish, chicken and mouse (Shih et al., 2010; Zhang et al., 2004; Zhang et al., 2003a). Hence, it is conceivable that the highly similar SoxB1 family members may all function as repressors through

interactions with the Groucho co-repressors. In this chapter, experiments were designed to test the interaction of Sox2 and Groucho.

3.2 Sox2 changes Groucho subcellular localization

Interaction between proteins can be tested by many methods including protein affinity chromatography, immunoprecipitation and yeast two hybrid assays. One rapid assay of protein interactions is to observe the subcellular localization change after co-expression in cells. Sox2 has three nuclear localization signals in its HMG DNA binding domain, therefore the majority of overexpressed Sox2 locate in the nucleus (Figure 3.2.1). The long forms of Groucho (Groucho 1-4) have casein kinase II/cdc2 phosphorylation sites and a nuclear localization sequence (CcN) which allows them also to locate to the nucleus, whereas Groucho5 lacks a nuclear localization signal and in both nuclear and cytoplasmic localization (Figure 3.2.2).

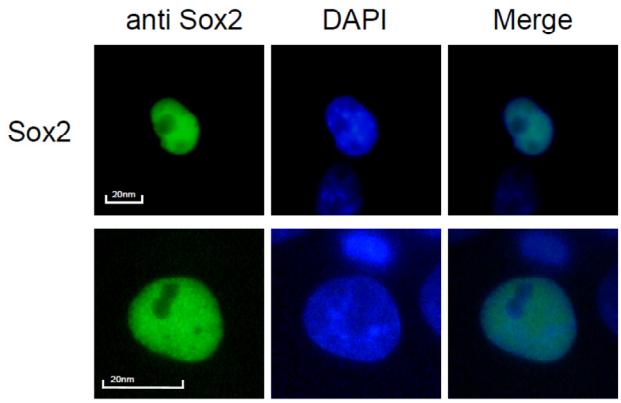


Figure 3.2.1 Sox2 subcellular localiziation

COS7 cells were transfected with Sox2 by electroporation and seeded on poly-D-lysine coated coverslips. After 24 hours, the cells were immunostained using anti-Sox2 antibody. DAPI staining identified the nucleus of the cells. Sox2 was detected in the nucleus.

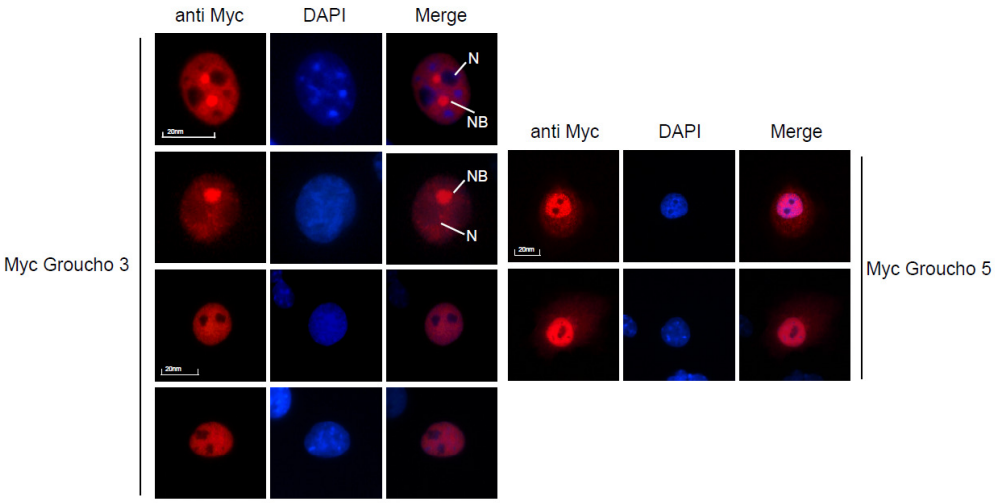


Figure 3.2.2 Groucho protein subcellular localization

COS7 cells were transfected with Myc-Groucho3 or Myc-Groucho5 by electorporation and seeded on poly-D-lysine coated coverslips. After 24 hours, the cells were immunostained using anti-Myc antibody. Groucho3 was detected in the nucleus and 10% of the Groucho3 stained cells exhibited nuclear bodies which did not co-localize with the nucleolus. Groucho5 was detected in the cytoplasm and nucleus, with a brighter staining in the nucleus than the cytoplasm. (N: Nucleolus; NB: Nuclear body)

According to the previous studies in our lab, the Sox3 and Groucho5 interaction was demonstrated using a nuclear translocation assay (PhD thesis, Zulfiqar Laghari, 2010). Here, the same experiment was carried out to investigate the interaction of Sox2 with Groucho5. As a positive control, COS7 cells were transfected with Sox3 and a Myc-tagged Groucho5 either alone or together and immunostaining was used to detect their subcellular localization (Figure 3.2.3). Sox3 localized to the nucleus and, also expressed alone, Myc-tagged Groucho5 was broadly distributed in cytoplasm and nucleus. When co-expressing both Sox3 and Groucho5, Groucho5 was mainly relocated to the nucleus together with Sox3 which implied an interaction with Sox3. These results agree with the previous studies in our lab.

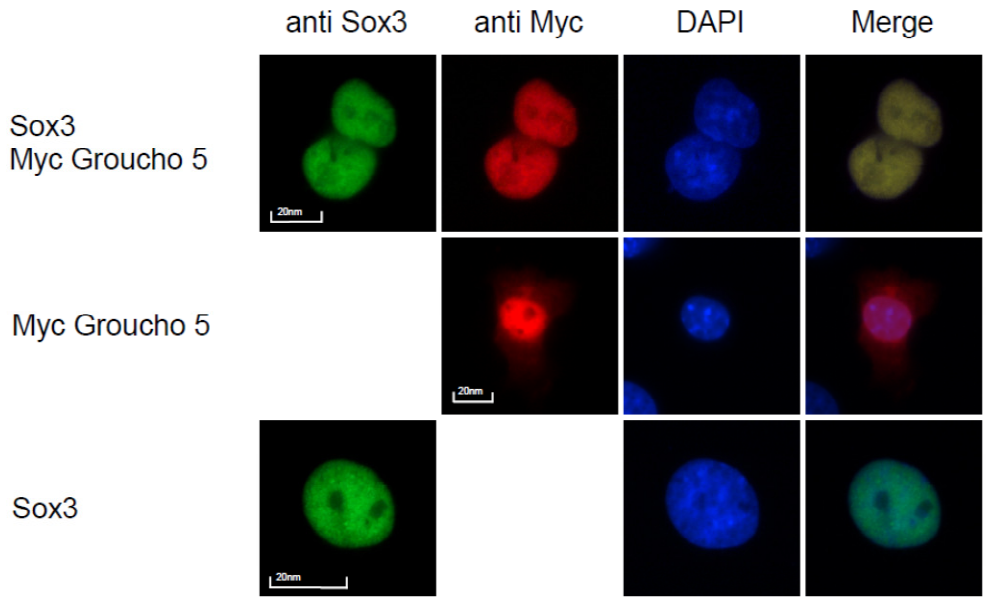


Figure 3.2.3 Sox3 translocates Groucho5 to the nucleus

COS7 cells were transfected with Sox3 or Myc-Groucho5 by electorporation and seeded on poly-D-lysine coated coverslips. After 24 hours, the cells were immunostained using anti-Sox3 and anti-Myc antibody. When expressed alone, Sox3 was detected in the nucleus while Myc-Groucho5 was detected in the cytoplasm and nucleus, with a brighter staining in the nucleus than the cytoplasm. Co-expressing Sox3 with Groucho5 caused Groucho5 to translocates into nucleus.

A similar analysis was carried out for Sox2 and Groucho5 interaction. When transfected alone, Groucho5 was distributed throughout the cells but staining was more intense in the nuclei (Figure 3.2.4.A). When both Sox2 and Groucho5 were overexpressed in COS7 cells, the nuclear Groucho5 was still detected whereas the cytoplasmic Groucho5 was significantly reduced (Figure 3.2.4.B; $p < 0.001$). This nuclear restricted localization of Groucho5 represents co-localization with Sox2. This co-localization of Sox2 and Groucho5 implied the interaction between two proteins. Quantification of these data demonstrated a distinct increase in the nucleus of cells that had nuclear only Groucho5 (Figure 3.2.4.C).

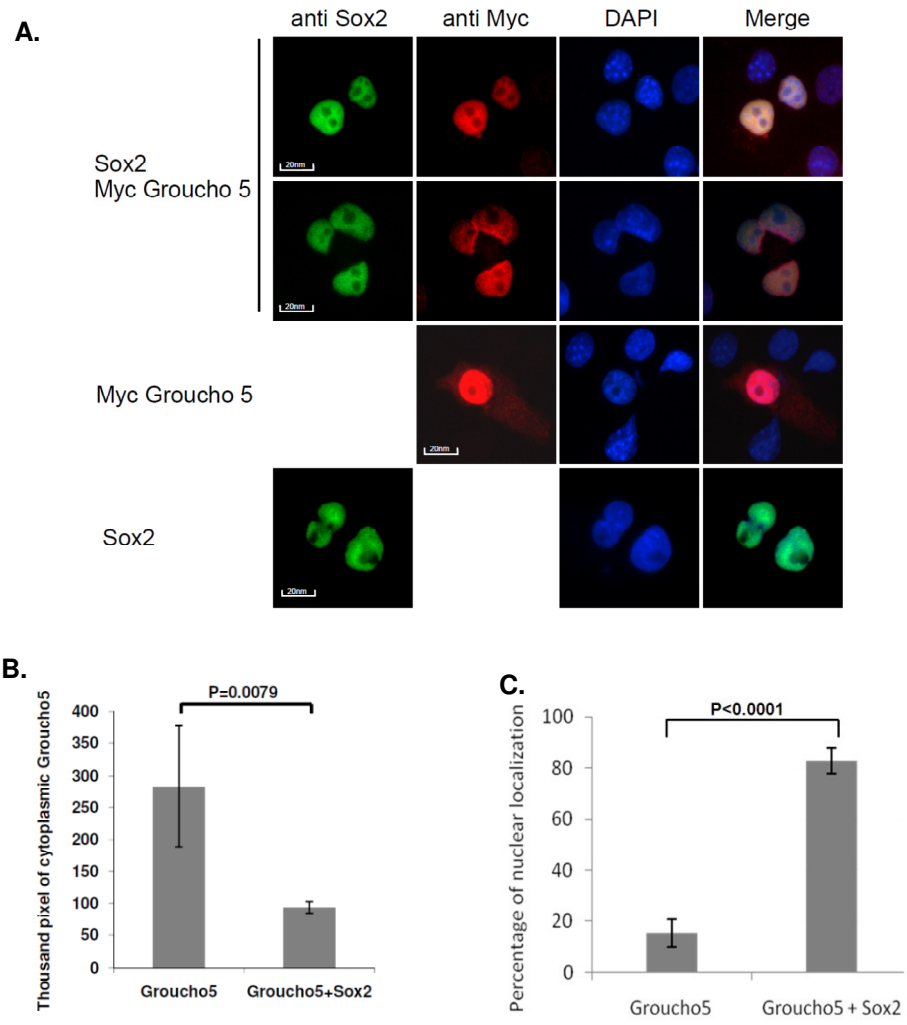


Figure 3.2.4 Sox2 translocates cytoplasmic Grouch5 to the nucleus

COS7 cells were transfected with Sox2, Myc-Groucho3 or both by electorporation and seeded on poly-D-lysine coated coverslips. After 24 hours, the cells were immunostained using anti-Sox2, anti-Myc or both antibodies. (A.) When co-transfected with Sox2, Groucho5 was detected in the nucleus, but no Groucho5 was detected in the cytoplasm. (B.) Quantification of cytoplasmic Groucho5. The unsaturated Groucho5 immunostaining picture were analysed by ImageJ software. The pixel number of cytoplasm was calculated as whole cell pixel number minus the nucleus pixel number. 15 cell pictures were analysed per set. (C) Quantification of Groucho5 nuclear localization. 100 Groucho5-expressing cells were counted. The number of cells express no cytoplasmic Groucho5 were counted and the percentage was calculated.

In order to test the interaction between Sox2 with Groucho, a construct encoding Sox2 and/or Myc-tagged Groucho3 was introduced into cells either alone or together and the subcellular localization of the proteins produced was detected using immunostaining. When cells were transfected with Myc-Groucho3 construct alone, Myc-Groucho3 protein inclusion bodies were seen, which did not co-localize with the nucleolus. When both Sox2 and the Myc-Groucho3 were introduced into COS7 cells simultaneously, the inclusion bodies of Groucho3 stained brighter in the periphery of than the centre and Sox2 now also formed inclusion bodies which co-localized with Groucho3 (Figure 3.2.5.A). This subcellular localization change of both Sox2 and Groucho3 when co-expressed in the cell implied interaction between the two proteins. To further quantify the frequency of the phenomena, the cells with or without Groucho3 nuclear bodies were counted in cells transfected with Groucho3 alone or with both Groucho3 and Sox2 co-expressed. The results in presented in Figure 3.2.5.B as percentage of the dots formed in the nucleus. When Groucho3 was expressed alone, 1% of cells had Groucho3 inclusion bodies with bright peripheral staining. When Groucho3 was co-transfected with Sox2, approximately 16% of Groucho3 positive nuclear bodies were colocalized with Sox2 and exhibited bright peripheral staining.

To sum up, the subcellular localization change of Grouchos after introducing Sox2 into cells suggests some kind of connection between Sox2 and Grouchos

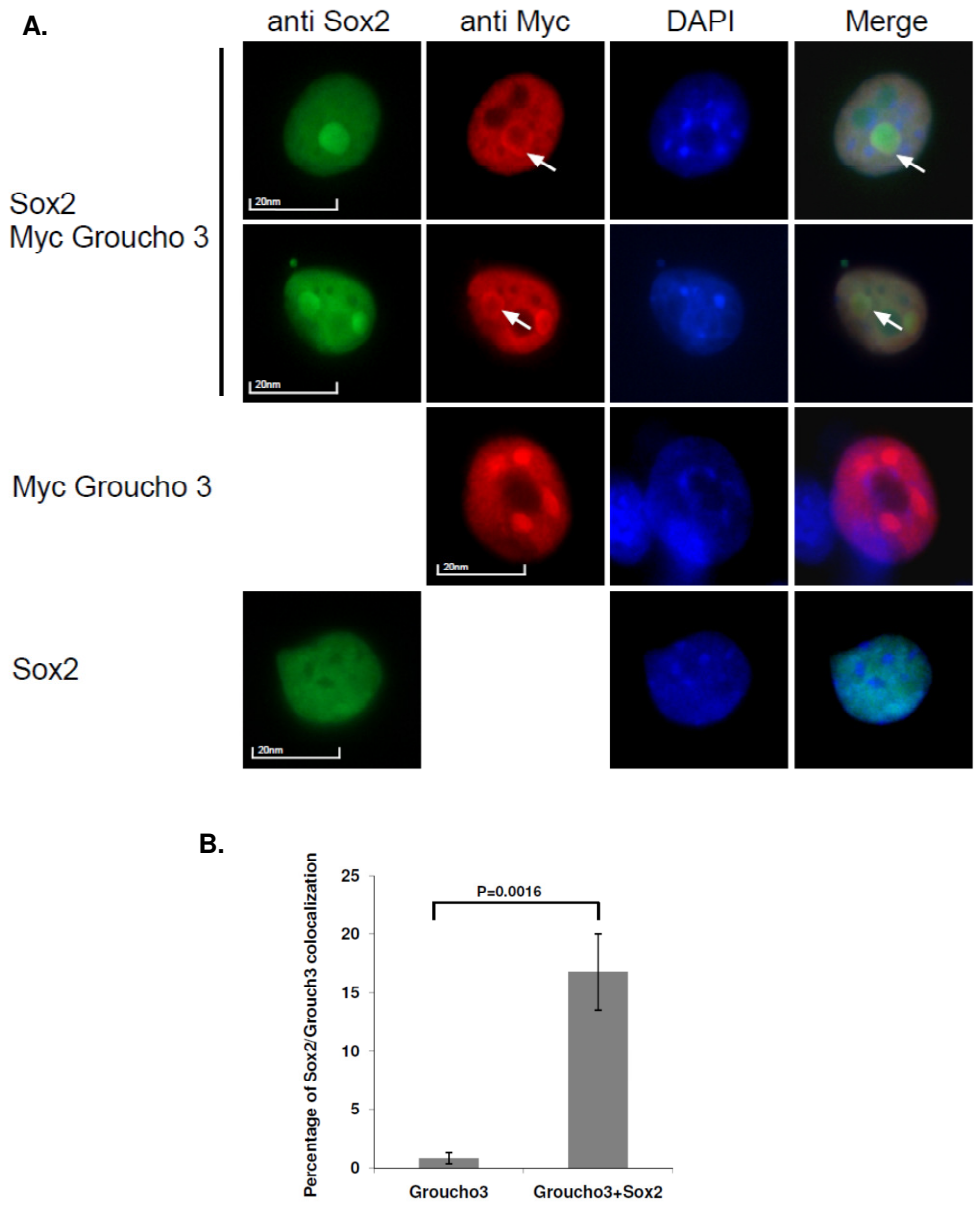


Figure 3.2.5 Sox2 co-localized in the nucleus with Groucho3

COS7 cells were transfected with either Sox2, Myc-Groucho3 or both constructs by electroporation and seeded on poly-D-lysine coated coverslips. After 24 hours, the cells were immunostained using anti-Sox2, anti-Myc or both antibodies. (A) When co-transfected with Groucho3, Sox2 formed inclusion bodies which colocalized with Groucho3. Groucho3 was detected in the nucleus and the inclusion bodies of Groucho3 only stained in the periphery inclusion bodies with bright edge were counted. The number of cells expressing Groucho3 was counted and the percentage of inclusion bodies with bright edge was calculated (first column). The number of cells co-localized Sox2 and Groucho3 in inclusion bodies was counted and the percentage was calculated (second column).

.

3.3 Groucho proteins interact with Sox2 as demonstrated by Co-immunoprecipitation

The change of Sox2 and Groucho subcellular localization when co-expressed implies interaction between these two molecules. However, the image under the microscope is a merged vision of the whole depth of the cell. Therefore, it could give a false positive result such as overlapping signal of different depths of focus and could also be due to indirect interaction via an intermediate protein complex. To further investigate the relationship between Sox2 and Grouchos, co-immunoprecipitation (Co-IP) was carried out to test the protein-protein interaction. Using the affinity of antibody and antigen, complexes in a cell lysate can be precipitated with a specific antibody conjugated to sepharose beads. The complexes are then separated by SDS-PAGE and analysed by western blot.

After transfection with appropriate plasmids, cells were recovered and harvested after being grown for 48 hours. The cell lysates were then used for co-immunoprecipitation assay. Here Myc antibody conjugated sepharose beads were used to precipitate the Myc tagged proteins. The precipitated samples were run through SDS-PAGE to separate the proteins according to their molecular weight and transferred onto a nitrocellulose membrane. Myc-Groucho was detected by Myc-antibodies and Sox2 detected using Sox2 antibodies.

3.3.1 Myc-bead specificity test

A control for the Myc-bead specificity was carried out in Sox2 and Myc-Groucho overexpressing cell lysates respectively (Figure 3.3.1). The

expression of Sox2 was shown in 5% whole cell extract (WEC). Lanes 1 and 2 show Sox2 expression at approximately 64kDa which is different from the predicted size of Sox2 (~34kDa). Since the Sox2 was overexpressed in the cell, it might trigger a specific degradation mechanism. The ubiquitin-proteasome pathway is one of the key protein degradation mechanisms (Ciechanover, 1998) and sumoylation is also involved in this pathway (Tatham et al., 2008). Hence the molecular weight difference of Sox2 might be caused by these post-transcriptional modification (Baltus et al., 2009b; Van Hoof et al., 2009). However, a single ubiquitination or sumoylation event only adds approximately 11kDa. The molecular weight difference of exogenous Sox2 is 30kDa larger than the predicted size. Hence other modifications seem likely.

After immunoprecipitation using Myc-beads, no Sox2 band was seen (lane 3, 4), which indicates that Myc-beads alone could not precipitate Sox2. The expression of Myc-Groucho3 and Myc-Groucho5 in 5% whole cell extract was not detectable (lane 6, 7). However, the Myc-beads could precipitate and concentrate Myc-tagged Groucho from the rest of whole cell extract (lane 9, 10) whereas the non-transfected cell lysate showed nothing has been precipitated (lane 8).

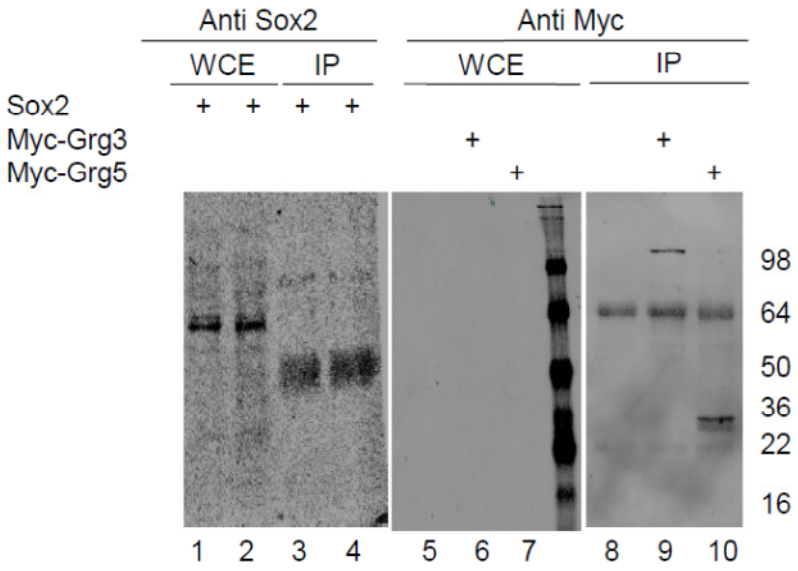


Figure 3.3.1 Myc beads specificity test

Sox2 expression in COS7 cells was detected using anti-Sox2 antibody in 5% whole cell extract (WCE). After immunoprecipitation (IP), the Myc beads did not precipitate overexpressed Sox2 (left panel). Myc-Groucho3 and Myc-Groucho5 expression could not detected by anti-Myc antibody in WCE (middle panel), but immunoprecipitation concentrated Myc-tagged proteins, which were detected by the anti-Myc antibody (right panel).

3.3.2 Co-immunoprecipiation of Sox2 and Groucho

To test the interaction between Sox2 and Groucho, COS7 cells were transfected with Sox2 and Myc-tagged Groucho3 or Groucho5. Co-immunoprecipitations were carried out with Myc-beads. The whole cell extract and the protein eluted from the Myc-beads after co-immunoprecipitation were analysed by western blotting with anti-Myc and anti-Sox2 antibodies. The exogenous expression of Myc-Groucho3, Myc-Groucho5 and Sox2 was faintly detected in 5% whole cell extract (Figure 3.3.2) (lane 1-4; 9-12). Myc-beads precipitated Myc-Groucho3 (lane 5, 6) and Myc-Groucho5 (lane 7, 8) suggesting that both Myc-Groucho3 and Myc-Groucho5 were expressed.

However, the molecular weight of Myc-Groucho5 is very close to the light chain of immunoglobulin (approximate 36kD), it is hard to distinguish the slightly higher Myc-Groucho5 bands and the bands of immunoglobulin (lane 7, 8). Although no obvious bands was detected by the Sox2 antibody in the whole cell lysate, the Sox2/Groucho interaction allowed Sox2 to be precipitated with Myc-beads and allowed Sox2 to be detected in lane 14 and 16. No similar molecular weight protein was precipitated by Myc-beads in the absence of co-transfected Sox2 (the negative control which had no Sox2 co-expression; lane 13, 15) which suggests the bands in lane 14 and 16 are genuinely Sox2. Nevertheless, the anti-Myc antibody detected a band at about the same position as Sox2 (lane 6, 8) which implied that the anti-Myc antibody might weakly cross-react non-specifically in this overexpression system. To further confirm the interaction of Sox2 and Groucho, it was necessary to co-immunoprecipitate endogenous Sox2.

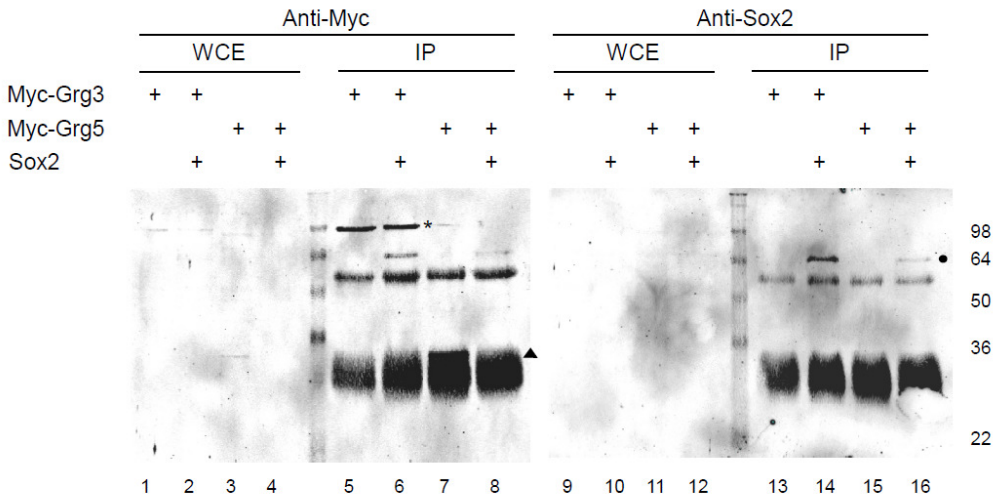


Figure 3.3.2 Co-immunoprecipitation of exogenous Sox2 through Myc-Groucho

COS7 cells were transfected with Sox2 and Myc-Groucho3 or Myc-Gruocho5. After 48 hours, the cell lysates were used for Myc immunoprecipitation. Western blots were carried out and detected with anti-Myc (lane 1-8) and anti-Sox2 antibodies (lane 9-16). The expression of transfected proteins was weakly detected in the whole cell extract (WEC; lane 1-4 and 9-12) whereas immunoprecipitation concentrated the proteins which could then be detected more clearly on the membrane (lane 5-8 and 13-16). Anti-Myc antibody detected the immunoprecipitated Myc-Groucho3 (*) and Myc-Groucho5 (▲). Anti-Sox2 antibody detected Sox2 (●) that had been precipitated with Groucho. (WCE: 5% of whole cell extract; IP: Immunoprecipitation; Grg3: Groucho3; Grg5: Groucho5.)

3.3.3 Co-immunoprecipiation of endogenous Sox2 and Groucho

In order to test the Grocho/Sox2 interaction in a more endogenous context, COS7 cells (which express endogenous Grouchos) were transfected with Myc-Sox2 and the interaction of Myc-Sox2 with endogenous Grouchos was tested by co-immunoprecipitation (Figure 3.3.3). The expression of Myc-Sox2 was shown in the whole cell extract (lane 8). After immunoprecipitation, Myc-Sox2 could only seen in the transfected cells (lane 5). A anti-pan TLE antibody was used that detected the C-terminus of long form Grouchos. Therefore, the anti-pan TLE antibody detected a high molecular weight band at approximate 98 kDa as predicted. In the whole cell extract, Grouchos could be detected in lysated of both transfected and non-transfected cells (lane 1, 2). After co-immunoprecipitation, only the Myc-Sox2 transfected cell lysate showed a Groucho band (lane 3). This indicates that the endogenous Grouchos in COS7 cells could also be precipitated together with Myc-Sox2 using Myc beads (Figure 3.3.3).

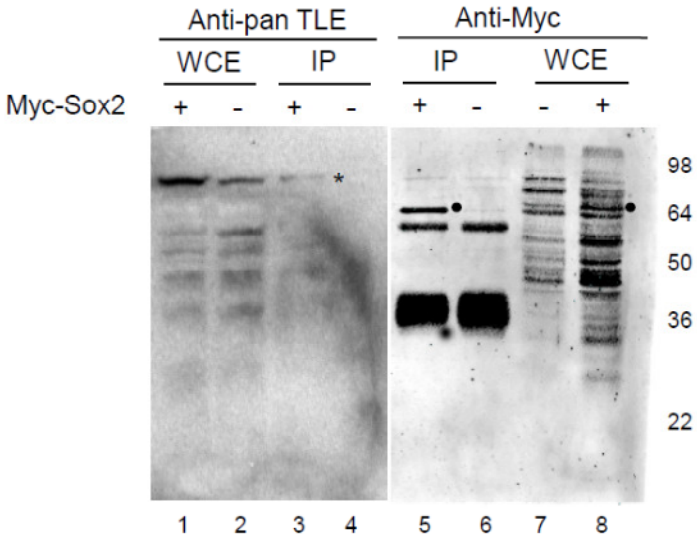


Figure 3.3.3 Co-immunoprecipitation of endogenous Groucho with Myc-Sox2 in COS7 cell

COS7 cells were transfected with Myc-Sox2. The expression of Myc-Sox2 was detected by anti-Myc antibody (lane 5-8). The Myc-beads did precipitate Myc-Sox2 after IP (lane 5, ●). The expression of endogenous Groucho was detected by anti-pan TLE antibody (lane 1-4). When overexpressing Myc-Sox2, the endogenous Groucho could be detected in the Myc-Sox2 pulled down complex (lane 3, *).

On the other hand, to test if the endogenous Sox2 also interacts with Groucho, human neural stem cells (hNSC) (which express high levels of endogenous Sox2) were used (Figure 3.3.4 A, B). Co-immunoprecipitation was carried out using Myc-Groucho3 and Myc-Grouch5 transfected hNSC cell lysates. The expression of Myc-Groucho3 and Myc-Groucho5 were shown in whole cell lysate (Figure 3.3.4 A lane 1, 2) and Myc-beads successfully precipitated both Myc-Groucho3 and Myc-Groucho5 (Figure 3.3.4 A lane 3, 4). Endogenous Sox2 expression was detected by anti-Sox2 antibody in whole cell extract at approximate 40 kDa (Figure 3.3.4 A lane 5, 6), which is close to the predicted Sox2 molecular weight. Several other bands were also detected by the anti-Sox2 antibody, one of them located at approximate 64 kDa, just like the

exogenous expression of Sox2 in COS7 cells. This implies the post-translational modification occurred also on the endogenous Sox2. However, the co-immunoprecipitation did not precipitate the endogenous Sox2 in this experiment (Figure 3.3.4 A lane 7, 8). Since the previous experiments all indicate that Sox2 might interact with Groucho, the failure of this co-immunoprecipitation might be because the interaction is not sufficiently stable. Previous studies have applied cross-linkers to solve this difficulty, such as Formaldehyde and DSP (Dithiobis [succinimidypropionate]) (Fairbanks, 1976; Fujita and Wade, 2004; Valerio Orlando, 1997). The study of the Oct4 and Nanog interaction is a good example for the use of DSP to overcome such an unstable interaction reference. Since both Oct4 and Nanog are partners of Sox2. I used DSP as a cross-linker and repeated the co-immunoprecipitation experiment. After treating with DSP, the cells lysates were used for co-immunoprecipitation. The expression of Myc-Groucho3 and Myc-Groucho5 could be detected by anti-Myc antibody (Figure 3.3.4 B lane 1, 2). Myc-beads precipitated Myc-Groucho3 and Myc-Groucho5 (Figure 3.3.4 B lane 3, 4). The endogenous Sox2 expression was detected by anti-Sox2 antibody in whole cell extract (Figure 3.3.4 B lane 5, 6, 7). After co-immunoprecipitation, endogenous Sox2 could be precipitated only when Myc-Groucho3 and Myc-Groucho5 was expressed (Figure 3.3.4 B lane 8, 9, 10). These results indicate that Grouchos interact weakly with Sox2.

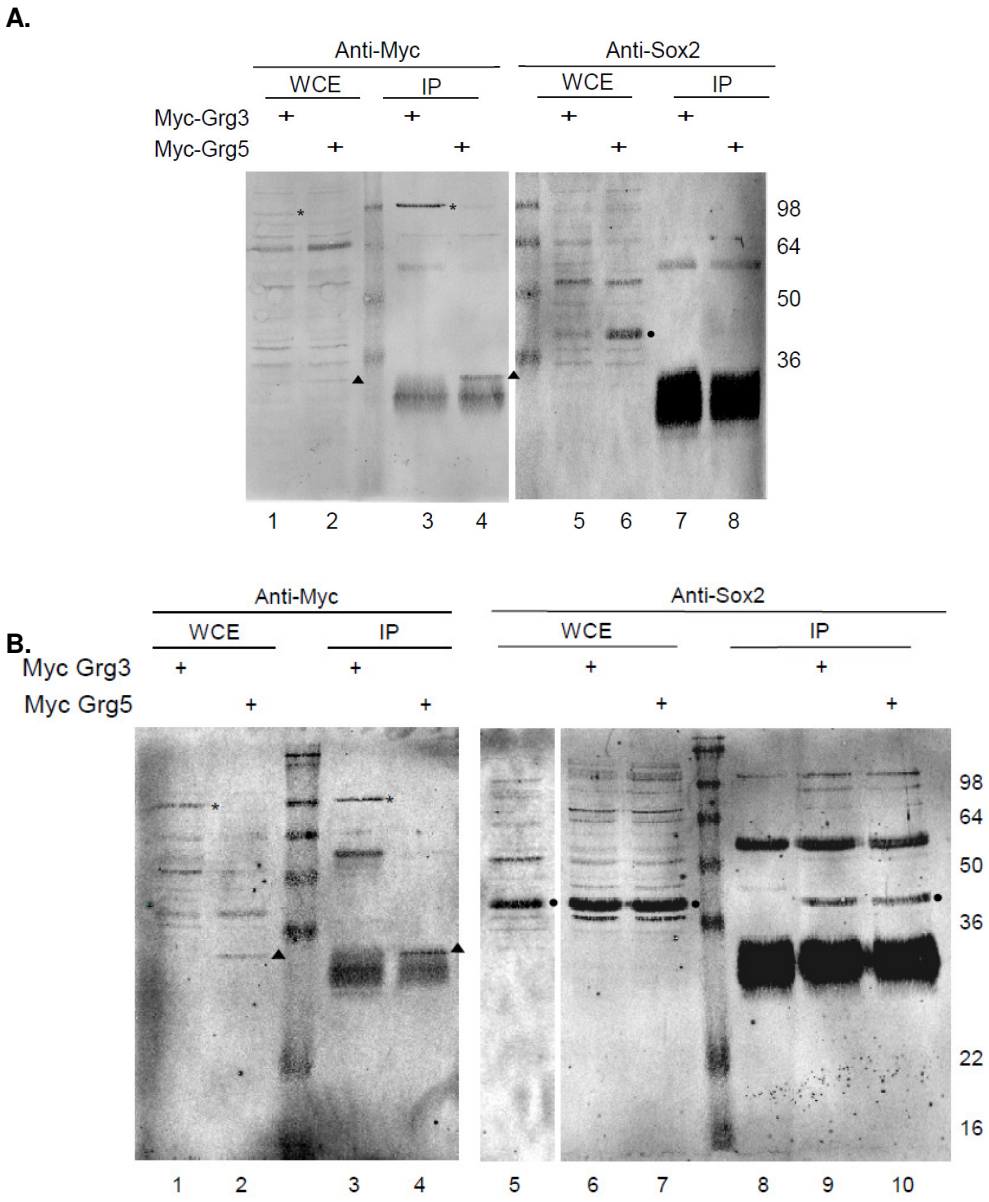


Figure 3.3.4 Co-immunoprecipitation of endogenous Sox2 and Myc-Groucho in hNSC cells

hNSCs were transfected with Myc-Groucho3 or Myc-Groucho5. After 48 hours, the non-treated cell lysates (A) or the DSP treated cell lystaes (B) were immunoprecipitated using Myc-beads. Myc-Groucho3, Myc-Groucho5 and endogenous Sox2 were detected in whole cell extract using anti-Myc (A lane 1, 2; B lane 1, 2) and anti-Sox2 antibodies (A lane 5, 6; B lane 5, 6, 7). Myc-Groucho3 (*) and Myc-groucho5 (▲) were immunoprecipitated and detected clearly by Myc-antibody. Endogenous Sox2 was detected in both Myc-Groucho3 and Myc-Groucho5 immunoprecipitated complex (B lane 9, 10) whereas the Myc beads alone cannot precipitate endogenous Sox2 (B lane 5). (WCE: 5% of whole cell extract; IP: Immunoprecipitation; Grg3: Groucho3; Grg5: Groucho5.)

3.3.4 Summary of co-immunoprecipitation experiments

In this section, the interaction between Sox2 and Groucho was demonstrated. However, the interaction revealed by the co-immunoprecipitation assay could be rather indirect. It is possible that the protein interaction demonstrated by Co-IP was not direct but through one or more intermediate proteins. To clarify this point further investigation was needed. The co-IP results show that Sox2 exists with Groucho3 and/or Groucho5, but the interaction with Groucho could be direct or indirect.

3.4 Groucho interacts directly with Sox2 in a yeast-two-hybrid assay

To test if the interaction between Sox2 and Groucho is direct, a modified yeast-two-hybrid was used. The yeast two-hybrid system has been used to investigate protein-protein interaction since 1989 (Field and Song, 1989). The advantage of using the yeast-two-hybrid system to investigate mouse protein interactions is that no potential intermediate mouse factors exist in the cellular environment. Therefore, indirect interactions are minimized. In other words, the yeast two-hybrid system could be used to test the direct interaction of proteins.

In this chapter, a modified yeast two-hybrid system was used. The GAL4 DNA-binding domain (BD) and activation domain (AD) were fused with Groucho and Sox2 respectively. The BD vector (pGBK-T7) and AD vector (pGAD-T7) carry leucine and tryptophan selection markers respectively (Figure 3.4.1). The AH190 modified yeast strain, which lacks the ability to make leucine, tryptophan, adenine and histidine, was used as the host cell in this system. The BD and AD carrying yeasts were selected on plates lacking both leucine and tryptophan (-2 SD plates). If a Groucho and Sox2 interaction took place in the yeast cells, the BD and AD of GAL4 would be recruited to the GAL4 binding promoter region which could then trigger downstream adenine and histidine synthase expression (Figure 3.4.1). Only when the interaction occurred could AH190 cells survive without an exogenous supply of adenine and histidine. This would therefore allow growth on plates which lack leucine, tryptophan, adenine and histidine (-4 SD plate).

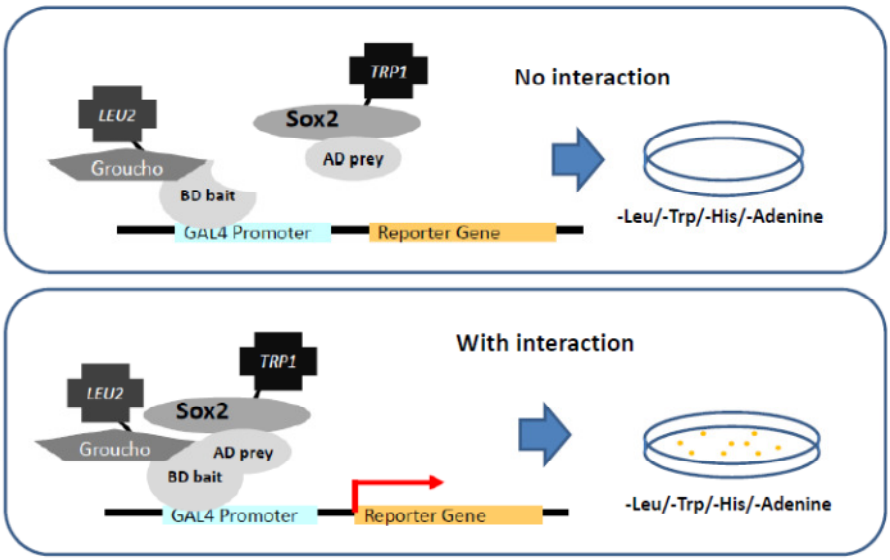


Figure 3.4.1 Schematic representation of the yeast two-hybrid system

Yeast carrying the BD and AD fusion protein express *LEU2* and *TRP1* that allows the yeast to grow on leucine and tryptophan "drop out" plates (-2 SD). To test the interaction of Groucho and Sox2, yeasts were spread on plates lacking leucine, tryptophan, histidine and adenine. The BD fusion protein will bind to the GAL4 promoter and the interaction between fusion proteins recruits the AD domain bring it close to the BD domain. Therefore, the interaction between the BD and the AD domain triggers down-stream reporter gene (adenine and histidine synthase) expression which allows the formation of yeast colonies.

To make the Sox2 AD construct, the full-length coding sequence of Sox2 with *EcoRI* and *BamHI* restriction enzyme sites on both ends was amplified using PCR using the pcDNA-Sox2 plasmid as template (appendix 1). In order to test the interaction of Sox2 with both long and short forms of Groucho, Groucho1 was used to represent the long form and Groucho5 was used as the short form. The Sox2 AD plasmid was then introduced into the host yeast containing either empty BD, Groucho1 BD and Groucho5 BD plasmid (plasmids constructed by Caroline Hirst). The expression of exogenous fusion proteins was detected using western blotting with appropriate antibodies (Figure 3.4.2). Anti-Myc antibody detected Myc tagged Groucho1 and

Groucho5 BD fusion protein (Figure 3.4.2.A lane 1, 2) whereas anti-HA antibody detected HA tagged Sox2 AD fusion protein (Figure 3.4.2.B lane 3, 4, 5). When grown on -4 SD plates, the BD vector carrying strain was used as a negative control and the Sox3 AD/ Groucho5 BD expressing strain was used as a positive control (Sox3 was shown to interact with Groucho5 in this assay previously; PhD thesis, Caroline Hirst, 2009) (Figure 3.4.3).

Unlike Sox2, yeast expressing Sox3 AD/ Groucho1 BD failed to grow on -4 plate in previous studies in our lab (PhD thesis, Caroline Hirst, 2009). This suggests that Sox2 might interact with long form Groucho more strongly than Sox3. The growth of Sox2 AD/ Grouchos BD yeast on -4 plates indicates that the interaction between Sox2 and Grouchos is likely to be direct.

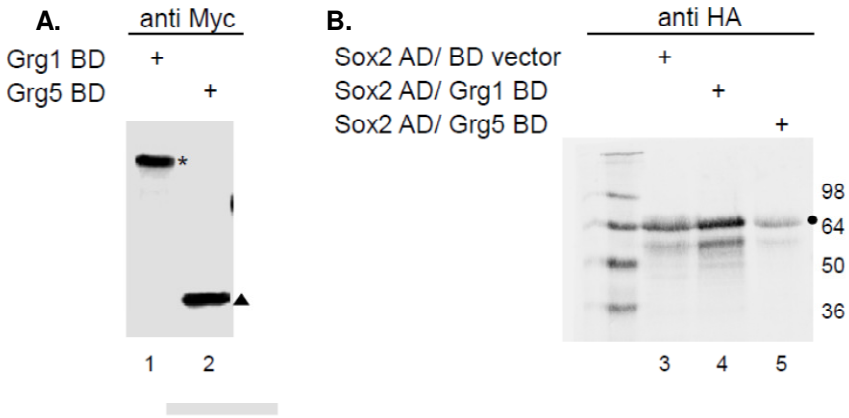


Figure 3.4.2 Sox2 AD expression

Yeast cells were analysed by SDS-PAGE and western bloting. (A) Anti-Myc antibody was used to detect Groucho1 (*) and Groucho5 (▲) BD fusion protein expression (Experiment performed by Caroline Hirst). (B) Sox2 AD fusion vector was transformed into AH190 yeasts carrying Groucho BD fusion vector. The expression of the Sox2 AD fusion protein was detected using anti-HA antibody (●).

Figure 3.4.3 Yeast two hybrid assay shows Sox2 interaction with Groucho

Transformed yeasts were grown on -4 SD plates. The colonies of the negative control BD vector and the positive control Sox3 AD/ Groucho5 BD. A single colony was picked and resuspended in 10ul (lst column of BD vector carrying yeasts) or 100ul (2nd column of BD vector carrying yeast and all other yeasts) of sterile distilled water. 10ul of the mixture was plated on -4 SD plates.

3.5 Groucho represses Sox2 transcriptional activity

Having demonstrated that Groucho and Sox2 interact, I then set out to determine if the interaction affected the function of Sox2 as a transcription factor. To test the transcriptional regulator activity of Sox2, a luciferase reporter assay was used. Firefly luciferase reporter plasmids contained three copies of Sox binding sites (3xSX) was used as the reporter of Sox2 transcriptional activity (Figure 3.5.1).

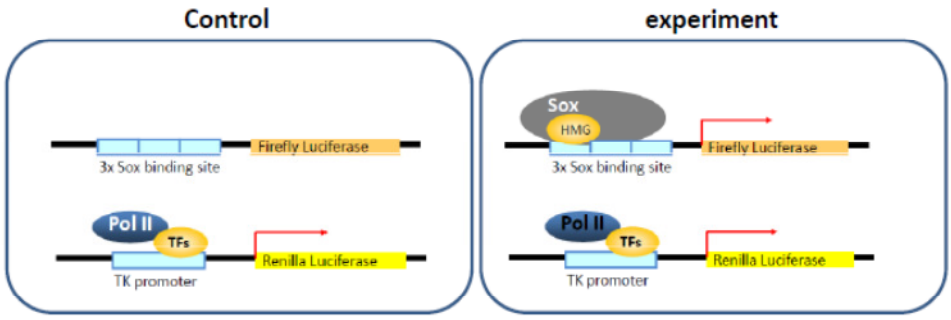


Figure 3.5.1 Schematic representation of luciferase assay

Cells transfected with 3xSX firefly and TK *Renilla* luciferase vectors which only express Renilla luciferase that continuously activated TK promoter in mammal cells act as controls. When co-expressed with Sox2 or Sox2 with Groucho, the firefly luciferase expression change indicates the transcriptional function of Sox2 on the 3xSX promoter.

The firefly luciferase reporter plasmid was constructed by Dr. Feist in pTATAluc vector plasmid (Kuhlbrodt et al., 1998). *Renilla* luciferase (with continuous activation from a tyrosine kinase (TK) promoter) (Promega) was used as the internal control in order to normalize the transfection efficiency. When transcription factors were co-transfected with both the *Renilla* and firefly luciferase reporters into cells, the relative luciferase activity could be calculated by the normalization equation:

$$\text{Relative luciferase activity} = \frac{(\text{Firefly}^{sample}/\textit{Renilla}^{sample})}{(\text{Firefly}^{control}/\textit{Renilla}^{control})}$$

In the equation, the control indicates the reading from cells transfected with *Renilla* and firefly luciferase reporters but no transcription factors.

Cells were transfected using liposome reagent and the luciferase assays were carried out 24 hours after transfection. Three wells were transfected with the same liposome reagent and DNA mixture as repeats in every individual test. Every test was repeated at least two times.

In order to test the optimal plasmid DNA quantity of pcDNA3 Sox2, different doses of Sox2 plasmid DNA were introduced into cells and analysed using the luciferase assay. Here, three cell lines, COS7, P19 and mouse embryonic stem cells (mES), were used for the test (Figure 3.5.2). Sox2 was found to only efficiently affect the 3xSX reporter in mES and P19 cells but not in COS7 cells. This suggests that Sox2 function on the 3xSX promoter might need specific co-factors that only exist in embryonic stem cells (mES) or embryonic carcinoma cells (P19) but not in COS7 cells. pcDNA3 Sox2 (1 µg) transfection increased the luciferase reporter expression approximate 2 fold compared to controls in both mES and P19 cells. However, compared to mES cells, P19 cells are easier to grow and transfect. For this reason, P19 cells were used for luciferase assay in the following work unless otherwise indicated. According to the results in mES and P19 cells, the transcriptional activator function of Sox2 reached maximum when 1 µg of pcDNA3 Sox2 plasmid was transfected. Therefore, from here onward, the amount of Sox2 in all luciferase assays was 1 µg unless stated otherwise.

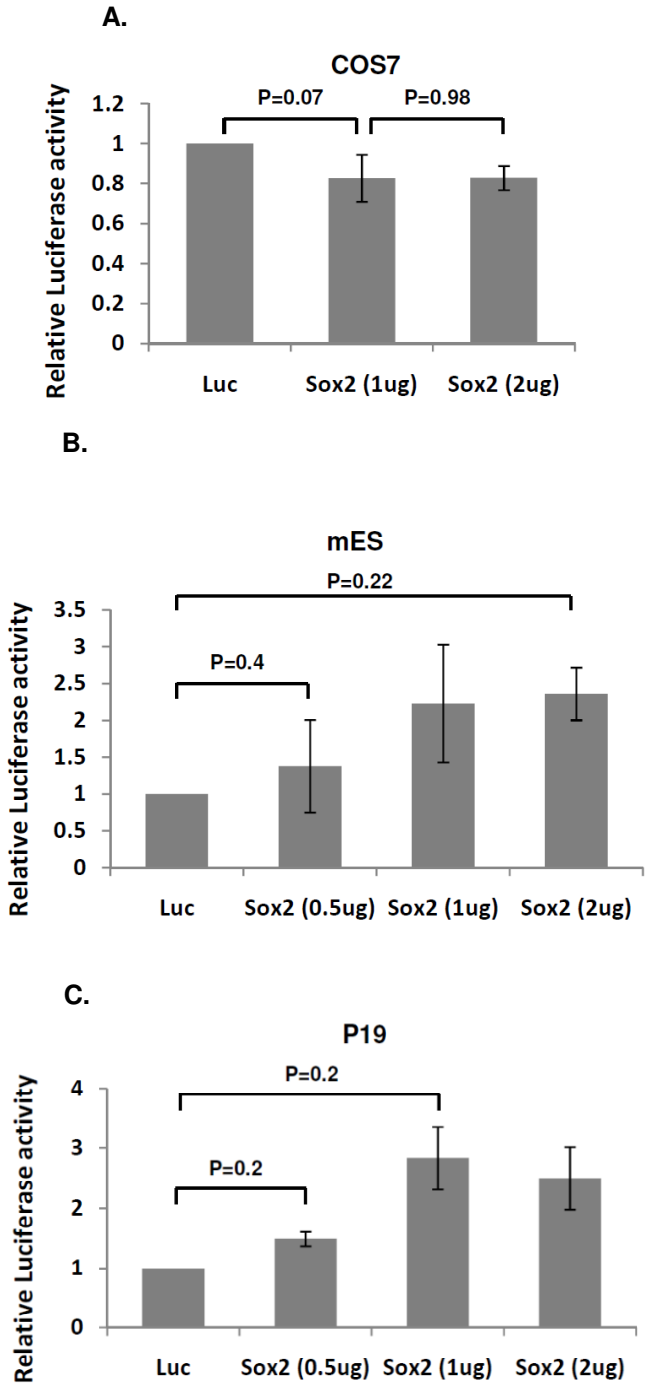


Figure 3.5.2 Sox2 acts as a transcriptional activator of a 3xSX luciferase reporter

Luciferase assays were carried out in COS7, mES, and P19 cells 24 hours after transfection. (A) In COS7 cells, Sox2 failed to activate luciferase reporter expression. In mES (B) and P19 (C) cells, Sox2 functioned as a transcriptional activator increasing the luciferase reporter expression. A paired t test was used to access significance. $p< 0.05$ is defined as significant difference.

In P19 cells, the 3xSX luciferase reporter was activated 2 fold when co-transfected with Sox2. However, when co-transfected with Groucho3, the Sox2-induced increase was ablated and the relative luciferase activity was decreased to the level of the control (Figure 3.5.3). This suggests that long form Grouchos, such as Groucho3, function as repressors of Sox2 transcriptional activation activity. The same effect was also seen when Sox2 was co-transfected with Groucho5. Groucho5 is a short form of Groucho, which has no C-terminal protein-protein interaction domain and can function as a dominant negative form of Groucho (Brantjes et al., 2001). However, the N-terminal Q domain in Groucho5 has been demonstrated as a protein-protein interaction domain that interacts with TCF (Pickles et al., 2002). Also, Groucho5 has been shown to function as a co-repressor of the androgen receptor (Yu et al., 2001). The luciferase assay results in this session, together with the data shown in chapter 3, indicate that Groucho5, the short form Groucho, could function as a co-repressor of Sox2 transcriptional activation activity.

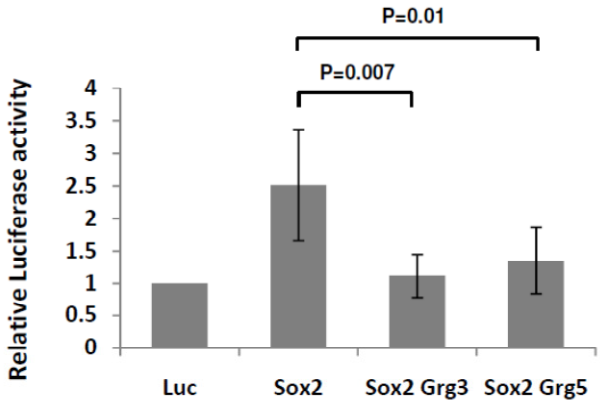


Figure 3.5.3 Groucho co-expression represses Sox2 transcriptional activity

3xSX promoter driven luciferase expression was increased when Sox2 was overexpressed in P19 cells. When Sox2 and Groucho3 or Groucho5 were co-expressed, the transcriptional activation by of Sox2 was repressed. A paired t test. $p< 0.05$ is defined as significant difference.

In order to test the affect of Groucho on endogenous Sox2 target promoters, luciferase reporter vectors containing the Nanog, Rex and GFAP promoter were constructed (Appendix 2-4). Previous studies had shown that Sox2 activates Nanog and Rex but represses GFAP expression (Cavallaro et al., 2008; Rodda et al., 2005; Shi et al., 2006).

To test if the Grouchos could repress the luciferase reporter expression by itself, Groucho3 or Groucho5 and the various promoter-containing luciferase reporters were introduced into cells without Sox2 (Figure 3.5.4). The results show that neither Groucho3 nor Groucho5 had any effect on 3xSX, Rex, Nanog and GFAP luciferase reporters at the level of transfection used.

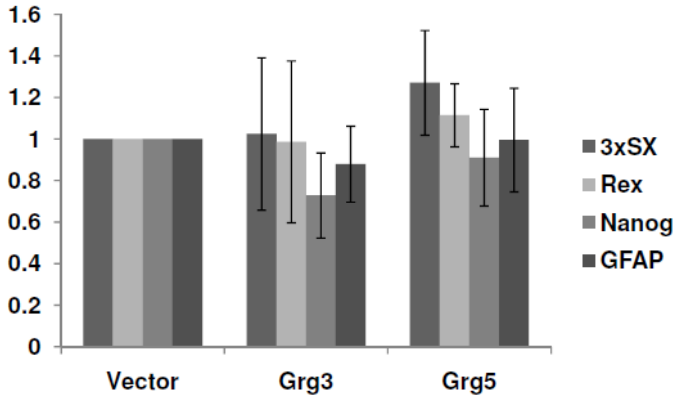


Figure 3.5.4 Groucho does not affect Sox2 target gene reporter expression

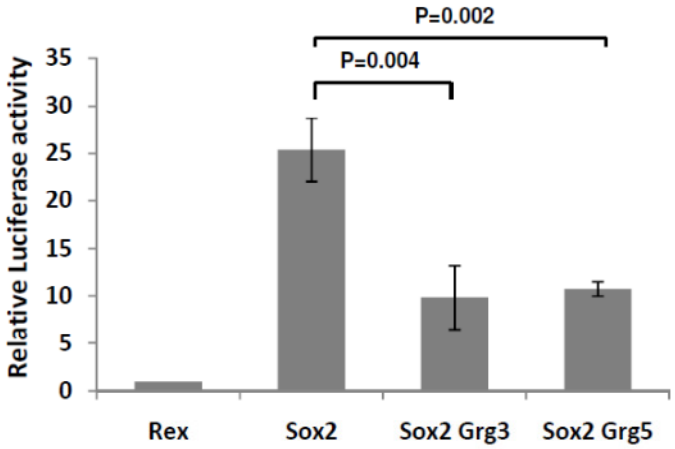
Luciferase assays were carried out on different Sox2 target gene promoter luciferase reporters 24 hours after transcription. In the presence of overexpressed Groucho3 or Groucho5 the luciferase reporter expression was not significantly different to the vector control (lysate from *Renilla* and firefly luciferase vectors transfected cells). $p < 0.05$ is defined as significant difference.

The transcriptional activity of Sox2 was first tested on the Rex luciferase reporter (Figure 3.5.5 A). Sox2 dramatically activated the Rex promoter-driven luciferase expression by over 20 fold when compared with the control. When co-expressed with Groucho3, the increase caused by Sox2 was suppressed by about 50% to only 10 times the control. Groucho5, the short form Groucho, suppressed Sox2 transcriptional activation activity to the same level as Groucho5. On the 3xSX luciferase reporter, Groucho3 or Groucho5 co-expression suppressed its transcriptional activation by Sox2 to the control level (Figure 3.5.3). However, transcription of the Rex reporter was still activated to approximately 10 times control level even when Groucho3 or Groucho5 were co-expressed with Sox2. These results suggest that the repression of Sox2 activation by Grouchos is restricted by context. When the Sox2 is able to induce a massive activation of a downstream gene, like Rex, the suppression

function of Groucho could only decrease the Sox2 transcriptional activation activity to a certain level but not eliminate it.

As for Nanog, Sox2 also acted as a transcriptional activator but not as effectively as on the Rex promoter (Figure 3.5.5 B). Groucho3 co-expression slightly suppressed Sox2 transcriptional activation activity whereas Groucho5 co-expression almost diminutal the Sox2 activation ability. These results show that Grouchos can repress Sox2 transcriptional activation, but the degree to which they achive this depends on the long-form or short-form of Groucho and the contact of target.

A.



B.

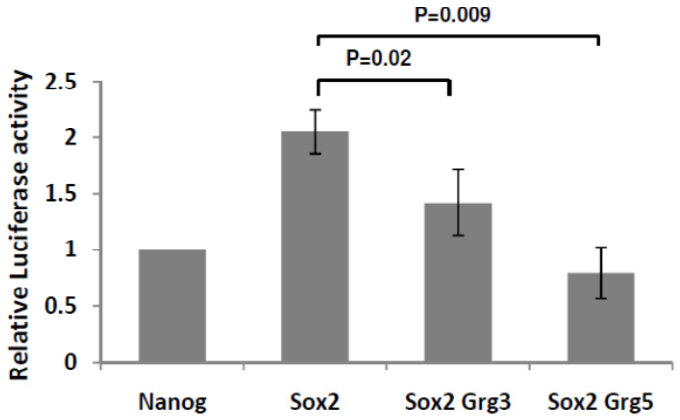


Figure 3.5.5 Grouchos repress Sox2 transcriptional activator function on target gene promoter reporter

Sox2 increased luciferase expression on both Rex (A) and Nanog (B) target gene promoters. However, when Sox2 was co-expressed with Groucho3 or Groucho5, the Sox2 activator function was repressed dramatically. $p < 0.05$ is defined as significant difference.

Cavallaro et al. (2008) demonstrated that GFAP expression was down regulated by Sox2 and also identified a Sox2 binding region on the GFAP promoter (Cavallaro et al., 2008). Here, the Sox2 binding region was cloned into a luciferase reporter and the transcriptional activity was tested. The result of my experiments of Sox2 function on their GFAP luciferase reporter were in

agreement with Cavallaro's data that Sox2 repressed the GFAP luciferase reporter expression to 50% of the control (Figure 3.5.6). The co-expression of Grouchos with Sox2 could further suppress the GFAP reporter by approximately 70%. These data suggest that Sox2 functions as a repressor and that the Grouchos can act as co-repressors to enhance repression of GFAP. In other words, Grouchos interfere with Sox2 transcriptional activation activity and also function as co-repressors of Sox2 transcriptional repression activity.

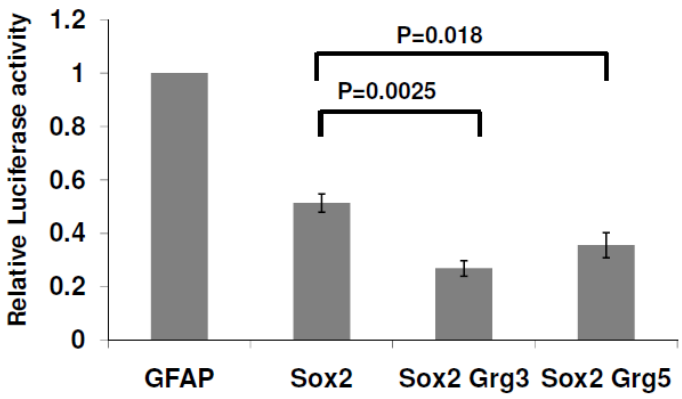


Figure 3.5.6 Down regulation of a GFAP-Luciferase reporter was enhanced by Groucho co-expression.

The expression of Sox2 repressed the luciferase activity driven by the GFAP promoter. The repression of GFAP reporter was even greater when Sox2 was co-expressed with Groucho3 or Groucho5. $p < 0.05$ is defined as significant difference.

3.6 Discussion

Although the SoxB1 family members are generally regarded as transcriptional activators, there is increasing evidence that SoxB1 factors act as bifunctional transcription factors. (Kan et al., 2004; Zhang et al., 2003a; Cavallaro et al., 2008; Muotri et al., 2005; Navarro et al., 2008). Unpublished data in our lab has demonstrated that the interaction of Sox3 and Groucho also represses the transcriptional activity of Sox3 (PhD thesis, Zulfiqar Laghari, 2010). Sox3 could translocate Groucho5 into the nucleus, in addition, the co-expression of Groucho5 interfered with Sox3 transcriptional activation activity on the 3xSX luciferase reporter. However, the mechanism of the repressor function of the SoxB1 family remained unknown.

In this chapter, the interaction between Sox2 and Groucho was tested firstly by observing the change of the subcellular localization. When overexpressing Sox2 with Groucho5, Groucho5 tended to accumulate in the nucleus instead of being distributed out through the cell. Nevertheless not all of the subcellular localization was changed by the co-expression (Figure 3.2.3 C; Figure 3.2.4 B). Co-IP was carried out to test the physical interaction between Sox2 and Groucho. COS7 cells were used as a neutral enviroment to test the interaction without other intermediate factors that would be presented in embryonic stem cells, for example. The results show that Sox2 could interact with both exogenous and endogenous Grouchos (Figure 3.3.2; Figure 3.3.3). However, the interaction of endogenous Sox2 with exogenous Groucho in human neural stem cells could be detected only after crosslinking (Figure 3.3.4). This indicates that the interaction of endogenous Sox2 with Groucho might be either very weak or only occurs very transiently. The yeast-two-hybrid system was

used to test the direct interaction of Sox2 and Groucho. In previous studies in our lab, transformation with sufficient Sox3 BD alone was enough to allow the host yeast cells to survive on -4 plates. In other words, the transcriptional activation of Sox3 could activate the expression of the reporter genes (PhD thesis, Caroline Hirst, 2009). This could cause a false positive result in yeast-two hybrid system. Therefore, the Sox2 AD but not Sox2 BD was used demonstrating the yeast-two hybrid assay in this chapter. The slow growth of the Sox2 AD/Groucho3 BD and Sox2 AD/ Groucho5 BD colonies implies a weak interaction of Sox2 with Groucho3 and Groucho5 (Figure 3.4.3).

Using a luciferase assay, it was shown that Grouchos suppress the Sox2 transcriptional activity both on artificial (3xSX) and endogenous (Nanog, Rex) Sox2 binding motifs (Figure 3.5.6). However, the ability of Sox2 to repress a GFAP luciferase reporter shows that Sox2 could also act as a transcriptional repressor. This result agrees with the data of Cavallaro et al., (2008). Co-expression of Groucho3 or Groucho5 with Sox2 only slightly enhanced repression of the GFAP reporter (Figure 3.5.7). This might be because the Sox2 repression complexes were already saturated with endogenous Groucho. Therefore, the effect of Groucho co-expression would not be obvious. These data indicate that Groucho functions as a co-repressor of Sox2 transcriptional repression activity.

Most transcription factors have been reported to interact with Groucho through the C-terminal WD domain (Brantjes et al., 2001). Since Groucho5, the short version of Groucho, has no C-terminal SP and WD domain, it had been defined as a de-repressor or dominant negative form of long form Groucho (Brantjes et al., 2001; Roose et al., 1998). However, some

transcription factors such as Runx2, Nolz1, and Pax5 interact with the N-terminus of Groucho proteins (Eberhard et al., 2000; Ji et al., 2009; Wang et al., 2004). This suggests that the short form Groucho is able to interact with transcription factors just like other members in the Groucho family. The data I present in this chapter indicate that Sox2 also interacts with both Groucho3 and Groucho5 (Figure 3.3.2). Therefore, both Groucho3 and Groucho5 could function as co-repressors with Sox2 (Figure 3.5.3)

In summary, all of the results in this chapter suggest an interaction between Sox2 and Groucho and this interaction helps Sox2 to function as a transcriptional repressor.

Chapter 4 Mapping the Sox2-Groucho interaction domain

4.1 Introduction

The mapping of protein-protein interaction domains is a critical step in protein function research. By mapping where in the protein the domain is located, a major protein-protein interaction domain could be revealed and a functional mutant could be generated. Comparison of the behaviour of the wild type protein and the "interaction domain mutant" (loss-of-function mutant) will shed a light on the *in vivo* role of the protein interaction. A previous study of Pax5 assigned the engrailed-like Groucho interaction domain to the region of amino acids 179-186 and the single point mutation of Y179E disrupted Groucho interaction (Figure 4.1.1) (Eberhard et al., 2000).

In this chapter, the mapping of the Sox2-Groucho interaction domain is demonstrated by various C-terminal deletions and point mutations.

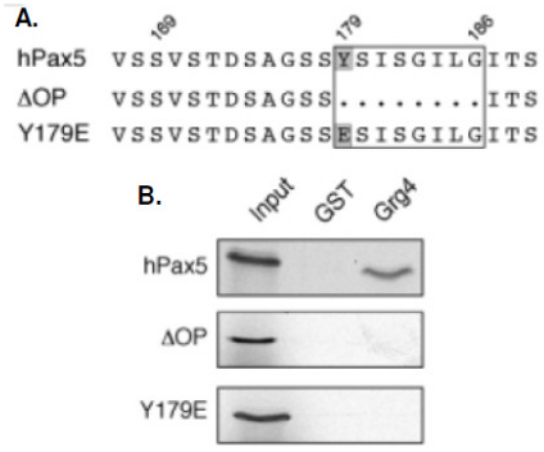


Figure 4.1.1 Mapping of the Pax5-Groucho interaction domain

(A) Schematic representing the wild type and mutant Pax5 sequence. (B) GST pull-down assay showing that ΔOP and Y179E lose the interaction with Groucho4 (Eberhard et al., 2000).

4.2 Sox2 C-terminal deletions

Many studies of the C-terminus of the Sox family of proteins have demonstrated that this is where the trans-activation domain or repression domain reside (Hosking et al., 1995; Pusch et al., 1998; Uchikawa et al., 1999). Studies in our lab during the time of my PhD (PhD thesis, Caroline Hirst, 2009, PhD thesis, Zulfiqar Laghari, 2010) mapped two regions of Sox3 which were essential for the Sox3-Groucho interaction (Figure 4.2.1).

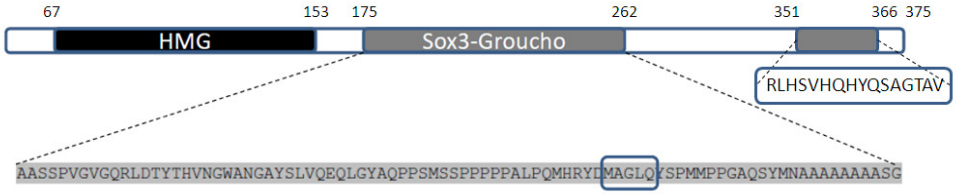


Figure 4.2.1 Schematic representation of potential Groucho interaction domains on mouse Sox3

The HMG domain of mouse Sox3 is shown as a black box (67-153) and the Sox3-Groucho interaction regions are shown as grey box (175-262 and 351-366). The Sox3-Groucho interaction region sequence from the yeast-two hybrid system (Caroline Hirst) is shown in the grey box beneath. The boxed sequences mark the predictive motifs which were sufficient in the subcellular localization change assay (Zulfiqar Laghari).

Since Sox2 and Sox3 are highly conserved in their amino acid sequence, it seemed likely that Sox2 also interacts with Groucho in a similar manner. In order to identify the possible Groucho interaction domain(s) of Sox2, six serial deletions of the Sox2 C-terminus were generated. The Sox2 C-terminal deletions were designed according to the alignment of Sox2 and Sox3 amino acid sequence of the broad Sox3-Groucho interaction region (Figure 4.2.2).

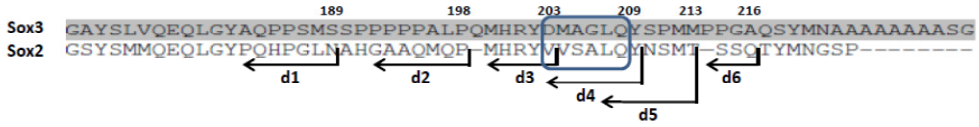


Figure 4.2.2 Schematic representation of the Sox2 C-terminal deletion design

The mouse Sox2 and Sox3 alignment is illustrated in the figure. The sequence of Sox3 is shadowed with grey. The boxed sequences mark the predictive motifs which were sufficient for Groucho interaction in the subcellular localization change assay in Sox3 (Zulfiqar Laghari). Sox2 C-terminal deletion designs are shown as d1-d6. The amino acid numbering of Sox2 is marked at the top.

Due to the high GC content of Sox2, the insert fragments of Sox2 C-terminal deletions were amplified by PCR in the presence of 10% DMSO. The Sox2 C-terminal deletions were cloned into pcDNA3 and pEGFPC1 vectors (Appendix 5-16). The pEGFPC1 Sox2 C-terminal deletions were introduced into COS7 cells and the expression of the GFP-tagged Sox2 deletions were detected by western blot (Figure 4.2.3). All of the Sox2 deletions retained the entire HMG DNA-binding domain where the NLS resident and localized in the nucleus (Figure 4.2.4).

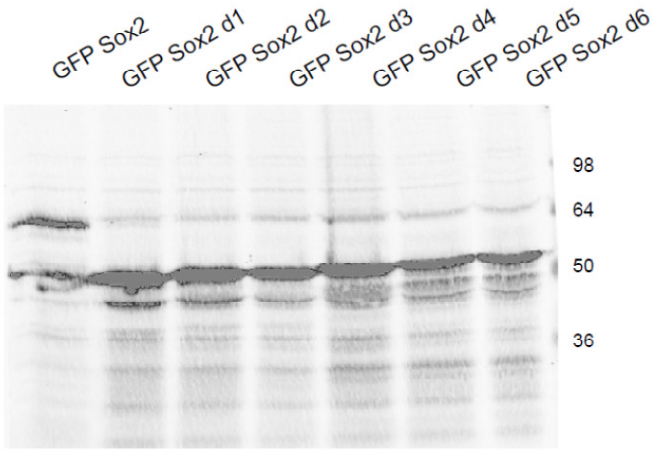


Figure 4.2.3 Expression of Sox2 C-terminal deletions

The GFP-tagged Sox2 C-terminal deletions were expressed in COS7 cells and detected by anti-GFP antibodies on a western blot membrane.

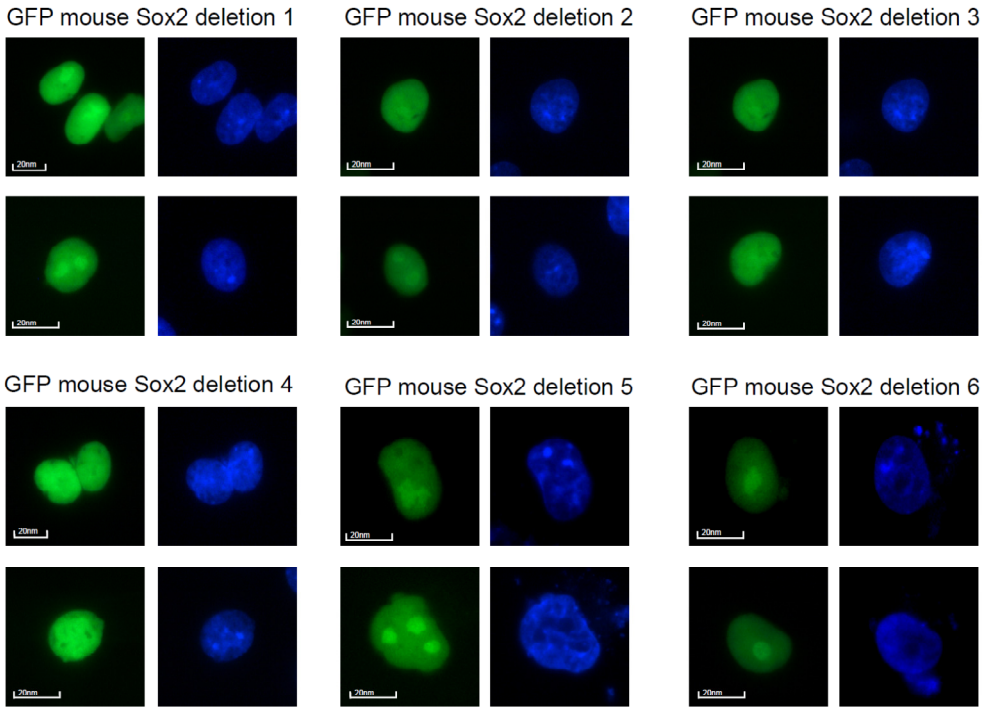


Figure 4.2.4 The subcellular localization of Sox2 C-terminal deletions

Sox2 C-terminal deletions were expressed in COS7 cells and the subcellular localization was observed after DAPI staining. The left panel of each set was GFP tagged Sox2 C-terminal deletions which show up green in colour. The right panel contains DAPI stained nucleus which show up blue in colour.

4.3 Sox2 deletions functional analyses

Firstly, analysis of subcellular localization was used to test the Sox2-Groucho interaction. COS7 cells were transfected with Myc tagged Groucho3 or Groucho5 alone or together with various Sox2 C-terminal deletions and the subcellular localization of Myc-Groucho3 or Groucho5 was detected by immunostaining with anti-Myc antibody. The Sox2 C-terminal deletions, d4-d6, changed the subcellular localization of Groucho3 (Figure 4.3.1) and Groucho5 (Figure 4.3.2) in a similar way to the wild type Sox2. Furthermore, the Sox2 C-terminal deletions, d4-d6, had almost the same level of translocation ability (50% in Groucho3, 80% in Groucho5). By contrast, the Sox2 C-terminal deletions, d1-d3, almost entirely lost the ability to change the subcellular localization of both Groucho3 and Groucho5. These data indicate that the amino acids between 203-209 might play an important role in the Sox2-Groucho interaction.

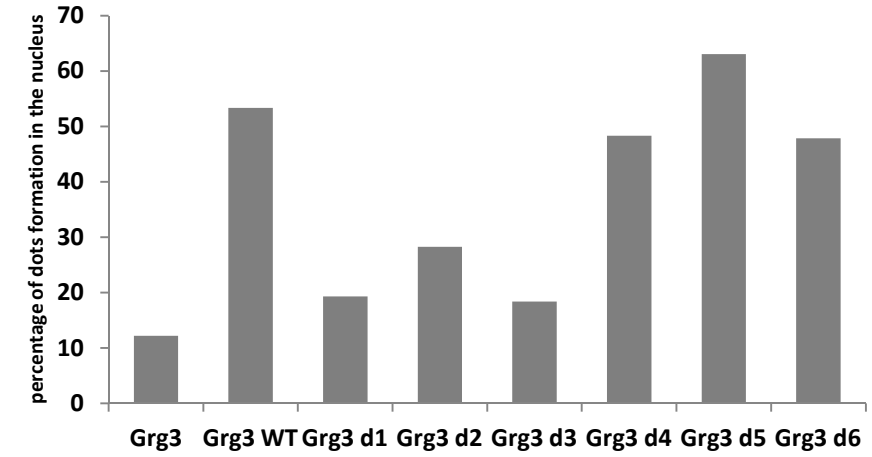


Figure 4.3.1 Subcellular localization of Groucho3 with Sox2 C-terminal deletions

COS7 cells were transfected with Myc-Groucho3 and various Sox2 C-terminal deletions. The subcellular localization of Myc-Groucho3 was detected by anti-Myc antibodies. Five fields across the coverslips (that of around 100 cells counted) were observed in every set of experiments. WT: Wild type Sox2; d1-6: Sox2 C-terminal deletion 1-6.

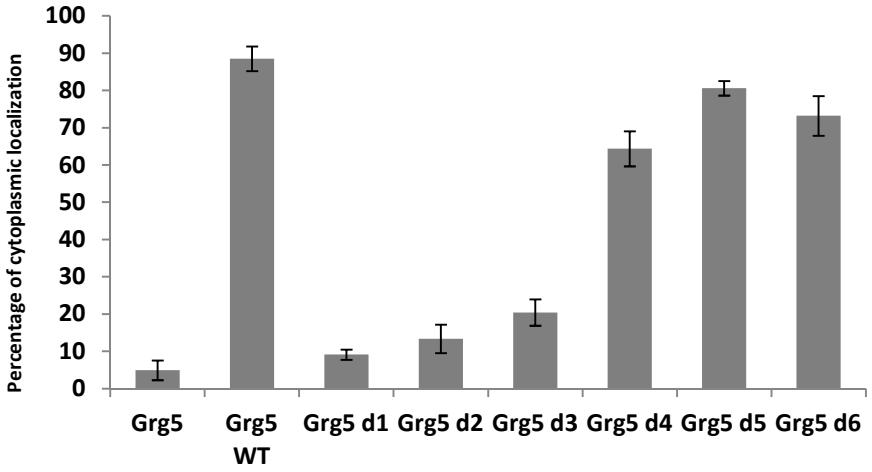


Figure 4.3.2 Subcellular localization of Groucho5 with Sox2 C-terminal deletions

COS7 cells were transfected with Myc-Groucho5 and various Sox2 C-terminal deletions. The subcellular localization of Myc-Groucho3 was detected by anti-Myc antibodies. Five fields across the coverslips (that of around 100 cells counted) were observed in every set of experiments. WT: Wild type Sox2; d1-6: Sox2 C-terminal deletion 1-6.

In order to test if the truncation of C-terminal Sox2 changed the transcriptional activity of Sox2, the 3xSX luciferase reporter assay was carried out (Figure 4.3.3). The transcriptional activity of Sox2 C-terminal deletions was generally weaker than that of wild type Sox2. However, the deletion d5 and d6 had slightly stronger transcriptional activity than the deletion d1-d4. Consistent with published data, these results imply that amino acids in the C-terminal region from 203-209 are important for transcriptional activation activity. This result indicates that the longer the truncation is the more serious the effect on the transcriptional activity function of Sox2.

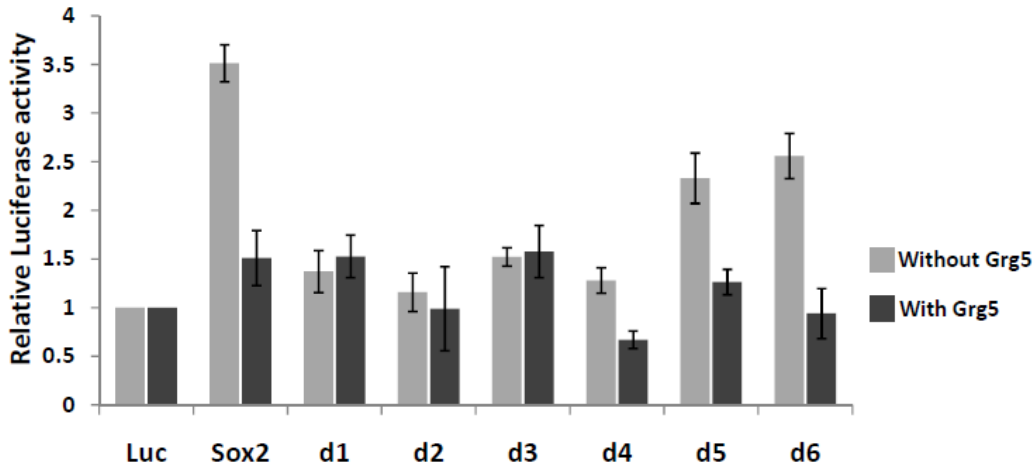


Figure 4.3.3 Groucho5 affects Sox2 C-terminal deletion d4-6 transcriptional activity

Sox2 C-terminal deletions have lower transcriptional activity than wild type Sox2 on 3xSX luciferase reporter. The transcriptional activity of activation Sox2 C-terminal deletion d4-d6 was repressed when co-expressed with Groucho5. Nevertheless, the representation of Groucho did not affect the transcriptional activity of Sox2 C-terminal deletion d1-d3.

If some of the C-terminal deletions remove the domain of Sox2 responsible for Groucho interaction this would predict that Groucho might no longer have any effect on Sox2 transcriptional activation activity. To analyse the function of Groucho5 on the Sox2 C-terminal deletions, Groucho5 was introduced into the luciferase assay system. As for wild type Sox2, the transcriptional activity of Sox2 C-terminal deletions d4-d6 was suppressed by Groucho5 to about 50% of its original value. By contrast, Groucho5 failed to affect the transcriptional activity of Sox2 C-terminal deletion d1-d3. However, the Sox2 deletion d1-d4 had lost much of their transcriptional activation activity which did not leave much opportunity for Groucho5 to exert suppressive function. But even so, Groucho5 still suppressed Sox2 deletion d4 transcriptional activation activity by 40%. The different response to Groucho5 is consistent with the mapping of the Groucho interaction domain to amino acids 203-209.

4.4 Targetted mutation of the Groucho-interaction domain of Sox2

Although the C-terminal deletions of Sox2 caused a loss of interaction with Groucho these truncations deleted almost one-third of the full length of Sox2. These deletions might therefore cause large conformation change and lead to functional difference. In order to minimize the difference between wild type and mutant and to define the function of the Groucho interaction region, point mutation was carried out to generate the loss-of-interaction mutants of Sox2. As mentioned in the previous section, the Sox2 C-terminal deletion d4 acted similarly to the wild type Sox2 in terms of Groucho5 interaction while the Sox2 C-terminal deletion d3 which has only 6 amino acids fewer than d4 acted differently to the wild type. Examination of this 6 amino acids sequence showed, there is a conserved motif, LQY, within the whole SoxB1 family (Figure 4.4.1).

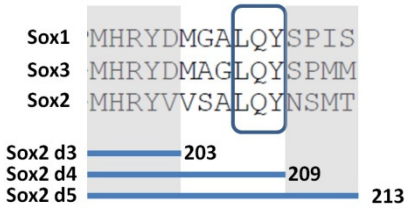


Figure 4.4.1 Schematic representation of LQY motif

The partial peptide sequence of the SoxB1 family is shown in the diagram. The conserved LQY motif is boxed. The C-terminal ends of various Sox2 C-terminal deletions d3-d5 are illustrated in the lower panel.

This implies the importance of the LQY motif in the SoxB1 family. On the other hand, Groucho interaction normally involves an hydrophobic aromatic amino acid (W, F, Y) and aliphatic amino acid (I, L, V) which are also present in the LQY motif. Therefore, site directed mutagenesis was carried out to replace the LQY motif with three neutral amino acids AAA (referred to as

$Sox2^{LQY/AAA}$). pcDNA3 Sox2 was used as template in the site directed mutagenesis and 5% DMSO was added to overcome the high GC content in Sox2. The resulting mutant was validated by sequencing.

In order to test the expression of $Sox2^{LQY/AAA}$, pcDNA3 $Sox2^{LQY/AAA}$ was transfected into COS7 cells followed by immunostaining using Sox2 antibody. The subcellular localization of $Sox2^{LQY/AAA}$ in COS7 cells was mainly in the nucleus similar to the wild type Sox2 (Figure 4.4.2).

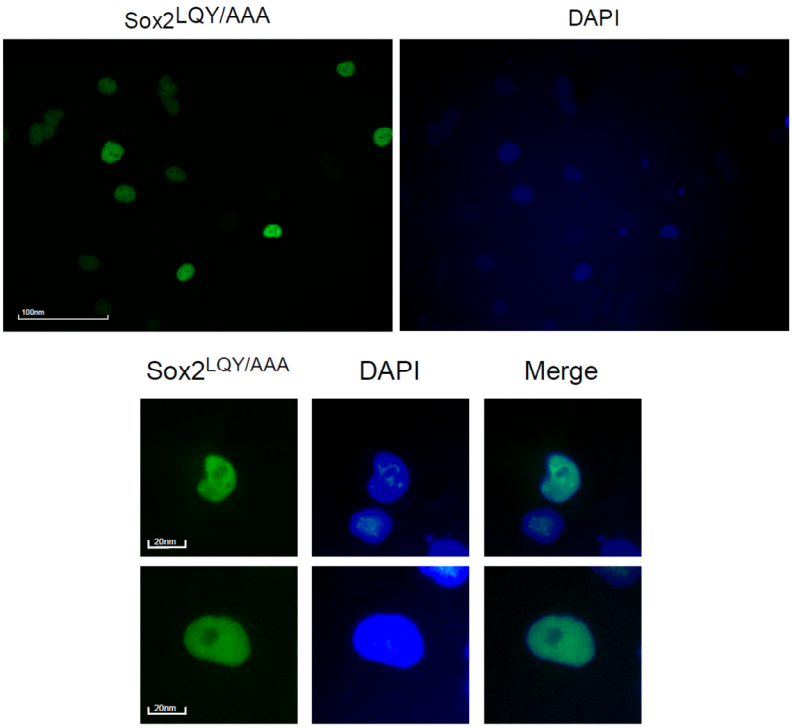


Figure 4.4.2 $Sox2^{LQY/AAA}$ subcellular localization

$Sox2^{LQY/AAA}$ was overexpressed in the COS7 cells. After 24 hours, the subcellular localization of $Sox2^{LQY/AAA}$ was detected using anti-Sox2 antibodies (left panel). The nucleus was stained using a DAPI mounting medium (middle panel). The right panel shows the merged image of green and blue.

In order to test if this mutation affected the ability of Sox2 to affect Groucho5 subcellular localization, Myc-Groucho5 was transfected in COS7 cells with either wild type Sox2 or $Sox2^{LQY/AAA}$ and immunostaining was

carried out using anti-Myc antibody. The results showed that Sox2$^{LQY/AAA}$ had a significantly reduced ability to change Groucho5 subcellular localization when compared to wild type Sox2 ($p<0.001$; Figure 4.4.3 A). Quantification demonstrated that Sox2$^{LQY/AAA}$ exhibited a 50% decrease in ability to alter the subcellular localization of Groucho5 with Sox2 wild type (Figure 4.4.3 B).

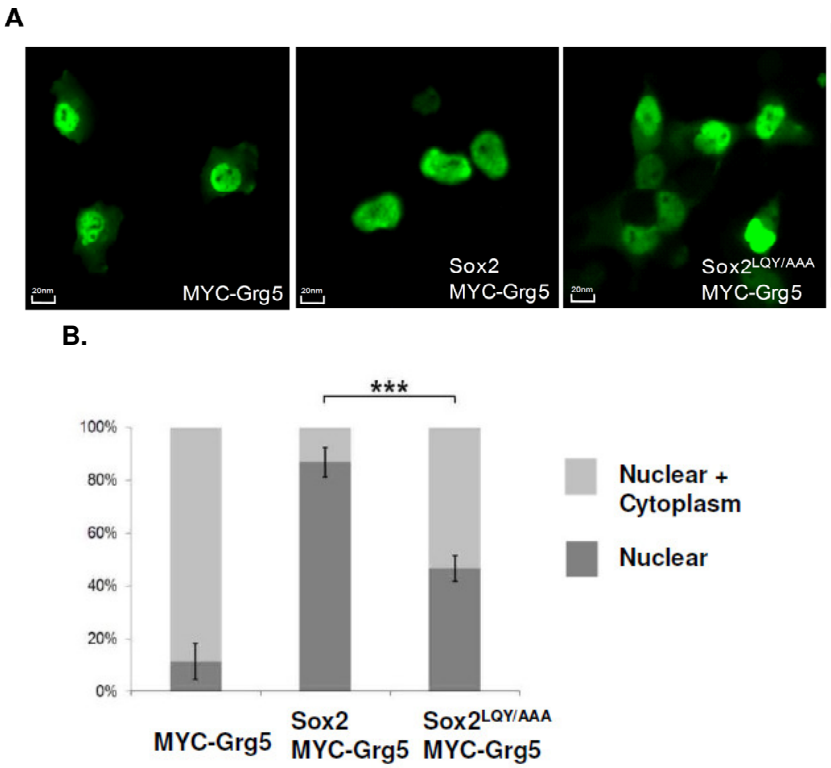


Figure 4.4.3 Sox2$^{LQY/AAA}$ mutant shows a reduced ability to translocate Groucho5

COS7 cells were transfected with Myc-Groucho5, Sox2 or Sox2$^{LQY/AAA}$ by electroporation and seeded on poly-D-lysine coated coverslips. After 24 hours, the cells were immunostained using anti-Myc antibodies. (A) Groucho5 was translocated into the nucleus by wild type Sox2 but not Sox2$^{LQY/AAA}$. (B) Quantification of Groucho5 nuclear localization. Five fields across the coverslips (that of around 100 cells counted) were observed in each experiment. (***: $p< 0.001$)

To further investigate if Sox2$^{LQY/AAA}$ could interact with Groucho5, Co-IP was carried out (Figure 4.4.4). Wild type Sox2 or Sox2$^{LQY/AAA}$ were introduced into COS7 cells either alone or with Myc-Groucho5. After Co-IP with Myc-

beads samples were analysed by western blot. The results showed that Sox2$^{LQY/AAA}$ failed to be pulled-down by Myc-Groucho5 whereas wild type Sox2 was co-precipitated. These data indicate that the Sox2$^{LQY/AAA}$ mutant has lost the ability to interact with Groucho5.

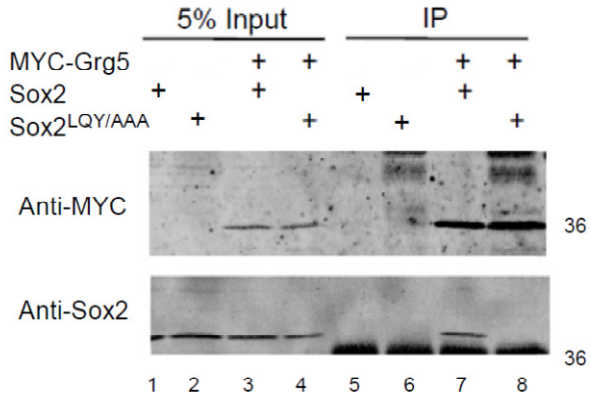


Figure 4.4.4 Sox2$^{LQY/AAA}$ lacks the ability to interact with Groucho5

COS7 cells were transfected with Myc-Groucho5, Sox2 or Sox2$^{LQY/AAA}$ by electroporation. Co-IP was carried out after 48 hours of culture using Myc-beads. The left panel shows the expression of exogenous proteins as 5% input. The right panel shows the proteins precipitated by MYC beads. Anti-MYC antibody detected the MYC-Groucho5 in the 5% input (lane 3, 4 upper panel) and the MYC beads precipitated samples (lane 7, 8 upper panel). Anti-Sox2 antibodies detected Sox2 expression in the 5% input (lane 1-4 lower panel). The wild type Sox2 was detected in the MYC-bead precipitated samples (lane 7 lower panel) but not the Sox2$^{LQY/AAA}$ mutant (lane 8 lower panel).

To test if the loss of Grouch5 interaction affects the transcriptional activity of Sox2, luciferase assays were carried out. Both Sox2 and Sox2$^{LQY/AAA}$ acted as transcriptional activators of the 3xSX luciferase reporter. In fact Sox2$^{LQY/AAA}$ acted similar with wild type Sox2 and activated the reporter expression more than 2 times (Figure 4.4.5 A; p=0.2071). As for the Rex luciferase reporter, Sox2$^{LQY/AAA}$ also activated Rex luciferase reporter to the level similar to wild type Sox2 (Figure 4.4.5 B; p=0.8821). These data indicate that Sox2$^{LQY/AAA}$ had no difference to wild type Sox2 in transcriptional activation function.

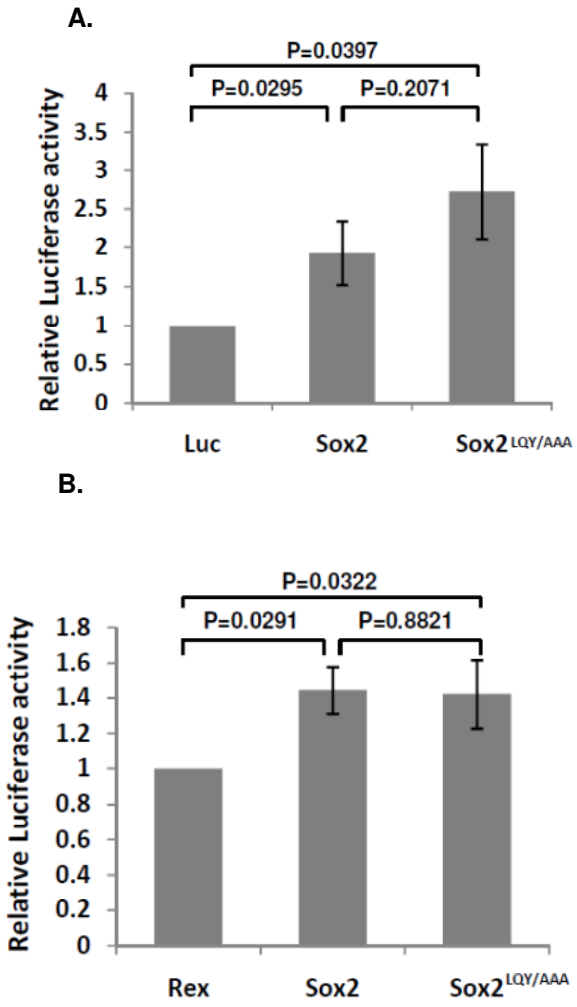


Figure 4.4.5 Sox2$^{LQY/AAA}$ retained transcriptional activation activity

P19 cells were transfected with 3xSX (A) or Rex (B) luciferase reporter. The transcriptional activity of Sox2 and Sox2$^{LQY/AAA}$ was tested by co-transfecting with the reporters. Both Sox2 and Sox2$^{LQY/AAA}$ activated reporters to a similar expression level.

As for the transcriptional repression function, Sox2$^{LQY/AAA}$ was tested in the GFAP luciferase reporter system. Interestingly, Sox2$^{LQY/AAA}$ failed to repress the GFAP luciferase reporter whereas the wild type Sox2 repressed the reporter to the level of 50%. On the other hand, Grouchos only further suppressed GFAP luciferase reporter with wild type Sox2 (p=0.0016) but not with Sox2$^{LQY/AAA}$ (p=0.0514) (Figure 4.4.6). These results imply that the

$Sox2^{LQY/AAA}$ mutant only loses its transcriptional repressor function but retains the activator function of Sox2. Therefore, the $Sox2^{LQY/AAA}$ mutant could be used as a loss-of-repressor-function form of Sox2 in further investigations.

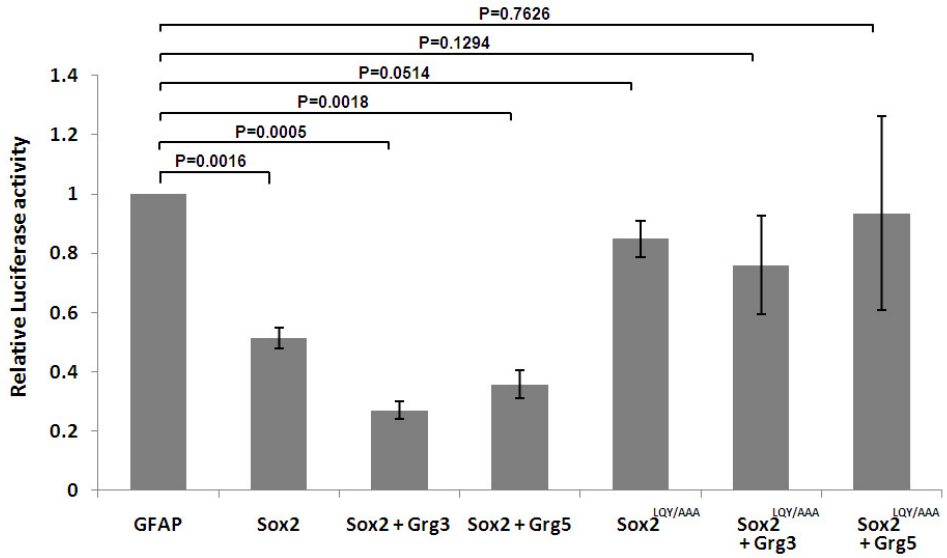


Figure 4.4.6 $Sox2^{LQY/AAA}$ lacks the ability to repress expression of a GFAP luciferase reporter

P19 cells were transfected with a GFAP luciferase reporter with various combinations of Sox2 and Groucho. The expression of the GFAP reporter was repressed by Sox2 co-expression and the Groucho co-expression suppressed the GFAP reporter even further. However, the $Sox2^{LQY/AAA}$ mutant alone or $Sox2^{LQY/AAA}$ mutant with Groucho did not affect the GFAP reporter expression.

4.5 Discussion

In this chapter, the Groucho-Sox2 interaction motif has been identified. The ability of Sox2 to repress expression of a GFAP luciferase reporter could be eliminated by mutating only the Groucho-Sox2 interaction motif, LQY, while the transcriptional activator function was retained (Figure 4.4.5; Figure 4.4.6). These results indicate that Sox2 acts as a bi-functional transcription factor through different motifs.

Previous studies on cSox3 in our lab (PhD thesis, Zulfiqar Laghari, 2010) have shown that there are two important domains involved in Sox3-Groucho interaction (domain1 and domain2 in Figure 4.5.1). A C-terminal deletion of chick Sox3 which lost domain1 was shown to lose the ability to translocate Groucho5. However, the internal deletion of domain1 and the point mutation of LQY in domain1 still retained the Groucho5 translocation ability. This suggests that another Groucho interaction domain exists in the internal deletion and point mutation. When delete both domain1 and domain2, the mutant chick Sox3 totally lost the ability to translocate Groucho5. This implies that not only LQY domain but the two domains (domain1 and domain2) are required for Sox3-Groucho5 interaction in chicken.

Although the domain2 in chick Sox3 is needed for Groucho interaction, the interaction assay in Sox2 deletions shows that is not necessary in mouse Sox2 (Figure 4.3.1; Figure 4.3.2). In mouse Sox2, the domain2 is interrupted by gaps and has less similarity than the one near the N-terminus (Table 4.5.1). Therefore, the interaction ability of the domain2 in mouse Sox2 might be weakened.

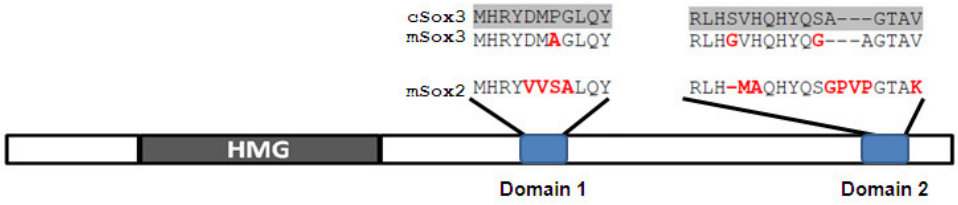


Figure 4.5.1 Schematic representation of Sox-Groucho interaction domain

The sequence of the two Sox-Groucho interaction domains in chick Sox3 are shown in grey. The homolog in mouse Sox3 and Sox2 are listed below cSox3 as mSox3 and mSox2. The various amino acids are marked in bold. The relative position of the domains in mouse Sox2 are illustrated.

sequence	identity	Gap frequency
cSox3 MHRYDMPGLQY mSox2 MHRYVVSALQY **** ***	63.6%	0%
cSox3 RLHSVHQHYQSA---GTA mSox2 RLHMA-QHYQSGPVPGTA *** ***** ***	61.1%	22.2%

Table 4.5.1 Alignment of Sox-Groucho interaction domain in chick Sox3 and mouse Sox2

Asterisks mark the amino acid identity between in cSox3 and mSox2. Dashed lines shows gaps. The identity and gap frequency were calculated by SIM – Alignment Tool for protein sequences, ExPASy (**http://www.expasy.ch/tools/sim-prot.html**).

To date, two Groucho interaction motifs have been defined, the WRPW/Y motif and the engrailed homology (FxIxxIL) interaction motif (Fisher et al., 1996; Goldstein et al., 2005; Kang et al., 2005). However, these two motifs interact mainly with the C-terminal WD domain of Groucho. Transcription factors that interact with the N-terminal Q domain of Groucho such as Six3 and TCF, were reported to function through an engrailed-like motif (Brantjes et al., 2001; Zhu et al., 2002). Runx2 (Wang et al., 2004) and Nolz1 (Ji et al., 2009) also interact with the N-terminus of Groucho5 but via a long, poorly defined region of C-terminal amino acids. Sequence comparison of the engrailed-like motif of Goosecoid, Pax5, Six3 and TCF1 revealed loose similarity of the first

hydrophobic aromatic amino acid (W, F, Y) and sixth aliphatic amino acid (I, L, V) (Table 4.5.2). The Sox2-Groucho interaction domain identified in this chapter also has loose similarity with the engrailed-like motif. Therefore, when the conserved LQY motif was mutated, a possible engrailed-like motif was also disrupted.

Gene	Engrailed like motif sequence
dEn	(172) SLAFSISNILS (182)
mGsc	(3) ASMFSIDNILA (13)
mPax5	(176) GSSYSISGILG (186)
mSix3	(85) TLNFSPEQVAS (95)
hTCF1	(210) VSWFTHPSLML (220)
mSox2	(199) MHRYVVSALQY (209)
	* ×

Table 4.5.2 Sequence comparison of the engrailed-like motif

The engrailed-like motif sequences are aligned in the list. An asterisk indicates the conserved hydrophobic aromatic amino acids (W, F, Y) and the cross indicates the conserved aliphatic amino acid. The species of genes are marked in front of the gene name (d: Drosophila; m: mouse; h: human) and the numbering of the amino acids are listed.

The C-terminal deletions of Sox2 might have dramatic structural changes which affect the interaction with Groucho. Therefore, the LQY mutant was made to minimize the change of the Sox2 protein structure. Using the I-TASSER service (**http://zhanglab.ccmb.med.umich.edu/I-TASSER/**), the structure and the predicted protein interaction residues of wild type and LQY mutated Sox2 were modelled (Figure 4.5.2). Although the LQY motif did not appear to be associated with the protein interaction residues, the YVV within the predicted engrailed-like motif did (Figure 4.5.2 A). The change of LQY into AAA inverts the surface position of the predicted engrailed-like motif (Figure 4.5.2 A; Figure 4.5.2 B). This might interrupt the function on engrailed-like motif function and weaken the interaction of Sox2 and Groucho. Further mutational analysis of this region is therefore merited.

A.

MYNMMETELKPPGPQQASGGGGGGGNATAAATGGNQKNSPDRVK**R**PMNAF
MVWSRGQRRKMAQENPKMHNSEISKRLGAEWKLLSETEKRPFIDEAKRL**R**
ALH**M**KEHPDY**KY**RPRRKTKTLMKKDKY**TLPGGLLAP**GGNSMASGVGVGAG
LGGGLNQRMDSYAHMNGWSNGSYSMMQEQLGYPQHPGLNAHGAAQMQPM**H**
R**YVV**SA**LQY**NSMTSSQT**YMN**GSPTYSMSYS**QQGTPGMALG**S**M**GS**V**V**K**SEA
SSSPPVVTSSSHSRAPCQAGDLRDMISMYLPGAEVPEPAAPSRLHMAQHY
QSGPVPGTAKYGTLPLSHM

B.

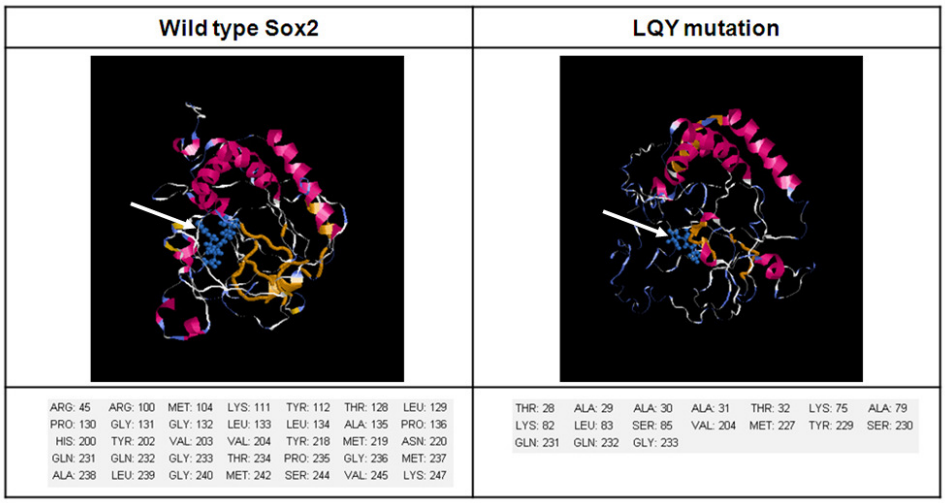


Figure 4.5.2 Sox2 protein interaction domain prediction

The amino acid sequence of wild type and LQY mutated Sox2 were sent to the I-TASSER website to analyse the possible structure. The sequence in (A) represents the predicted protein interaction residues of wild type Sox2 (marked red and bold). The 3D structure is shown in (B) The helix is marked in pink, protein interaction residues in yellow and the white arrows point LQY/AAA domain in blue. The protein interaction residues are listed below.

The results of this chapter indicate that there is an engrailed-like motif from 199-209 in mouse Sox2. Changing three amino acids (LQY) within the engrailed-like motif resulted a loss-of-function mutant of Sox2 ($Sox2^{LQY/AAA}$) in Groucho interaction and transcriptional repressor function by altering the engrailed-like motif surface position.

Chapter 5 Sox2 repressor function in human neural stem cells

5.1 Introduction

For over a decade, Sox2 has been regarded as a transcriptional activator that is widely expressed in the nervous system during development (Kamachi et al., 1995; Kamachi et al., 1998). In 2006, Takahashi and Yamanaka implicated Sox2 in their discovery of induced pluripotency. They successfully reprogrammed somatic cells to pluripotent cells by overexpressing only four key transcription factors Oct4, Klf4, c-myc and Sox2 in mouse embryonic fibroblasts (MEF) and named the ES-like cells "induced pluripotent stem (iPS) cells" (Takahashi and Yamanaka, 2006). Since then, many studies have focused on dissecting the core transcriptional network in embryonic stem cells. Genome wide searches for the DNA binding sites and genes regulated by Sox2, one of the "Yamanaka factors", have been described in recent studies. Boyer (2005) and Sharov (2008) identified the DNA binding sites of Oct4, Sox2 and Nanog using ChIP-chip (Chromatin immunoprecipitation on chip) in human and mouse embryonic stem cells respectively (Boyer et al., 2005; Sharov et al., 2008). Furthermore, Kim (2008) and Zhou (2007) used the next-generation sequencing technique, ChIP-PET (the combination of ChIP and paired-end tags) to analyse the DNA binding sites of Sox2 in human and mouse embryonic stem cells (Kim et al., 2008a; Zhao et al., 2007). ChIP-seq (ChIP-sequencing) was carried out to map the locations of Sox2 binding sites in mouse embryonic stem cells by Chen and colleagues (Zhou et al., 2007) and gene expression profiles were investigated in human embryonic stem cells in which Sox2

knock-down and embryonic carcinoma cells and Sox2 null mouse embryonic stem cells (Greber et al., 2007; Masui et al., 2007).

Together these studies show that the expression of a number of genes decreased in Sox2 knock–down cells. Among these genes were several encoding transcription factors (e.g., Oct, Nanog, STAT3) and components of the Tgf-β and Wnt signalling pathway (eg., Lefty2, Dkk1). This implicates Sox2 as a transcriptional activator, consistent with previous studies of Sox2 transcription factor function (Miyagi et al., 2009). On the other hand, expression of almost the same number of genes increased in Sox2 knock-down cells (623 up-regulated; 648 down-regulated; (Masui et al., 2007)) which means that Sox2 normally represses the expression of these genes, either directly or indirectly. Intriguingly, when compared with the Sox2 ChIP assay results, a number of these Sox2 down-regulated genes also been bound by Sox2 in their promoter regions. This suggests that a number of genes are directly repressed by Sox2 in the human or mouse embryonic stem cells.

Through a meta analysis of the results of these studies I identified several genes that were up-regulated in Sox2 knock-down cells and could also bind Sox2 potential binding sites (Table 5.1.1). some of these related to embryonic development, such as left-right axis formation (Bmp4 (Monsoro-Burq* and Douarin, 2000), Nodal (Lowe et al., 1996) and Zic3 (Gebbia et al., 1997)), eye development (Pax6 (Hever et al., 2006)) and brain formation (Hesx1 (Dattani et al., 1998), Id2 (Toma et al., 2000), Nfib (Deneen et al., 2006) and Nedd9 (Aquino et al., 2009)).

In order to test the role of Sox2 transcriptional repressor function in human neural stem cells (hNSC), wild type Sox2 and the Groucho-binding mutant

form ($Sox2^{LQY/AAA}$) were introduced into the human NSC line, RNA was then isolated for Affymetrix expression array. The results of three expression arrays are presented and discussed in this chapter.

Gene	Sharov, siRNA mES Expression array	Masui, inducible sox2 null mES Expression array	Sharov, mES ChIP	Chen Xi, mES ChIP	Kim, mES ChIP	Boyer, hES ChIP
Mppe1	√	√	√			
Nfib	√	√	√			
Slc6a1	√	√	√			
Tmsb4x	√	√	√			
Id4	√	√	√	√		
Esx1		√				√
Dazl		√	√			√
Etv5		√	√		√	√
Foxd3		√	√		√	√
Gbx2		√	√		√	√
Pax6		√	√		√	√
NEDD9		√				√
Bmp4		√	√		√	
Hesx1		√		√	√	√
Fgf4		√		√	√	
Gadd45b		√		√		
Nodal		√		√	√	
Id2		√		√		
Zic3		√	√	√		√

Table 5.1.1 Summary of possible Sox2 down-regulated target genes.

Possible Sox2 down-regulated genes are listed on the left. Expression array data originate from Sharov and Masui (Masui et al., 2007; Sharov et al., 2008). The ChIP assay data originate from Sharov, Chen, Kim and Boyer (Boyer et al., 2005; Chen et al., 2008; Kim et al., 2009b; Sharov et al., 2008). The ticks represent the up-regulation of genes in expression array or signals in ChIP assay.

5.2 Preparation of samples for expression array

In order to understand the transcriptional repression role of Sox2 in hNSCs, Sox2 and Sox2 $^{LQY/AAA}$ expression plasmid were introduced into hNSCs. In theory, the faster a gene's response to a transcription factor, the more likely the effect of the transcription factor function is direct. Hence, in order to obtain data that reflect possible Sox2 directly regulated genes, it was considered better to collect cells at the earliest time point of Sox2 expression.

To monitor the expression level of an exogenous gene in hNSCs, a GFP expression vector was co-transfected into hNSC with Sox2 and the cells expressing GFP were observed (Figure 5.2.1). The results show that exogenous gene expression increased dramatically 14 hours after transfection (Figure 5.2.2). Therefore, this time point was used to collect hNSC for the expression array analysis. Since the transfection efficiency of hNSCs was always below 100%, non-transfected cells still exist within the transfected cell population. The observation under the fluorescent microscope showed approximately 40% of the cell population were non-transfected, therefore the noise from the non-transfected cells would have a significant potential affect on the expression array data. In order to avoid the noise from these non-transfected cells, Fluorescence Activated Cell Sorting (FACS) was carried out on the cells 14 hours after transfection.

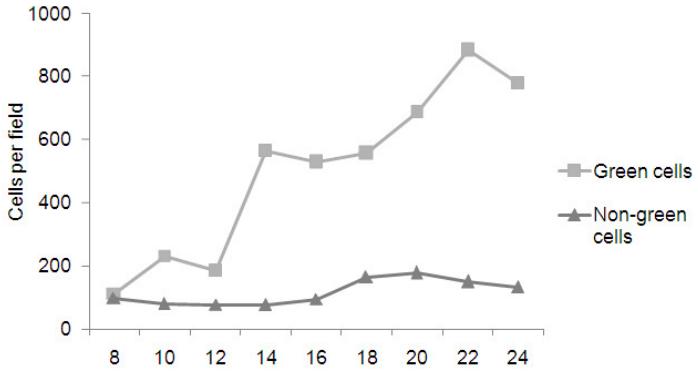


Figure 5.2.1 Timecourse of GFP expression in transfected hNSCs

hNSCs were transfected with GFP/Sox2 and the green and non-green fluorescent cells were observed at different time points (8-24 hours) and the cells counted.

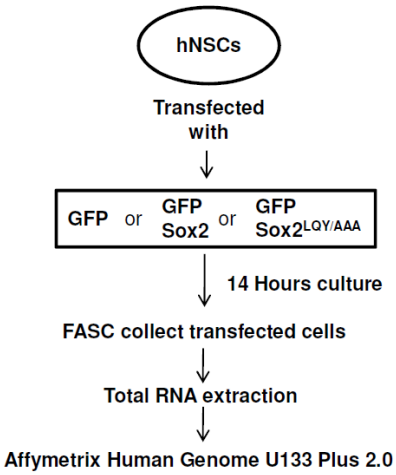
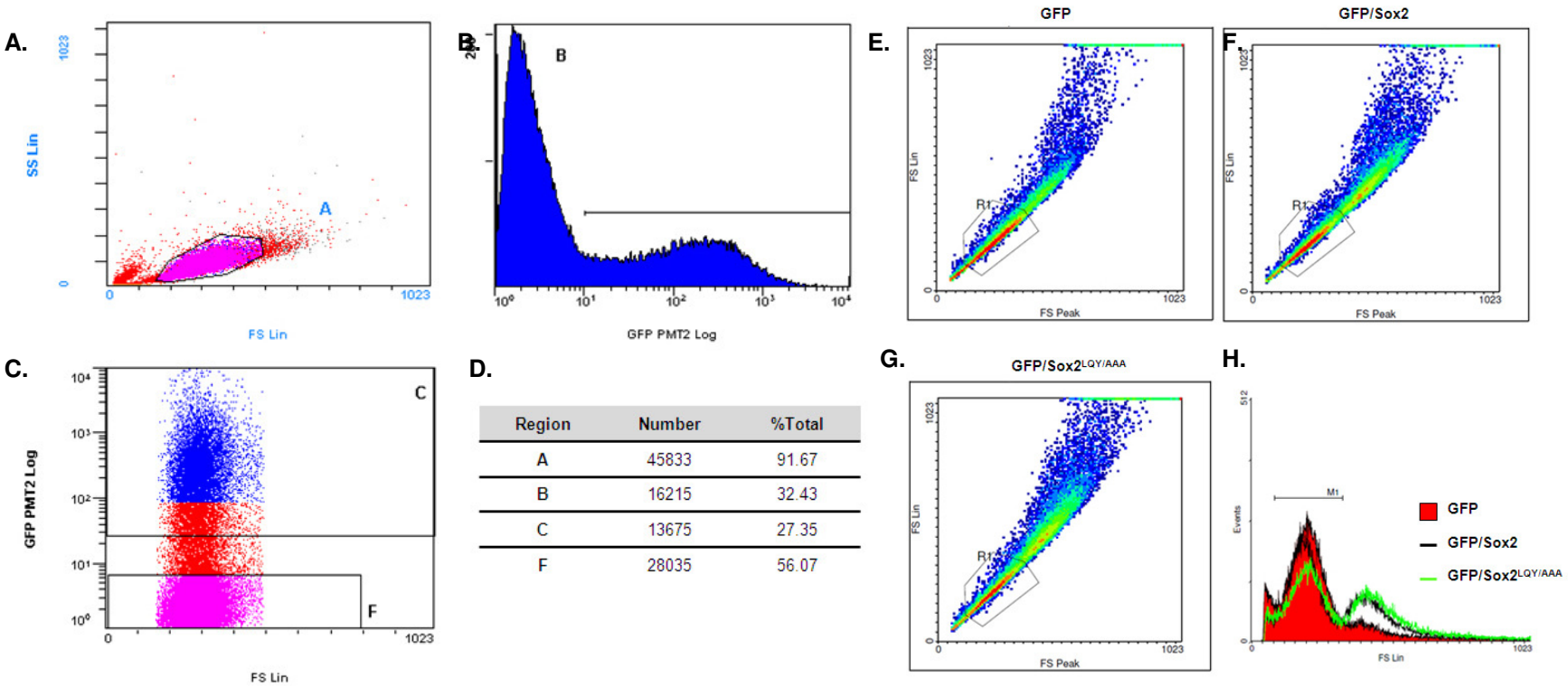


Figure 5.2.2 Flow chart of Affymetrix experiment

hNSCs were collected 14 hours after transfection with either GFP alone, GFP with wild type Sox2 or GFP with Sox2 $^{LQY/AAA}$. Fluoresence Activated Cell Sorting (FACS) was carried out to sort out the untransfected (GFP negative) cells. The GFP positive cells were collected and the total RNA extracted immediately. Total RNA was purified and applied to the Affymetrix expression array system.

During the electroporation procedure, it is very likely that cells take up GFP and Sox2 plasmid DNA simultaneously. Previous data in our lab suggested more than 90% co-transfection. Therefore, the cells expressing GFP are very likely to also express exogenous Sox2. According to this hypothesis, the Sox2

expressing cells could be sorted through the expression of GFP. FACS allowed the separation of the healthy and GFP positive cells from a group of cells (Figure 5.2.3).



Region	Number	%Total
A	45833	91.67
B	16215	32.43
C	13675	27.35
F	28035	56.07

Figure 5.2.3 FASC sorting charts

FACS detected the volume of cells by light scatter intensity (Figure 5.2.4 A). The healthy cells in region A were selected and the green strength of cells was detected (B). The fluorescence intensity of GFP expressing cells was measured and the cells which had higher fluorescence intensity (higher than 10 units) was shown in panel C. Cells which expressed fluorescence higher than 10 units were selected (region C in panel C) as the green positive cells whereas the low green level cells were selected as green negative (region F in panel C). Panel E-G show the distribution of fluorescence in GFP, GFP/Sox2 or GFP/Sox2$^{LQY/AAA}$ transfected cells. The cells in R1 region were collected. Panel H shows the relative graph of fluorescence intensity with cell numbers. The region R1 in panel E-G is marked as M1 in panel H.

The FACS technique detected the volume of cells by light scatter intensity (Figure 5.2.3 A). The parameters of every cell were shown as a dot on the SS/FS cytogram. More than 90% of cells were healthy in the transfected hNSCs population (Figure 5.2.3 D). The healthy cells in region A were selected and the green fluorescence level of cells were detected (Figure 5.2.3 B, C). Cells expressing more than 10 units of fluorescence were selected (region C in Figure 5.2.3 C, approximate 27% of region A selected cells) Figure 5.2.3 E-G shows the distribution of fluorescence in GFP, GFP/Sox2 or GFP/Sox2$^{LQY/AAA}$ transfected cells. The cells in R1 region were collected and the relative graph of fluorescence intensity with cell numbers is showed in Figure 5.2.3 H.

In order to avoid the degradation of RNA, cells were lysed immediately in TRI reagent. The total RNA of GFP positive cells was purified using RNeasy® Mini column (Qiagen). The quality of the total RNA was checked using a 2100 BioAnalizer (Agilent) (Figure 5.2.4). The RNA integrity number of all three sets of total RNA (GFP, GFP/Sox2 and GFP/ Sox2 $^{LQY/AAA}$) was 10 which indicated the quality of the total RNA was good enough.

The Affymetrix GeneChip® Human Genome U133 plus2 array was chosen to detect the expression profile in this chapter. There are up to 1.3 million different oligonucleotide probes attached on the chip which offers a rapid and thorough assay of human gene expression. Figure 5.2.2 illustrates the experimental strategy to determine the effect of Sox2 overexpression in hNSC using Affymetrix expression arrays. Further analysis of the expression array data will be discussed in the next section.

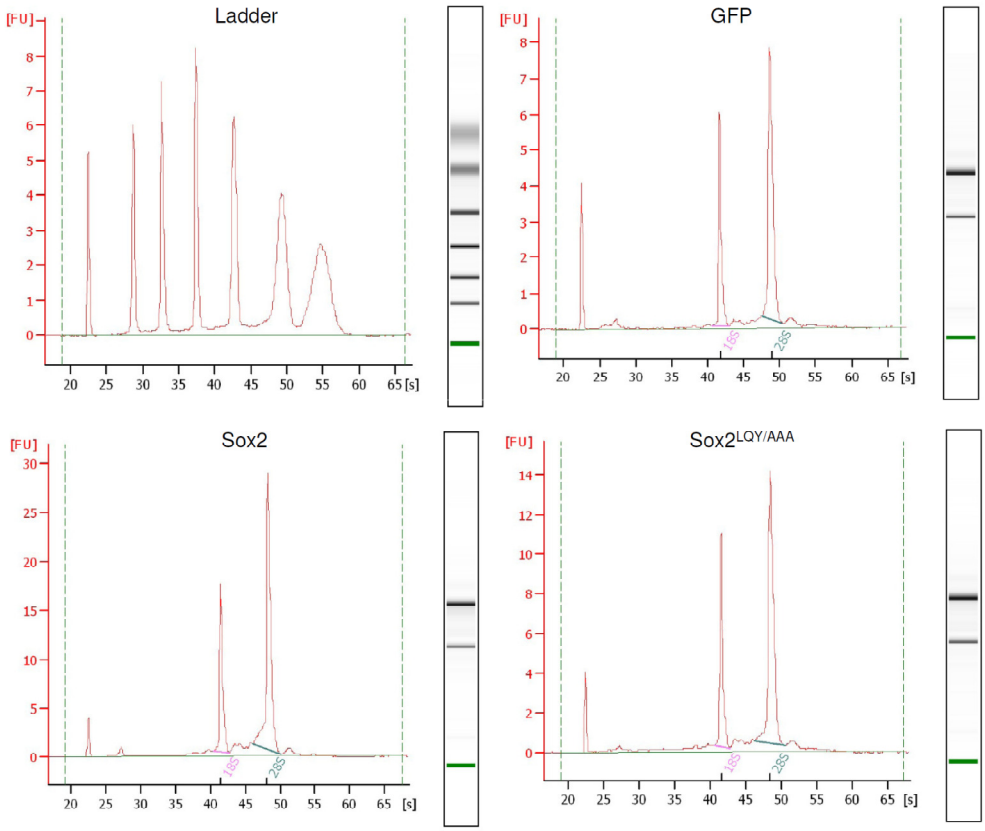


Figure 5.2.4 RNA quality checking

The total RNA of GFP, GFP/Sox2 and GFP/ Sox2 $^{LQY/AAA}$ transfeced hNSCs were analysed by 2100 Bio Analizer (Aglient). Simulated electrophoresis images are presented on the right.The ladder graph shows the marker peaks as control. The 18s and 28s peaks are shown in graphs from GFP, Sox2 and Sox2 $^{LQY/AAA}$ transfected cells.

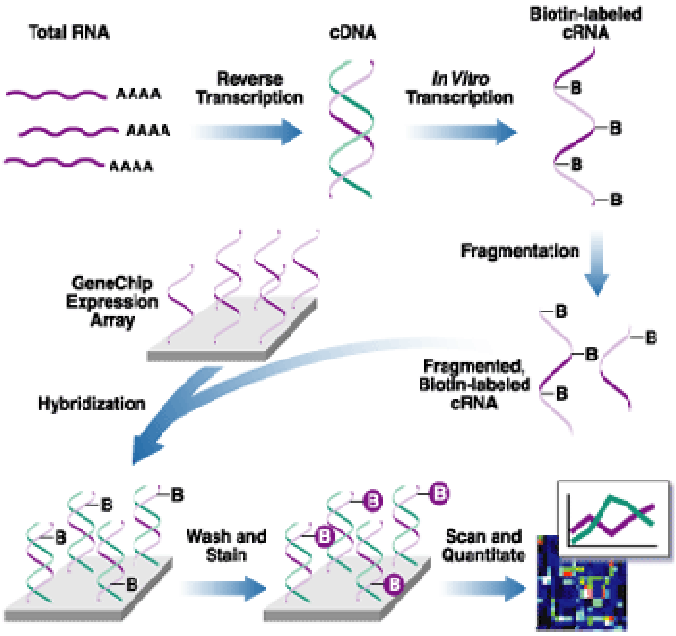


Figure 5.2.5 Diagram of Affymetrix GeneChip® Human Genome U133 plus2 array.

The total RNA from different experiments of hNSCs was subjected to analysis by the Affymetrix GeneChip® Human Genome U133 plus2 arrays. Reverse transcription using poly dT primers was carried out on the purified mRNA to generate cDNA. Biotin-labeled cRNA was prepared from the cDNA using an *in vitro* transcription method. The cRNA was fragmented before hybridization. After 16 hours of hybridization, the chip underwent an automated washing and staining in a fluidics station. Immediately following the washing and staining, the chip was scanned using the Affymetrix Microarray Suite. The software defines the probe cells and computes an intensity for each cell. (http://www.dkfz.de/gpcf/24.html)

5.3 Affymetrix array result

The quality control reports of Affymetrix arrays are listed in Table 5.3.1. The background signal shows the autofluorescence of the array surface and non-specific binding of target molecules. The ideal value of background signal is less than 100. The background signals of three arrays were approximate 50. The ideal value of the number of present cells is 40-50%. The number of present cells in three arrays was higher than 45%. The scale factor is the number that was used to normalize the signal. The ideal value of scale factor is around 3. The scale factors of three arrays fit this condition. The sig (3'/5') of β-Actin and GAPDH indicate the ratio of the labelling reaction and the value should be less than 3. The value of sig (3'/5') of β-Actin and GAPDH in all three arrays agreed with the standard. Together, the quality control reports showed all three arrays are successful and reliable.

Filename	GFP	GFP/Sox2	GFP/Sox2$^{QLY/AAA}$
Background	47.53	49.12	51.06
Number Present	47.30%	48.60%	45.50%
Scale Factor	3.171	2.883	3.54
Sig(3'/5') β-Actin	2.15	2.62	2.15
Sig(3'/5') Gapdh	1.01	1.06	1.04

Table 5.3.1 Quality control report of three arrays

The quality control results of three arrays are listed in the table. The background signal should be less than 50, number present should be 40-50%, scale factor should be around 3, sig(3'/5') of β-Actin and Gapdh should be less than 3.

Raw data from the Affymetrix arrays was normalized using the Affymetrix Microarray Suite software according to the reading of control probe sets. The Affymetrix array gave raw data of 54675 probe sets. Expression levels could be presented in different colours as a heat map. Figure 5.3.1 shows part of the heat maps for the arrays performed. The expression level of 372 genes was

increased following overexpression of wild type Sox2 or $Sox2^{LQY/AAA}$ including Hes5 (GFP: 139.672; GFP/Sox2: 1239.22; GFP/$Sox2^{LQY/AAA}$: 538.51) which has been shown to be up-regulated following Sox2 or $Sox2^{LQY/AAA}$ overexpression in mouse neural stem cells (Bani-Yaghoub et al., 2006). The expression level of 486 genes was decreased follow overexpression of wild type Sox2 when compare to the overexpression of GFP alone. This implies that Sox2 might repress the expression of these genes. Interestingly, some of the putative Sox2 repressed genes showed a higher expression level following overexpression of $Sox2^{LQY/AAA}$ when compared with overexpression of wild type Sox2. These genes included NNMT, PLAUR, LAMP and RAC3 (Table 5.3.2). Since $Sox2^{LQY/AAA}$ lacks the ability to interact with Groucho, the decrease of Sox2 suppression might relate to the co-repressor function of Groucho.

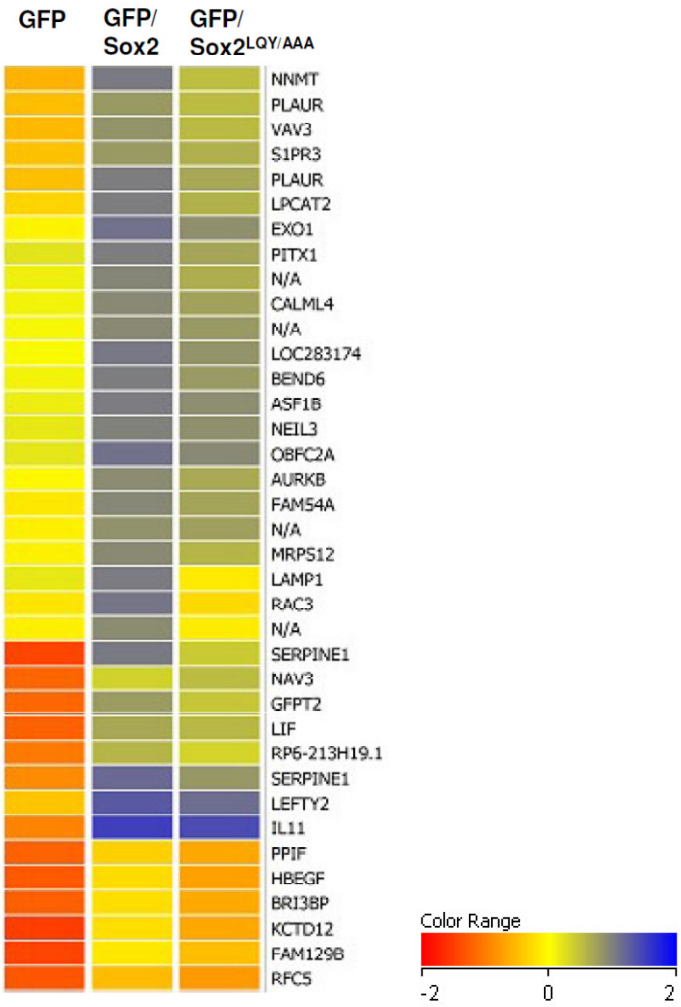


Figure 5.3.1 Partial of heat map of Affymetrix array result

The heat map represents the expression level of genes in hNSCs overexpression GFP, GFP/Sox2 or GFP/Sox2$^{LQY/AAA}$. Gene names are listed on the right. Red to blue colour represents the high to low expression level (colour scale shows on the bottom right). This part of the heat map, has been selected because it features genes that repressed by overexpression wild-type Sox2 but less so by the overexpression of Sox2$^{LQY/AAA}$.

Gene	Probe ID	GFP	GFP/Sox2	GFP/Sox2$^{LQY/AAA}$	Position
SOX2	213721_at	1901.465	2179.176	2202.344	chr3:182912396-182913912 (+) // 91.96 // q26.33
SOX2	213722_at	125.2528	159.6586	196.8431	chr3:182912396-182913912 (+) // 91.96 // q26.33
SOX2	214178_s_at	17.93751	1864.176	188.536	chr3:182912804-182913311 (-) // 94.67 // q26.33
SOX2	228038_at	7104.177	8169.044	8231.653	chr3:182914305-182914915 (+) // 97.13 // q26.33
NNMT	202237_at	359.1497	86.32508	137.84308	chr11:113672381-113688448 (+) // 98.84 // q23.2
PLAUR	210845_s_at	329.7068	106.50976	135.50977	chr19:48844571-48866539 (-) // 75.19 // q13.31
LAMP1	201551_s_at	183.50975	86.8374	242.59111	chr13:113008858-113024817 (+) // 95.43 // q34
RAC3	206103_at	250.84312	83.32363	272.356	chr17:77582856-77585025 (+) // 95.0 // q25.3

Table 5.3.2 Gene expression raw data

Raw data from all the probes of Sox2 and possible Sox2 repressed genes in each set of experiment are listed and underlined. The position of genes on the chromosome shows on the right column.

5.3.1 Signal for Sox2 probes on mRNAs

There were several probe sets for Sox2 on the U133 arrays. The raw data for the Sox2 probes on the arrays showed a wide variation in levels of increase of Sox2 expression in each array (Table 5.3.2 underlined). The possible reason of this inconsistent result might be the 3'-end bias appears quite obvious from the poly-dT primers reverse transcription reaction. The closer the probe to the 3'-end the stronger the signal will be. Figure 5.3.2 shows the relative position the Sox2 probes on the transcript. Since the wild-type Sox2 and Sox2$^{LQL/AAA}$ that were overexpressed were mouse genes, the probes that to detected the Sox2 coding region would match both human and mouse sequence, whereas the probes located in the for 5' and 3' UTR would only able to detect human Sox2. Together with the raw data of each set of probes, the 3'-end bias appears quite obvious. On the other hand, according to the raw data, the increase in Sox2 expression level could not clearly be determined. Therefore, further analysis of Sox2 expression of the same RNA sample was needed, in order to validate the experimental strategy.

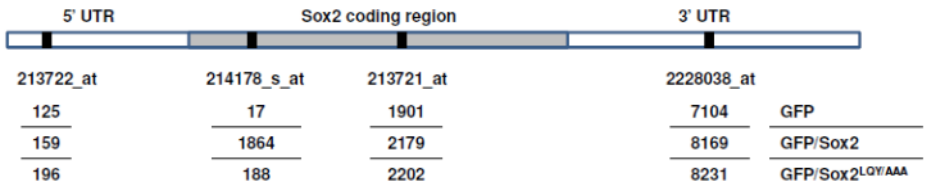


Figure 5.3.2 Sox2 probe position

Illustrated relative position of Sox2 probes on the Sox2 transcript. The raw data in each experiment is shown beneath the probe ID. The sequence of the probes in Sox2 coding region is matched with both mouse and human Sox2 gene, whereas the sequence of the probes in 5' and 3' UTR is matched only with human Sox2.

5.4 QRT-PCR analysis of Sox2 overexpression

In order to check the exact expression level of Sox2, QRT-PCR was carried out (Figure 5.4.1).

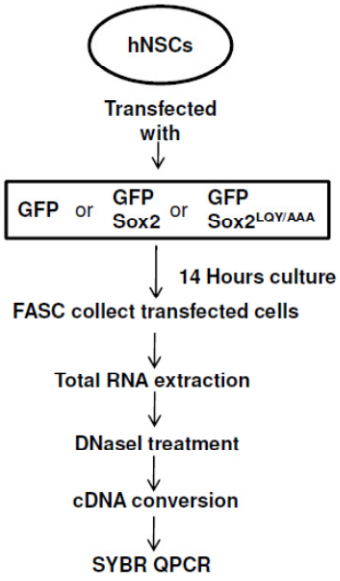


Figure 5.4.1 Illustration of sample preparation for SYBR QRT-PCR

hNSCs were transfected with GFP, GFP/Sox2 or GFP/Sox2$^{LQY/AAA}$. After 14 hours culture, the GFP expressed cells were collected by FACS sorting and the total RNA were extracted. DNaseI was used to digest the remaining genomic DNA. The total RNA were used to generate cDNA by reverse transcriptase. The cDNA samples were used as template and applied to SYBR QRT-PCR.

The same total RNA samples that were sent for Affymetrix expression array analysis were used to measure the Sox2 RNA level by QRT-PCR. The total RNA samples were used to generate cDNA by reverse transcription reaction (RT) with the mixture of random primer and poly-dT primer to avoid the 3'-end bias. The cDNA samples were then applied to SYBR Green QRT-PCR. Contamination with genomic DNA in the cDNA sample would affect the QRT-PCR result. Therefore, a control sample with no reverse transcriptase (non-RT) was included (Figure 5.4.2 A). If the RNA sample contained no genomic DNA, then the non-RT sample would not produce a signal in QRT-PCR. In order to

test if there were any primer dimer signals in the QRT-PCR, a no template control (NTC) was also subjected to QRT-PCR reaction. Each experiment was performed in triplicate.

In order to test if the overexpression of wild type mouse Sox2 or Sox2$^{LQY/AAA}$ was succesful, primers for the 5'-end, coding region and 3'-end of Sox2 were designed. Since the expression plasmid only contained the mouse Sox2 coding region sequence, primers for the non-coding regions were designed to target only human sequence whereas the primers for the coding region were designed to target both human and mouse Sox2 (Figure 5.4.2 B).

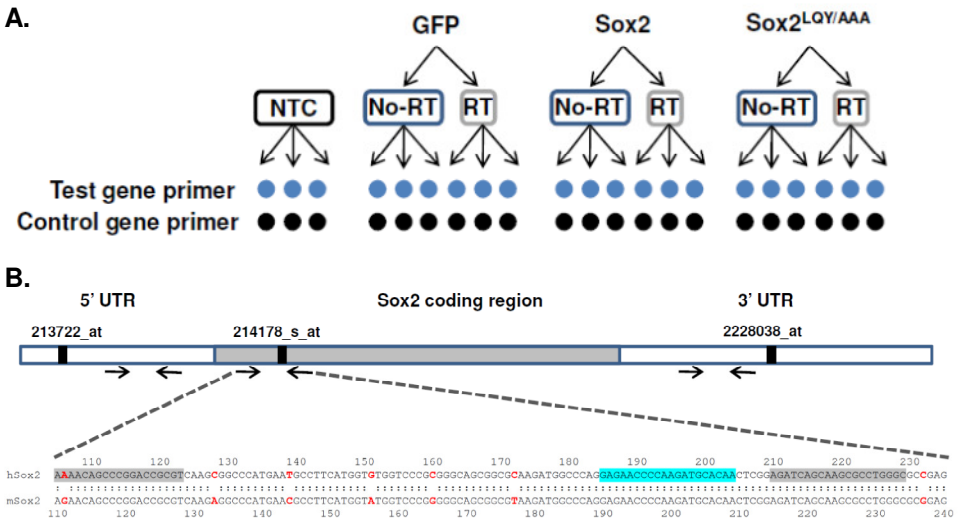


Figure 5.4.2 QRT-PCR experiment design

(A) The purified total RNA from GFP, GFP/Sox2 and GFP/Sox2$^{LQY/AAA}$ transfected hNSCs were used to generate cDNA by reverse transcriptase. Every set of reverse transcription has a no-RT (no reverse transcriptase) control in order to detect the genomic DNA contamination. The cDNA was used as template in QRT-PCR reaction. Every set of primer has a NTC (no template control) to test primer dimer existence. QRT-PCR using control gene primer (housekeeping gene, such as Actin) was carried out in order to normalize the expression level of test gene. (B) Sox2 5'-end, cording region and 3'-end QRT-PCR primers were designed as close to the Affymetrix probe as possible. The relative position of QPR-PCR primers are marked as arrows and the Affymetrix probes are marked as black band. Sox2 coding region primers were designed to match both human and mouse Sox2 and surround the Affymetrix probe set. The alignment of human and mouse sequence is showed beneath.

QRT-PCR monitors the PCR product in real time, reflected in the SYBR green signal on the plot (Figure 5.4.3 A). After 40 cycles of PCR reaction, the product was taken through a temperature gradient (50°C to 90°C) to denature the DNA product. This would show peaks at different time point on the melting curve plot when primer dimers exist in the PCR product. The melting curve of Sox2 coding region QRT-PCR reaction showed a single peak in every sample (Figure 5.4.3 B). This means the result of Sox2 coding region QRT-PCR is reliable. According to the QRT-PCR signal curve, the PCR products of GFP/Sox2 and GFP/Sox2$^{LQY/AAA}$ transfected hNSCs appeared earlier and higher than the GFP transfected hNSCs. This indicates that Sox2 expression levels in GFP/Sox2 and GFP/Sox2$^{LQY/AAA}$ transfected hNSCs were much higher than the GFP transfected set (Figure 5.4.3 A). As for the human Sox2 3'-end and 5'-end primers, the signal curves were similar in GFP, GFP/Sox2 and GFP/Sox2$^{LQY/AAA}$ transfected hNSCs (Figure 5.4.3 C, E). This result means the expression of endogenous human Sox2 transcripts in each set was similar. However, the melting curve of human Sox2 5'-end primers shows two peaks (Figure 5.4.3 F). This might imply that the primers are not specific. On the other hand, the melting curve of human Sox2 3'-end primers shows only one peak (figure 5.4.3 D) which means these primers are very specific and reliable.

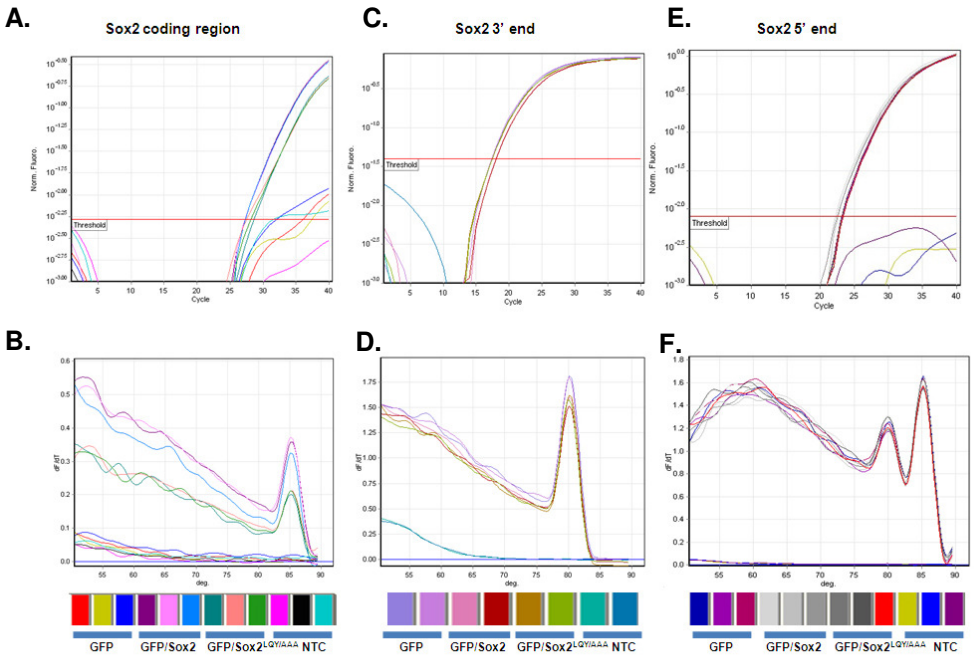


Figure 5.4.3 QRT-PCR checking Sox2 expression

The signal curves of QRT-PCR using Sox2 3'-end, coding region and 5'-end primers and no template control (NTC) in GFP, GFP/Sox2, GFP/Sox2$^{LQY/AAA}$ transfected hNSCs total RNA are shown in (A)(C)(E). The cycle threshold (Ct) value was determined according to the cutting through (lable as threshold on the plot) which was calculated by software from the QRT-PCR of the same primers on serial dilutions of template. The melting curves from 50°C to 90°C of each set of QRT-PCR reaction are shown in (B)(D)(F). The colour code of the plots is showed underneath.

The relative expression level of Sox2 was calculated according to the Pfaffl equation (2001) (Figure 5.4.4). The QRT-PCR result shows that the Sox2 expression level in GFP/Sox2 and GFP/Sox2$^{LQY/AAA}$ transfected hNSCs was more than 5 times higher than the GFP transfected hNSCs whereas the human Sox2 expression level are similar in all three sets. This result indicates that the overexpression of Sox2 was successful and the Affymetrix expression array data truly reflected the changes due to increasing levels of Sox2 in hNSCs. With respect to the array data for Sox2 expression shown in Table 5.3.2, these data also fit, since the probe for the 5' UTR and 3' UTR showed similar signal

in the presence or absence of overexpressed Sox2. In fact, the only anomaly is that one of the coding sequence probes failed to show a response in the $Sox2^{LQY/AAA}$ transfected cells. Therefore, we assume that this is a technical error. It is interesting to note, however, that the level of increase of Sox2 on the array for the coding sequence probes bears little resemblance to the QRT-PCR data. We can only calculate that this in a technical problem specific to Sox2 analysis since validation of QRT-PCR shown later was consistent with array data for other gain.

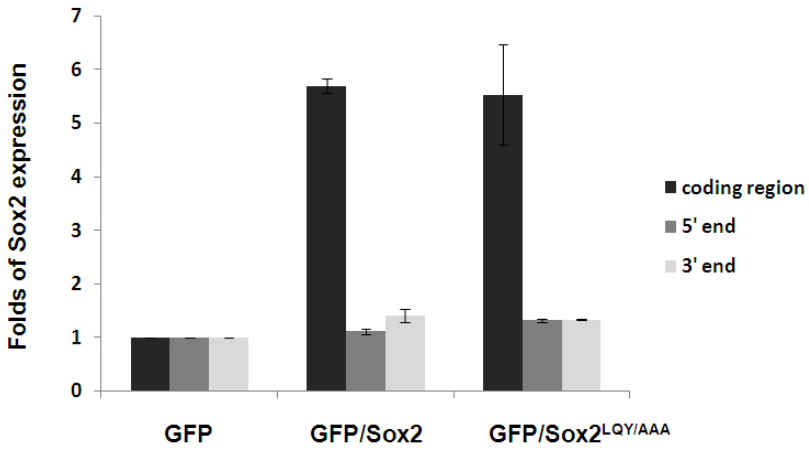


Figure 5.4.4 Quantification of QRT-PCR checking Sox2 expression

The relative expression level was calculated according to the equation in Pfaffl (2001). The result of QRT-PCR using Actin primers was used to normalize the result of Sox2 5'-end, coding region and 3'-end primers. The expression level of Sox2 in GFP/Sox2 and $GFP/Sox2^{LQY/AAA}$ were compared with GFP which was represented as one.

5.5 Affymetrix data analysis

Having established that the overexpression of wild type Sox2 and $Sox2^{LQY/AAA}$ was successful, the data from the Affymetrix expression arrays was analysed to determine the roles of Sox2 and Sox2 repressor function in hNSCs. The Affymetrix expression arrays give data from up to 54,676 probe sets. However, many probes gave very low signals which could not be distinguished from background noises. In accordance in Aris et al (2004), signals lower than 100 were be excluded in order to eliminate false results arising from such noise.

Here, I compare the data from cells overexpressing GFP, GFP+Sox2 or GFP+$Sox2^{LQY/AAA}$ as summarized below. Figure 5.5.1 shows the number of genes for which the expression level changed in cells overexpressing Sox2 or $Sox2^{LQY/AAA}$.

When the raw number of given genes from cells overexpressing GFP+Sox2 was divided by the value for the same genes from cells overexpressing GFP greater than 1.5, the gene was defined as up-regulated in GFP+Sox2 overexpressing cells (676 genes in total, approximately 6%). On the other hand, if the raw number of GFP divided by the one of GFP+Sox2 was greater than 1.5, the gene was defined as down-regulated in GFP+Sox2 overexpressing cells (786 genes in total, approximately 8%). The change in expression of most genes did not exceed 1.5 (8975 genes in total, approximately 86%). A similar trend was seen in the GFP+$Sox2^{LQY/AAA}$ set, $Sox2^{LQY/AAA}$ up-regulated 513 genes (approximately 5%), down-regulated 791 genes (approximately 7%) and most of the genes were unaffected (9194 genes in total, approximately 88%). This result implies that Sox2 or $Sox2^{LQY/AAA}$ overexpression affects the expression of less than 20% of genes in a largely similar manner. According to

the expression array data of Sox2 knock-down mice published by Masui et al. (2009), 623 genes were up-regulated and 648 genes were down-regulated. The similarity of these analyses to my study implies that the overexpression study is not causing significant artefactual effects. When the genes up-regulated by $Sox2^{LQY/AAA}$ were compared to the genes which had been up-regulated by Sox2, most of them were the same (555 genes, approximately 80%). Moreover, in $Sox2^{LQY/AAA}$ overexpressing cells the expression of 136 genes (19%) was not up-regulated to the level as wild-type Sox2 overexpressing cells. On the other hand, only four genes in $Sox2^{LQY/AAA}$ overexpressing cells have expression greater than wild-type Sox2 overexpressing cells (1%). Intriguingly, within the genes which were down-regulated by overexpressing wild-type Sox2, 53 genes (7%) were not down-regulated as much in the $Sox2^{LQY/AAA}$ overexpressing cells whereas only 16 genes (2%) were repressed more. This result is consistent with the loss-of-repressor function of $Sox2^{LQY/AAA}$ which was demonstrated in the previous chapters.

Set	Gene number	%total gene
GFP < Sox2	676	6%
GFP > Sox2	786	8%
GFP = Sox2	8975	86%
GFP < $Sox^{LQY/AAA}$	513	5%
GFP > $Sox^{LQY/AAA}$	791	7%
GFP = $Sox^{LQY/AAA}$	9194	88%
GFP < Sox2 < $Sox^{LQY/AAA}$	4	1%
GFP < Sox2 > $Sox^{LQY/AAA}$	136	19%
GFP < Sox2 = $Sox^{LQY/AAA}$	555	80%
GFP > Sox2 < $Sox^{LQY/AAA}$	53	7%
GFP > Sox2 > $Sox^{LQY/AAA}$	16	2%
GFP > Sox2 = $Sox^{LQY/AAA}$	735	91%

Table 5.5.1 Gene expression analysis

The raw number of GFP/Sox2 was divided by GFP for analysis the up-regulated genes. Raw number of GFP divided by GFP/Sox2 was used as down-regulated ratio. The definition of greater is the ratio larger than 1.5 and the rest of genes are defined as non-affected. The same method was applied to compare the expression of genes in GFP/$Sox2^{LQY/AAA}$ with GFP; Sox2 up-regulated genes with GFP/$Sox2^{LQY/AAA}$; Sox2 down-regulated genes with GFP/$Sox2^{LQY/AAA}$.

5.5.1 Sox2 repressed genes

Among genes repressed by Sox2 in hNSCs, the suppression of some was not as great in cells overexpressing $Sox2^{LQY/AAA}$ (Table 5.5.2; Figure 5.5.1). GFAP has been reported as a target of Sox2 repression (Cavallaro et al., 2008). In my data, GFAP was repressed by wild type Sox2. However, the expression level of GFAP in cells overexpressing $Sox2^{LQY/AAA}$ was not reduced as much as it was in cells overexpressing Sox2 (Figure 5.5.1). This implies that the $Sox2^{LQY/AAA}$ mutant might lost the repression function on GFAP. Since my previous data showed that the $Sox2^{LQY/AAA}$ mutant had lost its interaction with Grouchos, the mechanism of the decreasing of Sox2 repression in $Sox2^{LQY/AAA}$ overexpressed hNSCs might be related to Sox2-Groucho down-regulation machinery. A

similar trend could be seen for NNMT, SOX8, SERPINE1, HELLS, IGFBP2 and IGFBP3, which have also been reported to be Sox2 directly regulated genes (Masui et al., 2007; Greber et al., 2007; Chen et al., 2008; Kim et al., 2008; Boyer et al., 2005). Therefore, further validation of these genes was carried out as described in the next section.

Probe set	GFP	GFP/Sox2	GFP/Sox2$^{LQY/AAA}$	Gene	Sox2 repress	Sox2 binding
229259_at	554.7123	483.8145	608.1765	GFAP	Cavallaro	Cavallaro
203540_at	7772.967	5959.843	7339.053	GFAP		
202237_at	359.1497	86.32508	137.8431	NNMT		Chen
226913_s_at	577.1521	297.1707	492.8431	SOX8	Greber	
202627_s_at	464.5096	76.70228	105.2952	SERPINE1	Masui	Chen
202628_s_at	755.4006	85.8431	148.1116	SERPINE1		
223556_at	171.5098	101.195	109.5222	HELLS		
227350_at	553.1381	265.1765	297.3163	HELLS	Greber	Boyer/Kim
220085_at	260.1416	101.1764	157.527	HELLS		
202718_at	6033.509	3830.924	6291.625	IGFBP2		Boyer/Chen/Kim
210095_s_at	2637.51	1314.747	1433.139	IGFBP3		
205547_s_at	140.1765	61.0221	106.5252	TAGLN	Masui	Chen
200916_at	5701.422	4401.175	5856.467	TAGLN		

Table 5.5.2 Raw data of top choices of Sox2/Groucho regulated gene

The probe set and raw data of genes down-regulated in Sox2 overexpressing cells that were less affected in Sox2$^{LQY/AAA}$ overexpressing cells are listed in this table. The agreement of Sox2 induced down-regulation and existence of Sox2 binding sites of these genes in previous studies are listed. (Masui et al., 2007; Greber et al., 2007; Chen et al., 2008; Kim et al., 2008; Boyer et al., 2005)

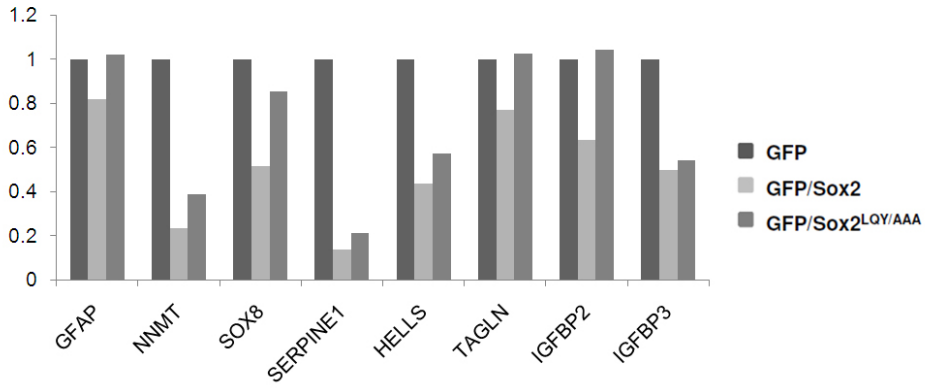


Figure 5.5.1 Top choices of Sox2/Groucho regulated gene expression.

The relative expression level of genes was calculated (normalized by the expression raw data of GFP). The relative value of genes which have raw data from more than one probe set was averaged.

5.5.2 GO term analysis of Affymetrix data

GO terms (gene ontology term) of Sox2 activated and repressed genes were analysed using the DAVID Bioinformatics Resources 6.7 website (http://david.abcc.ncifcrf.gov/; NIH) (Table 5.5.3). Overexpression of Sox2 in hNSCs activated 143 genes that related to transcriptional regulation and 37 genes that related to cell cycle. Surprisingly, not even one gene is related to transcriptional regulation amongst the Sox2 repressed genes. This suggests that the effects of Sox2 in hNSCs might function through activating downstream transcription factors. On the other hand, overexpression of Sox2 repressed 118 genes that related to cell cycle and 0 genes that related to transcriptional regulation.

Sox2 activated			
Gene ontology number	Gene ontology term	Gene number	%
GO:0045449	regulation of transcription	143	21.15385
GO:0006350	transcription	111	16.42012
GO:0051252	regulation of RNA metabolic process	106	15.68047
GO:0006355	regulation of transcription, DNA-dependent	104	15.38462
GO:0007166	cell surface receptor linked signal transduction	82	12.13018

Sox2 repressed			
Gene ontology number	Gene ontology term	Gene number	%
GO:0007049	cell cycle	118	14.99365
GO:0022402	cell cycle process	91	11.5629
GO:0006259	DNA metabolic process	84	10.67344
GO:0022403	cell cycle phase	78	9.911055
GO:0000279	M phase	67	8.513342

Table 5.5.3 Top name GO term of Sox2 regulation genes

The probe set list was analysed by DAVID Bioinformatics Resources 6.7 website (http://david.abcc.ncifcrf.gov/; NIH). The top name GO terms of Sox2 activated and repressed genes are showed in the list. The GO number, GO term, gene number and the percentage are listed in the table. Genes been categorised in different name classes share many genes in common. All the genes in transcription, regulation of RNA metabolic process, regulation of transcription, DNA-dependent and cell surface receptor linked signal transduction are including in regulation of transcription. Also, all the genes in cell cycle process, DNA metabolic process, cell cycle phase and M phase are included in cell cycle.

When compared with the data from previous studies of Chip-on-chip and expression arrays, 37 genes were identified as strong candidates for direct repression by Sox2 (Table 5.5.4). These genes were repressed by Sox2 more than 1.5 times in hNSCs and the repression was in agreement with data from Sox2 knock down mice or human embryonic stem cells. Also, their promoter regions had been shown to be bound by Sox2 in previous studies. Among these genes, IL11, CXCL14, CCL2 and CAMK1D are related to cytokine signal pathways, E2F8, EMP2 and AXL are related to cell cycle regulation and BASP1, PITX1, ARHGEF10 and GLIPR1 are related to neuron formation.

	Fold changes in Sox2 overexpressed hNSCs	Up-regulated in mouse Sox2 KD (Masui, 2007)	Up-regulated in human Sox2 KD (Greber, 2007)	Sox2 target
IL11	8.53		x	Chen
SERPINE1	5.32		x	Chen
LEFTY2	4.47	x		Chen/Kim
NNMT	4.26		x	Chen
CXCL14	3.59		x	Chen
E2F8	2.94	x		Chen
OLFML3	2.85	x		Boyer
SLC35F2	2.74		x	Chen
HMGA2	2.68		x	Chen
CCL2	2.67		x	Boyer
IER5L	2.39	x		Boyer
TNFRSF12A	2.36	x	x	Chen/Kim
BASP1	2.20	x		Chen
EMP2	2.16	x		Chen
CDK5R1	2.04		x	Chen
CAMK1D	2.02	x		Chen
DDIT4	2.01	x		Chen
LMCD1	1.96		x	Chen
TAGLN	1.93	x	x	Chen
AXL	1.91		x	Chen
RGS16	1.91		x	Kim
SLC31A2	1.87		x	Chen
ETS2	1.82		x	Chen
PITX1	1.77	x		Chen
RASSF1	1.71		x	Chen
ARHGEF10	1.68	x		Chen
GLIPR1	1.67		x	Chen
LTBP1	1.65		x	Chen
TCIRG1	1.65	x	x	Chen
SMAD7	1.63	x		Chen
SPP1	1.63		x	Chen/Sharov/Kim
ZDHHC14	1.61	x		Boyer/Kim
MYH9	1.54	x	x	Chen
LRRC8B	1.53	x		Chen

Table 5.5.4 Comparison of the Sox2 repressed genes with previous studies

The expression of genes that been repressed by Sox2 more than 1.5 times were compared with published data of expression arrays and ChIP on chip analysis (Masui et al., 2007; Greber et al., 2007; Chen et al., 2008; Kim et al., 2008; Boyer et al., 2005; Sharov et al., 2008).

5.6 QRT-PCR validation of Affymetrix data

In order to validate the results of the Affymetrix analysis, QRT-PCR was carried out to detect the expression level of the possible Sox2/Groucho regulated genes. The primers used in QRT-PCR were designed as near to the Affymetrix probe as possible to minimize the difference between QRT-PCR and Affymetrix expression array data due to their text difference. To avoid non-specific recognition and secondary structure of primers, the best primer pair for each gene was chosen from the suggestion of the primer-BLAST website (**http://www.ncbi.nlm.nih.gov/tools/primer-blast/**; NCBI).

In order to confirm the repression of Sox2 on these genes, the same transfection and RNA purification procedure was repeated on hNSCs (Figure 5.4.1). All the QRT-PCR results were normalized by the expression of the housekeeping gene, Actin. In Affymetrix expression array data, the expression level of these possible Sox2/Groucho regulated genes was decreased following overexpression of Sox2 but little or no decrease was seen in $Sox2^{LQY/AAA}$ overexpressing hNSCs (Figure 5.5.1). The results of QPT-PCR showed a similar trend except in the data of TAGLN which showed similar expression level in GFP, GFP/Sox2 and GFP/ $Sox2^{LQY/AAA}$ overexpressing hNSCs (Figure 5.6.1).

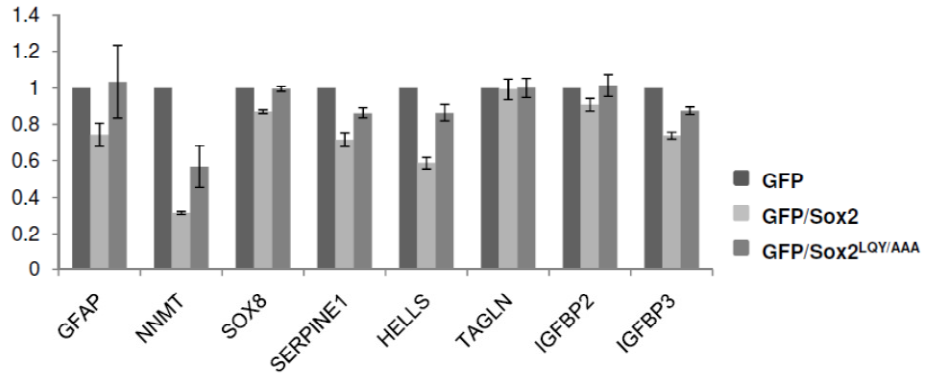
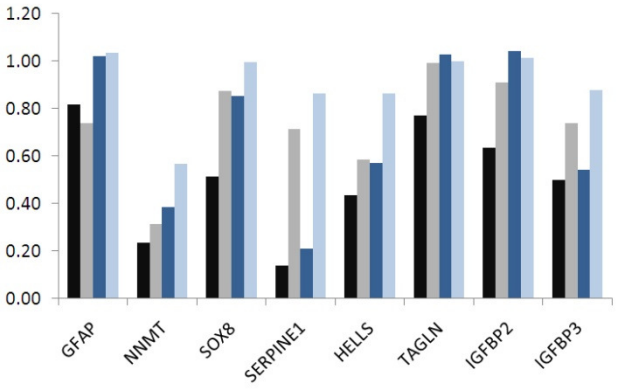


Figure 5.6.1 QRT-PCR result of possible target genes

The relative values of possible Sox2/Groucho regulated genes are represented. Triplicate QRT-PCR was carried out and the result was normalized by the housekeeping gene actin.

The QRT-PCR result showed the relative expression level of GFAP and NNMT in GFP, GFP/Sox2 and GFP/ $Sox2^{LQY/AAA}$ overexpresing hNSCs was very similar with the result of Affymetrix expression array. This implies that Sox2 represses the expression of GFAP and NNMT in hNSCs. The repression of SOX8, SERPINE1, HELLS, IGFPBP2 and IGFBP3 by Sox2 as detected by QRT-PCR was not as pronounced as see in the Affymetrix array. However, the results still showed a decrease in Sox2 overexpressing hNSCs and $Sox2^{LQY/AAA}$ overexpressing hNSCs showed milder repression of these genes. These QRT-PCR results confirm the data from Affymetrix expression array and imply that GFAP, NNMT, SOX8, SERPINE1, HELLS, IGFBP2 and IGFBP3 are possible Sox2/Groucho regulated target genes.



	Sox2		Sox2$^{LQY/AAA}$	
	Array	QRT-PCR	Array	QRT-PCR
GFAP	0.82	0.74	1.03	1.02
NNMT	0.23	0.31	0.57	0.39
SOX8	0.51	0.87	1.00	0.85
SERPINE1	0.14	0.71	0.87	0.21
HELLS	0.43	0.59	0.87	0.57
TAGLN	0.77	0.99	1.00	1.03
IGFBP2	0.63	0.91	1.01	1.04
IGFBP3	0.50	0.74	0.88	0.54

Figure 5.6.2 Comparison of Affymetrix data with QRT-PCR results of possible Sox2/Groucho regulated genes.

The relative values of Affymetrix data and QRT-PCR results of possible Sox2/Groucho regulated genes are showed as bar graph. The expression level of genes in GFP expressing cells was set as 1 which does not shown in the graph. The black and grey bars show the value from cells overexpressing wild type Sox2 and the dark blue and light blue bars show the value from cells overexpressing Sox2$^{LQY/AAA}$. Actual numbers of the relative value are listed in the table on the right.

5.7 Discussion

In this chapter, Sox2 regulated genes in hNSCs were identified using Affymetrix expression arrays. Although the Affymetrix signal of probes of Sox2 expression did not show a consistent increase in the level of Sox2 in overexpressing cells (Figure 5.3.2). QRT-PCR validation of the same RNA samples showed the expression of Sox2 in cells overexpressing GFP+Sox2 and GFP+Sox2$^{LQY/AAA}$ was consistently higher than cells overexpressing GFP (Figure 5.4.4). This inconsistency might be because of the 3'-end bias of the poly-dT primer used to generate the probes for the arrays. The validation by QRT-PCR showed a more than 5 times increase in Sox2 expression both the wild type Sox2 and Sox2$^{LQY/AAA}$ transfected cells. The overexpression of Sox2 or Sox2 $^{LQY/AAA}$ affected a similar number of genes (Table 5.5.1). The expression level of more than 80% of genes is similar in GFP, GFP/Sox2 and GFP/Sox2$^{LQY/AAA}$ overexpressing hNSCs. This implies that the overexpression did not have a general non-specific effect since these numbers are similar to the numbers seen in Sox2 knock-down experiments reference. Only 1% of genes that were activated in cells overexpressing wild type Sox2 should relatively lower expression in cells overexpressing Sox2$^{LQY/AAA}$. This implies that the Sox2$^{LQY/AAA}$ mutant retains its activation function. Among the Sox2-repressed genes, 7% showed lower repression by Sox2$^{LQY/AAA}$. This indicates that a part of the Sox2 repressor function is mediated through Groucho interaction. The decreases in gene expression might also be caused by indirect regulation by Sox2. Sox2 might regulate other transcription factors and affect the expression of down-stream genes indirectly. On the other hand, Groucho might not be the only co-repressor for Sox2 repression function. Previous studies had shown

that Sox2 can interact with transcriptional regulatory proteins such as mSin3A, HDAC1 and HDAC2 (Baltus et al., 2009a). Therefore, it is very likely that Sox2 functions as a transcriptional repressor through interaction with various co-repressors.

According to the ch-IP study of Boyer et al., (2005), SOX2, OCT4 and NANOG occupied 1303 actively transcribed genes and 957 inactive genes in human embryonic stem cells. Among the inactive genes a large part of them were transcription factors. In contrast to these data, our data indicates that Sox2 mainly activates transcription factors in human neural stem cells (Table 5.5.2). This implies a different role for Sox2 during embryonic stem cell differentiation. Interestingly, within the Sox2 regulated transcription factors, MEIS1 and PAX6 are inactivated in embryonic stem cells but activated in neural stem cells. Previous studies showed that both MEIS1 and PAX6 are important regulator of eye development (Zhang et al., 2003b). Meis1 was showed to directly regulate Pax6 in chicken (Zhang et al., 2002) and Pax6 and Sox2 function as partners to regulate lens development in chicken (Kamachi et al., 2001). Therefore, the switching of Sox2 transcriptional regulation role during embryonic stem cell is a necessary procedure for specific cell type differentiation.

Among the Sox2 repressed genes, 14% of genes are related to cell cycle (Table 5.5.2). Some of the Sox2 repressed genes also showed increased expression in Sox2 knock-out mouse embryonic stem cells (Masui et al., 2007). During embryogenesis, embryonic stem cells undergo rapid cell division in order to increase cell number for further differentiation (White and Dalton, 2005).

According to our data, the cell division rate might be slowed down through Sox2 in hNSCs.

GFAP has been shown to be directly repressed by Sox2 (Cavallaro et al., 2008). In chapter 4, luciferase assay data implied that repression of GFAP might be through a Sox2-Groucho interaction (Figure 4.4.6). Affymetrix expression array data and QRT-PCR validation show that Sox2 repressed GFAP expression whereas $Sox2^{LQY/AAA}$ did not affect the GFAP expression level (Figure 5.5.1; Figure 5.6.1). This confirms that Sox2 might regulate GFAP expression through a Groucho interaction. According to my Affymetrix expression array data and QRT-PCR results, I have identified 7 strong candidates that might also be down-regulated by Sox2 through Groucho interaction; namely NNMT, SOX8, SERPINE1, HELLS, IGFBP2 and IGFBP3. NNMT (nicotinamide N-methyltransferase) catalyzes the N-methylation of nictotinamide and other pyridines in liver (Aksoy et al., 1994). SERPINE1 (SERPIN peptidase inhibitor, clade E, member 1) also name as PAI1 (plasminogen activator inhibitor 1) inhibits plasminogen activation and negatively regulates fibrinolysis (Ginsburg et al., 1086; MEHTA and SHAPIRO, 2008). HELLS (helicase, lymphoid-specific) is involved in DNA strand separation during replication and transcription (Jarvis et al., 1996). IGFBPs (insulin-like growth factor-binding protein) bind with IGFs and affects cell growth and cancer cell differentiation. SOX8 belong to SOX E subgroup which is involved in neural crest development (Bell et al., 2000; Cheng et al., 2001), oligodengrocyte specification and differentiation (Stolt et al., 2005). All of these genes are repressed in Sox2 overexpressing hNSCs but not in

$Sox2^{LQY/AAA}$ overexpressed hNSCs. This result implies that Sox2 keeps genes silenced and this silencing requires Groucho interaction domain.

In summary, the data from expression arrays revealed that Sox2 down-regulates more genes then it up-regulates in human neural stem cells. The main group of genes that are activated by Sox2 overexpression is transcription factors, while it represses genes related to cell cycle. When the Sox2-Groucho interaction was lost the Sox2 repression of a group of genes was decreased. This suggests the importance of the Sox2-Goucho interaction in Sox2 transcriptional repression function. However, within the Sox2 repressed genes, 91% of them are not affected by the Groucho interaction mutant, $Sox2^{LQY/AAA}$. This implies that other repression mechanism might involve in Sox2 repression function, such as the interaction with HDACs, CtBP or other co-repressors. Also, it is possible that some of the gene repression resulted from indirect effects of Sox2 overexpression.

Chapter 6 Role of Sox2/Groucho interaction *in vivo*

6.1 Introduction

SoxB1 family members are the earliest known transcription factors expressed in response to neural induction (Rex et al., 1997). In Zebrafish, all members of the SoxB1 family are expressed before 24 hours post fertilized (hpf). Although the SoxB1 family used to be classified as transcriptional activator (Uchikawa et al., 1999), recent evidence indicates that the SoxB1 family plays a different role during early embryonic development. Sox3, one of the SoxB1 family members, has been shown to act as a transcriptional repressor in early *Xenopus* and Zebrafish embryos (Shih et al., 2010; Zhang et al., 2003a). Previous studies in our lab, injecting *sox3* mRNA at the one cell stage repressed organizer markers such as *squint*, *goosecoid* and *cordin* (Figure 6.1.1). However, the effect of *sox2* at these early stages of development was not studied. In this chapter, *sox2* was injected into one cell stage zebrafish embryos and the expression of organizer markers was analysed in order to provide an *in vivo* assay of its repressor functions.

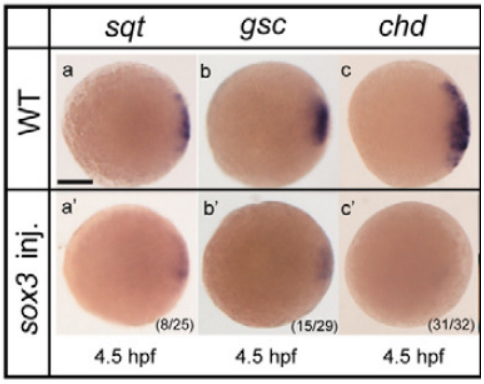


Figure 6.1.1 Sox3 represses markers of the Zebrafish organizer.

sox3 injected zebrafish embryos show dramatic reduction in the expression of the organizer markers, *squint* (*sqt*), *goosecoid* (*gsc*) and *chordin* (*chd*). (Shih et al., 2010)

6.2 Overexpression of *sox2* affects early zebrafish embryo development

In order to test the function of *sox2* in early zebrafish embryos, overexpression of zebrafish *sox2* in early zebrafish embryos was carried out. To do this, the first step was to create an expression vector which allowed the zebrafish *sox2* to be expressed in zebrafish embryos and allowed the expressed protein to be easily detected. Therefore, HA tagged zebrafish *sox2* and *sox2*$^{LQY/AAA}$ were constructed (Figure 6.2.1). Zebrafish *sox2* was amplified using pCS2-*sox2* plasmid as template with primers which contained XbaI and XhoI restriction sites (Table 6.2.1). The fragment was then inserted in pBUT2-HA vector which resulted in an HA tag being attached at the 3'end of the zebrafish *sox2* (Figure 6.2.1). The pBUT-zf*sox2*-HA plasmid was then used as template to create the zebrafish *sox2*$^{LQY/AAA}$ construct using site directed mutagenesis (Table 6.2.1).

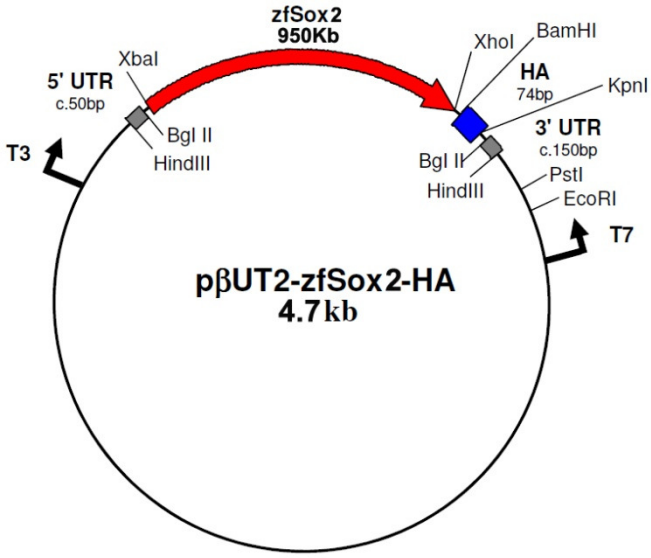


Figure 6.2.1 Zebrafish *sox2* and *sox2*$^{LQY/AAA}$ mutant expression vector map

Zebrafish *sox2* was cloned from pCS2-zf*sox2* and inserted into the pBUT2 expression vector with an HA tag at the 3' end. The *sox2*$^{LQY/AAA}$ mutant was generated using site direct mutagenesis kit using pBUT2-zf*sox2*-HA as the template.

Primer name	Primer sequence
zfSox2 HA F XbaI	TCTAGAatgtataacatgatggaa
zfSox2 HA R XhoI	CTCGAGacatatgcgataagggaat
zfSox2 LQY/AAA mut F	CGCTACGACATGAGCGCG**CTCCAGTA**CAACTCCATGACCAACTC
zfSox2 LQY/AAA mut R	GAGTTGGTCATGGAGTTG**TACTGGAG**CGCGCTCATGTCGTAGCG

Table 6.2.1 Primers used in creating zebrafish *sox2* and *sox2*$^{LQY/AAA}$ mutant expression vectors.

The primers used in constructing zebrafish *sox2* and *sox2*$^{LQY/AAA}$ mutant expression vectors are listed in the table. The capitalized letters in zf*sox2* HA F *XbaI* and zf*sox2* HA R *XhoI* mark the restriction sites. The bold letters in zf*sox2*$^{LQY/AAA}$ mut F and zf*sox2*$^{LQY/AAA}$ mut R mark the nucleotides that differ from wild type *sox2*.

The DNA sequence of pBUT2-zf*sox2*-HA and pBUT2-zf*sox2*$^{LQY/AAA}$-HA constructs was confirmed by GeneService DNA sequencing. *In vitro* mRNA transcription was carried out to generate mRNA of zebrafish *sox2*-HA and *sox2*$^{LQY/AAA}$–HA using T3 promoter oligonucletide as primer and EcoRI digested plasmid as template. The mRNA was injected into 1-2 cell stage zebrafish embryos and 4.5 hours after fertilization, the embryos were collected and lysed in deyolk buffer. The expression of zf*sox2*-HA and zf*sox2*$^{LQY/AAA}$-HA was detected by western blot using anti-HA antibody (Figure 6.2.2). Both zf*sox2*-HA and zf*sox2*$^{LQY/AAA}$-HA were clearly expressed in these injected zebrafish embryos.

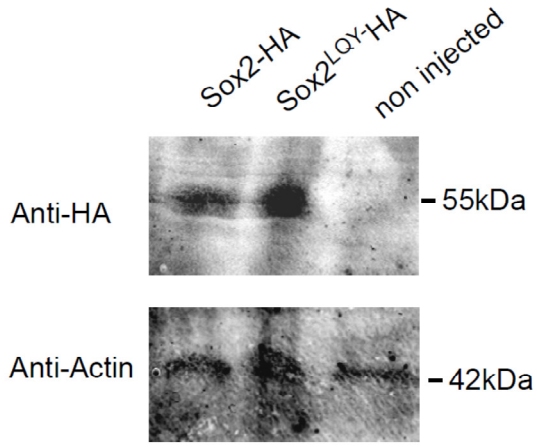


Figure 6.2.2 Zebrafish *sox2* and *sox2*$^{LQY/AAA}$ mutant expression in early embryos

The mRNA of zebrafish *sox2* or *sox2*$^{LQY/AAA}$ mutant was injected into 1-2 cell stage zebrafish embryos. 4.5 hours after fertilization, the embryos were collected and analysed using western blotting. Anti-HA antibody was used to detect the expression of *sox2*-HA or *sox2*$^{LQY/AAA}$-HA. Anti-Actin antibody was used to detect the endogenous actin as loading control.

6.3 Repressor function of *sox2* in early zebrafish embryo development

In the previous studies in our lab, overexpressing *sox3* at 1-2 cell stage resulted in headless embryos at 24 hpf (Figure 6.3.1; experiments performed by Cheng Liang, Kuo). When *sox3* mRNA was local-injected into ectoderm precursor cells at the 16-32 cell stage, duplications of the central nervous system were seen at 24 hpf (Figure 6.3.2) (Dee et al., 2008). The interpretation of these phenomena is that *sox3* repressed the organiser but at later stages induced formation of neural ectoderm. Due to the high similarity of *sox2* and sox3 (64% identity in 326 amino acid overlap; Figure 6.3.3), it was considered that *sox2* might also function as a repressor in early zebrafish embryos.

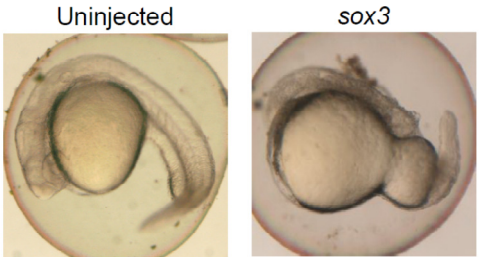


Figure 6.3.1 sox3 overexpression causes headless embryos.

Sox3 mRNA was injected in 1-2 cell stage zebrafish embryos. The picture was taken at 24 hpf. Pictures were taken by Cheng Liang, Kuo.

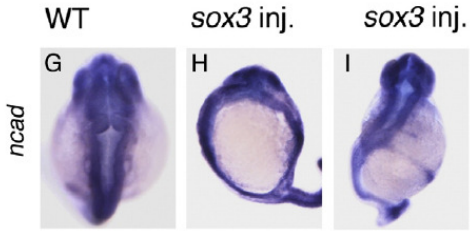


Figure 6.3.2 sox3 overexpression causes duplication of the central nervous system.

sox3 mRNA was injected into ectoderm precursors at the 16-32 cell stage. At 24 hpf, the embryos were collected and *in situ* hybridization was carried out to detect the neural marker *ncad*. The uninjected embryo is marked as WT in (G). The central nervous system duplication is shown in head (H) and tail (I) when overexpressing *sox3* (Dee et al., 2008).

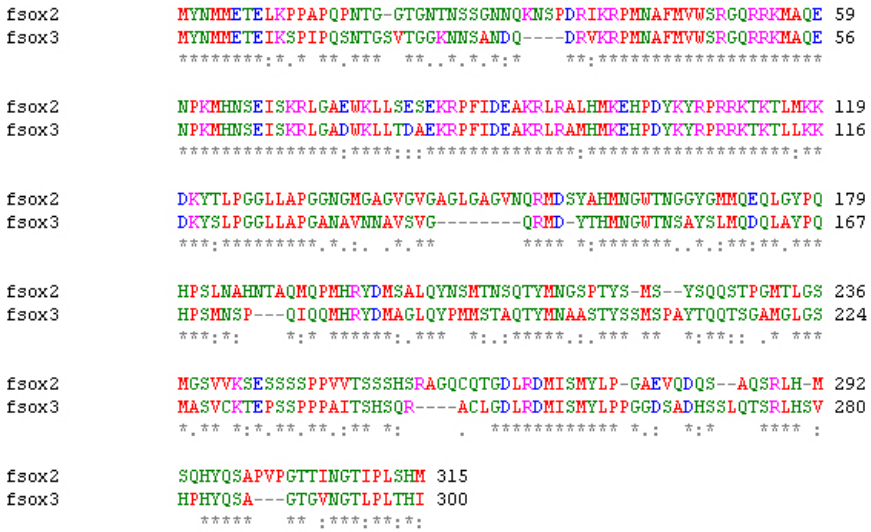


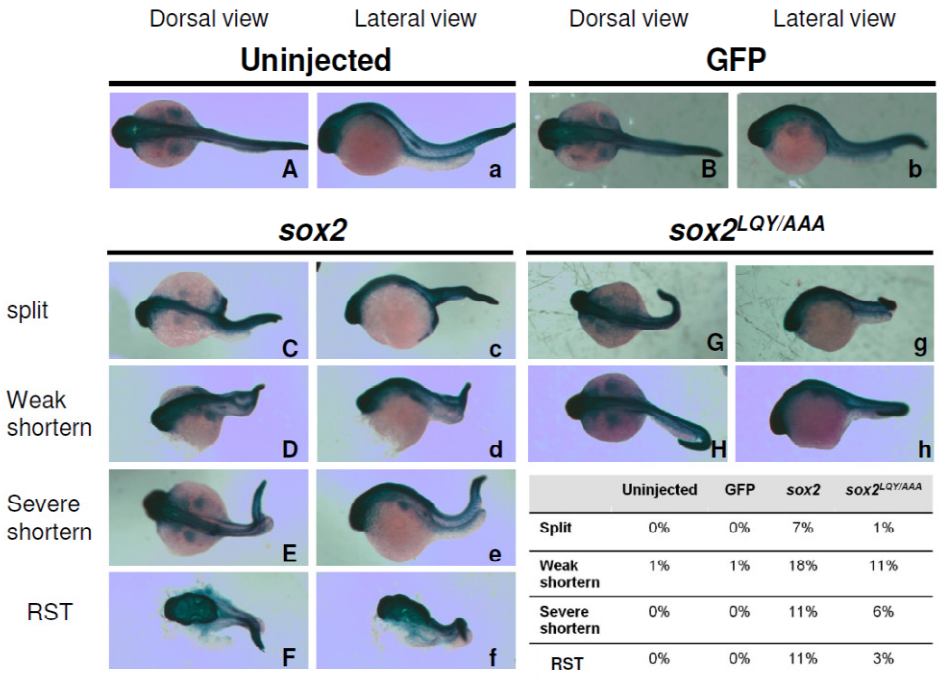
Figure 6.3.3 Alignment of zebrafish sox2 and sox3.

Amino acid sequence of zebrafish sox2 and sox3 alignment is shown. Analysis was carried out using ClustalW2 (**http://www.ebi.ac.uk/Tools/msa/clustalw2/**).

In order to test the function of the Sox2/Groucho interaction in early zebrafish embryo development, 300ng mRNA of zebrafish *sox2* or *sox2*$^{LQY/AAA}$ was injected into zebrafish embryos at the 1-2 cell stage and the neural marker *ncad* was detected using *in situ* hybridization at 24 hpf (Figure 6.3.4). The *ncad in situ* hybridization of uninjected embryos showed a normal central nervous tissue distribution (Figure 6.3.4 A, a). GFP injected embryos were used as a negative control and they showed no obvious difference in *ncad* expression as compared to uninjected embryos (Figure 6.3.4 B, b). However, *sox2* injected embryos showed various levels of abnormal neural development (47 out of 98 embryos; Figure 6.3.4 C-f). 7% of *sox2* overexpressing embryos showed duplication in central nervous system (Figure 6.3.4 C, c). 29% of *sox2* overexpressing embryos showed either weak or severe shortening of the anterior/posterior axis (Figure 6.3.4 D-e). 11% of the *sox2* overexpressing embryos showed reduced staining in the trunk (Figure 6.3.4 F, f). This implies

that overexpression of *sox2* could expand neuronal tissue and duplicate the central nervous system in the same way as was seen with *sox3*.

On the other hand, the overexpression of $sox2^{LQY/AAA}$ resulted in less abnormality (19 out of 89 embryos). Only 1% of the $sox2^{LQY/AAA}$ overexpressing embryos exhibited duplication of the central nervous system and 20% showed shortening of body length. This results implies that the loss of Groucho interaction reduces the effects of *sox2* on zebrafish central nervous development.



	Uninjected	GFP	*sox2*	$sox2^{LQY/AAA}$
Split	0%	0%	7%	1%
Weak shortern	1%	1%	18%	11%
Severe shortern	0%	0%	11%	6%
RST	0%	0%	11%	3%

Figure 6.3.4 *sox2* and $sox2^{LQY/AAA}$ affect embryonic development.

300ng mRNA of zebrafish *sox2* or $sox2^{LQY/AAA}$ was injected into zebrafish embryos at the 1-2 cell stage. At 24 hpf, the neural marker *ncad* was detected using *in situ* hybridization. The percentage of different appearance of embryos is listed in the table. RST: reduce signal in trunk. (32 embryos were uninjected, 29 GFP injected, 98 *sox2* injected and 89 $sox2^{LQY/AAA}$ injected.)

In order to test the effect of *sox2* on organizer formation, 50ng instead of 300ng of *sox2* or *sox2*$^{LQY/AAA}$ was injected into zebrafish embryos at the 1-2 cell stage. Previous experiments detected *ncad* expression in the injected embryos at 24 hpf which require higher *sox2* overexpression, presumably to compensate for the degradation of *sox2* protein during the longer developmental time. The organizer markers *goosecoid*, and *chordin* were detected at 4.5 hpf using *in situ* hybridization. Uninjected wild type embryos and GFP injected embryos were used as negative controls. Also, embryos overexpressing *sox3* were used as a positive control and to determine any functional differences to *sox2*. *Goosecoid* could be detected in the organizer by *in situ* hybridization at 4.5 hpf. In uninjected and GFP overexpressing embryos, the expression of *goosecoid* was located at the dorsal margin where the organizer is found (Figure 6.3.5 A, B). In contrast, *goosecoid* expression was hardly detected in most *sox3* overexpressing embryos (Figure 6.3.5 C). This result agrees with the previous studies in our lab (Shih et al., 2010). When *sox2* was overexpressed, a very different result was seen; a substantial expansion of *goosecoid* expression was observed (Figure 6.3.5.D). The expression of *goosecoid* was detected not only at the organizer but also over half of the animal pole in *sox2* overexpressing embryos. These results imply that *sox2* function is the opposite of *sox3* with respect to the effects on *goosecoid* expression in early zebrafish embryo. Why this might be so, is discussed below. When overexpressing *sox2*$^{LQY/AAA}$, the *goosecoid* expression pattern was similar to uninjected and GFP injected embryos (Figure 6.3.5 E). The loss of the ability to activate *goosecoid* expression by the *sox2*$^{LQY/AAA}$ mutant, suggests two possible models. First, the Sox2/Groucho interaction represses the transcriptional activator of *goosecoid*

expression. Alternatively, *sox2* might directly activate *goosecoid* expression and the $sox2^{LQY/AAA}$ mutation is causing the loss of an activation function of *sox2*. However, 20% of the $sox2^{LQY/AAA}$ overexpressing embryos showed similar *goosecoid* expression pattern to that seen in *sox2* overexpressing embryos. This suggests that the mutant form has not lost the ability to affect *goosecoid* expression seen for wild-type *sox2*, but that this ability is dramatically reduced.

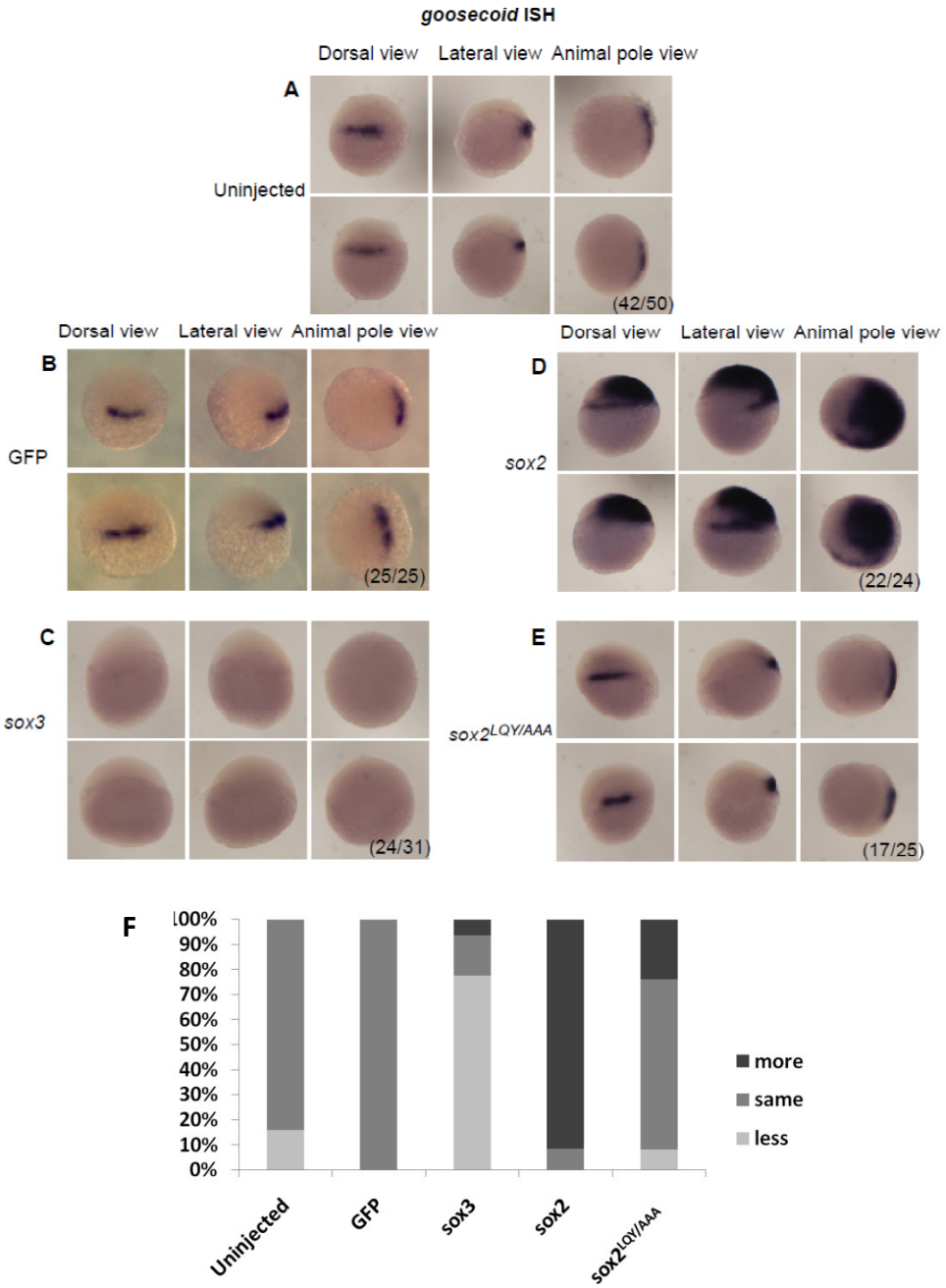


Figure 6.3.5 *goosecoid in situ* hybridization of *sox2* or *sox2*$^{LQY/AAA}$ overexpressing zebrafish embryos.

50ng of mRNA of GFP, *sox3, sox2* and *sox2*$^{LQY/AAA}$ was injected into zebrafish embryos at 1-2 cell stage. Embryos were collected at 4.5 hpf and *in situ* hybridization was carried out using a *goosecoid* probe. The numbers at the bottom-right show how many embryos exhibit expression patterns similar to that shown in the pictures (number of embryos have similar pattern/ number of total embryos). The expression of *goosecoid* was compared with uninjected embryos (as shown in A) and classified into three expression levels (more, same and less). The percentage of each expression level in different mRNA injected embryos is showed in F.

Another organizer marker, *chordin*, was also shown to be a target of *sox3* repressor function (Shih et al., 2010). The normal *chordin* expression pattern was detected by *in situ* hybridization in uninjected embryos and GFP injected embryos (Figure 6.3.6 A, B). The expression pattern of *chordin* was again located at the dorsal side of the margin in the region of the organizer. In the *sox3* overexpressing embryos, most showed very weak or no *chordin* expression at the dorsal margin (Figure 6.3.6 C). The definition of the strength of gene expression is dependent on the colour that was seen in the majority of embryos. The probes, colour developing reagents and colour developing time of *in situ* hybridization were all the same in the same batch of embryos. With these same conditions, the comparison of gene expression level could be distinguished by its colour stain after *in situ* hybridization. This result agrees with the previous studies in our lab that *chordin* was repressed when overexpressing *sox3* in early zebrafish embryos (Shih et al., 2010). When overexpressing *sox2* in embryos, the expression of *chordin* was only slightly weaker than in uninjected or GFP injected embryos (Figure 6.3.6). This implies that *sox2* might also down regulate *chordin* expression, but the repressor function of *sox2* on *chordin* is weaker than *sox3*. The *chordin* expression pattern in $sox2^{LQY/AAA}$ injected embryos was similar to uninjected or GFP injected embryos indicating that this weak repression of *chordin* is also Groucho dependent.

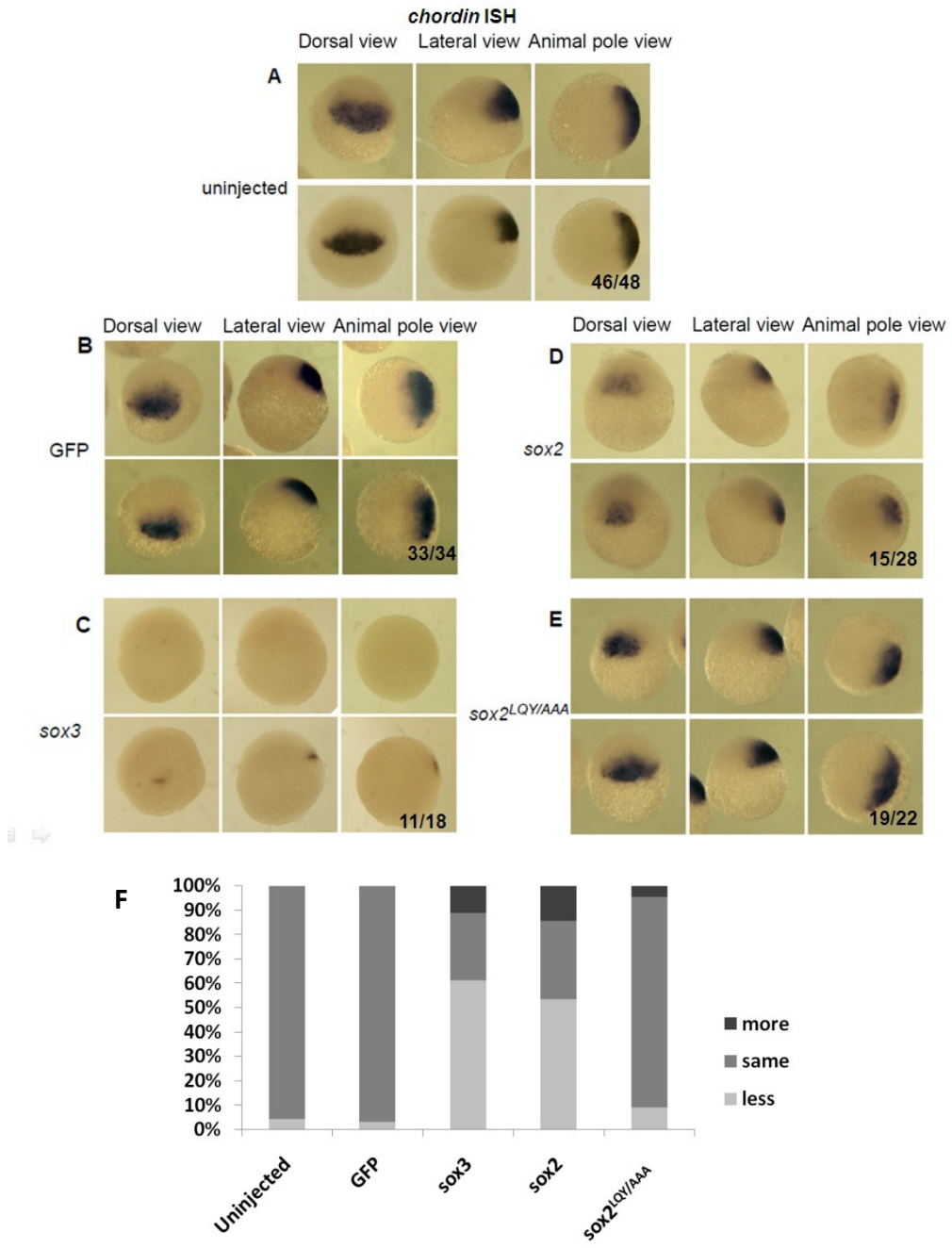


Figure 6.3.6 *chordin in situ* hybridization of *sox2* or $sox2^{LQY/AAA}$ overexpressing zebrafish embryos.

50ng of mRNA of GFP, *sox3, sox2* and $sox2^{LQY/AAA}$ was injected into zebrafish embryos at 1-2 cell stage respectively. Embryos were collected at 4.5 hpf and *in situ* hybridization was carried out using *chordin* probes. The number marked at the bottom-right shows how many embryos have similar expression pattern to that shown in the pictures (number of embryos have similar pattern/ number of total embryos). The expression of *chordin* was compared with uninjected embryos (as shown in A) and classified into three expression levels (more, same and less). The percentage of each expression level in different mRNA injected embryos is showed in F.

According to the previous studies in our lab, another gene that is regulated by *sox3* is *sox19b*. *sox19b*, also known as *sox31*, has only been identified in zebrafish (Girard et al., 2001). *sox19b* was categorized in the SoxB1 family due to the high conservation of its HMG domain and its C-terminal amino acid sequence. It has recently been shown, in early zebrafish embryos, that overexpression of *sox19b* results in impaired organizer formation (Hu et al., 2011). The normal expression of *sox19b* was detected in uninjected and GFP injected embryos by *in situ* hybridization (Figure 6.3.7 A, B). *sox19b* was distributed evenly in the animal cap of the whole embryo. In *sox3* overexpressing embryos, the area of *sox19b* expression was increased into a large patch on the embryo and the expression was stronger than the uninjected embryos (Figure 6.3.7 C). This result agrees with the previous studies in our lab that *sox19b* was expanded when overexpressing *sox3*. *sox2* overexpressing embryos showed a similar expression pattern to *sox3* overexpressing embryos (Figure 6.3.7 D). This indicates that *sox2* might function similarly with *sox3* in the regulation of *sox19b* expression. However, the Groucho interaction mutant, $sox2^{LQY/AAA}$, had no effect on the *sox19b* expression pattern whereas the wild-type *sox2* seems to activate the expression of *sox19b* (Fiugre 6.3.7 E). This implies that *sox2* might activate *sox19b* indirectly through its repressor function.

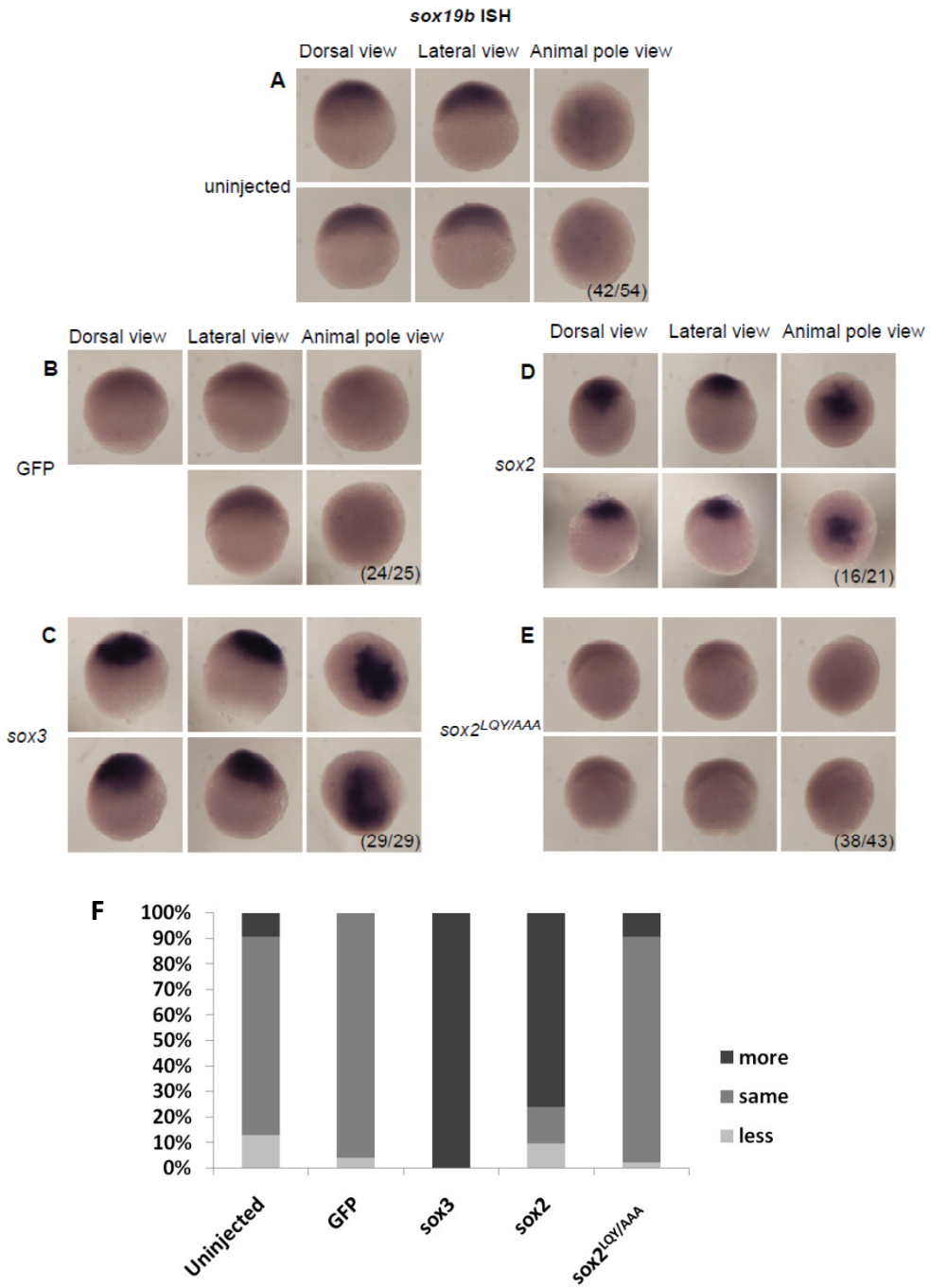


Figure 6.3.7 *sox19b in situ* hybridization of *sox2* or *sox2*$^{LQY/AAA}$ overexpressing zebrafish embryos.

50ng of mRNA of GFP, *sox3, sox2* and *sox2*$^{LQY/AAA}$ was injected into zebrafish embryos at 1-2 cell stage respectively. Embryos were collected at 4.5 hpf and *in situ* hybridization was carried out using *sox19b* probes. The number marked at the bottom-right shows how many embryos have similar expression pattern to that shown in the pictures (number of embryos have similar pattern/ number of total embryos). The expression of *sox19b* was compared with uninjected embryos (as shown in A) and classified into three expression levels (more, same and less). The percentage of each expression level in different mRNA injected embryos is showed in F.

6.4 Discussion

In this chapter, RNA encoding the *sox2* or *sox2*$^{LQY/AAA}$ mutant was injected into zebrafish embryos at 1-2 cell stage which allows the injected mRNA to be translated in the whole embryo. Therefore, the effects of overexpressing these proteins in the earliest stages of embryo development could be observed. However, in normal zebrafish embryo development, the expression time point of *sox2* is not as early as the 1-2 cell stage. The expression of *sox2* is slightly later than *sox3* in early zebrafish embryos (Okuda et al., 2006). Reverse transcriptase-polymerase chain reaction (RT-PCR) shows that *sox3* expression could be detected at the 32 cell stage whereas *sox2* expression could not be detected until 30% epiboly stage (Figure 6.4.1). In other words, there is no maternal/zygotic *sox2* expression at the 1-2 cell stage. At this stage of zebrafish development the overexpressed *sox2* might function on *soxb1* target genes or the genes that will be targeted by *sox2* at a later stage. Therefore, the effects of overexpressing *sox2* do not reflect its normal function in the zebrafish embryo development. However, this assay provides a way to distinguish any change of the function between the wild-type *sox2* and *sox2*$^{LQY/AAA}$. The results in this chapter show that *sox2*$^{LQY/AAA}$ acts differently to wild-type *sox2,* which suggests that the loss of Groucho interaction indeed alters the function of *sox2*.

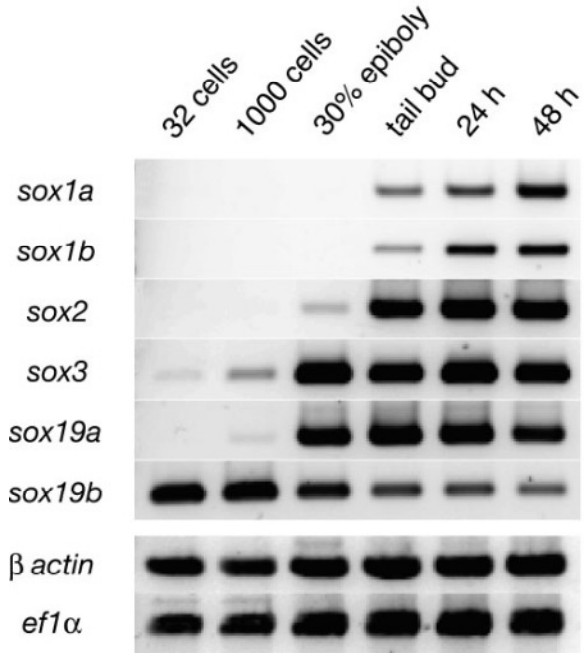


Figure 6.4.1 SoxB1 family member expression time point in early zebrafish embryos.

RT-PCR shows the temporal expression profiles of SoxB1 family members. The β actin and ef1α were used as controls (Okuda et al., 2006). Note that endogenous sox2 appears (30% epiboly) later than sox3 (32 cells) in early zebrafish embryos.

When looking at the general effects of *sox2* on zebrafish embryo development, wild-type *sox2* overexpressing embryos showed axis duplication and reduced *ncad* expression in the trunk, which was consistent with the effects of *sox3* overexpression (Figure 6.3.2; Figure 6.3.4 C-f). One effect of *sox3* overexpression is to restrict the formation of mesoderm through its repressor function, therefore causing the expansion of the central nervous system (Shih et al., 2010). This interpretation could also apply to the axis duplication of *sox2* overexpressing embryos. However, the $sox2^{LQY/AAA}$ mutant overexpressing embryos showed a less severe effect on the central nervous system. Since the LQY sequence is essential to the activities of wild-type *sox2*, it is likely that the same interaction is important for the function of endogenous *sox2*.

In situ hybridization analysis showed that overexpression of *sox2* affected the expression of *goosecoid*, *chordin* and *sox19b*. Wild-type *sox2* functioned

similarly to *sox3* in activating *sox19b* and repressing *chordin*, but leaded an opposite effect on *goosecoid* expression (Figure 6.3.5-7). Previous studies in our lab demonstrated that *sox3* can directly repress the expression of *chordin* and *goosecoid.* Similar effects of *sox2* and *sox3* on repressing *chordin* indicates that *sox2* might also repress *chordin* directly. The opposite effect of *sox2* on regulating *goosecoid* with *sox3* implies an indirect function of *sox2* on *goosecoid*. For *sox19b*, both *sox2* and *sox3* activate its expression which might be caused by the indirect effect of *sox2* and *sox3* transcriptional repression function. Nonetheless, $sox2^{LQY/AAA}$ mutant overexpressing embryos showed no obvious difference in the expression of any these genes compared to the unjected or GFP injected embryos. This indicates that the effect of overexpressed *sox2* on these genes depends on the LQY sequence which implies that the Sox2/Groucho interaction is required for *sox2* function in regulating these genes.

Since both *sox2* and *sox3* belong to the *soxb1* subgroup, they might have the same effect on regulating genes. In this chapter, the similar effect of *sox2* and *sox3* on repressing *chordin* and activating *sox19b* fits this hypothesis. However, an opposite effect of *sox2* and *sox3* was also observed in regulating *goosecoid*. This could be due to the protein structural and functional differences between *sox2* and *sox3*. Their highly similar HMG DNA-binding domain might lead them to bind to similar binding sites, but difference in other parts of the *sox2* or *sox3* proteins might result in different biological functions. According to my data, *sox2* and *sox3* could affect the same gene but they do not necessarily have the same effect on the gene.

The experiments in this chapter are primarily pilot experiments for further research. To confirm the primary observations, more experiments would be required. For example, the effect of specific gene overexpression on organizer marker expression was determined by the colour strength of *in situ* hybridization. Although all the probes and colour developing conditions were the same in every experiment, it is still not a truly quantitative and precise evaluation. QRT-PCR is a better method which might be used in the future to measure the level of gene expression. In addition, *in situ* hybridization does not provide evidence that any effects are direct or indirect. Further experiments for this purpose could include chromatin-immunoprecipitation (Chip) or electrophoretic mobility shift assay (EMSA) to confirm the binding of the transcription factor to the promoter region of target genes.

Data in this chapter show that the $sox2^{LQY/AAA}$ mutant acts very differently from the wild-type *sox2*, which strongly suggests that the Sox2/Groucho interaction is needed for the function of *sox2* in embryonic development (Figure 6.3.5-7). However, the effect of losing Groucho interaction might be indirect. Previous studies of *sox3* used fusion of the repression domain of engrailed or the activation domain of VP16 with *sox3* to mimic the transcriptional repression or activation function of *sox3* respectively (Shih et al., 2010; Zhang et al., 2004). Through this strategy, the direct function of transcriptional repression or activation function of *sox3* could be elucidated.

Since the lack of Groucho interaction of the $sox2^{LQY/AAA}$ mutant results in a different function from wild-type *sox2*, it would be very interesting to test if restoring the Groucho interaction, by attaching Groucho or a Groucho repression domain, could recover the repressor function. On the other hand,

replacing the LQY domain with a canonical Groucho binding domain (WRPW or FRPW) to enhance Sox2-Groucho interaction might result in a stronger Sox2 repression effect. This could be a good way to determine if my results are due to a direct effects of Sox2/Groucho interaction in target genes rather than an artificial change in the secondary structure of the Sox2 protein.

To sum up, in this chapter, the loss of Sox2/Groucho interaction mutant *sox2*$^{LQY/AAA}$ was showed to act distinctly from the wild-type sox2 in regulating *goosecoid*, *chordin* and *sox19b* expression. This implies that the Sox2/Groucho interaction might be crucial for the regular function of *sox2*. However, further research is needed to validate the results and confirm the direct effects of the Sox2/Groucho interaction.

Chapter 7 Discussion

In this thesis I have shown that Sox2 interacts with the transcriptional co-repressor, Groucho. By mapping, the motif by which Sox2 interacts with Groucho was defined. Disruption of the Sox2/Groucho interaction was analysed by changing only three adjacent amino acids, LQY. I have also shown that Sox2 acts as a transcriptional repressor in human neural stem cells, repressing almost as many genes as it activates. Mutation of the Sox2/Groucho interaction motif disrupted a small percentage of those repressive events. I also demonstrate that, in zebrafish, mutation of the Sox2/Groucho interaction motif has a dramatic impact on Sox2 function.

In this chapter, some of questions raised from this thesis will be discussed. The functional domains of SoxB1 proteins will be reviewed followed by a consideration of the regulation of the Sox2 bifunctional transcription activity. I also will discuss the role of Sox2 transcriptional repression in neural development.

7.1 The complexity of Sox2 function

7.1.1 Bifunctional transcription factors

Many transcription factors have roles both as activators and repressors which are therefore regarded as bifunctional transcription factors. *Krüppel* (*Kr*) is essential for the development of the thoracic and abdominal segments in early *Drosophila* embryos (Gaul and Jäckle, 1987). The N-terminus of *Kr* contains both transcriptional activation and repression domains (Licht et al., 1994). The action of *Kr* as a transcriptional activator or repressor depends on the context of its binding site (Frasch and Levine, 1987; Licht et al., 1990; Ruppert et al.,

1988). Likewise YY1, which has a similar zinc finger domain sequence to *Kr*, can repress in the absence and activate in the presence of the adenovirus protein E1A (Shi et al., 1991). Pax6, which is required for early eye determination in animals, was reported as a bifunctional transcription factor (Duncan et al., 1998). Pax6 contributes to the transcriptional activation of δ1-crystallin and repression of β-crystallin in the chicken lens. Pax6 functions as a repressor through binding to positive acting *cis* elements (Duncan et al., 1998; Richardson et al., 1995; Shinohara and Piatigorsky, 1976). Post-translational modification has been demonstrated to switch some transcriptional activators to repressors, such as Sp3 (specificity protein 3) (Valin and Gill, 2007). Sp3 represses transcription by competing with Sp1 for DNA binding (Hagen et al., 1994). However, in other cases, Sp3 functions as a transcriptional activator (Lania et al., 1997; Liang et al., 1996). Ross et al. reported that the SUMOylation of Sp3 is sufficient to convert Sp3 from a transcriptional activator to a repressor (Ross et al., 2002). Therefore, the dual function of transcriptional regulators is not a rare phenomenon.

7.1.2 Structure function of Sox2

As a bifunctional transcription factor, Sox2 might have independent regions which are responsible for either its transcriptional activator or repressor activity. A number of domains have been identified in Sox2. Next to the HMG DNA binding domain (39-126 aa) is located the group B homology motif (126-137 aa). At the C-terminus of Sox2 is a serine rich domain (205-263 aa) which includes the LEF homology domain (218-263 aa) (Figure 7.1.1). Serine rich domains have been reported to be important regulatory regions for some

transcription factors, such as v-Rel and CREB (Chen et al., 1999; Karin, 1992). The transcriptional activation region of Sox2 was first defined in the C-terminus which includes the serine rich domain (Bowles et al., 2000; Yuan et al., 1995). According to Gal4-luciferase data in chicken lens cells, the transcriptional activation domain might be located in aa 269-275.

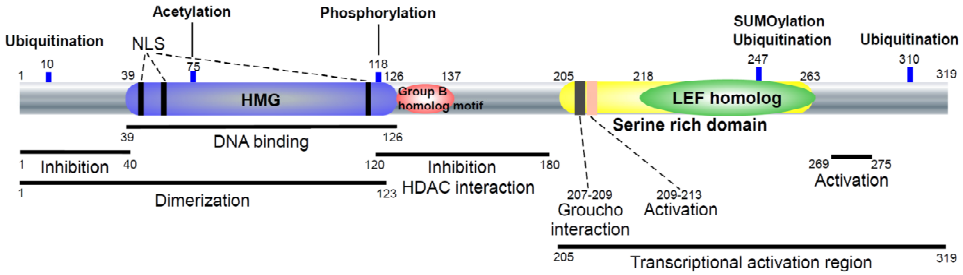


Figure 7.1.1 Sox2 protein structure and post-translational modification map.

The mouse Sox2 amino acid map of domains, functional motifs and post-translational modification sites is shown. (Baltus et al., 2009b; Ciechanover, 1998; Cox et al., 2010; Gao et al., 2009; Kamachi et al., 1999; Tsuruzoe et al., 2006; Van Hoof et al., 2009)

Also, putative internal inhibition domains might locate in the N-terminus (1-40 aa) and C-terminus (120-180 aa) (Figure 1.3.1 A, B) (Kamachi et al., 1998). As described in chapter 4, the luciferase assays have implied a possible transcriptional activation domain in the region of aa 209-213 (Figure 4.3.3). This complex map of Sox2 transcriptional activation function suggests that Sox2 not only acts as a transcriptional activator, but also regulates its transcriptional activation via an internal inhibition domain. Given the large number of genes affected by Sox2 *in vivo*, this subtle adjustment within Sox2 might be crucial to its function.

Some possible protein interaction domains in Sox2 have been identified in previous studies. The Sox2 self-associated domain was demonstrated as the N-terminus (1-123 aa). A putative HDAC1 and HDAC2 interaction region lies in

aa 120-180 (Cox et al., 2010). In chapter 3 and 4, the Groucho interaction domain was demonstrated in 207-209 aa. Therefore, the transcriptional repression function of Sox2 might be mediated via interaction with different co-repressors such as HDACs directly and Groucho. In order to further understand how Sox2 plays a role as a transcriptional repressor, we will also need to investigate the relationship of Sox2 with other co-repressors such as CtBP or chromatin remodelling factors.

7.1.3 Functional structure within SoxB1 subgroup proteins

As described in chapter 1, SoxB1 subgroup proteins have a highly conserved HMG DNA-binding domain. However, poly-alanine intervals exist in Sox1 and Sox3 but not Sox2. The similarity of the other putative functional domains was analysed by ClusterW (**http://www.ebi.ac.uk/Tools/msa/clustalw2/**) the results of which are discussed below. In the mouse SoxB1 subgroup, the similarity of the HMG DNA-binding domains is higher than 90%, which suggests that the preference of their binding sequence might be the same (Figure 7.1.2). The slightly lower similarity of the adjacent group B homology motif might contribute to the interaction with different partners and therefore regulate the different functions of the three members of the SoxB1 subgroup (Uchikawa et al., 1999). The similarity of the putative internal inhibition domain within SoxB1 subgroup proteins is approximate 40%. It has been reported that the internal inhibition domain can interfere with the transcriptional activation function of Sox1 and Sox2 (Kamachi et al., 1999). Since homology with Sox3 is similar this domain might function the same in Sox3. The similarity of the domain which was reported to be important for

interaction with co-repressor HDAC is 45-54% where the same Groucho interaction motif exists in all SoxB1 subgroup proteins suggesting that the interaction with the Groucho co-repressors might be shared by all SoxB1 subgroup proteins.

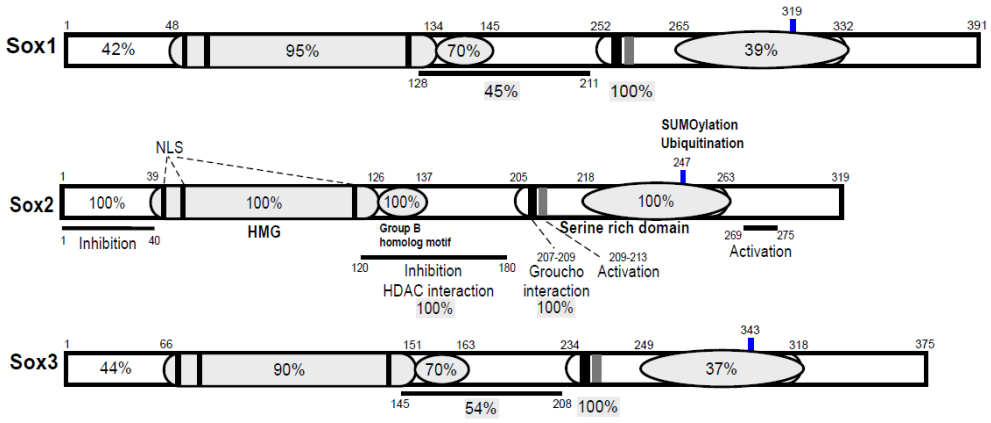


Figure 7.1.2 Comparison of organization of mouse SoxB1 subgroup proteins.

The similarity of inhibition domain, HMG domain, group B homolog motif, LEF homolog motif and Groucho interaction motif are labelled in the relative boxes. The similarity are calculated using Sox2 as 100 percent by ClustalW2.
(http://www.ebi.ac.uk/Tools/msa/clustalw2/).

If the structure and function of SoxB1 subgroup proteins are so alike, why have they all been retained during evolution? The possible explanation could be found from evolution point of view. Three possible theories have been proposed to interpret the fate of gene duplication (Figure 7.1.3) (Mazet and Shimeld, 2002). Degeneration eliminates a redundant gene; neofunctionalisation develops a new function of the duplicated gene; subfunctionalisation specifies the multi-function duplicated gene to different function. Unlike vertebrates, there is only one SoxB1 protein in *Drosophila* (*SoxNero*) and amphioxus (*AmphiSox1/2/3*) (Chao et al., 2007; Holland et al., 2000). This implies the duplication of SoxB1 genes during evolution. In vertebrates, all SoxB1 subgroup genes are not only expressed in the central

nervous system as in the invertebrate, but are also expressed in vertebrate-specific tissues, such as the epibranchial placodes (Ishii et al., 2001). This suggests that neofunctionalisation occurred during evolution from invertebrates to vertebrates (Mazet and Shimeld, 2002).

As described in chapter 1, the temporal and spatial patterns of experission of the three SoxB1 members in early embryo development is slightly different in different species. In chicken and zebrafish embryos, Sox3 is the first expressed SoxB1 gene whereas Sox2 and Sox3 are both expressed before Sox1 (Okuda et al., 2006; Wood and Episkopou, 1999). Sox1 expression could be detected in CNS, lens and urogenital ridge. Sox2 expression could be detected in inner cell mass, primitive ectoderm, trophoblast stem cells, CNS and lens. Sox3 expression could be detected in epiblast, CNS, lens and urogenital ridge (Miyagi et al., 2009). The genes regulated by the three SoxB1 subgroup proteins also vary (Kamachi et al., 2001; Nishiguchi et al., 1998; Rogers et al., 2009). Sox1 was reported to induce γ-crystallin expression (Nishiguchi et al., 1998), Sox2 is involved in FGF4, Oct4, Nanog, etc. expression (Kuroda et al., 2005; Okumura-Nakanishi et al., 2005; Yuan et al., 1995) and Sox3 regulates chordin, goosecoid, xvent etc. (Rogers et al., 2009). Although knockdown experiments show the redundancy of SoxB1 factors, knockout of Sox2 or Sox3 led to lethality of the early embryo (Avilion et al., 2003; Rizzoti et al., 2004). This suggests that they have distinct and critical function in embryonic development, implying that either neofunctionalisation or subfunctionalisation has occurred.

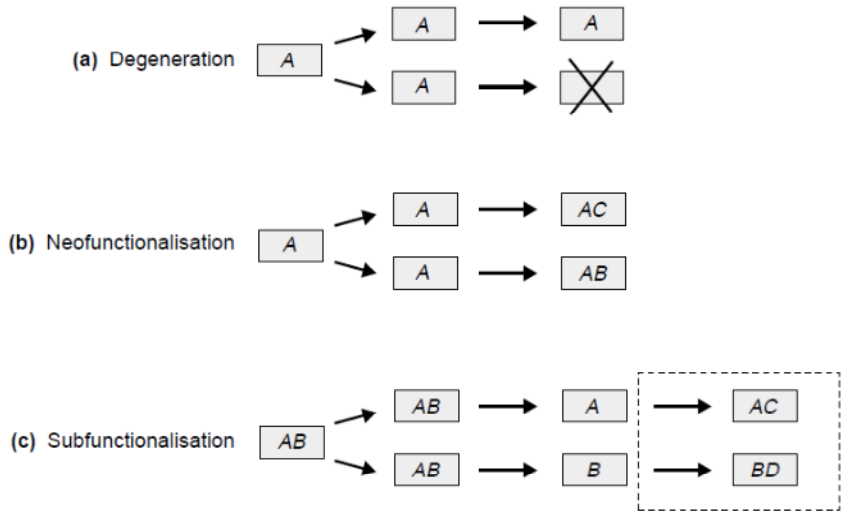


Figure 7.1.3 Scheme of fate of duplication genes.

(A) Degeneration model. One copy of duplication vanished. (B) Neofunctionalisation model. The duplicated genes acquire new functions. (C) Subfunctionalisation model. The multi-functional gene duplicates into two separate genes. The dash line box shows the possible subsequent neofunctionalisation of these genes (Mazet and Shimeld, 2002).

7.2 Regulation of Sox2 transcriptional function

If Sox2 is a bifunctional transcription factor, how could it determine when to act as an activator and when to act as a repressor? In the discussion section of chapter 5, the differing roles of Sox2 in affecting transcription factors in human neural stem cells and embryonic stem cells was mentioned. This suggests that the different type of cell or different developmental stage could alter Sox2 function. On the other hand, in human neural stem cells, Sox2 could function as a transcriptional activator of genes such as Hes5 and as a transcriptional repressor of genes such as GFAP simultaneously. This implies that Sox2 could be a bifunctional transcription factor depending on the context of target genes. In this section, some possible mechanisms are discussed.

7.2.1 Post-translational modification

Many transcription factors are modified after translation. Such modifications change the electronic charge or structure of the transcription factor, hence, potentially regulating their function. Sox2 has been reported to be subject to SUMOylation, ubiquitination, acetylation, phosphorylation and poly(ADP-ribosyl)ation. SUMOylation of proteins covalently conjugates one or many small ubiquitin-related modifier (SUMO) polypeptides to lysine residues (Wilkinson and Henley, 2010). Sox2 has a consensus SUMOylation motif (ψKxD/E; ψ is a large hydrophobic residue) at 246-249 aa within which the lysine 247 is the target residue of SUMOylation (Figure 7.1.1). SUMOylation has been reported to affect protein stability, alter protein-protein interactions and transcription factor function (El McHichi et al., 2010; Garcia-Dominguez and Reyes, 2009; Snow et al., 2010). SUMOylation was found to decrease the transcriptional activation activity of Sox2 via inhibiting Sox2 DNA binding ability (Figure 7.2.1 A) (Tsuruzoe et al., 2006). On the other hand, the SUMOylation site on Sox2 was demonstrated as a phosphorylation-dependent SUMOylation motif (PDSM; ψKxExxSP) (Van Hoof et al., 2009). The replacement of serine 249-251 to aspartic acid, which mimics constitutive phosphorylation, led to a high level of SUMOylation of lysine 245 in human Sox2.

SUMO modification could affect transcriptional regulatory proteins and lead to transcriptional repression through HDAC-mediated transcriptional repression (Yang and Sharrocks, 2004). The ETS domain transcription factor Elk1 was reported to be SUMOylated and recruits HDAC2 to repress target gene. When

the SUMOylation on Elk1 was removed, the transcriptional repression function of Elk1 was reversed.

A decrease in the protein level of a transcription factor could also represses its target gene expression. Ubiqutination is the most common post-translational modification involved in protein degradation. Similar to SUMOylation, ubiquitination also conjugates one or multi polypeptides on to lysine residues covalently. However, ubiquitination sites are not as conserved as SUMOylation sites. The most well-studied function of ubiquitination is to regulate the stability of a protein via proteasomal degradation (Ciechanover, 1998). Sox2 has been shown to be ubiquitinated but the specific ubiqutination sites on Sox2 have not been identified experimentally. However, using the prediction software UbPred (http://www.ubpred.org/) (Radivojac et al., 2010), three possible ubiquitination sites (lysine 10, 247, 310) were predicted (Figure 7.1.1). In addition, the acetylation of Sox2 lysine 75 has been demonstrated to retain the nuclear localization and reduce the levels of ubiquitination of Sox2 (Figure 7.2.1 A) (Baltus et al., 2009b). Moreover, the poly(ADP-ribosyl)ation of Sox2 was reported to regulate Sox2 protein levels and decrease activator of *FGF4* by Sox2 (Gao et al., 2009).

Since different post-translational modifications on transcription factors can lead to different functions, this has been proposed as a subtle functional adjustment of a transcription factor on its target genes. Post-translational modification can alter the surface charge and protein structure of transcription factors. These differences might change their protein-protein interaction and change the components of the transcription regulatory complex, therefore, leading to a different function of transcription factors.

7.2.2 Context of the regulatory region of Sox2 target genes

Bifunctional transcription factors could also determine their role through the specific context of target sequences (Boyle and Despres, 2010). As a transcription factor, Sox2 binds to specific sequences of target genes. It is possible that target genes contain the information which indicates to Sox2 whether to act as a transcriptional activator or repressor (Figure 7.2.1 B). A good example of the regulatory region context of a gene triggering different functions of a transcription factor is Sp3. Sp3 functions as a repressor when bound to multiple DNA-binding sites but acts as an activator when bound to single DNA-binding sites (Majello et al., 1997). Another factor, WRKY53 which was found in *Arabidopsis,* acts either as a transcriptional activator or repressor depending on the sequence surrounding the binding site on its target gene (Miao et al., 2004).

7.2.3 Existence of partners in the environment

As described in chapter 1, the DNA-binding ability of Sox2 alone does not appear to be strong enough to allow it to regulate its target genes. Therefore, cooperation with other transcription factors is necessary to stabilize the interaction. However, different tissues or cells have different gene expression profiles which create various micro-environments for transcription factors. The availability of partner proteins would therefore vary. On the other hand, Sox2 needs specific partners on different target genes. For example, Pax6 and Sox2 activate lens-specific *δ-crystallin* (Kamachi et al., 2001), whereas Oct4 and Sox2 activate another group of genes such as Nanog. There has to be a specific

binding site for Sox2 and its partner protein in close proximity. Therefore, whether the transcriptional regulation of Sox2 could have function on the target genes depends largely on the location of nearby transcription factor binding sites and whether there is a correct partner binding to the adjacent binding site (Figure 7.2.1 C).

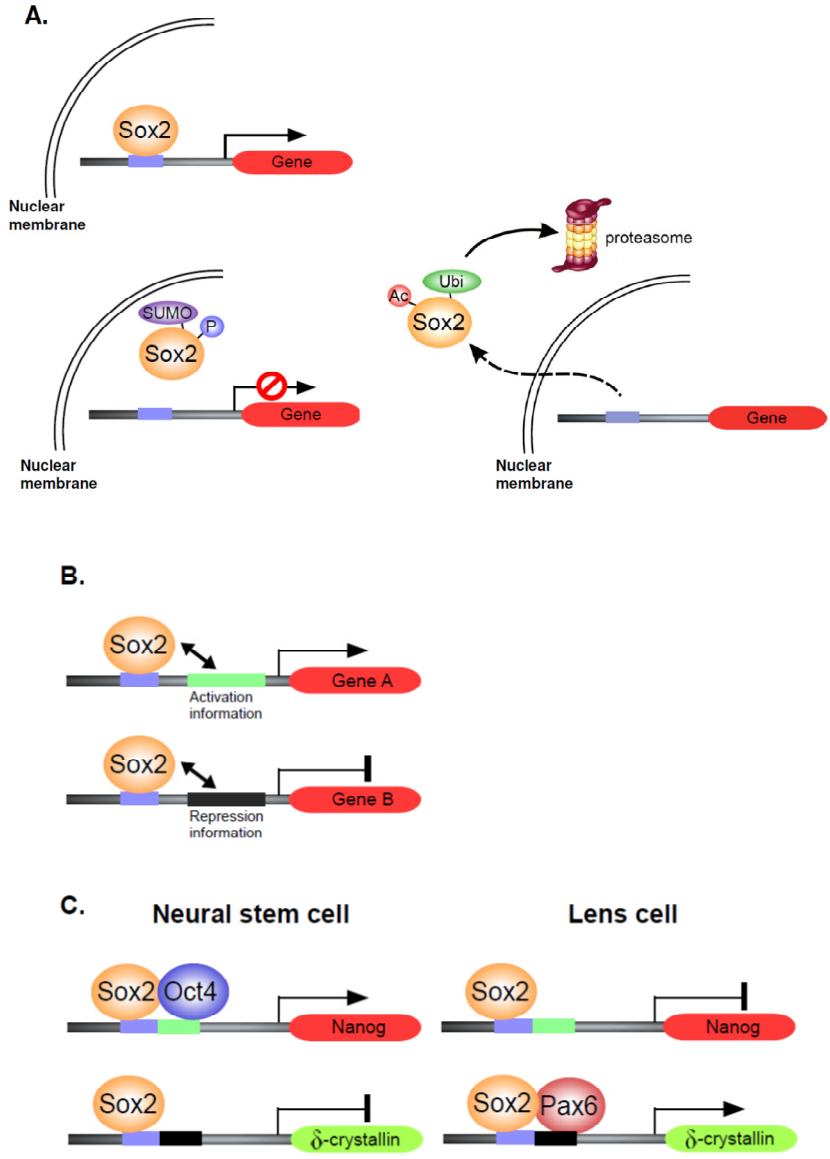


Figure 7.2.1 Scheme of the possible regulation of Sox2 transcriptional activity.

(A) Post-translational modification affects Sox2 transcriptional regulation. Phosphorylation dependant SUMOylation decreases Sox2 transcriptional activation. Acetylation-induced Sox2 ubiquitination leads to Sox2 export from the nucleus to the cytoplasm and processing by the proteasome. (B) Context of regulatory region of Sox2 target genes in front of transcription start point affects Sox2 transcriptional regulation. The flanking sequence of Sox2 binding sites or the secondary structure of the nearby DNA might contain the information for Sox2 to determine its transcriptional regulation role. (C) Existence of partner factors affects Sox2 transcriptional regulation. Different cell types have different gene expression profiles which supplies different partners for Sox2 to complete the transcriptional activation/repression of target genes.

7.3 Function of Sox2 transcriptional repression

7.3.1 Role of Sox2 in neural stem cells

The data presented in chapter 5 shows that approximate the same number of genes are repressed as are activated by overexpression of Sox2 (Table 5.5.1). This indicates the importance of Sox2 transcriptional repression in neural stem cells. Previous studies have shown that a constitutive activator form of SoxB1 subgroup proteins inhibited neural stem cell differentiation, while a constitutive repressor form of SoxB1 led to the early stages of differentiation (Bylund et al., 2003; Graham et al., 2003). This suggests that the ability of SoxB1 subgroup proteins to regulate the differentiation of neural stem cells depends on their transcriptional activator function. However, the differentiation induced by the constitutive repressor form of SoxB1 was not entirely complete. Fully differentiated neuron markers such as neurofilament or beta-tubulin were not detected. On the other hand, the constitutive activator form of SoxB1 proteins also did not fully inhibit the differentiation of neural stem cells, as shown by the expression of beta-tubulin. Together with the findings in this thesis, we propose a possible model of the mechanism of action of Sox2 in neural stem cells, as shown in Figure 7.3.1. The transcriptional activation function of Sox2 maintains neural stem cell features and inhibits differentiation via regulating down-stream transcription factors. The transcriptional repression function of Sox2 is required to inhibit differentiation through repressing effectors which will be activated after release from the neural stem cell state.

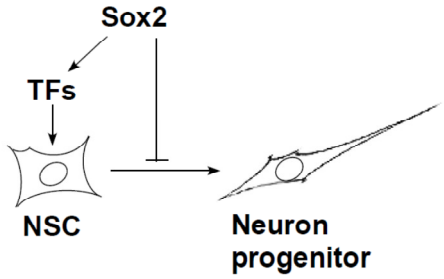


Figure 7.3.1 Putative model of the function of Sox2 in neural stem cell differentiation.

The transcriptional activation function of Sox2 activates the expression of down-stream transcription factors whereas the transcriptional repression function of Sox2 represses other effectors to inhibit differentiation of neural stem cells.

In chapter 5, I showed a relatively large proportion (143 out of 676 genes) of the genes activated by Sox2 in neural stem cells are transcription factors, whereas a large portion of Sox2 repressed genes are factors involved in regulating cell cycle (118 out of 786 genes). Surprisingly, amongst the Sox2 repressed genes not a single gene falls in the category of GO term as transcription factor. This indicates that Sox2 transcriptional activation and repression functions are very distinct in terms of their target genes in neural stem cells.

As discussed in chapter 5, Sox2 acts as a transcriptional activator of transcription factor genes in hNSCs (62 out of 704 genes in GO:0003700), but Sox2 was reported to bind to transcriptionally inactive transcription factor genes that have been implicated in developmental processes in human embryonic stem cells (Boyer et al., 2005). Although Sox2 also binds a small number of transcriptionally active transcription factor genes, this situation opposite to that I have described for human neural stem cells might result from the different stage of differentiation between embryonic stem cells and neural

stem cells. Embryonic stem cells are pluripotent whereas neural stem cells are multipotent cells. This implies that the role of Sox2 during embryonic development changes significantly. Further investigations are needed to dissect the transcriptional function of Sox2 more detail.

Sox2 has been reported to prevent neural stem cells from processing through differentiation which includes keeping self-renewal or proliferation of cells (Avilion et al., 2003; Bylund et al., 2003). Interestingly, in my data, Sox2 repressed many cell cycle related genes which might conflict with these previous findings. However, among the 118 cell cycle related genes, only 24 genes function as positive regulators of cell cycle, such as CCNA2 and CCNE1, while 18 genes function as negative regulator of cell cycle, such as BRCA1 and BRCA2. The rest of the genes are related to other cell cycle processes, such as chromosome segregation, centrosome duplication, etc., rather than the regulation of cell cycle. Similar with this finding, some cell cycle related genes such as PLAGL1, MYH9, NEK6, E2F8 and GSPT1are repressed by Sox2 both in human neural stem cells and mouse embryonic stem cells (Masui et al., 2007). Therefore, Sox2 might encourage self-renewal through repressing the negative regulation of genes of the cell cycle and adjusting the cell cycle rate by repressing the positive regulation genes. The possible hypothesis of different functions of Sox2 might be adjusted by the partners of Sox2 or other factors.

This indicates that Sox2 represses at least some of the same biological function in different species and developmental stage. The mechanisms behind this phenomemon need further investigation.

7.3.2 Sox2 in zebrafish

In human and mouse, sox2 expression is earlier than sox3 in the ectoderm. However, in chicken, zebrafish and frog, the earliest *sox2* expression is in the neural ectoderm whereas *sox3* is expressed at much earlier stages. Because of the high similarity within *soxb1* subgroup proteins, they might have redundant function on the target genes. *sox3* has been reported to regulate organizer formations in zebrafish through directly repressing the expressions of organizer marker genes (Shih et al., 2010). Therefore, it is difficult to study *sox2* endogenous biological function in zebrafish. We assume the effect of overexpressed *sox2* and *sox2* mutant in the 1-2 cell stage might lead to some functional effects on *sox3* target genes, even though these events might not be the normal function of *sox2*. Therefore, this system was used in chapter 6 to test the effects of Sox2-Groucho interaction.

The data represented in chapter 6 show that, like *sox3*, *sox2* represses *chordin* expression, but unlike *sox3* it also represses *sox19b* but it has opposite effect on *goosecoid*. In addition, the expression pattern of *sox2* at the 80% epiboly stage in zebrafish embryos (mid-gastrula stage) is restricted to the dorsal-anterior ectoderm whereas *sox3* is expressed in almost the whole prospective neural ectorderm (Figure 7.3.2). This indicates that the function of *sox2* in early zebrafish development might be slightly different from that of *sox3*.

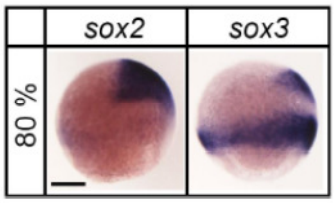


Figure 7.3.2 *sox2* and *sox3* expression pattern at 80% epiboly zebrafish embryo.

The 80% epiboly zebrafish embryos were processed through in situ hybridization using *sox2* or *sox3* probes. The pictures were took from lateral view. (Data were done by Yu-Haun Shih)

Although *sox3* plays an important role in early zebrafish development, *sox3* null mutant zebrafish which express only the dysfunctional N-terminus of *sox3* have no obvious phenotype when compared with the wild-type (unpublished data in our lab). However, using *sox3* morpholinos (MO1 and MO2) to knock-down endogenous *sox3* at the 1-2 cell stage produced a severe brain defect (Figure 7.3.3) (Dee et al., 2008). When the morpholino sequences were compared to the other SoxB1 subgroup members, MO2 showed high similarity with *sox2* which might have led to an off-target effect decreasing *sox2* leads simultaneously (Figure 7.3.4 A). In order to test if the off-target effect could happen on *sox2*, the 1-2 cell stage zebrafish embryos were co-injected with GFP-*sox2* or *sox3* alone with different combinations of *sox3* morpholinos (Figure 7.3.4 B-I). The results show that *sox2* expression indeed decreased when *sox3* morpholino MO2 were injected, but not MO1. Therefore, the brain defects in *sox3* morpholino (MO1 and MO2) injected zebrafish might be due to loss of both *sox2* and *sox3* rather than *sox3* alone. This indicates that *sox2* not only regulates the organizer but is also involved in brain development.

However, the unpublished data in our lab have shown that *sox3* induces *sox2* expression in a cell-autonomous manner. Therefore, a decrease of *sox3* might

also affect *sox2* expression levels indirectly. This possibility could be ruled out by the single knock-down of *sox2* or *sox3* by a more specific morpholino was done by Okuda et al. and the result showed no serious defective phenotype (Figure 7.3.5) (Okuda et al., 2010). These data suggest a compensat function between *sox2* and *sox3* during neural development, explaining the lack of phenotype in *sox3* mutant zebrafish. Nevertheless, a more detailed investigation of *sox2* and *sox3* function is needed to reveal their specific functions.

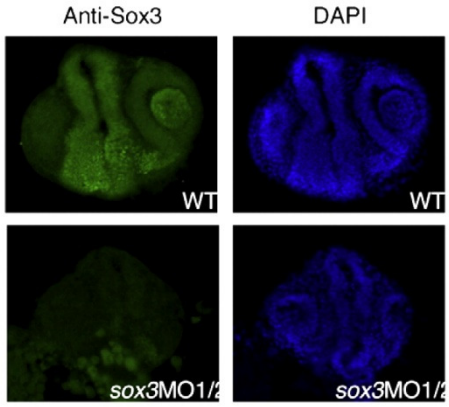


Figure 7.3.3 Brain defect caused by *sox3* morpholinos (MO1/2).

The immunostaining of brain section of wild-type and *sox3* morpholinos injected zebrafish brain was carried out using Anti-Sox3 antibody (From Dee et al., 2008).

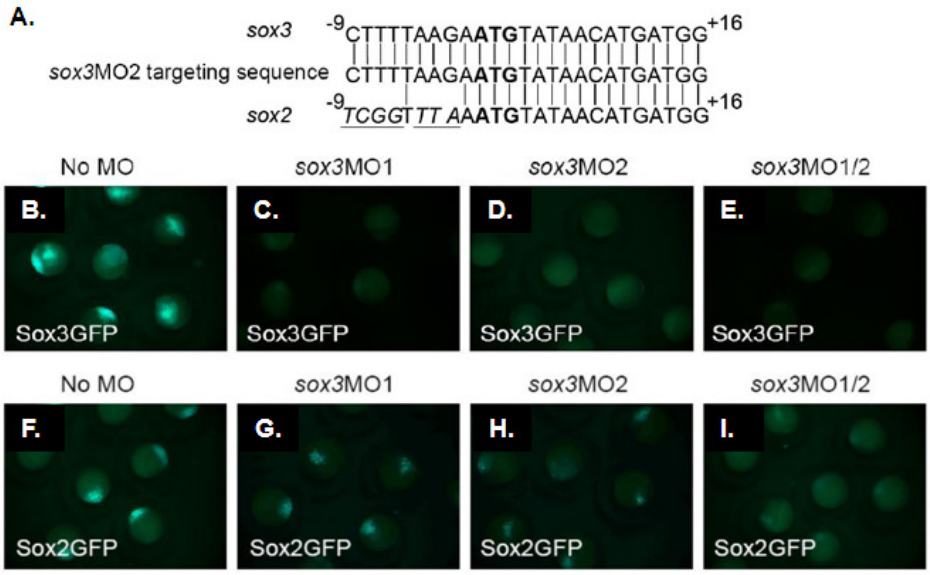


Figure 7.3.4 The specificity of *sox3* morpholinos.

(A) The alignment of *sox3* morpholino with *sox2* and *sox3* sequence. (B-E) Co-injection of 2.5ng *sox3* morpholinos (MO1 and MO2) alone or together with sox3-GFP at 1-2 cell stage and the image were took at 70% epiboly stage. (F-I) C0-injection of 2.5ng *sox3* morpholinos alone or together with sox2-GFP at 1-2 cell stage and the image were took at 70% epiboly stage. (pictures taken from PhD thesis, Yu-Huan Shih, 2009)

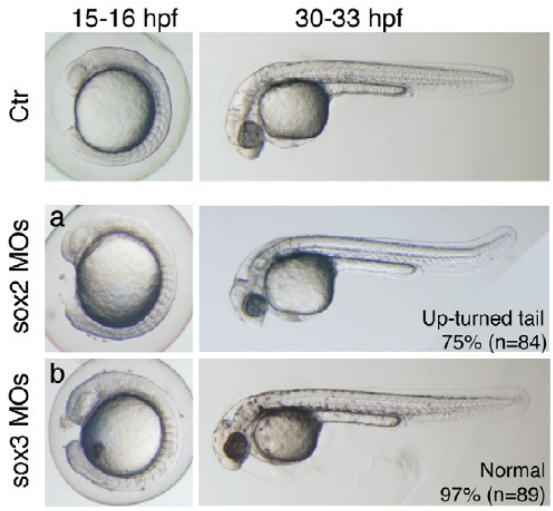


Figure 7.3.5 Specific knock-down of *sox2* and *sox3* have no serious effect on zebrafish.

Specific *sox2* or *sox3* morpholinos (0.9ng) were injected at 1 cell stage zebrafish embryos. The images were taken at15-16 hpf and 30-33 hpf respectively. Image was taken from Okuda et al., 2010.

7.4 Summary and perspectives

The data presented in this thesis indicate that Sox2 functions as a bifunctional transcription factor. Through interaction with Groucho, Sox2 acts as a transcriptional repressor of GFAP. However, the interaction of Sox2 with Groucho is required for regulation of only a small proportion of Sox2 target genes in neural stem cells and loss of their interaction has a dramatic effect on the function of Sox2 when overexpressed in early zebrafish embryo development.

There are several areas of further research identified in this thesis that could help understand the function of Sox2 more detail. First, in order to fully analyse the mechanism of Sox2 transcriptional repression function, other possible co-repressors, chromatin remodelers and the interaction with heterochromatin should be tested. Second, the mechanism by which Sox2 switches between transcriptional activator and repressor needs to be studied in the future.

As for the function of Sox2 in neural stem cells, the data in this thesis are the results of overexpression Sox2 and Sox2 mutant which lose Groucho/Sox2 interaction motif. Forcing cells to overexpress Sox2 might saturate the factors that Sox2 needs for its function or be toxic to the cells. Therefore, it might lead to an artificial result that not reflecting the normal function of Sox2. In order to understand the full function of endogenous Sox2, the expression array analysis of Sox2 knockdown cells is necessary

From the array data represented in this thesis, it seems that Sox2 transcriptional repression function affects the expression of many genes in neural stem cells. In order to confirm that those genes are genuine Sox2 repression targets, the

further studies of the candidates that selected from array data are needed. The expression level and the analysis of Sox2 binding sites in the candidate genes could confirm the relationship between Sox2 and the candidate genes.

Finally, the biological function of Sox2 transcriptional repressor function *in vivo* is needed. In zebrafish, further studies of the role of Sox2 in regulating the organizer formation and brain formation are needed to elucidate the mechanisms leading to the phenomena presented in chapter 6. On the other hand, Sox2 transcriptional activation and repression function affect genes which have distinct functions in neural stem cells; activating transcription factor, repressing cell cycle relative factors. Further investigations are needed to understand the role of these activities in neuronal differentiation and stem cell maintenance.

Chapter 8 Bibliography

Aksoy, S., Szumlanski, C.L., and Weinshilboum, R.M. (1994). Human liver nicotinamide N-methyltransferase. cDNA cloning, expression, and biochemical characterization. J Biol Chem *269*, 14835-14840.

Anderson, D.J., Gage, F.H., and Weissman, I.L. (2001). Can stem cells cross lineage boundaries? Nature Medicine *7*, 393-395.

Aquino, J.B., Lallemend, F., Marmigere, F., Adameyko, I., Golemis, E.A., and Ernfors, P. (2009). The retinoic acid inducible Cas-family signaling protein Nedd9 regulates neural crest cell migration by modulating adhesion and actin dynamics. neuroscience *162*, 14.

Aronson, B.D., Fisher, A.L., Blechman, K., Caudy, M., and Gergen, J.P. (1997). Groucho-dependent and -independent repression activities of Runt domain proteins. Mol Cell Biol *17*, 5581-5587.

Avilion, A.A., Nicolis, S.K., Pevny, L.H., Perez, L., Vivian, N., and Lovell-Badge, R. (2003). Multipotent cell lineages in early mouse development depend on SOX2 function. Genes & Development *17*, 126-140.

Bajoghli, B., Aghaallaei, N., Soroldoni, D., and Czerny, T. (2007). The roles of Groucho/Tle in left-right asymmetry and Kupffer's vesicle organogenesis. Dev Biol *303*, 347-361.

Baltus, G.A., Kowalski, M.P., Tutter, A.V., and Kadam, S. (2009a). A positive regulatory role for the mSIN3a/HDAC complex in pluripotency through nanog and sox2. J Biol Chem.

Baltus, G.A., Kowalski, M.P., Zhai, H., Tutter, A.V., Quinn, D., Wall, D., and Kadam, S. (2009b). Acetylation of sox2 induces its nuclear export in embryonic stem cells. Stem Cells *27*, 2175-2184.

Bani-Yaghoub, M., Tremblay, R.G., Lei, J.X., Zhang, D., Zurakowski, B., Sandhu, J.K., Smith, B., Ribecco-Lutkiewicz, M., Kennedy, J., Walker, P.R.*, et al.* (2006). Role of Sox2 in the development of the mouse neocortex. Dev Biol *295*, 52-66.

Beagle, B., and Johnson, G.V. (2010). AES/GRG5: more than just a dominant-negative TLE/GRG family member. Dev Dyn *239*, 2795-2805.

Bell, K.M., Western, P.S., and Sinclair, A.H. (2000). SOX8 expression during chick embryogenesis. Mechanisms of Development *94*, 4.

Bernier, G., Panitz, F., Zhou, X., Hollemann, T., Gruss, P., and Pieler, T. (2000). Expanded retina territory by midbrain transformation upon overexpression of Six6 (Optx2) in Xenopus embryos. Mech Dev *93*, 59-69.

Botquin, V., Hess, H., Fuhrmann, G., Anastassiadis, C., Gross, M.K., Vriend, G., and Scholer, H.R. (1998). New POU dimer configuration mediates antagonistic control of an osteopontin preimplantation enhancer by Oct-4 and Sox-2. Genes Dev *12*, 2073-2090.

Bowles, J., Schepers, G., and Koopman, P. (2000). Phylogeny of the SOX family of developmental transcription factors based on sequence and structural indicators. Dev Biol *227*, 239-255.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G.*, et al.* (2005). Core transcriptional regulatory circuitry in human embryonic stem cells. Cell *122*, 947-956.

Boyle, P., and Despres, C. (2010). Dual-function transcription factors and their entourage: unique and unifying themes governing two pathogenesis-related genes. Plant Signal Behav *5*, 629-634.

Brantjes, H., Roose, J., van De Wetering, M., and Clevers, H. (2001). All Tcf HMG box transcription factors interact with Groucho-related co-repressors. Nucleic Acids Res *29*, 1410-1419.

Bubnoff, A.v., and Cho, K.W.Y. (2001). Intracellular BMP Signaling Regulation
in Vertebrates: Pathway or Network? Developmental Biology *239*, 14.

Bylund, M., Andersson, E., Novitch, B.G., and Muhr, J. (2003). Vertebrate neurogenesis is counteracted by Sox1-3 activity. Nat Neurosci *6*, 1162-1168.

Cavallaro, M., Mariani, J., Lancini, C., Latorre, E., Caccia, R., Gullo, F., Valotta, M., DeBiasi, S., Spinardi, L., Ronchi, A.*, et al.* (2008). Impaired generation of mature neurons by neural stem cells from hypomorphic Sox2 mutants. Development *135*, 541-557.

Cavallo, R.A., Cox, R.T., Moline, M.M., Roose, J., Polevoy, G.A., Clevers, H., Peifer, M., and Bejsovec, A. (1998). Drosophila Tcf and Groucho interact to repress Wingless signalling activity. Nature *395*, 604-608.

Chao, A.T., Jones, W.M., and Bejsovec, A. (2007). The HMG-box transcription factor SoxNeuro acts with Tcf to control Wg/Wnt signaling activity. Development *134*, 989-997.

Chen, C., Agnes, F., and Gelinas, C. (1999). Mapping of a serine-rich domain essential for the transcriptional, antiapoptotic, and transforming activities of the v-Rel oncoprotein. Mol Cell Biol *19*, 307-316.

Chen, G., and Courey, A.J. (2000). Groucho/TLE family proteins and transcriptional repression. Gene *249*, 1-16.

Chen, X., Xu, H., Yuan, P., Fang, F., Huss, M., Vega, V.B., Wong, E., Orlov, Y.L., Zhang, W., Jiang, J.*, et al.* (2008). Integration of external signaling pathways with the core transcriptional network in embryonic stem cells. Cell *133*, 1106-1117.

Cheng, Y.C., Lee, C.J., Badge, R.M., Orme, A.T., and Scotting, P.J. (2001). Sox8 gene expression identifies immature glial cells in developing cerebellum and cerebellar tumours. Brain Res Mol Brain Res *92*, 193-200.

Chew, J.L., Loh, Y.H., Zhang, W., Chen, X., Tam, W.L., Yeap, L.S., Li, P., Ang, Y.S., Lim, B., Robson, P.*, et al.* (2005). Reciprocal transcriptional regulation of Pou5f1 and Sox2 via the Oct4/Sox2 complex in embryonic stem cells. Mol Cell Biol *25*, 6031-6046.

Choi, C.Y., Kim, Y.H., Kim, Y.O., Park, S.J., Kim, E.A., Riemenschneider, W., Gajewski, K., Schulz, R.A., and Kim, Y. (2005). Phosphorylation by the DHIPK2 protein kinase modulates the corepressor activity of Groucho. J Biol Chem *280*, 21427-21436.

Ciechanover, A. (1998). The ubiquitin-proteasome pathway: on protein death and cell life. EMBO J *17*, 7151-7160.

Collignon, J., Sockanathan, S., Hacker, A., Cohen-Tannoudji, M., Norris, D., Rastan, S., Stevanovic, M., Goodfellow, P.N., and Lovell-Badge, R. (1996). A comparison of the properties of Sox-3 with Sry and two related genes, Sox-1 and Sox-2. Development *122*, 509-520.

Courey, A.J., and Jia, S. (2001). Transcriptional repression: the long and the short of it. Genes Dev *15*, 2786-2796.

Cox, J.L., Mallanna, S.K., Luo, X., and Rizzino, A. (2010). Sox2 uses multiple domains to associate with proteins present in Sox2-protein complexes. PLoS ONE *5*, e15486.

Dattani, M.T., Martinez-Barbera, J.-P., Thomas, P.Q., Brickman, J.M., Gupta, R., Martensson, I.-L., Toresson, H., Fox, M., Wales, J.K.H., Hindmarsh, P.C.*, et al.* (1998). Mutations in the homeobox gene HESX1/Hesx1 associated with septo-optic dysplasia in human and mouse. nature genetics *19*, 9.

Dee, C.T., Hirst, C.S., Shih, Y.H., Tripathi, V.B., Patient, R.K., and Scotting, P.J. (2008). Sox3 regulates both neural fate and differentiation in the zebrafish ectoderm. Dev Biol *320*, 289-301.

Deneen, B., Ho, R., Lukaszewicz, A., Hochstim, C.J., Gronostajski, R.M., and Anderson, D.J. (2006). The transcription factor NFIA controls the onset of gliogenesis in the developing spinal cord. neuron *52*, 16.

Di Rocco, G., Gavalas, A., Popperl, H., Krumlauf, R., Mavilio, F., and Zappavigna, V. (2001). The recruitment of SOX/OCT complexes and the differential activity of HOXA1 and HOXB1 modulate the Hoxb1 auto-regulatory enhancer function. J Biol Chem *276*, 20506-20515.

Dubnicoff, T., Valentine, S.A., Chen, G., Shi, T., Lengyel, J.A., Paroush, Z., and Courey, A.J. (1997). Conversion of dorsal from an activator to a repressor by the global corepressor Groucho. Genes Dev *11*, 2952-2957.

Duncan, M.K., Haynes, J.I., 2nd, Cvekl, A., and Piatigorsky, J. (1998). Dual roles for Pax-6: a transcriptional repressor of lens fiber cell-specific beta-crystallin genes. Mol Cell Biol *18*, 5579-5586.

Eberhard, D., Jimenez, G., Heavey, B., and Busslinger, M. (2000). Transcriptional repression by Pax5 (BSAP) through interaction with corepressors of the Groucho family. EMBO J *19*, 2292-2303.

El McHichi, B., Regad, T., Maroui, M.A., Rodriguez, M.S., Aminev, A., Gerbaud, S., Escriou, N., Dianoux, L., and Chelbi-Alix, M.K. (2010). SUMOylation promotes PML degradation during encephalomyocarditis virus infection. J Virol *84*, 11634-11645.

Eminli, S., Utikal, J., Arnold, K., Jaenisch, R., and Hochedlinger, K. (2008). Reprogramming of neural progenitor cells into induced pluripotent stem cells in the absence of exogenous Sox2 expression. Stem Cells *26*, 2467-2474.

Fairbanks, A.J.L.a.G. (1976). Chemical probes of extended biological structures: Synthesis and properties of the cleavable protein cross-linking reagent [35S]dithiobis(succinimidyl propionate). Journal of Molecular Biology *104*, 19.

Fisher, A.L., and Caudy, M. (1998). Groucho proteins: transcriptional corepressors for specific subsets of DNA-binding transcription factors in vertebrates and invertebrates. Genes Dev *12*, 1931-1940.

Fisher, A.L., Ohsako, S., and Caudy, M. (1996). The WRPW motif of the hairy-related basic helix-loop-helix repressor proteins acts as a 4-amino-acid transcription repression and protein-protein interaction domain. Mol Cell Biol *16*, 2670-2677.

Forwood, J.K., Harley, V., and Jans, D.A. (2001). The C-terminal nuclear localization signal of the sex-determining region Y (SRY) high mobility group domain mediates nuclear import through importin beta 1. J Biol Chem *276*, 46575-46582.

Frasch, M., and Levine, M. (1987). Complementary patterns of even-skipped and fushi tarazu expression involve their differential regulation by a common set of segmentation genes in Drosophila. Genes Dev *1*, 981-995.

Fujita, N., and Wade, P.A. (2004). Use of bifunctional cross-linking reagents in mapping genomic distribution of chromatin remodeling complexes. Methods *33*, 81-85.

Gangemi, R.M., Griffero, F., Marubbi, D., Perera, M., Capra, M.C., Malatesta, P., Ravetti, G.L., Zona, G.L., Daga, A., and Corte, G. (2008). SOX2 Silencing in Glioblastoma Tumor Initiating Cells Causes Stop of Proliferation and Loss of Tumorigenicity. Stem Cells.

Gao, F., Kwon, S.W., Zhao, Y., and Jin, Y. (2009). PARP1 poly(ADP-ribosyl)ates Sox2 to control Sox2 protein levels and FGF4 expression during embryonic stem cell differentiation. J Biol Chem *284*, 22263-22273.

Gao, X., Chandra, T., Gratton, M.O., Quelo, I., Prud'homme, J., Stifani, S., and St-Arnaud, R. (2001). HES6 acts as a transcriptional repressor in myoblasts and can induce the myogenic differentiation program. J Cell Biol *154*, 1161-1171.

Garcia-Dominguez, M., and Reyes, J.C. (2009). SUMO association with repressor complexes, emerging routes for transcriptional control. Biochim Biophys Acta *1789*, 451-459.

Gasperowicz, M., and Otto, F. (2005). Mammalian Groucho homologs: redundancy or specificity? J Cell Biochem *95*, 670-687.

Gaul, U., and Jäckle, H. (1987). How to fill a gap in the Drosophila embryo. Trends in Genetics *3*, 127-131.

Gebbia, m., Ferrero, G.B., Pilia, G., Bassi, M.T., Aylsworth, A.S., Penman-Splitt, M., Bird, L.M., Bamforth, J.S., Burn, J., Schlessinger, D*., et al.* (1997). X-linked situs abnormalities recult from mutations in ZIC3. nature genetics *17*, 4.

Ginsburg, D., Zeheb, R., Yang, A.Y., Rafferty, U.M., Andreasen, P.A., Nielsen, L., Dano, K., Lebo, R.V., and Gelehrter, T.D. (1086). cDNA Cloning of Human Plasminogen Activator-inhibitor from Endothelial Cells. J Clin Invest *78*, 8.

Girard, F., Cremazy, F., Berta, P., and Renucci, A. (2001). Expression pattern of the Sox31 gene during zebrafish embryonic development. Mech Dev *100*, 71-73.

Goldstein, R.E., Cook, O., Dinur, T., Pisante, A., Karandikar, U.C., Bidwai, A., and Paroush, Z. (2005). An eh1-like motif in odd-skipped mediates recruitment of Groucho and repression in vivo. Mol Cell Biol *25*, 10711-10720.

Goldstein, R.E., Jimenez, G., Cook, O., Gur, D., and Paroush, Z. (1999). Huckebein repressor activity in Drosophila terminal patterning is mediated by Groucho. Development *126*, 3747-3755.

Graham, V., Khudyakov, J., Ellis, P., and Pevny, L. (2003). SOX2 Functions to Maintain Neural Progenitor Identity. *39*, 749-765.

Gray, S., and Levine, M. (1996). Transcriptional repression in development. Curr Opin Cell Biol *8*, 358-364.

Grbavec, D., and Stifani, S. (1996). Molecular interaction between TLE1 and the carboxyl-terminal domain of HES-1 containing the WRPW motif. Biochem Biophys Res Commun *223*, 701-705.

Greber, B., Lehrach, H., and Adjaye, J. (2007). Silencing of core transcription factors in human EC cells highlights the importance of autocrine FGF signaling for self-renewal. BMC Dev Biol *7*, 46.

Hagen, G., Muller, S., Beato, M., and Suske, G. (1994). Sp1-mediated transcriptional activation is repressed by Sp3. EMBO J *13*, 3843-3851.

Hanna-Rose, W., and Hansen, U. (1996). Active repression mechanisms of eukaryotic transcription repressors. Trends Genet *12*, 229-234.

Heimbucher, T., Murko, C., Bajoghli, B., Aghaallaei, N., Huber, A., Stebegg, R., Eberhard, D., Fink, M., Simeone, A., and Czerny, T. (2007). Gbx2 and Otx2 interact with the WD40 domain of Groucho/Tle corepressors. Mol Cell Biol *27*, 340-351.

Helen Brantjes, J.R., Marc van de Wetering and Hans Clevers (2001). All Tcf hmg box tarnscription factors interact with Groucho-related co-repressors. Nucleic Acids Research *29*, 10.

Hemmati-Brivanlou, A., Kelly, O.G., and Melton, D.A. (1994). Follistatin, an antagonist of activin, is expressed in the Spemann organizer and displays direct neuralizing activity. cell *77*, 283-295.

Hever, A., Williamson, K., and Heyningen, V.v. (2006). Developmental malformations of the eye: the role of PAX6, SOX2 and OTX2. clinical Genetics *69*, 12.

Holland, L.Z., Schubert, M., Holland, N.D., and Neuman, T. (2000). Evolutionary conservation of the presumptive neural plate markers AmphiSox1/2/3 and AmphiNeurogenin in the invertebrate chordate amphioxus. Dev Biol *226*, 18-33.

Hosking, B.M., Muscat, G.E., Koopman, P.A., Dowhan, D.H., and Dunn, T.L. (1995). Trans-activation and DNA-binding properties of the transcription factor, Sox-18. Nucleic Acids Res *23*, 2626-2628.

Hu, S., Wu, Z., Yan, Y., and Li, Y. (2011). Sox31 is involved in central nervous system anteroposterior regionalization through

regulating the organizer activity in zebrafish. Acta Biochim Biophys Sin *43*, 13.

Ishii, Y., Abu-Elmagd, M., and Scotting, P.J. (2001). Sox3 expression defines a common primordium for the epibranchial placodes in chick. Dev Biol *236*, 344-353.

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

Jarvis, C.D., Geiman, T., Vila-Storm, M.P., Osipovich, O., Akella, U., Candeias, S., Nathan, I., Durum, S.K., and Muegge, K. (1996). A novel putative helicase produced in early rnurine lymphocytes. Gene *169*, 5.

Ji, S.J., Periz, G., and Sockanathan, S. (2009). Nolz1 is induced by retinoid signals and controls motoneuron subtype identity through distinct repressor activities. Development *136*, 231-240.

Johnson, A.D. (1995). The price of repression. Cell *81*, 655-658.

Ju, B.G., Solum, D., Song, E.J., Lee, K.J., Rose, D.W., Glass, C.K., and Rosenfeld, M.G. (2004). Activating the PARP-1 sensor component of the groucho/ TLE1 corepressor complex mediates a CaMKinase IIdelta-dependent neurogenic gene activation pathway. Cell *119*, 815-829.

Kamachi, Y., Cheah, K.S., and Kondoh, H. (1999). Mechanism of Regulatory Target Selection by the SOX High-Mobility-Group Domain Proteins as Revealed by Comparison of SOX1/2/3 and SOX9. Mol Cell Biol *19*, 107-120.

Kamachi, Y., Sockanathan, S., Liu, Q., Breitman, M., Lovell-Badge, R., and Kondoh, H. (1995). Involvement of SOX proteins in lens-specific activation of crystallin genes. EMBO J *14*, 3510-3519.

Kamachi, Y., Uchikawa, M., Collignon, J., Lovell-Badge, R., and Kondoh, H. (1998). Involvement of Sox1, 2 and 3 in the early and subsequent molecular events of lens induction. Development *125*, 2521-2532.

Kamachi, Y., Uchikawa, M., and Kondoh, H. (2000). Pairing SOX off: with partners in the regulation of embryonic development. Trends Genet *16*, 182-187.

Kamachi, Y., Uchikawa, M., Tanouchi, A., Sekido, R., and Kondoh, H. (2001). Pax6 and SOX2 form a co-DNA-binding partner complex that regulates initiation of lens development. Genes Dev *15*, 1272-1286.

Kan, L., Israsena, N., Zhang, Z., Hu, M., Zhao, L.R., Jalali, A., Sahni, V., and Kessler, J.A. (2004). Sox1 acts through multiple independent pathways to promote neurogenesis. Dev Biol *269*, 580-594.

Kang, S.A., Seol, J.H., and Kim, J. (2005). The conserved WRPW motif of Hes6 mediates proteasomal degradation. Biochem Biophys Res Commun *332*, 33-36.

Karin, M. (1992). Signal transduction from cell surface to nucleus in development and disease. FASEB J *6*, 2581-2590.

Kent, J., Wheatley, S.C., Andrews, J.E., Sinclair, A.H., and Koopman, P. (1996). A male-specific role for SOX9 in vertebrate sex determination. Development *122*, 2813-2822.

Kim, J., Chu, J., Shen, X., Wang, J., and Orkin, S.H. (2008a). An extended transcriptional network for pluripotency of embryonic stem cells. Cell *132*, 1049-1061.

Kim, J., Lo, L., Dormand, E., and Anderson, D.J. (2003). SOX10 maintains multipotency and inhibits neuronal differentiation of neural crest stem cells. Neuron *38*, 17-31.

Kim, J.B., Sebastiano, V., Wu, G., Arauzo-Bravo, M.J., Sasse, P., Gentile, L., Ko, K., Ruau, D., Ehrich, M., van den Boom, D.*, et al.* (2009a). Oct4-induced pluripotency in adult neural stem cells. Cell *136*, 411-419.

Kim, J.B., Zaehres, H., Wu, G., Gentile, L., Ko, K., Sebastiano, V., Arauzo-Bravo, M.J., Ruau, D., Han, D.W., Zenke, M.*, et al.* (2008b). Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors. Nature *454*, 646-650.

Kim, M., Habiba, A., Doherty, J.M., Mills, J.C., Mercer, R.W., and Huettner, J.E. (2009b). Regulation of mouse embryonic stem cell neural differentiation by retinoic acid. Dev Biol.

Kishi, M., Mizuseki, K., Sasai, N., Yamazaki, H., Shiota, K., Nakanishi, S., and Sasai, Y. (2000). Requirement of Sox2-mediated signaling for differentiation of early Xenopus neuroectoderm. Development *127*, 791-800.

Knust, E., Schrons, H., Grawaen, F., and Campos-Ortega, J.A. (1992). Seven Genes of the Enhancer of split Complex of Drosophila melanogaster

Encode Helix-Loop-Helix Proteins. Genetics *132*, 505-518.

Kopp, J.L., Ormsbee, B.D., Desler, M., and Rizzino, A. (2008). Small Increases in the Level of Sox2 Trigger the Differentiation of Mouse Embryonic Stem Cells. Stem Cells *26*, 903-911.

Kormish, J.D., Sinner, D., and Zorn, A.M. (2010). Interactions between SOX factors and Wnt/beta-catenin signaling in development and disease. Dev Dyn *239*, 56-68.

Kuhlbrodt, K., Herbarth, B., Sock, E., Hermans-Borgmeyer, I., and Wegner, M. (1998). Sox10, a novel transcriptional modulator in glial cells. J Neurosci *18*, 237-250.

Kuroda, T., Tada, M., Kubota, H., Kimura, H., Hatano, S.Y., Suemori, H., Nakatsuji, N., and Tada, T. (2005). Octamer and Sox elements are required for transcriptional cis regulation of Nanog gene expression. Mol Cell Biol *25*, 2475-2485.

Laemmli, U.K. (1970). Cleavage of structural proteins during the assembly of the head of bacteriophage T4. Nature *227*, 680-685.

Lagutin, O., Zhu, C.C., Furuta, Y., Rowitch, D.H., McMahon, A.P., and Oliver, G. (2001). Six3 promotes the formation of ectopic optic vesicle-like structures in mouse embryos. Dev Dyn *221*, 342-349.

Lania, L., Majello, B., and De Luca, P. (1997). Transcriptional regulation by the Sp family proteins. Int J Biochem Cell Biol *29*, 1313-1323.

Laudet, V., Stehelin, D., and Clevers, H. (1993). Ancestry and diversity of the HMG box superfamily. Nucleic Acids Res *21*, 2493-2501.

Lee, T.I., and Young, R.A. (2000). Transcription of eukaryotic protein-coding genes. Annu Rev Genet *34*, 77-137.

Lefebvre, V., Dumitriu, B., Penzo-Mendez, A., Han, Y., and Pallavi, B. (2007). Control of cell fate and differentiation by Sry-related high-mobility-group box (Sox) transcription factors. Int J Biochem Cell Biol *39*, 2195-2214.

Lengerke, C., Fehm, T., Kurth, R., Neubauer, H., Scheble, V., Muller, F., Schneider, F., Petersen, K., Wallwiener, D., Kanz, L.*, et al.* (2011). Expression of the embryonic stem cell marker SOX2 in early-stage breast carcinoma. BMC Cancer *11*, 42.

Liang, Y., Robinson, D.F., Dennig, J., Suske, G., and Fahl, W.E. (1996). Transcriptional regulation of the SIS/PDGF-B gene in human osteosarcoma cells by the Sp family of transcription factors. J Biol Chem *271*, 11792-11797.

Licht, J.D., Grossel, M.J., Figge, J., and Hansen, U.M. (1990). Drosophila Kruppel protein is a transcriptional repressor. Nature *346*, 76-79.

Licht, J.D., Hanna-Rose, W., Reddy, J.C., English, M.A., Ro, M., Grossel, M., Shaknovich, R., and Hansen, U. (1994). Mapping and mutagenesis of the amino-terminal transcriptional repression domain of the Drosophila Kruppel protein. Mol Cell Biol *14*, 4057-4066.

Lindsley, E.B., and Grell, E.H. (1968). Genetic variation of Drosophila melanogaster. *627*.

Lledo, P.-M., Alonso, M., and Grubb, M.S. (2006). Adult neurogenesis and functional plasticity in neuronal circuits. Nat Rev Neurosci *7*, 179-193.

Loh, Y.H., Wu, Q., Chew, J.L., Vega, V.B., Zhang, W., Chen, X., Bourque, G., George, J., Leong, B., Liu, J.*, et al.* (2006). The Oct4 and Nanog transcription network regulates pluripotency in mouse embryonic stem cells. Nat Genet *38*, 431-440.

Lomant, A.J., and Fairbanks, G. (1976). Chemical probes of extended biological structures: synthesis and properties of the cleavable protein cross-linking reagent [35S]dithiobis(succinimidyl propionate). J Mol Biol *104*, 243-261.

Lowe, L.A., Supp, D.M., Sampath, K., Yokoyama, T., Wright, C.V.E., Potter, S.S., Overbeek, P., and Kuehn, M.R. (1996). Conserved left-right asymmetry of nodal expression and alterations in murine *situs inversus*. nature *381*, 4.

Lu, Y., Futtner, C., Rock, J.R., Xu, X., Whitworth, W., Hogan, B.L., and Onaitis, M.W. (2010). Evidence that SOX2 overexpression is oncogenic in the lung. PLoS One *5*, e11022.
Majello, B., De Luca, P., and Lania, L. (1997). Sp3 is a bifunctional transcription regulator with modular independent activation and repression domains. J Biol Chem *272*, 4021-4026.
Massague, J., and Chen, Y.G. (2000). Controlling TGF-beta signalling. Genes Dev *14*, 627-644.
Masui, S., Nakatake, Y., Toyooka, Y., Shimosato, D., Yagi, R., Takahashi, K., Okochi, H., Okuda, A., Matoba, R., Sharov, A.A.*, et al.* (2007). Pluripotency governed by Sox2 via regulation of Oct3/4 expression in mouse embryonic stem cells. Nat Cell Biol *9*, 625-635.
Matsushima, D., Heavner, W., and Pevny, L.H. (2011). Combinatorial regulation of optic cup progenitor cell fate by SOX2 and PAX6. Development *138*, 443-454.
Matys, V., Kel-Margoulis, O.V., Fricke, E., Liebich, I., Land, S., Barre-Dirrie, A., Reuter, I., Chekmenev, D., Krull, M., Hornischer, K.*, et al.* (2006). TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes. Nucleic Acids Res *34*, D108-110.
Mazet, F., and Shimeld, S.M. (2002). Gene duplication and divergence in the early evolution of vertebrates. Curr Opin Genet Dev *12*, 393-396.
McCulloch, E.A., and Till, J.E. (1960). The radiation sensitivity of normal mouse bone marrow cells, determined by quantitative marrow transplantation into irradiated mice. Radiat Res *13*, 115-125.
MEHTA, R., and SHAPIRO, A.D. (2008). Plasminogen activator inhibitor type 1 deficiency. Haemophilia *14*, 6.
Mertin, S., McDowall, S.G., and Harley, V.R. (1999). The DNA-binding specificity of SOX9 and other SOX proteins. Nucleic Acids Res *27*, 1359-1364.
Miao, Y., Laun, T., Zimmermann, P., and Zentgraf, U. (2004). Targets of the WRKY53 transcription factor and its role during leaf senescence in Arabidopsis. Plant Mol Biol *55*, 853-867.
Mitsui, K., Tokuzawa, Y., Itoh, H., Segawa, K., Murakami, M., Takahashi, K., Maruyama, M., Maeda, M., and Yamanaka, S. (2003). The homeoprotein Nanog is required for maintenance of pluripotency in mouse epiblast and ES cells. Cell *113*, 631-642.
Miyagi, S., Kato, H., and Okuda, A. (2009). Role of SoxB1 transcription factors in development. Cellular and Molecular Life Sciences *66*, 10.
Mizuseki, K., Kishi, M., Matsui, M., Nakanishi, S., and Sasai, Y. (1998). Xenopus Zic-related-1 and Sox-2, two factors induced by chordin, have distinct activities in the initiation of neural induction. Development *125*, 579-587.
Monsoro-Burq*, A.-H., and Douarin, N.L. (2000). Left-right asymmetry in BMP4 signalling pathway during
chick gastrulation. Mechanisms of Development *97*, 4.
Munoz-Sanjuan, I., and Brivanlou, A.H. (2002). Nerual induction, the default model and embryonic stem cells. Nature Reviews Neuroscience *3*, 271-280.
Muotri, A.R., Chu, V.T., Marchetto, M.C., Deng, W., Moran, J.V., and Gage, F.H. (2005). Somatic mosaicism in neuronal precursor cells mediated by L1 retrotransposition. Nature *435*, 903-910.
Murphy, F.V.t., Sweet, R.M., and Churchill, M.E. (1999). The structure of a chromosomal high mobility group protein-DNA complex reveals sequence-neutral mechanisms important for non-sequence-specific DNA recognition. EMBO J *18*, 6610-6618.
Navarro, P., Chambers, I., Karwacki-Neisius, V., Chureau, C., Morey, C., Rougeulle, C., and Avner, P. (2008). Molecular coupling of Xist regulation and pluripotency. Science *321*, 1693-1695.
Nichols, J., Zevnik, B., Anastassiadis, K., Niwa, H., Klewe-Nebenius, D., Chambers, I., Scholer, H., and Smith, A. (1998). Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4. Cell *95*, 379-391.
Nishiguchi, S., Wood, H., Kondoh, H., Lovell-Badge, R., and Episkopou, V. (1998). Sox1 directly regulates the gamma-crystallin genes and is essential for lens development in mice. Genes Dev *12*, 776-781.
Nishimoto, M., Fukushima, A., Okuda, A., and Muramatsu, M. (1999). The gene for the embryonic stem cell coactivator UTF1 carries a regulatory element which selectively interacts with a complex composed of Oct-3/4 and Sox-2. Mol Cell Biol *19*, 5453-5465.

Nowling, T., Bernadt, C., Johnson, L., Desler, M., and Rizzino, A. (2003). The co-activator p300 associates physically with and can mediate the action of the distal enhancer of the FGF-4 gene. J Biol Chem *278*, 13696-13705.

Nowling, T.K., Johnson, L.R., Wiebe, M.S., and Rizzino, A. (2000). Identification of the Transactivation Domain of the Transcription Factor Sox-2 and an Associated Co-activator. J Biol Chem *275*, 3810-3818.

Nutt, S.L., Heavey, B., Rolink, A.G., and Busslinger, M. (1999). Commitment to the B-lymphoid lineage depends on the transcription factor Pax5. Nature *401*, 556-562.

Ogryzko, V.V., Schiltz, R.L., Russanova, V., Howard, B.H., and Nakatani, Y. (1996). The transcriptional coactivators p300 and CBP are histone acetyltransferases. Cell *87*, 953-959.

Okuda, Y., Ogura, E., Kondoh, H., and Kamachi, Y. (2010). B1 SOX coordinate cell specification with patterning and morphogenesis in the early zebrafish embryo. PLoS Genet *6*, e1000936.

Okuda, Y., Yoda, H., Uchikawa, M., Furutani-Seiki, M., Takeda, H., Kondoh, H., and Kamachi, Y. (2006). Comparative genomic and expression analysis of group B1 sox genes in zebrafish indicates their diversification during vertebrate evolution. Dev Dyn *235*, 811-825.

Okumura-Nakanishi, S., Saito, M., Niwa, H., and Ishikawa, F. (2005). Oct-3/4 and Sox2 regulate Oct-3/4 gene in embryonic stem cells. J Biol Chem *280*, 5307-5317.

Onichtchouk, D., Chen, Y.G., Dosch, R., Gawantka, V., Delius, H., Massague, J., and Niehrs, C. (1999). Silencing of TGF-beta signalling by the pseudoreceptor BAMBI. nature *401*, 480-485.

Paroush, Z., Finley, R.L., Jr., Kidd, T., Wainwright, S.M., Ingham, P.W., Brent, R., and Ish-Horowicz, D. (1994). Groucho is required for Drosophila neurogenesis, segmentation, and sex determination and interacts directly with hairy-related bHLH proteins. Cell *79*, 805-815.

Perissi, V., Jepsen, K., Glass, C.K., and Rosenfeld, M.G. (2010). Deconstructing repression: evolving models of co-repressor action. Nat Rev Genet *11*, 109-123.

Pevny, L., and Placzek, M. (2005). SOX genes and neural progenitor identity. Curr Opin Neurobiol *15*, 7-13.

Pevny, L.H., Sockanathan, S., Placzek, M., and Lovell-Badge, R. (1998). A role for SOX1 in neural determination. Development *125*, 1967-1978.

Pfaffl, M.W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res *29*, e45.

Pickles, L.M., Roe, S.M., Hemingway, E.J., Stifani, S., and Pearl, L.H. (2002). Crystal structure of the C-terminal WD40 repeat domain of the human Groucho/TLE1 transcriptional corepressor. Structure *10*, 751-761.

Poulat, F., Girard, F., Chevron, M.P., Goze, C., Rebillard, X., Calas, B., Lamb, N., and Berta, P. (1995). Nuclear localization of the testis determining gene product SRY. J Cell Biol *128*, 737-748.

Pusch, C., Hustert, E., Pfeifer, D., Sudbeck, P., Kist, R., Roe, B., Wang, Z., Balling, R., Blin, N., and Scherer, G. (1998). The SOX10/Sox10 gene from human and mouse: sequence, expression, and transactivation by the encoded HMG domain transcription factor. Hum Genet *103*, 115-123.

Radivojac, P., Vacic, V., Haynes, C., Cocklin, R.R., Mohan, A., Heyen, J.W., Goebl, M.G., and Iakoucheva, L.M. (2010). Identification, analysis, and prediction of protein ubiquitination sites. Proteins *78*, 365-380.

Reis, L.M., Tyler, R.C., Schneider, A., Bardakjian, T., and Semina, E.V. (2010). Examination of SOX2 in variable ocular conditions identifies a recurrent deletion in microphthalmia and lack of mutations in other phenotypes. Mol Vis *16*, 768-773.

Ren, B., Chee, K.J., Kim, T.H., and Maniatis, T. (1999). PRDI-BF1/Blimp-1 repression is mediated by corepressors of the Groucho family of proteins. Genes Dev *13*, 125-137.

Rex, M., Orme, A., Uwanogho, D., Tointon, K., Wigmore, P.M., Sharpe, P.T., and Scotting, P.J. (1997). Dynamic expression of chicken Sox2 and Sox3 genes in ectoderm induced to form neural tissue. Dev Dyn *209*, 323-332.

Richardson, J., Cvekl, A., and Wistow, G. (1995). Pax-6 is essential for lens-specific expression of zeta-crystallin. Proc Natl Acad Sci U S A *92*, 4676-4680.

Rizzino, A. (2009). Sox2 and Oct-3/4: A Versatile Pair of Master Regulators that Orchestrate the Self-renewal and Pluripotency of Embryonic Stem Cells by Functioning as Molecular Rheostats. Wiley Interdiscip Rev Syst Biol Med *1*, 228-236.

Rizzoti, K., Brunelli, S., Carmignac, D., Thomas, P.Q., Robinson, I.C., and Lovell-Badge, R. (2004). SOX3 is required during the formation of the hypothalamo-pituitary axis. Nat Genet *36*, 247-255.

Rodda, D.J., Chew, J.L., Lim, L.H., Loh, Y.H., Wang, B., Ng, H.H., and Robson, P. (2005). Transcriptional regulation of nanog by OCT4 and SOX2. J Biol Chem *280*, 24731-24737.

Rogers, C.D., Harafuji, N., Archer, T., Cunningham, D.D., and Casey, E.S. (2009). Xenopus Sox3 activates sox2 and geminin and indirectly represses Xvent2 expression to induce neural progenitor formation at the expense of non-neural ectodermal derivatives. Mech Dev *126*, 42-55.

Rolink, A.G., Nutt, S.L., Melchers, F., and Busslinger, M. (1999). Long-term in vivo reconstitution of T-cell development by Pax5-deficient B-cell progenitors. Nature *401*, 603-606.

Roose, J., Molenaar, M., Peterson, J., Hurenkamp, J., Brantjes, H., Moerer, P., van de Wetering, M., Destree, O., and Clevers, H. (1998). The Xenopus Wnt effector XTcf-3 interacts with Groucho-related transcriptional repressors. Nature *395*, 608-612.

Ross, S., Best, J.L., Zon, L.I., and Gill, G. (2002). SUMO-1 modification represses Sp3 transcriptional activation and modulates its subnuclear localization. Mol Cell *10*, 831-842.

Ruppert, J.M., Kinzler, K.W., Wong, A.J., Bigner, S.H., Kao, F.T., Law, M.L., Seuanez, H.N., O'Brien, S.J., and Vogelstein, B. (1988). The GLI-Kruppel family of human genes. Mol Cell Biol *8*, 3104-3113.

Südbeck, P., Schmitz, M.L., Baeuerle, P.A., and Scherer, G. (1996). Sex reversal by loss of the C–terminal transactivation domain of human SOX9. Nature genetics *13*, 230-232.

Sasai, Y., Kageyama, R., Tagawa, Y., Shigemoto, R., and Nakanishi, S. (1992). Two mammalian helix-loop-helix factors structurally related to Drosophila hairy and Enhancer of split. Genes Dev *6*, 2620-2634.

Sasai, Y., Lu, B., Steinbeisser, H., Geissert, D., Gont, L.K., and De Robertis, E.M. (1994). Xenopus chordin: a novel dorsalizing factor activated by orgaizerspecific homeobox genes. cell *79*, 779-790.

Sekiya, T., and Zaret, K.S. (2007). Repression by Groucho/TLE/Grg proteins: genomic site recruitment generates compacted chromatin in vitro and impairs activator binding in vivo. Mol Cell *28*, 291-303.

Shamblott, M.J., Axelman, J., Wang, S., Bugg, E.M., Littlefield, J.W., Donovan, P.J., Blumenthal, P.D., Huggins, G.R., and Gearhart, J.D. (1998). Derivation of pluripotent stem cells from cultured human primordial germ cells. Proc Natl Acad Sci U S A *95*, 13726-13731.

Sharov, A.A., Masui, S., Sharova, L.V., Piao, Y., Aiba, K., Matoba, R., Xin, L., Niwa, H., and Ko, M.S. (2008). Identification of Pou5f1, Sox2, and Nanog downstream target genes with statistical confidence by applying a novel algorithm to time course microarray and genome-wide chromatin immunoprecipitation data. BMC Genomics *9*, 269.

Shi, W., Wang, H., Pan, G., Geng, Y., Guo, Y., and Pei, D. (2006). Regulation of the pluripotency marker Rex-1 by Nanog and Sox2. J Biol Chem *281*, 23319-23325.

Shi, Y., Seto, E., Chang, L.S., and Shenk, T. (1991). Transcriptional repression by YY1, a human GLI-Kruppel-related protein, and relief of repression by adenovirus E1A protein. Cell *67*, 377-388.

Shi, Y., Sun, G., Zhao, C., and Stewart, R. (2008). Neural stem cell self-renewal. Crit Rev Oncol Hematol *65*, 43-53.

Shih, Y.H., Kuo, C.L., Hirst, C.S., Dee, C.T., Liu, Y.R., Laghari, Z.A., and Scotting, P.J. (2010). SoxB1 transcription factors restrict organizer gene expression by repressing multiple events downstream of Wnt signalling. Development.

Shinohara, T., and Piatigorsky, J. (1976). Quantitation of delta-crystallin messenger RNA during lens induction in chick embryos. Proc Natl Acad Sci U S A *73*, 2808-2812.

Silvia Brunelli, E.S.C., Donald Bell, Richard Harland, and Robin Lovell-Badge (2003). Expression of SOX3 Throughout the Developing Central Nervous System Is Dependent on the Combined Action of Discrete, Evolutionarily Conserved Regulatory Elements. genesis *36*, 13.

Smith, W.C., and Harland, R.M. (1992). Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in Xenopus embryos. cell *70*, 829-840.

Snow, J.W., Kim, J., Currie, C.R., Xu, J., and Orkin, S.H. (2010). Sumoylation regulates interaction of FOG1 with C-terminal-binding protein (CTBP). J Biol Chem *285*, 28064-28075.

Song, H., Hasson, P., Paroush, Z., and Courey, A.J. (2004). Groucho oligomerization is required for repression in vivo. Mol Cell Biol *24*, 4341-4350.

Stifani, S., Blaumueller, C.M., Redhead, N.J., Hill, R.E., and Artavanis-Tsakonas, S. (1992). Human homologs of a Drosophila Enhancer of split gene product define a novel family of nuclear proteins. Nat Genet *2*, 119-127.

Stolt, C.C., Schmitt, S., Lommes, P., Sock, E., and Wegner, M. (2005). Impact of transcription factor Sox8 on oligodendrocyte specification in the mouse embryonic spinal cord. Developmental Biology *281*, 9.

Sudbeck, P., and Scherer, G. (1997). Two independent nuclear localization signals are present in the DNA-binding high-mobility group domains of SRY and SOX9. J Biol Chem *272*, 27848-27852.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663-676.

Tatham, M.H., Geoffroy, M.C., Shen, L., Plechanovova, A., Hattersley, N., Jaffray, E.G., Palvimo, J.J., and Hay, R.T. (2008). RNF4 is a poly-SUMO-specific E3 ubiquitin ligase required for arsenic-induced PML degradation. Nat Cell Biol *10*, 538-546.

Tetsuka, T., Uranishi, H., Imai, H., Ono, T., Sonta, S., Takahashi, N., Asamitsu, K., and Okamoto, T. (2000). Inhibition of nuclear factor-kappaB-mediated transcription by association with the amino-terminal enhancer of split, a Groucho-related protein lacking WD40 repeats. J Biol Chem *275*, 4383-4390.

Thirunavukkarasu, K., Mahajan, M., McLarren, K.W., Stifani, S., and Karsenty, G. (1998). Two domains unique to osteoblast-specific transcription factor Osf2/Cbfa1 contribute to its transactivation function and its inability to heterodimerize with Cbfbeta. Mol Cell Biol *18*, 4197-4208.

Thomson, J.A., Itskovitz-Eldor, J., Shapiro, S.S., Waknitz, M.A., Swiergiel, J.J., Marshall, V.S., and Jones, J.M. (1998). Embryonic stem cell lines derived from human blastocysts. Science *282*, 1145-1147.

Toma, J.G., El-Bizri, H., Barnabe-Heider, F., Aloyz, R., and Miller, F.D. (2000). Evidence that helix-loop-helix proteins collaborate with retinoblastoma tumor suppressor protein to regulate cortical neurogenesis. The Journal of Neuroscience *15*, 9.

Tsuruzoe, S., Ishihara, K., Uchimura, Y., Watanabe, S., Sekita, Y., Aoto, T., Saitoh, H., Yuasa, Y., Niwa, H., Kawasuji, M*., et al.* (2006). Inhibition of DNA binding of Sox2 by the SUMO conjugation. Biochem Biophys Res Commun *351*, 920-926.

Uchikawa, M., Kamachi, Y., and Kondoh, H. (1999). Two distinct subgroups of Group B Sox genes for transcriptional activators and repressors: their expression during embryonic organogenesis of the chicken. Mech Dev *84*, 103-120.

Urbánek, P., Wang, Z.-Q., Fetka, I., Wagner, E.F., and Busslinger, M. (1994). Complete block of early B cell differentiation and altered patterning of the posterior midbrain in mice lacking Pax5/BSAP. cell *79*, 901-912.

Uwanogho, D., Rex, M., Cartwright, E.J., Pearl, G., Healy, C., Scotting, P.J., and Sharpe, P.T. (1995). Embryonic expression of the chicken Sox2, Sox3 and Sox11 genes suggests an interactive role in neuronal development. Mech Dev *49*, 23-36.

Valerio Orlando, H.S.a.R.P. (1997). Analysis of Chromatin Structure byin VivoFormaldehyde Cross-Linking. Methods *11*, 10.

Valin, A., and Gill, G. (2007). Regulation of the dual-function transcription factor Sp3 by SUMO. Biochem Soc Trans *35*, 1393-1396.

Van Hoof, D., Munoz, J., Braam, S.R., Pinkse, M.W., Linding, R., Heck, A.J., Mummery, C.L., and Krijgsveld, J. (2009). Phosphorylation dynamics during early differentiation of human embryonic stem cells. Cell Stem Cell *5*, 214-226.

Wang, J., Rao, S., Chu, J., Shen, X., Levasseur, D.N., Theunissen, T.W., and Orkin, S.H. (2006a). A protein interaction network for pluripotency of embryonic stem cells. Nature *444*, 364-368.

Wang, T.W., Stromberg, G.P., Whitney, J.T., Brower, N.W., Klymkowsky, M.W., and Parent, J.M. (2006b). Sox3 expression identifies neural progenitors in persistent neonatal and adult mouse forebrain germinative zones. J Comp Neurol *497*, 88-100.

Wang, W., Wang, Y.G., Reginato, A.M., Glotzer, D.J., Fukai, N., Plotkina, S., Karsenty, G., and Olsen, B.R. (2004). Groucho homologue Grg5 interacts with the transcription factor Runx2-Cbfa1 and modulates its activity during postnatal growth in mice. Dev Biol *270*, 364-381.

Wegner, M., and Stolt, C.C. (2005). From stem cells to neurons and glia: a Soxist's view of neural development. Trends Neurosci *28*, 583-588.

Westerfield, M. (2000). The zebrafish book. A guide for the laboratory use of zebrafish (Danio reior). 4th edn (University of Oregon Press, Eugene).

White, J., and Dalton, S. (2005). Cell Cycle Control of Embryonic Stem Cells. Stem Cell Reviews *1*, 8.

Wilkinson, K.A., and Henley, J.M. (2010). Mechanisms, regulation and consequences of protein SUMOylation. Biochem J *428*, 133-145.

Wilson, M., and Koopman, P. (2002). Matching SOX: partner proteins and co-factors of the SOX family of transcriptional regulators. Curr Opin Genet Dev *12*, 441-446.

Wilson, P.A., and Hemmati-Brivanlou, A. (1995). Induction of epidermis and inhibition of neural fate by Bmp-4. Nature *376*, 331-333.

Wood, H.B., and Episkopou, V. (1999). Comparative expression of the mouse Sox1, Sox2 and Sox3 genes from pre-gastrulation to early somite stages. Mech Dev *86*, 197-201.

Yang, S.H., and Sharrocks, A.D. (2004). SUMO promotes HDAC-mediated transcriptional repression. Mol Cell *13*, 611-617.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., *et al.* (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917-1920.

Yu, X., Li, P., Roeder, R.G., and Wang, Z. (2001). Inhibition of androgen receptor-mediated transcription by amino-terminal enhancer of split. Mol Cell Biol *21*, 4614-4625.

Yuan, H., Corbi, N., Basilico, C., and Dailey, L. (1995). Developmental-specific activity of the FGF-4 enhancer requires the synergistic action of Sox2 and Oct-3. Genes & Development *9*, 2635-2645.

Zhang, C., Basta, T., Hernandez-Lagunas, L., Simpson, P., Stemple, D.L., Artinger, K.B., and Klymkowsky, M.W. (2004). Repression of nodal expression by maternal B1-type SOXs regulates germ layer formation in Xenopus and zebrafish. Dev Biol *273*, 23-37.

Zhang, C., Basta, T., Jensen, E.D., and Klymkowsky, M.W. (2003a). The beta-catenin/VegT-regulated early zygotic gene Xnr5 is a direct target of SOX3 regulation. Development *130*, 5609-5624.

Zhang, H., Levine, M., and Ashe, H.L. (2001). Brinker is a sequence-specific transcriptional repressor in the Drosophila embryo. Genes Dev *15*, 261-266.

Zhang, P., Jimenez, S.A., and Stokes, D.G. (2003b). Regulation of human COL9A1 gene expression. Activation of the proximal promoter region by SOX9. J Biol Chem *278*, 117-123.

Zhang, X., Chen, H.M., Jaramillo, E., Wang, L., and D'Mello, S.R. (2008). Histone deacetylase-related protein inhibits AES-mediated neuronal cell death by direct interaction. J Neurosci Res *86*, 2423-2431.

Zhang, X., Friedman, A., Heaney, S., Purcell, P., and Maas, R.L. (2002). Meis homeoproteins directly regulate Pax6 during vertebrate lens morphogenesis. Genes Dev *16*, 2097-2107.

Zhang, X., Yu, H., Yang, Y., Zhu, R., Bai, J., Peng, Z., He, Y., Chen, L., Chen, W., Fang, D., *et al.* (2010a). SOX2 in gastric carcinoma, but not Hath1, is related to patients' clinicopathological features and prognosis. J Gastrointest Surg *14*, 1220-1226.

Zhang, Y., Gao, S., and Wang, Z. (2010b). Structural and functional analysis of amino-terminal enhancer of split in androgen-receptor-driven transcription. Biochem J *427*, 499-511.

Zhao, M., Ren, C., Yang, H., Feng, X., Jiang, X., Zhu, B., Zhou, W., Wang, L., Zeng, Y., and Yao, K. (2007). Transcriptional profiling of human embryonic stem cells and embryoid bodies identifies HESRG, a novel stem cell gene. Biochem Biophys Res Commun *362*, 916-922.

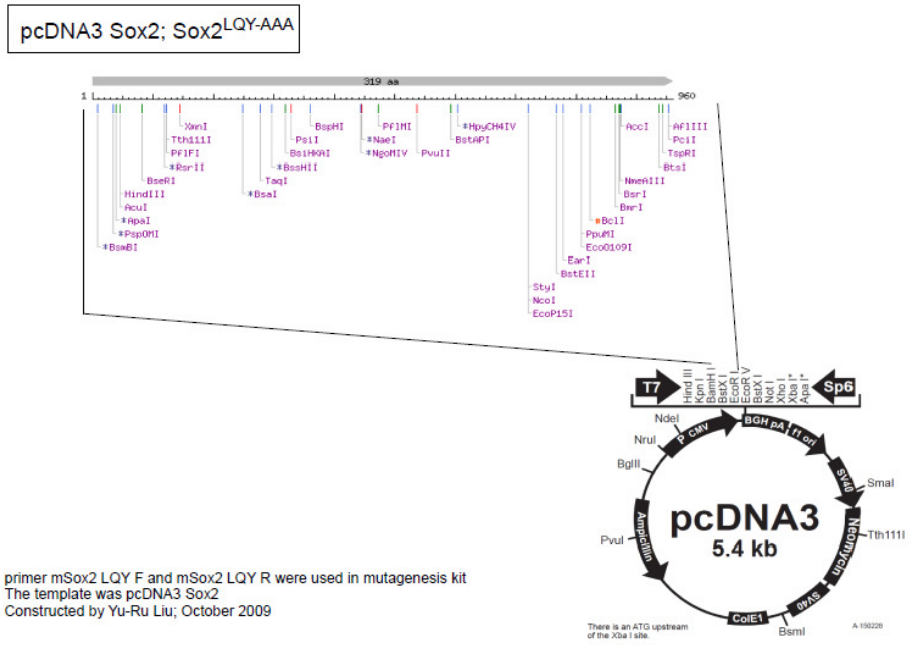
Zhao, S., Nichols, J., Smith, A.G., and Li, M. (2004). SoxB transcription factors specify neuroectodermal lineage choice in ES cells. Mol Cell Neurosci *27*, 332-342.

Zhou, Q., Chipperfield, H., Melton, D.A., and Wong, W.H. (2007). A gene regulatory network in mouse embryonic stem cells. Proc Natl Acad Sci U S A *104*, 16438-16443.

Zhu, C.C., Dyer, M.A., Uchikawa, M., Kondoh, H., Lagutin, O.V., and Oliver, G. (2002). Six3-mediated auto repression and eye development requires its interaction with members of the Groucho-related family of co-repressors. Development *129*, 2835-2849.

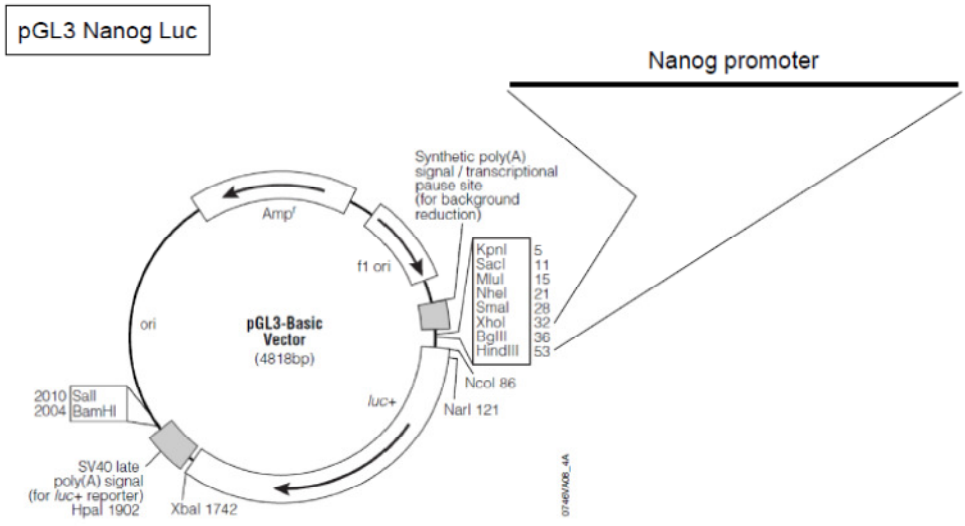
Chapter 9 Appendix

Appendix 1 pcDNA3-Sox2 / pcDNA3-Sox2$^{LQY/AAA}$



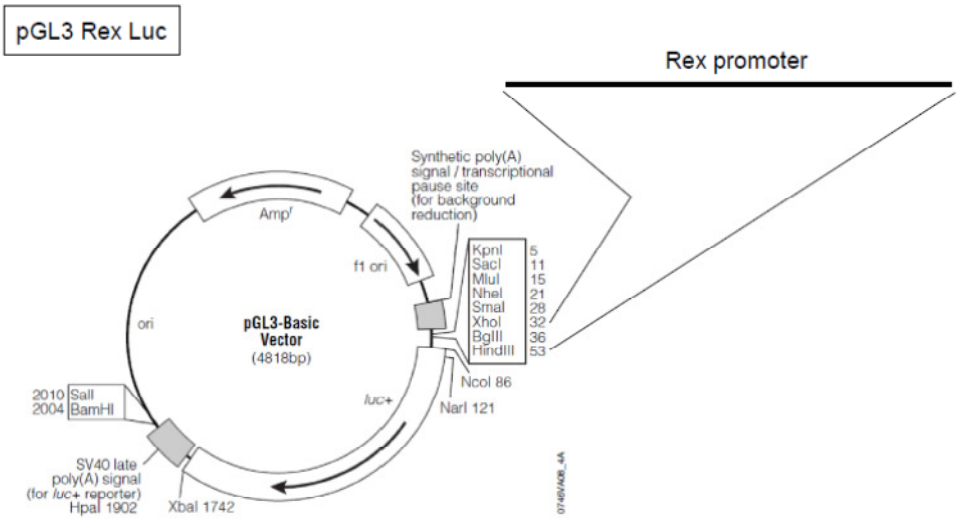
primer mSox2 LQY F and mSox2 LQY R were used in mutagenesis kit
The template was pcDNA3 Sox2
Constructed by Yu-Ru Liu; October 2009

Appendix 2 pGL3-Nanog



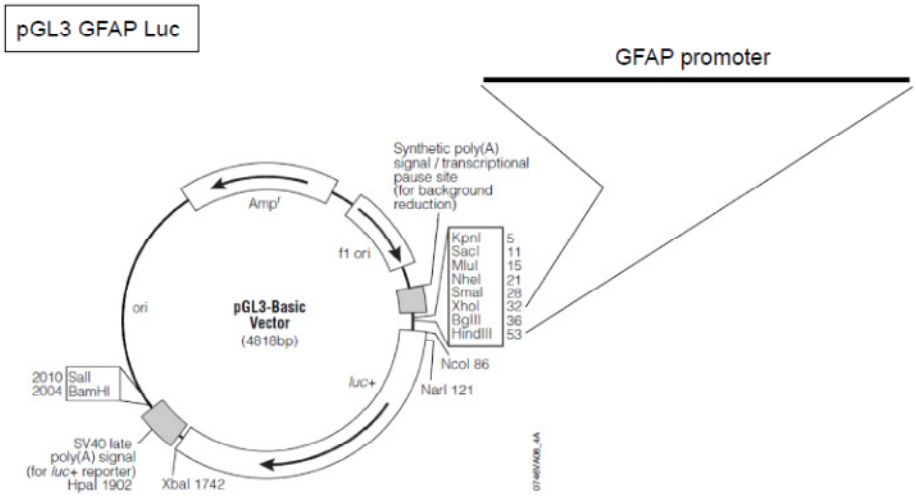
Primer were designed with XhoI and HindIII restriction sites according to *Rodda et. al. 2005*
Constructed by Yu-Ru Liu 2009, University of Nottingham

Appendix 3 pGL3-Rex



Primer were designed with XhoI and HindIII restriction sites according to *Shi et. al. 2006*
Constructed by Yu-Ru Liu 2009, University of Nottingham

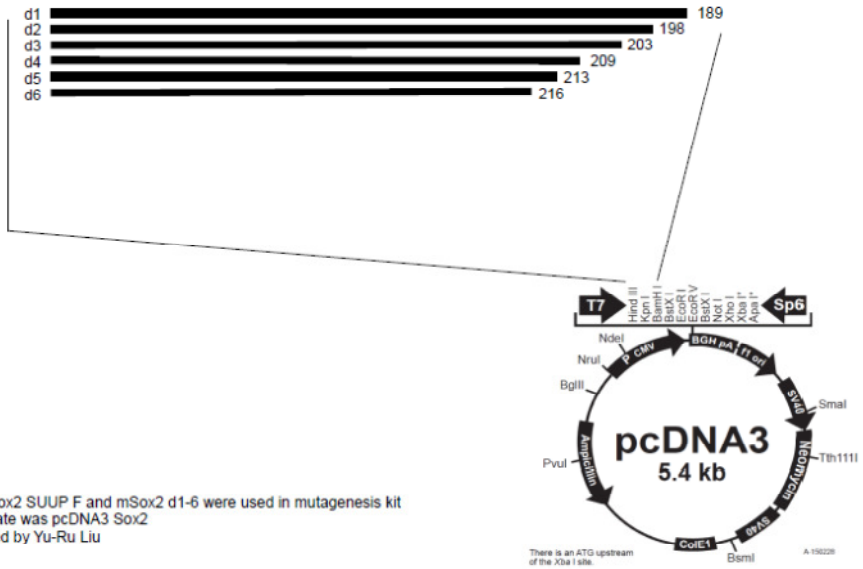
Appendix 4 pGL3-GFAP



Primer were designed with XhoI and HindIII restriction sites according to *Cavallaro et. al. 2008*
Constructed by Yu-Ru Liu 2009, University of Nottingham

Appendix 5 pcDNA3 Sox2 deletion 1-6

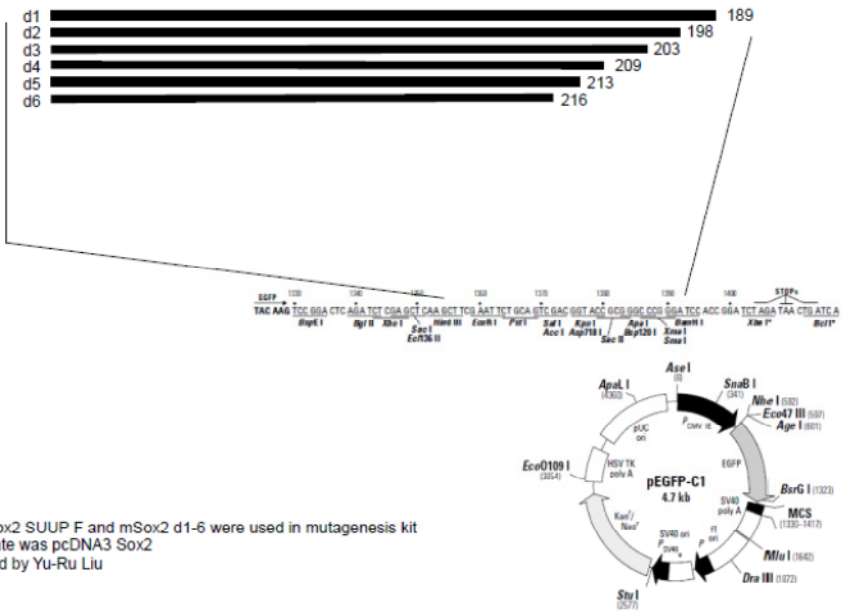
pcDNA3 Sox2 deletion1-6



primer mSox2 SUUP F and mSox2 d1-6 were used in mutagenesis kit
The template was pcDNA3 Sox2
Constructed by Yu-Ru Liu

Appendix 6 pEGFP-C1 Sox2 deletion 1-6

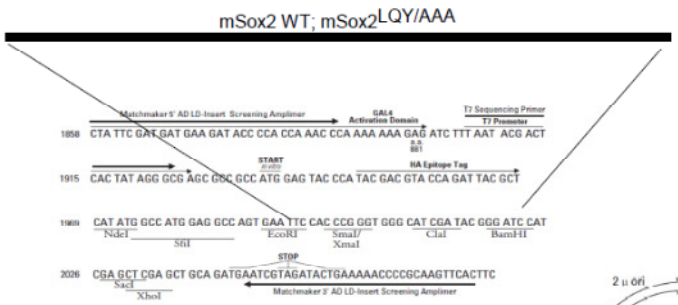
pEGFP-C1 Sox2 deletion1-6

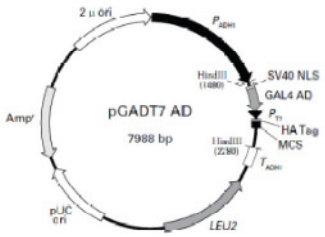


primer mSox2 SUUP F and mSox2 d1-6 were used in mutagenesis kit
The template was pcDNA3 Sox2
Constructed by Yu-Ru Liu

Appendix 7 pGADT7 Sox2; Sox2$^{LQY/AAA}$

pGADT7 Sox2; Sox2$^{LQY/AAA}$

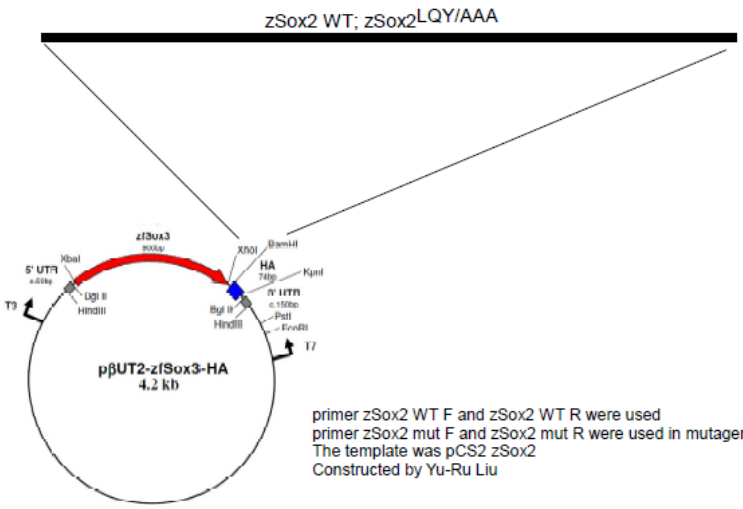




primer mSox2 WT F and mSox2 WT R were used in mutagenesis kit
The template was pcDNA3 Sox2
Constructed by Yu-Ru Liu

Appendix 8 pBUT2-zSox2; zSox2$^{LQY/AAA}$

pBUT2-zSox2; zSox2$^{LQY/AAA}$



primer zSox2 WT F and zSox2 WT R were used
primer zSox2 mut F and zSox2 mut R were used in mutagenesis kit
The template was pCS2 zSox2
Constructed by Yu-Ru Liu